Dynamic Nonlinear Econometric Models

Asymptotic Theory

Springer-Verlag Berlin Heidelberg GmbH

Benedikt M. Pötscher · Ingmar R. Prucha

Dynamic Nonlinear Econometric Models

Asymptotic Theory

Springer

Professor Benedikt M. Pötscher
University of Vienna
Department of Statistics, Operations Research
and Computer Science
Universitätsstraße 5
A-1010 Vienna, Austria

Professor Ingmar R. Prucha
University of Maryland at College Park
Department of Economics
Room 3105, Tydings Hall
College Park, Maryland 20742, USA

ISBN 978-3-642-08309-9

Cataloging-in-Publication Data applied for
Die Deutsche Bibliothek - CIP-Einheitsaufnahme
Pötscher, Benedikt M.: Dynamic nonlinear econometric models: asymptotic theory / Benedikt M. Pötscher; Ingmar R. Prucha.

97.07.00
ISBN 978-3-642-08309-9 ISBN 978-3-662-03486-6 (eBook)
DOI 10.1007/978-3-662-03486-6

Originally published by Springer-Verlag Berlin Heidelberg New York in 1997
Softcover reprint of the hardcover 1st edition 1997

Hardcoverdesign: Erich Kirchner, Heidelberg

SPIN 10576613 42/2202-5 4 3 2 1 0 - Printed on acid-free paper

BMP *Em memória de P.V.B.*

IRP *To the memory of W.P. and R.P.*

Preface

Many relationships in economics, and also in other fields, are both dynamic and nonlinear. A major advance in econometrics over the last fifteen years has been the development of a theory of estimation and inference for dynamic nonlinear models. This advance was accompanied by improvements in computer technology that facilitate the practical implementation of such estimation methods.

In two articles in *Econometric Reviews*, i.e., Pötscher and Prucha (1991a,b), we provided an expository discussion of the basic structure of the asymptotic theory of M-estimators in dynamic nonlinear models and a review of the literature up to the beginning of this decade. Among others, the class of M-estimators contains least mean distance estimators (including maximum likelihood estimators) and generalized method of moment estimators. The present book expands and revises the discussion in those articles. It is geared towards the professional econometrician or statistician.

Besides reviewing the literature we also presented in the above mentioned articles a number of then new results. One example is a consistency result for the case where the identifiable uniqueness condition fails. Another of these contributions was the introduction of the concept of L_p-approximability of a stochastic process by some (mixing) basis process. This approximation concept encompasses the concept of stochastic stability and the concept of near epoch dependence. Both of the latter approximation concepts had been used in the literature on the estimation of dynamic nonlinear econometric models, but the implications of the differences in these concepts were unclear at that time. Based on the encompassing approximation concept it was then possible to gain a better understanding of the differences and common grounds between the two approximation concepts. The encompassing framework made it, furthermore, possible to derive new results for the consistency and asymptotic normality of M-estimators of dynamic nonlinear models. Other contributions in the two review papers included improved consistency results for heteroskedasticity and autocorrelation robust variance-covariance matrix estimators in case of near epoch dependent data.

The theory presented in Pötscher and Prucha (1991a,b) and in the literature reviewed therein maintains catalogues of assumptions that are kept at a quite general and abstract level. As a consequence those catalogues of assumptions cover a wide range of applications. However this also means that for a specific estimation problem it is typically necessary to still expend considerable effort to verify if those assumptions are satisfied for the problem at hand. One of the features of this book is that we apply the general theory to an important more specific estimation problem. In particular, we analyze the full information maximum likelihood estimator of a dynamic nonlinear equation system. Apart from illustrating the applicability of the general theory, this analysis also provides new catalogues for the consistency and asymptotic normality of the nonlinear full information maximum likelihood estimator. We consider both the case of a correctly specified model and that of a misspecified model. An important question that seems natural when dealing with dynamic nonlinear systems is under which conditions the output process of such a system is L_p-approximable or near epoch dependent, given the input process has this property. In this book we provide several new results in this regard. Those results cover not only first order but also higher order dynamic systems.

As usual we would like to express our thanks to all who have contributed to the preparation of this monograph over the years. In particular we would like to thank Donald W.K. Andrews, Herman Bierens, Michael Binder, Immanuel M. Bomze, A. Ronald Gallant, David Pollard, and Halbert White for their helpful comments. Special thanks are due to Manfred Deistler and Harry H. Kelejian for their ongoing support and gracious advice on this as well as other research projects. We would also like to express our gratitude to Christian Cenker for expert advice on TeX-issues and to Birgit Ewald for helping with the preparation of the TeX-version of the manuscript. Finally we thank Michael Kumhof for his help in proof-reading the manuscript, and the editors of Springer-Verlag for their support and patience.

B. M. PÖTSCHER
I. R. PRUCHA

December 1996

Contents

1

INTRODUCTION

In the last twenty-five years considerable progress has been made in the theory of inference in nonlinear econometric models. A review of the literature up to the beginning of the 1980s is given in Burguete, Gallant and Souza (1982) and Amemiya (1983). For an account of related contributions in the statistics literature see, e.g., Humak (1983). The theory reviewed in these references assumes that the model is essentially static in nature and that the data generating process exhibits a certain degree of temporal homogeneity, e.g., some form of stationarity. Developments in recent years have focused on the extension of the theory to dynamic models, and in particular to situations where the data generating process can exhibit not only temporal dependence but also certain forms of temporal heterogeneity. As in the static case, this analysis also allows for model misspecification.

First progress towards a general theory for dynamic nonlinear econometric models was made by Bierens (1981, 1982a, 1984). His theory allows for temporal dependence in the data generating process and takes the dynamic structure of the model explicitly into account. Although Bierens does not take the data generating process to be stationary, his theory still assumes a certain degree of temporal homogeneity of the process. Bierens' analysis focuses mainly on least squares and robust estimation of a nonlinear regression model. Hansen (1982) considers generalized method of moments estimators in the context of dynamic models under the stronger homogeneity assumption that the data generating process is stationary. Many economic data exhibit, besides temporal dependence, also temporal heterogeneity. Therefore the asymptotic properties of estimators under such conditions have been analyzed by Domowitz and White (1982), White and Domowitz (1984), Bates and White (1985) and Domowitz (1985). Although the results of the latter papers can in principle be applied to processes generated by certain dynamic models, the results are not genuinely geared towards such models. In particular, in specifying the dependence properties of the data generating process, these papers do not explicitly take into account the dynamic structure of the model, as will be discussed in more detail below. Furthermore, as pointed out by Andrews (1987) and Pötscher and Prucha (1986a,b), some of the maintained assumptions in these papers are rather restrictive. Recent results by Wooldridge (1986), Gallant (1987a, Ch.7) and Gallant and White (1988) extend the theory of inference in dynamic nonlinear models to data generating processes that can exhibit both temporal

B. M. Pötscher et al., *Dynamic Nonlinear Econometric Models*

dependence and heterogeneity.

The estimators considered in the above cited literature are typically M-estimators, i.e., they are defined as the solution of a minimization (maximization) problem. M-estimators include, e.g., the least squares estimator, maximum likelihood estimators and generalized method of moments estimators.[1] A review of the above cited literature shows that the proofs employed to demonstrate the consistency and the asymptotic normality of M-estimators have a quite similar structure. The basic methods used in these proofs have their origin in numerous contributions in the statistics literature. More specifically, these methods date back to articles by Doob (1934), Cramér (1946), Wald (1949), and LeCam (1953), who consider the maximum likelihood estimator in the case of independent and identically distributed (i.i.d.) data and to the analysis of the least squares estimator by Jennrich (1969) and Malinvaud (1970), cf. also Hannan (1971), Robinson (1972), Wu (1981), and more recently Lai (1994). In his seminal article Huber (1967) analyzes the asymptotic properties of M-estimators in the case of i.i.d. data processes and allows for certain types of misspecification. Hoadley (1971) considers the asymptotic properties of the maximum likelihood estimator for independent and not necessarily identically distributed data processes. The statistics literature on the asymptotic properties of M-estimators for dependent data processes includes papers by Billingsley (1961), Silvey (1961), Roussas (1965), Crowder (1976), and Klimko and Nelson (1978), to mention a few. For surveys of the literature on maximum likelihood estimation and M-estimation see, e.g., Norden (1972, 1973) and Huber (1981). More recent contributions include Basawa and Koul (1988), Boente and Fraiman (1988), Dupačová and Wets (1988), Haberman (1989), Shapiro (1989), and Niemiro (1992).

The consideration of properties of M-estimators under misspecification arises naturally in the discussion of the behavior of test statistics under the alternative hypothesis, see, e.g., Silvey (1959). An explicit treatment of maximum likelihood estimators under misspecification is given in Foutz and Srivastava (1977, 1979), and White (1982), of Bayesian estimators in Berk (1966, 1970), of M-estimators in Huber (1967), and of nonlinear least squares estimators in Bunke and Schmidt (1980); see also White (1981), Humak (1983), and Schmidt (1987). In the time series and systems engineering literature this aspect has been analyzed under the heading of approximate systems modelling by Caines (1976, 1978), Caines and Ljung (1976), Ljung (1976a,b, 1978), Ljung and Caines (1979), Kabaila and Goodwin (1980), and Ploberger (1982a,b) for prediction error estimators, and by Pötscher (1987, 1991) for maximum likelihood estimators.

The econometrics literature cited above builds to a considerable extent

[1] The term M-estimator is used here in a more general meaning than, e.g., in the literature on robust estimation.

on the early contributions in the statistics literature, but like the research in time series analysis focuses more specifically on aspects necessary for a theory of inference for dynamic nonlinear models, which allows for temporal heterogeneity of the data generating process and for misspecification. For related work in the time series literature see, e.g., Priestley (1980, 1988), Tong (1983, 1990), Tjøstheim (1986, 1990), Kumar (1988), and Guégan (1994).[2]

Important ingredients in the typical proof of consistency and asymptotic normality of M-estimators are uniform laws of large numbers (ULLNs) and central limit theorems (CLTs). As a consequence, recent progress in the theory of inference in dynamic nonlinear models builds on progress in the derivation of ULLNs and CLTs. ULLNs for data generating processes, that are stationary or asymptotically stationary, are available in, e.g., LeCam (1953), Ranga Rao (1962), Jennrich (1969), Malinvaud (1970), Gallant (1977), Bierens (1981, 1982a, 1984, 1987), Amemiya (1985) and Pötscher and Prucha (1986a).[3] Hoadley's (1971) ULLN and its versions in Domowitz and White (1982) and White and Domowitz (1984) apply to temporally dependent and heterogeneous data generating processes. However, as pointed out by Andrews (1987) and Pötscher and Prucha (1986a,b) the maintained assumptions of this ULLN are restrictive (essentially requiring the random variables involved to be bounded). Andrews (1987) and Pötscher and Prucha (1986b, 1989, 1994b) introduce ULLNs for temporally dependent and heterogeneous processes under assumptions more appropriate for a theory of asymptotic inference in nonlinear econometric models. Furthermore, these papers specify the dependence structure in generic form in the sense that they assume the existence of laws of large numbers (LLNs) for certain "bracketing" functions of the data generating process, rather than to assume, e.g., a particular mixing property for the data generating process. As a consequence, these ULLNs are rather versatile tools that can be applied to processes with various dependence structures. Within the context of these ULLNs the dependence structure is relevant essentially only insofar as a LLN has to hold for the "bracketing" functions. The demonstration that a ULLN holds for a process with a particular dependence structure is therefore reduced to the demonstration that a LLN holds. For further results see also Andrews (1992), Newey (1991), and Pötscher and Prucha (1994a).

Since LLNs and CLTs are available for, e.g., α-mixing and ϕ-mixing processes it is tempting to simply postulate that the process of the endogenous

---

[2]In contrast to the econometrics literature on dynamic nonlinear models, this research concentrates more on the analysis of particular nonlinear models such as threshold models, bilinear models, etc.

[3]For independent random variables and in particular for i.i.d. random variables ULLNs have been established in the empirical process literature under much weaker conditions, see, e.g., Gänssler (1983) and Pollard (1984, 1990).

and exogenous variables is α-mixing or ϕ-mixing. This approach is used, e.g., in Domowitz and White (1982), White and Domowitz (1984), Bates and White (1985) and Domowitz (1985). However, as already alluded to above, in case the data are generated by a dynamic nonlinear model this assumption is not satisfactory. This is so, since then the endogenous variables typically depend on the infinite history of the exogenous variables and the disturbances. Now, even if the exogenous variables and the disturbances are α-mixing or ϕ-mixing, the endogenous variables need not inherit the same property, since α-mixing and ϕ-mixing are not necessarily preserved by transformations which involve the infinite past, see, e.g., Ibragimov and Linnik (1971), Chernick (1981), Andrews (1984), Athreya and Pantula (1986a,b), and Doukhan (1994).[4] Hence, an assumption that the process of endogenous and exogenous variables is α-mixing or ϕ-mixing does not seem to be adequate for a general treatment of dynamic models. In fact, as discussed next, it is possible to build a theory of asymptotic inference without these mixing conditions.

Intuitively one can expect LLNs and CLTs to hold for (functions of) the data generating process, provided both the dynamic system (generating the endogenous variables) and the process of exogenous variables and disturbances have a sufficiently "fading memory", even if the data generating process is not α-mixing or ϕ-mixing. This suggests that consistency and asymptotic normality results can also be obtained in such a context. The contributions of Bierens (1981, 1982a, 1984), Wooldridge (1986), Gallant (1987a, Ch.7), and Gallant and White (1988) can be viewed as a demonstration that this is indeed true under certain regularity conditions.[5] The basic approach taken in all these references is to show that LLNs and CLTs hold for (functions of) the data generating process by demonstrating that the (functions of the) data generating process can be approximated by processes with a sufficiently fading memory. However, these references differ in the approximation concept employed: Bierens' approximation concept leads to the definition of processes that are "stochastically stable w.r.t. an α-mixing [ϕ-mixing] base"; using this approximation concept he proves LLNs and CLTs for such processes. Wooldridge (1986), Gallant (1987a, Ch.7) and Gallant and White (1988) employ the concept of "near epoch dependence w.r.t. an α-mixing [ϕ-mixing] base", and then make use of a result by McLeish (1975a) that processes with such a dependence structure fall into the class of mixingales, for which LLNs and CLTs are available in

[4]Results that ensure that α-mixing or ϕ-mixing is preserved by such transformations only seem to be available under conditions that are unnecessarily restrictive for a satisfactory general theory of inference in dynamic nonlinear models.

[5]The contributions in this literature regarding limit theorems build on earlier contributions in probability theory, cf., e.g., Billingsley (1968, Section 21).

McLeish (1974, 1975a,b, 1977).[6]

Pötscher and Prucha (1991a,b) recently provided a survey and a critical discussion of these developments towards an asymptotic theory for M-estimators in dynamic nonlinear models. The first of these papers, Pötscher and Prucha (1991a), also introduced an encompassing framework for the approaches taken by Bierens on the one hand and by Gallant, White and Wooldridge on the other hand. These approaches differ mainly in the employed approximation concepts, i.e., stochastic stability versus near epoch dependence. The relationship between these approaches had not been explored in the literature before and a clear understanding of their relative merits was lacking. Apart from providing an understanding of the differences and common grounds between these rival approaches, the encompassing framework of Pötscher and Prucha (1991a,b) also resulted in LLNs and CLTs under simpler and weaker sets of assumptions than in, e.g., Gallant (1987a) and Gallant and White (1988). In turn this lead to catalogues of assumptions for consistency and asymptotic normality of M-estimators in dynamic nonlinear models that seem to be simpler than corresponding catalogues in, e.g., Gallant (1987a) and Gallant and White (1988). Pötscher and Prucha (1991b) also provided improved consistency results for heteroskedasticity and autocorrelation robust variance covariance matrix estimators in case of near epoch dependent data. These improvements over results available in the literature related to less restrictive assumptions on the feasible rate of increase of the truncation lag parameter. A further novel feature was that our results provided rates of convergence for the variance covariance matrix estimators, which are essential for the optimal selection of the truncation lag parameter.

The present book is an expanded and revised version of Pötscher and Prucha (1991a,b). The theory presented in those papers and in the literature reviewed therein is characterized by the fact that the maintained catalogues of assumptions are kept at a very general and abstract level in order to cover a wide range of applications. In this book we now also apply the general theory to an important concrete example. In particular, we illustrate the applicability of the general theory by deriving consistency and asymptotic normality results for the full information maximum likelihood estimator of a dynamic nonlinear simultaneous equation system. Apart form being illustrative, this analysis provides new catalogues of assumptions for consistency and asymptotic normality of the nonlinear full information maximum likelihood estimator. The contribution of the latter results over those available in the literature is that our results are based on a set of more specific low level assumptions on the model and the data generating process and not on a set of abstract high level assumptions. In

---

[6] The concept (though not the name) of near epoch dependence already appears in Billingsley (1968, Section 21); cf. also Blum and Rosenblatt (1957) and Ibragimov (1962).

the course of obtaining these new results it was also necessary to obtain conditions under which the output process of a dynamic nonlinear system is "weakly dependent" in the sense of being near epoch dependent, stochastically stable, or more generally satisfying the encompassing approximation concept of Pötscher and Prucha (1991a); see Theorems 6.10 - 6.12 in Chapter 6. This then facilitates, by applying existing LLNs and CLTs for such "weakly dependent" processes, the derivation of LLNs and CLTs for (functions of) the output process of the dynamic system.

In somewhat more detail, Theorem 6.10 generalizes results obtained in Bierens (1981, Ch. 5) to dynamic nonlinear systems, which are possibly only defined on a subset of Euclidean space rather than on the entire space – a situation which arises quite frequently in a nonlinear context. However, Theorem 6.10, as well as Theorem 6.11, are restricted to "contracting" systems. In contrast, Theorem 6.12 is substantially more general as it applies to "stable" systems and not only to "contracting" systems. This distinction becomes especially important when dealing with multivariate and/or higher order systems. In fact, Theorem 6.10, as well as Theorem 6.11, are not directly applicable to higher order systems. Another aspect of Theorems 6.10 - 6.12 and the accompanying discussion is that they help in clarifying some misconceptions in the recent literature pertaining to such results.

The book is organized as follows: Chapter 2 sets the stage for an asymptotic theory in nonlinear econometric models. Chapter 3 provides the basic modules for proving consistency of M-estimators. Chapter 4 is devoted to a discussion of possible extensions of the consistency results in Chapter 3. In Chapter 4 we also explore the limitations of the results of Chapter 3; in particular, we discuss the ramifications of misspecification for the assumption of identifiable uniqueness. A basic ingredient for consistency proofs are ULLNs for dependent and heterogeneous processes. Recent developments in this area are the subject of Chapter 5. Chapter 6 is devoted to a general theory of approximation concepts for stochastic processes, which incorporates the theory of stochastically stable processes as well as that of near epoch dependent processes. After a preliminary discussion in Section 6.1, these approximation concepts are introduced and basic properties of these concepts are explored in Section 6.2. The usefulness of these approximation concepts is illustrated by LLNs given in Section 6.3. (CLTs are discussed later in Chapter 10.) The behavior of the approximation concepts under nonlinear transformations is studied in Section 6.4. In this section we also provide sufficient conditions for a dynamic system such that the output process, i.e., the process of endogenous variables, satisfies the above mentioned approximation concepts. These results are crucial for deriving limit theorems for (functions of) the output process of a dynamic system. (They also play a prominent role in Chapter 14.) Chapter 7 presents catalogues of assumptions ensuring consistency of the least mean distance as well as of the generalized method of moments estimators. Chapter 8 discusses the basic structure of the asymptotic normality proof for M-estimators in nonlinear econometric models. Chapter 9 extends this discussion to nonstandard situations where, e.g., the objective function is not smooth or the nuisance parameter is infinite dimensional. Central limit theorems for dependent random variables, which form an important ingredient in the asymptotic normality proof, are discussed in Chapter 10. Chapter 11 presents asymptotic normality results for least mean distance and generalized method of moments estimators. A general discussion of heteroskedasticity and autocorrelation robust variance covariance matrix estimators is given in Chapter 12. The estimation of the variance covariance matrices for least mean distance and generalized method of moments estimators is treated in Chapter 13. Chapter 14 gives consistency and asymptotic normality results for the normal full information maximum likelihood estimator of a dynamic nonlinear simultaneous equation system. Concluding remarks are given in Chapter 15. All proofs are relegated to appendices.

2

MODELS, DATA GENERATING PROCESSES, AND ESTIMATORS

We start with a brief review of the basic structure of the classical estimation problem, which can be described as follows: The researcher observes a set of data assumed to be generated by a stochastic process. The probability law of this process is determined by a model. This "true" model is assumed to belong to a class of models where each model is indexed by a parameter. This parameter may either characterize the probability law of the stochastic process completely or only partially (e.g., it may only characterize the first and second moments). Apart from knowing that the true model belongs to the given model class, the value of the true parameter is not known. The parameter may be an element of a finite or infinite dimensional space. The estimation problem is then to infer the value of the true parameter (or of certain components of interest) on the basis of the observed data. Specific estimators are often derived from general principles such as the maximum likelihood principle or the method of moments. Given a particular estimator it is then of interest to analyze its performance.

A crucial assumption in the estimation problem described above is that the data have been generated by a member of the class of models under consideration. Since reality is likely to be more complex than any model, this assumption may be violated, i.e., the class of models under consideration may be misspecified. It is hence of interest to analyze the behavior of a given estimator also under the assumption of a data generating process that is not described by the given model class.

In the following we first formalize the above described general framework for M-estimators; for the sake of generality we allow for the presence of nuisance parameters. We then illustrate this framework in terms of the nonlinear least squares (NLS) estimator and the normal full information maximum likelihood (NFIML) estimator. Let $(\mathbf{z}_t : t \in \mathbf{N})$ be the data generating process defined on a probability space $(\Omega, \mathfrak{A}, P)$ with $\mathbf{z}_t$ taking its values in a non-empty measurable space $(Z, \mathfrak{Z})$, and let (B, ρ_B) and (T, ρ_T) be non-empty metric spaces. Typically, but not necessarily, $\mathbf{z}_t$ will have the interpretation of representing the vector of current and lagged endogenous and exogenous variables, and may also contain instrumental variables. The space Z will frequently be a Euclidean space or a subset

B. M. Pötscher et al., *Dynamic Nonlinear Econometric Models*

thereof. B is the space of parameters of interest and T will typically have the interpretation as the space of nuisance parameters. In semi-parametric or non-parametric applications B and T are typically not subsets of a Euclidean space. It is for this reason that we assume B and T only to be metric spaces rather than subsets of a Euclidean space.

Let $Q_n(z_1, ..., z_n, \tau, \beta)$ be a real valued function defined on $Z^n \times T \times B$ (where n denotes the sample size). Assume further that $Q_n(\mathbf{z}_1, ..., \mathbf{z}_n, \tau, \beta)$ is $\mathfrak{A}$-measurable for all $(\tau, \beta) \in T \times B$.[1] The M-estimators $\hat{\beta}_n$ corresponding to the objective function Q_n now satisfy for given estimators $\hat{\tau}_n$:

$$Q_n(\mathbf{z}_1, ..., \mathbf{z}_n, \hat{\tau}_n, \hat{\beta}_n) = \inf_{\beta \in B} Q_n(\mathbf{z}_1, ..., \mathbf{z}_n, \hat{\tau}_n, \beta), \tag{2.1}$$

i.e., they minimize the objective function over B.[2] Clearly, this setup also covers the case of misspecification. In the case of misspecification τ and β are parameters of the model class hypothesized by the researcher.

The above framework covers typical M-estimators for dynamic nonlinear equation systems. We illustrate this in the following examples in terms of the NLS estimator, and in terms of the NFIML and the nonlinear three stage least squares (N3SLS) estimator for a dynamic implicit nonlinear simultaneous equation system.

Example 1: Let $g_t : \mathbf{R} \times X \times A \to \mathbf{R}$ be Borel measurable functions, where $X \subseteq \mathbf{R}^{p_x}$ and $A \subseteq \mathbf{R}^{p_a}$ are Borel sets. Assume that the endogenous variables $\mathbf{y}_t$ are generated according to the following model

$$\mathbf{y}_t = g_t(\mathbf{y}_{t-1}, \mathbf{x}_t, \alpha_0) + \epsilon_t, \qquad t \in \mathbf{N}, \tag{2.2}$$

where $\mathbf{y}_0$ denotes some starting value. Assume further that the processes of the exogenous variables $(\mathbf{x}_t)$ and the disturbances (ϵ_t) are defined on $(\Omega, \mathfrak{A}, P)$ and take their values in X and $\mathbf{R}$, respectively; let $\alpha_0 \in A$ denote the true vector of regression parameters. The objective function of the NLS estimator is then given by

$$Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \beta) = n^{-1} \sum_{t=1}^{n} q_t(\mathbf{z}_t, \beta), \tag{2.3}$$

with

$$q_t(\mathbf{z}_t, \beta) = [\mathbf{y}_t - g_t(\mathbf{y}_{t-1}, \mathbf{x}_t, \alpha)]^2$$

[1] In the following we will sometimes simply write $Q_n(\tau, \beta)$ for $Q_n(z_1, \ldots, z_n, \tau, \beta)$ or $Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \tau, \beta)$ whenever the meaning of this expression is evident from the context.

[2] Technically speaking, for consistency and asymptotic normality results $\hat{\beta}_n$ has to satisfy (2.1) only asymptotically; cf., for example, Lemma 3.1 and Footnote 2 in Chapter 3. For a further relaxation of (2.1) see Section 4.4.

where $\mathbf{z}_t = (\mathbf{y}_t', \mathbf{y}_{t-1}', \mathbf{x}_t')'$ and $\beta = \alpha$.[3] □

Example 2: Let $f_t : Y \times Y \times X \times A \to E$ be Borel measurable functions, where $Y \subseteq \mathbf{R}^{p_y}$, $X \subseteq \mathbf{R}^{p_x}$, $A \subseteq \mathbf{R}^{p_\alpha}$, and $E \subseteq \mathbf{R}^{p_e}$ are Borel sets (and $p_y = p_e$). Let the process of the endogenous variables $(\mathbf{y}_t)$ be generated according to the following model

$$f_t(\mathbf{y}_t, \mathbf{y}_{t-1}, \mathbf{x}_t, \alpha_0) = \epsilon_t, \qquad t \in \mathbf{N}. \tag{2.4}$$

The processes of the exogenous variables $(\mathbf{x}_t)$ and the disturbances (ϵ_t) are defined on $(\Omega, \mathfrak{A}, P)$ and take their values in X and E, respectively; $\alpha_0 \in A$ denotes the true vector of system parameters. We assume that the model has a well-defined reduced form, i.e., for each $(y^\bullet, x, e, \alpha) \in Y \times X \times E \times A$ the equation $f_t(y, y^\bullet, x, \alpha) = e$ has a unique solution $y = g_t(y^\bullet, x, e, \alpha)$ where g_t is assumed to be measurable. Given an initial random variable $\mathbf{y}_0$ the process of endogenous variables $(\mathbf{y}_t)$ is then well-defined. We assume further that the vectors of disturbances ϵ_t are distributed i.i.d. normal with zero mean and variance covariance matrix Σ_0 and that the process (ϵ_t) is independent (jointly) of the process $(\mathbf{x}_t)$ and $\mathbf{y}_0$. To define the NFIML estimator properly we assume further that f_t is continuously differentiable w.r.t. y, that $\partial f_t / \partial y$ is nonsingular, and that Y is open in $\mathbf{R}^{p_y}$.[4]

The objective function of the NFIML estimator, i.e., the normal log-likelihood function conditional on the exogenous variables and on $\mathbf{y}_0$ is now (up to an additive constant and multiplied by $-1/n$) given by

$$Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \beta) = n^{-1} \sum_{t=1}^{n} q_t(\mathbf{z}_t, \beta), \tag{2.5}$$

with

$$q_t(\mathbf{z}_t, \beta) = -\ln |\det(\partial f_t / \partial y)| + (1/2) \ln \det(\Sigma) + (1/2) f_t' \Sigma^{-1} f_t,$$

where f_t and $\partial f_t / \partial y$ are evaluated at $(\mathbf{z}_t, \alpha)$, $\mathbf{z}_t = (\mathbf{y}_t', \mathbf{y}_{t-1}', \mathbf{x}_t')'$, and β is the vector composed of the elements of α and the diagonal and upper diagonal elements of Σ. Note that in (2.5) no nuisance parameter appears.

---

[3]As a general convention, in this book elements of Euclidean space are viewed as column vectors. However, for convenience of notation and where no confusion is possible, we will sometimes not distinguish between (v, w) and $(v', w')'$, where v and w are elements of Euclidean spaces, and where the prime denotes the transpose operation.

[4]We adopt the following convention: Let ϕ be a $s \times 1$ vector of real valued functions defined on (an open subset of) $\mathbf{R}^p$, let $v = (v_1, \ldots, v_p)' \in \mathbf{R}^p$, then $\partial \phi / \partial v = (\partial \phi / \partial v_1, \ldots, \partial \phi / \partial v_p)$ is the $s \times p$ matrix of first order partial derivatives.

The objective function of the N3SLS estimator is given by

$$Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \hat{\tau}_n, \beta) = \left[n^{-1} \sum_{t=1}^{n} q_t(\mathbf{z}_t, \beta)\right]' D(\hat{\tau}_n) \left[n^{-1} \sum_{t=1}^{n} q_t(\mathbf{z}_t, \beta)\right], \tag{2.6}$$

with

$$\begin{aligned} q_t(\mathbf{z}_t, \beta) &= f_t(\mathbf{y}_t, \mathbf{y}_{t-1}, \mathbf{x}_t, \alpha) \otimes \mathbf{a}_t, \\ D(\hat{\tau}_n) &= \left[\hat{\Sigma}_n \otimes n^{-1} \sum_{t=1}^{n} \mathbf{a}_t \mathbf{a}_t'\right]^{-1}, \end{aligned}$$

where $\mathbf{a}_t$ is the instrument vector, $\hat{\Sigma}_n$ is the two stage least squares estimator for Σ, $\hat{\tau}_n$ is the vector of diagonal and upper diagonal elements of $[\hat{\Sigma}_n \otimes n^{-1} \sum_{t=1}^{n} \mathbf{a}_t \mathbf{a}_t']^{-1}$ which is assumed to exist, $\mathbf{z}_t = (\mathbf{y}_t', \mathbf{y}_{t-1}', \mathbf{x}_t', \mathbf{a}_t')'$, and $\beta = \alpha$. □

As discussed above, the formal framework also allows for model misspecification. For example, the NLS, NFIML and N3SLS estimators defined by, respectively, objective functions (2.3), (2.5) and (2.6) remain well-defined estimators (in the sense of remaining well-defined statistics) even if the observed data are not generated by, respectively, model (2.2) or (2.4); in this case the parameter β appearing in the objective function may no longer characterize any aspect of the probability law of the data process.

We note that the general framework not only applies to dynamic nonlinear models of the form (2.2) or (2.4) as considered in the above examples, but more generally also to models containing higher order lags. For example, the framework also covers dynamic implicit nonlinear simultaneous equation systems of the form

$$f_t(\mathbf{y}_t, \ldots, \mathbf{y}_{t-l_t}, \mathbf{x}_t, \ldots, \mathbf{x}_{t-\bar{l}_t}, \alpha_o) = \epsilon_t, \qquad t \in \mathbf{N}.$$

For $l_t = l$ and $\bar{l}_t = \bar{l}$ this includes the case of a fixed lag length and for $l_t = t + l$ and $\bar{l}_t = t + \bar{l}$ $(l, \bar{l} \geq 0)$ the case of an increasing lag length, cf. also Chapter 5.

Estimators corresponding to objective functions of, respectively, the form

$$Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \hat{\tau}_n, \beta) = n^{-1} \sum_{t=1}^{n} q_t(\mathbf{z}_t, \hat{\tau}_n, \beta)$$

and

$$Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \hat{\tau}_n, \beta) = \vartheta_n \left(n^{-1} \sum_{t=1}^{n} q_t(\mathbf{z}_t, \hat{\tau}_n, \beta), \hat{\tau}_n, \beta\right),$$

where ϑ_n is some "distance" function, are usually referred to in the literature as least mean distance estimators and generalized method of moments estimators. For generalized method of moments estimators ϑ_n will often be a quadratic form in the moment vector $n^{-1}\sum_{t=1}^{n} q_t(\mathbf{z}_t, \hat{\tau}_n, \beta)$. The NLS and NFIML estimators defined in the above examples are special cases of least mean distance estimators, whereas the N3SLS estimator is a special case of a generalized method of moments estimator. We also note that in general quasi maximum likelihood estimators can be viewed as least mean distance estimators; in this case $q_t(\mathbf{z}_t, \beta)$ corresponds to the negative of the conditional log-likelihood in period t.

3

BASIC STRUCTURE OF THE CLASSICAL CONSISTENCY PROOF

In this chapter we describe the structure of the consistency proof for M-estimators in nonlinear econometric models as it has evolved from Jennrich (1969) and Malinvaud (1970). We shall refer to this proof as the classical consistency proof. The basic ideas date back to Doob (1934), Wald (1949) and LeCam (1953). The consistency proofs in the articles on asymptotic inference in nonlinear econometric models listed in Chapter 1 all share this common structure.

Let $Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \hat{\tau}_n, \beta)$ denote some criterion function as defined in Chapter 2 and let $\hat{\beta}_n$ be a corresponding M-estimator. The classical consistency proof deduces the limiting behavior of $\hat{\beta}_n$ from the limiting behavior of Q_n. In many cases there exist nonstochastic real valued functions $\bar{Q}_n$ defined on $T \times B$ such that the difference between $Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \hat{\tau}_n, \beta)$ and $\bar{Q}_n(\bar{\tau}_n, \beta)$ converges to zero (in a sense specified later) as the sample size tends to infinity; here $\bar{\tau}_n$ is nonstochastic and typically a population analogue of $\hat{\tau}_n$. The limiting behavior of $\hat{\beta}_n$ can then be analyzed by relating it to the limiting behavior of the minimizers $\bar{\beta}_n$ of $\bar{Q}_n(\bar{\tau}_n, \beta)$. In many cases $\bar{Q}_n$ will be taken as $EQ_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \bar{\tau}_n, \beta)$, or as $\lim_{n \to \infty} EQ_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \tau, \beta)$ evaluated at $\bar{\tau}_n$, or as $\lim_{n \to \infty} EQ_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \bar{\tau}_n, \beta)$, given the respective limits exist. Of course, in the last case the functions $\bar{Q}_n$ are independent of n.

In essence, the structure of the classical consistency proof has two ingredients. In the case of a correctly specified model with a compact parameter space and where $\bar{Q}_n \equiv \bar{Q}$ is continuous (and no nuisance parameter is present) the two ingredients are convergence of Q_n to $\bar{Q}$ uniformly over the parameter space and the existence of a unique minimizer of $\bar{Q}$ at the true parameter value. In more general cases essentially the same approach is employed subject to some modifications: As before, it is assumed that the difference between Q_n and $\bar{Q}_n$ converges to zero uniformly over the parameter space. However, the assumption that the true parameter value is a unique minimizer of $\bar{Q}$ (together with continuity and compactness) is replaced by an assumption that ensures that the minimizers of $\bar{Q}_n$ are essentially unique, as well as that the functions $\bar{Q}_n$ do not become too flat

B. M. Pötscher et al., *Dynamic Nonlinear Econometric Models*

at the minimizers. More formally, it is typically assumed that a sequence of minimizers $\bar{\beta}_n$ of $\bar{Q}_n$ has the following property (where the existence of the minimizers is implicitly assumed):

Definition 3.1. [1] *For a given sequence of functions $\bar{Q}_n : T \times B \to \mathbf{R}$ and a given (nonstochastic) sequence $\bar{\tau}_n \in T$ the sequence of minimizers $\bar{\beta}_n$ of $\bar{Q}_n(\bar{\tau}_n, \beta)$ is called identifiably unique, if for every $\epsilon > 0$:*

$$\liminf_{n\to\infty} \left[\inf_{\{\beta \in B : \rho_B(\beta, \bar{\beta}_n) \geq \epsilon\}} \bar{Q}_n(\bar{\tau}_n, \beta) - \bar{Q}_n(\bar{\tau}_n, \bar{\beta}_n) \right] > 0. \quad (3.1)$$

The above definition was introduced in White (1980) and Domowitz and White (1982). For the important special case where $\bar{Q}_n(\bar{\tau}_n, .)$ does not depend on the sample size, i.e., $\bar{Q}_n(\bar{\tau}_n, \beta) \equiv \bar{R}(\beta)$, identifiable uniqueness of $\bar{\beta}_n \equiv \bar{\beta}$ implies that $\bar{\beta}$ is the unique minimizer of $\bar{R}$. If furthermore B is compact and $\bar{R}$ is continuous (or, more generally, is lower semi-continuous), then identifiable uniqueness of $\bar{\beta}$ is equivalent to uniqueness of $\bar{\beta}$. However, if B is not compact then the existence of a unique minimizer $\bar{\beta}$ of $\bar{R}$ does not necessarily imply that $\bar{\beta}$ is identifiably unique, even if $\bar{R}$ is continuous. This is readily confirmed by considering the following example: $\bar{R}(\beta) = \beta^2/(1+\beta^2)^2$ with $B = \mathbf{R}$ and $\bar{\beta} = 0$. It is also easy to find examples where B is compact, $\bar{R}$ has a unique minimizer, but $\bar{R}$ is not lower semi-continuous, and where the unique minimizer is not identifiably unique. In general, identifiable uniqueness of some sequence of minimizers implies that the diameter of the set of minimizers of $\bar{Q}_n(\bar{\tau}_n, .)$ goes to zero as n tends to infinity, i.e., the difference between respective minimizers becomes negligible as the sample size increases. (Hence in this sense $\bar{\beta}_n$ is essentially unique.) Furthermore, if the sequence $\bar{\beta}_n$ is identifiably unique, then any other sequence of minimizers of $\bar{Q}_n(\bar{\tau}_n, .)$ is also identifiably unique.

In the context of maximum likelihood estimation of a correctly specified model, given stationarity, the identifiable uniqueness condition typically boils down to the condition that the true parameter is identified. The adequacy of the assumption of identifiable uniqueness in the case of misspecification will be discussed later, see Chapter 4.

The following lemma gives basic conditions for the convergence behavior of M-estimators in case where

$$R_n(\omega, \beta) = Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \hat{\tau}_n, \beta)$$

and

$$\bar{R}_n(\beta) = \bar{Q}_n(\bar{\tau}_n, \beta).$$

[1] Of course, this definition also includes the case where no nuisance parameter is present. Also, we adopt the convention that the infimum over the empty set is plus infinity.

Lemma 3.1. *Let $R_n : \Omega \times B \to \mathbf{R}$ and $\bar{R}_n : B \to \mathbf{R}$ be two sequences of functions such that a.s. [i.p.]*

$$\sup_B \left| R_n(\omega, \beta) - \bar{R}_n(\beta) \right| \to 0 \text{ as } n \to \infty. \tag{3.2}$$

Let $\bar{\beta}_n$ be an identifiably unique sequence of minimizers of $\bar{R}_n(\beta)$, then for any sequence $\hat{\beta}_n$ such that eventually[2]

$$R_n(\omega, \hat{\beta}_n) = \inf_B R_n(\omega, \beta) \tag{3.3}$$

holds, we have $\rho_B(\hat{\beta}_n, \bar{\beta}_n) \to 0$ a.s. [i.p.] as $n \to \infty$.[3]

In general, the above lemma does not imply that $\hat{\beta}_n$ converges, except if $\bar{\beta}_n$ converges. However, in case $\bar{\beta}_n \equiv \bar{\beta}$ the lemma implies that $\hat{\beta}_n \to \bar{\beta}$ as $n \to \infty$. A situation where $\bar{\beta}_n \equiv \bar{\beta}$ is (under certain regularity conditions) the case of maximum likelihood estimation of a correctly specified model where $\bar{\beta}$ is the true parameter and $\bar{R}_n$ is the expected log-likelihood. Another situation where $\bar{\beta}_n \equiv \bar{\beta}$ is the case where $\bar{R}_n \equiv \bar{R}$ is independent of n. (In the context of Example 2 this is, e.g., the case for the expected log-likelihood even if the model is misspecified if $f_t \equiv f$ and the data generating process $(\mathbf{y}_t', \mathbf{x}_t')'$ is strictly stationary.) Yet another situation where $\bar{\beta}_n \equiv \bar{\beta}$ is the case of least squares estimation of a nonlinear regression model where the response function corresponding to the value $\bar{\beta}$ of the regression parameter is the conditional mean of the dependent variable given the regressors.

Lemma 3.1 implies that (apart from a verification of the identifiable uniqueness condition) the classical consistency proof reduces to the verification that the objective function $R_n(\omega, \beta) = Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \hat{\tau}_n, \beta)$ satisfies the uniform convergence condition (3.2). In the case where no nuisance parameter is present and the objective function is of the form $R_n(\omega, \beta) = n^{-1} \sum_{t=1}^n q_t(\mathbf{z}_t, \beta)$ and $\bar{R}_n(\beta) = n^{-1} \sum_{t=1}^n E q_t(\mathbf{z}_t, \beta)$ the uniform convergence condition (3.2) boils down to a ULLN for $q_t(\mathbf{z}_t, \beta)$. For the case

---

[2]This means that there is an $\Omega_0 \in \mathfrak{A}$, $P(\Omega_0) = 1$, such that for any $\omega \in \Omega_0$ there is an $N(\omega)$ such that condition (3.3) holds for all $n \geq N(\omega)$. For the i.p. version of the lemma this could be relaxed to the requirement that (3.3) holds on sets $\Omega_n \in \mathfrak{A}$ with $P(\Omega_n) \to 1$. Note that both formulations amount to a slight relaxation of condition (2.1).

[3]The abbreviations "a.s." and "i.p." stand for, respectively, "almost surely" and "in probability" with respect to the probability measure P. For the convergence in probability version of this lemma it is implicitly assumed that the variables in any sequence, that is claimed to converge in probability, are measurable. This convention is adopted throughout the book. Sufficient conditions for $\hat{\beta}_n$ to be measurable are discussed in Lemma 3.4. A simple sufficient condition for the measurability of suprema (or infima) like in (3.2) is that the functions over which the supremum (or infimum) is taken are continuous on B for all $\omega \in \Omega$ and that B is separable, e.g., that B is compact and metrizable.

where a nuisance parameter is present, the estimator $\hat{\tau}_n$ will typically satisfy $\rho_T(\hat{\tau}_n, \bar{\tau}_n) \to 0$ as $n \to \infty$ (where $\bar{\tau}_n$ is typically some population analogue of $\hat{\tau}_n$). The following lemma reduces the verification of (3.2) to the verification of a similar condition that no longer involves the estimator $\hat{\tau}_n$ (regardless whether or not R_n takes the form of an average).

Lemma 3.2. *Given sequences of functions* $Q_n : Z^n \times T \times B \to \mathbf{R}$, $\bar{Q}_n : T \times B \to \mathbf{R}$, $\hat{\tau}_n : \Omega \to T$ *and a nonstochastic sequence* $\bar{\tau}_n \in T$, *let* $R_n(\omega, \beta) = Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \hat{\tau}_n, \beta)$ *and* $\bar{R}_n(\beta) = \bar{Q}_n(\bar{\tau}_n, \beta)$. *Let* $\{\bar{Q}_n : n \in \mathbf{N}\}$ *be uniformly equicontinuous on* $T \times B$ *(which is satisfied, e.g., if* $\{\bar{Q}_n : n \in \mathbf{N}\}$ *is equicontinuous on* $T \times B$ *and* T *and* B *are compact).*

(a) If $\rho_T(\hat{\tau}_n, \bar{\tau}_n) \to 0$ *a.s. [i.p.] as* $n \to \infty$ *and if a.s. [i.p.]*

$$\sup_{T \times B} \left| Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \tau, \beta) - \bar{Q}_n(\tau, \beta) \right| \to 0 \text{ as } n \to \infty \qquad (3.4)$$

then a.s. [i.p.]

$$\sup_{B} \left| R_n(\omega, \beta) - \bar{R}_n(\beta) \right| \to 0 \text{ as } n \to \infty. \qquad (3.2)$$

(b) The family $\{\bar{R}_n : n \in \mathbf{N}\}$ *is uniformly equicontinuous on* B.[4,5]

If B is compact and $\rho_T(\bar{\tau}_n, \bar{\tau}) \to 0$ as $n \to \infty$ for some $\bar{\tau} \in T$, then the assumption of uniform equicontinuity of $\{\bar{Q}_n : n \in \mathbf{N}\}$ in Lemma 3.2 can be weakened to equicontinuity. In case the objective function is of the form $R_n(\omega, \beta) = n^{-1} \sum_{t=1}^{n} q_t(\mathbf{z}_t, \hat{\tau}_n, \beta)$ and $\bar{R}_n(\beta) = n^{-1} \sum_{t=1}^{n} Eq_t(\mathbf{z}_t, \bar{\tau}_n, \beta)$ the uniform convergence condition (3.4) boils down to a ULLN for the functions $q_t(\mathbf{z}_t, \tau, \beta)$. That is, Lemma 3.2 implies that in this case the uniform

---

[4] A family of functions $\{f_n : n \in \mathbf{N}\}$ with $f_n : M_1 \to M_2$ where (M_1, ρ_1) and (M_2, ρ_2) are metric spaces is called equicontinuous at $x^\bullet \in M_1$ if for any $\epsilon > 0$ there exists a $\delta = \delta(x^\bullet, \epsilon) > 0$ such that $\sup_n \rho_2(f_n(x), f_n(x^\bullet)) < \epsilon$ for all $x \in M_1$ with $\rho_1(x, x^\bullet) < \delta$. The family is called equicontinuous on the subset M of M_1 if it is equicontinuous at all $x^\bullet \in M$. (This of course implies, but is in general not equivalent to equicontinuity of $\{f_n \mid M : n \in \mathbf{N}\}$ on M, where $f_n \mid M$ denotes the restriction of f_n to M. Both concepts coincide if M is an open subset of M_1 and if, in particular, $M = M_1$.) Furthermore, the family is called uniformly equicontinuous on the subset M of M_1 if it is equicontinuous on M and if $\delta(x^\bullet, \epsilon)$ for $x^\bullet \in M$ can be chosen independently of $x^\bullet$, i.e., $\delta = \delta(\epsilon)$. We note that in part of the mathematical literature the term equicontinuity is used to describe what we defined as uniform equicontinuity. Of course, if M_1 is a compact metric space then equicontinuity on M_1 is equivalent to uniform equicontinuity on M_1.

[5] Let (M_1, ρ_1) and (M_2, ρ_2) be two metric spaces. We endow the space $M_1 \times M_2$ with the metric $\rho((x, y), (x^\bullet, y^\bullet)) = \rho_1(x, x^\bullet) + \rho_2(y, y^\bullet)$, or if more convenient, with the metric $\bar{\rho}((x, y), (x^\bullet, y^\bullet)) = [\rho_1^2(x, x^\bullet) + \rho_2^2(y, y^\bullet)]^{1/2}$. Since $c_1 \rho \leq \bar{\rho} \leq c_2 \rho$ for suitable $0 < c_1 \leq c_2 < \infty$ the two metrics are of course metrically equivalent. In abuse of notation we write $\rho = \rho_1 + \rho_2$ and $\bar{\rho} = (\rho_1^2 + \rho_2^2)^{1/2}$.

convergence condition (3.2) for the objective function can again be verified from a ULLN. We note that equicontinuity of $\{\bar{Q}_n : n \in \mathbf{N}\}$ is typically obtained as a byproduct of a ULLN, cf. Chapter 5.

Lemma 3.2 is also useful in other contexts where Q_n does not necessarily have the interpretation of an objective function defining an estimator: Consider the case where $R_n(\omega) = Q_n(\mathbf{z}_1, \dots, \mathbf{z}_n, \hat{\tau}_n)$ and $\bar{R}_n = \bar{Q}_n(\bar{\tau}_n)$ do not depend on β. Then the lemma gives sufficient conditions under which $R_n(\omega) - \bar{R}_n \to 0$ a.s. [i.p.] as $n \to \infty$. Such situations, where we consider the convergence behavior of $Q_n(\mathbf{z}_1, \dots, \mathbf{z}_n, \hat{\tau}_n)$, arise, e.g., in the consistency proof of variance covariance matrix estimators or in the analysis of one-step M-estimators. In these cases $\hat{\tau}_n$ will typically be an estimator of non-nuisance and nuisance parameters.

In the following lemma we give sufficient conditions for (3.4) given the objective function is of the form

$$Q_n(\mathbf{z}_1, \dots, \mathbf{z}_n, \hat{\tau}_n, \beta) = \vartheta_n \left(S_n(\omega, \hat{\tau}_n, \beta), \hat{\tau}_n, \beta \right)$$

and

$$\bar{Q}_n(\tau, \beta) = \vartheta_n \left(\bar{S}_n(\tau, \beta), \tau, \beta \right).$$

Lemma 3.3. [6,7] *Let $S_n : \Omega \times T \times B \to C$, and $\bar{S}_n : T \times B \to C$ be sequences of functions where C is a subset of a Euclidean space, and let $\vartheta_n : C \times T \times B \to \mathbf{R}$. Let (I) $\{\vartheta_n : n \in \mathbf{N}\}$ be uniformly equicontinuous on $C \times T \times B$, or let (II) $\{\vartheta_n : n \in \mathbf{N}\}$ be equicontinuous on the subset $K \times T \times B$ of $C \times T \times B$, where T, B, and K are compact, and $\bar{S}_n(\tau, \beta) \in K$ for all $(\tau, \beta) \in T \times B$ and $n \in \mathbf{N}$.*

(a) If a.s. [i.p.]

$$\sup_{T \times B} \left| S_n(\omega, \tau, \beta) - \bar{S}_n(\tau, \beta) \right| \to 0 \text{ as } n \to \infty, \tag{3.5}$$

then a.s. [i.p.]

$$\sup_{T \times B} \left| \vartheta_n \left(S_n(\omega, \tau, \beta), \tau, \beta \right) - \vartheta_n \left(\bar{S}_n(\tau, \beta), \tau, \beta \right) \right| \to 0 \text{ as } n \to \infty. \tag{3.6}$$

(b) If $\{\bar{S}_n : n \in \mathbf{N}\}$ is equicontinuous [uniformly equicontinuous] on $T \times B$, then $\{\vartheta_n(\bar{S}_n(.), .) : n \in \mathbf{N}\}$ is equicontinuous [uniformly equicontinuous] on $T \times B$.

---

[6]Here and in the following $|x|$ denotes the Euclidean norm of a vector x.

[7]We remind the reader that the notion of equicontinuity of a family of functions on a subset as used in Lemma 3.3 is stronger than the notion of equicontinuity of the family of the restrictions of these functions to that subset; cf. Footnote 4.

Objective functions of the form considered in Lemma 3.3 arise in particular in the case of generalized method of moments estimators. For these estimators we typically have

$$S_n(\omega, \hat{\tau}_n, \beta) = n^{-1} \sum_{t=1}^{n} q_t(\mathbf{z}_t, \hat{\tau}_n, \beta)$$

and

$$\bar{S}_n(\bar{\tau}_n, \beta) = n^{-1} \sum_{t=1}^{n} E q_t(\mathbf{z}_t, \bar{\tau}_n, \beta),$$

where in the correctly specified case $\bar{S}_n$ is a vector of moments which equal zero at the true parameter value. (The functions ϑ_n have the interpretation of measuring the distance of respective moments from zero.) In this case condition (3.5) again boils down to a ULLN for the functions $q_t(\mathbf{z}_t, \tau, \beta)$.

In comparing conditions (I) and (II) we note that the latter seems to be more relevant in applications: In many cases ϑ_n will not depend on n and will be continuous, hence the equicontinuity condition in (II) is then automatically satisfied. Furthermore, if (3.5) is inferred from a ULLN for the functions $q_t(\mathbf{z}_t, \tau, \beta)$, as described above, then typical catalogues of assumptions for such ULLNs assume compactness of $T \times B$ and a dominance condition, which implies that all functions $\bar{S}_n$ take their values in a common compact set. Consequently, the remaining conditions in (II) will be satisfied automatically in such circumstances. Similarly as in the case of Lemma 3.2, we note that Lemma 3.3 is also useful in other contexts where $\vartheta_n(S_n, ., .)$ does not have the interpretation of an objective function defining an estimator.

Lemmata 3.1, 3.2 and 3.3 collect basic tools for the consistency proof of M-estimators. These tools have been used extensively in the literature. Explicitly stated versions of these lemmata in the econometrics literature can be found, e.g., in Amemiya (1973, 1983), White (1980), Bierens (1981), Domowitz and White (1982), Bates (1984), Bates and White (1985), Gallant and White (1988), and Pakes and Pollard (1989). Proofs of Lemmata 3.1 - 3.3 are given in Appendix A. Instead of proving Lemmata 3.2 and 3.3 directly we deduce both lemmata from Lemma A1 in Appendix A; this lemma may also be useful for generating alternative versions of Lemmata 3.2 and 3.3.

In Lemma 3.1 we have assumed that minimizers $\hat{\beta}_n$ of the objective function exist eventually. The following lemma gives sufficient conditions for the existence and measurability of minimizers. (Of course, the a.s. part of Lemma 3.1 holds even if the $\hat{\beta}_n$ are not measurable.) With $\mathfrak{B}(M)$ we denote the Borel σ-field on a metrizable space M and with $\mathfrak{Z}^n$ the product σ-field on Z^n.

Lemma 3.4. *Assume B is compact. Let $Q_n(z_1, \ldots, z_n, \tau, .)$ be a continuous function on B for each $(z_1, \ldots, z_n, \tau) \in Z^n \times T$ and let $Q_n(., \beta)$ be a $\mathfrak{Z}^n \otimes$*

$\mathfrak{B}(T)$-$\mathfrak{B}(\mathbf{R})$*-measurable function on* $Z^n \times T$ *for each* $\beta \in B$*. Then there exists a* $\mathfrak{Z}^n \otimes \mathfrak{B}(T)$-$\mathfrak{B}(B)$*-measurable function* $\beta_n = \beta_n(z_1, \ldots, z_n, \tau)$ *such that for all* $(z_1, \ldots, z_n, \tau) \in Z^n \times T$

$$Q_n(z_1, \ldots, z_n, \tau, \beta_n) = \inf_{\beta \in B} Q_n(z_1, \ldots, z_n, \tau, \beta)$$

holds.

Of course, the lemma ensures also the $\mathfrak{A}$-$\mathfrak{B}(B)$-measurability of $\hat{\beta}_n = \beta_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \hat{\tau}_n)$ given $\hat{\tau}_n$ is $\mathfrak{A}$-$\mathfrak{B}(T)$-measurable. For the case where B is a compact subset of Euclidean space the above lemma is given in Schmetterer (1966, Ch. 5), Lemma 3.3, and in Jennrich (1969), Lemma 2.[8] We note that measurable minimizers exist also in more general contexts than the one described in the above lemma; see, e.g., Pfanzagl (1969), Corollary 1.10, and Brown and Purves (1973).

The above discussion documents the importance of ULLNs in the context of consistency proofs. A detailed study of ULLNs will be given in Chapter 5. In the next chapter we comment on possible extensions and limitations of the theory presented above.

---

[8]The proof in Jennrich (1969) seems to be in error.

4

FURTHER COMMENTS ON CONSISTENCY PROOFS

In this chapter we comment further on the classical consistency proof, discuss possible extensions and point out limitations of the theory.[1] In the following let $R_n(\omega,\beta) = Q_n(\mathbf{z}_1,\ldots,\mathbf{z}_n,\hat{\tau}_n,\beta)$ be some objective function, let $\hat{\beta}_n$ be a corresponding M-estimator satisfying (2.1), and let $\bar{R}_n(\beta) = \bar{Q}_n(\bar{\tau}_n,\beta)$.

4.1 Transforming the Objective Function

Sometimes it is useful to replace R_n with a modified function R_n^* that defines the same estimator $\hat{\beta}_n$, but is easier to handle in the consistency proof. Of course, standard practices like transforming R_n monotonously or concentrating the objective function w.r.t. certain parameters serve this purpose. In particular, to avoid unnecessary moment requirements, it is often useful to consider modifications of the objective functions of the form

$$R_n^*(\omega,\beta) = R_n(\omega,\beta) - R_n(\omega,\beta_*)$$

where $\beta_* \in B$ is fixed; see, e.g., Huber (1967). We illustrate this within the context of least squares estimation of a nonlinear regression model. In particular, consider the following special case of the nonlinear regression model defined in Example 1 of Chapter 2 with $\beta_0 = \alpha_0$ and $\beta = \alpha$:

$$\mathbf{y}_t = g(\mathbf{x}_t,\beta_0) + \epsilon_t,$$

where $(\mathbf{x}_t)$ and (ϵ_t) are independent of each other, $(\mathbf{x}_t',\epsilon_t)'$ is strictly stationary and ergodic, $E(\epsilon_t) = 0$ and $0 < E[g(\mathbf{x}_t,\beta_0) - g(\mathbf{x}_t,\beta)]^2 < \infty$ for all $\beta \neq \beta_0$. Here the objective function of the least squares estimator is

$$\begin{aligned} R_n(\omega,\beta) &= n^{-1}\sum_{t=1}^{n}[\mathbf{y}_t - g(\mathbf{x}_t,\beta)]^2 \\ &= n^{-1}\sum_{t=1}^{n}\epsilon_t^2 + 2n^{-1}\sum_{t=1}^{n}\epsilon_t\,[g(\mathbf{x}_t,\beta_0) - g(\mathbf{x}_t,\beta)] \end{aligned}$$

[1]For the sake of simplicity we discuss mainly a.s. results in this chapter.

B. M. Pötscher et al., *Dynamic Nonlinear Econometric Models*

$$+ n^{-1} \sum_{t=1}^{n} [g(\mathbf{x}_t, \beta_0) - g(\mathbf{x}_t, \beta)]^2 .$$

The obvious choice for $\bar{R}_n(\beta)$ is $E[R_n(\omega, \beta)]$, which would require the additional assumption of a finite second moment for ϵ_t. Suppose now that for the analysis (but, of course, not for the computation) of the least squares estimator we replace $R_n(\omega, \beta)$ with

$$\begin{aligned} R_n^*(\omega, \beta) &= R_n(\omega, \beta) - R_n(\omega, \beta_0) \\ &= 2n^{-1} \sum_{t=1}^{n} \epsilon_t \, [g(\mathbf{x}_t, \beta_0) - g(\mathbf{x}_t, \beta)] \\ &\quad + n^{-1} \sum_{t=1}^{n} [g(\mathbf{x}_t, \beta_0) - g(\mathbf{x}_t, \beta)]^2 . \end{aligned}$$

Then the least squares estimator also minimizes this modified objective function, which does not involve powers of ϵ_t higher than the first. Thus by modifying the objective function it is possible to avoid the assumption of a finite second moment for ϵ_t in the analysis of the least squares estimator.

4.2 Weakening the Uniform Convergence Assumption

As is easily seen, a basic sufficient condition for $\rho_B(\hat{\beta}_n, \bar{\beta}_n) \rightarrow 0$ a.s. as $n \rightarrow \infty$ is that for each $\epsilon > 0$

$$\liminf_{n \to \infty} \left[\inf_{\{\beta \in B : \rho_B(\beta, \bar{\beta}_n) \geq \epsilon\}} R_n(\omega, \beta) - R_n(\omega, \bar{\beta}_n) \right] > 0 \text{ a.s.}, \tag{4.1}$$

i.e., outside the open balls of radius ϵ centered at $\bar{\beta}_n$ the functions R_n are larger than at $\bar{\beta}_n$, uniformly for large n, cf. Perlman (1972). The proof of Lemma 3.1 proceeds by deducing condition (4.1) from the identifiable uniqueness condition, i.e., from the condition that for each $\epsilon > 0$

$$\liminf_{n \to \infty} \left[\inf_{\{\beta \in B : \rho_B(\beta, \bar{\beta}_n) \geq \epsilon\}} \bar{R}_n(\beta) - \bar{R}_n(\bar{\beta}_n) \right] > 0, \tag{4.2}$$

and from the assumption that $R_n(\omega, \beta) - \bar{R}_n(\beta)$ converges uniformly to zero a.s.; in fact, under uniform convergence (4.1) and (4.2) are equivalent. The important distinguishing feature between (4.1) and (4.2) is that the latter is nonstochastic and hence is easier to handle.

Inspection of the proof of Lemma 3.1 shows that the assumption of uniform convergence is only sufficient but not necessary for deducing (4.1) from (4.2). In fact, instead of uniform convergence it is sufficient to assume that

$$\limsup_{n\to\infty} \sup_{B} \left[\bar{R}_n(\beta) - R_n(\omega,\beta)\right] \leq 0 \text{ a.s.,} \tag{4.3}$$

and

$$\limsup_{n\to\infty} \left[R_n(\omega,\bar{\beta}_n) - \bar{R}_n(\bar{\beta}_n)\right] \leq 0 \text{ a.s.} \tag{4.4}$$

Of course, (4.3) and (4.4) taken together are equivalent to:

$$\sup_{B} \left[\bar{R}_n(\beta) - R_n(\omega,\beta)\right] \to 0 \text{ a.s. as } n \to \infty, \tag{4.5}$$

and

$$R_n(\omega,\bar{\beta}_n) - \bar{R}_n(\bar{\beta}_n) \to 0 \text{ a.s. as } n \to \infty. \tag{4.6}$$

Condition (4.5) is a "one-sided" uniform convergence condition. Similarly as the uniform convergence condition (3.2) can typically be implied from ULLNs, as discussed in Chapter 3, the "one-sided" uniform convergence condition (4.5) can frequently be deduced from "one-sided" ULLNs. In the important special case where $\bar{\beta}_n \equiv \bar{\beta}$ condition (4.6) reduces to pointwise convergence at $\bar{\beta}$, and hence typically to a LLN. However, in the general case $\bar{\beta}_n$ could potentially be anywhere in the parameter space. Hence, given our ignorance about the position of $\bar{\beta}_n$ in the parameter space, to imply (4.6) in the general case it seems that we have to assume

$$\sup_{B} \left[R_n(\omega,\beta) - \bar{R}_n(\beta)\right] \to 0 \text{ a.s. as } n \to \infty, \tag{4.7}$$

which together with (4.5) implies (4.6). However, the two "one-sided" uniform convergence conditions (4.5) and (4.7) together are equivalent to uniform convergence of $R_n(\omega,\beta) - \bar{R}_n(\beta)$ to zero. Hence a weakening of the uniform convergence condition (3.2) as suggested in (4.5) and (4.6) only seems practical in the important special case where $\bar{\beta}_n \equiv \beta$. For further discussions see Pfanzagl (1969), Perlman (1972), Zaman (1985), Pollard (1991), and Liese and Vajda (1995).

In cases where uniform convergence of $R_n(\omega,\beta) - \bar{R}_n(\beta)$ fails, another strategy is based on the observation that

$$\liminf_{n\to\infty} \left\{ \inf_{\{\beta\in B:\rho_B(\beta,\bar{\beta}_n)\geq\varepsilon\}} \left[\left(R_n(\omega,\beta) - R_n(\omega,\bar{\beta}_n)\right)/b_n(\beta)\right] \right\} > 0 \text{ a.s.} \tag{4.8}$$

implies (4.1), and hence $\rho_B(\hat{\beta}_n,\bar{\beta}_n) \to 0$ a.s. as $n \to \infty$, given the functions b_n satisfy $\inf_n \inf_B b_n(\beta) > 0$. Let

$$R_n^o(\omega,\beta) = \left[R_n(\omega,\beta) - R_n(\omega,\bar{\beta}_n)\right]/b_n(\beta)$$

and

$$\bar{R}_n^o(\beta) = \left[\bar{R}_n(\beta) - \bar{R}_n(\bar{\beta}_n)\right]/b_n(\beta).$$

We may now attempt to verify, analogously as in the above discussion, (4.8) from an identifiable uniqueness condition applied to the rescaled functions $\bar{R}_n^o(\beta)$ and uniform (or one-sided uniform) convergence of the rescaled functions $R_n^o(\omega,\beta) - \bar{R}_n^o(\beta)$. The idea is that it may be easier to obtain uniform convergence after rescaling. Possible choices for the scaling functions are

$$b_n(\beta) = \left|\bar{R}_n(\beta)\right|$$

or

$$b_n(\beta) = 1 + \left|\bar{R}_n(\beta)\right|;$$

see Huber (1967), Perlman (1972), and Zaman (1991). Note that while $\bar{\beta}_n$ also minimizes $\bar{R}_n^o(\beta)$ it is typically not the case that $\hat{\beta}_n$ minimizes $R_n^o(\omega,\beta)$. This idea of rescaling the objective function has been used in the context of nonlinear regression models with nonstochastic regressors by Wu (1981) and Läuter (1987), see also Zaman (1989).

4.3 Uniform Convergence and Compactness

In case the parameter space is not compact the assumption of uniform convergence in Lemma 3.1 (or even the assumption of one-sided uniform convergence) may be difficult to meet. In this case a frequently used strategy is to demonstrate by ad hoc arguments that there exists a compact subset of the parameter space such that over the complement of the compact subset the objective function is eventually "large".[2] Typically this will be achieved by constructing the compact subset such that it contains the true parameter, or more generally the $\bar{\beta}_n$'s, and such that uniformly over the complement of the compact subset eventually $R_n(\omega,\beta) - R_n(\omega,\bar{\beta}_n)$ exceeds some positive constant (possibly depending on ω). As a consequence we can then restrict the analysis to the compact subset, since eventually any minimizer of the objective function over the entire parameter space must fall into the compact subset.

We illustrate this approach within the context of least squares estimation of a simple nonlinear regression model. Consider again the following special case of the nonlinear regression model defined in Example 1 of Chapter 2 with $\beta_0 = \alpha_0$ and $\beta = \alpha$:

$$\mathbf{y}_t = g(\mathbf{x}_t, \beta_0) + \epsilon_t,$$

where $(\mathbf{x}_t)$ and (ϵ_t) are independent of each other, $(\mathbf{x}_t', \epsilon_t)'$ is strictly stationary and ergodic, and where $E(\epsilon_t) = 0$ and $\sigma^2 = E(\epsilon_t^2) < \infty$.[3] Furthermore let the parameter space B be $\mathbf{R}$, and let $g(x,\beta)$ be measurable in the

[2] In principle the compact subset may depend on ω.

[3] The assumption of a finite second moment of ϵ_t is made for simplicity only. It could be avoided along the lines of the discussion in Section 4.1.

first argument and continuously differentiable in the second argument. If we assume, e.g., that

$$E\left[\inf_B [\partial g(\mathbf{x}_t,\beta)/\partial\beta]^2\right] > 0,^4 \tag{4.9}$$

we can construct a compact subset $B_* = [\beta_0 - M, \beta_0 + M]$ of the parameter space, such that the least squares estimator $\hat{\beta}_n$ falls eventually into B_*, as follows:

The objective function of the least squares estimator is given by

$$R_n(\omega,\beta) = n^{-1}\sum_{t=1}^{n}[\mathbf{y}_t - g(\mathbf{x}_t,\beta)]^2 .$$

We now have to find a constant $M > 0$ such that

$$\liminf_{n\to\infty}\left[\inf_{|\beta-\beta_0|>M}[R_n(\omega,\beta) - R_n(\omega,\beta_0)]\right] > 0 \text{ a.s.}$$

Let

$$c(x) = \inf_B [\partial g(x,\beta)/\partial\beta]^2$$

and $C = Ec(\mathbf{x}_t)$, then using the mean value theorem observe that

$$\begin{aligned}
&\liminf_{n\to\infty}\left\{\inf_{|\beta-\beta_0|>M}\left[n^{-1}\sum_{t=1}^{n}[g(\mathbf{x}_t,\beta) - g(\mathbf{x}_t,\beta_0)]^2\right]^{1/2}\right.\\
&\quad\left. -\ 2\left[n^{-1}\sum_{t=1}^{n}\epsilon_t^2\right]^{1/2}\right\}\\
\geq\ &\liminf_{n\to\infty}\left\{M\left[n^{-1}\sum_{t=1}^{n}c(\mathbf{x}_t)\right]^{1/2} - 2\left[n^{-1}\sum_{t=1}^{n}\epsilon_t^2\right]^{1/2}\right\}\\
=\ &MC^{1/2} - 2\sigma > 0 \text{ a.s.}
\end{aligned}$$

if $M > 2\sigma/C^{1/2}$, noting that $C > 0$ by assumption. Again using the mean value theorem and the Cauchy-Schwarz inequality we get for $M > 2\sigma/C^{1/2}$:

$$\begin{aligned}
&\liminf_{n\to\infty}\left\{\inf_{|\beta-\beta_0|>M}[R_n(\omega,\beta) - R_n(\omega,\beta_0)]\right\}\\
\geq\ &\liminf_{n\to\infty}\left\{\inf_{|\beta-\beta_0|>M}\left\{\left[n^{-1}\sum_{t=1}^{n}[g(\mathbf{x}_t,\beta) - g(\mathbf{x}_t,\beta_0)]^2\right]^{1/2}\right.\right.
\end{aligned}$$

[4] In the linear case this corresponds to the assumption $E(\mathbf{x}_t^2) > 0$.

$$\left[\left[n^{-1}\sum_{t=1}^{n}[g(\mathbf{x}_t,\beta)-g(\mathbf{x}_t,\beta_0)]^2\right]^{1/2} - 2\left[n^{-1}\sum_{t=1}^{n}\epsilon_t^2\right]^{1/2}\right]\Bigg\}\Bigg\}$$
$$\geq \; MC^{1/2}(MC^{1/2}-2\sigma) > 0.$$

Consistency of the least squares estimator $\hat{\beta}_n$ for β_0 can now be inferred from Lemma 3.1 applied to the compact subset $B_* = [\beta_0 - M, \beta_0 + M]$: First, observe that the identifiable uniqueness condition for the restricted problem becomes

$$\inf_{\epsilon\leq|\beta-\beta_0|\leq M} E\left[g(\mathbf{x}_t,\beta)-g(\mathbf{x}_t,\beta_0)\right]^2 > 0,$$

and is satisfied in view of (4.9) and the mean value theorem. Second, uniform convergence over B_* of $R_n(\omega,\beta)$ to $\bar{R}_n(\beta) = E[R_n(\omega,\beta)]$ follows from a standard ULLN for stationary and ergodic processes, if we assume additionally the standard dominance condition

$$E\sup_{B_*}\left[g(\mathbf{x}_t,\beta)-g(\mathbf{x}_t,\beta_0)\right]^2 < \infty.$$

For ULLNs for stationary and ergodic processes see, e.g., Ranga Rao (1962) and Pötscher and Prucha (1986a), Lemma A.2.

As suggested by the above example, demonstrations that there exists a compact subset of the parameter space to which the analysis can be restricted will typically be problem dependent. In the case of maximum likelihood estimation with a parameter space that is a closed subset of Euclidean space conditions for the existence of such a compact subset typically maintain that the densities decline to zero along any sequence of parameters whose norm tends to infinity; compare, e.g., Wald (1949), Huber (1967), Perlman (1972). Of course, the assumption that the parameter space is a closed subset of Euclidean space is not satisfied in such basic cases as in the case of maximum likelihood estimation of the mean μ and the positive definite variance covariance matrix Σ of a normally distributed random vector. Still, a suitable compact subset of the form

$$\{(\mu,\Sigma) : |\mu| \leq c_1,\ \lambda_{\min}(\Sigma) \geq c_2,\ \lambda_{\max}(\Sigma) \leq c_3\}$$

with $c_1 < \infty$, $0 < c_2 < c_3 < \infty$ can be found upon utilizing the specific structure of the normal density (where $\lambda_{\min}$ and $\lambda_{\max}$ denote the smallest and largest eigenvalues); cf. also the discussion after Theorem 14.1.

An alternative to the strategy of reducing the consistency problem to a compact subset of the parameter space is to extend the estimation problem to a compactification of the parameter space; see for example Bahadur (1967), Huber (1967), Perlman (1972), Heijmans and Magnus (1986a, 1987). Although compactifications are typically easy to construct, one then has to deduce, e.g., the identifiable uniqueness condition of the extended problem from the original problem. Typically, this will require similar arguments as those necessary for demonstrating that the problem can be restricted to a compact subset of the parameter space.

4.4 Approximate M-Estimators

Inspection of the proof of Lemma 3.1 reveals that it also holds for approximate M-estimators. We say that $\hat{\beta}_n$ is an approximate M-estimator if the sequence $\hat{\beta}_n$ eventually satisfies

$$R_n(\omega, \hat{\beta}_n) \leq \inf_B R_n(\omega, \beta) + \delta_n, \tag{4.10}$$

where $\delta_n \to 0$ as $n \to \infty$ is a given sequence of positive numbers, cf., e.g., Wald (1949). To ensure the existence of approximate M-estimators we assume here that eventually $\inf_B R_n(\omega, \beta) > -\infty$.[5]

We note that not only Lemma 3.1, but also the discussions in Sections 4.1 - 4.3, apply to approximate M-estimators. This is reassuring, since any numerically calculated M-estimator will be an approximate M-estimator rather than an M-estimator in the strict sense. The existence of measurable approximate M-estimators is guaranteed under general conditions, see, e.g., Brown and Purves (1973).

4.5 Limitations: An Illustrative Example

In this section we review exemplarily the basic structure of the consistency proof of maximum likelihood estimators in autoregressive moving average (ARMA) models. Although ARMA models are linear, the corresponding likelihood function is highly nonlinear. It turns out that (unless the model class is restricted in an unnatural way) the consistency proof is complicated. For instance, as will be explained in more detail below, the objective function $R_n(\omega, \beta)$ does in general not converge uniformly (or one-sided uniformly) to some limiting function $\bar{R}(\beta)$, even after restricting the problem to some compact space. The message from the example is that any catalogue of assumptions that allows the direct application of the classical consistency proof will, while covering probably many models of interest, not be applicable to all situations of interest. Hence problem specific methods have to be employed in such cases.

The consistency proof for ARMA models (without restricting the model class in an unnatural way) is due to Hannan (1973) for the univariate case and Dunsmuir and Hannan (1976) for the multivariate case. Some lacunae

[5]This assumption could be avoided by a more general formulation of (4.10), see, e.g., Pfanzagl (1969) or Brown and Purves (1973).

in the latter paper are closed in Deistler, Dunsmuir and Hannan (1978) and Pötscher (1987). The latter paper also deals with the misspecified case and general parameter restrictions; for estimation of ARMA models under misspecification see also Ploberger (1982b) and Pötscher (1991). For a general reference on the estimation of ARMA models see Hannan and Deistler (1988).

Consider the univariate ARMA(1,1) model

$$a(L)\mathbf{y}_t = b(L)\epsilon_t, \qquad t \in \mathbf{Z}, \tag{4.11}$$

where L is the lag operator, $a(z) = 1 + az$, $b(z) = 1 + bz$, $|a| < 1$, $|b| \leq 1$, and where ϵ_t is white noise, i.e., $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = \sigma^2 > 0$ and $E\epsilon_t\epsilon_r = 0$ for $t \neq r$. The unique weakly stationary solution of (4.11) is given by

$$\mathbf{y}_t = k(L)\epsilon_t \tag{4.12}$$

where

$$k(z) = \sum_{i=0}^{\infty} k_i z^i = a^{-1}(z)b(z)$$

is the transfer function. Observe that $\mathbf{y}_t$ depends on (a, b) only through $k(z)$, that is (k, σ^2) are the "intrinsic" parameters rather than (a, b, σ^2). The following discussion focuses on the estimation of (k, σ^2).[6] The parameter space is then

$$B = \{\beta = (k, \sigma^2) : k(z) = (1 + az)^{-1}(1 + bz), |a| < 1, |b| \leq 1, 0 < \sigma^2 < \infty\}.$$

This space can be endowed with a suitable metric. Let $\beta_0 = (k_0, \sigma_0^2)$ be the true parameter. As the objective function $R_n(\omega, \beta)$ defining the estimator for β_0 we take the normal quasi log-likelihood function (multiplied by $-1/n$). A detailed study of the properties of the objective function R_n can be found in Deistler and Pötscher (1984).

We note that B is not compact, which reflects the fact that $|a| < 1$ and $0 < \sigma^2 < \infty$. (In a multivariate version of (4.11) there is also another source for non-compactness, as the parameters in (4.11) can also become arbitrarily large.) To transfer the problem to a compact parameter space neither strategy suggested in Section 4.3 is readily and directly applicable. However, a combination of both strategies works: One can extend the objective function $R_n(\omega, \beta)$ to a larger space $B_{**} \supseteq B$ in a suitable way, one can find a compact subset $B_* \subseteq B_{**}$ of the larger space and then show that minimizing $R_n(\omega, \beta)$ over B_* leads to the same answer as minimizing $R_n(\omega, \beta)$ over B (for large n a.s.).

[6]Given consistent estimators for (k, σ^2) it is then possible to obtain consistent estimators for (a, b) given that the parameters (a, b) are identifiable in the parameter space under consideration. Of course, one could work directly with (a, b, σ^2), but this would not simplify, but rather complicate the analysis.

As mentioned above, a major difficulty arises since even on the compact set B_* the objective function $R_n(\omega,\beta)$ does not in general converge uniformly (or one-sided uniformly) to its asymptotic counterpart $\bar{R}(\beta)$.[7] The non-uniformness arises for example near $b = 1$ and $b = -1$, that is for models with a root of the moving average part near the unit circle. The approach taken in the literature cited above is to prove a more subtle version of (4.3) by modifying $R_n(\omega,\beta)$ and $\bar{R}(\beta)$ such that the modified functions converge uniformly over suitable subsets of the parameter space and then by letting the "degree of modification" decline to zero.[8] Consistency is then obtained by verifying (4.4). We note also that the verification of (4.4) is involved; for a detailed discussion see Pötscher (1987) and Dahlhaus and Pötscher (1989).

4.6 Identifiable Uniqueness

As mentioned in Chapter 3, in case $\bar{R}_n \equiv \bar{R}$ is continuous (or lower semicontinuous) and the parameter space B is compact, the identifiable uniqueness condition is equivalent to the existence of a unique minimizer of $\bar{R}$. As a consequence, the identifiable uniqueness condition can be checked conveniently by inspection of a single function. The following lemma gives conditions under which, also in the case where $\bar{R}_n$ depends on n, a simplified version of the identifiable uniqueness condition can be found.

Lemma 4.1. *Given a sequence of functions $\bar{R}_n : B \to \mathbf{R}$, let B be compact and $\{\bar{R}_n : n \in \mathbf{N}\}$ be equicontinuous on B. Furthermore let $\bar{\beta}_n$ be a sequence of minimizers of $\bar{R}_n$ with $\bar{\beta}_n \to \bar{\beta}$ as $n \to \infty$. Then $\bar{\beta}_n$ is identifiably unique w.r.t. $\bar{R}_n$ iff*

$$\liminf_{n\to\infty} \left[\bar{R}_n(\beta) - \bar{R}_n(\bar{\beta}_n)\right] > 0 \text{ for all } \beta \neq \bar{\beta}$$

(or equivalently iff $\liminf_{n\to\infty}[\bar{R}_n(\beta) - \bar{R}_n(\bar{\beta})] > 0$ for all $\beta \neq \bar{\beta}$).

In the case of least mean distance or generalized method of moments estimation equicontinuity of $\{\bar{R}_n : n \in \mathbf{N}\}$ is typically obtained as a byproduct from a ULLN (via Lemmata 3.2 or 3.3). Of course, an important special case where $\bar{\beta}_n \to \bar{\beta}$ is the case $\bar{\beta}_n \equiv \bar{\beta}$.

As discussed in Chapter 3, identifiable uniqueness implies essential uniqueness of the minimizers $\bar{\beta}_n$ of $\bar{R}_n$ in the sense that the diameter of the set of minimizers of $\bar{R}_n$ goes to zero as n tends to infinity. In the correctly specified case this assumption seems typically to be reasonable as it

[7] We note that in this example the function $\bar{R}$ takes its values in $\mathbf{R} \cup \{+\infty\}$.

[8] This modification of the objective function is highly problem specific and exploits heavily the structure of ARMA models.

frequently just boils down to identifiability of the true parameter. In the following we discuss in more detail the adequacy of the identifiable uniqueness assumption under misspecification: For simplicity assume for the moment that $\bar{R}_n \equiv \bar{R}$. Then identifiable uniqueness implies the existence of a unique minimizer of $\bar{R}$, and hence – loosely speaking – the existence of a unique best approximate model. It seems that there is no genuine reason why, in general, this should necessarily be the case under misspecification. Indeed, within the context of ARMA models Kabaila (1983) has demonstrated the existence of multiple minimizers of $\bar{R}$ if ARMA models of too low an order are fitted to higher order ARMA processes. Example 3 in Pötscher (1991) provided even an instance where $\bar{R}$ is constant over the entire parameter space and hence is minimized at every parameter value. On the other hand Ploberger (1982a) showed that the best approximating ARMA model is unique provided the amount of misspecification (measured in terms of the spectral measure) is small.

Freedman and Diaconis (1982) consider the "simple" problem of estimating a location parameter from an i.i.d. sample, say $\mathbf{z}_1, \ldots, \mathbf{z}_n$. In their example the objective function is of the form

$$R_n(\omega, \beta) = n^{-1} \sum_{t=1}^{n} \rho(\mathbf{z}_t - \beta),$$

where ρ is a certain function that is symmetric about zero and where the $\mathbf{z}_t$'s have a density that is symmetric about zero. They show that $\bar{R}(\beta) = E[R_n(\omega, \beta)]$ has two minima located symmetrically around zero and hence the identifiable uniqueness condition is violated (and the M-estimator is inconsistent for the true value zero).[9] Donoho and Liu (1988) show that for certain minimum distance estimators the Freedman-Diaconis effect can occur even if the hypothesized and true densities are arbitrarily close.

Given the above discussion it seems of interest to consider the properties of M-estimators without the assumption of identifiable uniqueness. In the context of misspecified ARMA models this question has been analyzed in Ploberger (1982b) for prediction error estimators, and in Pötscher (1987, 1991) for maximum likelihood estimators. Within that context $\bar{R}_n \equiv \bar{R}$ and $\hat{\beta}_n$ is the estimator for the transfer function and the innovation variance, cf. Section 4.5. They demonstrate that $\hat{\beta}_n$ still converges to the set of minimizers of $\bar{R}$, where the ARMA systems corresponding to the minimizers have the interpretation of best approximators to the true ARMA process

[9] We note that in this example the location model is correctly specified. The effect stems from the fact that the objective function differs from the true likelihood. Another interesting feature of this example is that although $\partial \bar{R}(\beta)/\partial \beta = 0$ at the true parameter value also the M-estimator of type II is inconsistent. However a consistently started one-step M-estimator is consistent. For further discussions of one-step M-estimators see, e.g., Bickel (1975), Pötscher and Prucha (1986a) and Prucha and Kelejian (1984).

within the given model class. In the following we obtain a similar result within the present context that also covers the case where $\bar{R}_n$ may depend on n. We introduce the following regularity condition for level sets.

Definition 4.1. *For a given sequence of functions $\bar{R}_n : B \to \mathbf{R}$ and a sequence $\bar{c}_n \in \mathbf{R}$ the level sets $\bar{B}_n = \{\beta \in B : \bar{R}_n(\beta) \leq \bar{c}_n\}$ are said to be regular if $\bar{B}_n \neq \emptyset$ for $n \in \mathbf{N}$ and if for every $\epsilon > 0$*

$$\liminf_{n\to\infty} \left[\inf_{\{\beta\in B:\rho_B(\beta,\bar{B}_n)\geq\epsilon\}} \bar{R}_n(\beta) - \sup_{\beta\in\bar{B}_n} \bar{R}_n(\beta) \right] > 0. \qquad (4.13)$$

For example, if $\bar{R}_n \equiv \bar{R}$ is lower semi-continuous and B is compact, then every sequence of level sets $\bar{B}_n$ with $\bar{c}_n \equiv \bar{c} \geq \min_B \bar{R}$ is regular. Definition 4.1 clearly generalizes the definition of identifiable uniqueness. The following lemma generalizes Lemma 3.1.

Lemma 4.2. *Let $R_n : \Omega \times B \to \mathbf{R}$ and $\bar{R}_n : B \to \mathbf{R}$ be two sequences of functions such that a.s. [i.p.]*

$$\sup_B \left| R_n(\omega,\beta) - \bar{R}_n(\beta) \right| \to 0 \text{ as } n \to \infty. \qquad (4.14)$$

Let $\bar{c}_n = \inf_B \bar{R}_n(\beta) > -\infty$ and let the corresponding level sets $\bar{B}_n$ be regular. Then for any sequence $\hat{\beta}_n$ such that eventually[10]

$$R_n(\omega,\hat{\beta}_n) \leq \inf_B R_n(\omega,\beta) + \delta_n, \qquad (4.15)$$

where $\delta_n > 0$ and $\delta_n \to 0$ as $n \to \infty$, i.e., for any sequence of almost minimizers $\hat{\beta}_n$, we have $\rho_B(\hat{\beta}_n, \bar{B}_n) \to 0$ a.s. [i.p.] as $n \to \infty$.[11]

The above lemma tells us that the distance between (almost) minimizers of $R_n(\omega,\beta)$ and the set of minimizers of $\bar{R}_n(\beta)$ becomes negligible. Generalizations similar to the generalizations of Lemma 3.1 discussed in Section 4.2 are also possible for the above lemma. Furthermore, in cases where uniform

---

[10]For the i.p. version of the lemma this could be relaxed to the requirement that (4.15) holds on sets $\Omega_n \in \mathfrak{A}$ with $P(\Omega_n) \to 1$.

[11]The formulation of the lemma is such that the existence of $\hat{\beta}_n$ is implicitly assumed. We note, however, that the existence of $\hat{\beta}_n$ satisfying (4.15) eventually follows in fact from the a.s. version of (4.14) and the assumption that $\bar{c}_n = \inf_B \bar{R}_n(\beta) > -\infty$, since they together imply that eventually $\inf_B R_n(\omega,\beta) > -\infty$. Similarly, the existence of $\hat{\beta}_n$ satisfying (4.15) on sets $\Omega_n \in \mathfrak{A}$ with $P(\Omega_n) \to 1$ – cf. also Footnote 10 – follows in fact from the i.p. version of (4.14) and the assumption that $\bar{c}_n = \inf_B \bar{R}_n(\beta) > -\infty$.

convergence over the entire parameter space B does not hold, strategies for reducing the parameter space as discussed in Section 4.3 may be combined with Lemma 4.2.

If $\bar{R}_n \equiv \bar{R}$ is lower semi-continuous and B is compact the regularity assumption on the level sets $\bar{B}_n$ and the assumption $\bar{c}_n > -\infty$ in Lemma 4.2 are always satisfied, cf. the discussion after Definition 4.1. Hence, in this important case, Lemma 4.2 provides a complete description of the convergence behavior of $\hat{\beta}_n$ without requiring any further assumptions beyond the uniform convergence condition (4.14). In the general case, where $\bar{R}_n$ may depend on n or where $\bar{R}_n \equiv \bar{R}$ is not lower semi-continuous or where B is not compact, Lemma 4.2 also covers many further situations where the identifiable uniqueness assumption is violated. We note, however, that in the general case the identifiable uniqueness assumption can also fail in ways not covered by Lemma 4.2. For example, consider a situation where each $\bar{R}_n$, although having a unique minimizer $\bar{\beta}_n$, becomes flatter and flatter at $\bar{\beta}_n$ as n increases. Then the regularity condition on the level sets $\bar{B}_n$ in Lemma 4.2 – which in this example is equivalent to the identifiable uniqueness condition due to the uniqueness of the minimizers – will typically be violated. In this and other situations where Lemma 4.2 fails to apply, not much can be said about the convergence behavior of $\hat{\beta}_n$ in general, as it will depend on the particularities of the estimation problem at hand. In the situation just discussed it may in some instances be possible to renormalize R_n and $\bar{R}_n$ such that Lemma 3.1 applies to the renormalized functions and to establish convergence of $\rho_B(\hat{\beta}_n, \bar{\beta}_n)$ to zero in this fashion. In other instances $\hat{\beta}_n$ will in fact not satisfy $\rho_B(\hat{\beta}_n, \bar{\beta}_n) \to 0$ a.s. or i.p.

Sometimes in situations where the identifiable uniqueness condition fails but Lemma 4.2 applies, ensuring convergence of $\hat{\beta}_n$ to the set of minimizers of $\bar{R}_n$, one is interested only in certain functions, ζ_n say, of the parameter β that are constant over the set of parameter values that minimize $\bar{R}_n$. Such functions are reminiscent of "estimable" functions of an unidentifiable parameter in linear models and they can be estimated consistently (under regularity conditions) even if the identifiable uniqueness condition is violated.

Corollary 4.3. *Suppose the assumptions of Lemma 4.2 hold. Let* $\zeta_n : B \to \mathbf{R}^k$ *be a sequence of functions that is constant on* $\bar{B}_n$ *and that is uniformly equicontinuous on the subset* $\bigcup\{\bar{B}_n : n \in \mathbf{N}\}$ *of* B.[12] *Then* $\zeta_n(\hat{\beta}_n) - \bar{\zeta}_n \to 0$

[12]Recall that uniform equicontinuity on a subset is a stronger condition than uniform equicontinuity of the family of restrictions to the same subset. The uniform equicontinuity property required in the corollary clearly follows if ζ_n is equicontinuous on a compact set containing $\bigcup\{\bar{B}_n : n \in \mathbf{N}\}$. It trivially also follows from uniform equicontinuity of ζ_n on B (which in case B is compact reduces to equicontinuity on B). The (uniform) equicontinuity conditions simplify to corresponding (uniform) continuity conditions if ζ_n does not depend on n.

a.s. [i.p.] as $n \to \infty$, *where* $\bar{\zeta}_n$ *is the value of* ζ_n *on* $\bar{B}_n$.

A related result is given in Gallant (1987b). Of course, Corollary 4.3 also holds if the functions ζ_n take their values in an arbitrary metric space. Furthermore, inspection of the proof of Corollary 4.3 shows that the essential ingredient is – apart from the properties of ζ_n assumed in Corollary 4.3 – the property $\rho_B(\hat{\beta}_n, \bar{B}_n) \to 0$ a.s. [i.p.]. Hence, a result like Corollary 4.3 can be obtained for situations not covered by Lemma 4.2 as long as the sequence of estimators "converges" to a sequence of sets for which the functions ζ_n satisfy properties analogous to the ones listed in Corollary 4.3.

Corollary 4.3 shows how consistency of "estimable" functions can be deduced from $\rho_B(\hat{\beta}_n, \bar{B}_n) \to 0$. It is worth pointing out that even if this crucial condition cannot be established, it may still sometimes be possible to prove the consistency of "estimable" functions along the following lines: Suppose the entire estimation problem is reparameterized in terms of a new parameter, say θ, such that the "underidentification" disappears. (One may think of θ as composed of suitable identifiable functions of the original parameter β.) If it is then possible to establish consistency of estimators for θ directly in the new parameterization without recourse to the β-parameterization, consistency for functions of interest, which depend on β only through θ (in a continuous fashion), will then follow immediately. Of course, a number of technical complications may be encountered in actually implementing this approach in a particular case.[13]

---

[13]Essentially such an approach is used in establishing consistency for (normal quasi) maximum likelihood estimators of transfer functions of multivariate ARMA models. In this case the estimable functions in θ are the transfer function coefficients. For references see Section 4.5.

5

UNIFORM LAWS OF LARGE NUMBERS

As documented in Chapters 3 and 4 a basic ingredient for typical consistency proofs is that the difference between the objective function and its nonstochastic counterpart converges to zero uniformly over the parameter space (or, if the approach in Section 4.3 is followed, at least over a suitably chosen subset). In many cases uniform convergence of the objective function will follow from a uniform law of large numbers (ULLN), either directly or via Lemmata 3.2 and 3.3.

As mentioned in Chapter 1 there is a considerable body of literature on ULLNs for stationary or asymptotically stationary processes. However, for a proper asymptotic theory for dynamic nonlinear models we need ULLNs that apply to temporally dependent and heterogeneous processes. Hoadley (1971) introduced a ULLN that allows for independent and not necessarily identically distributed data processes. This ULLN (or some version of it) has been used widely in the econometrics literature, see, e.g., White (1980), Domowitz and White (1982), Levine (1983), White and Domowitz (1984), Bates and White (1985) and Domowitz (1985). However, Andrews (1987) and Pötscher and Prucha (1986a,b) point out that the assumptions maintained by this ULLN, and hence the catalogues of assumptions in the papers utilizing this ULLN, are severe and preclude the analysis of many estimators and models of interest in econometrics, since the assumptions of this ULLN essentially require that the random variables involved are bounded. (We note, however, that since the proofs of the theorems regarding consistency in the above papers by Bates, Domowitz, Levine and White follow the structure of the classical consistency proof, Hoadley's ULLN could be readily replaced by some alternative ULLN. Of course, this would require corresponding modifications in the catalogues of assumptions maintained in these papers.) The above observation and the extension of the asymptotic theory to dynamic models and temporally dependent and heterogeneous data generating processes necessitated the development of ULLNs that are applicable in such settings. Such ULLNs have been introduced by Andrews (1987) and Pötscher and Prucha (1986b, 1989, 1994b) and will be presented below.[1]

---

[1]For simplicity of presentation the results in this chapter are only presented

B. M. Pötscher et al., *Dynamic Nonlinear Econometric Models*

5.1 ULLNs for Dependent and Heterogeneous Processes

For the following discussion let (Θ, ρ) be a non-empty metric space and let the measurable space $(Z, \mathfrak{Z})$, in which the process $(\mathbf{z}_t)$ takes its values, be a metrizable space with its corresponding Borel σ-field. Furthermore, let $q_t : Z \times \Theta \to \mathbf{R}$ be $\mathfrak{Z}$-measurable for each $\theta \in \Theta$, $t \in \mathbf{N}$. ULLNs provide conditions under which sums of the form

$$n^{-1} \sum_{t=1}^{n} [q_t(\mathbf{z}_t, \theta) - Eq_t(\mathbf{z}_t, \theta)]$$

converge to zero uniformly over the parameter space Θ. In applying ULLNs within the context of Chapter 3 the space Θ will typically correspond to $T \times B$ or B (or some subset thereof). We introduce the following assumptions.

Assumption 5.1. *Θ is compact.*

Assumption A. *For each $\theta \in \Theta$ there exists an $\eta > 0$ such that $\rho(\theta, \theta^\bullet) \leq \eta$ implies*

$$|q_t(\mathbf{z}_t, \theta^\bullet) - q_t(\mathbf{z}_t, \theta)| \leq b_t(\mathbf{z}_t) h\left[\rho(\theta, \theta^\bullet)\right], \text{ for all } t \in \mathbf{N}, \text{ a.s.},$$

where $b_t : Z \to [0, \infty)$ and $h : [0, \infty) \to [0, \infty)$ are such that $b_t(\mathbf{z}_t)$ is $\mathfrak{A}$-measurable,

$$\sup_n n^{-1} \sum_{t=1}^{n} Eb_t(\mathbf{z}_t) < \infty,$$

$h(y) \downarrow h(0) = 0$ as $y \downarrow 0$, and η, b_t, h and the null set may depend on θ.

Assumption B. *$q_t(z, \theta) = \sum_{k=1}^{K} r_{kt}(z) s_{kt}(z, \theta)$, where the r_{kt} are $\mathfrak{Z}$-measurable real functions for all $t \in \mathbf{N}$ and $1 \leq k \leq K$, and the family $\{s_{kt}(z, \theta) : t \in \mathbf{N}\}$ of real functions is equicontinuous on $Z \times \Theta$ for all $1 \leq k \leq K$, i.e., for each $(z^\bullet, \theta^\bullet) \in Z \times \Theta$ we have*

$$\sup_{t \in \mathbf{N}} |s_{kt}(z, \theta) - s_{kt}(z^\bullet, \theta^\bullet)| \to 0 \text{ as } (z, \theta) \to (z^\bullet, \theta^\bullet).$$

Assumption C. *(i) Let $d_t(z) = \sup_{\theta \in \Theta} |q_t(z, \theta)|$, then*

$$\sup_n n^{-1} \sum_{t=1}^{n} E\left[d_t(\mathbf{z}_t)^{1+\gamma}\right] < \infty$$

---

for sequences of random variables. However, they can be readily extended to triangular arrays.

for some $\gamma > 0$, *and*

(ii) $\sup_n n^{-1} \sum_{t=1}^{n} E\left|r_{kt}(\mathbf{z}_t)\right| < \infty$ *for* $1 \leq k \leq K$.

Assumption D. *The sequence* $\{\bar{H}_n^z : n \in \mathbf{N}\}$ *is tight on* Z *where* $\bar{H}_n^z = n^{-1} \sum_{t=1}^{n} H_t^z$ *and* H_t^z *is the distribution of* $\mathbf{z}_t$. *(That is,*

$$\lim_{m \to \infty} \sup_n n^{-1} \sum_{t=1}^{n} P(\mathbf{z}_t \notin K_m) = 0$$

for some sequence of compact sets $K_m \subseteq Z$.*)*

Assumption 5.2. *(Local Laws of Large Numbers) For* $\eta > 0$ *let*

$$q_t^*(z, \theta; \eta) = \sup_{\rho(\theta, \theta^\bullet) < \eta} q_t(z, \theta^\bullet)$$

and

$$q_{t*}(z, \theta; \eta) = \inf_{\rho(\theta, \theta^\bullet) < \eta} q_t(z, \theta^\bullet).$$

For each $\theta \in \Theta$ *and* $\eta > 0$ *small enough (i.e.,* $0 < \eta \leq \eta(\theta)$*) the functions* $q_t^*(z, \theta; \eta)$ *and* $q_{t*}(z, \theta; \eta)$ *are real valued and* $\mathfrak{Z}$*-measurable and the random variables* $q_t^*(\mathbf{z}_t, \theta; \eta)$ *and* $q_{t*}(\mathbf{z}_t, \theta; \eta)$ *satisfy a strong [weak] LLN for all* $\eta > 0$ *small enough.*

In assuming the existence of a LLN for the "bracketing" functions $q_t^*(\mathbf{z}_t, \theta; \eta)$ and $q_{t*}(\mathbf{z}_t, \theta; \eta)$ in Assumption 5.2 it is implicitly maintained that their expectations exist and are finite. The $\mathfrak{Z}$-measurability and finiteness of q_t^* and q_{t*} maintained in Assumption 5.2 follows automatically under Assumptions B and 5.1, since the map $\theta \to q_t(z, \theta)$ is continuous for every $z \in Z$ and since Θ is compact and metrizable. Similarly, under Assumptions B and 5.1 the $\mathfrak{Z}$-measurability and finiteness of $d_t(z)$ is automatically ensured. We note furthermore that Andrews (1987) and Pötscher and Prucha (1989) refer to local LLNs as pointwise LLNs. However, in this book the term pointwise LLNs is used to describe LLNs for $q_t(\mathbf{z}_t, \theta)$.

The following ULLNs follow from Andrews (1987) and Pötscher and Prucha (1989), respectively:[2]

Theorem 5.1. *If Assumptions 5.1, 5.2 and A hold, then* $Eq_t(\mathbf{z}_t, \theta)$ *exists and is finite and*

(a) $\sup_{\theta \in \Theta} \left|n^{-1} \sum_{t=1}^{n} [q_t(\mathbf{z}_t, \theta) - Eq_t(\mathbf{z}_t, \theta)]\right| \to 0$ *a.s. [i.p.] as* $n \to \infty$,

(b) $\{n^{-1} \sum_{t=1}^{n} Eq_t(\mathbf{z}_t, \theta) : n \in \mathbf{N}\}$ *is equicontinuous on* Θ.

[2]The above catalogue of assumptions can be weakened, see Andrews (1987) and Pötscher and Prucha (1986b, 1989, 1994b).

Theorem 5.2. [3] *If Assumptions 5.1, 5.2 and B, C, D hold, then* $Eq_t(\mathbf{z}_t, \theta)$ *exists and is finite and*

(a) $\sup_{\theta \in \Theta} |n^{-1} \sum_{t=1}^{n} [q_t(\mathbf{z}_t, \theta) - Eq_t(\mathbf{z}_t, \theta)]| \to 0$ *a.s. [i.p.] as* $n \to \infty$,

(b) $\{n^{-1} \sum_{t=1}^{n} Eq_t(\mathbf{z}_t, \theta) : n \in \mathrm{N}\}$ *is equicontinuous on* Θ.

An important feature of both of the above ULLNs expressed in Assumption 5.2 is that they transform local laws of large numbers into uniform laws of large numbers. It is for that reason that those ULLNs have been called generic. In a particular application the local laws of large numbers postulated in Assumption 5.2 will typically have to be deduced from more basic assumptions on the probability law of the process $(\mathbf{z}_t)$. For example, if the process $(\mathbf{z}_t)$ is α-mixing or ϕ-mixing then the random variables $q_t^*(\mathbf{z}_t, \theta; \eta)$ and $q_{t*}(\mathbf{z}_t, \theta; \eta)$ have the same respective properties and hence corresponding LLNs that are readily available in the literature can be used to imply Assumption 5.2. However, as pointed out in Chapter 1, the assumption of an α-mixing or ϕ-mixing data generating process seems generally not to be adequate for a treatment of dynamic models. LLNs tailored to the needs of an asymptotic theory for such models have been discussed in McLeish (1975a), Bierens (1981, 1982a, 1984, 1987), Gallant (1987a, Ch.7), Andrews (1988), Gallant and White (1988), Hansen (1991), De Jong (1995a), and Davidson and De Jong (1995), and are the subject of Sections 6.3 and 6.5.

Assumption A maintains that the functions $q_t(\mathbf{z}_t, \theta)$ satisfy a generalized Lipschitz condition with respect to θ and a moment condition on the generalized Lipschitz bound. In practice this assumption will often be deduced from suitable differentiability conditions with respect to θ and dominance conditions on the first derivative. In contrast, Assumption B maintains for the functions $q_t(z, \theta)$ a continuity type condition jointly with respect to z and θ. The assumption allows for discontinuities in z given those discontinuities can be "separated" from the parameters θ.[4] Of course, it also contains the case where $\{q_t : t \in \mathrm{N}\}$ is equicontinuous on $Z \times \Theta$ (e.g., if we put $K = 1$, $r_{kt} \equiv 1$). We note that Assumptions A and B, and hence Theorems 5.1 and 5.2, complement each other. Practically speaking Assumption A imposes a stronger "smoothness" condition w.r.t. the parameters while Assumption B imposes a stronger "smoothness" condition w.r.t. the data. The first part of Assumption C is a standard dominance condition, the second part is, e.g., automatically satisfied if the r_{kt} are indicator functions. Sufficient conditions for Assumption D will be given in Assumptions D1,

---

[3]The joint equicontinuity condition maintained in Assumption B (and some of the other assumptions) can be weakened if $Eq_t(\mathbf{z}_t, \theta)$, $Eq_t^*(\mathbf{z}_t, \theta; \eta)$ and $Eq_{t*}(\mathbf{z}_t, \theta; \eta)$ do not depend on t, which is, for example, the case in a stationary or i.i.d. environment with $q_t \equiv q$. For results of this type see Jennrich (1969), Theorem 2, and Pötscher and Prucha (1986a), Lemma A.2.

[4]As a consequence, this assumption covers, e.g., the case of the log-likelihood function of a Tobit model where the r_{kt} represent indicator functions.

D2, D' and D1' below. For a further discussion of the assumptions see Andrews (1987) and Pötscher and Prucha (1989). Compare also the discussion in Section 6.6.

Within the context of dynamic models we need to consider the application of ULLNs to functions of the form $q_t(\mathbf{w}_t, \ldots, \mathbf{w}_{t-l_t}, \theta)$, where $\mathbf{w}_t$ takes its values in a Borel subset W of $\mathbf{R}^{p_w}$. Typically (but not necessarily) $\mathbf{w}_t$ will be a vector composed of current endogenous and exogenous variables. The case of a fixed lag length corresponds to $l_t = l$; the case of an increasing lag length corresponds, e.g., to $l_t = t - l$ ($l \geq 0$). The latter case arises, e.g., in the context of quasi maximum likelihood estimation of dynamic models where the disturbances follow a moving average process. Also models with an increasing lag length occur for example in the literature on dynamic factor demand with an endogenous capital depreciation rate; see, e.g., Epstein and Denny (1980) and Prucha and Nadiri (1988, 1996).

To apply Theorems 5.1 and 5.2 to functions $q_t(\mathbf{w}_t, \ldots, \mathbf{w}_{t-l_t}, \theta)$ we define in case of a fixed lag length $\mathbf{z}_t = (\mathbf{w}_t', \ldots, \mathbf{w}_{t-l}')'$. In case of an increasing lag length we define $\mathbf{z}_t = (\mathbf{w}_t', \mathbf{w}_{t-1}', \ldots)'$ (if $\mathbf{w}_{-s}$ is not defined for some $s \geq 0$ we may, e.g., set it equal to an arbitrary element of W). In both cases we can then redefine $q_t(\mathbf{w}_t, \ldots, \mathbf{w}_{t-l_t}, \theta)$ as $q_t(\mathbf{z}_t, \theta)$. In case of a fixed lag length the space Z is generally a subset of $\prod_{i=0}^{l} W \subseteq \mathbf{R}^{(l+1)p_w}$. The reason for allowing Z to be a subset, rather than the entire product space $\prod_{i=0}^{l} W$, is that $q_t(., \theta)$ may not be defined on the entire product space but only on the subset Z (with the process $(\mathbf{z}_t)$ also taking its values only in Z). Analogously, in case of an increasing lag length Z is generally a subset of $\prod_{i=0}^{\infty} W \subseteq \mathbf{R}^{\infty}$.

Assumptions A, B and C are basically conditions on the functions q_t, while Assumption D represents a condition on the distribution of the random variables $\mathbf{z}_t$. If $\mathbf{z}_t = (\mathbf{w}_t', \ldots, \mathbf{w}_{t-l}')'$ or $\mathbf{z}_t = (\mathbf{w}_t', \mathbf{w}_{t-1}', \ldots)'$ it seems of interest to consider the meaning of this condition in terms of the originally given random variables $\mathbf{w}_t$, and to provide sufficient conditions for Assumption D in terms of the process $(\mathbf{w}_t)$.

Assumption D1. *W is a Borel subset of $\mathbf{R}^{p_w}$ and Z is a relatively closed subset in $\prod_{i=0}^{\infty} W$. Furthermore the family $\{\bar{H}_n^w = n^{-1}\sum_{t=1}^{n} H_t^w : n \in \mathbf{N}\}$ is tight on W, where H_t^w is the distribution of $\mathbf{w}_t$.*[5]

Assumption D2. *W is a closed subset of $\mathbf{R}^{p_w}$ and Z is a relatively closed subset in $\prod_{i=0}^{\infty} W$.*[6] *Furthermore, for a monotone function $s : [0, \infty) \to [0, \infty)$ with $s(x) \to \infty$ as $x \to \infty$ we have $\sup_n n^{-1}\sum_{t=1}^{n} Es(|\mathbf{w}_t|) < \infty$.*

---

[5]Note that this assumption postulates tightness of the family $\{\bar{H}_n^w : n \in \mathbf{N}\}$ on W, and not just tightness of this family viewed as distributions on the larger space $\mathbf{R}^{p_w}$, cf. Billingsley (1968), Lemma 2 and Problem 8 in Section 6.

[6]Observe that, since W is closed, Z is closed in $\mathbf{R}^{\infty}$.

In Lemmata C1 and C2 in Appendix C we show that either one of Assumptions D1 and D2 is sufficient for Assumption D. Assumptions D1 and D2 are given for the case where $\mathbf{z}_t = (\mathbf{w}_t', \mathbf{w}_{t-1}', \ldots)'$. Of course, the case of a fixed lag length may be regarded as a special situation of the case of an increasing lag length. For simplicity of presentation we have given the above conditions for the general case only. If one wishes to consider the case of a fixed lag length separately, then the above conditions apply with $\prod_{i=0}^{\infty} W$ replaced by $\prod_{i=0}^{l} W$.

It follows from Parthasarathy (1967, p.29) that a sufficient condition for the family $\{\bar{H}_n^w : n \in \mathbf{N}\}$ in Assumption D1 to be tight on W is that $\bar{H}_n^w$ converges weakly to some distribution, say $\bar{H}^w$, as $n \to \infty$. A typical choice for the function $s(x)$ in Assumption D2 is $s(x) = x^p$, $p > 0$, or $s(x) = \ln(1+x)$. In both cases Assumption D2 reduces to a weak moment-type condition. Furthermore, the relative closedness assumption for Z is trivially satisfied if $Z = \prod_{i=0}^{\infty} W$ or $\prod_{i=0}^{l} W$.

Assumption D2 maintains that W and hence Z is closed. For more general formulations of this assumption see Pötscher and Prucha (1989), Assumption 5A', and Pötscher and Prucha (1994b), Remark 5.

The common feature of Assumptions D, D1 and D2 is that they exclude situations where some mass of the average probability distribution $\bar{H}_n^z$ of $\mathbf{z}_t$ escapes a sequence of compact sets in Z. This is achieved in Assumption D1 by requiring the analogous property for the average probability distribution $\bar{H}_n^w$ of $\mathbf{w}_t$ and in Assumption D2 by requiring closedness of Z and by placing bounds on moments of certain functions of $\mathbf{w}_t$.

Returning to the general case where $\mathbf{z}_t$ is not necessarily of the form $\mathbf{z}_t = (\mathbf{w}_t', \mathbf{w}_{t-1}', \ldots)'$ we note that Theorems 5.1 and 5.2 only ensure that $n^{-1}\sum_{t=1}^{n}[q_t(\mathbf{z}_t,\theta) - Eq_t(\mathbf{z}_t,\theta)]$ converges to zero uniformly on Θ. In certain cases a stronger result, namely that also $n^{-1}\sum_{t=1}^{n} q_t(\mathbf{z}_t,\theta)$ converges to some finite limit uniformly on Θ is useful. Clearly, in order to obtain this stronger result we need some kind of asymptotic stationarity of the process and of the functions q_t.

Assumption D'. *The process $(\mathbf{z}_t)$ is asymptotically stationary in the sense that the probability measures $\bar{H}_n^z$ converge weakly to some probability measure $\bar{H}^z$ on Z. Furthermore, $\bar{H}^z$ and each of the $\bar{H}_n^z$ are tight on Z.*

We note that Assumption D' implies Assumption D, see Theorem 8 in Billingsley (1968, p.241). The second half of Assumption D' is automatically satisfied if Z is a Borel subset of Euclidean space, or more generally, a Borel subset of a complete and separable metrizable space, see Parthasarathy (1967, p.29). We can now obtain the following ULLN corresponding to Theorem 5.2.

Theorem 5.3. *Let $\{q_t(z,\theta) : t \in \mathbf{N}\}$ be equicontinuous on $Z \times \Theta$ and let*

$q_t(z,\theta)$ converge pointwise to some function $q(z,\theta)$ on $Z \times \Theta$ as $t \to \infty$. Then under Assumptions 5.1, 5.2, C(i), D' and given

$$\sup_n n^{-1} \sum_{t=1}^{n} E\left[d(\mathbf{z}_t)^{1+\delta}\right] < \infty$$

for some $\delta > 0$, where $d(z) = \sup_{\theta\in\Theta} |q(z,\theta)|$, the conclusions of Theorem 5.2 hold. Furthermore $\int q(z,\theta)d\bar{H}^z$ exists, is finite and continuous on Θ, and

$$\sup_{\theta\in\Theta} \left| n^{-1} \sum_{t=1}^{n} q_t(\mathbf{z}_t,\theta) - \int q(z,\theta)d\bar{H}^z \right| \to 0 \text{ a.s. [i.p.] as } n \to \infty. \quad (5.1)$$

The above theorem is a generalization of Theorem 2 in Pötscher and Prucha (1989), which in turn generalizes results in Bierens (1981, 1982a, 1984, 1987) and Pötscher and Prucha (1986a). If $\mathbf{z}_t = (\mathbf{w}_t', \ldots, \mathbf{w}_{t-l}')'$ or $\mathbf{z}_t = (\mathbf{w}_t', \mathbf{w}_{t-1}', \ldots)'$ it seems again of interest to give a sufficient condition for Assumption D' in terms of the process $\mathbf{w}_t$. That the following assumption is sufficient for Assumption D' is established in Lemma C3 in Appendix C.

Assumption D1'. *For all $k \geq 0$, $\bar{H}_n^{w,k}$ converges weakly to some probability measure $\bar{H}^{w,k}$ on $\prod_{i=0}^{k} W$, W a Borel subset of $\mathbf{R}^{p_w}$, and Z is relatively closed in $\prod_{i=0}^{\infty} W$, where $\bar{H}_n^{w,k} = n^{-1}\sum_{t=1}^{n} H_t^{w,k}$ and $H_t^{w,k}$ denotes the distribution of $(\mathbf{w}_t', \ldots, \mathbf{w}_{t-k}')'$.*

Of course, in case of a fixed lag length it suffices to put $k = l$ and to replace $\prod_{i=0}^{\infty} W$ with $\prod_{i=0}^{l} W$ in the above assumption.

5.2 Further Remarks on ULLNs

(i) In light of our discussion in Section 4.2 it may be of interest to obtain uniform convergence results for $c_n^{-1}\sum_{t=1}^{n}[q_t(\mathbf{z}_t,\theta) - Eq_t(\mathbf{z}_t,\theta)]$ and "one-sided" versions of the above results. We note that the above results readily generalize to those situations, see Andrews (1987, 1992), Pötscher and Prucha (1989, 1994a,b) and Pollard (1991).

(ii) The common structure of the proofs of Theorems 5.1 and 5.2 is to imply uniform convergence from the local LLNs by verifying the so-called "first moment continuity condition", i.e.,

$$\sup_n n^{-1} \sum_{t=1}^{n} E \sup_{\rho(\theta,\theta^\bullet)<\delta} |q_t(\mathbf{z}_t,\theta) - q_t(\mathbf{z}_t,\theta^\bullet)| \to 0 \text{ as } \delta \to 0 \quad (5.2)$$

for each $\theta^{\bullet} \in \Theta$, cf. Andrews (1987), and Pötscher and Prucha (1989, 1994b). We note that uniform convergence can also be implied from pointwise LLNs (i.e., LLNs for $q_t(\mathbf{z}_t, \theta)$) by verifying the following stronger condition:

$$\sup_n n^{-1} \sum_{t=1}^{n} E \sup_{\theta^{\bullet} \in \Theta} \sup_{\rho(\theta, \theta^{\bullet}) < \delta} |q_t(\mathbf{z}_t, \theta) - q_t(\mathbf{z}_t, \theta^{\bullet})| \to 0 \text{ as } \delta \to 0. \quad (5.3)$$

ULLNs which are closely related to Theorems 5.1 and 5.2 but are based on pointwise LLNs rather than on local LLNs have recently been derived by Newey (1991), Andrews (1992) and Pötscher and Prucha (1994a). All of these papers use stochastic equicontinuity concepts that are closely related to (5.2) and (5.3). For an extensive discussion of various forms of stochastic equicontinuity concepts see Pötscher and Prucha (1994a).[7] We note furthermore that the results given in Chapter 6 below not only provide sufficient conditions for local LLNs but clearly also for pointwise LLNs. Therefore, the asymptotic theory for M-estimators developed in this book could as well have been based on ULLNs using pointwise LLNs rather than on ULLNs using local LLNs.

(iii) ULLNs for totally bounded and not only for compact parameter spaces have been derived in Andrews (1992) and Pötscher and Prucha (1994a). However, as discussed in the latter reference, from a mathematical point of view uniform convergence results on a totally bounded parameter space are not really more general than those on a compact parameter space.

---

[7]The concept of stochastic equicontinuity in its various forms has been extensively used in the literature on convergence of stochastic processes, see, e.g., Billingsley (1968), and Pollard (1984).

6

APPROXIMATION CONCEPTS AND LIMIT THEOREMS

In this chapter we provide formalizations of the notion that a stochastic process has a "fading memory". Some of these formalizations employ concepts of approximation of one process by another process. The aim of these formalizations is to define classes of processes that – while still satisfying limit theorems (LLNs and CLTs) – are broad and cover, in particular, processes that are generated from a dynamic system.[1] In Section 6.1 we start with a discussion of the limitations of the concept of α-mixing [ϕ-mixing], followed by the definition of L_p-approximability of a stochastic process in Section 6.2. This approximation concept was introduced in Pötscher and Prucha (1991a). It encompasses the approximation concept of stochastic stability and near epoch dependence, and helps to clarify the relationship between these concepts. In Section 6.3 we then discuss LLNs for L_p-approximable and near epoch dependent processes. (The discussion of CLTs is deferred to Chapter 10.) Frequently we are interested in limit theorems for a function of an L_p-approximable [near epoch dependent] process. E.g., when proving consistency via the use of a ULLN, we need to establish local LLNs, i.e., LLNs for the "bracketing" functions $q_t^*(\mathbf{z}_t, \theta; \eta)$ and $q_{t*}(\mathbf{z}_t, \theta; \eta)$. If the underlying process $(\mathbf{z}_t)$ is L_p-approximable [near epoch dependent] this can be accomplished by making use of results that show under which circumstances functions preserve the L_p-approximability [near epoch dependence] property. Preservation results of this type are the subject of Section 6.4. In considering dynamic systems it is important to know when the process generated by the system will satisfy the L_p-approximability [near epoch dependence] property. Hence, in Section 6.4 we also provide sufficient conditions for dynamic systems under which the output process is L_p-approximable [near epoch dependent]. Since limit theorems for L_p-approximable [near epoch dependent] processes are available (cf. Section 6.3 and Chapter 10), such results are fundamental for the derivation of limit theorems for (func-

[1]For simplicity of presentation the results in this chapter are only presented for sequences of random variables. However, they can be readily extended to triangular arrays.

tions of) processes that are generated by dynamic systems. Several of these results, and in particular those that cover higher order systems, are new and have not been available in the literature previously. Finally, in Section 6.5 we utilize the results developed in this chapter to give sets of sufficient conditions which ensure that $q_t(\mathbf{z}_t, \theta)$ satisfies a local LLN, i.e., we provide sufficient conditions for Assumption 5.2 in Chapter 5.

6.1 Dynamic Models and Mixing Processes

On an intuitive level it is obvious that, in order to establish limit theorems for a stochastic process, it is typically necessary that the process has a "fading" memory. The notion of a fading memory has been formalized in various different ways in the statistics and probability theory literature. One important formalization is the concept of α-mixing and ϕ-mixing. In essence, α-mixing and ϕ-mixing is a notion of "asymptotic independence". For background information on α-mixing and ϕ-mixing see the recent monograph by Doukhan (1994).

Definition 6.1. [2] *Let $(\xi_t)_{t\in\mathbf{Z}}$ be a stochastic process on $(\Omega, \mathfrak{A}, P)$ that takes its values in some measurable space. Let $\mathfrak{A}^i_{-\infty}$ be the σ-field generated by $\xi_i, \xi_{i-1}, \ldots$ and let $\mathfrak{A}^\infty_k$ be the σ-field generated by $\xi_k, \xi_{k+1}, \ldots$; define*

$$\alpha(j) = \sup_{k\in\mathbf{Z}} \sup \left\{ |P(F \cap G) - P(F)P(G)| : F \in \mathfrak{A}^k_{-\infty},\ G \in \mathfrak{A}^\infty_{k+j} \right\},$$

$$\phi(j) = \sup_{k\in\mathbf{Z}} \sup \left\{ |P(G|F) - P(G)| : F \in \mathfrak{A}^k_{-\infty},\ G \in \mathfrak{A}^\infty_{k+j},\ P(F) > 0 \right\}.$$

If $\alpha(j)$ [$\phi(j)$] goes to zero as j approaches infinity, we call the process (ξ_t) α-mixing [ϕ-mixing].

Clearly, the mixing coefficients $\alpha(j)$ and $\phi(j)$ are measures of the memory of the process. Every ϕ-mixing process is α-mixing.[3] Simple examples of ϕ-mixing and hence α-mixing processes are independent and m-dependent processes. Other examples of α-mixing processes are strictly invertible Gaussian ARMA processes.[4] Under appropriate moment condi-

---

[2]In order to apply this definition to processes $(\xi_t)_{t\in\mathbf{N}}$ we can put ξ_s equal to some constant for $s \leq 0$.

[3]A reason for considering the subclass of ϕ-mixing processes separately is that limit theorems under the assumption of a ϕ-mixing process can sometimes be derived under weaker additional conditions than is possible under the assumption of an α-mixing process; cf. Theorem 6.4 for example.

[4]For conditions under which linear processes are α-mixing [ϕ-mixing] see, e.g., Ibragimov and Linnik (1971), Chanda (1974), Gorodetskii (1977), Withers (1981a), Pham and Tran (1985), and Mokkadem (1986).

tions α-mixing and ϕ-mixing processes satisfy LLNs and CLTs.

A useful feature of α-mixing [ϕ-mixing] processes is that measurable functions of finitely many elements of the process are themselves α-mixing [ϕ-mixing]. For example, if we are interested in LLNs for the "bracketing" functions $q_t^*(\mathbf{z}_t, \theta; \eta)$ and $q_{t*}(\mathbf{z}_t, \theta; \eta)$ and if $\mathbf{z}_t$ is α-mixing [ϕ-mixing], then $q_t^*(\mathbf{z}_t, \theta; \eta)$ and $q_{t*}(\mathbf{z}_t, \theta; \eta)$ are also α-mixing [ϕ-mixing], and they will hence satisfy LLNs under appropriate dominance conditions.

In many applications $\mathbf{z}_t$ will be composed of current endogenous and exogenous variables and finitely many lags thereof. Clearly, in this case $\mathbf{z}_t$ will be α-mixing [ϕ-mixing] if the process of endogenous and exogenous variables has that property. Now, in static models the dependent variables will typically be a function of a finite number of exogenous variables and disturbances. Consequently, if the process of exogenous variables and disturbances is α-mixing [ϕ-mixing] then also the process of the endogenous and exogenous variables will have this property. Hence for static models the assumption that $\mathbf{z}_t$ is α-mixing [ϕ-mixing] seems to be "problem adequate", as it can be reduced to mixing conditions on the exogenous variables and the disturbances only. However, in dynamic models the endogenous variables will typically depend on the infinite history of the exogenous variables and disturbances. Since α-mixing [ϕ-mixing] is not necessarily preserved under transformations which involve the infinite past, the endogenous variables need not be α-mixing [ϕ-mixing], even if the exogenous variables and disturbances have that property.[5] Consequently, within the context of dynamic nonlinear models the assumption that the process of endogenous and exogenous variables, and hence $\mathbf{z}_t$, is α-mixing [ϕ-mixing] does not seem to be "problem adequate". We note further that even if the process of endogenous and exogenous variables is α-mixing [ϕ-mixing], but $\mathbf{z}_t$ contains an increasing number of lagged values and hence effectively includes the entire past of the endogenous and exogenous variables, then $\mathbf{z}_t$ may not be α-mixing [ϕ-mixing].

The assumption that the data generating process is α-mixing [ϕ-mixing] has nevertheless been used widely in the econometrics literature, see, e.g., Domowitz and White (1982), White and Domowitz (1984), Bates and White (1985) and Domowitz (1985). The above discussion suggests that the assumption of an α-mixing [ϕ-mixing] data generating process, while not ruling out dynamic models, is not genuinely geared towards such models, since for dynamic models results that ensure that the process of endogenous variables is in fact α-mixing [ϕ-mixing] seem only to be available under conditions that are unnecessarily restrictive for an asymptotic the-

[5]It has been shown that even simple AR(1) processes with i.i.d. disturbances are not necessarily α-mixing or ϕ-mixing, see, e.g., Ibragimov and Linnik (1971), Chernick (1981), Andrews (1984), and Athreya and Pantula (1986a,b). Whether or not an AR(1) process with i.i.d. disturbances is α-mixing or ϕ-mixing depends crucially on properties of the distribution of the disturbances.

ory for dynamic nonlinear models.[6] The above discussion suggests further that within the context of dynamic nonlinear models we will typically be confronted with processes $(\mathbf{z}_t)$ (or functions thereof) which effectively depend on the infinite past of some basis process, say $(\mathbf{e}_t)$. That is, we will be confronted with establishing limit theorems for functions of the form $h_t(\mathbf{e}_t, \mathbf{e}_{t-1}, \ldots)$. The process $(\mathbf{e}_t)$ may correspond directly to the process of the exogenous variables and disturbances, but may, e.g., also correspond to innovations that generate the exogenous variables and disturbances. For the following discussion in this chapter it is irrelevant which particular interpretation is given to the basis process $(\mathbf{e}_t)$. In rational expectations models we may be confronted more generally with functions of the form $h_t(\ldots, \mathbf{e}_{t+1}, \mathbf{e}_t, \mathbf{e}_{t-1}, \ldots)$. The formal analysis in the subsequent sections will also cover this case.

Intuitively speaking, we expect a LLN (and also a CLT) for functions $h_t(\mathbf{e}_t, \mathbf{e}_{t-1}, \ldots)$ to hold, even if $h_t(\mathbf{e}_t, \mathbf{e}_{t-1}, \ldots)$ is not α-mixing or ϕ-mixing, as long as the process $(\mathbf{e}_t)$ is sufficiently "mixing" and the functions h_t are such that they put "declining weights" on high lags of $\mathbf{e}_t$, thus ensuring a "fading" memory of the process $h_t(\mathbf{e}_t, \mathbf{e}_{t-1}, \ldots)$. The approach taken in Bierens (1981, 1982a, 1984, 1987), Wooldridge (1986), Gallant (1987a, Ch.7), Gallant and White (1988), and Pötscher and Prucha (1991a,b) towards the analysis of dynamic nonlinear models is similar in that they assume an α-mixing [ϕ-mixing] condition for the process $(\mathbf{e}_t)$ and in that they formalize the above notion of "declining weights" by approximating the functions $h_t(\mathbf{e}_t, \mathbf{e}_{t-1}, \ldots)$ by functions that only depend on finitely many elements of $(\mathbf{e}_t)$. The assumption that the basis process $(\mathbf{e}_t)$ is α-mixing [ϕ-mixing] seems not to be unreasonable, since $\mathbf{e}_t$ may effectively be interpreted as an innovation.

On an intuitive level the approximation concepts used in the papers by Bierens on the one hand and in the work of Gallant, White and Wooldridge on the other hand seem closely related. However on a technical level the differences and similarities are less than obvious and had not been explored in the literature prior to Pötscher and Prucha (1991a). In the following section we will discuss the differences and similarities in detail.

6.2 Approximation Concepts

In the following we first define formally the approximation concept mentioned above.

---

[6]Cf. Footnote 5. For positive results see Mokkadem (1987), Doukhan (1994) and the references given in Footnote 4. These results typically make assumptions on the distribution of the disturbance process and/or a Markovian assumption.

Definition 6.2. *Let* $(\mathbf{v}_t)_{t\in\mathbf{N}}$ *and* $(\mathbf{e}_t)_{t\in\mathbf{Z}}$ *be stochastic processes defined on* $(\Omega, \mathfrak{A}, P)$ *that take their values in* $\mathbf{R}^{p_v}$ *and* $\mathbf{R}^{p_e}$*, respectively. Then the process* $(\mathbf{v}_t)$ *is called*

(a) L_0*-approximable by the basis process* $(\mathbf{e}_t)$ *if there exist measurable functions* $h_t^m : \mathbf{R}^{(2m+1)p_e} \to \mathbf{R}^{p_v}$ *such that for every* $\delta > 0$ *we have*

$$\limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} P\left(|\mathbf{v}_t - h_t^m(\mathbf{e}_{t+m},\ldots,\mathbf{e}_{t-m})| > \delta\right) \to 0 \text{ as } m \to \infty,$$

(b) L_p*-approximable,* $0 < p < \infty$*, by the basis process* $(\mathbf{e}_t)$ *if there exist measurable functions* $h_t^m : \mathbf{R}^{(2m+1)p_e} \to \mathbf{R}^{p_v}$ *such that*

$$\limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} \left\|\mathbf{v}_t - h_t^m(\mathbf{e}_{t+m},\ldots,\mathbf{e}_{t-m})\right\|_p \to 0 \text{ as } m \to \infty.^7$$

For ease of notation we shall often write $\mathbf{h}_t^m$ for $h_t^m(\mathbf{e}_{t+m},\ldots,\mathbf{e}_{t-m})$ and we shall refer to $\mathbf{h}_t^m$ as an L_p-approximator for $\mathbf{v}_t$. Clearly, if $\mathbf{v}_t$ has second moments, then the conditional mean is the best approximator in the L_2-norm. Hence, we can then choose $\mathbf{h}_t^m = E(\mathbf{v}_t|\mathbf{e}_{t+m},\ldots,\mathbf{e}_{t-m})$ without loss of generality in the definition of L_2-approximability. Furthermore, as shown below, the conditional mean can be chosen without loss of generality as the approximator in the definition of L_p-approximability even for $p \neq 2$, given a suitable moment condition is satisfied. However, especially when considering nonlinear transformations of $\mathbf{v}_t$, it turns out to be convenient to also allow for approximators other than the conditional mean. (Also it may sometimes be easier to find approximators that are different from the conditional mean by ad hoc arguments.)

The approximation concept introduced in Definition 6.2 is in the spirit of similar concepts which have been used in the probability literature for the derivation of limit theorems for dependent processes by, e.g., Blum and Rosenblatt (1957), Ibragimov (1962), Billingsley (1968) and McLeish (1975a,b). The concept of L_0-approximability is (up to inessential details) identical to Bierens' (1981) concept of stochastic stability. Within the context of stationary processes the concept of L_2-approximability was used by Billingsley (1968) and Bierens (1983); the latter author refers to this concept as "ν-stability in L_2".

---

[7] For a random variable ξ taking its values in Euclidean space we define for $0 < p < \infty : \|\xi\|_p = [\int |\xi|^p\, dP]^{1/p}$ where $|.|$ is the Euclidean norm. Furthermore, $\|\xi\|_\infty$ denotes the essential supremum of $|\xi|$. Of course, $\|\xi\|_p$ is a norm only for $p \geq 1$. We have chosen the term "L_0-approximable" for the property introduced in part (a) of the definition, since the space of all P-equivalence classes of measurable real valued functions endowed with the topology of convergence in probability is sometimes denoted by $L_0(\Omega, \mathfrak{A}, P)$ and can be viewed as a "natural limit" of the spaces $L_p(\Omega, \mathfrak{A}, P)$ for $p \to 0$.

Based on McLeish (1975a,b) the following concept of near epoch dependence is used in Wooldridge (1986), Gallant (1987a, Ch.7) and Gallant and White (1988).

Definition 6.3. *Let* $(\mathbf{v}_t)_{t\in\mathbf{N}}$ *and* $(\mathbf{e}_t)_{t\in\mathbf{Z}}$ *be stochastic processes defined on* $(\Omega, \mathfrak{A}, P)$ *that take their values in* $\mathbf{R}^{p_v}$ *and* $\mathbf{R}^{p_e}$*, respectively. Then the process* $(\mathbf{v}_t)$ *is called near epoch dependent of size* $-q$ *on the basis process* $(\mathbf{e}_t)$*, if the sequence*

$$\nu_m = \sup_t \|\mathbf{v}_t - E(\mathbf{v}_t|\mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m})\|_2$$

is of size $-q$*,* $q > 0$*.*[8,9]

As can be seen immediately from the above definitions, near epoch dependence implies L_2-approximability. Near epoch dependence is more stringent than L_2-approximability in two respects. First, for near epoch dependence the quantities $\|\mathbf{v}_t - E(\mathbf{v}_t|\mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m})\|_2$ themselves have to become small, while for L_2-approximability this is only required from the Cesàro-sums of these quantities. Second, and more importantly, in the definition of near epoch dependence the quantities $\|\mathbf{v}_t - E(\mathbf{v}_t|\mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m})\|_2$ are assumed to decline at a certain rate (uniformly in t).

Andrews (1988) generalized the concept of near epoch dependence. He calls a process L_p-near epoch dependent, $1 \le p \le 2$, if there exist sequences of constants (d_t) and (ν_m) such that

$$\|\mathbf{v}_t - E(\mathbf{v}_t|\mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m})\|_p \le d_t \nu_m$$

with

$$\sup_n n^{-1} \sum_{t=1}^{n} d_t < \infty$$

and $\nu_m \to 0$ as $m \to \infty$. No rate of convergence is specified for the sequence (ν_m). Clearly any L_p-near epoch dependent process is also L_p-approximable.

The above discussion shows that the approximation concepts employed in the econometrics literature on dynamic nonlinear models, i.e., stochastic stability, ν-stability in L_2, near epoch dependence as well as L_p-near

---

[8] We define, as is common in the econometrics literature, a sequence (ν_m) to be of size $-q$, $q > 0$, if ν_m is $O(m^{-\lambda})$ for some $\lambda > q$. The original definition given in McLeish (1975a) is actually slightly more general.

[9] Of course, we could have defined near epoch dependence of $(\mathbf{v}_t)$ on $(\mathbf{e}_t)$ by the requirement that $\sup_t \|\mathbf{v}_t - h_t^m(\mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m})\|_2$ is of size $-q$, thus paralleling the definition of L_p-approximability. However, it seems that the definition as given here is more commonly used in the literature. In any case, if second moments of $\mathbf{v}_t$ exist, then both versions coincide in view of the minimum mean square error property of the conditional expectation.

epoch dependence, can all be viewed as special cases of the concept of L_p-approximability. It also turns out that L_p-approximability is all that is needed for a weak LLN, given mixing properties of the basis process, see Section 6.3.

The following theorem shows that the concepts of L_p-approximability for different values of p are equivalent under a suitable moment condition.

Theorem 6.1. *(a) Suppose* $(\mathbf{v}_t)$ *is* $L_{p^\bullet}$*-approximable by* $(\mathbf{e}_t)$ *for some* $0 \leq p^\bullet < \infty$*, then* $(\mathbf{v}_t)$ *is also* L_p*-approximable by* $(\mathbf{e}_t)$ *for any* p *with* $0 \leq p \leq p^\bullet$*. (In fact, any* $L_{p^\bullet}$*-approximator* $\mathbf{h}_t^m$ *is also an* L_p*-approximator for* $0 \leq p \leq p^\bullet$*.)*

(b) Suppose $(\mathbf{v}_t)$ *is* L_0*-approximable by* $(\mathbf{e}_t)$ *and*

$$\sup_n n^{-1} \sum_{t=1}^{n} \|\mathbf{v}_t\|_{p^\bullet}^{1+\gamma} < \infty$$

for some $0 < p^\bullet < \infty$ *and* $\gamma > 0$*, then the process* $(\mathbf{v}_t)$ *is* L_p*-approximable by* $(\mathbf{e}_t)$ *for any* p*,* $0 \leq p < p^\bullet$*.*

(c) Suppose the assumptions of (b) hold with $p^\bullet > 1$*, then the conditional mean* $E(\mathbf{v}_t|\mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m})$ *is an* L_p*-approximator for any* p*,* $0 \leq p < p^\bullet$*.*

Clearly,

$$\sup_t E\,|\mathbf{v}_t|^{p^\bullet} < \infty$$

is a sufficient condition for the moment condition in parts (b) and (c) of the above theorem; furthermore,

$$\sup_n n^{-1} \sum_{t=1}^{n} E\,|\mathbf{v}_t|^{p^\bullet} < \infty$$

is sufficient if $p^\bullet > 1$ (which follows from Lyapunov's inequality and by choosing $\gamma > 0$ such that $1 + \gamma < p^\bullet$). We note that

$$\sup_n n^{-1} \sum_{t=1}^{n} E\,|\mathbf{v}_t|^{p^\bullet} < \infty$$

for $p^\bullet > 1$ is a typical moment condition in LLNs for non-identically distributed processes, cf. Section 6.3.

Theorem 6.1 helps to clarify the relationship between the above discussed approximation concepts used in the econometrics literature on dynamic nonlinear models. These approximation concepts measure the approximation error in different L_p-norms or in terms of probabilities. Part (a) of the theorem states the obvious fact that the stringency of the approximation concepts increases with p. Part (b) of the theorem states the less obvious fact that under a suitable moment condition the differences in the approximation concepts pertaining to the chosen distance measures $\|.\|_p$ vanish. As a consequence we see that, given

$$\sup_n n^{-1} \sum_{t=1}^{n} \|\mathbf{v}_t\|_{2+\epsilon}^{1+\gamma} < \infty$$

for some $\epsilon > 0$ and $\gamma > 0$, Bierens' concept of stochastic stability and the concept of near epoch dependence used by Gallant, McLeish, White and Wooldridge essentially differ only in that the latter concept prescribes a rate of decline for the approximation errors.

Part (c) of Theorem 6.1 tells us that for L_p-approximable processes, $0 \leq p < \infty$, the conditional mean is always an L_p-approximator provided the moment condition holds for a $p^\bullet > 1$ and $p^\bullet > p$. Of course, part (a) of Theorem 6.1 already implies that the conditional mean is an L_p-approximator for $0 \leq p \leq 2$ given the process is L_2-approximable.

6.3 Laws of Large Numbers for L_p-Approximable and Near Epoch Dependent Processes

The following theorem describes a basic strategy for deriving LLNs for L_p-approximable processes.

Theorem 6.2. *Let* $(\mathbf{v}_t)$ *be* L_1*-approximable by* $(\mathbf{e}_t)$.

(a) Suppose there exist L_1*-approximators* $(\mathbf{h}_t^m)$ *that satisfy a weak LLN for each* $m \in \mathrm{N}$, *then also* $(\mathbf{v}_t)$ *satisfies a weak LLN.*[10]

(b) Suppose that

$$\sup_n n^{-1} \sum_{t=1}^{n} E\,|\mathbf{v}_t|^{1+\epsilon} < \infty$$

for some $\epsilon > 0$, *then there exist* L_1*-approximators* $(\mathbf{h}_t^m)$ *which, for each* $m \in \mathrm{N}$, *are bounded in absolute value uniformly in* t.

Part (a) of the above theorem shows that in order to prove a LLN for an L_1-approximable process it suffices to prove a LLN for the approximators. LLNs for $\mathbf{h}_t^m = h_t^m(\mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m})$ are typically deduced from mixing conditions on the basis process $(\mathbf{e}_t)$ and moment conditions on $\mathbf{h}_t^m$. By design $\mathbf{h}_t^m$ only depends on finitely many elements of $(\mathbf{e}_t)$. Hence mixing

[10] Recall that the assumption that a process, say, (ξ_t) satisfies a LLN implicitly maintains that $E\,|\xi_t| < \infty$.

conditions on $(\mathbf{e}_t)$ will typically carry over to the process $(\mathbf{h}_t^m)$; e.g., if $(\mathbf{e}_t)$ is α-mixing, then also $(\mathbf{h}_t^m)$ is α-mixing. The moment condition

$$\sup_n n^{-1} \sum_{t=1}^{n} E \left|\mathbf{v}_t\right|^{1+\epsilon} < \infty$$

for some $\epsilon > 0$ in part (b) of the above theorem is common in LLNs for non-identically distributed processes. Part (b) of the theorem tells us that under such a moment condition we can choose bounded approximators $\mathbf{h}_t^m$. This is convenient since in proving a LLN for these $\mathbf{h}_t^m$ we do not have to worry about moment conditions for $\mathbf{h}_t^m$. Of course, a general sufficient condition for $(\mathbf{h}_t^m)$ to satisfy a weak LLN is that the variance covariance matrix of $n^{-1}\sum_{t=1}^{n} \mathbf{h}_t^m$ goes to zero as $n \to \infty$. Given boundedness of $\mathbf{h}_t^m$ (uniformly in t) this is clearly implied by

$$n^{-2} \sum_{1 \le t < s \le n} \operatorname{cov}(\mathbf{h}_t^m, \mathbf{h}_s^m) \to 0 \text{ as } n \to \infty,$$

which essentially means that the covariance function dies out.[11]

As a corollary to Theorem 6.2 we obtain, e.g., the following weak LLN in case the basis process is α-mixing. Since every ϕ-mixing process is α-mixing, the result automatically also applies to the case of ϕ-mixing basis processes. As suggested by the above discussion the proof proceeds by verifying that

$$n^{-2} \sum_{1 \le t < s \le n} \operatorname{cov}(\mathbf{h}_t^m, \mathbf{h}_s^m) \to 0 \text{ as } n \to \infty$$

from the mixing properties of $\mathbf{h}_t^m$ (which are implied by the mixing properties of the basis process).

Theorem 6.3. *Suppose* $\sup_n n^{-1}\sum_{t=1}^{n} E\left|\mathbf{v}_t\right|^{1+\epsilon} < \infty$ *for some* $\epsilon > 0$. *Let* $(\mathbf{v}_t)$ *be* L_0*-approximable by* $(\mathbf{e}_t)$ *and let* $(\mathbf{e}_t)$ *be* α*-mixing, then* $(\mathbf{v}_t)$ *satisfies a weak LLN.*

The above LLN is slightly more general than a LLN introduced by Andrews (1988) for L_1-near epoch dependent processes. The above LLN also generalizes a LLN by Bierens (1982a, 1984) in that here the process $(\mathbf{v}_t)$ is not assumed to be asymptotically stationary. Analogously to Bierens (1982a, 1984) the α-mixing condition, i.e., $\alpha(j) \to 0$ as $j \to \infty$, can be

[11] If one puts $\mathbf{e}_t = \mathbf{v}_t$ in Theorem 6.2 then the approximators $\mathbf{h}_t^m$ in part (b) of the theorem become truncated versions of $\mathbf{v}_t$. Hence Theorem 6.2 can be viewed as a generalization of standard proof techniques for weak LLNs in which one first truncates the random variables involved and then shows that the covariance function of the truncated random variables dies out.

slightly weakened. For reasons of simplicity we do not present this generalization here. As a point of interest we note further that in light of the moment condition in the above theorem the condition of L_0-approximability maintained in this theorem is in fact equivalent to L_1-approximability and hence only seemingly weaker, cf. Theorem 6.1. Hence the above LLN as well as Bierens' LLNs for stochastically stable processes are effectively based on L_1-approximability.

Andrews (1988) uses a different approach based on L_p-mixingales to obtain his version of the above LLN. In fact, Andrews derives his version of the above LLN as a special case of a LLN for L_p-mixingales. In the following definition let $(\xi_t)_{t\in\mathbf{Z}}$ be a sequence of random variables on $(\Omega, \mathfrak{A}, P)$ taking their values in Euclidean space and let $(\mathfrak{A}_t)_{t\in\mathbf{Z}}$ be a nondecreasing sequence of sub-σ-fields of $\mathfrak{A}$. In the subsequent discussion the following definition will be applied to the centered process $\xi_t = \mathbf{v}_t - E(\mathbf{v}_t)$ (where we can put $\mathbf{v}_t$ equal to zero for $t \leq 0$); the sub-σ-fields $\mathfrak{A}_t$ will typically be taken as the σ-fields generated by $\{\mathbf{e}_t, \mathbf{e}_{t-1}, \ldots\}$.

Definition 6.4. *The sequence $(\xi_t, \mathfrak{A}_t)$ is an L_p-mixingale, $1 \leq p \leq 2$, if there exist constants $c_t \geq 0$, $t \geq 1$, and $\psi_m \geq 0$, $m \geq 0$, such that $\psi_m \to 0$ as $m \to \infty$ and for all $t \geq 1$ and $m \geq 0$ we have*

$$\left\| E(\xi_t | \mathfrak{A}_{t-m}) \right\|_p \leq c_t \psi_m \tag{6.1}$$

and

$$\left\| \xi_t - E(\xi_t | \mathfrak{A}_{t+m}) \right\|_p \leq c_t \psi_{m+1}. \tag{6.2}$$

L_2-mixingales were introduced in McLeish (1975a) under the name of mixingales. The generalized notion of L_p-mixingales given in the above definition was introduced by Andrews (1988). Clearly every L_p-mixingale has zero mean. The constants c_t are typically measures of the magnitude of the random variables ξ_t, e.g., $c_t = \left\| \xi_t \right\|_p$. For limit theorems boundedness conditions as, e.g.,

$$\sup_n n^{-1} \sum_{t=1}^{n} c_t < \infty \tag{6.3}$$

are typically assumed. In abuse of terminology we shall also call $\mathbf{v}_t$ an L_p-mixingale if $\xi_t = \mathbf{v}_t - E(\mathbf{v}_t)$ is an L_p-mixingale.

Andrews' approach for proving his version of Theorem 6.3 parallels the approach in McLeish (1975a) for strong LLNs and can now be described as follows: He first derives a weak LLN for L_1-mixingales. He then shows that any process $(\mathbf{v}_t)$ that is L_1-near epoch dependent on some α-mixing basis process and that satisfies

$$\sup_t E\left|\mathbf{v}_t\right|^{1+\epsilon} < \infty$$

for some $\epsilon > 0$ is an L_1-mixingale, with $c_t = \|\mathbf{v}_t - E(\mathbf{v}_t)\|_1 \leq 2\|\mathbf{v}_t\|_1$ satisfying (6.3), and hence the weak LLN for L_1-mixingales applies. Inspection of Andrews' proof also shows that his approach can be generalized to provide an alternative proof of Theorem 6.3 as it stands. The modifications in Andrews' proof involve that (i) L_1-near epoch dependence is weakened to L_1-approximability, (ii) the moment condition

$$\sup_t E|\mathbf{v}_t|^{1+\epsilon} < \infty$$

is weakened to

$$\sup_n n^{-1} \sum_{t=1}^{n} E|\mathbf{v}_t|^{1+\epsilon} < \infty,$$

and (iii) the L_1-mixingale conditions (6.1) and (6.2) as well as the boundedness condition (6.3) are weakened to the following conditions with $p = 1$:

$$\limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} \|E(\xi_t|\mathfrak{A}_{t-m})\|_p \to 0 \text{ as } m \to \infty, \tag{6.4}$$

and

$$\limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} \|\xi_t - E(\xi_t|\mathfrak{A}_{t+m})\|_p \to 0 \text{ as } m \to \infty.\text{[12]} \tag{6.5}$$

This minor modification of Andrews' approach as outlined above shows also – apart from providing an alternative proof of Theorem 6.3 via mixingales – that the concept of a process which is L_p-approximable w.r.t. an α-mixing base fits nicely into the generalized framework of L_p-mixingales expressed in (6.4) and (6.5). The splitting of the bounds on the r.h.s. of (6.1) and (6.2) into c_t and ψ_m seems unnecessary in order to derive weak LLNs. Of course, the proof of Theorem 6.3 given in Appendix D shows that the result can be obtained without resorting to the theory of mixingales.

The consistency results in Gallant (1987a, Ch.7) and Gallant and White (1988) are based on the following strong LLN for processes that are near epoch dependent on an α-mixing or ϕ-mixing basis process. This LLN is derived in Theorem 3.1 in McLeish (1975a).

Theorem 6.4. *Let* $(\mathbf{v}_t)$ *be near epoch dependent of size* $-1/2$ *on* $(\mathbf{e}_t)$ *with*

$$\sum_{t=1}^{\infty} \|\mathbf{v}_t\|_r^2 / t^2 < \infty$$

[12] Andrews (1988) proves his weak LLN for L_1-mixingales under conditions (6.1)-(6.3) and uniform integrability. Inspection of the proof shows that the weak LLN also holds under (6.4), (6.5) and uniform integrability, or under (6.4), (6.5) and $\sup_n n^{-1} \sum_{t=1}^{n} E|\mathbf{v}_t|^{1+\epsilon} < \infty$.

for some $r \geq 2$. Suppose $(\mathbf{e}_t)$ is α-mixing with mixing coefficients of size $-r/(r-2)$ and $r > 2$ holds, or ϕ-mixing with mixing coefficients of size $-r/(2r-2)$, then $(\mathbf{v}_t)$ satisfies a strong LLN.

Of course, a sufficient condition for

$$\sum_{t=1}^{\infty} \|\mathbf{v}_t\|_r^2 / t^2 < \infty$$

is

$$\sup_t E |\mathbf{v}_t|^r < \infty.$$

In proving the above theorem McLeish first shows that any process $(\mathbf{v}_t)$ that is near epoch dependent on some α-mixing or ϕ-mixing basis process and that satisfies the above moment condition is an L_2-mixingale with mixingale coefficients ψ_m of size $-1/2$, if the approximation errors ν_m are of size $-1/2$ and the α-mixing or ϕ-mixing coefficients are of the size as given in the theorem. Then he applies his strong LLN for L_2-mixingales.

In comparing Theorems 6.3 and 6.4 we note that the latter theorem gives a strong (as opposed to a weak) LLN at the expense of stronger assumptions: Theorem 6.4 requires rates of convergence for the mixing coefficients $\alpha(j)$ and $\phi(j)$ as well as for the approximation error ν_m, while such conditions are not postulated in Theorem 6.3. Furthermore Theorem 6.4 requires the existence of at least second moments of the process $(\mathbf{v}_t)$ while Theorem 6.3 only requires the existence of a moment slightly higher than the first. For further LLNs for L_p-near epoch dependent processes and L_p-mixingales see Hansen (1991), De Jong (1995a), and Davidson and De Jong (1995).

6.4 Preservation of Approximation Concepts under Transformation

As already mentioned at the beginning of this chapter one is frequently confronted with the need to establish limit theorems for functions of an L_p-approximable or near epoch dependent process. For example, as discussed in Chapters 3 and 5, in the course of proving consistency of an estimator we may want to establish a ULLN for functions $q_t(\mathbf{z}_t, \theta)$. To this end we need to establish LLNs for the "bracketing" functions $q_t^*(\mathbf{z}_t, \theta; \eta)$ and $q_{t*}(\mathbf{z}_t, \theta; \eta)$. In principle we could obtain these LLNs by simply assuming that $q_t^*(\mathbf{z}_t, \theta; \eta)$ and $q_{t*}(\mathbf{z}_t, \theta; \eta)$ satisfy the conditions placed on $(\mathbf{v}_t)$ in Theorems 6.3 or 6.4. However, it seems of interest to explore more basic conditions under which this will be the case. Given $(\mathbf{z}_t)$ is L_p-approximable by [near epoch dependent on] a mixing basis process, such more basic conditions then essentially amount to conditions under which L_p-approximability [near epoch dependence] is preserved under transformations.

In the following we hence consider general conditions under which transformations of L_p-approximable [near epoch dependent] processes are L_p-approximable [near epoch dependent], see Theorems 6.5 - 6.7 and Corollary 6.8.[13] Combining these transformation results with Theorems 6.3 and 6.4 will then give LLNs for functions of L_p-approximable [near epoch dependent] processes. Similarly, combining the transformation results with the results in Chapter 10 will provide CLTs for such processes.

The transformation results just discussed start from the assumption that $(\mathbf{z}_t)$ is L_p-approximable [near epoch dependent]. Typically (but not necessarily) $\mathbf{z}_t$ is composed of endogenous and exogenous variables. This now raises the question for which classes of dynamic systems $(\mathbf{z}_t)$ will actually possess the L_p-approximability [near epoch dependence] property. Hence in this section we also give conditions for a dynamic system under which the process of endogenous variables, i.e., the output process, can be shown to be L_p-approximable by [near epoch dependent on] a mixing basis process if the process of exogenous variables and disturbances, i.e., the input process, has the analogous property, see Theorems 6.10 - 6.12. Combining these theorems with Lemma 6.9 then gives conditions for the L_p-approximability [near epoch dependence] of $(\mathbf{z}_t)$. These results are important for developing an asymptotic theory for dynamic nonlinear models in that they enable us to apply the established machinery of limit theorems for L_p-approximable [near epoch dependent] processes to the output process of a dynamic nonlinear model.

We start with the process $(\mathbf{v}_t)$ defined on $(\Omega, \mathfrak{A}, P)$ that takes its values in a Borel subset V of $\mathbf{R}^{p_v}$.[14] By allowing $(\mathbf{v}_t)$ to possibly take its values in a proper subset of $\mathbf{R}^{p_v}$ we can cover situations where the transforming functions are defined only on that subset. *We emphasize that in the following V should be viewed as a metric space in its own right with the induced Euclidean metric.* We use the notation $\bar{H}_n^v = n^{-1} \sum_{t=1}^n H_t^v$ where H_t^v is the distribution of $\mathbf{v}_t$ on V.

Assumption 6.1. *The family of functions $\{g_t : t \in \mathbf{N}\}$, where $g_t : V \rightarrow \mathbf{R}$, is equicontinuous on V.*

Theorem 6.5. *Suppose $(\mathbf{v}_t)$ is L_0-approximable by the basis process $(\mathbf{e}_t)$.*

---

[13]The reason for working with L_p-approximable [near epoch dependent] processes rather than with L_p-mixingales is that the class of L_p-mixingales is not closed under many nonlinear transformations.

[14]Since V is a Borel subset, $\mathbf{v}_t$ can equally well be regarded as a measurable function taking its values in $\mathbf{R}^{p_v}$ and hence the theory in the preceding sections applies.

Suppose $\{\bar{H}_n^v : n \in \mathbf{N}\}$ *is tight on* V *and that Assumption 6.1 holds. Then* $(g_t(\mathbf{v}_t))$ *is* L_0*-approximable by* $(\mathbf{e}_t)$. *If, additionally,*

$$\sup_n n^{-1} \sum_{t=1}^{n} \|g_t(\mathbf{v}_t)\|_{p^\bullet}^{1+\gamma} < \infty$$

for some $0 < p^\bullet < \infty$ *and some* $\gamma > 0$, *then* $(g_t(\mathbf{v}_t))$ *is* L_p*-approximable by* $(\mathbf{e}_t)$ *for* $0 \leq p < p^\bullet$.

Of course, if V is a closed subset of $\mathbf{R}^{p_v}$ then the above tightness condition is implied by, e.g., a mild moment condition like

$$\sup_n n^{-1} \sum_{t=1}^{n} E\,|\mathbf{v}_t|^\gamma < \infty$$

for some $\gamma > 0$, cf. the related discussion in Chapter 5. We note further that the theorem holds if the tightness condition is dropped but equicontinuity is replaced by uniform equicontinuity in Assumption 6.1, cf. the remark after Lemma D4 in Appendix D. Furthermore, Theorem 6.5 can be generalized to discontinuous functions, provided the set of "discontinuity" points is small, see Lemma D4 for details.

Theorem 6.5 and the subsequent preservation results given in Theorems 6.6, 6.7 and Corollary 6.8 are formulated for real valued functions g_t. The more general case of functions taking values in $\mathbf{R}^k$, $k \geq 1$, is immediately reduced to this case by applying the theorems to the components of the functions and by appealing to Lemma 6.9.

The following Lipschitz-type assumption has been used by Gallant (1987a, Ch.7) and Gallant and White (1988) in order to establish near epoch dependence of functions of near epoch dependent processes. Before giving the result on the transformation of near epoch dependent processes we first use this Lipschitz-type assumption for a further result on the transformation of L_p-approximable processes.

Assumption 6.2. [15] *The family of Borel measurable functions* $\{g_t : t \in \mathbf{N}\}$, *where* $g_t : V \to \mathbf{R}$, *is such that*

$$|g_t(v) - g_t(v^\bullet)| \leq B_t(v, v^\bullet)\,|v - v^\bullet|$$

for all $(v, v^\bullet) \in V \times V$ *and* $B_t : V \times V \to [0, \infty)$ *is Borel measurable.*

[15] Of course, without imposing further conditions on B_t, Assumption 6.2 does not restrict the class of Borel measurable functions g_t. However, in the following additional conditions will be placed on B_t.

Theorem 6.6. *(a) Suppose* $(\mathbf{v}_t)$ *is* L_0*-approximable by the basis process* $(\mathbf{e}_t)$. *Suppose Assumption 6.2 holds with*

$$\limsup_{m\to\infty} \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} E\left[B_t(\mathbf{v}_t, \mathbf{h}_t^m)^{\epsilon}\right] < \infty$$

for some $\epsilon > 0$ *and for some* L_0*-approximators* $\mathbf{h}_t^m$ *of* $(\mathbf{v}_t)$, *where the approximators take their values in* V. *Then* $(g_t(\mathbf{v}_t))$ *is* L_0*-approximable by* $(\mathbf{e}_t)$. *If, additionally,*

$$\sup_n n^{-1} \sum_{t=1}^{n} \|g_t(\mathbf{v}_t)\|_{p^\bullet}^{1+\gamma} < \infty$$

for some $0 < p^\bullet < \infty$ *and some* $\gamma > 0$, *then* $(g_t(\mathbf{v}_t))$ *is* L_p*-approximable by* $(\mathbf{e}_t)$ *for* $0 \leq p < p^\bullet$.

(b) Suppose $(\mathbf{v}_t)$ *is* L_p*-approximable by the basis process* $(\mathbf{e}_t)$ *for some* $0 \leq p < \infty$. *Suppose Assumption 6.2 holds with* $B_t(v, v^\bullet)$ *equal to some constant* $c < \infty$ *for all* $(v, v^\bullet) \in V \times V$ *and all* $t \in \mathbf{N}$. *Then* $(g_t(\mathbf{v}_t))$ *is* L_p*-approximable by* $(\mathbf{e}_t)$.

Of course, since every L_s-approximable process, $0 \leq s < \infty$, is L_0-approximable (and any L_s-approximator is an L_0-approximator) Theorems 6.5 and 6.6(a) also give conditions under which functions of L_s-approximable processes are L_p-approximable, $0 \leq p < p^\bullet$, by the same basis process. We note furthermore that the results in Theorems 6.5 and 6.6 contain Theorems 5.1.4 and 5.1.5 in Bierens (1981) for differentiable transformations that are independent of t as special cases.

The following theorem provides a preservation result for near epoch dependence. Contrary to L_p-approximability, this concept involves a rate of decline for the approximation errors. In order to be able to infer a rate of decline for the approximation errors of $g_t(\mathbf{v}_t)$ from the rate of decline of the approximation errors of $\mathbf{v}_t$ it seems necessary to place some Lipschitz-type condition, as, e.g., expressed in Assumption 6.2, (plus a moment condition on the Lipschitz bound) on the transforming function g_t. Continuity alone, as expressed in Assumption 6.1, is certainly not sufficient. Part (a) of the following result is given in Gallant (1987a, pp.498-499) for the case $V = \mathbf{R}^{p_v}$ only; cf. also Gallant and White (1988), Theorem 4.2.

Theorem 6.7. *Let* $\mathbf{h}_t^m$ *be some* L_s*-approximators of* $(\mathbf{v}_t)$ *based on* $(\mathbf{e}_t)$ *for some* $s \geq 1$, *where the approximators take their values in* V *and the sequence of approximation errors* $\sup_t \|\mathbf{v}_t - \mathbf{h}_t^m\|_s$ *is of size* $-q$ *for some* $q > 0$. *Suppose Assumption 6.2 holds and that* $E\left|g_t(\mathbf{v}_t)\right|^2 < \infty$ *for all* $t \in \mathbf{N}$.

(a) Assume that

$$\sup_m \sup_t \|B_t(\mathbf{v}_t, \mathbf{h}_t^m)\|_{s/(s-1)} < \infty$$

and

$$\sup_m \sup_t \|B_t(\mathbf{v}_t, \mathbf{h}_t^m)\,|\mathbf{v}_t - \mathbf{h}_t^m|\,\|_{s^\bullet} < \infty$$

for some $s^\bullet > 2$, *then* $(g_t(\mathbf{v}_t))$ *is near epoch dependent of size* $-(q/2)(s^\bullet - 2)/(s^\bullet - 1)$ *on* $(\mathbf{e}_t)$.

(b) Assume that $B_t(v, v^\bullet)$ *is equal to some constant* $c < \infty$ *for all* $(v, v^\bullet) \in V \times V$ *and all* $t \in \mathbf{N}$. *Then* $\sup_t \|g_t(\mathbf{v}_t) - g_t(\mathbf{h}_t^m)\|_s$ *is of size* $-q$.

The following corollary is given in Lemma 1 in Andrews (1991a).[16]

Corollary 6.8. *Suppose Assumption 6.2 holds with* $B_t(v, v^\bullet)$ *equal to some constant* $c < \infty$ *for all* $(v, v^\bullet) \in V \times V$ *and all* $t \in \mathbf{N}$, *and that* $E\,|g_t(\mathbf{v}_t)|^2 < \infty$ *for all* $t \in \mathbf{N}$. *If* $(\mathbf{v}_t)$ *is near epoch dependent on* $(\mathbf{e}_t)$ *of size* $-q$, *then* $(g_t(\mathbf{v}_t))$ *is near epoch dependent on* $(\mathbf{e}_t)$ *of size* $-q$.

In part (a) of Theorem 6.7 there are trade-offs between moment requirements (expressed by $s^\bullet$) and the loss in the speed – due to the transformation – by which the approximation errors go to zero. Even if arbitrarily high moments are assumed (i.e., $s^\bullet$ arbitrarily large), the guaranteed rate of decline of the approximation errors for $(g_t(\mathbf{v}_t))$ is at most half of that of the approximation errors for $(\mathbf{v}_t)$.

Theorems 6.6(a) and 6.7 explicitly assume that there exist approximators that take their values only in V. We show in Lemma D3 in Appendix D that given there exist L_p-approximators that take their values in $\mathbf{R}^{p_v}$, there also exist L_p-approximators that take their values in V. Hence the assumption concerning the range of the approximators in Theorem 6.6(a) is not restrictive. In the remark following Lemma D3 in Appendix D we show furthermore that given there exist L_p-approximators that take their values in $\mathbf{R}^{p_v}$ with approximation errors $\sup_t \|\mathbf{v}_t - \mathbf{h}_t^m\|_p$ of size $-q$, then there exist also L_p-approximators that take their values in V with approximation errors of the same size. Hence the assumption concerning the range of the approximators in Theorem 6.7 is also not restrictive. Of course, if the

[16]The proof of Lemma 1 in Andrews (1991a) is incorrect unless the set V is additionally assumed to be convex, since Andrews implicitly uses the argument that the conditional expectation of a V-valued random variable again takes its values in V. This, however, is in general only the case if V is convex. The proof of the corollary given in the appendix covers general sets V and thus restores the first part of Andrews' lemma.

original approximators are the conditional means, as, e.g., in the definition of near epoch dependence, then the modified approximators will typically not be conditional means (unless V is convex).

Comparing Theorem 6.5 on the one hand and Theorems 6.6(a) and 6.7(a) on the other hand it seems that the conditions involved in Theorem 6.5 are simpler to check than those of Theorems 6.6(a) and 6.7(a) in that the conditions in the former theorem do not involve an approximator $\mathbf{h}_t^m$. Theorems 6.6(b), 6.7(b) and Corollary 6.8 also provide relatively simple conditions under which L_p-approximability and near epoch dependence concepts are preserved under transformation, however, under the stronger assumption of uniformly bounded Lipschitz constants. Of course, combining Theorems 6.5 - 6.7 and Corollary 6.8 with Theorems 6.3 and 6.4, respectively, gives LLNs for functions $g_t(\mathbf{v}_t)$ of L_p-approximable or near epoch dependent processes. This will be discussed in more detail in Section 6.5 below.

In applications we may ask if, e.g., a stacked vector of current and lagged endogenous and exogenous variables is L_p-approximable [near epoch dependent] given the endogenous and exogenous variables are L_p-approximable [near epoch dependent]. In parts (a) and (a') of the following lemma it is shown that this is in fact the case. Parts (b) and (b') demonstrate that the components of L_p-approximable [near epoch dependent] processes are again L_p-approximable [near epoch dependent]. In parts (d) and (d') we consider situations where an L_p-approximable [near epoch dependent] process is sampled every k-th period. It is shown that the sampled process is L_p-approximable [near epoch dependent] with respect to an appropriately defined basis process. Parts (c) and (c') deal with the "inverse" situation where a process is built up from a set of L_p-approximable [near epoch dependent] processes. Again, it is shown that the new process is then also L_p-approximable [near epoch dependent].

Lemma 6.9. [17] *(a) Let $(\xi_t)_{t\in\mathbf{N}}$ and $(\eta_t)_{t\in\mathbf{N}}$ be L_p-approximable, for some $0 \le p < \infty$, by $(\mathbf{e}_t^\xi)_{t\in\mathbf{Z}}$ and $(\mathbf{e}_t^\eta)_{t\in\mathbf{Z}}$, respectively. Then*

$$\left((\xi_t', \dots, \xi_{t-l}', \eta_t', \dots, \eta_{t-k}')'\right)_{t\in\mathbf{N}}$$

is L_p-approximable by $((\mathbf{e}_t^{\xi\prime}, \mathbf{e}_t^{\eta\prime})')_{t\in\mathbf{Z}}$, where $\xi_0, \dots, \xi_{1-l}$ and $\eta_0, \dots, \eta_{1-k}$ are p-fold integrable random variables.

(a') Let $(\xi_t)_{t\in\mathbf{N}}$ and $(\eta_t)_{t\in\mathbf{N}}$ be quadratically integrable processes that are near epoch dependent of size $-q$, $q > 0$, on $(\mathbf{e}_t^\xi)_{t\in\mathbf{Z}}$ and $(\mathbf{e}_t^\eta)_{t\in\mathbf{Z}}$, respectively. Then

$$\left((\xi_t', \dots, \xi_{t-l}', \eta_t', \dots, \eta_{t-k}')'\right)_{t\in\mathbf{N}}$$

[17] The various stochastic processes in this lemma are assumed to be defined on a common probability space and to take their values in Euclidean spaces.

is near epoch dependent of size $-q$ *on* $((\tilde{\mathbf{e}}_t^{\xi\prime}, \tilde{\mathbf{e}}_t^{\eta\prime})')_{t\in\mathbf{Z}}$, *where* $\xi_0, \ldots, \xi_{1-l}$ *and* $\eta_0, \ldots, \eta_{1-k}$ *are quadratically integrable random variables, and*

$$\tilde{\mathbf{e}}_t^{\xi} = (\mathbf{e}_t^{\xi\prime}, \xi_0', \ldots, \xi_{1-l}')' \qquad \textit{and} \qquad \tilde{\mathbf{e}}_t^{\eta} = (\mathbf{e}_t^{\eta\prime}, \eta_0', \ldots, \eta_{1-k}')'$$

for $t = 1$, *and*

$$\tilde{\mathbf{e}}_t^{\xi} = (\mathbf{e}_t^{\xi\prime}, 0, \ldots, 0)' \qquad \textit{and} \qquad \tilde{\mathbf{e}}_t^{\eta} = (\mathbf{e}_t^{\eta\prime}, 0, \ldots, 0)'$$

for $t \neq 1$. *If additionally also*

$$\left\| \xi_j - E\left(\xi_j \mid \mathbf{e}_{j+m}^{\xi}, \ldots, \mathbf{e}_{j-m}^{\xi}\right) \right\|_2, \quad 1 - l \leq j \leq 0,$$

and

$$\left\| \eta_j - E\left(\eta_j \mid \mathbf{e}_{j+m}^{\eta}, \ldots, \mathbf{e}_{j-m}^{\eta}\right) \right\|_2, \quad 1 - k \leq j \leq 0,$$

are of size $-q$, *then*

$$\left((\xi_t', \ldots, \xi_{t-l}', \eta_t', \ldots, \eta_{t-k}')'\right)_{t\in\mathbf{N}}$$

is near epoch dependent of size $-q$ *on* $((\mathbf{e}_t^{\xi\prime}, \mathbf{e}_t^{\eta\prime})')_{t\in\mathbf{Z}}$.

(b) Let $((\xi_t', \eta_t')')_{t\in\mathbf{N}}$ *be* L_p*-approximable, for some* $0 \leq p < \infty$, *by* $(\mathbf{e}_t)_{t\in\mathbf{Z}}$, *then* $(\xi_t)_{t\in\mathbf{N}}$ *and* $(\eta_t)_{t\in\mathbf{N}}$ *are each* L_p*-approximable by* $(\mathbf{e}_t)_{t\in\mathbf{Z}}$.

(b') Let $((\xi_t', \eta_t')')_{t\in\mathbf{N}}$ *be near epoch dependent of size* $-q$, $q > 0$, *on* $(\mathbf{e}_t)_{t\in\mathbf{Z}}$, *then* $(\xi_t)_{t\in\mathbf{N}}$ *and* $(\eta_t)_{t\in\mathbf{N}}$ *are each near epoch dependent of size* $-q$ *on* $(\mathbf{e}_t)_{t\in\mathbf{Z}}$.

(c) Let $(\xi_t^i)_{t\in\mathbf{N}}$, $i = 0, \ldots, k-1$, *be* L_p*-approximable, for some* $0 \leq p < \infty$, *by* $(\mathbf{e}_t^{\xi})_{t\in\mathbf{Z}}$, *where the* ξ_t^i *have the same dimension.*[18] *Consider the process* $(\eta_t)_{t\in\mathbf{N}}$ *defined by*

$$\eta_{(\tau-1)k+i+1} = \xi_\tau^i, \qquad \tau \in \mathbf{N}, \textit{ and } i = 0, \ldots, k-1.$$

Then $(\eta_t)_{t\in\mathbf{N}}$ *is* L_p*-approximable by* $(\mathbf{e}_t^{\eta})_{t\in\mathbf{Z}}$, *where*

$$\mathbf{e}_{(\tau-1)k+i+1}^{\eta} = \mathbf{e}_\tau^{\xi}, \qquad \tau \in \mathbf{Z}, \textit{ and } i = 0, \ldots, k-1.$$

(c') Let $(\xi_t^i)_{t\in\mathbf{N}}$, $i = 0, \ldots, k-1$, *be quadratically integrable processes that are near epoch dependent of size* $-q$, $q > 0$, *on* $(\mathbf{e}_t^{\xi})_{t\in\mathbf{Z}}$, *where the* ξ_t^i *have the same dimension. Consider the process* $(\eta_t)_{t\in\mathbf{N}}$ *defined by*

$$\eta_{(\tau-1)k+i+1} = \xi_\tau^i, \qquad \tau \in \mathbf{N}, \textit{ and } i = 0, \ldots, k-1.$$

---

[18] The assumption that the processes (ξ_t^i) are L_p-approximable by a "common" basis process $(\mathbf{e}_t^{\xi})$, rather than by "individual" basis processes $(\mathbf{e}_t^i)$, can be made without loss of generality, since in the latter case we can always take $\mathbf{e}_t^{\xi} = (\mathbf{e}_t^{0\prime}, \ldots, \mathbf{e}_t^{k-1\prime})'$ as the "common" basis process.

Then $(\eta_t)_{t\in\mathbf{N}}$ *is near epoch dependent of size* $-q$ *on* $(\mathbf{e}_t^\eta)_{t\in\mathbf{Z}}$*, where*

$$\mathbf{e}^\eta_{(\tau-1)k+i+1} = \mathbf{e}^\xi_\tau, \qquad \tau \in \mathbf{Z},\ \textit{and}\ i = 0,\ldots,k-1.$$

(d) Let $(\xi_t)_{t\in\mathbf{N}}$ *be* L_p*-approximable, for some* $0 \le p < \infty$*, by* $(\mathbf{e}_t^\xi)_{t\in\mathbf{Z}}$*. Consider the processes* $(\eta_t^i)_{t\in\mathbf{N}}$ *defined by*

$$\eta_t^i = \xi_{(t-1)k+i+1}, \qquad t \in \mathbf{N},\ \textit{and}\ i = 0,\ldots,k-1,$$

where $k \ge 1$*. Then, for every* $i = 0,\ldots,k-1$*, the process* $(\eta_t^i)_{t\in\mathbf{N}}$ *is* L_p*-approximable by* $(\mathbf{e}_t^\eta)_{t\in\mathbf{Z}}$ *where*

$$\mathbf{e}_t^\eta = \left(\mathbf{e}^{\xi\prime}_{(t-1)k+1},\ldots,\mathbf{e}^{\xi\prime}_{(t-1)k+k}\right)'.$$

(d') Let $(\xi_t)_{t\in\mathbf{N}}$ *be a quadratically integrable process that is near epoch dependent of size* $-q$*,* $q > 0$*, on* $(\mathbf{e}_t^\xi)_{t\in\mathbf{Z}}$*. Consider the processes* $(\eta_t^i)_{t\in\mathbf{N}}$ *defined by*

$$\eta_t^i = \xi_{(t-1)k+i+1}, \qquad t \in \mathbf{N},\ \textit{and}\ i = 0,\ldots,k-1,$$

where $k \ge 1$*. Then, for every* $i = 0,\ldots,k-1$*, the process* $(\eta_t^i)_{t\in\mathbf{N}}$ *is near epoch dependent of size* $-q$ *on* $(\mathbf{e}_t^\eta)_{t\in\mathbf{Z}}$ *where*

$$\mathbf{e}_t^\eta = \left(\mathbf{e}^{\xi\prime}_{(t-1)k+1},\ldots,\mathbf{e}^{\xi\prime}_{(t-1)k+k}\right)'.$$

The next three theorems give sets of conditions under which the output process of a dynamic system is L_p-approximable [near epoch dependent], given the input process has the corresponding property. We note that in Theorems 6.10 - 6.12 below $\mathbf{v}_t$ should typically be interpreted as the vector of endogenous variables, while $\mathbf{w}_t$ should be given the interpretation of the vector of exogenous variables and disturbances. A key condition in Theorem 6.10 is that the functions g_t are contraction mappings (for fixed input values). An important generalization of this theorem, that relaxes this condition, will be given in Theorem 6.12 below.

Theorem 6.10. *Let* $(\mathbf{v}_t)_{t\in\mathbf{N}}$ *and* $(\mathbf{w}_t)_{t\in\mathbf{Z}}$ *be stochastic processes taking their values in Borel subsets* V *and* W *of* $\mathbf{R}^{p_v}$ *and* $\mathbf{R}^{p_w}$*, respectively, and let* $g_t : V \times W \to V$ *be functions for* $t \in \mathbf{N}$*. Suppose that* $(\mathbf{v}_t)$ *is generated according to the dynamic system*

$$\mathbf{v}_t = g_t(\mathbf{v}_{t-1}, \mathbf{w}_t), \quad t \in \mathbf{N},$$

where $\mathbf{v}_0$ *is some initial random variable taking its values in* V*. Assume that for all* $(v, v^\bullet) \in V \times V, (w, w^\bullet) \in W \times W$*, and* $t \in \mathbf{N}$

$$|g_t(v,w) - g_t(v^\bullet, w^\bullet)| \le d_v\,|v - v^\bullet| + d_w\,|w - w^\bullet|$$

holds where the global Lipschitz constants satisfy $0 \leq d_v < 1$ *and* $0 \leq d_w < \infty$.

(a) If $\|\mathbf{w}_t\|_r < \infty$ *for* $t \in \mathbf{N}$ *and* $\|\mathbf{v}_0\|_r < \infty$ *for some* $r \geq 1$*, then* $\|\mathbf{v}_t\|_r < \infty$ *for* $t \in \mathbf{N}$*. If additionally*

$$\sup_{t \geq 1} \|\mathbf{w}_t\|_r < \infty$$

and

$$\sup_{t \geq 1} |g_t(\bar{v}, \bar{w})| < \infty$$

holds for some elements $\bar{v} \in V$, $\bar{w} \in W$*, then also*

$$\sup_{t \geq 1} \|\mathbf{v}_t\|_r < \infty.$$

(b) If

$$\sup_{t \geq 1} \|\mathbf{w}_t\|_r < \infty$$

and $\|\mathbf{v}_0\|_r < \infty$ *for some* $r \geq 1$*, and if* $(\mathbf{w}_t)_{t \in \mathbf{N}}$ *is* L_r*-approximable by some process* $(\mathbf{e}_t)_{t \in \mathbf{Z}}$*, then* $(\mathbf{v}_t)$ *is* L_r*-approximable by* $(\mathbf{e}_t)$.

(c) If

$$\sup_{t \geq 1} \|\mathbf{w}_t\|_2 < \infty$$

and $\|\mathbf{v}_0\|_2 < \infty$*, and if* $(\mathbf{w}_t)_{t \in \mathbf{N}}$ *is near epoch dependent of size* $-q$ *on some process* $(\mathbf{e}_t)_{t \in \mathbf{Z}}$*, then* $(\mathbf{v}_t)$ *is near epoch dependent of size* $-q$ *on* $(\tilde{\mathbf{e}}_t)$*, where* $\tilde{\mathbf{e}}_t = (\mathbf{e}_t', 0)'$ *for* $t \neq 1$ *and* $\tilde{\mathbf{e}}_1 = (\mathbf{e}_1', \mathbf{v}_0')'$*. If additionally also* $\|\mathbf{v}_0 - E(\mathbf{v}_0 | \mathbf{e}_m, \ldots, \mathbf{e}_{-m})\|_2$ *is of size* $-q$*, then* $(\mathbf{v}_t)$ *is near epoch dependent of size* $-q$ *on* $(\mathbf{e}_t)$*.* [19]

A slightly more general version of this theorem is given as Lemma D5 in Appendix D. Note that by choosing $\mathbf{e}_t = \mathbf{w}_t$ the above theorem provides as a special case conditions for the L_r-approximability and near epoch dependence of $(\mathbf{v}_t)$ w.r.t. the input process $(\mathbf{w}_t)$. In this case the proof actually shows that the approximation errors

$$\limsup_{n \to \infty} n^{-1} \sum_{t=1}^{n} \|\mathbf{v}_t - \mathbf{h}_t^m\|_r$$

and

$$\sup_{t \geq 1} \|\mathbf{v}_t - \mathbf{h}_t^m\|_2$$

[19] Of course, in case $\mathbf{v}_0$ is nonstochastic, near epoch dependence on $(\tilde{\mathbf{e}}_t)$ coincides with near epoch dependence on $(\mathbf{e}_t)$, and the condition that $\|\mathbf{v}_0 - E(\mathbf{v}_0 \mid \mathbf{e}_m, \ldots, \mathbf{e}_{-m})\|_2$ is of size $-q$ is automatically satisfied.

underlying the result in part (b) and the first result in part (c), respectively, decay at an exponential rate as $m \to \infty$. The choice $\mathbf{e}_t = \mathbf{w}_t$ in applying the above theorem is particularly useful when $\mathbf{w}_t$ is α-mixing or ϕ-mixing.

The above theorem requires that the dynamic system is contracting in the sense that for every w the functions $v \to g_t(v, w)$ have Lipschitz bounds, which are strictly less than 1 (uniformly in t). Frequently the contraction condition on g_t, i.e., $d_v < 1$, and the condition $d_w < \infty$ will be established by showing that

$$\sup\left\{\left|\text{stac}_{i=1}^{p_v}\left[\mathbf{i}_i' \frac{\partial g_t}{\partial v}\left(v^i, w^i\right)\right]\right| : v^i \in V,\ w^i \in W,\ \ i = 1, \ldots, p_v,\ t \in \mathbf{N}\right\} < 1$$

and

$$\left|\frac{\partial g_t}{\partial w}\right| \leq d < \infty,$$

for some constant d, where the stac-operator creates a matrix consisting of the rows shown as the arguments of the operator[20], and where $\mathbf{i}_i$ denotes the i-th column of the $p_v \times p_v$ identity matrix. The above sufficient conditions are derived by applying the mean value theorem to g_t. Since the mean value theorem has to be applied to each component of g_t separately, this results in different sets of mean values at which the derivatives of each component of g_t are evaluated. It is for this reason that we have to allow for a different argument list for each of the rows of the matrix generated by the stac-operator. In case of a univariate model, i.e., $p_v = 1$, the above conditions simplify to $|\partial g_t / \partial v| \leq \text{const} < 1$, and $|\partial g_t / \partial w| \leq d < \infty$.

Certain dynamic systems that do not satisfy the contraction condition of Theorem 6.10 may nevertheless be brought under the scope of this theorem by the following transformation device: Suppose we can find a family of transformations $S_t : V \to \bar{V}$, $t \geq 0$, with $\bar{V}$ a Borel subset of $\mathbf{R}^{p_v}$, S_t and S_t^{-1} Borel measurable, such that the transformed system

$$\bar{\mathbf{v}}_t = \bar{g}_t(\bar{\mathbf{v}}_{t-1}, \mathbf{w}_t), \qquad t \in \mathbf{N},$$

satisfies the conditions of Theorem 6.10. Here $\bar{\mathbf{v}}_t = S_t \mathbf{v}_t$ and $\bar{g}_t(\bar{v}, w) = S_t(g_t(S_{t-1}^{-1}\bar{v}, w))$ for all $(\bar{v}, w) \in \bar{V} \times W$. If the transformations S_t^{-1} are well-behaved such that they satisfy (component-wise) the assumptions of Theorem 6.5 or 6.6 [Corollary 6.8], then we obtain the L_p-approximability [near epoch dependence] of $(\mathbf{v}_t)$ from the L_p-approximability [near epoch dependence] of $(\bar{\mathbf{v}}_t)$, which itself follows as a consequence of Theorem 6.10

[20]That is, let $a_{i.}$ denote the i-th row of a $m \times n$ matrix A, then $\text{stac}_{i=1}^{m} a_{i.} = A$. Unless stated otherwise we also adopt the convention that $|A|$ denotes the smallest matrix norm that is compatible with the Euclidean vector norm, i.e., $|A|$ is the square root of the largest eigenvalue of $A'A$.

applied to the transformed system. A similar remark also applies to Theorems 6.11 and 6.12 given below. A particular instance where this transformation device can be applied successfully is the following: Suppose the original system satisfies the assumptions of Theorem 6.10 if the Euclidean norm $|v| = (v'v)^{1/2}$ is replaced by another norm $(v'Av)^{1/2}$, where A is some symmetric and positive definite matrix. Now choose $S_t \equiv S$ as a linear transformation on $\mathbf{R}^{p_v}$ satisfying $S'S = A$. Then the transformed system satisfies all the assumptions of Theorem 6.10 (w.r.t. the Euclidean norm), and clearly also S^{-1} satisfies (component-wise) the assumptions of Theorems 6.5, 6.6 and Corollary 6.8. Hence, in this instance the transformation device readily delivers the L_p-approximability [near epoch dependence] of $(\mathbf{v}_t)$.

Theorem 6.10 generalizes the discussion in Bierens (1981), Section 5.1.3, for univariate dynamic systems of the form $\mathbf{v}_t = \phi(\mathbf{v}_{t-1}) + \mathbf{w}_t$. Gallant (1987a, pp.502-503) considers a univariate system of the form $\mathbf{v}_t = \phi(\mathbf{v}_{t-1}, \mathbf{x}_t) + \epsilon_t$ with $\mathbf{v}_0 = 0$. Given $|\partial\phi/\partial v| \leq d_1 < 1$, and $|\partial\phi/\partial x| \leq d_2 < \infty$, and the appropriate moment conditions, we readily obtain L_p-approximability of $\mathbf{v}_t$ by [near epoch dependence of $\mathbf{v}_t$ on] $\mathbf{w}_t = (\mathbf{x}_t', \epsilon_t)'$ from Theorem 6.10. That is, the variables $\mathbf{v}_t$ can be approximated by random variables $\mathbf{h}_t^m$ which only depend on $\mathbf{x}_{t+m}, \epsilon_{t+m}, \ldots, \mathbf{x}_{t-m}, \epsilon_{t-m}$. Gallant claims near epoch dependence of $\mathbf{v}_t$ on $(\mathbf{x}_t', \epsilon_t)'$ to hold only under the assumption $|\partial\phi/\partial v| \leq d_1 < 1$. However, his proof, which uses arguments similar to those in Bierens and in the proof of Theorem 6.10, only shows that the variables $\mathbf{v}_t$ can be approximated by random variables which depend not only on $\mathbf{x}_{t+m}, \epsilon_{t+m}, \ldots, \mathbf{x}_{t-m}, \epsilon_{t-m}$, but also on $\mathbf{x}_{t-m-1}, \mathbf{x}_{t-m-2}, \ldots$ Furthermore, for $\mathbf{v}_t$ generated more generally by $\mathbf{v}_t = \phi(\mathbf{v}_{t-1}, \mathbf{x}_t, \epsilon_t)$, Gallant (1987a, pp.481-482) also claims near epoch dependence of $\mathbf{v}_t$ on $\mathbf{w}_t = (\mathbf{x}_t', \epsilon_t)'$ to hold under the assumptions $|\partial\phi/\partial v| \leq d_1 < 1$ and $|\partial\phi/\partial\epsilon| \leq d_2 < \infty$. Again his proof only shows that the variables $\mathbf{v}_t$ can be approximated by random variables which depend not only on $\mathbf{x}_{t+m}, \epsilon_{t+m}, \ldots, \mathbf{x}_{t-m}, \epsilon_{t-m}$, but also on $\mathbf{x}_{t-m-1}, \mathbf{x}_{t-m-2}, \ldots$ We note that in order to deduce the desired conclusion from Theorem 6.10 one would have to add the condition $|\partial\phi/\partial x| \leq d_3 < \infty$.[21] Therefore both claims in Gallant (1987a) remain unproven in general. However, we will show below that the claims are correct for the case of convex V; see the discussion after Theorem 6.11.

Kuan and White (1994), Proposition 4.4, present a result that is similar to the last claim of part (c) of Theorem 6.10 above.[22] However, the proof

---

[21] Gallant and White (1988, pp.29-31) discuss the same model and establish the near epoch dependence of $\mathbf{v}_t$ on $\mathbf{w}_t = (\mathbf{x}_t', \epsilon_t)'$ assuming $|\partial\phi/\partial v| \leq d_1 < 1$, and (a variant of) $|\partial\phi/\partial\epsilon| \leq d_2 < \infty$ and $|\partial\phi/\partial x| \leq d_3 < \infty$.

[22] Compared with Theorem 6.10(c), Kuan and White assume only the existence of second moments of $\mathbf{w}_t$ and not their boundedness in t, but assume $\mathbf{v}_0$ to be a bounded random variable. Furthermore they also make the stronger assumption

of their result is not completely correct, unless the sets V and W are, e.g., additionally assumed to be convex. Kuan and White's argument involves the evaluation of the function g, which is defined on $V \times W$, at conditional expectations of $\mathbf{v}_t$ and $\mathbf{w}_t$. However, those conditional expectations need not take their values in V and W, unless these sets are convex. In contrast, the method of proof employed in establishing Theorem 6.10 only requires V and W to be Borel sets.

Inspection of the proof of Theorem 6.10 reveals that the Lipschitz condition maintained by that theorem is also needed in this form even in the special case where $\mathbf{e}_t$ is chosen to be equal to $\mathbf{w}_t$. While the proof of Proposition 4.4 in Kuan and White (1994) only works for convex sets, analyzing the method of proof shows that it has the advantage that in the special case where $\mathbf{e}_t = \mathbf{w}_t$ and where V is convex the functions g_t have to possess the Lipschitz property only w.r.t. the first argument. This leads to the following theorem, which is closely related to Theorem 6.10.

Theorem 6.11. *Let $(\mathbf{v}_t)_{t\in\mathbf{N}}$ and $(\mathbf{w}_t)_{t\in\mathbf{Z}}$ be stochastic processes taking their values in Borel subsets V and W of $\mathbf{R}^{p_v}$ and $\mathbf{R}^{p_w}$, respectively, and let $g_t : V \times W \to V$ be Borel measurable functions for $t \in \mathbf{N}$. Assume that V is convex. Suppose that $(\mathbf{v}_t)$ is generated according to the dynamic system*

$$\mathbf{v}_t = g_t(\mathbf{v}_{t-1}, \mathbf{w}_t), \quad t \in \mathbf{N},$$

where $\mathbf{v}_0$ is some initial random variable taking its values in V. Assume that for all $(v, v^\bullet) \in V \times V$, $w \in W$, and $t \in \mathbf{N}$

$$|g_t(v, w) - g_t(v^\bullet, w)| \le d_v \, |v - v^\bullet|$$

holds where the global Lipschitz constant satisfies $0 \le d_v < 1$. If

$$\sup_{t\ge 0} \|\mathbf{v}_t\|_2 < \infty,$$

then $(\mathbf{v}_t)$ is near epoch dependent of size $-q$ on $(\mathbf{w}_t)$ for any $q > 0$ (and hence $(\mathbf{v}_t)$ is also L_2-approximable by $(\mathbf{w}_t)$).

Applying the mean value theorem we obtain the following sufficient condition for the contraction condition on g_t in Theorem 6.11:

$$\sup\left\{\left|\operatorname{stac}_{i=1}^{p_v}\left[\mathbf{i}_i' \frac{\partial g_t}{\partial v}\left(v^i, w\right)\right]\right| : v^i \in V,\, w \in W,\;\; i = 1, ..., p_v,\, t \in \mathbf{N}\right\} < 1,$$

---

that the function g does not depend on t, that g is bounded, and that V is compact. (The assumed boundedness of $\mathbf{v}_0$ is already implied by the implicit assumption that $\mathbf{v}_0$ has to take its values in the compact set V.) The Lipschitz conditions in their paper are formulated differently, but clearly are equivalent to the Lipschitz condition in Theorem 6.10.

where $\mathbf{i}_i$ denotes the i-th column of the $p_v \times p_v$ identity matrix; cf. the discussion after Theorem 6.10. In the case of a univariate model, i.e., $p_v = 1$, this condition simplifies to $|\partial g_t/\partial v| \leq$ const < 1. Theorem 6.11 thus shows that in case the set V is convex, Gallant's claims discussed above actually turn out to be correct; in fact they are correct under the single condition $|\partial\phi/\partial v| \leq d_1 < 1$ (provided one can establish the moment condition in Theorem 6.11).

Since dynamic systems of higher order can always be reformulated as dynamic systems of order one, it seems suggestive that Theorems 6.10 and 6.11 also cover dynamic systems of higher order. Unfortunately, this is not the case even for linear systems, as the following simple example shows. For purposes of illustration consider the following autoregressive model of order two

$$\mathbf{y}_t = a\mathbf{y}_{t-1} + b\mathbf{y}_{t-2} + \epsilon_t.$$

Rewriting the system in stacked notation as a system of order one, i.e., in companion form, leads to

$$\mathbf{v}_t = A\mathbf{v}_{t-1} + \mathbf{w}_t$$

where

$$\mathbf{v}_t = \begin{bmatrix} \mathbf{y}_t \\ \mathbf{y}_{t-1} \end{bmatrix}, \; \mathbf{w}_t = \begin{bmatrix} \epsilon_t \\ 0 \end{bmatrix}, \; A = \begin{bmatrix} a & b \\ 1 & 0 \end{bmatrix}.$$

Now choose, e.g., $v = (1,0)'$, $v^\bullet = (0,0)'$, and $w = w^\bullet$. In order for the contraction condition, i.e., $d_v < 1$, in Theorems 6.10 and 6.11 to be satisfied we would need to have

$$|Av - Av^\bullet| \leq d_v\,|v - v^\bullet| < |v - v^\bullet|\,.$$

However, $|v - v^\bullet| = |(1,0)'| = 1$ is never larger than $|Av - Av^\bullet| = |(a,1)'| = (1+a^2)^{1/2}$ regardless of the values of the parameters a and b. That is, the contraction condition $d_v < 1$ is always violated for the stacked system of order one (even if the autoregressive model of order two is stable). Stated differently, the contraction condition $d_v < 1$ amounts to

$$|A| = (\lambda_{\max}(A'A))^{1/2} \leq d_v < 1,$$

which can never be satisfied because

$$\begin{aligned} |A|^2 &= \sup\{v'A'Av : |v| = 1\} \\ &\geq (1,0)A'A(1,0)' = (1+a^2). \end{aligned}$$

It is easy to see that the above discussion extends to nonlinear systems.[23] That is, the contraction condition $d_v < 1$ of Theorems 6.10 and 6.11 can

[23]Consider the nonlinear system $\mathbf{y}_t = \phi_t(\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-p}, \mathbf{x}_t)$ where $\mathbf{x}_t$ denotes here (observed as well as unobserved) inputs. After reformulating the system in the usual way as a first order system, $\mathbf{v}_t = g_t(\mathbf{v}_{t-1}, \mathbf{w}_t)$, choose $v = (1,0,\ldots,0)'$, $v^\bullet = (0,\ldots,0)'$, and $w = w^\bullet$, and observe that $|v - v^\bullet| = |(1,0,\ldots,0)'| = 1$ while $|g_t(v,w) - g_t(v^\bullet, w^\bullet)| = |(\gamma,1,0,\ldots,0)| = (\gamma^2+1)^{1/2}$ for some γ.

never be satisfied for the companion form of a higher order dynamic system, regardless of whether the system is linear or nonlinear.

It should be noted that also for multivariate systems of order one, which do not result from a reduction of a higher order system, the contraction condition $d_v < 1$ will frequently be violated even if the system is "stable". For example, for triangular linear systems the usual stability condition only puts restrictions on the diagonal elements of the system matrix, while the contraction condition also restricts the off-diagonal elements.

Given the above observations about the restrictive nature of the contraction condition maintained by Theorems 6.10 and 6.11, it seems imperative for the development of a general asymptotic estimation theory for dynamic systems to establish analogous results that cover "stable" dynamic systems including higher order dynamic systems. Such a result is given below in Theorem 6.12. To motivate the conditions of this theorem consider, e.g., the linear system

$$\mathbf{v}_t = g(\mathbf{v}_{t-1}, \mathbf{w}_t) = G\mathbf{v}_{t-1} + \mathbf{w}_t.$$

Assume that the system is stable, i.e., all characteristic roots of G are less than one in absolute value. The contraction condition $d_v < 1$ would require $|G| \leq d_v < 1$ to hold, which is a much stronger condition than stability, as was pointed out above. Now consider the model iterated k times

$$\begin{aligned} \mathbf{v}_{t+k-1} &= g^{(k)}(\mathbf{v}_{t-1}, \mathbf{w}_t, \ldots, \mathbf{w}_{t+k-1}) \\ &= G^k \mathbf{v}_{t-1} + \sum_{i=0}^{k-1} G^{k-1-i} \mathbf{w}_{t+i}. \end{aligned}$$

Since G was assumed to be stable it follows that G^k converges to zero as $k \to \infty$, and hence for k sufficiently large $|G^k| \leq d_v < 1$. That is, the contraction condition holds for the system iterated k times, i.e., for $g^{(k)}$. The idea behind the proof of Theorem 6.12 is now to apply, in a first step, Theorem 6.10 to the subprocesses of $(\mathbf{v}_t)$, which are generated by the iterated system when initialized by $\mathbf{v}_i$, $i = 0, \ldots, k-1$, respectively, thus establishing L_r-approximability [near epoch dependence] of the subprocesses. The idea is further to show, in the second step of the proof, that one can recover L_r-approximability [near epoch dependence] of $(\mathbf{v}_t)$ from the corresponding property of the subprocesses. Theorem 6.12 shows that this idea works indeed even for "stable" dynamic nonlinear systems, where "stability" means that after a sufficiently large number of iterations an iterate of the system is contracting.[24] (Of course, in light of the above discussion, this notion of "stability" is satisfied by any linear system that is stable in the usual sense.)

---

[24]We note that Theorem 6.12 can be slightly generalized by using Lemma D5 instead of Theorem 6.10 in the proof.

We emphasize again that Theorem 6.12 – in contrast to Theorems 6.10 and 6.11 – covers "stable " and not only contracting systems. In particular, Theorem 6.12 also applies to dynamic nonlinear systems of higher order. As discussed above, the contraction condition in Theorems 6.10 and 6.11 can *never* be satisfied for the companion form of a higher order system.[25] However, this is not so for the stability condition used by Theorem 6.12. A "stable" higher order system will typically give rise to an equivalent reduced system of order one, where – after a sufficiently large number of iterations – an iterate of the companion form is contracting. (Recall that a stable linear system of arbitrary order leads to a stable companion form, and hence to a system of order one, an iterate of which is contracting.)

Theorem 6.12. *Let $(\mathbf{v}_t)_{t\in\mathbf{N}}$ and $(\mathbf{w}_t)_{t\in\mathbf{Z}}$ be stochastic processes taking their values in Borel subsets V and W of $\mathbf{R}^{p_v}$ and $\mathbf{R}^{p_w}$, respectively, and let $g_t : V \times W \to V$ be functions for $t \in \mathbf{N}$. Suppose that $(\mathbf{v}_t)$ is generated according to the dynamic system*

$$\mathbf{v}_t = g_t(\mathbf{v}_{t-1}, \mathbf{w}_t), \quad t \in \mathbf{N},$$

where $\mathbf{v}_0$ is some initial random variable taking its values in V. For $t \in \mathbf{N}$ and $k \in \mathbf{N}$ define the functions $g_t^{(k+1)}$, representing the iterations of the dynamic system, by the following recursions:

$$g_t^{(k+1)}(v, w_1, \ldots, w_{k+1}) = g_{t+k}\left(g_t^{(k)}(v, w_1, \ldots, w_k), w_{k+1}\right),$$

where $g_t^{(1)} = g_t$. Assume that there exists a $k^* \geq 1$ such that for all $(v, v^\bullet) \in V \times V$, $(w_1, \ldots, w_{k^*}, w_1^\bullet, \ldots, w_{k^*}^\bullet) \in \prod_{i=1}^{2k^*} W$, and $t \in \mathbf{N}$

$$\left|g_t^{(k^*)}(v, w_1, \ldots, w_{k^*}) - g_t^{(k^*)}(v^\bullet, w_1^\bullet, \ldots, w_{k^*}^\bullet)\right| \tag{6.6}$$

$$\leq d_v \left|v - v^\bullet\right| + d_w \left\| \begin{bmatrix} w_1 - w_1^\bullet \\ \cdot \\ \cdot \\ \cdot \\ w_{k^*} - w_{k^*}^\bullet \end{bmatrix} \right\|$$

[25] A possible alternative route to circumvent the problem that companion forms of higher order dynamic systems never satisfy the contraction condition in Theorems 6.10 and 6.11 could be based on the transformation device discussed after Theorem 6.10. This would require that we find transformations S_t such that the transformed companion form satisfies Theorem 6.10 or 6.11 and such that S_t^{-1} preserves L_p-approximability [near epoch dependence]. For a stable *linear* system this is always possible by choosing S as a linear transformation such that $\left|SGS^{-1}\right|$ is less than one – cf., e.g., Lemma 5.6.10 in Horn and Johnson (1985) and note that this lemma also holds with the norm $|||\,.\,|||_1$ replaced by the spectral norm $|\,.\,|$ used here. However, it is less than clear for which classes of dynamic nonlinear systems such transformations can be constructed and how this could be done in a systematic way.

holds where the global Lipschitz constants satisfy $0 \leq d_v < 1$ *and* $0 \leq d_w < \infty$.

(a) If $\|\mathbf{w}_t\|_r < \infty$ *for* $t \in \mathbf{N}$ *and* $\|\mathbf{v}_i\|_r < \infty$, $i = 0, \ldots, k^* - 1$, *for some* $r \geq 1$, *then* $\|\mathbf{v}_t\|_r < \infty$ *for* $t \in \mathbf{N}$. *If additionally*

$$\sup_{t \geq 1} \|\mathbf{w}_t\|_r < \infty$$

and

$$\sup_{t \geq 1} \left| g_t^{(k^*)}(\bar{v}, \bar{w}_1, \ldots, \bar{w}_{k^*}) \right| < \infty$$

holds for some elements $\bar{v} \in V$, $(\bar{w}_1, \ldots, \bar{w}_{k^*}) \in \prod_{i=1}^{k^*} W$, *then also*

$$\sup_{t \geq 1} \|\mathbf{v}_t\|_r < \infty.$$

(b) If

$$\sup_{t \geq 1} \|\mathbf{w}_t\|_r < \infty$$

and $\|\mathbf{v}_i\|_r < \infty$, $i = 0, \ldots, k^* - 1$, *for some* $r \geq 1$, *and if* $(\mathbf{w}_t)_{t \in \mathbf{N}}$ *is* L_r*-approximable by some process* $(\mathbf{e}_t)_{t \in \mathbf{Z}}$, *then* $(\mathbf{v}_t)$ *is* L_r*-approximable by* $(\mathbf{e}_t)$.

(c) If

$$\sup_{t \geq 1} \|\mathbf{w}_t\|_2 < \infty$$

and $\|\mathbf{v}_i\|_2 < \infty$, $i = 0, \ldots, k^* - 1$, *and if* $(\mathbf{w}_t)_{t \in \mathbf{N}}$ *is near epoch dependent of size* $-q$ *on some process* $(\mathbf{e}_t)_{t \in \mathbf{Z}}$, *then* $(\mathbf{v}_t)$ *is near epoch dependent of size* $-q$ *on* $(\tilde{\mathbf{e}}_t)$, *where* $\tilde{\mathbf{e}}_t = (\mathbf{e}_t', 0, \ldots, 0)'$ *for* $t \neq 1$ *and* $\tilde{\mathbf{e}}_1 = (\mathbf{e}_1', \mathbf{v}_0', \ldots, \mathbf{v}_{k^*-1}')'$. *If additionally also*

$$\|\mathbf{v}_i - E(\mathbf{v}_i | \mathbf{e}_{i+m}, \ldots, \mathbf{e}_{i-m})\|_2, \quad i = 0, \ldots, k^* - 1,$$

are of size $-q$, *then* $(\mathbf{v}_t)$ *is near epoch dependent of size* $-q$ *on* $(\mathbf{e}_t)$.[26]

Theorem 6.12 maintains the conditions $\|\mathbf{v}_i\|_r < \infty$ for $i = 0, \ldots, k^* - 1$. This condition can be simplified – as is easily seen – to the single condition $\|\mathbf{v}_0\|_r < \infty$, provided the system maps g_t are Lipschitz functions, cf. Remark (iii) after Lemma D5 in Appendix D.

The above theorem requires that the dynamic system is stable in the sense that for some k^* the functions $v \to g_t^{(k^*)}(v, w_1, \ldots, w_{k^*})$ satisfy condition (6.6) with $d_v < 1$ and $d_w < \infty$. Applying the mean value theorem it

[26]We note again that in case $\mathbf{v}_0, \ldots, \mathbf{v}_{k^*-1}$ are nonstochastic, near epoch dependence on $(\tilde{\mathbf{e}}_t)$ coincides with near epoch dependence on $(\mathbf{e}_t)$, and the condition that the approximation errors $\|\mathbf{v}_i - E(\mathbf{v}_i | \mathbf{e}_{i+m}, \ldots, \mathbf{e}_{i-m})\|_2$ are of size $-q$ is automatically satisfied.

is readily seen that a sufficient condition for this to hold is that

$$\sup\left\{\left|\operatorname{stac}_{i=1}^{p_v}\left[\mathbf{i}_i'\frac{\partial g_t^{(k^*)}}{\partial v}\left(v^i,w_1^i,\ldots,w_{k^*}^i\right)\right]\right|:\right.$$
$$\left. v^i\in V,\, w_j^i\in W,\ \ j=1,...,k^*,\ \ i=1,...,p_v,\, t\in\mathbf{N}\right\}<1$$

and

$$\sup\left\{\left|\frac{\partial g_t^{(k^*)}}{\partial w_l}\left(v,w_1,\ldots,w_{k^*}\right)\right|:\right.$$
$$\left. v\in V,\ \ w_j\in W,\, j=1,...,k^*,\, t\in\mathbf{N}\right\}<\infty,\qquad l=1,...,k^*,$$

where $\mathbf{i}_i$ denotes the i-th column of the $p_v\times p_v$ identity matrix; cf. the discussion after Theorem 6.10.

The functions $g_t^{(k^*)}$ are defined recursively in terms of $g_{t+k^*-1},...,g_t$. Utilizing this definition we can also give the following sufficient conditions for the above two conditions:

$$\sup\left\{\left|\operatorname{stac}_{i=1}^{p_v}\left[\mathbf{i}_i'\prod_{l=1}^{k^*}\frac{\partial g_{t+k^*-l}}{\partial v}\left(g_t^{(k^*-l)}(v^i,w_1^i,...,w_{k^*-l}^i),w_{k^*-l+1}^i\right)\right]\right|:\right.$$
$$\left. v^i\in V,\ \ w_j^i\in W,\ \ j=1,...,k^*,\ \ i=1,...,p_v,\ \ t\in\mathbf{N}\right\}<1$$

with $g_t^{(0)}(v)=v$ and

$$|\partial g_t/\partial v|\le c_1<\infty,\qquad |\partial g_t/\partial w|\le c_2<\infty.$$

Theorems 6.10 - 6.12 are important ingredients for the development of laws of large numbers and central limit theorems for (functions of) processes generated by dynamic systems, in that they can be readily combined with existing laws of large numbers and central limit theorems for L_r-approximable and near epoch dependent processes. For example, combining Theorem 6.12 with Theorems 6.3, 6.4, or 10.2 immediately gives laws of large numbers or central limit theorems for such processes. We shall also make use of this fact in Chapter 14, where we consider the estimation of dynamic nonlinear models by quasi maximum likelihood methods.

6.5 Illustrations of Local Laws of Large Numbers

In this section we illustrate how the results obtained in the previous section can be utilized to establish sufficient conditions that ensure that $q_t(\mathbf{z}_t,\theta)$

satisfies a local LLN, i.e., conditions that imply LLNs for the "bracketing" functions $q_t^*(\mathbf{z}_t, \theta; \eta)$ and $q_{t*}(\mathbf{z}_t, \theta; \eta)$. Clearly, a first set of sufficient conditions could be obtained from Theorems 6.3 and 6.4 if we simply put $\mathbf{v}_t$ equal to $q_t^*(\mathbf{z}_t, \theta; \eta)$ and $q_{t*}(\mathbf{z}_t, \theta; \eta)$ in these theorems. This would amount to postulating the L_p-approximability or near epoch dependence conditions maintained in those theorems directly for $q_t^*(\mathbf{z}_t, \theta; \eta)$ and $q_{t*}(\mathbf{z}_t, \theta; \eta)$. It seems of interest to be able to avoid such a "black box" assumption and to have available more basic sets of sufficient conditions for local LLNs that place the L_p-approximability or near epoch dependence assumptions directly on the data generating process $(\mathbf{z}_t)$ itself.[27] We first consider the practically most important case where $\mathbf{z}_t$ is a vector of finite dimension. More specifically we assume that Z is a Borel subset of $\mathbf{R}^{p_z}$ and the associated σ-field $\mathfrak{Z}$ is then the induced Borel σ-field. Also recall from Chapter 5 that (Θ, ρ) is a metric space, that $q_t(., \theta)$ is $\mathfrak{Z}$-measurable for each $\theta \in \Theta$ and $t \in \mathbf{N}$, and that $d_t(z) = \sup_{\theta \in \Theta} |q_t(z, \theta)|$. The following three theorems give sets of more basic conditions ensuring that $q_t(\mathbf{z}_t, \theta)$ satisfies a local LLN.[28] The first theorem is based on an equicontinuity assumption for q_t and follows as a corollary to Theorems 6.3 and 6.5.

Theorem 6.13. *Let Θ be compact and Z be a Borel subset of $\mathbf{R}^{p_z}$. Suppose $(\mathbf{z}_t)$ is L_0-approximable by an α-mixing basis process and*

$$\sup_n n^{-1} \sum_{t=1}^{n} E\left[d_t(\mathbf{z}_t)^{1+\gamma}\right] < \infty$$

for some $\gamma > 0$. If $\{\bar{H}_n^z : n \in \mathbf{N}\}$ is tight on Z and if the family $\{q_t(z, \theta) : t \in \mathbf{N}\}$ is equicontinuous on $Z \times \Theta$, then for any $\theta \in \Theta$ and any $\eta > 0$ the functions $q_t^*(z, \theta; \eta)$ and $q_{t*}(z, \theta; \eta)$ are real valued and Borel measurable and $q_t^*(\mathbf{z}_t, \theta; \eta)$ and $q_{t*}(\mathbf{z}_t, \theta; \eta)$ satisfy a weak LLN, i.e., the "in probability" version of Assumption 5.2 is satisfied.

The two subsequent theorems are based on a Lipschitz-type condition for q_t. The next theorem follows as a corollary to Theorems 6.3 and 6.6.

---

[27]In case $\mathbf{z}_t$ consists of endogenous and exogenous variables, where the endogenous variables are generated from a dynamic system, the L_p-approximability of $(\mathbf{z}_t)$ by [near epoch dependence of $(\mathbf{z}_t)$ on] a mixing process can then be implied from the analogous properties of the processes of exogenous variables and disturbances, e.g., via Theorem 6.12.

[28]If $\mathbf{z}_t = (\mathbf{w}_t', \ldots, \mathbf{w}_{t-l}')'$ and if $(\mathbf{w}_t)$ is L_p-approximable then in view of Lemma 6.9 also $(\mathbf{z}_t)$ has the same property. Hence, if for example $\mathbf{w}_t$ represents the vector of exogenous and endogenous variables, the approximability conditions in the subsequent Theorems 6.13 and 6.14 are effectively put on the endogenous and exogenous process. Also the tightness of $\{\bar{H}_n^z : n \in \mathbf{N}\}$ follows from tightness of $\{\bar{H}_n^w : n \in \mathbf{N}\}$ (if Z is relatively closed in $\prod_{i=0}^{l} W$), cf. Lemma C1 in Appendix C.

Theorem 6.14. [29] *Let Z be a Borel subset of $\mathbf{R}^{p_z}$ and let $q_t^*(z,\theta;\eta)$ and $q_{t*}(z,\theta;\eta)$ be real valued and Borel measurable for any $\theta \in \Theta$ and $\eta > 0$ small enough. Suppose $(\mathbf{z}_t)$ is L_0-approximable by an α-mixing basis process $(\mathbf{e}_t)$ and*

$$\sup_n n^{-1} \sum_{t=1}^{n} E\left[d_t(\mathbf{z}_t)^{1+\gamma}\right] < \infty$$

for some $\gamma > 0$, where $d_t(\mathbf{z}_t)$ is $\mathfrak{A}$-measurable. Suppose there exist Borel measurable functions $B_t : Z \times Z \to [0,\infty)$ such that for all $z, z^\bullet \in Z$ and $\theta \in \Theta$

$$|q_t(z,\theta) - q_t(z^\bullet,\theta)| \leq B_t(z,z^\bullet)\,|z - z^\bullet|\,.$$

Suppose further that

$$\limsup_{m\to\infty} \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} E\left[B_t(\mathbf{z}_t, \mathbf{h}_t^m)^\epsilon\right] < \infty$$

for some $\epsilon > 0$ and some L_0-approximators $\mathbf{h}_t^m$ of $(\mathbf{z}_t)$ based on $(\mathbf{e}_t)$, where the approximators take their values in Z, then $q_t^*(\mathbf{z}_t,\theta;\eta)$ and $q_{t*}(\mathbf{z}_t,\theta;\eta)$ satisfy a weak LLN, and hence the "in probability" version of Assumption 5.2 is satisfied.

By applying McLeish's strong LLN, given above as Theorem 6.4, Gallant (1987a, Ch.7) and Gallant and White (1988) give strong local LLNs under the high level assumption that $q_t^*(\mathbf{z}_t,\theta;\eta)$ and $q_{t*}(\mathbf{z}_t,\theta;\eta)$ are near epoch dependent. They also present in Gallant (1987a, pp.498-499) and in Gallant and White (1988), Theorem 4.2, results on the transformation of near epoch dependent processes, which allow the derivation of more basic conditions for strong local LLNs. As discussed, Gallant's and White's transformation results assume that the transformations are defined on entire Euclidean spaces. The following theorem now provides basic conditions for strong local LLNs for transformations that are not necessarily defined on the entire Euclidean space. It is obtained as a corollary to Theorems 6.4 and 6.7(a).

Theorem 6.15. *Let Z be a Borel subset of $\mathbf{R}^{p_z}$, let $q_t^*(z,\theta;\eta)$ and $q_{t*}(z,\theta;\eta)$ be real valued and Borel measurable for any $\theta \in \Theta$ and $\eta > 0$ small enough, and assume that*

$$\sum_{t=1}^{\infty} \|d_t(\mathbf{z}_t)\|_r^2 / t^2 < \infty$$

[29] If Θ is compact the $\mathfrak{A}$-measurability of $d_t(\mathbf{z}_t)$ follows automatically from the measurability assumptions on the "bracketing" functions q_t^* and q_{t*}, since then $d_t(\mathbf{z}_t)$ can be expressed as the maximum of the absolute values of a finite number of "bracketing" functions. The same remark also applies to Theorem 6.15.

for some $r \geq 2$, where $d_t(\mathbf{z}_t)$ is $\mathfrak{A}$-measurable. Suppose there exist Borel measurable functions $B_t : Z \times Z \to [0, \infty)$ such that for all $z, z^\bullet \in Z$ and $\theta \in \Theta$

$$|q_t(z,\theta) - q_t(z^\bullet,\theta)| \leq B_t(z,z^\bullet)\,|z - z^\bullet|\,.$$

For some $s \geq 1$ let $\mathbf{h}_t^m$ be an L_s-approximator of $(\mathbf{z}_t)$ based on $(\mathbf{e}_t)$, where the approximators take their values in Z and the sequence of approximation errors $\sup_t \|\mathbf{z}_t - \mathbf{h}_t^m\|_s$ is of size $-q$ for some $q > 1$. Suppose further that

$$\sup_m \sup_t \|B_t(\mathbf{z}_t, \mathbf{h}_t^m)\|_{s/(s-1)} < \infty$$

and

$$\sup_m \sup_t \|B_t(\mathbf{z}_t, \mathbf{h}_t^m)\,|\mathbf{z}_t - \mathbf{h}_t^m|\,\|_{(2q-1)/(q-1)} < \infty.$$

If the basis process $(\mathbf{e}_t)$ is α-mixing with mixing coefficients of size $-r/(r-2)$ and $r > 2$, or ϕ-mixing with mixing coefficients of size $-r/(2r-2)$, then $q_t^*(\mathbf{z}_t, \theta; \eta)$ and $q_{t*}(\mathbf{z}_t, \theta; \eta)$ satisfy a strong LLN, and hence the "almost sure" version of Assumption 5.2 is satisfied.

In the above discussion of sufficient conditions for local laws of large numbers we have assumed that $\mathbf{z}_t$ is a finite vector. In case of, e.g., a model with moving average disturbances, $\mathbf{z}_t$ may be of the form $\mathbf{z}_t = (\mathbf{w}_t', \mathbf{w}_{t-1}', \ldots)'$. Hence in the following we will outline how the above results can be generalized to the case where Z is a Borel subset of $\mathbf{R}^\infty$. Of course, in order to obtain generalizations of Theorems 6.13 - 6.15 one only has to generalize the results on L_p-approximability [near epoch dependence] of functions of L_p-approximable [near epoch dependent] processes given in Theorems 6.5 - 6.7 and Corollary 6.8. As a first step one has to generalize the concept of L_p-approximability [near epoch dependence] for processes that take their values in $\mathbf{R}^\infty$. In defining the approximation error we have to replace the Euclidean metric by some appropriate metric. For definiteness we choose the metric to be $d(x,y) = \sum_{i=1}^\infty |x_i - y_i| \,/[2^i(1 + |x_i - y_i|)]$, $x = (x_1, x_2, \ldots) \in \mathbf{R}^\infty$, $y = (y_1, y_2, \ldots) \in \mathbf{R}^\infty$. Then Theorem 6.5 carries over if q_t is equicontinuous w.r.t. $d(.,.)$ as the metric on Z. Similarly Theorems 6.6 and 6.7 can be generalized if the Lipschitz-type condition holds w.r.t. $d(.,.)$ as the metric on Z; concerning such a result for near epoch dependence under a Lipschitz-type condition see Theorem 4.2 in Gallant and White (1988). Intuitively speaking, for $q_t^*(\mathbf{z}_t, \theta; \eta)$ and $q_{t*}(\mathbf{z}_t, \theta; \eta)$ with $\mathbf{z}_t = (\mathbf{w}_t', \mathbf{w}_{t-1}', \ldots)'$ to be L_p-approximable [near epoch dependent] and to satisfy a LLN if $\mathbf{w}_t$ is L_p-approximable [near epoch dependent] it seems necessary that the dependence of the functions q_t on arguments corresponding to high lags is weak. This is formally expressed in terms of equicontinuity or Lipschitz-type assumptions for q_t w.r.t. the metric $d(.,.)$, which itself puts declining weights on coordinates corresponding to high lags.

6.6 Comparison of ULLNs for L_p-Approximable and Near Epoch Dependent Processes

As discussed in Chapter 5, the ULLNs by Andrews (1987) and Pötscher and Prucha (1989) – given in that chapter as Theorems 5.1 and 5.2 – maintain different smoothness conditions for the functions $q_t(z, \theta)$. Those smoothness conditions – corresponding to Theorems 5.1 and 5.2 – are expressed by Assumptions A and B. Assumption A maintains that the functions q_t satisfy a Lipschitz-type condition with respect to θ. No smoothness condition is imposed with respect to the data. Assumption B maintains a continuity type condition jointly with respect to z and θ for the functions q_t. Both ULLNs maintain that the functions $q_t(\mathbf{z}_t, \theta)$ satisfy local LLNs. In Section 6.5 of this chapter we have developed basic sufficient conditions for such local LLNs to hold within the context of L_p-approximable and near epoch dependent processes. Those sufficient conditions also put smoothness conditions on the functions q_t. It seems of interest to discuss the total resulting smoothness conditions when combining the ULLNs in Chapter 5 with the local LLNs in Section 6.5.

For ease of discussion we explicitly restate the smoothness condition maintained by, respectively, Theorem 6.13 and by Theorems 6.14 and 6.15.

Assumption E. *The family $\{q_t : t \in \mathbf{N}\}$ is equicontinuous on $Z \times \Theta$.*

Assumption F. *There exist Borel measurable functions $B_t : Z \times Z \to [0, \infty)$ such that for all $z, z^\bullet \in Z$ and $\theta \in \Theta$*

$$|q_t(z, \theta) - q_t(z^\bullet, \theta)| \leq B_t(z, z^\bullet) \, |z - z^\bullet| \, .$$

(The class of Borel measurable functions B_t is then restricted in Theorems 6.14 and 6.15 by dominance conditions. Those dominance conditions also involve approximators $\mathbf{h}_t^m$.)

Assumptions B and E fit together well. More specifically Assumption E is a special case of Assumption B with $K = 1$ and $r_{kt} \equiv 1$. Hence no additional smoothness condition is needed for the ULLN given in Theorem 5.2 (apart from the restriction that $K = 1$ and $r_{kt} \equiv 1$) if the local LLN for $q_t(\mathbf{z}_t, \theta)$ is implied by Theorem 6.13. In comparing Assumption A with Assumptions E and F we see that both Assumptions E and F postulate an additional smoothness condition for the functions $q_t(z, \theta)$, since Assumption A imposes no smoothness condition w.r.t. z. If for the ULLN given as Theorem 5.1 the local LLN for $q_t(\mathbf{z}_t, \theta)$ is implied via Theorem 6.14 or 6.15, then q_t has to satisfy a Lipschitz-type condition in both arguments (plus dominance conditions). As a practical matter we may often attempt to verify those Lipschitz-type conditions from a differentiability assumption (plus dominance conditions for the first order derivatives). Against this background Theorem 5.2 together with Theorem 6.13, which both require essentially only that $q_t(z, \theta)$ is equicontinuous, seems more readily applicable (also since the assumptions do not involve approximators $\mathbf{h}_t^m$).

We note further that Theorem 5.2 together with Theorem 6.13 extend the approach taken by Bierens (1981, 1982a, 1984): Bierens assumed that $q_t \equiv q$ does not depend on t and that q is continuous. This assumption is weakened here to equicontinuity of q_t. Bierens assumed furthermore that $(\mathbf{z}_t)$ is asymptotically stationary, which essentially rules out heterogeneity. This assumption is weakened to tightness which, as discussed, can typically be implied by a weak moment condition. (Of course, by assuming $q_t \equiv q$ and $(\mathbf{z}_t)$ to be asymptotically stationary, Bierens is able to show also that $n^{-1} \sum_{t=1}^{n} Eq_t(\mathbf{z}_t, \theta)$ converges to a finite limit; cf. also Theorem 5.3.)

The equicontinuity assumption can be weakened further via Lemma D4 in Appendix D. This lemma generalizes Theorem 6.5 by allowing, to some extent, for non-equicontinuity of the transforming functions (and in particular for some forms of discontinuity). Furthermore, as pointed out by Newey (1987), equicontinuity may often be obtained through a suitable redefinition of the data.

7

CONSISTENCY: CATALOGUES OF ASSUMPTIONS

In the previous chapters we have discussed the basic structure of the classical consistency proof for M-estimators and have established basic modules that can be employed for consistency proofs in dynamic nonlinear models. Various catalogues of assumptions that imply the consistency of M-estimators in dynamic nonlinear models can be obtained by combining respective modules. In the following we illustrate this by specifying two alternative catalogues of assumptions for the consistency of a general class of M-estimators, which includes least mean distance and generalized method of moments estimators. For a further illustrative application of the respective modules see Chapter 14, which contains a derivation of the asymptotic properties of the (quasi) NFIML estimator of a dynamic implicit nonlinear simultaneous equation system.

We continue to maintain the general setup of the estimation problem introduced in Chapter 2 and consider M-estimators $\hat{\beta}_n$ corresponding to objective functions of the form:

$$R_n(\omega,\beta) = Q_n(\mathbf{z}_1,\ldots,\mathbf{z}_n,\hat{\tau}_n,\beta) = \vartheta_n\left(n^{-1}\sum_{t=1}^{n} q_t(\mathbf{z}_t,\hat{\tau}_n,\beta),\hat{\tau}_n,\beta\right) \quad (7.1)$$

where $q_t : Z \times T \times B \to \mathbf{R}^{p_q}$, $\vartheta_n : \mathbf{R}^{p_q} \times T \times B \to \mathbf{R}$, and $\hat{\tau}_n : \Omega \to T$ is an estimator for the nuisance parameter. The class of M-estimators defined by (7.1) is fairly general and includes generalized method of moments estimators as well as least mean distance estimators; the latter correspond to the case where $p_q = 1$ and $\vartheta_n(c,\tau,\beta) \equiv c$. For some sequence $\bar{\tau}_n \in T$ (typically a population analogue of $\hat{\tau}_n$) let $\bar{\beta}_n$ denote a sequence of minimizers of

$$\bar{R}_n(\beta) = \bar{Q}_n(\bar{\tau}_n,\beta) = \vartheta_n\left(n^{-1}\sum_{t=1}^{n} Eq_t(\mathbf{z}_t,\bar{\tau}_n,\beta),\bar{\tau}_n,\beta\right), \quad (7.2)$$

where the existence of the minimizers is implicitly assumed. The existence of the expectation on the r.h.s. of (7.2) will follow from Assumption 7.1(c) below.

We now introduce two sets of assumptions for the consistency of $\hat{\beta}_n$. The first set is based on an equicontinuity assumption on the functions q_t and the second set is based on Lipschitz-type assumptions for those functions. Assumptions common to both sets are collected in Assumption 7.1. Those specific to the two sets are given in Assumption 7.2 and 7.3, respectively. In the following q_{ti} denotes the i-th component of q_t.

Assumption 7.1. *(a) Z is a Borel subset of $\mathbf{R}^{p_z}$ (where the associated σ-field $\mathfrak{Z}$ is the induced Borel σ-field) and T and B are compact metric spaces.*

(b) $\{\vartheta_n : n \in \mathbf{N}\}$ is equicontinuous on $\mathbf{R}^{p_q} \times T \times B$. (In case $\vartheta_n \equiv \vartheta$ this clearly reduces to continuity of ϑ on $\mathbf{R}^{p_q} \times T \times B$.)

(c) Let

$$d_t(z) = \max_{1 \le i \le p_q} \sup_{T \times B} |q_{ti}(z, \tau, \beta)|,$$

then

$$\sup_n n^{-1} \sum_{t=1}^{n} E\left[d_t(\mathbf{z}_t)^{1+\gamma}\right] < \infty$$

for some $\gamma > 0$.

(d) $(\mathbf{z}_t)$ is L_0-approximable by some α-mixing basis process, say, $(\mathbf{e}_t)$.

(e) $\rho_T(\hat{\tau}_n, \bar{\tau}_n) \to 0$ i.p. as $n \to \infty$.

Assumption 7.2. *$\{q_t : t \in \mathbf{N}\}$ is equicontinuous on $Z \times T \times B$. (In case $q_t \equiv q$ this clearly reduces to continuity of q on $Z \times T \times B$.) Furthermore $\{n^{-1} \sum_{t=1}^{n} H_t^z : n \in \mathbf{N}\}$ is tight on Z (which is, e.g., the case if*

$$\sup_n n^{-1} \sum_{t=1}^{n} E\left|\mathbf{z}_t\right|^{\gamma^\bullet} < \infty$$

for some $\gamma^\bullet > 0$ and if Z is closed in $\mathbf{R}^{p_z}$).

Assumption 7.3. *(a) For each $(\tau, \beta) \in T \times B$ the functions $q_t(., \tau, \beta)$ are $\mathfrak{Z}$-measurable, and for each $(\tau, \beta) \in T \times B$ there exists an $\eta > 0$ such that $\rho((\tau, \beta), (\tau^\bullet, \beta^\bullet)) \le \eta$ implies*[1]

$$|q_t(\mathbf{z}_t, \tau^\bullet, \beta^\bullet) - q_t(\mathbf{z}_t, \tau, \beta)| \le b_t(\mathbf{z}_t) h\left[\rho\left((\tau, \beta), (\tau^\bullet, \beta^\bullet)\right)\right],$$

for all $t \in \mathbf{N}$, a.s., where $b_t : Z \to [0, \infty)$ and $h : [0, \infty) \to [0, \infty)$ are such that $b_t(\mathbf{z}_t)$ is $\mathfrak{A}$-measurable,

$$\sup_n n^{-1} \sum_{t=1}^{n} E b_t(\mathbf{z}_t) < \infty,$$

---

[1]The metric ρ denotes here, e.g., $\rho_T + \rho_B$; cf. Footnote 5 in Chapter 3.

$h(y) \downarrow h(0) = 0$ as $y \downarrow 0$, and η, b_t, h and the null set may depend on (τ, β).

(b) There exist Borel measurable functions $B_t : Z \times Z \rightarrow [0, \infty)$ such that for all $z, z^\bullet \in Z$ and $(\tau, \beta) \in T \times B$

$$|q_t(z, \tau, \beta) - q_t(z^\bullet, \tau, \beta)| \leq B_t(z, z^\bullet)\,|z - z^\bullet|$$

with

$$\limsup_{m \to \infty} \limsup_{n \to \infty} n^{-1} \sum_{t=1}^{n} E\left[B_t(\mathbf{z}_t, \mathbf{h}_t^m)^\epsilon\right] < \infty$$

for some $\epsilon > 0$ and some L_0-approximators $\mathbf{h}_t^m$ of $(\mathbf{z}_t)$, where the approximators take their values in Z.

(c) The functions $q_{ti}^*(z, \tau, \beta; \eta)$ and $q_{ti*}(z, \tau, \beta; \eta)$, $i = 1, ..., p_q$, are finite and Borel measurable for any $(\tau, \beta) \in T \times B$ and $\eta > 0$ small enough.

(d) The functions

$$\sup_{T \times B} \left| n^{-1} \sum_{t=1}^{n} q_t(\mathbf{z}_t, \tau, \beta) - n^{-1} \sum_{t=1}^{n} E q_t(\mathbf{z}_t, \tau, \beta) \right|$$

and

$$\sup_{T \times B} \left| \vartheta_n \left(n^{-1} \sum_{t=1}^{n} q_t(\mathbf{z}_t, \tau, \beta), \tau, \beta \right) - \vartheta_n \left(n^{-1} \sum_{t=1}^{n} E q_t(\mathbf{z}_t, \tau, \beta), \tau, \beta \right) \right|$$

are $\mathfrak{A}$-measurable.

We now introduce the following result concerning the consistency of $\hat{\beta}_n$.

Theorem 7.1. [2] *Suppose Assumptions 7.1 and either 7.2 or 7.3 hold. Then*

$$\sup_B \left| R_n(\omega, \beta) - \bar{R}_n(\beta) \right| \rightarrow 0 \text{ i.p. as } n \rightarrow \infty \qquad (7.3)$$

and $\{\bar{R}_n : n \in \mathbf{N}\}$ is equicontinuous on B. Furthermore, let $\bar{\beta}_n$ be an identifiably unique sequence of minimizers of $\bar{R}_n(\beta)$ and let $\hat{\beta}_n$ be any sequence of estimators such that eventually[3]

$$R_n(\omega, \hat{\beta}_n) = \inf_B R_n(\omega, \beta) \qquad (7.4)$$

---

[2]The measurability of $\sup_B \left| R_n(\omega, \beta) - \bar{R}_n(\beta) \right|$, which is implicitly assumed in (7.3) – cf. Footnote 3 in Chapter 3 – is automatically guaranteed under Assumptions 7.1 and 7.2.

[3]This could be relaxed to the requirement that (7.4) holds on sets $\Omega_n \in \mathfrak{A}$ with $P(\Omega_n) \rightarrow 1$.

holds. Then $\hat{\beta}_n$ *is consistent for* $\bar{\beta}_n$, *i.e.,* $\rho_B(\hat{\beta}_n, \bar{\beta}_n) \to 0$ *i.p. as* $n \to \infty$.

The above theorem follows from Lemma 3.1, where Assumptions 7.1 and either 7.2 or 7.3 are used to ensure that $R_n(\omega, \beta) - \bar{R}_n(\beta)$ converges to zero uniformly over the parameter space B via Lemmata 3.2 - 3.3 and, respectively, Theorems 5.2, 6.13 and Theorems 5.1, 6.14.[4] The theorem also covers the case where no nuisance parameter τ is present.[5]

As discussed in Section 4.6, the assumption that an identifiably unique sequence of minimizers exists is restrictive under misspecification. The above theorem can be readily generalized along the lines that $\hat{\beta}_n$ converges to the sets of minimizers of $\bar{R}_n(\beta)$ by using Lemma 4.2 in place of Lemma 3.1. Of course, Theorem 7.1 also holds for approximate M-estimators, cf. Sections 4.4 and 4.6.

In applications the equicontinuity condition in Assumption 7.2 will be easier to verify than the Lipschitz-type conditions in Assumption 7.3, in particular also since the latter involve approximators $\mathbf{h}_t^m$; cf. Section 6.6 for a related discussion. The equicontinuity condition in Assumption 7.2 matches with that maintained in the ULLN by Pötscher and Prucha (1989) and in Theorem 6.13. The Lipschitz-type conditions in Assumption 7.3 match with those in the ULLN by Andrews (1987) and in Theorem 6.14.

Strong consistency results analogous to the weak consistency result given in Theorem 7.1 can be obtained if instead of Theorems 6.13 or 6.14 we use Theorem 6.15 to imply local LLNs. As can be seen from an inspection of Theorem 6.15 the resulting catalogues of assumptions are rather complex. The approach taken in Gallant (1987a, Ch.7) and Gallant and White (1988) leads to one of these catalogues.

---

[4]In the case of generalized method of moments estimators q_t is vector valued and hence the ULLNs are applied to each component.

[5]This case can be incorporated into the framework of the theorem by formally viewing the objective function as a function on $T \times B$, where T can be chosen as an arbitrary compact subset of some Euclidean space, and by setting $\hat{\tau}_n = \bar{\tau}_n = \tau_0$, where τ_0 is an arbitrary element of T.

8

BASIC STRUCTURE OF THE ASYMPTOTIC NORMALITY PROOF

In this chapter we describe the basic structure underlying the derivation of the asymptotic distribution of M-estimators in nonlinear econometric models. As remarked in Chapter 1, the basic methods used in this derivation date back to Doob (1934), Cramér (1946), LeCam (1953), Huber (1967) and Jennrich (1969), to mention a few; for a more extensive bibliography see Norden (1972, 1973) and the references in Chapter 1. The asymptotic normality proofs in the articles on nonlinear econometric models listed in Chapter 1 all share this common structure. The basic idea is to express the estimator as a linear function of the score vector by means of a Taylor series expansion and then to derive the asymptotic distribution of the estimator from the asymptotic distribution of the score vector.

We maintain the basic setup as described in Chapter 2. Let $Q_n : Z^n \times T \times B \to \mathbf{R}$ denote some objective function and let $(\mathbf{z}_t)$ be a stochastic process taking its values in Z. Let $\hat{\beta}_n$ and $\hat{\tau}_n$ be estimators with the typical interpretation of estimators of the parameters of interest and of nuisance parameters, respectively. In the following $\hat{\beta}_n$ need not necessarily be a minimizer of Q_n, but only an approximate solution of a set of corresponding first order conditions. As discussed in Chapters 2 and 7 the objective function will often take the form

$$Q_n(\mathbf{z}_1, \dots, \mathbf{z}_n, \hat{\tau}_n, \beta) = \vartheta_n \left(S_n(\omega, \hat{\tau}_n, \beta), \hat{\tau}_n, \beta\right) \tag{8.1}$$

with

$$S_n(\omega, \tau, \beta) = n^{-1} \sum_{t=1}^{n} q_t(\mathbf{z}_t, \tau, \beta)$$

where $q_t : Z \times T \times B \to \mathbf{R}^{p_q}$ and $\vartheta_n : \mathbf{R}^{p_q} \times T \times B \to \mathbf{R}$. The class of least mean distance estimators corresponds to $p_q = 1$ and $\vartheta_n(c, \tau, \beta) \equiv c$. The class of generalized method of moments estimators corresponds to the case where S_n denotes a vector of sample moments and ϑ_n represents some "distance" function like a quadratic form (in the vector of sample moments).

The subsequent assumption describes the basic setup employed in deriving the asymptotic distribution of $\hat{\beta}_n$. We have formulated this assumption

B. M. Pötscher et al., *Dynamic Nonlinear Econometric Models*

in a rather general fashion such that it is also applicable in certain non-ergodic situations, since the added generality does not significantly increase the complexity of the derivation of the asymptotic distribution result. The leading and classical case is the situation where the normalizing sequences M_n and N_n behave like $n^{1/2}I$.

Assumption 8.1. [1] *(a) The parameter spaces T and B are measurable subsets of Euclidean space $\mathbf{R}^{p_\tau}$ and $\mathbf{R}^{p_\beta}$, respectively.*

(b) $Q_n(\mathbf{z}_1, ..., \mathbf{z}_n, \tau, \beta)$ is $\mathfrak{A}$-measurable for all $(\tau, \beta) \in T \times B$ and $Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, ., .)$ is a.s. twice continuously partially differentiable at every point (τ, β) in the interior of $T \times B$ (where the exceptional null set does not depend on (τ, β)).

(c) The estimators $(\hat{\tau}_n, \hat{\beta}_n)$ take their values in $T \times B$. There exists a (non-random) sequence $(\bar{\tau}_n, \bar{\beta}_n) \in T \times B$, which is eventually uniformly in the interior[2] of $T \times B$, such that $\hat{\beta}_n - \bar{\beta}_n = o_p(1)$, $\hat{\tau}_n - \bar{\tau}_n = o_p(1)$ and $M_n(\hat{\tau}_n - \bar{\tau}_n) = O_p(1)$ for a sequence of (possibly random) square matrices M_n, which are non-singular with probability tending to one.

(d) The sequence $\hat{\beta}_n$ satisfies

$$nN_n^{+\prime}\nabla_{\beta'}Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \hat{\tau}_n, \hat{\beta}_n) = o_p(1)$$

for a sequence of (possibly random) square matrices N_n, which are non-singular with probability tending to one. (I.e., $\hat{\beta}_n$ satisfies the normalized first order conditions up to an error of magnitude $o_p(1)$.)[3]

(e) For all sequences of random vectors $(\tilde{\tau}_n, \tilde{\beta}_n)$ with $\tilde{\tau}_n - \bar{\tau}_n = o_p(1)$ and $\tilde{\beta}_n - \bar{\beta}_n = o_p(1)$ we have

$$nN_n^{+\prime}\nabla_{\beta\beta}Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \tilde{\tau}_n, \tilde{\beta}_n)N_n^{+} - C_n = o_p(1)$$

---

[1]We note that because of Assumption 8.1(c) there exists a sequence of sets $\Omega_n \in \mathfrak{A}$ with $P(\Omega_n) = 1$ such that $(\hat{\tau}_n, \hat{\beta}_n)$ and $(\tilde{\tau}_n, \tilde{\beta}_n)$ belong to the interior of $T \times B$ for $\omega \in \Omega_n$. Therefore the derivatives of Q_n evaluated at $(\hat{\tau}_n, \hat{\beta}_n)$ and $(\tilde{\tau}_n, \tilde{\beta}_n)$ considered below are well-defined at least for $\omega \in \Omega_n$. As usual, in the sequel we will often use the notation $\xi_n = o_p(a_n)$ or $\xi_n = O_p(a_n)$ even if the variables ξ_n are only well defined on sets $\Omega_n \in \mathfrak{A}$ with $P(\Omega_n) \to 1$ as $n \to \infty$.

[2]I.e., there exists an $\epsilon > 0$ such that the Euclidean distance from $(\bar{\tau}_n, \bar{\beta}_n)$ to the complement of $T \times B$ relative to $p_\tau + p_\beta$-dimensional Euclidean space exceeds ϵ for all large n.

[3]Let f be a $s \times 1$ vector of real valued functions defined on an open subset of $\mathbf{R}^p \times \mathbf{R}^r$, let $x = (x_1, \ldots, x_p)' \in \mathbf{R}^p$ and $y = (y_1, \ldots, y_r)' \in \mathbf{R}^r$. Then $\nabla_x f = (\partial f/\partial x_1, \ldots, \partial f/\partial x_p)$ is the $s \times p$ matrix of first order partial derivatives w.r.t. x and $\nabla_{x'} f = (\nabla_x f)'$. If $s = 1$, then $\nabla_{xy} f = \nabla_y(\nabla_{x'} f)$ denotes the $p \times r$ matrix of second order partial derivatives. More generally, if $s \geq 1$, then $\nabla_{xy} f = \nabla_y(\text{vec}(\nabla_{x'} f))$ denotes the $ps \times r$ matrix of second order partial derivatives. $\nabla_{xx} f$ and $\nabla_{yy} f$ are defined analogously. Furthermore, for a matrix A we denote the Moore-Penrose inverse by A^+. The inverse of a matrix A is denoted as usual by A^{-1}.

for a sequence of (possibly random) matrices C_n, which are non-singular with probability tending to one and satisfy $|C_n| = O_p(1)$ and $|C_n^+| = O_p(1)$.[4]

(f) For all sequences $(\tilde{\tau}_n, \tilde{\beta}_n)$ as in (e) we have

$$nN_n^{+\prime}\nabla_{\beta\tau}Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \tilde{\tau}_n, \tilde{\beta}_n)M_n^+ = o_p(1).$$

(g) There exists a sequence of (possibly random) matrices D_n with $|D_n| = O_p(1)$, such that

$$-nN_n^{+\prime}\nabla_{\beta'}Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \bar{\tau}_n, \bar{\beta}_n) = D_n\zeta_n + o_p(1)$$

where ζ_n and ζ are random vectors satisfying $\zeta_n \xrightarrow{D} \zeta$. (Here $\xrightarrow{D}$ denotes convergence in distribution.)

As remarked above, in the classical case the normalizing sequences will be of the form $M_n = n^{1/2}I$ and $N_n = n^{1/2}I$, C_n and D_n will be non-random, and ζ will be normally distributed. In non-ergodic situations, however, we may need to consider norming sequences other than $n^{1/2}I$ and limiting distributions other than the normal distribution, e.g., variance mixtures of normals. Furthermore, random matrix norming may sometimes be necessary in such situations. For recent contributions to the theory of non-ergodic models see, e.g., Basawa and Scott (1983), Park and Phillips (1988, 1989), Phillips (1989) and Wooldridge (1986).

Clearly, if $\hat{\beta}_n$ is an interior minimizer of $Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \hat{\tau}_n, \beta)$, or more generally, if $\hat{\beta}_n$ is a solution of the first order conditions, then Assumption 8.1(d) is trivially satisfied (for any choice of N_n). As discussed in Chapter 3, if $\hat{\beta}_n$ is a minimizer of $Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \hat{\tau}_n, \beta)$ then $\bar{\beta}_n$ will frequently be a minimizer of a non-random analogue $\bar{Q}_n$ of Q_n. Furthermore, Assumption 8.1 also covers the case where no nuisance parameter is present.[5] We return to a more detailed discussion of Assumption 8.1 after Corollary 8.2 below.

In the next lemma we give the asymptotic distribution of approximate solutions $\hat{\beta}_n$ of the first order conditions $\nabla_{\beta'}Q_n = 0$. As discussed in more detail below, the lemma can be readily reformulated for approximate solutions of estimating equations, say $F_n = 0$, which need not represent a set of first order conditions. We note further that for non-random C_n, M_n and N_n Assumption 8.1 implies $C_n^+ = C_n^{-1}$, $M_n^+ = M_n^{-1}$ and $N_n^+ = N_n^{-1}$ for large n; in this case we shall therefore always write C_n^{-1}, M_n^{-1} and N_n^{-1} ignoring finitely many n.

---

[4]Cf. Footnote 20 in Chapter 6 concerning the definition of the norm of a matrix.

[5]To incorporate this case into the framework of Assumption 8.1 we may formally view the objective function as a function on $T \times B$, where T can be chosen as an arbitrary subset of some Euclidean space with $\text{int}(T) \neq \emptyset$, and by setting $\hat{\tau}_n = \bar{\tau}_n = \tau_0$, where τ_0 is an arbitrary element of $\text{int}(T)$.

Lemma 8.1. *Given Assumption 8.1 holds, then*

$$N_n(\hat{\beta}_n - \bar{\beta}_n) = C_n^+ D_n \zeta_n + o_p(1)$$

where $\zeta_n \xrightarrow{D} \zeta$. *If furthermore* $C_n^+ D_n \to A$ *i.p., A non-random, then*

$$N_n(\hat{\beta}_n - \bar{\beta}_n) \xrightarrow{D} A\zeta.$$

More generally, for any sequence of (possibly random) matrices G_n *with* $|G_n| = O_p(1)$ *we have*

$$G_n N_n(\hat{\beta}_n - \bar{\beta}_n) = G_n C_n^+ D_n \zeta_n + o_p(1)$$

where $\zeta_n \xrightarrow{D} \zeta$. *If furthermore* $G_n C_n^+ D_n \to A^\bullet$ *i.p.,* $A^\bullet$ *non-random, then*

$$G_n N_n(\hat{\beta}_n - \bar{\beta}_n) \xrightarrow{D} A^\bullet \zeta.$$

The lemma provides a basic representation of the normalized parameter estimator (up to an error of magnitude $o_p(1)$) as a linear function of a random vector that converges in distribution. The lemma shows furthermore that linear transformations of the normalized parameter vector have an analogous representation where the approximation error remains of magnitude $o_p(1)$ as long as the transformation matrices G_n are bounded in probability.

Lemma 8.1 establishes a distributional convergence result for $N_n(\hat{\beta}_n - \bar{\beta}_n)$ in case $C_n^+ D_n$ converges to a limit, say A. In such a case the lemma implies that the cumulative distribution function of $N_n(\hat{\beta}_n - \bar{\beta}_n)$ converges to the cumulative distribution function of $A\zeta$ (in continuity points of the latter distribution function). This justifies the use of the latter distribution function as an approximation to the former one. If, e.g., ζ is normally distributed we obtain a normal approximation for the distribution of $N_n(\hat{\beta}_n - \bar{\beta}_n)$. In the general case, where $C_n^+ D_n$ does not converge, it is less obvious in which sense the distributions of $N_n(\hat{\beta}_n - \bar{\beta}_n)$ and of $C_n^+ D_n \zeta$ are "close" to each other. In the practically most important case where C_n and D_n can be chosen to be non-random the following is a consequence of Skorohod's Representation Theorem (Billingsley (1979, p.337)): There exist random variables $\tilde{\zeta}_n$ and $\tilde{\zeta}$ defined on some probability space such that $\tilde{\zeta}_n \stackrel{D}{=} \zeta_n$, $\tilde{\zeta} \stackrel{D}{=} \zeta$, where $\stackrel{D}{=}$ denotes equality in distribution, and $\tilde{\zeta}_n \to \tilde{\zeta}$ a.s. Because of boundedness of $C_n^{-1} D_n$ we then have

$$\begin{aligned} C_n^{-1} D_n \zeta_n &\stackrel{D}{=} C_n^{-1} D_n \tilde{\zeta}_n = C_n^{-1} D_n \tilde{\zeta} + o(1), \\ C_n^{-1} D_n \tilde{\zeta} &\stackrel{D}{=} C_n^{-1} D_n \zeta, \end{aligned}$$

and

$$N_n(\hat{\beta}_n - \bar{\beta}_n) = C_n^{-1} D_n \zeta_n + o_p(1).$$

The above three equations describe in a somewhat obscure way the "closeness" between the distribution of $N_n(\hat{\beta}_n - \bar{\beta}_n)$ and of $C_n^{-1} D_n \zeta$. In Lemma F2 and Corollaries F3 and F4 in Appendix F we discuss in detail under which circumstances the difference between the cumulative distribution functions of $N_n(\hat{\beta}_n - \bar{\beta}_n)$ and of $C_n^{-1} D_n \zeta$ converges to zero. In the most important special case where ζ is distributed $N(0, \Sigma)$, with Σ positive definite, it follows – under Assumption 8.1 – from Corollary F4(b) that the difference between the cumulative distribution function of $N_n(\hat{\beta}_n - \bar{\beta}_n)$ and that of a $N(0, C_n^{-1} D_n \Sigma D_n' C_n^{-1\prime})$ distribution converges pointwise to zero if C_n, D_n are non-random and the smallest eigenvalues of $C_n^{-1} D_n D_n' C_n^{-1\prime}$ are bounded away from zero, thus justifying the use of the latter distribution as an approximation to the former one. (Of course, the discussion in this paragraph also applies to the distributional relationship between $G_n N_n(\hat{\beta}_n - \bar{\beta}_n)$ and $G_n C_n^{-1} D_n \zeta$, if additionally the matrices G_n are non-random and bounded.)

Both in the above lemma as well as in the above discussion the distributions approximating the distributions of $N_n(\hat{\beta}_n - \bar{\beta}_n)$ (or $G_n N_n(\hat{\beta}_n - \bar{\beta}_n)$) will in general depend on the sample size n. It is clearly of interest to isolate circumstances under which the approximating distributions can be chosen to be independent of n. Of course, the simple but important case where the matrices $C_n^+ D_n$ (or $G_n C_n^+ D_n$) converge in probability to a non-random matrix is such an instance and has already been isolated in the lemma. The following corollary covers further important cases. Part (a) and part (b) of the corollary will be useful in the context of least mean distance and generalized method of moments estimators, respectively.

Corollary 8.2. [6] *Let Assumption 8.1 hold.*

(a) Let the matrices D_n be square and non-singular with probability tending to one and let $|D_n^+| = O_p(1)$, then

$$D_n^+ C_n N_n(\hat{\beta}_n - \bar{\beta}_n) = \zeta_n + o_p(1) \xrightarrow{D} \zeta.$$

(a') More generally, let the matrices D_n have full column rank with probability tending to one and let $|(D_n' D_n)^+| = O_p(1)$, then

$$(D_n' D_n)^+ D_n' C_n N_n(\hat{\beta}_n - \bar{\beta}_n) = \zeta_n + o_p(1) \xrightarrow{D} \zeta.$$

(b) Let the matrices C_n and D_n be non-random, let D_n have full row-rank (except possibly for finitely many n) and let $\left|(D_n D_n')^{-1}\right| = O(1)$. If ζ

[6] Let A be a symmetric nonnegative definite matrix, then $A^{1/2}$ denotes the unique symmetric and nonnegative definite square root of A. If A is furthermore nonsingular, then $A^{-1/2}$ denotes $(A^{-1})^{1/2}$.

is distributed as $N(0,\Sigma)$, Σ *non-singular, then*

$$(C_n^{-1}D_n\Sigma D_n'C_n^{-1\prime})^{-1/2}N_n(\hat{\beta}_n-\bar{\beta}_n)\xrightarrow{D}N(0,I).$$

(b') More generally, let the matrices C_n *and* D_n *be non-random, let* $\operatorname{rank}(D_n)=d$ *be constant (except possibly for finitely many n), and let* $|(D_nD_n')^+|=O(1)$. *If* ζ *is distributed as* $N(0,\Sigma)$, Σ *non-singular, then*

$$U_n'[(C_n^{-1}D_n\Sigma D_n'C_n^{-1\prime})^+]^{1/2}N_n(\hat{\beta}_n-\bar{\beta}_n)\xrightarrow{D}N(0,\operatorname{diag}(I_d,0)),$$

where U_n *is an orthogonal matrix of eigenvectors of* $C_n^{-1}D_n\Sigma D_n'C_n^{-1\prime}$ *whose first d columns correspond to the non-zero eigenvalues.*

The corollary implies, e.g., the following if ζ is distributed $N(0,\Sigma)$: Under the assumptions of part (a), (a'), (b) and (b'), respectively,

$$(\hat{\beta}_n-\bar{\beta}_n)'N_n'C_n'D_n^{+\prime}\Sigma^+D_n^+C_nN_n(\hat{\beta}_n-\bar{\beta}_n),$$

$$(\hat{\beta}_n-\bar{\beta}_n)'N_n'C_n'D_n(D_n'D_n)^+\Sigma^+(D_n'D_n)^+D_n'C_nN_n(\hat{\beta}_n-\bar{\beta}_n),$$

$$(\hat{\beta}_n-\bar{\beta}_n)'N_n'(C_n^{-1}D_n\Sigma D_n'C_n^{-1\prime})^{-1}N_n(\hat{\beta}_n-\bar{\beta}_n)$$

and

$$(\hat{\beta}_n-\bar{\beta}_n)'N_n'(C_n^{-1}D_n\Sigma D_n'C_n^{-1\prime})^+N_n(\hat{\beta}_n-\bar{\beta}_n)$$

are approximately chi-square distributed with numbers of degrees of freedom equal to $\operatorname{rank}(\Sigma)$, $\operatorname{rank}(\Sigma)$, p_β and d. This observation is of interest for the construction of test statistics like the Wald test statistic.

Typically, C_n plays the role of an asymptotic version of the Hessian matrix of the criterion function, and D_nD_n' can be – loosely speaking – thought of as the variance covariance matrix of the score vector (if the variance covariance matrix of ζ is the identity matrix). Of course, in the absence of normalizing assumptions the matrices D_n are not uniquely determined. For example, if D_n converges to D, say, then D_n can be absorbed into ζ_n.[7]

We next discuss Assumption 8.1 in more detail. Clearly, in case $\bar{\beta}_n\equiv\bar{\beta}$ and $\bar{\tau}_n\equiv\bar{\tau}$ the condition that $(\bar{\tau}_n,\bar{\beta}_n)$ lies uniformly in the interior of $T\times B$ postulated in Assumption 8.1(c) reduces to the condition that $(\bar{\tau},\bar{\beta})$ lies in the interior of $T\times B$. Furthermore, in that case we can (possibly after redefining $(\hat{\tau}_n,\hat{\beta}_n)$ on $\Omega-\Omega_n$ with $P(\Omega_n)\to 1$) reduce the parameter spaces T and B to suitably small Euclidean balls centered at $\bar{\tau}$ and $\bar{\beta}$, respectively, and perform the analysis on the Cartesian product of these two balls. An analogous reduction of the original parameter spaces to Euclidean balls

[7] Alternatively, if D_n is square and nonsingular with probability tending to one and if $|D_n^+|=O_p(1)$ then we could absorb D_n into C_n and N_n by replacing the matrices C_n, D_n, M_n, and N_n in Assumption 8.1 by $D_n^+C_nD_n^{+\prime}$, I, M_n, and $D_n'N_n$, respectively.

contained in the original parameter spaces can be made if, e.g., $\bar{\tau}_n$ and $\bar{\beta}_n$ happen to remain uniformly in the interior of the Cartesian product of two Euclidean balls. Therefore, in many cases, it is possible to reduce the asymptotic normality proof to the case of a compact parameter space, which contains $(\bar{\tau}_n, \bar{\beta}_n)$ uniformly in its interior, even if the original parameter space is not compact. This reduction to a compact parameter space is particularly helpful in establishing the convergence of the Hessian blocks as expressed in Assumption 8.1(e),(f), and is discussed in more detail below.

Inspection of the proof of Lemma 8.1 shows that $N_n(\hat{\beta}_n - \bar{\beta}_n)$ is asymptotically equivalent to a linear function of the score $\nabla_{\beta'} Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \bar{\tau}_n, \bar{\beta}_n)$ and of $\hat{\tau}_n - \bar{\tau}_n$, cf. equation (F.2') in Appendix F. Therefore, the asymptotic distribution of $\hat{\beta}_n$ would depend in general on that of the score vector as well as on that of $\hat{\tau}_n$. Assumption 8.1(f), in conjunction with $M_n(\hat{\tau}_n - \bar{\tau}_n) = O_p(1)$, eliminates this dependence on the asymptotic distribution of $\hat{\tau}_n$, i.e., it ensures that τ can be interpreted as a nuisance parameter.[8]

The convergence in distribution of ζ_n postulated in Assumption 8.1(g) will usually be deduced from a CLT and ζ will then be normally distributed. In case of least mean distance estimators we will usually be able to establish a CLT for the normalized score vector itself and D_n will be a square matrix with $|D_n^{-1}| = O(1)$, i.e.,

$$-nD_n^{-1} N_n^{+\prime} \nabla_{\beta'} Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \bar{\tau}_n, \bar{\beta}_n) = \zeta_n \xrightarrow{D} N(0, I)$$

will hold. For generalized method of moments estimators it is necessary to linearize the normalized score vector in terms of the vector of sample moments prior to the application of a CLT. This will then lead to the representation

$$-nN_n^{+\prime} \nabla_{\beta'} Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \bar{\tau}_n, \bar{\beta}_n) = D_n \zeta_n + o_p(1)$$

where D_n is now no longer a square matrix, unless the dimension of β equals the number of sample moments used in defining the estimator.

Lemma 8.1 can also be applied to derive the asymptotic distribution of estimators $\hat{\beta}_n$ as long as they correspond to solutions of a system of equations

$$F_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \hat{\tau}_n, \hat{\beta}_n) = 0,$$

or more generally satisfy

$$nN_n^{+\prime} F_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \hat{\tau}_n, \hat{\beta}_n) = o_p(1),$$

[8] Inspection of the proof of Lemma 8.1 shows that the assumption $M_n(\hat{\tau}_n - \bar{\tau}_n) = O_p(1)$ and Assumption 8.1(f) can be replaced by the higher level condition $nN_n^{+\prime} \nabla_{\beta\tau} Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \tilde{\tau}_n, \tilde{\beta}_n)(\hat{\tau}_n - \bar{\tau}_n) = o_p(1)$.

where F_n need not represent a set of first order conditions. Inspection of the proof shows that the lemma can be cast in terms of F_n by replacing $\nabla_{\beta'}Q_n, \nabla_{\beta\beta}Q_n$ and $\nabla_{\beta\tau}Q_n$ with $F_n, \nabla_\beta F_n$ and $\nabla_\tau F_n$, respectively, and by maintaining partial continuous differentiability of F_n instead of twice partial continuous differentiability of Q_n in Assumption 8.1(b).

The matrix $\nabla_{\beta\beta}Q_n$, and hence $N_n^{+\prime}\nabla_{\beta\beta}Q_n N_n^+$, is symmetric. Consequently, if Assumption 8.1(e) is satisfied for a sequence C_n, then it is also satisfied for a suitable sequence of symmetric matrices, e.g., for $(C_n + C_n')/2$.[9] Therefore C_n could have been restricted to be symmetric in Lemma 8.1.[10] However, the matrix $\nabla_\beta F_n$ will not necessarily be symmetric. Hence by allowing C_n to be nonsymmetric, Lemma 8.1 can be readily applied to estimators $\hat{\beta}_n$ satisfying

$$nN_n^{+\prime}F_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \hat{\tau}_n, \hat{\beta}_n) = o_p(1)$$

as discussed above.

The convergence conditions for the Hessian matrix in Assumptions 8.1(e), (f) will frequently be derived via uniform convergence. The following discussion of sufficient conditions focuses on the classical case where $M_n = n^{1/2}I, N_n = n^{1/2}I$, and C_n and D_n are non-random. In view of Lemma 3.2 the following Assumption 8.2 is then clearly sufficient for the convergence conditions on the Hessian matrix formulated in Assumptions 8.1(e) and (f). That is, Assumption 8.2 implies that for all sequences $(\tilde{\tau}_n, \tilde{\beta}_n)$ with $(\tilde{\tau}_n, \tilde{\beta}_n) - (\bar{\tau}_n, \bar{\beta}_n) = o_p(1)$ we have

$$\nabla_{\beta\beta}Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \tilde{\tau}_n, \tilde{\beta}_n) - C_n = o_p(1)$$

and

$$\nabla_{\beta\tau}Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \tilde{\tau}_n, \tilde{\beta}_n) = o_p(1).$$

Here $C_n = C_{1n}(\bar{\tau}_n, \bar{\beta}_n)$ where $C_{1n}(.,.)$ is as in Assumption 8.2.

Assumption 8.2. [11] *(a) There exist subsets T' and B' of the interior*

---

[9]Since $\nabla_{\beta\beta}Q_n$ is symmetric clearly $nN_n^{+\prime}\nabla_{\beta\beta}Q_n N_n^+ - C_n' = o_p(1)$ by Assumption 8.1(e), and hence $(C_n + C_n')/2 - C_n = o_p(1)$. It now follows immediately from Lemma F1 that Assumption 8.1(e) also holds with $(C_n + C_n')/2$ replacing C_n.

[10]If one imposes symmetry of C_n in Assumption 8.1(e), then it can be shown that Lemma 8.1 can be obtained even without the assumption of norm boundedness of C_n in Assumption 8.1(e). To obtain Corollary 8.2(a) and (a') one then has to replace the norm boundedness assumption on D_n^+ and $(D_n'D_n)^+$ by the norm boundedness of $D_n^+C_n$ and $(D_n'D_n)^+D_n'C_n$, respectively. For Corollary 8.2(b),(b'), however, norm boundedness of C_n is needed nevertheless.

[11]We note that the convergence conditions on the Hessian blocks in Assumption 8.1(e),(f) are equivalent to the condition that the Hessian blocks converge uniformly over any shrinking sequence of neighborhoods of $(\bar{\tau}_n, \bar{\beta}_n)$. Hence the

of T and B, respectively, such that $(\bar{\tau}_n, \bar{\beta}_n)$ is eventually uniformly in the interior of $T' \times B'$.[12]

(b) There exists a sequence of non-random $p_\beta \times p_\beta$ matrices $C_{1n}(\tau, \beta)$ defined on $T' \times B'$ such that

$$\sup_{T' \times B'} |\nabla_{\beta\beta} Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \tau, \beta) - C_{1n}(\tau, \beta)| = o_p(1)$$

and $\{C_{1n} : n \in \mathbf{N}\}$ is uniformly equicontinuous on $T' \times B'$.

(c) There exists a sequence of non-random $p_\beta \times p_\tau$ matrices $C_{2n}(\tau, \beta)$ defined on $T' \times B'$ such that

$$\sup_{T' \times B'} |\nabla_{\beta\tau} Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \tau, \beta) - C_{2n}(\tau, \beta)| = o_p(1),$$

$C_{2n}(\bar{\tau}_n, \bar{\beta}_n) = o(1)$, and $\{C_{2n} : n \in \mathbf{N}\}$ is uniformly equicontinuous on $T' \times B'$.

Clearly, given Assumption 8.1(c), Assumption 8.2(a) is trivially satisfied for $T' = \text{int}(T)$ and $B' = \text{int}(B)$. However, the verification of Assumptions 8.2(b),(c) will typically be easier if we find subsets T' and B' that are compact. Under Assumption 8.1(c) this is of course always possible in case $\bar{\tau}_n \to \bar{\tau}$ and $\bar{\beta}_n \to \bar{\beta}$ by choosing T' and B' as sufficiently small closed Euclidean balls centered at $\bar{\tau}$ and $\bar{\beta}$, respectively. Clearly, the provision of uniform equicontinuity in Assumption 8.2(b),(c) can be replaced by equicontinuity, if $T' \times B'$ is compact.

As discussed in detail in Chapter 11, in case of least mean distance estimators or generalized method of moments estimators we can frequently imply Assumption 8.2 from ULLNs. Recall that for least mean distance estimators the objective function Q_n is of the form

$$Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \tau, \beta) = n^{-1} \sum_{t=1}^{n} q_t(\mathbf{z}_t, \tau, \beta).$$

The uniform convergence conditions in Assumption 8.2 can then be implied from ULLNs for the elements of $\nabla_{\beta\beta} Q_n$ and $\nabla_{\beta\tau} Q_n$ if we put $C_{1n}(\tau, \beta) = E\nabla_{\beta\beta} Q_n$ and $C_{2n}(\tau, \beta) = E\nabla_{\beta\tau} Q_n$ for $(\tau, \beta) \in T' \times B'$. Equicontinuity of $C_{1n}(\tau, \beta)$ and $C_{2n}(\tau, \beta)$ is typically obtained as a by-product of ULLNs, cf.

uniform convergence condition on $T' \times B'$ maintained in Assumption 8.2 could be further weakened. Cf., e.g., Heijmans and Magnus (1986b), Weiss (1971, 1973) and Wooldridge (1986), Section 3.4.

[12]I.e., there exists an $\epsilon > 0$ such that the Euclidean distance from $(\bar{\tau}_n, \bar{\beta}_n)$ to the complement of $T' \times B'$ relative to $p_\tau + p_\beta$-dimensional Euclidean space exceeds ϵ for all large n. Furthermore observe that $(\tilde{\tau}_n, \tilde{\beta}_n)$ falls into $T' \times B'$ on sets Ω_n with $P(\Omega_n) \to 1$, given $(\tilde{\tau}_n, \tilde{\beta}_n) - (\bar{\tau}_n, \bar{\beta}_n) = o_p(1)$ and Assumption 8.2(a) holds.

Chapters 3 and 5. Of course, if $T' \times B'$ is compact, equicontinuity already implies uniform equicontinuity.

For generalized method of moments estimators the objective function is of the form

$$Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \tau, \beta) = \vartheta_n(S_n(\omega, \tau, \beta), \tau, \beta),$$

where

$$S_n(\omega, \tau, \beta) = n^{-1} \sum_{t=1}^{n} q_t(\mathbf{z}_t, \tau, \beta)$$

has the interpretation of a vector of moments. As will be discussed in more detail in Chapter 11, under the assumptions of that chapter it is easily checked that the matrices of second order derivatives $\nabla_{\beta\beta} Q_n$ and $\nabla_{\beta\tau} Q_n$ are, respectively, of the form

$$G_{1n}(S_n, \nabla_\beta S_n, \nabla_{\beta\beta} S_n, \tau, \beta)$$

and

$$G_{2n}(S_n, \nabla_\beta S_n, \nabla_\tau S_n, \nabla_{\beta\tau} S_n, \tau, \beta),$$

where G_{1n} and G_{2n} are appropriate functions defined on $E_1 \times T' \times B'$ and $E_2 \times T' \times B'$, and where E_1 and E_2 are Euclidean spaces of appropriate dimensions, cf. equation (11.2) in Chapter 11.[13] If (i) G_{1n} and G_{2n} are uniformly equicontinuous on $E_1 \times T' \times B'$ and $E_2 \times T' \times B'$, respectively, and (ii) the elements of S_n, $\nabla_\beta S_n$, $\nabla_\tau S_n$, $\nabla_{\beta\tau} S_n$, and $\nabla_{\beta\beta} S_n$ satisfy ULLNs on $T' \times B'$, then it follows from Lemma 3.3 that the convergence conditions in Assumption 8.2 are satisfied if we define

$$C_{1n}(\tau, \beta) = G_{1n}(ES_n, E\nabla_\beta S_n, E\nabla_{\beta\beta} S_n, \tau, \beta),$$

$$C_{2n}(\tau, \beta) = G_{2n}(ES_n, E\nabla_\beta S_n, E\nabla_\tau S_n, E\nabla_{\beta\tau} S_n, \tau, \beta)$$

for $(\tau, \beta) \in T' \times B'$ and if C_{2n} satisfies $C_{2n}(\bar{\tau}_n, \bar{\beta}_n) = o_p(1)$. Furthermore, the (uniform) equicontinuity of C_{1n} and C_{2n} on $T' \times B'$ required in Assumption 8.2 can then also be deduced from the corresponding property of the restrictions of ES_n, $E\nabla_\beta S_n$, $E\nabla_\tau S_n$, $E\nabla_{\beta\tau} S_n$, and $E\nabla_{\beta\beta} S_n$ to $T' \times B'$ in view of Lemma 3.3(b). Frequently, ULLNs employ dominance conditions which then entail that ES_n, $E\nabla_\beta S_n$, $E\nabla_\tau S_n$, $E\nabla_{\beta\tau} S_n$, and $E\nabla_{\beta\beta} S_n$ are bounded on $T' \times B'$ uniformly in n, and hence lie in compact subsets $K_1, \ldots, K_5$ of respective Euclidean spaces. This observation allows for an important simplification: Given $T' \times B'$ is also compact, inspection of Lemma 3.3 shows that it then suffices in (i) that G_{1n} and

[13] The functions G_{1n} and G_{2n} could in fact be defined on $E_1 \times \mathrm{int}(T \times B)$ and $E_2 \times \mathrm{int}(T \times B)$, respectively, cf. Chapter 11. However, for the following discussion we view G_{1n} and G_{2n} as functions defined only on $E_1 \times T' \times B'$ and $E_2 \times T' \times B'$, respectively.

G_{2n} are only equicontinuous on the subsets $K_1 \times K_2 \times K_5 \times T' \times B'$ and $K_1 \times K_2 \times K_3 \times K_4 \times T' \times B'$ of $E_1 \times T' \times B'$ and $E_2 \times T' \times B'$, respectively. Furthermore, in this situation equicontinuity of G_{1n} and G_{2n} on these respective subsets is easily seen to be satisfied, if $\vartheta_n(c, \tau, \beta)$ is twice continuously partially differentiable on an open set containing $K_1 \times T' \times B'$, if the derivatives $\nabla_c \vartheta_n(c, \tau, \beta)$, $\nabla_{cc} \vartheta_n(c, \tau, \beta)$, $\nabla_{\beta c} \vartheta_n(c, \tau, \beta)$, $\nabla_{\tau c} \vartheta_n(c, \tau, \beta)$, $\nabla_{\beta\tau} \vartheta_n(c, \tau, \beta)$, and $\nabla_{\beta\beta} \vartheta_n(c, \tau, \beta)$ are equicontinuous on the subset $K_1 \times T' \times B'$ of $\mathbf{R}^{p_q} \times T' \times B'$, and if these derivatives are bounded on $K_1 \times T' \times B'$ uniformly in n. Of course, if additionally $\vartheta_n \equiv \vartheta$, then these equicontinuity conditions reduce to continuity conditions and the uniform boundedness condition is automatically satisfied since $K_1 \times T' \times B'$ is compact.

9

ASYMPTOTIC NORMALITY UNDER NONSTANDARD CONDITIONS

The standard approach for deriving the asymptotic distribution of M-estimators outlined in the previous chapter relies on the assumption that the objective function Q_n is twice continuously differentiable w.r.t. both the parameter of interest β and the nuisance parameter τ. (Or, if the estimator $\hat{\beta}_n$ is derived as an approximate solution of a set of estimating equations $F_n = 0$, it is maintained that F_n is continuously differentiable.) In a number of applications this smoothness assumption is too stringent. E.g., if the objective function corresponds to the least absolute deviation estimator or Huber's M-estimator this assumption is violated. Also in a semiparametric context, where τ represents an infinite dimensional nuisance parameter that varies in a metric space T which is not a subset of Euclidean space, the notion of differentiability w.r.t. τ may not be available, although Q_n may be smooth as a function of β for every given value of τ. Such situations can often be handled by a refinement of the argument underlying Lemma 8.1. The basic idea is again to show that the (normalized) estimator $\hat{\beta}_n$ is asymptotically equivalent to a linear transformation of the score vector evaluated at the true parameter, and then to invoke a CLT for the score vector. Of course, in the absence of the smoothness assumptions of Chapter 8, establishing such a linear transformation is now more delicate (and, if Q_n is not differentiable at all, special care has to be given to defining the notion of a score vector properly). The linearization is frequently attempted by showing that the objective function can – in a certain sense – be replaced by its asymptotic counterpart $\bar{Q}_n$ and by exploiting the usually greater degree of smoothness of the latter function in the linearization argument. The fact that $\bar{Q}_n$ is frequently a smooth function, even when Q_n is not, originates from the fact that $\bar{Q}_n$ is frequently equal to EQ_n or $\lim_{n\to\infty} EQ_n$ and taking expectations is a smoothing operation. This approach was pioneered by Daniels (1961) and Huber (1967). For a modern exposition see Pollard (1985). The following discussion will be informal.

To fix ideas we follow Huber (1967) and assume that no nuisance parameter is present and that the estimator $\hat{\beta}_n$ satisfies

$$\nabla_{\beta'} Q_n(\omega, \hat{\beta}_n) = o_p(n^{-1/2}).$$

In particular, we assume here that the objective function Q_n is differentiable but not necessarily twice continuously differentiable. This covers, e.g., Huber's M-estimator. (The argument given below can be adapted to cover estimators derived from non-differentiable objective functions, see, e.g., Pollard (1985).) We also assume for simplicity that $\nabla_{\beta'} Q_n(\omega, \beta)$ converges in probability to

$$\lambda(\beta) = \lim_{n\to\infty} E\nabla_{\beta'} Q_n(\omega, \beta)$$

and that $\hat{\beta}_n \to \bar{\beta}$ in probability where $\lambda(\bar{\beta}) = 0$. Although $\nabla_{\beta'} Q_n$ is not assumed to be continuously differentiable, the function λ will frequently be continuously differentiable, due to the smoothing effect of taking expectations. Clearly

$$\begin{aligned} o_p(1) &= n^{1/2}\nabla_{\beta'} Q_n(\omega, \hat{\beta}_n) \qquad (9.1)\\ &= n^{1/2}\left[\nabla_{\beta'} Q_n(\omega, \hat{\beta}_n) - \lambda(\hat{\beta}_n) - \nabla_{\beta'} Q_n(\omega, \bar{\beta})\right] \\ &\quad + n^{1/2}\left[\lambda(\hat{\beta}_n) + \nabla_{\beta'} Q_n(\omega, \bar{\beta})\right]. \end{aligned}$$

If we can establish that

$$n^{1/2}\left[\nabla_{\beta'} Q_n(\omega, \hat{\beta}_n) - \lambda(\hat{\beta}_n) - \nabla_{\beta'} Q_n(\omega, \bar{\beta})\right] = o_p(1) \qquad (9.2)$$

we get

$$o_p(1) = n^{1/2}\left[\lambda(\hat{\beta}_n) + \nabla_{\beta'} Q_n(\omega, \bar{\beta})\right]. \qquad (9.3)$$

Recall that λ is assumed to be continuously differentiable and that $\lambda(\bar{\beta}) = 0$. From (9.3) we now get via a Taylor expansion of λ that asymptotically $n^{1/2}(\hat{\beta}_n - \bar{\beta})$ is a linear function of the score vector, i.e.,

$$n^{1/2}(\hat{\beta}_n - \bar{\beta}) = -\left[\nabla_{\beta}\lambda(\bar{\beta})\right]^{-1} n^{1/2}\nabla_{\beta'} Q_n(\omega, \bar{\beta}) + o_p(1). \qquad (9.4)$$

Asymptotic normality of $n^{1/2}(\hat{\beta}_n - \bar{\beta})$ now follows upon establishing asymptotic normality of $n^{1/2}\nabla_{\beta'} Q_n(\omega, \bar{\beta})$ via a CLT.

Of course, the verification of (9.2) is non-trivial and represents a crucial step in this approach. Huber (1967) provides conditions under which (9.2) holds in the i.i.d. context. Clearly, (9.2) also holds if the random functions $n^{1/2}[\nabla_{\beta'} Q_n(\omega, \beta) - \lambda(\beta)]$ are stochastically equicontinuous at $\bar{\beta}$, i.e., if for each $\epsilon > 0$

$$\limsup_{n\to\infty} P\left(\sup_{\rho_B(\beta,\bar{\beta})<\delta} \left| n^{1/2}\left[\nabla_{\beta'} Q_n(\omega, \beta) - \lambda(\beta)\right] - n^{1/2}\left[\nabla_{\beta'} Q_n(\omega, \bar{\beta}) - \lambda(\bar{\beta})\right]\right| > \epsilon \right)$$

goes to zero as $\delta \to 0$.[1] Of course, this raises the question how to verify the stochastic equicontinuity property. In the context of independent observations the theory of empirical processes provides a number of techniques for verifying stochastic equicontinuity and we may draw on these results, see, e.g., Pollard (1984, 1985). For dependent observations some sets of sufficient conditions can be found in Andrews (1989b, 1991a, 1993), Andrews and Pollard (1989, 1994), Arcones and Yu (1994), Doukhan, Massart and Rio (1995), and Hansen (1996).

Next consider the case where Q_n depends on a possibly infinite dimensional nuisance parameter τ and is not differentiable w.r.t. this parameter. Assume for simplicity that $\bar{\tau}_n \equiv \bar{\tau}$. If Q_n is twice continuously differentiable w.r.t. β for each value of τ and if $\hat{\beta}_n$ satisfies

$$o_p(1) = n^{1/2}\nabla_{\beta'}Q_n(\omega, \hat{\tau}_n, \hat{\beta}_n),$$

then using a Taylor expansion w.r.t. β only we arrive at

$$o_p(1) = n^{1/2}\nabla_{\beta'}Q_n(\omega, \hat{\tau}_n, \bar{\beta}_n) + \nabla_{\beta\beta}Q_n(\omega, \hat{\tau}_n, \{\tilde{\beta}_n^i\})n^{1/2}(\hat{\beta}_n - \bar{\beta}_n), \quad (9.5)$$

where $\nabla_{\beta\beta}Q_n(\omega, \hat{\tau}_n, \{\tilde{\beta}_n^i\})$ denotes the matrix whose j-th row is the j-th row of $\nabla_{\beta\beta}Q_n$ evaluated at $(\omega, \hat{\tau}_n, \tilde{\beta}_n^j)$ and where $\tilde{\beta}_n^j$ are mean values. The second term on the r.h.s. of (9.5) can be handled exactly as in the proof of Lemma 8.1. However, the first term can now no longer easily be shown to be asymptotically normal by appealing directly to a CLT, since it still contains the estimator $\hat{\tau}_n$ for the nuisance parameter τ. If we can show that

$$n^{1/2}\nabla_{\beta'}Q_n(\omega, \hat{\tau}_n, \bar{\beta}_n) - n^{1/2}\nabla_{\beta'}Q_n(\omega, \bar{\tau}, \bar{\beta}_n) = o_p(1) \quad (9.6)$$

holds, then (9.5) becomes

$$o_p(1) = n^{1/2}\nabla_{\beta'}Q_n(\omega, \bar{\tau}, \bar{\beta}_n) + \nabla_{\beta\beta}Q_n(\omega, \hat{\tau}_n, \{\tilde{\beta}_n^i\})n^{1/2}(\hat{\beta}_n - \bar{\beta}_n) \quad (9.7)$$

and we can then proceed as in the proof of Lemma 8.1 and appeal to a CLT to establish the asymptotic normality of the score vector evaluated at $(\bar{\tau}, \bar{\beta}_n)$, which then implies the asymptotic normality of $\hat{\beta}_n$. Of course, verifying (9.6) is not trivial. Again, if $\hat{\tau}_n - \bar{\tau} \to 0$ i.p. (9.6) is implied if, e.g., the random functions $n^{1/2}\nabla_{\beta'}Q_n(\omega, \tau, \bar{\beta}_n)$ are stochastically equicontinuous at $\bar{\tau}$, i.e., if for each $\epsilon > 0$

$$\limsup_{n\to\infty} P\left(\sup_{\rho_T(\tau,\bar{\tau})<\delta}\left|n^{1/2}\nabla_{\beta'}Q_n(\omega, \tau, \bar{\beta}_n) - n^{1/2}\nabla_{\beta'}Q_n(\omega, \bar{\tau}, \bar{\beta}_n)\right| > \epsilon\right)$$

goes to zero as $\delta \to 0$.

[1] The stochastic equicontinuity condition basically controls the modulus of continuity of the random function under consideration and is essentially a tightness condition for the sequence of random functions.

In the case of generalized method of moments estimators the objective function is a nonlinear function of a vector of sample moments. In this case it proves useful – rather than to attempt to directly verify that the normalized score vector is stochastically equicontinuous – to first linearize the score vector with respect to the vector of sample moments. Under certain assumptions it then suffices to establish stochastic equicontinuity of the vector of sample moments. This is essentially the approach taken by Andrews (1989a).

10

CENTRAL LIMIT THEOREMS

A key ingredient for the asymptotic normality proof, as outlined in Chapter 8, is that the normalized score vector can be expressed as a linear function of random variables ζ_n which converge in distribution, cf. Assumption 8.1(g). In this chapter we present central limit theorems (CLTs) which can be used to imply this distributional convergence of ζ_n in the important case where ζ_n can be expressed as a normalized sum of random variables. We give two alternative CLTs.

The first CLT presented applies to random variables that form a martingale difference sequence. This CLT is particularly relevant for the case of a correctly specified maximum likelihood problem. More specifically, if the objective function is the correctly specified log-likelihood function corresponding to the process $\mathbf{z}_t$, then the score vector (at the true parameter) can be written as the sum of respective scores per observation conditional on past information, which form a martingale difference sequence under weak regularity conditions. A further instance, in which the score vector is composed of martingale differences, is the case where a dynamic nonlinear regression model is estimated by the least squares method and the model is correctly specified in the sense that the response function at the true parameter represents the mean of the endogenous variable conditional on current explanatory variables and on all lagged endogenous and explanatory variables. However, if the model is correctly specified only in the sense that the response function at the true parameter value represents the mean of the endogenous variable conditional on *partial* information (e.g., conditional on current and lagged explanatory variables only) then the least squares score vector will in general not be composed of martingale differences, since the disturbances in the true model may then be autocorrelated. Note that in both cases the regression model is correctly specified in the sense that the model correctly describes the response function.[1]

[1]Hence, if a model is correctly specified in the sense that the model captures the probability structure of the data generating process, but the objective function is not the true likelihood function, then the random variables of which the score vector is composed may or may not possess the martingale difference property. (We also note that it is not impossible – although unlikely – for the random variables of which the score vector is composed to have the martingale difference

B. M. Pötscher et al., *Dynamic Nonlinear Econometric Models*

In the general case, including the case of misspecified models, the random variables of which the score vector is composed do not necessarily have the structure of a martingale difference sequence. Nevertheless CLTs can be established if these random variables exhibit a suitable form of "weak dependence" over time. The second CLT given will apply to such situations.

10.1 A Central Limit Theorem for Martingale Differences

In this section we introduce a CLT for martingale differences that are L_0-approximable by an α-mixing basis process. This theorem is based on a martingale difference CLT by McLeish (1974); cf. Theorem 3.2 in Hall and Heyde (1980). It generalizes the CLT for stochastically stable martingale differences given in Bierens (1981, 1984) in that it relaxes an asymptotic stationarity assumption made in these papers. Bierens' CLT was based on Brown's (1971) martingale difference CLT.

As in Chapter 6, $(\mathbf{v}_t)_{t\in\mathbf{N}}$ and $(\mathbf{e}_t)_{t\in\mathbf{Z}}$ denote stochastic processes defined on $(\Omega, \mathfrak{A}, P)$ that take their values in $\mathbf{R}^{p_v}$ and $\mathbf{R}^{p_e}$, respectively. Furthermore, let

$$V_n = E\left[\left(\sum_{t=1}^{n}\mathbf{v}_t\right)\left(\sum_{t=1}^{n}\mathbf{v}_t'\right)\right].$$

Clearly, if $(\mathbf{v}_t)_{t\in\mathbf{N}}$ is a martingale difference sequence then

$$V_n = \sum_{t=1}^{n} E\mathbf{v}_t\mathbf{v}_t'.$$

Theorem 10.1. *Let* $(\mathbf{v}_t)$ *be a martingale difference sequence (w.r.t. some filtration* $(\mathfrak{F}_t)$*) and let* $(\mathbf{v}_t)$ *be* L_0*-approximable by* $(\mathbf{e}_t)$ *where* $(\mathbf{e}_t)$ *is* α*-mixing. Assume that*

$$\sup_n n^{-1}\sum_{t=1}^{n} E\left|\mathbf{v}_t\right|^{2+\delta} < \infty$$

for some $\delta > 0$ *holds.*

(a) If $\liminf_{n\to\infty}\lambda_{\min}(n^{-1}V_n) > 0$*, then*

$$V_n^{-1/2}\sum_{t=1}^{n}\mathbf{v}_t \xrightarrow{D} N(0, I).$$

---

property in certain misspecified cases.)

(b) If $n^{-1}V_n \to V^*$, then

$$n^{-1/2}\sum_{t=1}^{n}\mathbf{v}_t \xrightarrow{D} N(0,V^*).$$

(V^* may be singular. In this case $N(0,V^*)$ denotes the degenerate normal distribution with zero mean and variance covariance matrix V^*.)

Part (b) of the theorem is essentially a multivariate version of Bierens' (1981, 1984) CLT. Part (a) generalizes this CLT in that it does not require convergence of the (normalized) variance covariance matrices and thus also applies to processes which do not satisfy this kind of asymptotic stationarity condition. Clearly, if $n^{-1}V_n$ does not converge, we can not expect to obtain a CLT by using $n^{-1/2}$ as normalizing constants. Part (a) shows that the matrix $V_n^{-1/2}$ represents the proper normalization.[2]

In proving the above CLT we verify, in particular, that the "norming condition" of Theorem 3.2 in Hall and Heyde (1980) holds. This condition essentially boils down to a LLN for $\mathbf{v}_t\mathbf{v}_t'$. Given the maintained assumptions of the above CLT, this LLN is derived from the LLN for L_0-approximable processes presented in Theorem 6.3.

10.2 A Central Limit Theorem for Functions of Mixing Processes

As remarked above, the assumption that the random variables of which the score vector is composed form a martingale difference will typically be satisfied only in certain correctly specified cases. To derive asymptotic normality results for the general case, including the misspecified case, we need a CLT that does not rely on the martingale assumption. Also, as discussed in more detail in Chapter 11 below, the random variables of which the score vector is composed will under misspecification typically depend on n, i.e., form a doubly indexed array. Before giving the CLT we hence have to extend the definition of near epoch dependence to doubly indexed arrays.[3]

Definition 10.1. *Let $(\mathbf{v}_{t,n} : t \in \mathbf{N},\ n \in \mathbf{N})$ and $(\mathbf{e}_t)_{t\in\mathbf{Z}}$ be stochastic processes that take their values in $\mathbf{R}^{p_v}$ and $\mathbf{R}^{p_e}$, respectively. Then the process $(\mathbf{v}_{t,n})$ is called near epoch dependent of size $-q$ on the basis process*

[2]We note that V_n^{-1} and $V_n^{-1/2}$ may not be well-defined for finitely many n.

[3]We note that the CLT for martingale difference sequences presented in Theorem 10.1 can also be formulated for doubly indexed arrays.

$(\mathbf{e}_t)$ *if the sequence*

$$\nu_m = \sup_n \sup_t \|\mathbf{v}_{t,n} - E(\mathbf{v}_{t,n} \mid \mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m})\|_2$$

is of size $-q$, $q > 0$.

The following CLT is applicable to processes which are near epoch dependent on an α-mixing or ϕ-mixing basis process. We define similarly as before

$$V_n = E\left[\left(\sum_{t=1}^{n} \mathbf{v}_{t,n}\right)\left(\sum_{t=1}^{n} \mathbf{v}'_{t,n}\right)\right].$$

Theorem 10.2. *Let* $(\mathbf{v}_{t,n})$ *be near epoch dependent of size* -1 *on* $(\mathbf{e}_t)$ *where* $(\mathbf{e}_t)$ *is* α*-mixing with mixing coefficient of size* $-2r/(r-2)$ *or* ϕ*-mixing with mixing coefficients of size* $-r/(r-1)$*, for some* $r > 2$*. Assume that* $E\mathbf{v}_{t,n} = 0$ *and* $\sup_n \sup_t E\left|\mathbf{v}_{t,n}\right|^r < \infty$.

(a) If $\liminf_{n\to\infty} \lambda_{\min}(n^{-1}V_n) > 0$*, then*

$$V_n^{-1/2} \sum_{t=1}^{n} \mathbf{v}_{t,n} \xrightarrow{D} N(0, I).$$

(b) If $n^{-1}V_n \to V^*$*, then*

$$n^{-1/2} \sum_{t=1}^{n} \mathbf{v}_{t,n} \xrightarrow{D} N(0, V^*).$$

(V^* *may be singular. In this case* $N(0, V^*)$ *denotes the degenerate normal distribution with zero mean and variance covariance matrix* V^**.)*

Theorem 10.2(a) is a multivariate version of Corollary 4.4 in Wooldridge (1986), see also Theorem 5.3 in Gallant and White (1988). A more restrictive version of this theorem is also given in Theorem 2 in Gallant (1987a, p.519). The proof of Corollary 4.4 in Wooldridge (1986) is similar in spirit to the proof of a CLT for dependent processes in Withers (1981b, 1983). For related results see De Jong (1995b).

McLeish (1975b, 1977) presents functional CLTs for L_2-mixingales. Under the assumptions of Theorem 10.2 the process $(\mathbf{v}_{t,n})$ can be shown to be an L_2-mixingale, see McLeish (1975a), Theorem 3.1, and Gallant and White (1988), Lemma 3.14. Hence McLeish's functional CLTs for L_2-mixingales contain functional CLTs (and hence CLTs) for near epoch dependent processes. However, in order to obtain *functional* CLTs, McLeish (1975b, 1977) maintains somewhat stronger asymptotic stationarity assumptions than maintained in Theorem 10.2(a), cf. also Wooldridge and White (1988).[4]

We note that Theorem 10.2(a) does not require convergence of the variance covariance matrices $n^{-1}V_n$ and thus (in this sense) Theorems 10.1(a) and 10.2(a) allow for the same degree of heterogeneity. In both Theorem 10.1(a) and 10.2(a) the degree of heterogeneity is limited by two factors: (i) The assumption that the smallest eigenvalues of $n^{-1}V_n$ are bounded away from zero and (ii) the respective boundedness assumptions on moments of order higher than two, which imply that the largest eigenvalues of $n^{-1}V_n$ are bounded from above.[5] Hence, although Theorems 10.1(a) and 10.2(a) do not require convergence of the variance covariance matrices $n^{-1}V_n$, both theorems maintain implicitly that $c_1 I \leq n^{-1}V_n \leq c_2 I$, with c_1, c_2 positive and finite.[6] (Of course, under Theorems 10.1(b) and 10.2(b) we only have $n^{-1}V_n \leq c_2 I$.)

As the LLN for near epoch dependent processes given in Theorem 6.4, also Theorem 10.2 shows trade-offs between the moment condition on the process $(\mathbf{v}_{t,n})$ and the size of the mixing coefficients. As compared to that LLN also the size of the approximation error in the definition of near epoch dependence has to be smaller for the CLT to hold.

Clearly, in situations where the stochastic process under consideration can be established to be a martingale difference sequence, Theorem 10.1 is preferable to Theorem 10.2 in that the former theorem requires only L_0-approximability, no rate of decline for the mixing coefficients, and a somewhat less stringent moment condition.

---

[4]Further recent contributions to functional CLTs for dependent processes include Peligrad (1981), Herrndorf (1983, 1984), and Doukhan, Massart and Rio (1994); see also Eberlein and Taqqu (1986).

[5]More specifically, under the assumptions of Theorem 10.1 the boundedness of $n^{-1}V_n$ follows from Lyapunov's inequality since $V_n = \Sigma_{t=1}^n E\mathbf{v}_t\mathbf{v}_t'$. Under the assumptions of Theorem 10.2 $(\mathbf{v}_{t,n} : t \in \mathbf{N})$ is a mixingale of size -1 with coefficients $c_{tn} = \max\{\|\mathbf{v}_{t,n}\|_r, 1\}$ for each $n \in \mathbf{N}$; cf. McLeish (1975a), Theorem 3.1, and Gallant and White (1988), Lemma 3.14. Since c_{tn} is bounded the boundedness of $n^{-1}V_n$ follows from McLeish's (1975a) inequality; cf. also Theorem 3.11 in Gallant and White (1988).

[6]The inequality $c_1 I \leq n^{-1}V_n$ with $c_1 > 0$ may only hold for all but finitely many $n \in \mathbf{N}$.

11

ASYMPTOTIC NORMALITY: CATALOGUES OF ASSUMPTIONS

Based on the discussion in Chapters 8 and 10 it is now possible to provide various sets of sufficient conditions for the asymptotic normality of M-estimators in dynamic nonlinear models. In Sections 11.1 and 11.2 we establish the asymptotic normality of least mean distance estimators and of generalized method of moments estimators under exemplary catalogues of assumptions. In Section 11.3 we relate these results to those available in the econometrics literature and provide further remarks.

The approach taken in the following is to establish asymptotic normality via Lemma 8.1 and Corollary 8.2 by verifying the conditions maintained by the lemma and corollary from lower level conditions. In both Sections 11.1 and 11.2 we present a catalogue of assumptions for the case where a martingale structure (for the random variables of which the score vector is composed) is available as well as a catalogue of assumptions for the general case, including the misspecified case. The reason for providing a separate catalogue of assumptions for the case where a martingale structure is available is that we can obtain asymptotic normality under weaker assumptions by exploiting this martingale structure.

As discussed in Chapter 8 we consider M-estimators $\hat{\beta}_n$ corresponding to an objective function of the form

$$Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \hat{\tau}_n, \beta) = \vartheta_n\left(S_n(\omega, \hat{\tau}_n, \beta), \hat{\tau}_n, \beta\right) \tag{11.1}$$

with

$$S_n(\omega, \tau, \beta) = n^{-1} \sum_{t=1}^{n} q_t(\mathbf{z}_t, \tau, \beta)$$

where $q_t : Z \times T \times B \to \mathbf{R}^{p_q}$ and $\vartheta_n : \mathbf{R}^{p_q} \times T \times B \to \mathbf{R}$. As discussed, the class of least mean distance estimators corresponds to $p_q = 1$ and $\vartheta_n(c, \tau, \beta) \equiv c$. The class of generalized method of moments estimators corresponds to the case where S_n denotes a vector of sample moments and ϑ_n represents a "distance" function. The following assumptions will be used both for least mean distance and generalized method of moments estimators.

Assumption 11.1. *(a) The parameter spaces T and B are measurable subsets of Euclidean space $\mathbf{R}^{p_\tau}$ and $\mathbf{R}^{p_\beta}$, respectively.*

(b) For each $(\tau,\beta) \in T \times B$ the functions $q_t(.,\tau,\beta)$ are $\mathfrak{Z}$-measurable. For each $z \in Z$ the functions $q_t(z,.,.)$ are twice continuously partially differentiable at every point (τ,β) in the interior of $T \times B$.

(c) The estimators $(\hat{\tau}_n, \hat{\beta}_n)$ take their values in $T \times B$ and $\hat{\beta}_n$ satisfies the first order conditions up to an error of order less than $n^{-1/2}$, i.e.,

$$\nabla_{\beta'} Q_n(\mathbf{z}_1,\ldots,\mathbf{z}_n,\hat{\tau}_n,\hat{\beta}_n) = o_p(n^{-1/2}).^1$$

(d) The sequence of estimators $(\hat{\tau}_n, \hat{\beta}_n)$ satisfies $\hat{\beta}_n - \bar{\beta}_n = o_p(1)$ and $\hat{\tau}_n - \bar{\tau}_n = O_p(n^{-1/2})$ for some (non-random) sequence $(\bar{\tau}_n, \bar{\beta}_n)$. The sequence $(\bar{\tau}_n, \bar{\beta}_n)$ lies uniformly in the interior of a compact set $T' \times B'$, which itself is contained in the interior of $T \times B$.[2]

(e) The set Z is a Borel subset of $\mathbf{R}^{p_z}$ (with $\mathfrak{Z}$ the induced Borel σ-field). The process $(\mathbf{z}_t)$ is L_0-approximable by some α-mixing basis process, say, $(\mathbf{e}_t)$.

(f) The sequence $\{n^{-1}\sum_{t=1}^n H_t^z : n \in \mathbf{N}\}$ is tight on Z, where H_t^z denotes the distribution of $\mathbf{z}_t$.

We remark, similarly as in Chapter 8, that the parameter spaces T and B in the above assumption need not represent the original parameter spaces of the estimation problem under consideration, but can be appropriately chosen reductions of the original spaces. E.g., if $\bar{\tau}_n \equiv \bar{\tau}$ and $\bar{\beta}_n \equiv \bar{\beta}$ then T and B can be chosen as neighborhoods of $\bar{\tau}$ and $\bar{\beta}$, respectively. We note further that the results in this chapter clearly also apply to estimation problems that do not contain a nuisance parameter, cf. the corresponding discussion in Chapter 8. In the following we shall write for ease of notation $\underline{S}_n$, $\nabla_\beta \underline{S}_n$, $\nabla_\tau \underline{S}_n$, $\nabla_{\beta\tau} \underline{S}_n$, $\nabla_{\beta\beta} \underline{S}_n$, respectively, for S_n, $\nabla_\beta S_n$, $\nabla_\tau S_n$, $\nabla_{\beta\tau} S_n$, $\nabla_{\beta\beta} S_n$ evaluated at $(\omega, \bar{\tau}_n, \bar{\beta}_n)$.

---

[1]Note that in view of part (d) of the assumption the derivatives $\nabla_{\beta'} Q_n(\mathbf{z}_1,\ldots,\mathbf{z}_n,\hat{\tau}_n,\hat{\beta}_n)$ are well defined with probability tending to one, cf. also Footnote 1 in Chapter 8.

[2]Clearly, if we can establish that the sequence $(\bar{\tau}_n, \bar{\beta}_n)$ is contained in a compact set, which is a subset of the interior of $T \times B$, then we can always find a compact set $T' \times B'$, which is also a subset of the interior of $T \times B$, such that now $(\bar{\tau}_n, \bar{\beta}_n)$ lies uniformly in the interior of $T' \times B'$. The latter property of $T' \times B'$ will turn out to be convenient in what follows and therefore part (d) of the assumption is formulated w.l.o.g. in this seemingly more restrictive form.

11.1 Asymptotic Normality of Least Mean Distance Estimators

For least mean distance estimators Assumptions 8.1(a)-(d) with $M_n = n^{1/2}I$ and $N_n = n^{1/2}I$ are clearly implied by Assumptions 11.1(a)-(d), since in this case

$$Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \tau, \beta) = S_n(\omega, \tau, \beta) = n^{-1} \sum_{t=1}^{n} q_t(\mathbf{z}_t, \tau, \beta).$$

Assumption 11.1 and the following assumption allow us to establish uniform convergence of the Hessian blocks, which is essential for the verification of Assumption 8.1(e),(f); cf. the discussion after Assumption 8.2.

Assumption 11.2. *The family $\{f_t : t \in \mathbf{N}\}$ is equicontinuous on $Z \times T' \times B'$ and*

$$\sup_n n^{-1} \sum_{t=1}^{n} E \left[\sup_{T' \times B'} |f_t(\mathbf{z}_t, \tau, \beta)|^{1+\gamma} \right] < \infty$$

for some $\gamma > 0$, where f_t denotes the restriction to $Z \times T' \times B'$ of any of the components of $\nabla_{\beta\tau} q_t$ or $\nabla_{\beta\beta} q_t$. (Here $T' \times B'$ is the compact subset employed in Assumption 11.1(d).)[3]

Define on $T' \times B'$ the matrices $C_{1n}(\tau, \beta)$ and $C_{2n}(\tau, \beta)$ by

$$C_{1n}(\tau, \beta) = E \nabla_{\beta\beta} S_n(\omega, \tau, \beta)$$

and

$$C_{2n}(\tau, \beta) = E \nabla_{\beta\tau} S_n(\omega, \tau, \beta).$$

The following lemma then gives uniform convergence of the Hessian blocks.

Lemma 11.1. *Let Assumptions 11.1 and 11.2 hold, then*

$$\sup_{T' \times B'} |\nabla_{\beta\beta} Q_n - C_{1n}| \to 0$$

and

$$\sup_{T' \times B'} |\nabla_{\beta\tau} Q_n - C_{2n}| \to 0$$

in probability as $n \to \infty$. Furthermore, $\{C_{1n} : n \in \mathbf{N}\}$ and $\{C_{2n} : n \in \mathbf{N}\}$ are equicontinuous on $T' \times B'$.

[3] Clearly these equicontinuity conditions are satisfied if $\nabla_{\beta\tau} q_t$ and $\nabla_{\beta\beta} q_t$ are equicontinuous on $Z \times \text{int}(T \times B)$.

To imply the remaining conditions of Assumption 8.1(e)-(g) as well as those of Corollary 8.2(a) we maintain furthermore Assumption 11.3 and either Assumption 11.4 or Assumption 11.5 below.

Assumption 11.3. *(a)* $E\nabla_{\beta}\underline{S}_n = 0$,
(b) $E\nabla_{\beta\tau}\underline{S}_n = 0$,
(c) $\liminf_{n\to\infty} \lambda_{\min}(E\nabla_{\beta\beta}\underline{S}_n) > 0$, *and*
(d) $\liminf_{n\to\infty} \lambda_{\min}(nE(\nabla_{\beta'}\underline{S}_n\nabla_{\beta}\underline{S}_n)) > 0$.

Note that the expectations in Assumption 11.3 exist given Assumptions 11.2 and 11.4 or 11.5 hold.

Assumption 11.4. *Let* $\bar{\tau}_n = \bar{\tau}$, $\bar{\beta}_n = \bar{\beta}$ *and let* $\nabla_{\beta}q_t(\mathbf{z}_t, \bar{\tau}, \bar{\beta})$ *be a martingale difference sequence (with respect to some filtration* $(\mathfrak{F}_t)$*) satisfying*

$$\sup_n n^{-1} \sum_{t=1}^{n} E\left|\nabla_{\beta}q_t(\mathbf{z}_t, \bar{\tau}, \bar{\beta})\right|^{2+\delta} < \infty$$

for some $\delta > 0$. *Furthermore, let* $\{\nabla_{\beta}q_t(z, \bar{\tau}, \bar{\beta}) : t \in \mathbf{N}\}$ *be equicontinuous on* Z.

Assumption 11.5. *Let* $\nabla_{\beta}q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n)$ *be near epoch dependent of size* -1 *on the basis process* $(\mathbf{e}_t)$, *which is assumed to be* α*-mixing with mixing coefficients of size* $-2r/(r-2)$ *or* ϕ*-mixing with mixing coefficients of size* $-r/(r-1)$, *for some* $r > 2$. *Furthermore, let*

$$\sup_n \sup_t E\left|\nabla_{\beta}q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n)\right|^r < \infty.$$

Sufficient conditions for Assumption 11.5 will be discussed in Section 11.3. We now have the following asymptotic normality result for the least mean distance estimator.

Theorem 11.2. *(a) Let Assumptions 11.1, 11.2, 11.3, and either 11.4 or 11.5 hold. Then*

$$n^{1/2}(\hat{\beta}_n - \bar{\beta}_n) = C_n^{-1}D_n\zeta_n + o_p(1)$$

with

$$\zeta_n \xrightarrow{D} N(0, I),$$

and

$$n^{1/2}D_n^{-1}C_n(\hat{\beta}_n - \bar{\beta}_n) \xrightarrow{D} N(0, I),$$

where C_n *and* D_n *are given by*

$$C_n = E\nabla_{\beta\beta}\underline{S}_n$$

and

$$D_n = (nE(\nabla_{\beta'}\underline{S}_n\nabla_\beta\underline{S}_n))^{1/2}.$$

Furthermore we have

$$|C_n| = O(1),\ \left|C_n^{-1}\right| = O(1),\ |D_n| = O(1) \text{ and } \left|D_n^{-1}\right| = O(1),$$

and hence $\hat{\beta}_n$ *is* $n^{1/2}$*-consistent for* $\bar{\beta}_n$*.*

(b) Let Assumptions 11.1, 11.2, 11.3(a)-(c), and either 11.4 or 11.5 hold. Assume further that $nE(\nabla_{\beta'}\underline{S}_n\nabla_\beta\underline{S}_n) \to \Lambda$ *(where* Λ *is not necessarily positive definite). Then*

$$n^{1/2}(\hat{\beta}_n - \bar{\beta}_n) = C_n^{-1}\zeta_n + o_p(1)$$

with

$$\zeta_n \xrightarrow{D} N(0,\Lambda),$$

and

$$n^{1/2}C_n(\hat{\beta}_n - \bar{\beta}_n) \xrightarrow{D} N(0,\Lambda),$$

where C_n *is as in part (a). Furthermore we have*

$$|C_n| = O(1),\ \left|C_n^{-1}\right| = O(1),$$

and hence $\hat{\beta}_n$ *is* $n^{1/2}$*-consistent for* $\bar{\beta}_n$*.*

The difference between parts (a) and (b) of Theorem 11.2 is that in part (b) the assumption that the smallest eigenvalue of $nE(\nabla_{\beta'}\underline{S}_n\nabla_\beta\underline{S}_n)$ is bounded away from zero is exchanged for the assumption that the matrix $nE(\nabla_{\beta'}\underline{S}_n\nabla_\beta\underline{S}_n)$ converges. Additional variants of the distributional convergence results in the above theorem are collected in Theorem H1 in Appendix H.

The condition $E\nabla_\beta\underline{S}_n = 0$ in Assumption 11.3 will generally be satisfied automatically under the regularity conditions used to establish consistency of $\hat{\beta}_n$. For example, if $\hat{\beta}_n$ is a minimizer of

$$Q_n(\mathbf{z}_1,\ldots,\mathbf{z}_n,\hat{\tau}_n,\beta) = S_n(\omega,\hat{\tau}_n,\beta),$$

then – under the regularity conditions set forth in Chapters 3 and 7 – $\hat{\beta}_n$ will be consistent for the minimizers $\bar{\beta}_n$ of

$$\bar{Q}_n(\bar{\tau}_n,\beta) = EQ_n(\mathbf{z}_1,\ldots,\mathbf{z}_n,\bar{\tau}_n,\beta) = ES_n(\omega,\bar{\tau}_n,\beta).$$

Hence $\nabla_\beta E\underline{S}_n = E\nabla_\beta\underline{S}_n = 0$, as $\bar{\beta}_n$ is an interior point (assuming that interchanging the order of integration and differentiation is permitted).

Clearly also under Assumption 11.4 the condition $E\nabla_\beta \underline{S}_n = 0$ in Assumption 11.3 is automatically satisfied. The condition $E\nabla_{\beta\tau}\underline{S}_n = 0$ in Assumption 11.3 ensures – as discussed in Chapter 8 – that the parameter τ is a nuisance parameter, i.e., that the distribution of $\hat{\tau}_n$ does not affect the distribution of $n^{1/2}(\hat{\beta}_n - \bar{\beta}_n)$ asymptotically.

The assumption that τ is a nuisance parameter may not be an innocuous assumption in the presence of misspecification (either of the model or of the objective function). E.g., in the case of robust estimation of location based on a symmetric loss function the scale parameter is generally not a nuisance parameter, if the true distribution is asymmetric.

Assumption 11.4 (together with Assumption 11.1(e),(f)) or Assumption 11.5, respectively, provide the essential ingredients for the asymptotic normality of the score vector $\nabla_\beta \underline{S}_n$. Assumption 11.4 postulates a martingale structure for the score vector. As discussed, this assumption seems appropriate if, e.g., Q_n is the correctly specified log-likelihood function, or the nonlinear least squares objective function based on a model which accurately captures the behavior of the conditional expectation of the dependent variables given all predetermined variables. Within this context the assumption $\bar{\tau}_n \equiv \bar{\tau}$, $\bar{\beta}_n \equiv \bar{\beta}$ seems reasonable, as $\bar{\tau}_n$, $\bar{\beta}_n$ will usually coincide with the true parameter values in a correctly specified case. Assumption 11.5 allows for misspecified situations as well as for correctly specified models with an autocorrelated score vector, as it does not require a martingale structure for $\nabla_\beta \underline{S}_n$. Under misspecification there is in general also no reason why $\bar{\tau}_n$ and $\bar{\beta}_n$ would not depend on n, except in a stationary environment.

11.2 Asymptotic Normality of Generalized Method of Moments Estimators

For generalized method of moments estimators the objective function is of the form (11.1) where ϑ_n measures the "distance" of the vector of sample moments from zero. The leading case is where $\vartheta_n(c, \tau, \beta) = c' P_n(\tau, \beta) c$ is a quadratic form. The properties of ϑ_n expressed in Assumption 11.6(a),(b) below are designed to cover the leading case, but allow also for more general specifications of ϑ_n. The condition on the derivative w.r.t. β of ϑ_n in Assumption 11.6(b) is clearly satisfied in the leading case or if ϑ_n does not depend on β, as is frequently the case. The assumption that $\nabla_c \vartheta_n(0, \tau, \beta) = 0$ is a natural one if we think of $\vartheta_n(c, \tau, \beta)$ as measuring the distance of c from zero. Clearly this assumption is satisfied in the leading case. Assumption 11.1(a)-(d) and Assumption 11.6(a) clearly imply Assumption 8.1(a)-(d) with $M_n = n^{1/2} I$ and $N_n = n^{1/2} I$. Assumption 11.6(a),(c) in conjunction with Assumption 11.1 allow us to establish uniform convergence of the Hessian blocks, which is essential for the verification of Assumption 8.1(e),(f);

cf. the discussion after Assumption 8.2.

Assumption 11.6. *(a) The functions ϑ_n are twice continuously partially differentiable on $\mathbf{R}^{p_q} \times \operatorname{int}(T \times B)$. The derivatives $\nabla_c \vartheta_n(c,\tau,\beta)$, $\nabla_{cc}\vartheta_n(c,\tau,\beta)$, $\nabla_{\beta c}\vartheta_n(c,\tau,\beta)$, $\nabla_{c\tau}\vartheta_n(c,\tau,\beta)$, $\nabla_{\beta\tau}\vartheta_n(c,\tau,\beta)$ as well as $\nabla_{\beta\beta}\vartheta_n(c,\tau,\beta)$ are equicontinuous on $\mathbf{R}^{p_q} \times \operatorname{int}(T \times B)$, and are bounded uniformly in n on every compact subset of $\mathbf{R}^{p_q} \times \operatorname{int}(T \times B)$.*

(b) $\nabla_c \vartheta_n(0,\tau,\beta) = 0$ and $\nabla_\beta \vartheta_n(0,\tau,\beta) = 0$ for all $(\tau,\beta) \in T' \times B'$.

(c) The family $\{f_t : t \in \mathbf{N}\}$ is equicontinuous on $Z \times T' \times B'$ and

$$\sup_n n^{-1} \sum_{t=1}^{n} E\left[\sup_{T' \times B'} |f_t(\mathbf{z}_t,\tau,\beta)|^{1+\gamma}\right] < \infty$$

for some $\gamma > 0$, where f_t denotes the restriction to $Z \times T' \times B'$ of any of the components of q_t, $\nabla_\beta q_t$, $\nabla_\tau q_t$, $\nabla_{\beta\tau} q_t$, or $\nabla_{\beta\beta} q_t$. (Here $T' \times B'$ is the compact subset employed in Assumption 11.1(d).)[4]

As remarked in Chapter 8 the blocks $\nabla_{\beta\beta} Q_n$ and $\nabla_{\beta\tau} Q_n$ of the Hessian matrix are of the form

$$G_{1n}(S_n, \nabla_\beta S_n, \nabla_{\beta\beta} S_n, \tau, \beta)$$

and

$$G_{2n}(S_n, \nabla_\beta S_n, \nabla_\tau S_n, \nabla_{\beta\tau} S_n, \tau, \beta),$$

respectively, where G_{1n} and G_{2n} are each defined on the Cartesian product of a Euclidean space (of appropriate dimension) with $\operatorname{int}(T \times B)$. More specifically,

$$\begin{aligned} & G_{1n}(S_n, \nabla_\beta S_n, \nabla_{\beta\beta} S_n, \tau, \beta) \\ = \; & (\nabla_c \vartheta_n \otimes I)\nabla_{\beta\beta} S_n + \nabla_{\beta'} S_n \nabla_{cc}\vartheta_n \nabla_\beta S_n \\ & + \nabla_{\beta'} S_n \nabla_{c\beta}\vartheta_n + \nabla_{\beta c}\vartheta_n \nabla_\beta S_n + \nabla_{\beta\beta}\vartheta_n, \end{aligned} \tag{11.2a}$$

$$\begin{aligned} & G_{2n}(S_n, \nabla_\beta S_n, \nabla_\tau S_n, \nabla_{\beta\tau} S_n, \tau, \beta) \\ = \; & (\nabla_c \vartheta_n \otimes I)\nabla_{\beta\tau} S_n + \nabla_{\beta'} S_n \nabla_{cc}\vartheta_n \nabla_\tau S_n \\ & + \nabla_{\beta'} S_n \nabla_{c\tau}\vartheta_n + \nabla_{\beta c}\vartheta_n \nabla_\tau S_n + \nabla_{\beta\tau}\vartheta_n, \end{aligned} \tag{11.2b}$$

where the derivatives of ϑ_n are evaluated at (S_n, τ, β).

Define on $T' \times B'$ the matrices $C_{1n}(\tau,\beta)$ and $C_{2n}(\tau,\beta)$ by

$$C_{1n}(\tau,\beta) = G_{1n}(ES_n, E\nabla_\beta S_n, E\nabla_{\beta\beta} S_n, \tau, \beta)$$

[4]Clearly these equicontinuity conditions are satisfied if q_t, $\nabla_\beta q_t$, $\nabla_\tau q_t$, $\nabla_{\beta\tau} q_t$, and $\nabla_{\beta\beta} q_t$ are equicontinuous on $Z \times \operatorname{int}(T \times B)$.

and

$$C_{2n}(\tau,\beta) = G_{2n}(ES_n, E\nabla_\beta S_n, E\nabla_\tau S_n, E\nabla_{\beta\tau} S_n, \tau, \beta).$$

The following lemma then gives the uniform convergence of the Hessian blocks.

Lemma 11.3. *Let Assumptions 11.1 and 11.6 hold. Then*

(a) $\sup_{T'\times B'} |S_n - ES_n|$, $\sup_{T'\times B'} |\nabla_\beta S_n - E\nabla_\beta S_n|$, $\sup_{T'\times B'} |\nabla_\tau S_n - E\nabla_\tau S_n|$, $\sup_{T'\times B'} |\nabla_{\beta\tau} S_n - E\nabla_{\beta\tau} S_n|$ *and* $\sup_{T'\times B'} |\nabla_{\beta\beta} S_n - E\nabla_{\beta\beta} S_n|$ *converge to zero in probability as* $n \to \infty$. *Furthermore, the restrictions of* $\{ES_n : n \in \mathbf{N}\}$, $\{E\nabla_\beta S_n : n \in \mathbf{N}\}$, $\{E\nabla_\tau S_n : n \in \mathbf{N}\}$, $\{E\nabla_{\beta\tau} S_n : n \in \mathbf{N}\}$ *and* $\{E\nabla_{\beta\beta} S_n : n \in \mathbf{N}\}$ *to* $T' \times B'$ *are equicontinuous on* $T' \times B'$.

(b) $\sup_{T'\times B'} |\nabla_{\beta\beta} Q_n - C_{1n}| \to 0$ *and* $\sup_{T'\times B'} |\nabla_{\beta\tau} Q_n - C_{2n}| \to 0$ *in probability as* $n \to \infty$. *Furthermore,* $\{C_{1n} : n \in \mathbf{N}\}$ *and* $\{C_{2n} : n \in \mathbf{N}\}$ *are equicontinuous on* $T' \times B'$.

For generalized method of moments estimators the score vector is not in the form of a Cesàro sum. Hence, to be able to apply a CLT, this vector must first be linearized. The following lemma accomplishes this task. As discussed in Chapter 7, we have $\bar{Q}_n(\tau,\beta) = \vartheta_n(ES_n(\omega,\tau,\beta),\tau,\beta)$. Note that $ES_n(\omega,\tau,\beta)$ is finite for $(\tau,\beta) \in T' \times B'$ in view of Assumption 11.6.

Lemma 11.4. *Let Assumptions 11.1 and 11.6 hold and assume that* $\underline{S}_n - E\underline{S}_n = O_p(n^{-1/2})$. *Then*

$$\begin{aligned} n^{1/2}&\left[\nabla_\beta Q_n(\mathbf{z}_1,\ldots,\mathbf{z}_n,\bar{\tau}_n,\bar{\beta}_n) - \nabla_\beta \bar{Q}_n(\bar{\tau}_n,\bar{\beta}_n)\right] \\ &= \nabla_c \vartheta_n(E\underline{S}_n,\bar{\tau}_n,\bar{\beta}_n) n^{1/2}(\nabla_\beta \underline{S}_n - E\nabla_\beta \underline{S}_n) \\ &+ n^{1/2}(\underline{S}_n - E\underline{S}_n)' \left[\nabla_{cc}\vartheta_n(E\underline{S}_n,\bar{\tau}_n,\bar{\beta}_n) E\nabla_\beta \underline{S}_n \right. \\ &\qquad \left. + \nabla_{c\beta}\vartheta_n(E\underline{S}_n,\bar{\tau}_n,\bar{\beta}_n)\right] + o_p(1). \end{aligned}$$

To imply the remaining conditions of Assumption 8.1(e)-(g) as well as those of Corollary 8.2(b) for the generalized method of moments estimator we maintain furthermore Assumption 11.7 and either Assumption 11.8 or Assumption 11.9 below.

Assumption 11.7. *(a)* $E\underline{S}_n = 0$,

(b) $E\nabla_\tau \underline{S}_n = 0$,

(c) $\liminf_{n\to\infty} \lambda_{\min}(\nabla_{cc}\vartheta_n(0,\bar{\tau}_n,\bar{\beta}_n)) > 0$,

(d) $\liminf_{n\to\infty} \lambda_{\min}(E\nabla_{\beta'} \underline{S}_n E\nabla_\beta \underline{S}_n) > 0$,

(e) $\liminf_{n\to\infty} \lambda_{\min}(nE\underline{S}_n\underline{S}_n') > 0$.

Note that the expectations in Assumption 11.7 exist given Assumptions 11.6 and 11.8 or 11.9 hold.

Assumption 11.8. *Let $\bar{\tau}_n = \bar{\tau}$, $\bar{\beta}_n = \bar{\beta}$ and let $q_t(\mathbf{z}_t, \bar{\tau}, \bar{\beta})$ be a martingale difference sequence (with respect to some filtration $(\mathfrak{F}_t)$) satisfying*

$$\sup_n n^{-1} \sum_{t=1}^{n} E \left| q_t(\mathbf{z}_t, \bar{\tau}, \bar{\beta}) \right|^{2+\delta} < \infty$$

for some $\delta > 0$.

Assumption 11.9. *Let $q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n)$ be near epoch dependent of size -1 on the basis process $(\mathbf{e}_t)$, which is assumed to be α-mixing with mixing coefficients of size $-2r/(r-2)$ or ϕ-mixing with mixing coefficients of size $-r/(r-1)$, for some $r > 2$. Furthermore, let*

$$\sup_n \sup_t E \left| q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n) \right|^{r} < \infty.$$

Sufficient conditions for Assumption 11.9 will be discussed in Section 11.3. We now have the following asymptotic normality result for the generalized method of moments estimator.

Theorem 11.5. *(a) Let Assumptions 11.1, 11.6, 11.7, and either 11.8 or 11.9 hold. Then*

$$n^{1/2}(\hat{\beta}_n - \bar{\beta}_n) = C_n^{-1} D_n \zeta_n + o_p(1)$$

with

$$\zeta_n \xrightarrow{D} N(0, I),$$

and

$$(C_n^{-1} D_n D_n' C_n^{-1\prime})^{-1/2} n^{1/2}(\hat{\beta}_n - \bar{\beta}_n) \xrightarrow{D} N(0, I),$$

where C_n and D_n are given by

$$C_n = E\nabla_{\beta'} \underline{S}_n \left[\nabla_{cc} \vartheta_n(0, \bar{\tau}_n, \bar{\beta}_n) \right] E\nabla_{\beta} \underline{S}_n$$

and

$$D_n = E\nabla_{\beta'} \underline{S}_n \left[\nabla_{cc} \vartheta_n(0, \bar{\tau}_n, \bar{\beta}_n) \right] (nE\underline{S}_n \underline{S}_n')^{1/2}.$$

Furthermore we have

$$|C_n| = O(1), \quad \left|C_n^{-1}\right| = O(1), \quad |D_n| = O(1) \text{ and } \left|(D_n D_n')^{-1}\right| = O(1),$$

and hence $\hat{\beta}_n$ is $n^{1/2}$-consistent for $\bar{\beta}_n$.

(b) Let Assumptions 11.1, 11.6, 11.7(a)-(d), and either 11.8 or 11.9 hold. Assume further that

$$nE\underline{S}_n\underline{S}_n' \to \Lambda$$

(where Λ is not necessarily positive definite). Then

$$n^{1/2}(\hat{\beta}_n - \bar{\beta}_n) = C_n^{-1}D_n\zeta_n + o_p(1)$$

and

$$n^{1/2}C_n(\hat{\beta}_n - \bar{\beta}_n) = D_n\zeta_n + o_p(1)$$

with

$$\zeta_n \xrightarrow{D} N(0,\Lambda),$$

where C_n is as in part (a) and

$$D_n = E\nabla_{\beta'}\underline{S}_n\left[\nabla_{cc}\vartheta_n(0,\bar{\tau}_n,\bar{\beta}_n)\right].$$

Furthermore we have

$$|C_n| = O(1),\ |C_n^{-1}| = O(1),\ |D_n| = O(1) \text{ and } |(D_nD_n')^{-1}| = O(1),$$

and hence $\hat{\beta}_n$ is $n^{1/2}$-consistent for $\bar{\beta}_n$.

The difference between parts (a) and (b) of Theorem 11.5 is that in part (b) the assumption that the smallest eigenvalue of $nE\underline{S}_n\underline{S}_n'$ is bounded away from zero is exchanged for the assumption that $nE\underline{S}_n\underline{S}_n'$ converges to some matrix. Additional variants of the distributional convergence results in the above theorem are collected in Theorem H2 in Appendix H.

Theorem 11.5 shows that the asymptotic distribution of $\hat{\beta}_n$ depends on ϑ_n only through $\nabla_{cc}\vartheta_n(0,\bar{\tau}_n,\bar{\beta}_n)$. Hence, from the point of view of asymptotic efficiency, there is no loss in considering only generalized method of moments estimators based on a quadratic "distance" function $c'P_nc$ with $P_n = \nabla_{cc}\vartheta_n(0,\bar{\tau}_n,\bar{\beta}_n)$. This was pointed out by Newey (1988) in a special case. Furthermore, if also $\dim(S_n) = \dim(\beta)$ holds, then $C_n^{-1}D_n$ reduces in parts (a) and (b) of Theorem 11.5 to, respectively, $(E\nabla_\beta\underline{S}_n)^{-1}(nE\underline{S}_n\underline{S}_n')^{1/2}$ and $(E\nabla_\beta\underline{S}_n)^{-1}$, and hence the asymptotic distributional result for $\hat{\beta}_n$ does not depend on the particular form of ϑ_n at all. The intuitive explanation for this simplification is that in case $\dim(S_n) = \dim(\beta)$ we will typically be able to set the sample moments simultaneously to zero and hence the choice of the "distance" function ϑ_n will have no effect.

The conditions in Assumption 11.7(a),(b) that $E\underline{S}_n = 0$ and $E\nabla_\tau\underline{S}_n = 0$ are crucial for two reasons.[5] First, if they are violated τ will generally not be a nuisance parameter (even in the simple case $\vartheta_n(c,\tau,\beta) = c'Pc$ where P is

[5]Clearly under Assumption 11.8 the condition $E\underline{S}_n = 0$ in Assumption 11.7(a) is automatically satisfied.

a matrix that does not depend on (τ, β), but where, of course, S_n depends on (τ, β)). This follows as then $C_{2n}(\bar{\tau}_n, \bar{\beta}_n)$ is in general not $o(1)$. Hence, if the parameter τ is present and overidentifying moment restrictions are used in an estimation problem, i.e., if $\dim(S_n) > \dim(\beta)$, then misspecification is generally a serious problem for generalized method of moments estimators as in this case we typically cannot expect $E\underline{S}_n = 0$. Consequently, the asymptotic distributional results of Theorem 11.5 will not hold and, in particular, the asymptotic distribution of $\hat{\beta}_n$ will then in general depend on the distribution of $\hat{\tau}_n$.

Second, assume for the discussion in this paragraph that the parameter τ is not present or, more generally, that we can establish $C_{2n}(\bar{\tau}_n, \bar{\beta}_n) = o(1)$ – i.e., that τ is a nuisance parameter – from some source other than $E\underline{S}_n = 0$ and $E\nabla_\tau \underline{S}_n = 0$. Then the assumption $E\underline{S}_n = 0$ is still crucial in order to obtain the results given in Theorem 11.5. This is so since $E\underline{S}_n = 0$ is also used in the proof to show that the normalized score vector can be expressed asymptotically as a linear function of $\underline{S}_n$, to which a CLT is then applied. If $E\underline{S}_n \neq 0$ then Theorem 11.5 is, of course, no longer valid, but we can still obtain an alternative asymptotic normality result for $\hat{\beta}_n$. In the following we outline assumptions under which a derivation of such an alternative result is possible. As is evident from Lemma 11.4, given Assumptions 11.1 and 11.6, we can again represent the normalized score vector $n^{1/2}\nabla_\beta Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \bar{\tau}_n, \bar{\beta}_n)$ up to an error of magnitude $o_p(1)$ as a linear function of $(\nabla_\beta \underline{S}_n - E\nabla_\beta \underline{S}_n)$ and $(\underline{S}_n - E\underline{S}_n)$. Asymptotic normality of the normalized score then follows if the vector made up from the components of $(\nabla_\beta \underline{S}_n - E\nabla_\beta \underline{S}_n)$ and $(\underline{S}_n - E\underline{S}_n)$ satisfies a CLT, and if $n^{1/2}\nabla_\beta \bar{Q}_n(\bar{\tau}_n, \bar{\beta}_n) = o(1)$.[6] The latter condition holds automatically given $\bar{\beta}_n$ minimizes $\bar{Q}_n(\bar{\tau}_n, \beta)$, which will typically be the case. The CLT can again be deduced from lower level conditions, e.g., via Theorem 10.2. Under conditions, similar in spirit to Assumption 11.7(c)-(e), the relevant matrices C_n and D_n will satisfy the remaining requirements of Assumption 8.1 and we obtain a result similar to Theorem 11.5. However, it is important to note that the formulas for C_n and D_n are now different from those given in Theorem 11.5. For an alternative catalogue of assumptions for the misspecified case (and where the parameter τ is not present) see Gallant and White (1988, Ch.4).

As discussed above, if misspecification is present in the context of generalized method of moments estimation (with overidentifying moment restrictions), then τ will typically not be a nuisance parameter. Therefore, if misspecification is suspected, it may be prudent to attempt to somehow treat τ together with β as parameters of interest, i.e., to try to "eliminate the nuisance parameters by converting them into parameters of interest". We shall discuss such a strategy next. If τ is not a nuisance parameter, in-

[6]We note that Assumption 11.6(b) is not needed in this context.

formation about the asymptotic distribution of $\hat{\beta}_n$ can typically be obtained only if we are able to get a handle on the joint asymptotic distribution of the normalized score vector $n^{1/2}\nabla_\beta Q_n$ and of $n^{1/2}(\hat{\tau}_n - \bar{\tau}_n)$, cf. Chapter 8. One approach to achieve this is to (i) augment the first order conditions $\nabla_\beta Q_n = 0$ by the equations that define the estimator for τ, (ii) define $\hat{\beta}_n$ and $\hat{\tau}_n$ as the solution of the augmented system of equations, and (iii) derive the joint asymptotic distribution of $\hat{\beta}_n$ and $\hat{\tau}_n$ from the augmented system of equations defining $\hat{\beta}_n$ and $\hat{\tau}_n$.

We now return with our discussion to the case where $E\underline{S}_n = 0$. Recall that for the asymptotic normality result given in Theorem 11.5 we have maintained $n^{1/2}$-consistency of $\hat{\tau}_n$. We note that in the important special case, where the "distance" function ϑ_n – but not the vector of sample moments S_n – depends on τ, this assumption can be dropped, as was pointed out by Newey (1988). The proof proceeds by first linearizing the normalized score vector evaluated at $(\hat{\tau}_n, \hat{\beta}_n)$ with respect to $\hat{\beta}_n$, which yields an equation of the form (9.5). As discussed in Chapter 9 this leaves one with the task of establishing asymptotic normality of the normalized score vector evaluated at $\hat{\tau}_n$ and $\bar{\beta}_n$. Further linearization of this score vector with respect to the vector of sample moments makes it possible to express it in the form of a matrix, which depends on $\hat{\tau}_n$, times the normalized vector of sample moments $n^{1/2}\underline{S}_n$. Note that $\underline{S}_n$ does not depend on $\hat{\tau}_n$. This makes it possible to establish asymptotic normality of $\hat{\beta}_n$ from a CLT for $\underline{S}_n$ without the assumption of $n^{1/2}$-consistency of $\hat{\tau}_n$, since the asymptotic behavior of the matrix premultiplying $n^{1/2}\underline{S}_n$ can now be deduced from uniform convergence results and consistency of $\hat{\tau}_n$ alone.[7]

11.3 Further Discussion and Comparison of Assumptions

(i) In the following we provide a sufficient condition for Assumptions 11.5 or 11.9 with $f_t = \nabla_\beta q_t$ or $f_t = q_t$, respectively, in terms of near epoch dependence of the underlying data process $(\mathbf{z}_t)$.

Assumption 11.10. *(a) Assume that $Z = \mathbf{R}^{p_z}$ and that for some $s^\bullet > 2$ the process $(\mathbf{z}_t)$ is near epoch dependent of size $-2(s^\bullet - 1)/(s^\bullet - 2)$ on the basis process $(\mathbf{e}_t)$, which is assumed to be α-mixing with mixing coefficients of size $-2r/(r-2)$ or ϕ-mixing with mixing coefficients of size $-r/(r-1)$, for some $r > 2$.*

[7] Since we have not made use of differentiability of the objective function w.r.t. τ in the above argument, this assumption can be dropped. In fact, τ could here be an infinite dimensional parameter.

(b) $|f_t(z,\bar{\tau}_n,\bar{\beta}_n) - f_t(z^\bullet,\bar{\tau}_n,\bar{\beta}_n)| \le B_{tn}(z,z^\bullet)|z - z^\bullet|$ *for all* $(z, z^\bullet)$, *where* $B_{tn} : Z \times Z \to [0,\infty)$ *is Borel measurable and satisfies*

$$\sup_t \sup_n \sup_m \|B_{tn}(\mathbf{z}_t, \mathbf{h}_t^m)\|_2 < \infty$$

and

$$\sup_t \sup_n \sup_m \|B_{tn}(\mathbf{z}_t, \mathbf{h}_t^m)\, |\mathbf{z}_t - \mathbf{h}_t^m|\, \|_{s^\bullet} < \infty$$

with $\mathbf{h}_t^m = E(\mathbf{z}_t|\mathbf{e}_{t+m},\ldots,\mathbf{e}_{t-m})$.

(c) $\sup_n \sup_t E\left|f_t(\mathbf{z}_t,\bar{\tau}_n,\bar{\beta}_n)\right|^r < \infty$.

Given Assumption 11.1, Assumption 11.10 implies Assumptions 11.5 and 11.9, respectively. This follows from an extension of Theorem 6.7(a) in Chapter 6 to arrays or from Theorem 4.2 in Gallant and White (1988). In the more general case of a Borel set $Z \subseteq \mathbf{R}^{p_z}$ Assumption 11.10 has to be modified in that $\mathbf{h}_t^m$ has to be chosen to fall into Z. This is always possible if $(\mathbf{z}_t)$ is near epoch dependent as shown in Lemma D3 in Appendix D and the remark following that lemma. However, this may complicate the analysis.

(ii) Bierens (1981, 1984) established asymptotic normality for the nonlinear least squares estimator and for robust M-estimators in a dynamic nonlinear regression model under the assumption that the data generating process is stochastically stable w.r.t. an i.i.d. or ϕ-mixing basis process. (As discussed in Chapter 6 stochastic stability is up to inessential details the same as L_0-approximability.) In both papers Bierens maintains an asymptotic stationarity condition.[8] His results are furthermore based on the assumption that the error process forms a martingale difference sequence. Theorem 11.2 under Assumptions 11.1 - 11.4 hence extends Bierens' results by covering a wider class of objective functions and by allowing for more heterogeneity in the data processes, in that Bierens' asymptotic stationarity condition is avoided. The asymptotic normality results in Gallant (1987a, Ch.7) and Gallant and White (1988, Ch.5) cover both least mean distance and generalized method of moments estimators and also avoid Bierens' asymptotic stationarity condition. Gallant and White (1988) do not consider nuisance parameters in their analysis. Gallant's (1987a) and Gallant and White's (1988) asymptotic normality results differ from Theorems 11.2 and 11.5 in that their entire theory is built on the concept of near epoch dependence. They also do not isolate results for cases where a martingale structure is available. Furthermore, in their basic catalogues of assumptions Gallant (1987a) and Gallant and White (1988) directly assume near epoch dependence of all functions of the data generating process

[8] For the case of strictly stationary processes asymptotic normality of generalized method of moments estimators has been established in Hansen (1982); cf. also Newey (1985a).

that are required to satisfy a LLN or CLT in the proof of their asymptotic normality result; no explicit catalogue of assumptions for asymptotic normality is given which is directly based on near epoch dependence of the underlying data generating process. Of course, Proposition 1 in Gallant (1987a, Ch.7), Theorem 4.2 in Gallant and White (1988), or Theorem 6.7(a) in Chapter 6 could be used to provide such sufficient conditions. This would then result in a catalogue of assumptions which would be on a comparable level with the assumptions in Sections 11.1, 11.2 and in remark (i) above. However, the resulting catalogue of assumptions would be rather complex, since the approximators $\mathbf{h}_t^m$ then appear explicitly in these assumptions, cf., e.g., Theorem 6.15 and the discussion in Chapter 7. Hence, an advantage of the approach in Sections 11.1 and 11.2 seems to be that it leads to a more manageable catalogue of sufficient conditions for asymptotic normality. This becomes especially transparent for the case where a martingale difference structure is available, since then the catalogue of assumptions only uses the weaker concept of L_0-approximability of $(\mathbf{z}_t)$ rather than near epoch dependence of $(\mathbf{z}_t)$; cf. Assumptions 11.1 - 11.4 and Assumptions 11.1, 11.6 - 11.7 for least mean distance estimators and generalized method of moments estimators, respectively. Thus, in the case where a martingale difference structure is available, assumptions concerning the size of the approximation error appearing in the definition of near epoch dependence can be avoided altogether. Furthermore, in contrast to Gallant (1987a) and Gallant and White (1988), the approach in Sections 11.1 and 11.2 requires near epoch dependence assumptions only on $\nabla_\beta q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n)$ for least mean distance estimators or on $q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n)$ for generalized method of moments estimators in the case when no martingale difference structure is available.

(iii) The asymptotic normality results for least mean distance and generalized method of moments estimators in Gallant (1987a, Ch.7) and Gallant and White (1988, Ch.5) differ from Theorem 11.2 and 11.5 also in that Andrews' (1987) ULLN, i.e., Theorem 5.1 in Chapter 5, is used to establish uniform convergence of the Hessian blocks. In the preceding Sections 11.1 and 11.2 this uniform convergence of the Hessian blocks was established using Pötscher and Prucha's (1989) ULLN given in Theorem 5.2 in Chapter 5 (where the local LLNs are implied via Theorem 6.13 from the assumption that $(\mathbf{z}_t)$ is L_0-approximable). Clearly, we could have also used Andrews' (1987) ULLN, i.e., Theorem 5.1, in place of Theorem 5.2 to derive this uniform convergence. Basically we would then have to trade the equicontinuity assumption on f_t w.r.t. (z, τ, β) in Assumptions 11.2 and 11.6 for a Lipschitz-type assumption on f_t w.r.t. (τ, β). Such Lipschitz-type conditions will frequently be implied from bounds on the derivative. Since f_t involves already second order derivatives of q_t, we would then typically have to assume the existence of third order derivatives of q_t w.r.t. (τ, β) as compared to only second order derivatives needed by the approach based on the ULLN given in Theorem 5.2. Although Andrews' ULLN in its generic form, i.e., Theorem 5.1, does not impose any "smoothness" conditions on f_t w.r.t. z, a smoothness condition w.r.t. z is – as discussed in Section 6.6 – typically needed to imply the existence of local LLNs (assumed by that theorem) from L_0-approximability or near epoch dependence of $(\mathbf{z}_t)$. Hence, comparing the total smoothness conditions needed to establish asymptotic normality of M-estimators for dynamic nonlinear econometric models, it does not seem that using Theorem 5.1 in place of Theorem 5.2 would give a better catalogue of assumptions.

(iv) In case of generalized method of moments estimation based on a quadratic "distance" function asymptotic normality results can be obtained without requiring q_t to be twice but only once continuously differentiable. This was pointed out by Andrews (1991c) and is possible, since in case of a quadratic "distance" function the score vector takes the form

$$\nabla_{\beta'} Q_n(\hat{\tau}_n, \hat{\beta}_n) = \nabla_{\beta'} S_n(\omega, \hat{\tau}_n, \hat{\beta}_n) P_n(\hat{\tau}_n, \hat{\beta}_n) S_n(\omega, \hat{\tau}_n, \hat{\beta}_n).$$

Instead of expanding the score vector around $(\bar{\tau}_n, \bar{\beta}_n)$ one only expands $S_n(\omega, \hat{\tau}_n, \hat{\beta}_n)$. This again leads to a representation of $n^{1/2}(\hat{\beta}_n - \bar{\beta}_n)$ as a linear function of $n^{1/2}\underline{S}_n = n^{1/2} S_n(\omega, \bar{\tau}_n, \bar{\beta}_n)$. A CLT for $\underline{S}_n$ then delivers asymptotic normality of $\hat{\beta}_n$.[9]

(v) Any estimator solving the first order conditions corresponding to the objective function of a least mean distance estimator could be artificially recast as a generalized method of moments estimator by converting the problem of solving $\nabla_\beta Q_n = 0$ into the problem of minimizing, e.g., $\nabla_\beta Q_n \nabla_{\beta'} Q_n$. This approach allows us in principle to subsume least mean distance estimators within the class of quadratic generalized method of moments estimators, cf. Andrews (1989a,c). Proceeding as in remark (iv) above, one would then obtain results comparable to those, e.g., in Theorem 11.2. We could also attempt to subsume least mean distance estimators as special cases of the larger class of generalized method of moments estimators with "distance" functions that are not necessarily quadratic. We could then attempt to apply one of the asymptotic normality results for this larger class of estimators as, e.g., Theorem 11.5. However, a disadvantage of this approach is that then higher differentiability conditions have to be assumed as compared to the case where the estimator is treated directly as a least mean distance estimator as is, e.g., the case in Theorem 11.2 (or as a *quadratic* generalized method of moments estimator as is discussed in remark (iv) above). It seems that to derive asymptotic normality results for least mean distance estimators under minimal conditions we need to treat them separately from the *general* class of generalized method of moments estimators with "distance" functions that are not necessarily quadratic.

(vi) One of the assumptions for asymptotic normality of $\hat{\beta}_n$ maintained

[9]The discussion in this remark assumes that τ is a nuisance parameter and that $E\underline{S}_n = 0$.

here and in the literature on dynamic nonlinear econometric models, cf. Gallant (1987a) and Gallant and White (1988), namely that there exist non-random sequences $\bar{\tau}_n$ and $\bar{\beta}_n$ such that $\hat{\tau}_n - \bar{\tau}_n \to 0$ and $\hat{\beta}_n - \bar{\beta}_n \to 0$ holds, is restrictive in misspecified situations. To appreciate the implications of this assumption consider, for simplicity only, the case of a least mean distance estimator in a stationary context with no nuisance parameter present: $\bar{\beta}_n \equiv \bar{\beta}$ will then typically be a minimizer of $\bar{Q}(\beta) \equiv \bar{Q}_n(\beta) = EQ_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \beta)$ and will not depend on n. Convergence of $\hat{\beta}_n$ to $\bar{\beta}$ can usually only be expected if $\bar{\beta}$ is identifiably unique, which implies in particular that $\bar{\beta}$ is the unique minimizer of $\bar{Q}$. However, this uniqueness is, as discussed in Section 4.6, not natural under misspecification. Now, if uniqueness fails to hold, then $\hat{\beta}_n$ will typically "converge" to the set of "pseudo true" values which minimize $\bar{Q}$ rather than to a single point; cf. Section 4.6. Asymptotic normality of $\hat{\beta}_n$ will then typically fail. In fact, in such a case the distribution of $\hat{\beta}_n$ itself (without renormalization by $n^{1/2}$) may have a non-degenerate limiting distribution. The form of this limiting distribution will, however, depend heavily on the geometric structure of the set of "pseudo true" values.[10] The problem is to some extent similar to the problem of finding the limiting distribution of estimators of non-identified parameters in a well-specified context, see, e.g., Phillips (1989).

---

[10]This set could be finite or infinite. See Pötscher (1991), Example 3, for a situation where this set is infinite.

12

HETEROSKEDASTICITY AND AUTOCORRELATION ROBUST ESTIMATION OF VARIANCE COVARIANCE MATRICES

12.1 An Outline of the Variance Covariance Matrix Estimation Problem

Inspection of the asymptotic normality results for least mean distance and generalized method of moments estimators given in, e.g., Theorems 11.2(a) and 11.5(a) shows that in both cases a matrix of the form $C_n^{-1} D_n D_n' \, C_n^{-1\prime}$ acts as an asymptotic variance covariance matrix of $n^{1/2}(\hat{\beta}_n - \bar{\beta}_n)$, where C_n and D_n are given in those theorems. For purposes of inference we need estimators of C_n and D_n. Inspection of the matrices C_n reveals that these matrices are essentially composed of terms of the form $n^{-1}\sum_{t=1}^{n} E\mathbf{w}_{t,n}$, where $\mathbf{w}_{t,n} = \mathbf{w}_t(\bar{\tau}_n, \bar{\beta}_n)$ equals $\nabla_{\beta\beta} q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n)$ in the case of least mean distance estimators or equals $\nabla_{\beta} q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n)$ in the case of generalized method of moments estimators. The matrices D_n – apart from containing similar terms – also contain an expression of the form

$$n^{-1} E\left[\left(\sum_{t=1}^{n} \mathbf{v}_{t,n}\right)\left(\sum_{t=1}^{n} \mathbf{v}_{t,n}'\right)\right],$$

where $\mathbf{v}_{t,n} = \mathbf{v}_t(\bar{\tau}_n, \bar{\beta}_n)$ equals $\nabla_{\beta'} q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n)$ in the case of least mean distance estimators or equals $q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n)$ in the case of generalized method of moments estimators. The expressions of the form $n^{-1}\sum_{t=1}^{n} E\mathbf{w}_{t,n}$ will typically be estimated by $n^{-1}\sum_{t=1}^{n} \hat{\mathbf{w}}_{t,n}$ where $\hat{\mathbf{w}}_{t,n} = \mathbf{w}_t(\hat{\tau}_n, \hat{\beta}_n)$. Consistency of such estimators can be derived from ULLNs and from consistency of $(\hat{\tau}_n, \hat{\beta}_n)$ in a rather straightforward manner via Lemma 3.2.[1] The esti-

[1]We note that for most of the terms of the form $n^{-1}\Sigma_{t=1}^{n} \hat{\mathbf{w}}_{t,n}$ the relevant uniform convergence results used to establish their convergence are typically already established in the course of the asymptotic normality proof, cf., e.g., Lemmata 11.1 and 11.3.

mation of

$$\Psi_n = n^{-1}E\left[\left(\sum_{t=1}^{n}\mathbf{v}_{t,n}\right)\left(\sum_{t=1}^{n}\mathbf{v}'_{t,n}\right)\right]$$

reduces to a similar problem in the important special case where $(\mathbf{v}_{t,n})$ is a martingale difference array (or, more generally is uncorrelated and has mean zero), since then the expression for Ψ_n reduces to $n^{-1}\sum_{t=1}^{n} E\mathbf{v}_{t,n}\mathbf{v}'_{t,n}$. As discussed above $(\mathbf{v}_{t,n})$ will have a martingale difference structure in certain correctly specified cases. In the general case, however, $(\mathbf{v}_{t,n})$ will typically be autocorrelated (with autocorrelation of unknown form) and hence the estimation of Ψ_n is more involved.

Observe that Ψ_n can be written equivalently as

$$\Psi_n = n^{-1}\sum_{t=1}^{n} E\mathbf{v}_{t,n}\mathbf{v}'_{t,n} + \sum_{j=1}^{n-1} n^{-1}\sum_{t=1}^{n-j}\left[E\mathbf{v}_{t,n}\mathbf{v}'_{t+j,n} + E\mathbf{v}_{t+j,n}\mathbf{v}'_{t,n}\right]. \tag{12.1}$$

To highlight the complications arising in the estimation of Ψ_n in case $(\mathbf{v}_{t,n})$ is autocorrelated assume for the moment that $(\mathbf{v}_{t,n}) = (\mathbf{v}_t)$ is a zero mean stationary process (with absolutely summable covariance function). Then the estimand

$$\Psi_n = n^{-1}E\left[\left(\sum_{t=1}^{n}\mathbf{v}_t\right)\left(\sum_{t=1}^{n}\mathbf{v}'_t\right)\right]$$

converges to

$$E\mathbf{v}_t\mathbf{v}'_t + \sum_{j=1}^{\infty}\left[E\mathbf{v}_t\mathbf{v}'_{t+j} + E\mathbf{v}_{t+j}\mathbf{v}'_t\right],$$

which equals 2π times the value of the spectral density of $(\mathbf{v}_t)$ at frequency zero. Thus estimation of this expression essentially involves the estimation of an infinite number of covariances of the process $(\mathbf{v}_t)$. For the case where $(\mathbf{v}_t)$ is stationary and observable this problem of estimating the spectral density (at frequency zero) has been studied extensively in the time series literature, cf., e.g., Anderson (1971, Ch.9), Priestley (1981, Ch.6,7). The "naive" estimator

$$n^{-1}\sum_{t=1}^{n}\mathbf{v}_t\mathbf{v}'_t + \sum_{j=1}^{n-1}(n-j)^{-1}\sum_{t=1}^{n-j}\left[\mathbf{v}_t\mathbf{v}'_{t+j} + \mathbf{v}_{t+j}\mathbf{v}'_t\right]$$

where each $E\mathbf{v}_t\mathbf{v}'_{t+j}$ is estimated unbiasedly by $(n-j)^{-1}\sum_{t=1}^{n-j}\mathbf{v}_t\mathbf{v}'_{t+j}$ is well-known to be inconsistent. (For simplicity assume for the moment that $\mathbf{v}_{t,n} = \mathbf{v}_t$ is observable; in general, as shown below, replacing $\mathbf{v}_{t,n}$ with the observable $\hat{\mathbf{v}}_{t,n}$ will not affect the consistency result under mild regularity conditions.) Intuitively speaking, the reason for the inconsistency is that the estimator is a sum of n terms, each with a variance roughly of the order

$O((n-j)^{-1})$. The variance of the estimator is now roughly n times as large, i.e., of the order O(1). (Of course, we have here ignored the correlation between the individual terms contributing to the sum.) It should also be noted that the estimators $(n-j)^{-1}\sum_{t=1}^{n-j} \mathbf{v}_t\mathbf{v}_{t+j}'$ have a large variance for j close to n.

One of the standard approaches taken in the time series literature to obtain consistent estimators is to reduce the variance of the estimator by excluding some of the sample moments $(n-j)^{-1}\sum_{t=1}^{n-j} \mathbf{v}_t\mathbf{v}_{t+j}'$ from the formula for the "naive" estimator (or more generally by "down-weighing" those sample moments). It seems natural to exclude or down-weigh the sample moments corresponding to lags j close to n. This is achieved by introducing weights into the formula for the "naive" estimator. Down-weighing of the sample covariances has the effect of reducing the variance of the estimator at the expense of introducing a bias. The search for consistent estimators hence amounts to finding weighting schemes such that asymptotically both the variance and the bias vanish. We note that the alternative "naive" estimator

$$n^{-1}\sum_{t=1}^{n} \mathbf{v}_t\mathbf{v}_t' + \sum_{j=1}^{n-1} n^{-1}\sum_{t=1}^{n-j}\left[\mathbf{v}_t\mathbf{v}_{t+j}' + \mathbf{v}_{t+j}\mathbf{v}_t'\right],$$

which equals 2π times the periodogram (calculated from $\mathbf{v}_1,\ldots,\mathbf{v}_n$) evaluated at frequency zero, is still inconsistent, although some moderate down-weighing of the sample moments $(n-j)^{-1}\sum_{t=1}^{n-j} \mathbf{v}_t\mathbf{v}_{t+j}'$ takes place already.

Motivated by the above discussion, and in keeping with a long tradition in spectral density estimation, we consider estimators for Ψ_n of the form

$$\begin{aligned}\hat{\Psi}_n &= w(0,n)\left[n^{-1}\sum_{t=1}^{n} \hat{\mathbf{v}}_{t,n}\hat{\mathbf{v}}_{t,n}'\right] \\ &+ \sum_{j=1}^{n-1} w(j,n)\left[n^{-1}\sum_{t=1}^{n-j}\left[\hat{\mathbf{v}}_{t,n}\hat{\mathbf{v}}_{t+j,n}' + \hat{\mathbf{v}}_{t+j,n}\hat{\mathbf{v}}_{t,n}'\right]\right],\end{aligned} \tag{12.2}$$

where $\hat{\mathbf{v}}_{t,n} = \mathbf{v}_t(\hat{\tau}_n,\hat{\beta}_n)$ and the weights $w(j,n)$ are real numbers.[2] Sufficient conditions for consistency of this class of estimators will be discussed in the next section.

The above discussion suggests furthermore that it would be of interest to consider alternative approaches for the estimation of Ψ_n based on alternative techniques used in the time series literature for the estimation of a spec-

[2]In (12.2) we have chosen to normalize the sample moments by n rather than by $n-j$ as is usual in the literature. By redefining the weights as $w(j,n)(n-j)/n$ we could have equivalently expressed $\hat{\Psi}_n$ in terms of sample moments normalized by $n-j$.

tral density (at frequency zero).[3] For example, an alternative approach is autoregressive spectral density estimation, see, e.g., Parzen (1983), Priestley (1981, Ch.7,8). In this approach the spectral density of the time series is estimated by the spectral density calculated from an autoregressive model fitted to this series. For the resulting estimator to be consistent the order of the fitted autoregressive model must be allowed to increase with the sample size (at an appropriate rate). For asymptotic properties of this estimator in the context of stationary processes see Berk (1974), Hannan and Kavalieris (1986) and Hannan and Deistler (1988, Ch.7), cf. also Parzen (1983). It hence is of interest to systematically explore the merits of such an estimation procedure within the present context where stationarity of $(\mathbf{v}_{t,n})$ is not maintained. See Den Haan and Levin (1996) for a recent contribution.

The above outlined difficulties in the estimation of Ψ_n due to the fact that essentially an infinite number of "parameters" has to be estimated do not arise if $E\mathbf{v}_{t,n}\mathbf{v}'_{t+j,n} = 0$ for $j > m$ and all n and if m is known (e.g., if $(\mathbf{v}_{t,n})$ has zero mean and is m-dependent with m known); cf., e.g., Hansen (1982) and Hansen and Singleton (1982). In this case Ψ_n is given by

$$n^{-1}\sum_{t=1}^{n} E\mathbf{v}_{t,n}\mathbf{v}'_{t,n} + \sum_{j=1}^{m} n^{-1}\sum_{t=1}^{n-j}\left[E\mathbf{v}_{t,n}\mathbf{v}'_{t+j,n} + E\mathbf{v}_{t+j,n}\mathbf{v}'_{t,n}\right]$$

which can typically be estimated consistently by

$$n^{-1}\sum_{t=1}^{n} \hat{\mathbf{v}}_{t,n}\hat{\mathbf{v}}'_{t,n} + \sum_{j=1}^{m} n^{-1}\sum_{t=1}^{n-j}\left[\hat{\mathbf{v}}_{t,n}\hat{\mathbf{v}}'_{t+j,n} + \hat{\mathbf{v}}_{t+j,n}\hat{\mathbf{v}}'_{t,n}\right].$$

More generally, if a parametric model for the autocorrelation structure of $(\mathbf{v}_{t,n})$ is known to the researcher, this model can be used to obtain a parametric estimator for Ψ_n, which will typically be superior to $\hat{\Psi}_n$. This situation occurs, for example, in the case of ordinary least squares estimation of a regression model with disturbances that follow an autoregressive process of order m.

12.2 Sufficient Conditions for Consistency

In the following we present results concerning the consistency of the estimators $\hat{\Psi}_n$. We first give a lemma that is of a generic nature and provides basic conditions for consistency of $\hat{\Psi}_n$ for a wide variety of dependence

[3]Note that $\Gamma_n(j)$ with $\Gamma_n(j) = n^{-1}\Sigma_{t=1}^{n-j} E\mathbf{v}_{t,n}\mathbf{v}'_{t+j,n}$ for $0 \leq j \leq n-1$, $\Gamma_n(j) = 0$ for $j \geq n$, and $\Gamma_n(-j) = \Gamma_n(j)'$, has formally all the properties of a covariance function. Hence, for fixed sample size n, the problem of estimating Ψ_n is in fact formally identical to the estimation of a spectral density $\phi_n(\omega) = (2\pi)^{-1}\Sigma_{s=-\infty}^{\infty}\Gamma_n(s)\exp(-i\omega s)$ at frequency zero from estimators for $\Gamma_n(j)$.

structures of the underlying stochastic process. We then give more specific results under the assumption of near epoch dependence. Those results summarize and extend consistency results given in Newey and West (1987), Gallant (1987a) and Gallant and White (1988). A further feature is that we also provide rates of convergence for $\hat{\Psi}_n - \Psi_n$. Closely related results are also given in Andrews (1991b). We note that all these results on consistency of $\hat{\Psi}_n$ are intimately related to consistency results for spectral density estimators in the time series literature.

We need the following assumption, which postulates smoothness conditions for $\mathbf{v}_t(\tau, \beta)$ and $n^{1/2}$-consistency of $(\hat{\tau}_n, \hat{\beta}_n)$.[4] This assumption also covers the case where no nuisance parameter τ is present, cf. the similar discussion in Chapters 8 and 11.

Assumption 12.1. *Let T and B be measurable subsets of Euclidean space $\mathbf{R}^{p_\tau}$ and $\mathbf{R}^{p_\beta}$, respectively, and let $\mathbf{v}_t(\tau, \beta)$ be random vectors taking their values in $\mathbf{R}^{p_v}$ for each $(\tau, \beta) \in T \times B$ and $t \in \mathbf{N}$. Assume that $\mathbf{v}_t(.,.)$ is a.s. continuously partially differentiable at every point (τ, β) in the interior of $T \times B$ (and where the exceptional null set does not depend on (τ, β)). Let $(\hat{\tau}_n, \hat{\beta}_n)$ be a sequence of estimators that take their values in $T \times B$ and satisfy*

$$(\hat{\tau}_n, \hat{\beta}_n) - (\bar{\tau}_n, \bar{\beta}_n) = O_p(n^{-1/2}),$$

where $(\bar{\tau}_n, \bar{\beta}_n) \in T \times B$ is a (non-random) sequence which is eventually uniformly in the interior[5] of $T' \times B'$, which itself is contained in the interior of $T \times B$.

Note that only weights $w(j, n)$ with $0 \leq j \leq n-1$ enter the formula for $\hat{\Psi}_n$. Hence, without loss of generality, we can always assume that there exists an index ℓ_n with $0 < \ell_n \leq n$ such that $w(j, n) = 0$ for all $j \geq \ell_n$; for definiteness, we will always take ℓ_n to be the smallest such index.[6]

As discussed in more detail below, the rate at which the truncation lag ℓ_n approaches infinity as compared to n will be crucial for the consistency

[4]This assumption (together with a boundedness condition on the second moments and a condition on the weighting scheme) is used to show that the estimator $\hat{\Psi}_n$ differs from the corresponding expression where $\hat{\mathbf{v}}_{t,n}$ is replaced with $\mathbf{v}_{t,n}$ only by a term which is $o_p(1)$. More generally, to establish that replacing $\hat{\mathbf{v}}_{t,n}$ with $\mathbf{v}_{t,n}$ has no effect asymptotically, it suffices to establish consistency of $(\hat{\tau}_n, \hat{\beta}_n)$ and a uniform stochastic equicontinuity property of $w(0,n)[n^{-1}\Sigma_{t=1}^{n}\mathbf{v}_t\mathbf{v}_t'] + \Sigma_{j=1}^{n-1}w(j,n)[n^{-1}\Sigma_{t=1}^{n-j}(\mathbf{v}_t\mathbf{v}_{t+j}' + \mathbf{v}_{t+j}\mathbf{v}_t')]$ as a function of (τ, β), cf., e.g., Andrews (1992) and Pötscher and Prucha (1994a).

[5]I.e., there exists an $\epsilon > 0$ such that the Euclidean distance from $(\bar{\tau}_n, \bar{\beta}_n)$ to the complement of $T' \times B'$ relative to the $p_\tau + p_\beta$-dimensional Euclidean space exceeds ϵ for all large n. Note also that here $T' \times B'$ need not be compact.

[6]We have $\ell_n > 0$, since we exclude the degenerate case $w(j, n) = 0$ for all j.

of $\hat{\Psi}_n$. Recall that $\mathbf{v}_{t,n} = \mathbf{v}_t(\bar{\tau}_n, \bar{\beta}_n)$ and $\hat{\mathbf{v}}_{t,n} = \mathbf{v}_t(\hat{\tau}_n, \hat{\beta}_n)$. We define

$$
\begin{aligned}
a(0,n) &= \left[E \left| n^{-1} \sum_{t=1}^{n} [\mathbf{v}_{t,n}\mathbf{v}_{t,n}' - E\mathbf{v}_{t,n}\mathbf{v}_{t,n}'] \right|^2 \right]^{1/2}, \\
a(j,n) &= \left[E \left| n^{-1} \sum_{t=1}^{n-j} [\mathbf{v}_{t,n}\mathbf{v}_{t+j,n}' - E\mathbf{v}_{t,n}\mathbf{v}_{t+j,n}' \right.\right. \\
&\qquad \left.\left. + \mathbf{v}_{t+j,n}\mathbf{v}_{t,n}' - E\mathbf{v}_{t+j,n}\mathbf{v}_{t,n}'] \right|^2 \right]^{1/2}
\end{aligned}
$$

for $j \geq 1$,

$$
\begin{aligned}
\eta(0,n) &= \left| n^{-1} \sum_{t=1}^{n} E\mathbf{v}_{t,n}\mathbf{v}_{t,n}' \right|, \\
\eta(j,n) &= \left| n^{-1} \sum_{t=1}^{n-j} [E\mathbf{v}_{t,n}\ \mathbf{v}_{t+j,n}' + E\mathbf{v}_{t+j,n}\mathbf{v}_{t,n}'] \right|
\end{aligned}
$$

for $j \geq 1$. The terms $a(j,n)$ can be interpreted as a measure of the distance between the empirical and theoretical second moments of the process $(\mathbf{v}_{t,n})$. Furthermore define

$$
\Delta_t = \sup \{ |\mathbf{v}_t(\tau,\beta)|,\ |\nabla_\beta \mathbf{v}_t(\tau,\beta)|,\ |\nabla_\tau \mathbf{v}_t(\tau,\beta)| : (\tau,\beta) \in T' \times B' \},
$$

where T' and B' are the subsets given in Assumption 12.1.

Lemma 12.1. *Let Assumption 12.1 hold. Assume that* $\|\mathbf{v}_{t,n}\|_2 < \infty$ *for all* $t \in \mathbf{N}$ *and* $n \in \mathbf{N}$, *and that*

$$
n^{-1} \sum_{t=1}^{n} \Delta_t^2 = O_p(1).
$$

(a) Then $\hat{\Psi}_n - \Psi_n = O_p(\gamma_n)$ *where* $\gamma_n = \max\{\gamma_{1n}, \gamma_{2n}, \gamma_{3n}\}$ *with*

$$
\gamma_{1n} = n^{-1/2} \sum_{j=0}^{n-1} |w(j,n)|,
$$

$$
\gamma_{2n} = \sum_{j=0}^{n-1} |w(j,n)|\, a(j,n)
$$

and

$$
\gamma_{3n} = \sum_{j=0}^{n-1} |w(j,n) - 1|\, \eta(j,n).
$$

Hence, if γ_{1n}, γ_{2n}, *and* γ_{3n} *converge to zero as* $n \to \infty$, *we have* $\hat{\Psi}_n - \Psi_n \to 0$ *i.p. as* $n \to \infty$.

(b) If the weights are bounded (i.e., $\sup_n \sup_j |w(j,n)| < \infty$ *) then*

$$\gamma_{1n} = O(\ell_n / n^{1/2}).$$

Hence $\lim_{n\to\infty} \gamma_{1n} = 0$ *if* $\ell_n = o(n^{1/2})$.

(c) If the weights are bounded then

$$\gamma_{2n} = O\left(\sum_{j=0}^{\ell_n - 1} a(j,n)\right).$$

Hence $\lim_{n\to\infty} \gamma_{2n} = 0$ *if* $\sum_{j=0}^{\ell_n-1} a(j,n) = o(1)$.

(d) If the weights are bounded, $\lim_{n\to\infty} w(j,n) = 1$ *for all* $j \geq 0$, *and if*

$$\sum_{j=0}^{\infty} \sup_{n\geq 1} \eta(j,n) < \infty$$

then

$$\lim_{n\to\infty} \gamma_{3n} = 0.$$

(e) If $|w(j,n) - 1| \leq \delta_j \kappa_n$ *for* $0 \leq j < n$ *and if*

$$\sum_{j=0}^{\infty} \delta_j \sup_{n\geq 1} \eta(j,n) < \infty$$

then

$$\gamma_{3n} = O(\kappa_n).$$

Hence $\lim_{n\to\infty} \gamma_{3n} = 0$ *if* $\kappa_n = o(1)$.

Part (a) of Lemma 12.1 gives a bound γ_n for the order of magnitude of $\hat{\Psi}_n - \Psi_n$. Parts (b), (c) and (d) (or (e)) provide a range for the rate of increase of the truncation lag ℓ_n that ensures consistency of $\hat{\Psi}_n$. More specifically, parts (b) and (c) imply an upper bound on the rate of ℓ_n. Part (d) as well as part (e) (with $\kappa_n = o(1)$) imply a lower bound, namely that $\ell_n \to \infty$ as $n \to \infty$. (This follows from $\lim_{n\to\infty} w(j,n) = 1$ for all $j \geq 0$.)

The proof of Lemma 12.1 is based on the inequality

$$\left|\hat{\Psi}_n - \Psi_n\right| \leq \left|\hat{\Psi}_n - \bar{\Psi}_n\right| + \left|\bar{\Psi}_n - E\bar{\Psi}_n\right| + \left|E\bar{\Psi}_n - \Psi_n\right| \tag{12.3}$$

where the pseudo-estimator

$$\begin{aligned} \bar{\Psi}_n &= w(0,n)\left[n^{-1}\sum_{t=1}^{n} \mathbf{v}_{t,n}\mathbf{v}_{t,n}'\right] \\ &\quad + \sum_{j=1}^{n-1} w(j,n)\left[n^{-1}\sum_{t=1}^{n-j}(\mathbf{v}_{t,n}\mathbf{v}_{t+j,n}' + \mathbf{v}_{t+j,n}\mathbf{v}_{t,n}')\right] \end{aligned}$$

is obtained from $\hat{\Psi}_n$ by replacing $\hat{\mathbf{v}}_{t,n}$ with the unobserved $\mathbf{v}_{t,n}$. The first term on the right hand side of (12.3) captures the effect that $(\bar{\tau}_n, \bar{\beta}_n)$ has to be estimated by $(\hat{\tau}_n, \hat{\beta}_n)$, and is shown to be $O_p(\gamma_{1n})$. Part (b) of the lemma shows that this term is $O(\ell_n/n^{1/2})$ under the mild condition that the weights are bounded. The second term represents the deviation of $\bar{\Psi}_n$ from its mean. This term is shown to be $O_p(\gamma_{2n})$ by bounding its variance. Intuitively it is clear that for the variance to tend to zero we want ℓ_n to be small relative to n, in order to keep the number of sample covariances in the formula for $\bar{\Psi}_n$ small relative to n. This is confirmed by an inspection of the bound γ_{2n}: The terms $a(j,n)$ appearing in the formula for γ_{2n} measure the quality of the estimators $n^{-1}\sum_{t=1}^{n-j}\mathbf{v}_{t,n}\mathbf{v}'_{t+j,n}$ of the theoretical second moments $n^{-1}\sum_{t=1}^{n-j}E\mathbf{v}_{t,n}\mathbf{v}'_{t+j,n}$ at lag j. The faster $a(j,n)$ tends to zero for each j (i.e., the better the second moments can be estimated) the faster ℓ_n may grow (i.e., the more lags can be included into $\bar{\Psi}_n$) while still retaining the convergence of γ_{2n} to zero. Hence, convergence of the second term on the right hand side of (12.3) to zero typically implies an upper bound on the rate at which ℓ_n can increase to infinity. The third term on the right hand side of (12.3) finally is a bias term and its order of magnitude is bounded by γ_{3n}. The terms $\eta(j,n)$ reflect the size of the theoretical second moments. They are required to decline as $j \to \infty$ via the condition

$$\sum_{j=0}^{\infty}\sup_{n\geq 1}\eta(j,n) < \infty.$$

This condition will typically only be satisfied for processes that are essentially centered at zero and is then a condition on the memory of the process. The larger the terms $\eta(j,n)$, the larger will typically be the bias $|E\bar{\Psi}_n - \Psi_n|$ due to the down-weighing (and exclusion for $j \geq \ell_n$) of second moments. In general, for the bias to vanish asymptotically the weights $w(j,n)$ will have to converge to unity and, in particular, ℓ_n will have to go to infinity. Clearly, the slower the decline of $\eta(j,n)$ as $j \to \infty$, the faster ℓ_n will have to go to infinity in order to retain the property that the bias converges to zero. Hence, in general, convergence of the third term on the right hand side of (12.3) to zero implies a lower bound on the rate of ℓ_n. Summarizing, we note that in order to obtain consistency we have to balance the variance and the bias of the estimator, and this entails upper and lower bounds on the rate of increase of ℓ_n.[7]

The weights $w(j,n)$ are often obtained from a kernel $W : \mathbf{R} \to [-1,1]$ via $w(j,n) = W(j/\ell_n^*)$, with $W(0) = 1$, $W(x) = W(-x)$ for all x, $W(x) = 0$ for $|x| > 1$ and bandwidth parameter ℓ_n^*, $0 < \ell_n^* < n$. The boundedness assumption on $w(j,n)$ is then automatically satisfied. Typically the kernels W are also at least piecewise continuous. Commonly used kernels like

[7] The proof in fact shows that in case $\hat{\mathbf{v}}_{t,n} = \mathbf{v}_{t,n}$ the convergence of $\hat{\Psi}_n - \Psi_n$ to zero in Lemma 12.1(a) holds in the L_2-sense and not only in probability.

the rectangular, the Bartlett, the Parzen, and the Blackman-Tukey kernel satisfy all these conditions, cf. Brockwell and Davis (1991, Ch.10), Priestley (1981, Ch.6). These kernels are in fact positive on $(-1,1)$ and hence $\ell_n = [\ell_n^*]+1$ or $\ell_n = [\ell_n^*]$, where $[\ell_n^*]$ denotes the largest integer not greater than ℓ_n^*.[8] For any kernel W we have $\ell_n \leq [\ell_n^*]+1$. If also $W(x) \neq 0$ holds in a neighborhood of zero we have furthermore $0 < \text{const}*\ell_n^* \leq \ell_n$ and hence any rate of convergence expressed in terms of ℓ_n can be expressed equivalently in terms of ℓ_n^*.

Given the weights $w(j,n)$ are obtained from a kernel W, it is possible to determine the rate κ_n and the constants δ_j in Lemma 12.1(e) as follows: Suppose $W(x)$ is ρ times continuously differentiable in a neighborhood of zero with i-th derivatives $W^{(i)}$ satisfying $W^{(i)}(0) = 0$ for $i < \rho$. From a Taylor expansion we obtain

$$|W(x) - W(0)| = \left|W^{(\rho)}(\zeta)\right| |x|^{\rho} / \rho! \leq \text{const} * |x|^{\rho}$$

for all $|x| < \epsilon$ for some $\epsilon > 0$ small enough, observing that $|\zeta| \leq |x|$ and that $W^{(\rho)}$ is continuous and hence bounded on a sufficiently small neighborhood of zero. Since $W(x)$ is bounded and $|x|^{\rho} \geq \epsilon^{\rho} > 0$ for $|x| \geq \epsilon$ we obtain

$$|W(x) - W(0)| \leq \text{const} * |x|^{\rho}$$

even for all $|x| \leq 1$, possibly after changing the constant on the right hand side. Consequently

$$|w(j,n) - 1| = |W(j/\ell_n^*) - W(0)| \leq \text{const} * (j/\ell_n^*)^{\rho}$$

for $0 \leq j < n$. Hence we can choose $\kappa_n = (\ell_n^*)^{-\rho}$ and $\delta_j = \text{const}*j^{\rho}$. More generally, we arrive at the same result if $W(x)$ is only assumed to satisfy

$$\lim_{x \to 0} |W(x) - 1| / |x|^{\rho} < \infty$$

where $\rho > 0$ (and where ρ need not be an integer).[9] Clearly, the larger the value of ρ for which this condition is satisfied, the flatter and smoother the kernel will be at zero. For a given sequence of $\ell_n^* \to \infty$, the bound γ_{3n} for the bias term will converge to zero the faster, the larger the value of ρ can be chosen subject to the restriction

$$\sum_{j=0}^{\infty} j^{\rho} \sup_{n \geq 1} \eta(j,n) < \infty.$$

[8]The tuning constant of the Blackman-Tukey kernel is assumed (as usual) to be in the interval $[0, 0.25]$.

[9]Note that for such a kernel $W(x) \neq 0$ in a neighborhood of zero. Hence ℓ_n and ℓ_n^* are, as discussed above, of the same order of magnitude.

That is, the rate γ_{3n} guaranteed by Lemma 12.1 for the bias term is limited by two facts: On the one hand by the smoothness and flatness degree of W, and on the other hand by the rate of decline of the covariance function as expressed in the decline of $\sup_{n\geq 1}\eta(j,n)$. I.e., if W satisfies the above condition for ρ_1 and $\sup_{n\geq 1}\eta(j,n)$ satisfies

$$\sum_{j=0}^{\infty} j^{\rho_2}\sup_{n\geq 1}\eta(j,n) < \infty,$$

then the "best" feasible choice is $\kappa_n = (\ell_n^*)^{-\rho}$ with $\rho = \min(\rho_1,\rho_2)$. In the context of mean zero stationary processes $\sup_{n\geq 1}\eta(j,n)$ equals the absolute value of the covariance function. The "summability" condition

$$\sum_{j=0}^{\infty} j^{\rho}\sup_{n\geq 1}\eta(j,n) < \infty$$

on the covariance function can also be expressed, as can be shown, as a smoothness condition on the associated spectral density function. Hence the size of the bias of $\hat{\Psi}_n$ is determined by two factors, the smoothness of the spectral density and the smoothness and flatness of the kernel. We note that the above discussion of the bias term is based on well-known results in Parzen (1957) for the bias of spectral density estimators, see also Anderson (1971, Ch.9).

The following result is obtained as a corollary to Lemma 12.1 applied to near epoch dependent processes. The corollary does not maintain that the weights are generated from a kernel.

Corollary 12.2. *Let Assumption 12.1 hold. Assume that*

$$\sup_n \sup_t \left\|\mathbf{v}_{t,n}'\mathbf{v}_{t,n}\right\|_r < \infty$$

for some $r > 2$ and that

$$n^{-1}\sum_{t=1}^{n} E\Delta_t^2 = O(1).$$

Let $(\mathbf{v}_{t,n})$ be a zero mean process which is near epoch dependent of size $-2(r-1)/(r-2)$ on a basis process, which is α-mixing with mixing coefficients of size $-r/(r-2)$ or ϕ-mixing with mixing coefficients of size $-r/(r-1)$. Assume that the weights $w(j,n)$ are bounded, and $\lim_{n\to\infty} w(j,n) = 1$ for each $j \geq 0$ (which implies $\ell_n \to \infty$). Then

$$\gamma_{1n} = O(\ell_n/n^{1/2}),$$

$$\gamma_{2n} = O(\ell_n^{3/2}/n^{1/2})$$

and

$$\gamma_{3n} = o(1).$$

Hence, $\hat{\Psi}_n - \Psi_n \to 0$ *i.p. as* $n \to \infty$ *if the weighting scheme satisfies additionally* $\ell_n = o(n^{1/3})$.

The corollary thus gives $n^{1/3}$ as an upper "bound" on the range of orders of magnitude for ℓ_n that ensure consistency of $\hat{\Psi}_n$. We see from the corollary that the bound γ_{1n} for the order of magnitude of $|\hat{\Psi}_n - \bar{\Psi}_n|$ is always of a smaller order than γ_{2n}, the bound for the "variance" term. Hence the upper "bound" on the range of orders of magnitude for ℓ_n that ensure consistency of $\hat{\Psi}_n$ according to Corollary 12.2 is only determined by γ_{2n}. To obtain $\gamma_{3n} = o(1)$ the condition $\lim_{n\to\infty} w(j,n) = 1$ for each $j \geq 0$ is needed. This implies $\ell_n \to \infty$ as $n \to \infty$, which provides the lower "bound" for the rate of increase of ℓ_n. The rate $O(\ell_n^{3/2}/n^{1/2})$ for γ_{2n} is obtained under near epoch dependence by establishing that $a(j,n)^2 \leq \text{const}{*}(j{+}1)/n$. Of course, improvement of this bound would lead to an improvement in the rate for γ_{2n}. For example, for the class of linear processes with absolutely summable coefficients and based on an i.i.d. noise sequence with finite fourth moments it is known that $a(j,n)^2 \leq \text{const}/n$, cf. Anderson (1971), Corollary 8.3.1; using this bound in Lemma 12.1 leads to a rate $\gamma_{2n} = O(\ell_n/n^{1/2})$, and hence to the weaker condition $\ell_n = o(n^{1/2})$ for consistency. (Note, however, that in this example the rate can be improved even further by a refinement of the technique of proof as discussed at the end of this section.)

The corollary only shows the bound for the bias term γ_{3n} to be $o(1)$, but does not give information on the rate of convergence of this term, and hence of $\hat{\Psi}_n - \Psi_n$, to zero. Such a rate is established in the following corollary under additional assumptions on the weights and under stricter memory conditions on the process $(\mathbf{v}_{t,n})$. Let $\mathcal{W}(\rho)$ for $\rho > 0$ denote the class of kernels $W : \mathbf{R} \to [-1,1]$ satisfying $W(0) = 1$, $W(x) = W(-x)$ for all x, $W(x) = 0$ for $|x| > 1$, and

$$\lim_{x\to 0} |W(x) - 1| \,/\, |x|^{\rho} < \infty.$$

Recall that, as discussed above, for this class of kernels ℓ_n and ℓ_n^* are of the same order of magnitude and hence the rates appearing in the following corollary could be formulated equivalently in terms of ℓ_n.

Corollary 12.3. *Let Assumption 12.1 hold. Assume that*

$$\sup_n \sup_t \left\| \mathbf{v}'_{t,n}\mathbf{v}_{t,n} \right\|_r < \infty$$

for some $r > 2$ *and that*

$$n^{-1} \sum_{t=1}^{n} E\Delta_t^2 = O(1).$$

Let $(\mathbf{v}_{t,n})$ *be a zero mean process which is near epoch dependent of size* $\min\{-2(r-1)/(r-2), -(\rho+1)\}$ *on a basis process, which is* α*-mixing with mixing coefficients of size* $-(\rho+1)r/(r-2)$ *or* ϕ*-mixing with mixing coefficients of size* $-(\rho+1)r/(r-1)$*, and let* $w(j,n) = W(j/\ell_n^*)$ *with* $W(.) \in \mathcal{W}(\rho)$ *and* $\rho > 0$*. Then*

$$\gamma_{1n} = O(\ell_n^*/n^{1/2}),$$

$$\gamma_{2n} = O\left((\ell_n^*)^{3/2}/n^{1/2}\right)$$

and

$$\gamma_{3n} = O\left((\ell_n^*)^{-\rho}\right).$$

Hence, if $\ell_n^* \to \infty$ *and* $\ell_n^* = o(n^{1/3})$ *then* $\hat{\Psi}_n - \Psi_n \to 0$ *i.p. as* $n \to \infty$*; in fact*

$$\hat{\Psi}_n - \Psi_n = O_p(\gamma_n)$$

where

$$\gamma_n = \max\left((\ell_n^*)^{3/2}/n^{1/2}, (\ell_n^*)^{-\rho}\right).$$

Furthermore, the optimal rate for γ_n *is* $n^{-\rho/(2\rho+3)}$*, which is achieved for* $\ell_n^* \sim n^{1/(2\rho+3)}$.

The rate of the bound γ_n for the order of magnitude of $\hat{\Psi}_n - \Psi_n$ is solely determined by γ_{2n}, which controls the variance of $\bar{\Psi}_n$, and by γ_{3n}, which controls the bias of $\bar{\Psi}_n$, since γ_{2n} dominates γ_{1n}. The optimal rate for γ_n is hence obtained by equating the orders of magnitude of γ_{2n} and γ_{3n}. Clearly, in cases where the size requirements in the above corollary are satisfied with a value ρ_2 in place of ρ, and where the kernel W belongs to $\mathcal{W}(\rho_1)$, the corollary is applicable with $\rho = \min(\rho_1, \rho_2)$. Similarly as in the discussion preceding Corollary 12.2 we note that ρ_1 measures the smoothness and flatness of the kernel W at $x = 0$, whereas ρ_2 (together with r) determines the rate of decline of the covariance function (which can be expressed in terms of smoothness of the spectral density in the stationary case). Corollary 12.3 hence shows that $\rho = \min(\rho_1, \rho_2)$ together with ℓ_n^* determines the rate of convergence of γ_n.

The Bartlett, the Parzen, and the Blackman-Tukey kernel belong to $\mathcal{W}(\rho)$ for all $\rho > 0$ less than or equal to 1, 2, and 2, respectively, see Anderson (1971, p.527). Hence the optimal rate for γ_n that is achievable with the Bartlett, the Parzen, and the Blackman-Tukey kernels is $n^{-1/5}, n^{-2/7}$, and $n^{-2/7}$, respectively, given the memory conditions in Corollary 12.3 are satisfied for ρ equal to 1, 2, and 2, respectively. The rectangular kernel belongs to $\mathcal{W}(\rho)$ for all $0 < \rho < \infty$. The optimal rate for γ_n that is achievable with the rectangular kernel is therefore only determined by the memory conditions on the process $\mathbf{v}_{t,n}$; for example, it is arbitrarily close to $n^{-1/2}$ for processes which satisfy the memory conditions in Corollary 12.3 for arbitrarily large ρ.

Corollary 12.2 and 12.3 extend results in Newey and West (1987), Gallant (1987a) and Gallant and White (1988). The result in Newey and West (1987) is for α-mixing and ϕ-mixing processes only, whereas Gallant (1987a) and Gallant and White (1988) give results for near epoch dependent processes. All these results give $\ell_n = o(n^{1/4})$ as a condition for consistency.[10] Corollary 12.2 improves these results by showing that $\ell_n = o(n^{1/3})$ suffices for consistency in the context of near epoch dependent processes, while maintaining the same size requirements for the approximation errors ν_m as in Gallant (1987a) and Gallant and White (1988).[11] Furthermore, the size requirements on the α-mixing coefficients are weaker than those in Newey and West (1987), Gallant (1987a) and Gallant and White (1988). In contrast to these references and to Corollary 12.2, Corollary 12.3 also provides a (bound for the) rate of convergence for $\hat{\Psi}_n$.

Corollary 12.2 and 12.3 are in the spirit of consistency results for spectral density estimators.[12] From the time series literature, e.g., Parzen (1957) and Anderson (1971, Ch.9), it is known that for kernels W as in Corollary 12.3 and for a stationary time series $(\mathbf{v}_{t,n}) = (\mathbf{v}_t)$ where $\mathbf{v}_t$ is observable, the term $\left|\bar{\Psi}_n - E\bar{\Psi}_n\right|$ is even $O(\ell_n^*/n)$, and hence $\ell_n^* = o(n)$ and $\ell_n^* \to \infty$ are sufficient for the consistency of $\bar{\Psi}_n$. (Given $\mathbf{v}_t$ is observable, $\bar{\Psi}_n$ is observable and $\gamma_{1n} = 0$.) As a consequence the optimal rate for γ_n is then improved. These results are derived by establishing a tighter bound for the order of magnitude of $\left|\bar{\Psi}_n - E\bar{\Psi}_n\right|$ than γ_{2n} in terms of the fourth order cumulants of the process and a mixing condition formulated in terms of those cumulants.[13] (For comparison, recall from the discussion after Corollary 12.2 that, in case of linear processes with absolutely summable coefficients and based on an i.i.d. noise sequence with finite fourth moments, Lemma 12.1 implies $\gamma_{2n} = O(\ell_n/n^{1/2}) = O(\ell_n^*/n^{1/2})$.) Andrews (1991b) uses Parzen's results to derive analogous ones for the case where $\mathbf{v}_t$ is not observable but has to be estimated by $\hat{\mathbf{v}}_{t,n}$, by showing that the bound $\gamma_{1n} = O(\ell_n^*/n^{1/2})$ can be improved to a bound which is $O(\ell_n^*/n)$. He then extends those results to certain heteroskedastic processes that satisfy a

---

[10] White (1984) and White and Domowitz (1984) give consistency results for α-mixing and ϕ-mixing processes with $\ell_n = o(n^{1/3})$. However, Newey and West (1987) point out that their proofs are incorrect; cf. also Gallant and White (1988, pp.99).

[11] Quah (1990) also gives consistency results with $\ell_n = o(n^{1/3})$ but only for α-mixing processes and (a variant of) ϕ-mixing processes. He furthermore suggests that his results could also be used to show that the results in Gallant and White (1988) for near epoch dependent processes remain intact with $\ell_n = o(n^{1/3})$ rather than with $\ell_n = o(n^{1/4})$. However, such a demonstration based on his results seems to be less than obvious.

[12] See, e.g., Priestley (1981, Ch.6) and the references cited therein.

[13] This improvement of the bound for $|\bar{\Psi}_n - E\bar{\Psi}_n|$ is not limited to the stationary case. In the generic setting of Lemma 12.1 this would, however, lead to rather involved conditions.

similar mixing condition formulated in terms of fourth order cumulants. Related results are also given in Keener, Kmenta and Weber (1991) and Hansen (1992).

12.3 Further Remarks

(i) Without further conditions on the weights $w(j,n)$ the estimators $\hat{\Psi}_n$ are not necessarily nonnegative definite. For example, weights generated from the rectangular kernel with $\ell_n < n$ may give estimates that are not nonnegative definite. To obtain a condition that ensures nonnegative definiteness of $\hat{\Psi}_n$ first note that $\hat{\Psi}_n$ can be written as

$$\int_{-\pi}^{\pi} \tilde{W}_n(\omega) I_n(\omega) d\omega$$

where

$$\tilde{W}_n(\omega) = (2\pi)^{-1} \sum_{j=-n+1}^{n-1} w(|j|, n) \exp(-i\omega j)$$

is the Fourier-transform of the weights, and where

$$I_n(\omega) = n^{-1} \left[\sum_{t=1}^{n} \hat{\mathbf{v}}_{t,n} \exp(-i\omega t) \right] \left[\sum_{t=1}^{n} \hat{\mathbf{v}}_{t,n} \exp(-i\omega t) \right]^*$$

is the periodogram of $\hat{\mathbf{v}}_{t,n}$. Since by definition $I_n(\omega)$ is always nonnegative definite, it follows that $\hat{\Psi}_n$ is nonnegative definite if the Fourier-transform of the weights $\tilde{W}_n(\omega)$ is nonnegative for all $\omega \in [-\pi, \pi]$.[14] This condition is equivalent to the existence of constants $c(j,n)$ such that

$$w(j,n) = \sum_{k=j}^{n-1} c(k,n) c(k-j,n).^{15}$$

This fact is well-known in the time series literature. The latter condition is also given in Newey and West (1987) and Gallant and White (1988).

(ii) We have excluded kernels W which have unbounded support from our analysis for simplicity of presentation. In particular, the sufficient conditions in Lemma 12.1(b)-(e) are primarily designed to cover weights generated from kernels with bounded support. We note that commonly used

[14] E.g., the weights generated from the Bartlett kernel or the Parzen kernel satisfy this condition.

[15] Since $\tilde{W}_n(\omega)$ is even, nonnegative and a trigonometric polynomial it can formally be viewed as the spectral density of a moving average process. It hence has a representation of the form $\tilde{W}_n(\omega) = |\sum_{j=0}^{n-1} c(j,n) \exp(-i\omega j)|^2$ from which the representation of the $w(j,n)$ in terms of the $c(j,n)$ follows.

kernels like the rectangular, the Bartlett, the Parzen, and the Blackman-Tukey kernel all have bounded support. A similar analysis for kernels with unbounded support is possible starting directly from Lemma 12.1(a). For a discussion which allows for kernels with unbounded support see, e.g., Priestley (1981, Ch.6) and Andrews (1991b).

(iii) The question which kernel W and bandwidth parameter ℓ_n^* to choose has been discussed extensively in the time series literature, see, e.g., Priestley (1981, Ch.7). For stationary processes it is not only possible to bound the mean-square error of the estimator, but also to find an asymptotic expansion for the mean-square error. The leading term in this expansion depends on the kernel W and on the bandwidth parameter ℓ_n^*. One can then attempt to minimize the leading term for the mean-square error with respect to W and ℓ_n^*. It has been shown that the Bartlett-Priestley (or quadratic) kernel is optimal in this sense (in the class of kernels having a nonnegative Fourier transform), see Priestley (1981, Ch.7) and the references therein. The expression for the corresponding optimal sequence of bandwidth parameters ℓ_n^* depends on characteristics of the underlying data generating process. A feasible implementation of the optimal bandwidth parameter sequence ℓ_n^* requires estimation of these unknown characteristics. In the context of variance covariance matrix estimation such procedures are discussed in Andrews (1991b) and Newey and West (1994). It seems that the choice of the kernel is not overly important for the mean-square error of the estimator, the choice of the bandwidth sequence being the more crucial factor, cf. Priestley (1981, p.449).

(iv) For stationary (mean zero) processes the size of the bias of $\bar{\Psi}_n$ (and of $\hat{\Psi}_n$) depends on the smoothness of the spectral density of the data generating process, as was discussed above. Translated to a finite samples situation this indicates that it is easier to estimate a "flat" spectral density than a "wildly fluctuating" one. On this basis Press and Tukey (1956) suggested to "prewhiten" the data process. More specifically, they suggested to first subject the data to a linear transformation which makes the series look more like an uncorrelated series and to estimate the spectral density of the transformed data. Dividing this estimate by the square of the absolute value of the spectral characteristic of the linear transformation (i.e., by reversing the linear transformation in the frequency domain) then yields an estimate of the spectral density of the original data. Of course, the above procedure requires sufficient knowledge about the shape of the spectral density function to be able to come up with the appropriate transformation. If a priori knowledge is not available this transformation has to be estimated from the data. One way to obtain such a linear transformation is to fit an autoregressive model to the data. The residuals from this model then represent the transformed data. Such a procedure is, e.g., discussed in Priestley (1981, p.603). Andrews and Monahan (1992) also consider estimators for Ψ_n of this type. For a proposal using autoregressive moving average models at the prewhitening stage see Lee and Phillips (1994). For a further discussion of practical issues pertaining to the estimation of spectral density functions (and hence of $\hat{\Psi}_n$) see Priestley (1981, Ch.7).

(v) Corollary 12.2 and 12.3 assume $E\mathbf{v}_{t,n} = 0$.[16] Under misspecification this assumption may often be violated. In the context of Theorems 11.2 and 11.5 the weaker condition $\sum_{t=1}^{n} E\mathbf{v}_{t,n} = 0$ is maintained.[17] Given this condition, $\hat{\Psi}_n$ will (under weak additional assumptions) asymptotically overestimate Ψ_n as it – loosely speaking – estimates Ψ_n plus a squared bias term. If $\hat{\Psi}_n$ is then used to construct confidence regions, the confidence regions will be conservative. For related discussions regarding the estimation of variance covariance matrices under misspecification see Gallant (1987a, Ch.7) and Gallant and White (1988, Ch.6).

(vi) For Monte Carlo simulations that compare the small sample properties of alternative estimators of Ψ_n see Andrews (1991b), Andrews and Monahan (1992), and Newey and West (1994); cf. also Keener, Kmenta and Weber (1991).

---

[16]Lemma 12.1 does not explicitly require $E\mathbf{v}_{t,n} = 0$. However the summability conditions in parts (d) and (e) will typically only be satisfied if the process essentially has zero mean.

[17]As discussed in detail in Chapter 11, this weaker assumption will typically hold in case of least mean distance estimators even under misspecification. For generalized method of moments estimators, however, even this weaker assumption will typically be violated under misspecification.

13

CONSISTENT VARIANCE COVARIANCE MATRIX ESTIMATION: CATALOGUES OF ASSUMPTIONS

In this chapter we provide consistency results for the estimation of the variance covariance matrices of least mean distance and generalized method of moments estimators as given in parts (a) of Theorems 11.2 and 11.5. Consistency results for the estimation of the variance covariance matrices as given in parts (b) of those theorems can be obtained analogously.

13.1 Estimation of the Variance Covariance Matrix of Least Mean Distance Estimators

The asymptotic variance covariance matrix of the least mean distance estimator in Theorem 11.2(a) is of the form $C_n^{-1} D_n D_n' C_n^{-1\prime}$ where

$$C_n = E\nabla_{\beta\beta}\underline{S}_n$$

and

$$D_n D_n' = nE(\nabla_{\beta'}\underline{S}_n \nabla_{\beta}\underline{S}_n).$$

Consistent estimation of the asymptotic variance covariance matrix hence can be reduced to consistent estimation of C_n and $D_n D_n'$. In this section let $\mathbf{v}_t(\tau,\beta) = \nabla_{\beta'} q_t(\mathbf{z}_t,\tau,\beta)$ and hence $\hat{\mathbf{v}}_{t,n} = \nabla_{\beta'} q_t(\mathbf{z}_t,\hat{\tau}_n,\hat{\beta}_n)$. Consider the following estimators:

$$\begin{aligned}\hat{C}_n &= n^{-1}\sum_{t=1}^{n}\nabla_{\beta\beta}q_t(\mathbf{z}_t,\hat{\tau}_n,\hat{\beta}_n),\\ \hat{\Phi}_n &= n^{-1}\sum_{t=1}^{n}\hat{\mathbf{v}}_{t,n}\hat{\mathbf{v}}_{t,n}' = n^{-1}\sum_{t=1}^{n}\nabla_{\beta'}q_t(\mathbf{z}_t,\hat{\tau}_n,\hat{\beta}_n)\nabla_{\beta}q_t(\mathbf{z}_t,\hat{\tau}_n,\hat{\beta}_n),\end{aligned}$$

B. M. Pötscher et al., *Dynamic Nonlinear Econometric Models*

$$
\begin{aligned}
\hat{\Psi}_n &= w(0,n)\left[n^{-1}\sum_{t=1}^{n}\hat{\mathbf{v}}_{t,n}\hat{\mathbf{v}}'_{t,n}\right] \\
&\quad + \sum_{j=1}^{n-1} w(j,n)\left[n^{-1}\sum_{t=1}^{n-j}\left[\hat{\mathbf{v}}_{t,n}\hat{\mathbf{v}}'_{t+j,n} + \hat{\mathbf{v}}_{t+j,n}\hat{\mathbf{v}}'_{t,n}\right]\right] \\
&= w(0,n)\left[n^{-1}\sum_{t=1}^{n}\nabla_{\beta'}q_t(\mathbf{z}_t,\hat{\tau}_n,\hat{\beta}_n)\nabla_{\beta}q_t(\mathbf{z}_t,\hat{\tau}_n,\hat{\beta}_n)\right] \\
&\quad + \sum_{j=1}^{n-1} w(j,n)\Bigg[n^{-1}\sum_{t=1}^{n-j}\Big[\nabla_{\beta'}q_t(\mathbf{z}_t,\hat{\tau}_n,\hat{\beta}_n)\nabla_{\beta}q_{t+j}(\mathbf{z}_{t+j},\hat{\tau}_n,\hat{\beta}_n) \\
&\quad + \nabla_{\beta'}q_{t+j}(\mathbf{z}_{t+j},\hat{\tau}_n,\hat{\beta}_n)\nabla_{\beta}q_t(\mathbf{z}_t,\hat{\tau}_n,\hat{\beta}_n)\Big]\Bigg],
\end{aligned}
$$

where $\hat{\Psi}_n$ is of the form given in (12.2). The estimator $\hat{\Phi}_n$ is appropriate in the case where $\mathbf{v}_{t,n} = \nabla_{\beta'}q_t(\mathbf{z}_t,\bar{\tau}_n,\bar{\beta}_n)$ has mean zero and is uncorrelated (e.g., a martingale difference), whereas $\hat{\Psi}_n$ has to be used if $\mathbf{v}_{t,n}$ is autocorrelated. The subsequent assumptions strengthen Assumptions 11.2, 11.3 and 11.5.

Assumption 11.2*. *The family $\{f_t : t \in \mathbf{N}\}$ is equicontinuous on $Z \times T' \times B'$ and*

$$\sup_n n^{-1}\sum_{t=1}^{n} E\left[\sup_{T'\times B'} |f_t(\mathbf{z}_t,\tau,\beta)|^{1+\gamma}\right] < \infty$$

for some $\gamma > 0$, where f_t denotes the restriction to $Z\times T'\times B'$ of any of the components of $\nabla_{\beta\tau}q_t$, $\nabla_{\beta\beta}q_t$, or $\nabla_{\beta'}q_t\nabla_{\beta}q_t$. (Here $T'\times B'$ is the compact subset employed in Assumption 11.1(d).)

Assumption 11.3*. *(a)* $E\nabla_{\beta}q_t(\mathbf{z}_t,\bar{\tau}_n,\bar{\beta}_n) = 0$,

(b) $E\nabla_{\beta\tau}\underline{S}_n = 0$,

(c) $\liminf_{n\to\infty}\lambda_{\min}(E\nabla_{\beta\beta}\underline{S}_n) > 0$, *and*

(d) $\liminf_{n\to\infty}\lambda_{\min}(nE(\nabla_{\beta'}\underline{S}_n\nabla_{\beta}\underline{S}_n)) > 0$.

Assumption 11.5*. *Let $\nabla_{\beta}q_t(\mathbf{z}_t,\bar{\tau}_n,\bar{\beta}_n)$ be near epoch dependent of size $-2(r-1)/(r-2)$ on the basis process $(\mathbf{e}_t)$, which is assumed to be α-mixing with mixing coefficients of size $-2r/(r-2)$ or ϕ-mixing with mixing coefficients of size $-r/(r-1)$, for some $r > 2$. Furthermore, let*

$$\sup_n\sup_t E\left|\nabla_{\beta}q_t(\mathbf{z}_t,\bar{\tau}_n,\bar{\beta}_n)\right|^{2r} < \infty.$$

We also will make use of the following assumption.

Assumption 13.1. *Let*

$$\sup_n n^{-1} \sum_{t=1}^{n} E\left[\sup_{T' \times B'} |\nabla_\beta q_t(\mathbf{z}_t, \tau, \beta)|^2\right] < \infty,$$

$$\sup_n n^{-1} \sum_{t=1}^{n} E\left[\sup_{T' \times B'} |\nabla_{\beta\tau} q_t(\mathbf{z}_t, \tau, \beta)|^2\right] < \infty,$$

and

$$\sup_n n^{-1} \sum_{t=1}^{n} E\left[\sup_{T' \times B'} |\nabla_{\beta\beta} q_t(\mathbf{z}_t, \tau, \beta)|^2\right] < \infty.$$

(Here $T' \times B'$ is the compact subset employed in Assumption 11.1(d).)

We can now give the following consistency results for the estimators of the variance covariance matrices of least mean distance estimators.[1]

Theorem 13.1. *(a) Let Assumptions 11.1, 11.2*, 11.3, and 11.4 hold, then*

$$\hat{C}_n^{-1} \hat{\Phi}_n \hat{C}_n^{-1\prime} - C_n^{-1} D_n D_n' C_n^{-1\prime} \to 0 \text{ i.p. as } n \to \infty.$$

(b) Let Assumptions 11.1, 11.2, 11.3, 11.5*, and 13.1 hold. Assume furthermore that the weights $w(j,n)$ are bounded, $\lim_{n\to\infty} w(j,n) = 1$ for all $j \geq 0$, and that $\ell_n = o(n^{1/3})$, then*

$$\hat{C}_n^{-1} \hat{\Psi}_n \hat{C}_n^{-1\prime} - C_n^{-1} D_n D_n' C_n^{-1\prime} \to 0 \text{ i.p. as } n \to \infty.$$

The assumptions in part (a) of Theorem 13.1, i.e., for the estimation of the variance covariance matrix if $\mathbf{v}_{t,n} = \nabla_{\beta'} q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n) = \nabla_{\beta'} q_t(\mathbf{z}_t, \bar{\tau}, \bar{\beta})$ is a martingale difference, are only slightly stronger than the assumptions for the corresponding asymptotic normality result in Theorem 11.2(a). The only additional requirement is that also the elements of $\nabla_{\beta'} q_t \nabla_\beta q_t$ satisfy

[1]The assumptions maintained by part (a) of Theorem 13.1 are clearly stronger than what would be needed for establishing $\hat{C}_n^{-1} \hat{\Phi}_n \hat{C}_n^{-1\prime} - C_n^{-1} D_n D_n' C_n^{-1\prime} \to 0$ i.p., since this result could essentially be derived from uniform convergence results, consistency of the estimators $(\hat{\tau}_n, \hat{\beta}_n)$, and uncorrelatedness of $\mathbf{v}_t(\bar{\tau}_n, \bar{\beta}_n)$ alone, as discussed in Section 12.1. The assumptions of the theorem are, however, such that the asymptotic normality result of Theorem 11.2(a) applies, thus ensuring that $C_n^{-1} D_n D_n' C_n^{-1\prime}$ actually is the asymptotic variance covariance matrix of $\hat{\beta}_n$. We note furthermore that Theorem 13.1 also covers the case where no nuisance parameter τ is present, cf. the similar discussion in Chapters 8 and 11.

the conditions of Assumption 11.2. Of course, in cases where $C_n = D_n D_n'$ we only need to estimate either C_n or $D_n D_n'$. Such a case arises, e.g., in the context of maximum likelihood estimation of a correctly specified model. If $\hat{C}_n^{-1}$ is then used as an estimator for the variance covariance matrix, consistency follows already as a by-product of the asymptotic normality result in Theorem 11.2(a) and hence no further additional conditions are necessary.[2]

For part (b) of Theorem 13.1, i.e., for the case where the scores $\mathbf{v}_{t,n} = \nabla_{\beta'} q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n)$ may be autocorrelated, we have to strengthen the near epoch dependence condition and the moment conditions (Assumptions 11.5* and 13.1) as compared with the corresponding asymptotic normality result in Theorem 11.2(a). Furthermore, and more importantly, we have to strengthen the condition

$$E\nabla_\beta \underline{S}_n = n^{-1} \sum_{t=1}^{n} E\nabla_\beta q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n) = 0$$

for all n, maintained in Assumption 11.3(a), to the condition that

$$E\nabla_\beta q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n) = 0$$

for all t, maintained in Assumption 11.3*(a). Recall from the discussion after Theorem 11.2 that the condition $E\nabla_\beta \underline{S}_n = 0$ is usually automatically satisfied as a consequence of the definition of $\bar{\beta}_n$ as a minimizer of $ES_n(\omega, \bar{\tau}_n, \beta)$ (which in turn is usually a consequence of the definition of $\hat{\beta}_n$). Assumption 11.3*(a) is in general stronger. However, Assumption 11.3*(a) coincides with Assumption 11.3(a) if $(\bar{\tau}_n, \bar{\beta}_n) \equiv (\bar{\tau}, \bar{\beta})$ for all n, which is, e.g., usually the case if $(\bar{\tau}, \bar{\beta})$ represents the "true" parameter in a correctly specified model. Assumption 11.3*(a) also reduces to Assumption 11.3(a) if $E\nabla_\beta q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n)$ does not depend on t.

---

[2]This, of course, does not imply that $\hat{C}_n^{-1}$ is always the preferred estimator. Calzolari and Panattoni (1988) and Prucha (1984) report on Monte Carlo simulations that analyze the small sample properties of alternative variance covariance estimators in the context of maximum likelihood estimation of a simultaneous equation model with independent errors. They find that the variance covariance matrix estimates may differ considerably in small samples depending on which form of the variance covariance matrix estimator is employed.

13.2 Estimation of the Variance Covariance Matrix of Generalized Method of Moments Estimators

The asymptotic variance covariance matrix of the generalized method of moments estimator in Theorem 11.5(a) is of the form $C_n^{-1} D_n D_n' C_n^{-1\prime}$ where

$$C_n = E\nabla_{\beta'}\underline{S}_n \left[\nabla_{cc}\vartheta_n(0, \bar{\tau}_n, \bar{\beta}_n)\right] E\nabla_{\beta}\underline{S}_n$$

and

$$D_n D_n' = E\nabla_{\beta'}\underline{S}_n \left[\nabla_{cc}\vartheta_n(0, \bar{\tau}_n, \bar{\beta}_n)\right] \left(nE\underline{S}_n\underline{S}_n'\right) \left[\nabla_{cc}\vartheta_n(0, \bar{\tau}_n, \bar{\beta}_n)\right] E\nabla_{\beta}\underline{S}_n.$$

Consistent estimation of the asymptotic variance covariance matrix hence can be reduced to consistent estimation of C_n and $D_n D_n'$. In this section let $\mathbf{v}_t(\tau, \beta) = q_t(\mathbf{z}_t, \tau, \beta)$ and hence $\hat{\mathbf{v}}_{t,n} = q_t(\mathbf{z}_t, \hat{\tau}_n, \hat{\beta}_n)$. Consider the following estimators:

$$\hat{C}_n = \nabla_{\beta'}\hat{S}_n \nabla_{cc}\hat{\vartheta}_n \nabla_{\beta}\hat{S}_n,$$

where

$$\begin{aligned}
\nabla_{\beta'}\hat{S}_n &= n^{-1}\sum_{t=1}^{n}\nabla_{\beta'}q_t(\mathbf{z}_t, \hat{\tau}_n, \hat{\beta}_n), \\
\nabla_{cc}\hat{\vartheta}_n &= \nabla_{cc}\vartheta_n(0, \hat{\tau}_n, \hat{\beta}_n),
\end{aligned}$$

and

$$\begin{aligned}
\hat{\Phi}_n &= n^{-1}\sum_{t=1}^{n}\hat{\mathbf{v}}_{t,n}\hat{\mathbf{v}}_{t,n}' = n^{-1}\sum_{t=1}^{n} q_t(\mathbf{z}_t, \hat{\tau}_n, \hat{\beta}_n) q_t(\mathbf{z}_t, \hat{\tau}_n, \hat{\beta}_n)', \\
\hat{\Psi}_n &= w(0,n)\left[n^{-1}\sum_{t=1}^{n}\hat{\mathbf{v}}_{t,n}\hat{\mathbf{v}}_{t,n}'\right] \\
&\quad + \sum_{j=1}^{n-1} w(j,n)\left[n^{-1}\sum_{t=1}^{n-j}\left[\hat{\mathbf{v}}_{t,n}\hat{\mathbf{v}}_{t+j,n}' + \hat{\mathbf{v}}_{t+j,n}\hat{\mathbf{v}}_{t,n}'\right]\right] \\
&= w(0,n)\left[n^{-1}\sum_{t=1}^{n} q_t(\mathbf{z}_t, \hat{\tau}_n, \hat{\beta}_n) q_t(\mathbf{z}_t, \hat{\tau}_n, \hat{\beta}_n)'\right] \\
&\quad + \sum_{j=1}^{n-1} w(j,n)\left[n^{-1}\sum_{t=1}^{n-j}\left[q_t(\mathbf{z}_t, \hat{\tau}_n, \hat{\beta}_n) q_{t+j}(\mathbf{z}_{t+j}, \hat{\tau}_n, \hat{\beta}_n)'\right.\right. \\
&\quad \left.\left. + q_{t+j}(\mathbf{z}_{t+j}, \hat{\tau}_n, \hat{\beta}_n) q_t(\mathbf{z}_t, \hat{\tau}_n, \hat{\beta}_n)'\right]\right],
\end{aligned}$$

where $\hat{\Psi}_n$ is of the form given in (12.2). The estimator $\hat{\Phi}_n$ is appropriate for estimating $nE\underline{S}_n\underline{S}_n'$ in the case where $\mathbf{v}_{t,n} = q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n)$ has mean zero and is uncorrelated (e.g., a martingale difference), whereas $\hat{\Psi}_n$ has to be used if $\mathbf{v}_{t,n}$ is autocorrelated. The subsequent assumptions strengthen Assumptions 11.6, 11.7 and 11.9.

Assumption 11.6*. *(a) Assumption 11.6(a) holds.*
(b) Assumption 11.6(b) holds.
(c) The family $\{f_t : t \in \mathbf{N}\}$ is equicontinuous on $Z \times T' \times B'$ and

$$\sup_n n^{-1} \sum_{t=1}^{n} E\left[\sup_{T' \times B'} |f_t(\mathbf{z}_t, \tau, \beta)|^{1+\gamma}\right] < \infty$$

for some $\gamma > 0$, where f_t denotes the restriction to $Z \times T' \times B'$ of any of the components of q_t, $\nabla_\beta q_t$, $\nabla_\tau q_t$, $\nabla_{\beta\tau} q_t$, $\nabla_{\beta\beta} q_t$, or $q_t q_t'$. (Here $T' \times B'$ is the compact subset employed in Assumption 11.1(d).)

Assumption 11.7*. *(a) $Eq_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n) = 0$,*
(b) $E\nabla_\tau \underline{S}_n = 0$,
(c) $\liminf_{n\to\infty} \lambda_{\min}(\nabla_{cc}\vartheta_n(0, \bar{\tau}_n, \bar{\beta}_n)) > 0$,
(d) $\liminf_{n\to\infty} \lambda_{\min}(E\nabla_{\beta'} \underline{S}_n E\nabla_\beta \underline{S}_n) > 0$,
(e) $\liminf_{n\to\infty} \lambda_{\min}(nE\underline{S}_n \underline{S}_n') > 0$.

Assumption 11.9*. *Let $q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n)$ be near epoch dependent of size $-2(r-1)/(r-2)$ on the basis process $(\mathbf{e}_t)$, which is assumed to be α-mixing with mixing coefficients of size $-2r/(r-2)$ or ϕ-mixing with mixing coefficients of size $-r/(r-1)$, for some $r > 2$. Furthermore, let*

$$\sup_n \sup_t E\left|q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n)\right|^{2r} < \infty.$$

We will also make use of the following assumption.

Assumption 13.2. *Let*

$$\sup_n n^{-1} \sum_{t=1}^{n} E\left[\sup_{T' \times B'} |q_t(\mathbf{z}_t, \tau, \beta)|^2\right] < \infty,$$

$$\sup_n n^{-1} \sum_{t=1}^{n} E\left[\sup_{T' \times B'} |\nabla_\beta q_t(\mathbf{z}_t, \tau, \beta)|^2\right] < \infty,$$

and

$$\sup_n n^{-1} \sum_{t=1}^{n} E\left[\sup_{T' \times B'} |\nabla_\tau q_t(\mathbf{z}_t, \tau, \beta)|^2\right] < \infty.$$

(Here $T' \times B'$ is the compact subset employed in Assumption 11.1(d).)

We can now give the following consistency results for the estimators of the variance covariance matrices of generalized method of moments estimators.[3]

Theorem 13.2. *(a) Let Assumptions 11.1, 11.6*, 11.7, and 11.8 hold, then*

$$\hat{C}_n^{-1}\nabla_{\beta'}\hat{S}_n\nabla_{cc}\hat{\vartheta}_n\hat{\Phi}_n\nabla_{cc}\hat{\vartheta}_n\nabla_{\beta}\hat{S}_n\hat{C}_n^{-1\prime} - C_n^{-1}D_nD_n'C_n^{-1\prime} \to 0 \ i.p.$$

as $n \to \infty$.

(b) Let Assumptions 11.1, 11.6, 11.7, 11.9*, and 13.2 hold. Assume furthermore that the weights $w(j,n)$ are bounded, $\lim_{n\to\infty} w(j,n) = 1$ for all $j \geq 0$, and that $\ell_n = o(n^{1/3})$, then*

$$\hat{C}_n^{-1}\nabla_{\beta'}\hat{S}_n\nabla_{cc}\hat{\vartheta}_n\hat{\Psi}_n\nabla_{cc}\hat{\vartheta}_n\nabla_{\beta}\hat{S}_n\hat{C}_n^{-1\prime} - C_n^{-1}D_nD_n'C_n^{-1\prime} \to 0 \ i.p.$$

as $n \to \infty$.

The assumptions in part (a) of Theorem 13.2, i.e., for the estimation of the variance covariance matrix if $\mathbf{v}_{t,n} = q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n) = q_t(\mathbf{z}_t, \bar{\tau}, \bar{\beta})$ is a martingale difference, are only slightly stronger than the assumptions for the corresponding asymptotic normality result in Theorem 11.5(a). The only additional requirement is that also the elements of $q_t q_t'$ satisfy the conditions of Assumption 11.6(c). For part (b) of Theorem 13.2, i.e., for the case where $\mathbf{v}_{t,n} = q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n)$ may be autocorrelated, we have to strengthen the near epoch dependence condition and the moment conditions (Assumptions 11.9* and 13.2) as compared with the corresponding asymptotic normality result in Theorem 11.5(a). Furthermore we have to strengthen the condition

$$E\underline{S}_n = n^{-1}\sum_{t=1}^{n} Eq_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n) = 0$$

for all n, maintained in Assumption 11.7(a), to the condition that

$$Eq_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n) = 0$$

for all t, maintained in Assumption 11.7*(a). Recall from the discussion after Theorem 11.5 that the condition $E\underline{S}_n = 0$ is typically only satisfied for correctly specified models. Assumption 11.7*(a) will coincide with Assumption 11.7(a) if $(\bar{\tau}_n, \bar{\beta}_n) \equiv (\bar{\tau}, \bar{\beta})$, which is, e.g., usually the case under correct specification. Assumption 11.7*(a) also reduces to Assumption 11.7(a) if $Eq_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n)$ does not depend on t.

---

[3] An analogous remark as in Footnote 1 also applies here.

14

QUASI MAXIMUM LIKELIHOOD ESTIMATION OF DYNAMIC NONLINEAR SIMULTANEOUS SYSTEMS

In this chapter we derive consistency and asymptotic normality results for the quasi normal full information maximum likelihood (quasi-NFIML) estimator of the parameters of a dynamic implicit nonlinear simultaneous equation system. The qualifier "quasi" is used to indicate that the objective function employed to define the estimator may differ from the actual likelihood of the data. There are two major purposes for this chapter. One purpose is to further illustrate how the tools reviewed and developed in the previous chapters can be applied to this more concrete problem. The second purpose is to introduce new results concerning the asymptotic properties of the quasi-NFIML estimator of a *dynamic* implicit nonlinear simultaneous equation system. These results complement and extend those given in Amemiya (1977, 1982) and Gallant and Holly (1980) for the NFIML estimator of *static* systems. Apart from permitting the system to be dynamic we also allow for temporal heterogeneity of the data generating process.

We note that while the discussion in this chapter illustrates the usefulness of the tools reviewed and developed in this book, it also shows that applying the basic theorems regarding consistency and asymptotic normality of M-estimators given in the previous chapters to a concrete problem can still require considerable effort.

We consider in the following the estimation of the parameters of the following dynamic implicit nonlinear simultaneous equation system:

$$f_t(\mathbf{y}_t, \ldots, \mathbf{y}_{t-l}, \mathbf{x}_t, \alpha) = \epsilon_t, \qquad t \in \mathbf{N}, \tag{14.1}$$

where the processes of the endogenous variables $\mathbf{y}_t$, exogenous variables $\mathbf{x}_t$, and disturbances ϵ_t take their values in, respectively, $\mathbf{R}^{p_y}$, $\mathbf{R}^{p_x}$, and $\mathbf{R}^{p_e}$ with $p_y = p_e$.[1] The parameter vector α is an element of a Borel set $A \subseteq \mathbf{R}^{p_a}$ and $f_t : Z \times A \to \mathbf{R}^{p_e}$ is Borel measurable where $Z = \mathbf{R}^{(l+1)p_y+p_x}$.

[1]In referring to the variables $\mathbf{x}_t$ as exogenous we only want to indicate that in the context of system (14.1) those variables are viewed as input variables. No further assumptions about the stochastic nature of $\mathbf{x}_t$ should be associated with

Furthermore, f_t has a nonsingular matrix of partial derivatives $\nabla_y f_t$ on $Z \times A$. We now define, as usual, the objective function of the quasi-NFIML estimator as (cf. Example 2 in Chapter 2):

$$R_n(\omega,\beta) = Q_n(\mathbf{z}_1,\ldots,\mathbf{z}_n,\beta) = n^{-1}\sum_{t=1}^{n} q_t(\mathbf{z}_t,\beta), \tag{14.2}$$

with

$$q_t(\mathbf{z}_t,\beta) = -\ln|\det(\nabla_y f_t)| + (1/2)\ln\det(\Sigma) + (1/2)f_t'\Sigma^{-1}f_t,$$

where f_t and $\nabla_y f_t$ are evaluated at $(\mathbf{z}_t,\alpha)$ with $\mathbf{z}_t = (\mathbf{y}_t',\ldots,\mathbf{y}_{t-l}',\mathbf{x}_t')'$. The parameter space for Σ is a subset of the set of all symmetric positive definite $p_e \times p_e$ matrices. It is convenient to describe the matrix Σ in terms of σ, the vector of its diagonal and upper diagonal elements. (In order to denote the dependence of Σ on σ we shall also write sometimes $\Sigma(\sigma)$.) The resulting parameter space for σ is denoted by $S \subseteq \mathbf{R}^{p_e(p_e+1)/2}$. The vector β is then composed of the elements of α and σ. The parameter space B, on which $q_t(z,.)$ is defined, is taken as $A \times S$.

We note that the setup and assumptions postulated so far in this chapter will be assumed to hold throughout the entire chapter and the corresponding appendix.

In the following we first introduce a general consistency result for the quasi-NFIML estimator allowing for misspecification. The sources for misspecification in the sense that (14.2) is only a quasi log-likelihood and not the true (conditional) log-likelihood can be manifold. For example, one source might be that the system (14.1) is misspecified. Another source may be that, while (14.1) is correctly specified, the actual disturbance distribution is not normal. Yet another source might be that (14.2) only represents a partial and not the true (conditional) log-likelihood, as will be discussed in more detail later.

Second, we will then consider in more detail the special case where the system of simultaneous equations (14.1) is correctly specified. For this case we will derive consistency and asymptotic normality results for the quasi-NFIML estimator, as well as results concerning the consistent estimation of the asymptotic variance covariance matrix of the quasi-NFIML estimator. As a byproduct we also introduce results concerning the L_p-approximability of the data generating process.

---

this terminology.

14.1 A General Consistency Result for Quasi-NFIML Estimators

In this section we utilize Theorem 7.1 to introduce a consistency result for the quasi-NFIML estimator of a dynamic implicit nonlinear simultaneous equation system. For that result we shall not make the assumption that $\mathbf{z}_t$ follows (14.1), i.e., we shall allow for the system (14.1) to be misspecified.[2] Consider the following assumptions:

Assumption 14.1. *(a) The parameter space* $A \subseteq \mathbf{R}^{p_a}$ *is compact.*

(b) The parameter space $S \subseteq \mathbf{R}^{p_e(p_e+1)/2}$ *is compact.* [3]

(c) For some $\gamma > 0$

$$\sup_n n^{-1} \sum_{t=1}^{n} E\left[\sup_A |f_t(\mathbf{z}_t, \alpha)|^{2+2\gamma}\right] < \infty,$$

$$\sup_n n^{-1} \sum_{t=1}^{n} E\left[\sup_A |\ln|\det(\nabla_y f_t(\mathbf{z}_t, \alpha))||^{1+\gamma}\right] < \infty.$$

(d) For some $\gamma > 0$

$$\sup_n n^{-1} \sum_{t=1}^{n} E|\mathbf{y}_t|^{\gamma} < \infty.$$

(e) For some $\gamma > 0$

$$\sup_n n^{-1} \sum_{t=1}^{n} E|\mathbf{x}_t|^{\gamma} < \infty.$$

(f) $\{f_t : t \in \mathbf{N}\}$ *and* $\{\ln|\det(\nabla_y f_t)| : t \in \mathbf{N}\}$ *are equicontinuous on* $Z \times A$.

(g) $\sup_t |f_t(z, \alpha)| < \infty$ *for all* $(z, \alpha) \in Z \times A$.

Assumption 14.2. *The process* $(\mathbf{y}_t', \mathbf{x}_t')'$ *is* L_0*-approximable by some* α*-mixing basis process* $(\mathbf{e}_t)_{t\in\mathbf{Z}}$.

The following theorem now follows from Theorem 7.1. To prove the theorem we verify that Assumptions 14.1 and 14.2 imply Assumptions 7.1 and 7.2, cf. Appendix K.

---

[2] As a consequence, for that result it is not necessary to assume that a well-defined reduced form exists.

[3] Since $\Sigma(\sigma)$ is positive definite for any $\sigma \in S$, compactness of S is equivalent to S being a closed subset of a set of the form $\{\sigma \in \mathbf{R}^{p_e(p_e+1)/2} : c_1 \leq \lambda_{\min}(\Sigma(\sigma)), \lambda_{\max}(\Sigma(\sigma)) \leq c_2\}$ with $0 < c_1 \leq c_2 < \infty$.

Theorem 14.1. [4] *Let*

$$R_n(\omega,\beta) = n^{-1}\sum_{t=1}^{n} q_t(\mathbf{z}_t,\beta)$$

be the objective function of the quasi-NFIML estimator defined in (14.2) and

$$\bar{R}_n(\beta) = n^{-1}\sum_{t=1}^{n} Eq_t(\mathbf{z}_t,\beta).$$

Suppose Assumptions 14.1 and 14.2 hold, then

$$\sup_B \left|R_n(\omega,\beta) - \bar{R}_n(\beta)\right| \to 0 \text{ i.p. as } n \to \infty,$$

and $\{\bar{R}_n : n \in \mathbf{N}\}$ is equicontinuous on B. Furthermore, let $\bar{\beta}_n$ be an identifiably unique sequence of minimizers of $\bar{R}_n(\beta)$ and let $\hat{\beta}_n$ be any sequence of quasi-NFIML estimators, i.e.,

$$R_n(\omega,\hat{\beta}_n) = \inf_B R_n(\omega,\beta) \tag{14.3}$$

holds for all $\omega \in \Omega$ and $n \in \mathbf{N}$.[5] *Then $\hat{\beta}_n$ is consistent for $\bar{\beta}_n$, i.e., $|\hat{\beta}_n - \bar{\beta}_n| \to 0$ i.p. as $n \to \infty$.*

We note again that the above theorem does not assume that $\mathbf{y}_t$ is actually generated by (14.1). Thus the notion of a true parameter vector is vacuous in this context. Therefore, even if the minimizers $\bar{\beta}_n$ in Theorem 14.1 are independent of n, i.e., $\bar{\beta}_n \equiv \bar{\beta}$, the parameter vector $\bar{\beta}$ may have no interpretation other than that of a minimizer of $\bar{R}_n(\beta)$. We note further that even in the case where $\mathbf{y}_t$ is generated by (14.1) (with, say, i.i.d. disturbances), $\bar{\beta}_n$ need not coincide with the true parameter vector β_0. Hence Theorem 14.1 does not imply that $\hat{\beta}_n$ converges to the true parameter vector β_0. That is, as already discussed in Amemiya (1977) and Phillips (1982) for the special case of a static nonlinear simultaneous equation model, in general the quasi-NFIML estimator does not converge to β_0 even if (14.1) holds, unless certain further conditions are satisfied; cf. the discussion in the following section.

Theorem 14.1 also holds for approximate M-estimators, cf. Section 4.4. Furthermore, the assumption that B is equal to $A \times S$ can be relaxed. Clearly, the theorem also holds if B is only a compact subset of $A \times S$, and thus a corresponding version of the theorem also covers the case of cross restrictions between α and σ.

---

[4]We note that here and in the following theorems the existence of $\bar{R}_n$ is ensured by the maintained assumptions.

[5]The existence of measurable $\hat{\beta}_n$ satisfying (14.3) is ensured by Lemma 3.4.

In the following discussion we show how in the context of Theorem 14.1 the compactness assumption for S can be removed. Loosely speaking, the basic strategy is to demonstrate that the functions R_n and $\bar{R}_n$ become large outside some compact subset of the parameter space, and then to apply Theorem 14.1 to the estimation problem restricted to the compact subset; cf. also Section 4.3. This strategy is formally implemented in Lemma K1 in Appendix K. For the ensuing discussion, which is based on this lemma, we assume that Assumptions 14.1(a),(c)-(g) and 14.2 are in force, and that S now corresponds to the set of all symmetric positive definite $p_e \times p_e$ matrices.[6] We also maintain additionally that

$$\inf_n \inf_A \lambda_{\min}\left(n^{-1}\sum_{t=1}^{n} Ef_t f_t'\right) > 0.$$

Since, in contrast to Theorem 14.1, the parameter space $B = A \times S$ is now no longer compact, the existence of a minimizer $\bar{\beta}_n$ of $\bar{R}_n(\beta)$ follows no longer trivially from the continuity of $\bar{R}_n$. The existence of such minimizers can, however, be proven, as is demonstrated in Lemma K1(b) in Appendix K. As in Theorem 14.1 we assume that those minimizers are identifiably unique. Similarly, the existence of minimizers of $R_n(\omega,\beta)$ is now no longer ensured. We hence define the quasi-NFIML estimator $\hat{\beta}_n$ to be the minimizer of $R_n(\omega,\beta)$, if a minimizer exists, and assign to it some fixed (arbitrary) element of B otherwise.[7] Lemma K1 now shows that there exists a compact subset, say S_*, of S and a sequence of sets $\Omega_n \in \mathfrak{A}$ with $P(\Omega_n) \to 1$ as $n \to \infty$ such that (i) $\bar{\beta}_n$ belongs to the compact subset $B_* = A \times S_*$ of $B = A \times S$, (ii) $R_n(\omega,\beta)$ attains its minimum on $B = A \times S$ whenever $\omega \in \Omega_n$, and all the minimizers of $R_n(\omega,\beta)$ over $B = A \times S$ belong to $B_* = A \times S_*$ whenever $\omega \in \Omega_n$. Thus, while for reasons of mathematical definiteness it was necessary to define $\hat{\beta}_n(\omega)$ also for $\omega \in \Omega$ for which a minimizer of $R_n(\omega,\beta)$ does not exist, the lemma shows that $\hat{\beta}_n(\omega)$ is in fact a minimizer of $R_n(\omega,\beta)$ over $B = A \times S$, except on a sequence of ω-sets whose probability tends to zero. It also shows that the $\hat{\beta}_n(\omega)$'s take their values in the compact set $B_* = A \times S_*$, except on a sequence of ω-sets whose probability tends to zero. Theorem 14.1 applied to the restricted compact parameter space $B_* = A \times S_*$ shows the consistency of the restricted quasi-NFIML estimators, i.e., of minimizers of $R_n(\omega,\beta)$ over $B_* = A \times S_*$.[8] Since $\hat{\beta}_n$ differs from such a restricted quasi-NFIML estimator only on ω-sets whose probability tends to zero, this also implies

[6] The following discussion also holds if S corresponds to a relatively closed subset of the set of all symmetric positive definite $p_e \times p_e$ matrices.

[7] The existence of a measurable $\hat{\beta}_n$ satisfying the above definition can be deduced from Corollary 1 in Brown and Purves (1973).

[8] Note that the assumption of identifiable uniqueness of $\bar{\beta}_n$ w.r.t. B clearly implies identifiable uniqueness of $\bar{\beta}_n$ w.r.t. the restricted parameter space B_*.

the consistency of $\hat{\beta}_n$ for $\bar{\beta}_n$.

14.2 Asymptotic Results for the Quasi-NFIML Estimator in Case of a Correctly Specified System

In the previous section we allowed for misspecification of the system (14.1). In this section we will discuss in more detail the case where $\mathbf{y}_t$ is actually generated by (14.1).[9]

14.2.1 Sufficient Conditions for L_p-Approximability of the Data Generating Process

Theorem 14.1 postulates in Assumption 14.2 a form of weak dependence of the data generating process. In particular, this assumption maintains that the process $(\mathbf{y}_t', \mathbf{x}_t')'$ is L_0-approximable by an α-mixing basis process. This assumption may be considered a high-level condition. Given that in Theorem 14.1 the process $\mathbf{y}_t$ is not assumed to be generated by (14.1), thus allowing for misspecification, it seemed necessary to impose the weak dependence assumption directly on $(\mathbf{y}_t', \mathbf{x}_t')'$, as we then typically have no further information on the mechanism that generates the data. However, if (14.1) is correctly specified in the sense that $\mathbf{y}_t$ is actually generated by this system, then sufficient conditions for Assumption 14.2 can be given in terms of the input processes $\mathbf{x}_t$ and ϵ_t. More specifically, consider the following assumptions.

Assumption 14.3. *For each* $(y_{-1}', \ldots, y_{-l}', x', e')' \in \mathbf{R}^{lp_y+p_x+p_e}$, *each* $\alpha \in A$, *and each* $t \geq 1$, *the equation*

$$f_t(y, y_{-1}, \ldots, y_{-l}, x, \alpha) = e$$

has a unique solution, say

$$y = g_t(y_{-1}, \ldots, y_{-l}, x, e, \alpha).$$

Furthermore the function $g_t : \mathbf{R}^{lp_y+p_x+p_e} \times A \to \mathbf{R}^{p_y}$ *is Borel measurable.*

The above assumption implies that the system (14.1) has a well-defined reduced form. (Recall that, although not expressed in the notation, we have assumed that $p_y = p_e$.)

[9] Of course, the statement that $\mathbf{y}_t$ is generated by (14.1) is vacuous unless some further restrictions (e.g., independence or weak dependence conditions) are placed on the disturbance process.

Assumption 14.4. *The process of the endogenous variables* $\mathbf{y}_t$ *is generated by (14.1) for some* $\alpha = \alpha_0 \in A$, *starting from the initial random variables* $\mathbf{y}_0, \ldots, \mathbf{y}_{1-l}$.

Under Assumptions 14.3 and 14.4 the process of the endogenous variables can be described in terms of the reduced form by

$$\mathbf{y}_t = g_t\left(\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-l}, \mathbf{x}_t, \epsilon_t, \alpha_0\right), \qquad t \in \mathbf{N}.$$

For the purpose of obtaining sufficient conditions for Assumption 14.2 it proves useful to rewrite the dynamic system given by the reduced form equivalently as the following dynamic system of order one:

$$\mathbf{v}_t = \phi_t\left(\mathbf{v}_{t-1}, \mathbf{w}_t\right), \qquad t \in \mathbf{N},$$

with $\mathbf{v}_t = (\mathbf{y}_t', \ldots, \mathbf{y}_{t-l+1}')'$, $\mathbf{w}_t = (\mathbf{x}_t', \epsilon_t')'$, and where the functions $\phi_t : \mathbf{R}^{lp_y+p_x+p_e} \rightarrow \mathbf{R}^{lp_y}$ are defined as

$$\phi_t(v, w) = \begin{bmatrix} g_t(y_{-1}, \ldots, y_{-l}, x, e, \alpha_0) \\ y_{-1} \\ \cdot \\ \cdot \\ \cdot \\ y_{-l+1} \end{bmatrix}, \qquad t \in \mathbf{N},$$

with $v = (y_{-1}', \ldots, y_{-l}')' \in \mathbf{R}^{lp_y}$ and $w = (x', e')' \in \mathbf{R}^{p_x+p_e}$. Note that ϕ_t depends also on α_0, although we do not express this in the notation. Additionally, we define for $t \in \mathbf{N}$ and $k \in \mathbf{N}$ the functions $\phi_t^{(k+1)}$, representing iterations of the dynamic system, in terms of the following recursions:

$$\phi_t^{(k+1)}(v, w_1, \ldots, w_{k+1}) = \phi_{t+k}(\phi_t^{(k)}(v, w_1, \ldots, w_k), w_{k+1}),$$

for all $v \in \mathbf{R}^{lp_y}$ and $w_i \in \mathbf{R}^{p_x+p_e}$, $i = 1, \ldots, k+1$, and where $\phi_t^{(1)} = \phi_t$. Now consider the following assumptions. For an interpretation of and motivation for the conditions maintained in these assumptions see the discussion surrounding Theorem 6.12 in Chapter 6.

Assumption 14.5. *The reduced form satisfies the following Lipschitz conditions with* $V = \mathbf{R}^{lp_y}$ *and* $W = \mathbf{R}^{p_x+p_e}$*:*

(a) There exist constants c_1 *and* c_2 *with* $0 \leq c_1 < \infty$ *and* $0 \leq c_2 < \infty$ *such that for all* $(v, v^\bullet) \in V \times V$, $(w, w^\bullet) \in W \times W$, *and* $t \in \mathbf{N}$

$$\left|g_t(v, w, \alpha_0) - g_t(v^\bullet, w^\bullet, \alpha_0)\right| \leq c_1 \left|v - v^\bullet\right| + c_2 \left|w - w^\bullet\right|,$$

i.e., $g_t(.,., \alpha_0)$ *is globally Lipschitz.*

(b) There exists an integer $k^* \geq 1$ and constants d_1 and d_2 with $0 \leq d_1 < 1$ and $0 \leq d_2 < \infty$ such that

$$\left|\phi_t^{(k^*)}(v, w_1, \ldots, w_{k^*}) - \phi_t^{(k^*)}(v^\bullet, w_1^\bullet, \ldots, w_{k^*}^\bullet)\right|$$
$$\leq d_1 |v - v^\bullet| + d_2 \left\| \begin{bmatrix} w_1 - w_1^\bullet \\ \cdot \\ \cdot \\ \cdot \\ w_{k^*} - w_{k^*}^\bullet \end{bmatrix} \right\|$$

for all $(v, v^\bullet) \in V \times V$, $(w_1, \ldots, w_{k^*}, w_1^\bullet, \ldots, w_{k^*}^\bullet) \in \prod_{i=1}^{2k^*} W$, and $t \in \mathbf{N}$.

Assumption 14.6. *(a)* $\|\mathbf{y}_{-i}\|_1 < \infty$ *for* $i = 0, \ldots, l-1$.
(b) $\sup_{t \geq 1} \|\mathbf{x}_t\|_1 < \infty$.
(c) $\sup_{t \geq 1} \|\epsilon_t\|_1 < \infty$.

We now have the following result concerning the L_0-approximability of the process $(\mathbf{y}_t', \mathbf{x}_t')'$ based on Theorem 6.12.

Lemma 14.2. *Suppose $(\mathbf{x}_t', \epsilon_t')'$ is L_1-approximable by some basis process $(\mathbf{e}_t)$.*[10] *Then under Assumptions 14.3 - 14.6 the process $(\mathbf{y}_t', \mathbf{x}_t')'$ is L_1-approximable, and hence L_0-approximable, with respect to the basis process $(\mathbf{e}_t)$. Furthermore, $\|\mathbf{y}_t\|_1 < \infty$ for $t \geq 1$.*

Assumptions 14.3 - 14.6 thus provide sufficient conditions for Assumption 14.2 if the basis process $\mathbf{e}_t$ is α-mixing. Of course, if $(\mathbf{x}_t', \epsilon_t')'$ is α-mixing, then $(\mathbf{x}_t', \epsilon_t')'$ itself (with, say, $(\mathbf{x}_t', \epsilon_t')' = 0$ for $t < 1$) can be taken as the basis process $(\mathbf{e}_t)$.

The proof of Lemma 14.2 is based on Theorem 6.12. The Lipschitz condition in Assumption 14.5(a) together with Assumption 14.6(a) is used to establish that $\|\mathbf{v}_i\|_r < \infty$, $i = 0, \ldots, k^* - 1$, an assumption maintained by Theorem 6.12. Assumption 14.5(b) is essentially a stability condition for the system; cf. the discussion after Theorem 6.12. Given the function $g_t(y_{-1}, \ldots, y_{-l}, x, e, \alpha_0)$ is continuously partially differentiable in $(y_{-1}, \ldots, y_{-l}, x, e)$, Assumption 14.5(b) is satisfied if the following holds:

$$\sup\left\{\left|\operatorname{stac}_{i=1}^{p_y}\left[\mathbf{i}_i' \frac{\partial \phi_t^{(k^*)}}{\partial v}\left(v^i, w_1^i, \ldots, w_{k^*}^i\right)\right]\right| : \right.$$
$$\left. v^i \in V,\ w_j^i \in W,\ \ j = 1, \ldots, k^*,\ i = 1, \ldots, p_y,\ t \in \mathbf{N}\right\} < 1$$

---

[10]Recall from Theorem 6.1(b) that L_0-approximability of $(\mathbf{x}_t', \epsilon_t')'$ implies L_1-approximability given $\sup_n n^{-1} \sum_{t=1}^n E\left|(\mathbf{x}_t', \epsilon_t')'\right|^\gamma < \infty$ for some $\gamma > 1$ holds.

and

$$\sup\left\{\left|\frac{\partial\phi_t^{(k^*)}}{\partial w_l}(v, w_1, \ldots, w_{k^*})\right| : v \in V,\ w_j \in W,\ j = 1, \ldots, k^*,\ t \in \mathbf{N}\right\} < \infty, \qquad l = 1, \ldots, k^*,$$

where the stac-operator creates a matrix consisting of the rows shown as the arguments of the operator, and where $\mathbf{i}_i$ denotes the i-th column of the $p_y \times p_y$ identity matrix. The above sufficient conditions are derived by applying the mean value theorem to $\phi_t^{(k^*)}$. The reason that we have to allow for a different argument list for each of the rows of the matrix generated by the stac-operator is again the fact that the mean value theorem has to be applied to each element of $\phi_t^{(k^*)}$ separately; cf. also the discussion after Theorems 6.10 and 6.12.

The functions $\phi_t^{(k^*)}$ are defined recursively in terms of $\phi_{t+k^*-1}, \ldots, \phi_t$. Utilizing this definition we can also give the following sufficient conditions for the above two conditions:

$$\sup\left\{\left|\mathrm{stac}_{i=1}^{p_y}\left[\mathbf{i}_i' \prod_{l=1}^{k^*} \frac{\partial\phi_{t+k^*-l}}{\partial v}\left(\phi_t^{(k^*-l)}(v^i, w_1^i, \ldots, w_{k^*-l}^i), w_{k^*-l+1}^i\right)\right]\right| : v^i \in V,\ w_j^i \in W,\ j = 1, \ldots, k^*,\ i = 1, \ldots, p_y,\ t \in \mathbf{N}\right\} < 1$$

with $\phi_t^{(0)}(v) = v$ and

$$\sup\left\{\left|\frac{\partial\phi_t}{\partial v}(v, w)\right| : v \in V,\ w \in W,\ t \in \mathbf{N}\right\} < \infty,$$
$$\sup\left\{\left|\frac{\partial\phi_t}{\partial w}(v, w)\right| : v \in V,\ w \in W,\ t \in \mathbf{N}\right\} < \infty.$$

The latter two conditions are, of course, equivalent to

$$\sup\left\{\left|\frac{\partial g_t}{\partial v}(v, w, \alpha_0)\right| : v \in V,\ w \in W,\ t \in \mathbf{N}\right\} < \infty,$$
$$\sup\left\{\left|\frac{\partial g_t}{\partial w}(v, w, \alpha_0)\right| : v \in V,\ w \in W,\ t \in \mathbf{N}\right\} < \infty.$$

We note that the last two conditions also imply Assumption 14.5(a), i.e., the global Lipschitz property of $g_t(.,., \alpha_0)$.

Theorem 14.1 also maintains in Assumption 14.1(d) a moment condition on the process $(\mathbf{y}_t)$. The following result states that under Assumptions 14.3 - 14.6 this moment condition is already implied by the other assumptions maintained by Theorem 14.1.

Lemma 14.3. *Given Assumptions 14.1(g) and 14.3 - 14.6 hold, Assumption 14.1(d) is satisfied. (In fact,* $\sup_{t \geq 1} \|\mathbf{y}_t\|_1 < \infty$ *holds.)*

We note that in the important special case where f_t does not depend on t Assumption 14.1(g) is trivially satisfied.

14.2.2 Consistency

We now revisit the consistency of the quasi-NFIML estimator in the case where $\mathbf{y}_t$ is generated by (14.1), i.e., in the case where the system is correctly specified as expressed by Assumptions 14.3 and 14.4.

One consequence of the correct specification of the system is, as was just discussed in the preceding subsection, that we can provide sufficient conditions for the high-level condition of weak dependence formulated as Assumption 14.2. Another important consequence of the correct specification of the system is that in this case we have available the notion of a true parameter vector β_0, which then raises the important question whether or not $\hat{\beta}_n$ is consistent for β_0. Of course, if we simply *postulate* that β_0 – taking the role of $\bar{\beta}_n$ in Theorem 14.1 – is the (identifiably unique) minimizer of $\bar{R}_n$, we obtain consistency of $\hat{\beta}_n$ for β_0 from Theorem 14.1 in a trivial way. We will formulate such a result as the first of the subsequent theorems. In the second of the two subsequent theorems the property that β_0 minimizes $\bar{R}_n$ will be established from sufficient conditions.

Given $\bar{\beta}_n = \beta_0$, we can simplify the identifiable uniqueness condition via Lemma 4.1, since now $\bar{\beta}_n$ does not depend on n. We emphasize, however, that for nonlinear systems of the form (14.1) the assumption of $\bar{R}_n$ being minimized at β_0 is not innocuous, even if the system (14.1) is correctly specified, since this assumption will not hold in general without further restrictions on the class of admissible distributions of the disturbances; cf. the discussion in Amemiya (1977) and Phillips (1982).

In the following theorems we also replace Assumptions 14.1(d) and 14.2 with sufficient conditions by employing Lemmata 14.2 and 14.3. To this effect we introduce the following assumptions.[11]

Assumption 14.6*. *(a)* $\|\mathbf{y}_{-i}\|_1 < \infty$ *for* $i = 0, \ldots, l-1$.
(b) $\sup_{t \geq 1} \|\mathbf{x}_t\|_{1+\delta} < \infty$ *for some* $\delta > 0$.
(c) $\sup_{t \geq 1} \|\epsilon_t\|_{1+\delta} < \infty$ *for some* $\delta > 0$.

[11] Assumption 14.6* strengthens some of the moment conditions in Assumption 14.6 slightly and is introduced mainly for convenience. We note that Theorem 14.4 also holds with Assumption 14.6* replaced by Assumption 14.6, if L_0-approximability of $(\mathbf{x}_t', \epsilon_t')'$ in Assumption 14.7 is strengthened to L_1-approximability.

Assumption 14.7. *The process* $(\mathbf{x}_t', \epsilon_t')'$ *is* L_0*-approximable by some* α*-mixing basis process* $(\mathbf{e}_t)_{t\in\mathbf{Z}}$.

Theorem 14.4. *Let*

$$R_n(\omega,\beta) = n^{-1}\sum_{t=1}^{n} q_t(\mathbf{z}_t,\beta)$$

be the objective function of the quasi-NFIML estimator defined in (14.2) and

$$\bar{R}_n(\beta) = n^{-1}\sum_{t=1}^{n} Eq_t(\mathbf{z}_t,\beta).$$

Suppose Assumptions 14.1(a)-(c),(f),(g), 14.3 - 14.5, 14.6, and 14.7 hold. Then we have*

$$\sup_B \left|R_n(\omega,\beta) - \bar{R}_n(\beta)\right| \to 0 \text{ i.p. as } n \to \infty,$$

and $\{\bar{R}_n : n \in \mathbf{N}\}$ *is equicontinuous on* B.

Furthermore, let $\hat{\beta}_n$ *be any sequence of quasi-NFIML estimators, i.e.,*

$$R_n(\omega,\hat{\beta}_n) = \inf_B R_n(\omega,\beta)$$

holds for all $\omega \in \Omega$ *and* $n \in \mathbf{N}$. *Then:*

(a) If $\bar{R}_n$ *is minimized at* β_0 *for every* $n \in \mathbf{N}$ *and if*

$$\liminf_{n\to\infty} \left[\bar{R}_n(\beta) - \bar{R}_n(\beta_0)\right] > 0 \text{ for all } \beta \neq \beta_0,$$

then $\hat{\beta}_n$ *is consistent for* β_0, *i.e.,* $|\hat{\beta}_n - \beta_0| \to 0$ *i.p. as* $n \to \infty$.

(b) If $\bar{R}\,(\beta) = \lim_{n\to\infty} \bar{R}_n(\beta)$ *exists for all* $\beta \in B$ *and if* β_0 *is the unique minimizer of* $\bar{R}$, *then* $\hat{\beta}_n$ *is consistent for* β_0, *i.e.,* $|\hat{\beta}_n - \beta_0| \to 0$ *i.p. as* $n \to \infty$.

Returning to the discussion before Theorem 14.4 we now consider the following assumption on the distribution of the disturbances. As shown below, this assumption ensures, in particular, that β_0 minimizes $\bar{R}_n$.

Assumption 14.8. *(a) The disturbances* ϵ_t *are* $N(0,\Sigma(\sigma_0))$ *with* $\sigma_0 \in S$.

(b) ϵ_t *is independent of* $\{\epsilon_{t-1},\ldots,\epsilon_1,\mathbf{x}_t,\ldots,\mathbf{x}_1,\mathbf{y}_0,\ldots,\mathbf{y}_{1-l}\}$ *for each* $t \geq 1$.

(c) The Jacobian $\nabla_y f_t(.,\alpha_0)$ *is continuous on* Z *for every* $t \geq 1$.

Assumption 14.8(a),(b) implies that the disturbances ϵ_t are in fact i.i.d. Assumption 14.8(a),(b) is, for example, certainly satisfied if we assume that the disturbance process (ϵ_t) is i.i.d. $N(0, \Sigma(\sigma_0))$ and that this process is independent (jointly) of the process of exogenous variables $(\mathbf{x}_t)$ and the initial random variables $\mathbf{y}_0, \ldots, \mathbf{y}_{1-l}$. Of course, Assumption 14.8(a),(b) is more general. For example, Assumption 14.8(b) is satisfied if $(\mathbf{x}_t)$ is generated by a system of the form

$$\mathbf{x}_t = g_t^x \left(\mathbf{x}_{t-1}, \ldots, \mathbf{x}_{t-m}, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-k}, \mathbf{u}_t\right),$$

and if ϵ_t is independent of

$$\{\epsilon_{t-1}, \ldots, \epsilon_1, \mathbf{u}_t, \ldots, \mathbf{u}_1, \mathbf{y}_0, \ldots, \mathbf{y}_{1-\max(k,l)}, \mathbf{x}_0, \ldots, \mathbf{x}_{1-m}\}.$$

Given Assumptions 14.3 and 14.4 hold, Assumption 14.8(b) is equivalent to the assumption that ϵ_t is independent of

$$\{\mathbf{y}_{t-1}, \ldots, \mathbf{y}_1, \mathbf{y}_0, \ldots, \mathbf{y}_{1-l}, \mathbf{x}_t, \ldots, \mathbf{x}_1\}$$

for each $t \geq 1$. Assumption 14.8(c) is a technical assumption used to ensure that we can obtain the conditional density of $\mathbf{y}_t$ given

$$\{\mathbf{y}_{t-1}, \ldots, \mathbf{y}_1, \mathbf{y}_0, \ldots, \mathbf{y}_{1-l}, \mathbf{x}_t, \ldots, \mathbf{x}_1\}$$

from the (conditional) density of ϵ_t via the transformation technique; cf. Theorem 17.2 in Billingsley (1979).

Lemma 14.5. *Suppose Assumptions 14.3, 14.4 and 14.8 hold.*

(a) Then the conditional density of $\mathbf{y}_t$ *given* $\mathbf{y}_{t-1} = y_{t-1}, \ldots, \mathbf{y}_{1-l} = y_{1-l}, \mathbf{x}_t = x_t, \ldots, \mathbf{x}_1 = x_1$ *is also the conditional density of* $\mathbf{y}_t$ *given* $\mathbf{y}_{t-1} = y_{t-1}, \ldots, \mathbf{y}_{t-l} = y_{t-l}, \mathbf{x}_t = x_t$ *and equals*

$$\begin{aligned} & \pi_t^Y(y_t \mid y_{t-1}, \ldots, y_{1-l}, x_t \ldots, x_1; \beta_0) = \pi_t^Y(y_t \mid y_{t-1}, \ldots, y_{t-l}, x_t; \beta_0) \\ = \; & (2\pi)^{-p_e/2} \left|\det(\nabla_y f_t)\right| \left[\det\left(\Sigma(\sigma_0)\right)\right]^{-1/2} \exp\left(-f_t'\Sigma(\sigma_0)^{-1} f_t/2\right) \end{aligned}$$

where f_t *and* $\nabla_y f_t$ *are evaluated at* (z_t, α_0)*, where* $z_t = (y_t', \ldots, y_{t-l}', x_t')'$ *and* $\beta_0 = (\alpha_0', \sigma_0')'$*. (Consequently,*

$$q_t(z_t, .) = -\ln\left[\pi_t^Y(y_t \mid y_{t-1}, \ldots, y_{t-l}, x_t; .)\right] - (p_e/2)\ln(2\pi).)$$

(b) Furthermore,

$$E\left[q_t(\mathbf{z}_t, \beta) - q_t(\mathbf{z}_t, \beta_0) \mid \mathbf{y}_{t-1} = y_{t-1}, \ldots, \mathbf{y}_{t-l} = y_{t-l}, \mathbf{x}_t = x_t\right],$$

which equals

$$E\left[q_t(\mathbf{z}_t, \beta) - q_t(\mathbf{z}_t, \beta_0) \mid \mathbf{y}_{t-1} = y_{t-1}, \ldots, \mathbf{y}_{1-l} = y_{1-l}, \mathbf{x}_t = x_t, \ldots, \mathbf{x}_1 = x_1\right],$$

is minimized over $B = A \times S$ *at* $\beta = \beta_0 = (\alpha_0', \sigma_0')'$ *for every* $t \geq 1$*. If* $E\left|q_t(\mathbf{z}_t, \beta_0)\right| < \infty$*, then* $Eq_t(\mathbf{z}_t, \beta)$ *is minimized at* $\beta = \beta_0$*. If* $E\left|q_t(\mathbf{z}_t, \beta_0)\right| < \infty$ *for all* $t \geq 1$*, then* $\bar{R}_n(\beta)$ *is minimized at* $\beta = \beta_0$ *for all* $n \geq 1$*.*

The conditional expectations in the above lemma are defined as, respectively,

$$\int [q_t(z_t,\beta) - q_t(z_t,\beta_0)]\, \pi_t^Y(y_t \mid y_{t-1},\dots,y_{t-l},x_t;\beta_0)dy_t$$

and

$$\int [q_t(z_t,\beta) - q_t(z_t,\beta_0)]\, \pi_t^Y(y_t \mid y_{t-1},\dots,y_{1-l},x_t,\dots,x_1;\beta_0)dy_t.$$

The integrals are well-defined, possibly assuming the value $+\infty$, since they both represent the Kullback-Leibler divergence between the densities $\pi_t^Y(y_t \mid y_{t-1},\dots,y_{t-l},x_t;\beta)$ and $\pi_t^Y(y_t \mid y_{t-1},\dots,y_{t-l},x_t;\beta_0)$. Further, given $E\,|q_t(\mathbf{z}_t,\beta_0)| < \infty$ the expectation $Eq_t(\mathbf{z}_t,\beta)$ is well-defined for all $\beta \in B$, possibly assuming the value $+\infty$. (In fact, in the above lemma the condition $E\,|q_t(\mathbf{z}_t,\beta_0)| < \infty$ could be replaced by the weaker condition $Eq_t^-(\mathbf{z}_t,\beta_0) < \infty$, where q_t^- denotes the negative part of q_t.) A similar remark applies to $\bar{R}_n(\beta)$.

Part (b) of the above lemma establishes that β_0 minimizes $\bar{R}_n$, which is a key condition for consistency of $\hat{\beta}_n$ for β_0, and which had been postulated as an assumption in Theorem 14.4 above. Part (a) of the lemma permits, under Assumptions 14.3, 14.4 and 14.8, the interpretation of the objective function (14.1) defining $\hat{\beta}_n$ as a partial log-likelihood in the sense of Cox (1975), but not necessarily as the true log-likelihood of the data (conditional on $\mathbf{x}_1,\dots,\mathbf{x}_n$ and $\mathbf{y}_0,\dots,\mathbf{y}_{1-l}$). To see this observe that the likelihood, say, $\pi^{Y,X}$ of $(\mathbf{y}_n',\dots,\mathbf{y}_{1-l}',\mathbf{x}_n',\dots,\mathbf{x}_1')'$ conditional on the initial values $\mathbf{y}_0 = y_0,\dots,\mathbf{y}_{1-l} = y_{1-l}$ can be factored as follows:[12]

$$\begin{aligned}
&\pi^{Y,X}(y_n,\dots,y_1,x_n,\dots,x_1 \mid y_0,\dots,y_{1-l};\beta) \\
= &\prod_{t=1}^{n} \pi_t^{X,Y}(x_t,y_t \mid y_{t-1},\dots,y_{1-l},x_{t-1},\dots,x_1;\beta) \\
= &\left[\prod_{t=1}^{n} \pi_t^{X}(x_t \mid y_{t-1},\dots,y_{1-l},x_{t-1},\dots,x_1;\beta)\right] \\
&\left[\prod_{t=1}^{n} \pi_t^{Y}(y_t \mid y_{t-1},\dots,y_{1-l},x_t,\dots,x_1;\beta)\right],
\end{aligned}$$

where $\pi_t^{X,Y}, \pi_t^X$ and π_t^Y denote the conditional densities of $(\mathbf{x}_t',\mathbf{y}_t')'$, $\mathbf{x}_t$, and $\mathbf{y}_t$, respectively. The second term on the right hand side of the last equality represents – in the terminology of Cox (1975) – the partial likelihood function. By Lemma 14.5 we now see that the negative logarithm of

[12]For this discussion of the relationship between partial and conditional likelihood only we assume that a joint density of $(\mathbf{y}_n',\dots,\mathbf{y}_{1-l}',\mathbf{x}_n',\dots,\mathbf{x}_1')'$ exists – an assumption that is not maintained otherwise.

the partial likelihood function (normalized by the sample size and up to an irrelevant additive constant) equals our objective function (14.1) defining $\hat{\beta}_n$. We note that in general the partial likelihood does not coincide with the likelihood conditional on $\mathbf{x}_1, \ldots, \mathbf{x}_n$ (and the initial values) since the first term on the right side of the above equation is in general not equal to the marginal density of $\mathbf{x}_1, \ldots, \mathbf{x}_n$ (conditional on the initial values). Since in general the first term will also depend on the parameter vector β, this implies that in general maximizing the partial likelihood

$$\prod_{t=1}^{n} \pi_t^Y \left(y_t \mid y_{t-1}, \ldots, y_{1-l}, x_t, \ldots, x_1; \beta\right)$$

is not equivalent to maximizing the true likelihood

$$\pi^{Y,X} \left(y_n, \ldots, y_1, x_n, \ldots, x_1 \mid y_0, \ldots, y_{1-l}; \beta\right).$$

For further discussions see Wong (1986) and Slud (1992). It seems that this distinction between partial and true likelihood function has not always been observed in the econometrics literature. Of course, e.g., in the special case where $\{\epsilon_1, \ldots, \epsilon_n\}$ is independent (jointly) of $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and the initial values, the partial likelihood function and the likelihood function conditional on $\mathbf{x}_1, \ldots, \mathbf{x}_n$ (and on the initial values) coincide, and then this distinction vanishes.

Combining Theorem 14.4 with Lemma 14.5 yields the following consistency result. We emphasize that, in contrast to Theorem 14.4, in the following theorem the key condition that β_0 minimizes $\bar{R}_n$ over B is not simply postulated, but rather is a consequence of the assumptions.

Theorem 14.6. [13] *Let*

$$R_n(\omega, \beta) = n^{-1} \sum_{t=1}^{n} q_t(\mathbf{z}_t, \beta)$$

be the objective function of the quasi-NFIML estimator defined in (14.2) and

$$\bar{R}_n(\beta) = n^{-1} \sum_{t=1}^{n} E q_t(\mathbf{z}_t, \beta).$$

---

[13] It seems that within the context of a correctly specified system with i.i.d. normally distributed disturbances the estimator $\hat{\beta}_n$ would typically be called the NFIML estimator in the econometrics literature. The preceding discussion implies, however, that under the assumptions of the theorem the estimator $\hat{\beta}_n$ can not necessarily be interpreted as the NFIML estimator, but only as the partial-NFIML estimator. To avoid introducing new terminology, we have chosen to continue using the term quasi-NFIML estimator for $\hat{\beta}_n$ also in this situation.

Suppose Assumptions 14.1(a)-(c),(f),(g), 14.3 - 14.5, 14.6(a),(b), 14.7, and 14.8 hold. Then we have*

$$\sup_B \left| R_n(\omega, \beta) - \bar{R}_n(\beta) \right| \to 0 \text{ i.p. as } n \to \infty,$$

$\{\bar{R}_n : n \in \mathbf{N}\}$ *is equicontinuous on* B, *and* β_0 *minimizes* $\bar{R}_n$ *over* B *for every* $n \in \mathbf{N}$.

Let $\hat{\beta}_n$ *be any sequence of quasi-NFIML estimators, i.e.,*

$$R_n(\omega, \hat{\beta}_n) = \inf_B R_n(\omega, \beta)$$

holds for all $\omega \in \Omega$ *and* $n \in \mathbf{N}$. *Then:*

(a) If

$$\liminf_{n \to \infty} \left[\bar{R}_n(\beta) - \bar{R}_n(\beta_0) \right] > 0 \text{ for all } \beta \neq \beta_0,$$

then $\hat{\beta}_n$ *is consistent for* β_0, *i.e.,* $|\hat{\beta}_n - \beta_0| \to 0$ *i.p. as* $n \to \infty$.

(b) If $\bar{R}(\beta) = \lim_{n \to \infty} \bar{R}_n(\beta)$ *exists for all* $\beta \in B$ *and if* $\bar{R}$ *has a unique minimizer in* B, *then this unique minimizer of* $\bar{R}$ *is* β_0 *and* $\hat{\beta}_n$ *is consistent for* β_0, *i.e.,* $|\hat{\beta}_n - \beta_0| \to 0$ *i.p. as* $n \to \infty$.

If β_0 is identified in the parameter space B, in the sense that $\beta_0 \neq \beta$ implies $R_n(\omega, \beta_0) \neq R_n(\omega, \beta)$ with positive probability, then β_0 is the unique minimizer of $\bar{R}_m(\beta)$ over B for all $m \geq n$, see Lemma K2 in Appendix K. As discussed in Chapter 3, identifiability of β_0 in this sense is, in general, not sufficient to imply the stronger identifiable uniqueness condition

$$\liminf_{n \to \infty} \left[\bar{R}_n(\beta) - \bar{R}_n(\beta_0) \right] > 0 \text{ for all } \beta \neq \beta_0.$$

However, if $\bar{R}_n \equiv \bar{R}$ does not depend on n, then (within the context of Theorem 14.6) this condition reduces to the requirement that β_0 is the unique minimizer of $\bar{R}$, and then identifiability of β_0 (in the above sense) becomes equivalent to identifiable uniqueness of β_0.

14.2.3 Asymptotic Normality and Variance Covariance Matrix Estimation

In this subsection we provide results concerning the asymptotic distribution of the quasi-NFIML estimator in the case where the system is correctly specified as expressed by Assumptions 14.3 and 14.4. We also discuss how the asymptotic variance covariance matrix of the quasi-NFIML estimator can be estimated consistently. These results rely on Theorems 11.2 and 13.1. We introduce the following assumptions.[14]

[14]As is evident from the discussion in Chapter 8, given consistency it would actually suffice to postulate the subsequent conditions for asymptotic normality

Assumption 14.9. *The true parameter vector* $\beta_0 = (\alpha_0', \sigma_0')'$ *is contained in the interior of* $B = A \times S$.

Assumption 14.10. *(a) For every* $z \in Z$ *and* $t \geq 1$ *the functions* $f_t(z,.)$ *and* $\nabla_y f_t(z,.)$ *are twice continuously partially differentiable in* α *on* $\operatorname{int}(A)$.

(b) The families

$$\{\nabla_\alpha f_t : t \in \mathbf{N}\}, \; \{\nabla_{\alpha\alpha} f_t : t \in \mathbf{N}\}, \; \{\nabla_{\alpha\alpha} [\ln |\det(\nabla_y f_t)|] : t \in \mathbf{N}\}$$

are equicontinuous on $Z \times \operatorname{int}(A)$.

(c) $\{\nabla_\alpha [\ln |\det(\nabla_y f_t(z, \alpha_0))|] : t \in \mathbf{N}\}$ *is equicontinuous on* Z.

(d) $\sup_{t \geq 1} |\nabla_\alpha f_t(z,\alpha)| < \infty$, $\sup_{t \geq 1} |\nabla_{\alpha\alpha} f_t(z,\alpha)| < \infty$ *for each* $(z, \alpha) \in Z \times \operatorname{int}(A)$.

The first of the above assumptions is, of course, a standard assumption for establishing asymptotic normality. This assumption can also accommodate situations where the parameters in the system of equations are postulated to satisfy equality restrictions by defining α as the vector of free parameters, possibly after a reparameterization.[15] The second assumption maintains smoothness conditions guaranteeing differentiability of the objective function w.r.t. the parameters. The equicontinuity conditions also postulated in this assumption (together with similar equicontinuity conditions in Assumption 14.1) are used to establish, among other things, ULLNs for the Hessian of the objective function. The boundedness conditions in the second assumption above are automatically satisfied if the system functions f_t do not depend on t.

We next compute the score and the Hessian of the objective function. It is readily seen that the components corresponding to period t of the score and the Hessian of the objective function are given by:[16]

$$\nabla_{\beta'} q_t(z,\beta) = \begin{bmatrix} \nabla_{\alpha'} q_t(z,\beta) \\ \nabla_{\sigma'} q_t(z,\beta) \end{bmatrix} \tag{14.4a}$$

---

to hold only in a neighborhood around the true parameter vector. However, since the true parameter vector is unknown, these conditions would then have to be assumed to hold for all possible values of the true parameter vector anyway. For this reason, we have chosen to employ immediately a slightly stronger "global" form of the assumptions. Of course, a "local" version of the results can immediately be recovered upon redefining the parameter space as an appropriate neighborhood of the true parameter vector.

[15] Clearly, asymptotic normality results analogous to the ones given below can also be obtained for situations where the equality restrictions on β also involve the variance covariance matrix parameters σ. This is not pursued in the following in order to avoid increasing the complexity of the notation.

[16] For conventions regarding ∇_β, $\nabla_{\beta'}$, $\nabla_{\beta\beta}$, etc., see Footnote 3 in Chapter 8.

and

$$\nabla_{\beta\beta}q_t(z,\beta) = \begin{bmatrix} \nabla_{\alpha\alpha}q_t(z,\beta) & \nabla_{\alpha\sigma}q_t(z,\beta) \\ \nabla_{\sigma\alpha}q_t(z,\beta) & \nabla_{\sigma\sigma}q_t(z,\beta) \end{bmatrix} \tag{14.4b}$$

where

$$\begin{aligned}
\nabla_{\alpha'}q_t(z,\beta) &= -\nabla_{\alpha'}\left[\ln\left|\det\left(\nabla_y f_t(z,\alpha)\right)\right|\right] + \left(\nabla_{\alpha'}f_t(z,\alpha)\right)\Sigma^{-1}f_t(z,\alpha), \\
\nabla_{\sigma'}q_t(z,\beta) &= (1/2)\nabla_{\sigma'}\mathrm{vec}(\Sigma^{-1})\left[-\mathrm{vec}\left(\Sigma\right) + \mathrm{vec}\left(f_t(z,\alpha)f_t'(z,\alpha)\right)\right], \\
\nabla_{\alpha\alpha}q_t(z,\beta) &= -\nabla_{\alpha\alpha}\left[\ln\left|\det\left(\nabla_y f_t(z,\alpha)\right)\right|\right] \\
&\quad + \left(\nabla_{\alpha'}f_t(z,\alpha)\right)\Sigma^{-1}\left(\nabla_\alpha f_t(z,\alpha)\right) \qquad (14.4\text{c}) \\
&\quad + \left[\left(f_t'(z,\alpha)\Sigma^{-1}\right)\otimes I_{p_a}\right]\nabla_{\alpha\alpha}f_t(z,\alpha), \\
\nabla_{\alpha\sigma}q_t(z,\beta) &= \left(\nabla_{\sigma\alpha}q_t(z,\beta)\right)' = \left(f_t'(z,\alpha)\otimes\nabla_{\alpha'}f_t(z,\alpha)\right)\nabla_\sigma\mathrm{vec}\left(\Sigma^{-1}\right), \\
\nabla_{\sigma\sigma}q_t(z,\beta) &= (1/2)\nabla_{\sigma'}\mathrm{vec}\left(\Sigma^{-1}\right)\left(\Sigma\otimes\Sigma\right)\nabla_\sigma\mathrm{vec}\left(\Sigma^{-1}\right) \\
&\quad + (1/2)\left\{\left[\mathrm{vec}\left(f_t(z,\alpha)f_t'(z,\alpha)\right) - \mathrm{vec}(\Sigma)\right]'\otimes I_{p_e(p_e+1)/2}\right\} \\
&\quad \nabla_{\sigma\sigma}\mathrm{vec}\left(\Sigma^{-1}\right).
\end{aligned}$$

To establish the ULLNs for the Hessian matrix mentioned above, we also make use of moment conditions on the components making up the Hessian matrix. These moment conditions, together with moment conditions on the components of the score vector, are collected in the next assumption. The moment conditions for the score in parts (b) and (d) of this assumption are needed for the central limit theorem for the score vector. Of course, under the assumption that ϵ_t is distributed $N(0,\Sigma(\sigma_0))$ part (d) follows already from part (a). The condition in part (c) assists in verifying the martingale difference property of the score vector.

Assumption 14.11. *(a) For some* $\gamma > 0$

$$\sup_n n^{-1}\sum_{t=1}^n E\left[\sup_{\alpha\in\mathrm{int}(A)}\left|\nabla_\alpha f_t(\mathbf{z}_t,\alpha)\right|^{2+2\gamma}\right] < \infty,$$

$$\sup_n n^{-1}\sum_{t=1}^n E\left[\sup_{\alpha\in\mathrm{int}(A)}\left|f_t(\mathbf{z}_t,\alpha)\right|^{1+\gamma}\left|\nabla_{\alpha\alpha}f_t(\mathbf{z}_t,\alpha)\right|^{1+\gamma}\right] < \infty,$$

$$\sup_n n^{-1}\sum_{t=1}^n E\left[\sup_{\alpha\in\mathrm{int}(A)}\left|\nabla_{\alpha\alpha}\left[\ln\left|\det\left(\nabla_y f_t(\mathbf{z}_t,\alpha)\right)\right|\right]\right|^{1+\gamma}\right] < \infty.$$

(b) For some $\delta > 0$

$$\sup_n n^{-1}\sum_{t=1}^n E\left|\nabla_\alpha\left[\ln\left|\det\left(\nabla_y f_t(\mathbf{z}_t,\alpha_0)\right)\right|\right]\right|^{2+\delta} < \infty.$$

(c) For every $t \geq 1$

$$E\left[\sup_{\alpha\in\mathrm{int}(A)}\left|\nabla_\alpha\left[\ln\left|\det\left(\nabla_y f_t(\mathbf{z}_t,\alpha)\right)\right|\right]\right|\right] < \infty.$$

(d) For some $\delta > 0$

$$\sup_n n^{-1} \sum_{t=1}^{n} E\left[|\nabla_\alpha f_t(\mathbf{z}_t, \alpha_0)|^{2+\delta} |\epsilon_t|^{2+\delta} \right] < \infty,$$

$$\sup_n n^{-1} \sum_{t=1}^{n} E |\epsilon_t|^{4+2\delta} < \infty.$$

We introduce the following additional assumptions.

Assumption 14.12. *(a)*

$$\liminf_{n\to\infty} \lambda_{\min} \left[n^{-1} \sum_{t=1}^{n} E\nabla_{\beta\beta} q_t(\mathbf{z}_t, \beta_0) \right] > 0.$$

(b)

$$\liminf_{n\to\infty} \lambda_{\min} \left[n^{-1} \sum_{t=1}^{n} E(\nabla_{\beta'} q_t(\mathbf{z}_t, \beta_0) \nabla_\beta q_t(\mathbf{z}_t, \beta_0)) \right] > 0.$$

Assumption 14.13. *For every* $t \geq 1$

$$E\left[\sup_{\beta \in \text{int}(B)} |\nabla_\beta \exp(-q_t(\mathbf{z}_t, \beta))| / \exp(-q_t(\mathbf{z}_t, \beta_0)) \right] < \infty,$$

$$E\left[\sup_{\beta \in \text{int}(B)} |\nabla_{\beta\beta} \exp(-q_t(\mathbf{z}_t, \beta))| / \exp(-q_t(\mathbf{z}_t, \beta_0)) \right] < \infty.$$

The nonsingularity type condition for the Hessian matrix postulated in Assumption 14.12(a) above is an instance of a condition typically maintained in the context of M-estimation. It delivers the "necessary" curvature of the objective function, such that the deviation of the estimator from the true parameter vector can be expressed asympotically as a linear function of the score vector.[17] Assumption 14.12(b) seems to be necessary for

[17] Gallant and Holly (1980) do not maintain a condition like Assumption 14.12, but rather argue on p.716 in the proof of their Lemma 1 that this condition is implied by the assumption that the true parameter value is a unique minimizer of the limiting objective function. However, this argument is not correct since positive definiteness of the matrix of second order derivatives is a sufficient, but not a necessary condition for a critical point to be a (local) minimizer.

establishing a central limit theorem for the score vector in the general non-stationary case, cf. the discussion following Theorem 11.2. Under additional assumptions including normality and the regularity conditions formulated in Assumption 14.13, however, we will have

$$E\nabla_{\beta\beta}q_t(\mathbf{z}_t, \beta_0) = E\left(\nabla_{\beta'}q_t(\mathbf{z}_t, \beta_0)\nabla_{\beta}q_t(\mathbf{z}_t, \beta_0)\right)$$

and then Assumption 14.12(b) will reduce to Assumption 14.12(a). A sufficient condition for Assumption 14.12(a) that can be interpreted more readily will be given in Lemma 14.10 below.

The asymptotic variance covariance matrix of the quasi-NFIML estimator will turn out to be composed of the matrices C_n and D_n^2, where

$$\begin{aligned} C_n &= n^{-1}\sum_{t=1}^{n} E\nabla_{\beta\beta}q_t(\mathbf{z}_t, \beta_0), \\ D_n &= \left[n^{-1}\sum_{t=1}^{n} E\left(\nabla_{\beta'}q_t(\mathbf{z}_t, \beta_0)\nabla_{\beta}q_t(\mathbf{z}_t, \beta_0)\right)\right]^{1/2}. \end{aligned}$$

We introduce the following estimators for C_n and D_n^2:

$$\begin{aligned} \hat{C}_n &= n^{-1}\sum_{t=1}^{n} \nabla_{\beta\beta}q_t(\mathbf{z}_t, \hat{\beta}_n), \\ \hat{\Phi}_n &= n^{-1}\sum_{t=1}^{n} \nabla_{\beta'}q_t(\mathbf{z}_t, \hat{\beta}_n)\nabla_{\beta}q_t(\mathbf{z}_t, \hat{\beta}_n). \end{aligned}$$

To establish consistency of $\hat{\Phi}_n$ for D_n^2 we will need ULLNs to hold for the functions making up D_n. The following additional assumption is needed for this purpose only.

Assumption 14.14. *(a) The family* $\{\nabla_\alpha[\ln|\det(\nabla_y f_t)|] : t \in \mathbf{N}\}$ *is equicontinuous on* $Z \times \operatorname{int}(A)$.

(b) $\sup_{t\geq 1}|\nabla_\alpha[\ln|\det(\nabla_y f_t(z,\alpha))|]| < \infty$ *for each* $(z,\alpha) \in Z \times \operatorname{int}(A)$.

(c) For some $\gamma > 0$

$$\begin{aligned} &\sup_n n^{-1}\sum_{t=1}^{n} E\left[\sup_{\alpha\in\operatorname{int}(A)} |f_t(\mathbf{z}_t, \alpha)|^{4+4\gamma}\right] < \infty, \\ &\sup_n n^{-1}\sum_{t=1}^{n} E\left[\sup_{\alpha\in\operatorname{int}(A)} |\nabla_\alpha\left[\ln|\det\left(\nabla_y f_t(\mathbf{z}_t, \alpha)\right)|\right]|^{2+2\gamma}\right] < \infty, \\ &\sup_n n^{-1}\sum_{t=1}^{n} E\left[\sup_{\alpha\in\operatorname{int}(A)} |f_t(\mathbf{z}_t, \alpha)|^{2+2\gamma}\,|\nabla_\alpha f_t(\mathbf{z}_t, \alpha)|^{2+2\gamma}\right] < \infty. \end{aligned}$$

We now give three theorems concerning the asymptotic normality of the quasi-NFIML estimator. In the first of the subsequent theorems we postulate that β_0 is the (identifiably unique) minimizer of $\bar{R}_n$ and that the sequence $\nabla_\beta q_t(\mathbf{z}_t, \beta_0)$ forms a martingale difference sequence. In the second and third of the subsequent theorems those properties are established from sufficient conditions. In the latter theorem we also imply Assumption 14.12(b) from Assumption 14.12(a) by utilizing Assumption 14.13.

Theorem 14.7. *Let*

$$R_n(\omega, \beta) = n^{-1} \sum_{t=1}^{n} q_t(\mathbf{z}_t, \beta)$$

be the objective function of the quasi-NFIML estimator defined in (14.2) and

$$\bar{R}_n(\beta) = n^{-1} \sum_{t=1}^{n} E q_t(\mathbf{z}_t, \beta).$$

Suppose Assumptions 14.1(a)-(c),(f),(g), 14.3 - 14.5, 14.6, 14.7, 14.9, 14.10, 14.11(a),(b),(d), 14.12 hold, suppose $\bar{R}_n$ is minimized at β_0 for every $n \in \mathbf{N}$,*

$$\liminf_{n\to\infty} \left[\bar{R}_n(\beta) - \bar{R}_n(\beta_0)\right] > 0 \text{ for all } \beta \neq \beta_0,$$

and suppose

$$E[\nabla_\beta q_t(\mathbf{z}_t, \beta_0) \mid \mathfrak{F}_{t-1}] = 0 \text{ a.s. for } t \geq 1,$$

where $\mathfrak{F}_{t-1}$ is the σ-field generated by $\{\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{1-l}, \mathbf{x}_t, \ldots, \mathbf{x}_1\}$. Let $\hat{\beta}_n$ be any sequence of quasi-NFIML estimators, i.e.,

$$R_n(\omega, \hat{\beta}_n) = \inf_B R_n(\omega, \beta)$$

holds for all $\omega \in \Omega$ and $n \in \mathbf{N}$.

(a) Then

$$n^{1/2} D_n^{-1} C_n (\hat{\beta}_n - \beta_0) \xrightarrow{D} N(0, I),$$

where

$$C_n = n^{-1} \sum_{t=1}^{n} E \nabla_{\beta\beta} q_t(\mathbf{z}_t, \beta_0)$$

and

$$D_n = \left[n^{-1} \sum_{t=1}^{n} E\left(\nabla_{\beta'} q_t(\mathbf{z}_t, \beta_0) \nabla_\beta q_t(\mathbf{z}_t, \beta_0)\right) \right]^{1/2}.$$

Furthermore, we have $|C_n| = O(1)$, $\left|C_n^{-1}\right| = O(1)$, $|D_n| = O(1)$ and $\left|D_n^{-1}\right| = O(1)$, and hence $\hat{\beta}_n$ is $n^{1/2}$-consistent for β_0.

(b) If in addition Assumption 14.14 holds, then

$$\hat{C}_n^{-1}\hat{\Phi}_n\hat{C}_n^{-1} - C_n^{-1}D_n^2C_n^{-1} \to 0 \text{ i.p. as } n \to \infty.$$ [18]

If D_nD_n' converges to some matrix, say, Λ, the asymptotic normality result in part (a) can be cast into the form $n^{1/2}C_n(\hat{\beta}_n - \beta_0) \xrightarrow{D} N(0, \Lambda)$. Under convergence of D_nD_n', this result can in fact be obtained without making use of Assumption 14.12(b), cf. Theorem 11.2(b). A similar remark applies to the next theorem.

In Theorem 14.7 we have assumed both that $\bar{R}_n$ is minimized at β_0 and that $\nabla_\beta q_t(\mathbf{z}_t, \beta_0)$ is a martingale difference sequence. In the special case of a static model with nonstochastic exogenous variables and where the disturbances ϵ_t are independently distributed, the martingale difference property follows automatically from the assumption that β_0 minimizes $\bar{R}_n$ (provided $E\nabla_\beta q_t(\mathbf{z}_t, \beta) = \nabla_\beta E q_t(\mathbf{z}_t, \beta)$ holds, which is guaranteed, e.g., under Assumptions 14.1(c), 14.11(a),(c)).

We return again to the case of a general dynamic model. In contrast to Theorem 14.7, the following theorem maintains Assumption 14.8 which postulates, in particular, normality of the disturbances. The following theorem does not postulate the martingale difference property of $\nabla_\beta q_t(\mathbf{z}_t, \beta_0)$ and the property that $\bar{R}_n$ is minimized at β_0 as assumptions. Rather, the proof shows that these two properties can now be established with the help of Assumption 14.8.

Theorem 14.8. *Let*

$$R_n(\omega, \beta) = n^{-1}\sum_{t=1}^{n} q_t(\mathbf{z}_t, \beta)$$

be the objective function of the quasi-NFIML estimator defined in (14.2) and

$$\bar{R}_n(\beta) = n^{-1}\sum_{t=1}^{n} E q_t(\mathbf{z}_t, \beta).$$

Suppose Assumptions 14.1(a)-(c),(f),(g), 14.3 - 14.5, 14.6(a),(b), 14.7 - 14.10, 14.11(a)-(c), 14.12 hold, and suppose that*

$$\liminf_{n\to\infty} \left[\bar{R}_n(\beta) - \bar{R}_n(\beta_0)\right] > 0 \text{ for all } \beta \neq \beta_0.$$

Let $\hat{\beta}_n$ be any sequence of quasi-NFIML estimators, i.e.,

$$R_n(\omega, \hat{\beta}_n) = \inf_B R_n(\omega, \beta)$$

[18] We note that Assumption 14.14 overrides in part some of the other assumptions maintained in the theorem. For simplicity of presentation we have not further condensed the list of maintained assumptions.

holds for all $\omega \in \Omega$ *and* $n \in \mathbf{N}$.

(a) Then

$$n^{1/2} D_n^{-1} C_n (\hat{\beta}_n - \beta_0) \xrightarrow{D} N(0, I),$$

where

$$C_n = n^{-1} \sum_{t=1}^{n} E \nabla_{\beta\beta} q_t(\mathbf{z}_t, \beta_0)$$

and

$$D_n = \left[n^{-1} \sum_{t=1}^{n} E \left(\nabla_{\beta'} q_t(\mathbf{z}_t, \beta_0) \nabla_{\beta} q_t(\mathbf{z}_t, \beta_0) \right) \right]^{1/2}.$$

Furthermore, we have $|C_n| = O(1)$, $|C_n^{-1}| = O(1)$, $|D_n| = O(1)$ *and* $|D_n^{-1}| = O(1)$, *and hence* $\hat{\beta}_n$ *is* $n^{1/2}$*-consistent for* β_0.

(b) If in addition Assumption 14.14 holds, then

$$\hat{C}_n^{-1} \hat{\Phi}_n \hat{C}_n^{-1} - C_n^{-1} D_n^2 C_n^{-1} \to 0 \text{ i.p. as } n \to \infty.$$

The following theorem differs from Theorem 14.8 in that it additionally maintains Assumption 14.13. Consequently, as remarked above,

$$E \nabla_{\beta\beta} q_t(\mathbf{z}_t, \beta_0) = E \left(\nabla_{\beta'} q_t(\mathbf{z}_t, \beta_0) \nabla_{\beta} q_t(\mathbf{z}_t, \beta_0) \right).$$

Hence Assumption 14.12(a) coincides with Assumption 14.12(b), which can therefore be dropped from the list of assumptions.

Theorem 14.9. *Let*

$$R_n(\omega, \beta) = n^{-1} \sum_{t=1}^{n} q_t(\mathbf{z}_t, \beta)$$

be the objective function of the quasi-NFIML estimator defined in (14.2) and

$$\bar{R}_n(\beta) = n^{-1} \sum_{t=1}^{n} E q_t(\mathbf{z}_t, \beta).$$

Suppose Assumptions 14.1(a)-(c),(f),(g), 14.3 - 14.5, 14.6(a),(b), 14.7 - 14.10, 14.11(a)-(c), 14.12(a), 14.13 hold, and suppose that*

$$\liminf_{n \to \infty} \left[\bar{R}_n(\beta) - \bar{R}_n(\beta_0) \right] > 0 \text{ for all } \beta \neq \beta_0.$$

Let $\hat{\beta}_n$ *be any sequence of quasi-NFIML estimators, i.e.,*

$$R_n(\omega, \hat{\beta}_n) = \inf_{B} R_n(\omega, \beta)$$

holds for all $\omega \in \Omega$ *and* $n \in \mathbf{N}$.

(a) Then

$$n^{1/2} C_n^{1/2} (\hat{\beta}_n - \beta_0) \xrightarrow{D} N(0, I),$$

where

$$C_n = n^{-1} \sum_{t=1}^{n} E \nabla_{\beta\beta} q_t(\mathbf{z}_t, \beta_0).$$

Furthermore, we have $|C_n| = O(1)$, $\left|C_n^{-1}\right| = O(1)$, *and hence* $\hat{\beta}_n$ *is* $n^{1/2}$*-consistent for* β_0. *Furthermore,*

$$E \nabla_{\beta\beta} q_t(\mathbf{z}_t, \beta_0) = E\left(\nabla_{\beta'} q_t(\mathbf{z}_t, \beta_0) \nabla_{\beta} q_t(\mathbf{z}_t, \beta_0)\right)$$

for every $t \geq 1$ *holds and hence* $C_n = D_n D_n' = D_n^2$, *where* D_n *is as in Theorem 14.8.*

(b) $\hat{C}_n^{-1} - C_n^{-1} \to 0$ *i.p. as* $n \to \infty$.

(c) If in addition Assumption 14.14 holds, then also $\hat{\Phi}_n^{-1} - C_n^{-1} \to 0$ *i.p. and* $\hat{C}_n^{-1} \hat{\Phi}_n \hat{C}_n^{-1} - C_n^{-1} \to 0$ *i.p. as* $n \to \infty$.

Under the assumptions of Theorems 14.7 and 14.8 the asymptotic variance covariance matrix of $\hat{\beta}_n$ is given by $C_n^{-1} D_n^2 C_n^{-1}$. Within the context of Theorem 14.9 this expression simplifies to C_n^{-1}, observing that in this case $C_n = D_n^2$. Parts (b) and (c) of this theorem provide three alternative consistent estimators for the asymptotic variance covariance matrix. (Note that for consistency of $\hat{C}_n^{-1}$ the additional Assumption 14.14 is not needed.[19])

The asymptotic normality results given above maintain in Assumption 14.12(a) that the Hessian matrix of the negative log-likelihood function is "uniformly" positive definite. Since the Hessian matrix of the log-likelihood function is a complicated expression involving the system function f_t as well as various of its derivatives w.r.t. α and y, this assumption is not easy to interpret in terms of the underlying system. In the following lemma we will provide a sufficient condition that only involves the first order derivative of f_t w.r.t. α and that has a natural interpretation as a "persistent excitation" condition. This generalizes an analogous sufficient condition in Amemiya (1977), who considered a special case of our model corresponding to a static system without cross-equation parameter restrictions and with nonstochastic and – loosely speaking – asymptotically stationary regressors. For the following lemma we need an additional assumption.

Assumption 14.15. *(a) For every* $(y_{-1}', \ldots, y_{-l}', x')' \in \mathbf{R}^{lp_y + p_x}$ *and* $t \geq 1$ *the functions* $f_t(y, y_{-1}, \ldots, y_{-l}, x, \alpha)$ *are twice continuously partially differentiable w.r.t.* y *and* α *on* $\mathbf{R}^{p_y} \times \operatorname{int}(A)$.

[19] Compare Footnote 2 in Chapter 13 concerning references that provide insight into the relative merits of these alternative variance covariance matrix estimators.

(b) For all $k = 1, \ldots, p_a$ and $t \geq 1$

$$E\left\{\left|\nabla_y(\partial/\partial\alpha_k)f_t(\mathbf{z}_t,\alpha_0)\left[\nabla_y f_t(\mathbf{z}_t,\alpha_0)\right]^{-1}\right|^{1+\delta}\right\} < \infty$$

for some $\delta > 0$, where α_k denotes the k-th component of α.

Lemma 14.10. [20] *Suppose the assumptions for part (a) of Theorem 14.9 hold, except for Assumption 14.12(a). Suppose further that Assumption 14.15 holds. Then the following "persistent excitation" condition*

$$\liminf_{n\to\infty} \lambda_{\min}\left[n^{-1}\sum_{t=1}^{n} E\left\{E\left[\nabla_{\alpha'} f_t(\mathbf{z}_t,\alpha_0)|\mathfrak{F}_{t-1}\right]E\left[\nabla_\alpha f_t(\mathbf{z}_t,\alpha_0)|\mathfrak{F}_{t-1}\right]\right\}\right] > 0, \tag{14.5}$$

where $\mathfrak{F}_{t-1}$ is the σ-field generated by $\{\mathbf{y}_{t-1},\ldots,\mathbf{y}_{1-l},\mathbf{x}_t,\ldots,\mathbf{x}_1\}$, implies that Assumption 14.12(a), i.e.,

$$\liminf_{n\to\infty} \lambda_{\min}\left[n^{-1}\sum_{t=1}^{n} E\nabla_{\beta\beta}q_t(\mathbf{z}_t,\beta_0)\right] > 0,$$

holds.[21]

Clearly, if

$$F_n = n^{-1}\sum_{t=1}^{n} E\left\{E\left[\nabla_{\alpha'} f_t(\mathbf{z}_t,\alpha_0)\mid\mathfrak{F}_{t-1}\right]E\left[\nabla_\alpha f_t(\mathbf{z}_t,\alpha_0)\mid\mathfrak{F}_{t-1}\right]\right\}$$

converges to some limiting matrix, then the "persistent excitation" condition (14.5) reduces to the condition of nonsingularity of the limiting matrix. To relate the "persistent excitation" condition (14.5) to Condition 6 in Amemiya (1977) assume for the moment that this convergence indeed takes place and that we consider the case of a static system without cross-equation parameter restrictions and with nonstochastic exogenous variables. Let α be partitioned as $\alpha = (\alpha^{(1)\prime},\ldots,\alpha^{(p_y)\prime})'$, where $\alpha^{(i)}$ denotes the subset of parameters actually appearing in the i-th equation. The absence of cross-equation parameter restrictions then implies that

$$\begin{aligned} F_n &= n^{-1}\sum_{t=1}^{n} \operatorname{diag}_i\left(E\nabla_{\alpha^{(i)}} f_{ti}(\mathbf{z}_t,\alpha_0)\right)' \operatorname{diag}_i\left(E\nabla_{\alpha^{(i)}} f_{ti}(\mathbf{z}_t,\alpha_0)\right) \\ &= \operatorname{diag}_i\left(n^{-1}\sum_{t=1}^{n} E\nabla_{\alpha^{(i)}} f_{ti}(\mathbf{z}_t,\alpha_0)' E\nabla_{\alpha^{(i)}} f_{ti}(\mathbf{z}_t,\alpha_0)\right), \end{aligned}$$

[20] We note that Assumption 14.15 overrides in part some of the other assumptions maintained in Theorem 14.9 and in the lemma. Again, for simplicity of presentation we have not further condensed the list of maintained assumptions.

[21] Since under the maintained assumptions $C_n = D_n^2$ it follows that also Assumption 14.12(b) holds.

where the operator diag_i creates a block diagonal matrix with p_y diagonal blocks. Condition 6 in Amemiya (1977) is now readily seen to correspond to the assumption that $\lim_{n\to\infty} F_n$ is nonsingular. If, for example, the model is further specialized to a linear seemingly unrelated regression model, i.e.,

$$f_{ti}(\mathbf{z}_t, \alpha_0) = \mathbf{y}_{ti} - \mathbf{x}_t^{(i)}\alpha^{(i)}$$

where $\mathbf{y}_{ti}$ denotes the i-th endogenous variable in period t and $\mathbf{x}_t^{(i)}$ denotes the row vector of (nonstochastic) exogenous variables in the i-th equation in period t, then this condition reduces to nonsingularity of

$$\lim_{n\to\infty} n^{-1}\sum_{t=1}^{n} \mathbf{x}_t^{(i)\prime}\mathbf{x}_t^{(i)}$$

for $i = 1, \ldots, p_y$.

For further interpretation of Lemma 14.10 we note that clearly the following condition is equivalent to condition (14.5):

$$\liminf_{n\to\infty} \lambda_{\min}(\Psi_n) > 0, \tag{14.6}$$

where

$$\Psi_n = n^{-1}\sum_{t=1}^{n} E\left\{E\left[\nabla_{\alpha'} f_t(\mathbf{z}_t, \alpha_0)|\mathfrak{F}_{t-1}\right]\Sigma(\sigma_0)^{-1}E\left[\nabla_\alpha f_t(\mathbf{z}_t, \alpha_0)|\mathfrak{F}_{t-1}\right]\right\}.$$

From Theorem 11.5 we now see that under suitable regularity conditions Ψ_n is the inverse of the asymptotic variance covariance matrix of the best N3SLS estimator, where the instruments are taken to be the elements of $E[\nabla_\alpha f_t(\mathbf{z}_t, \alpha_0) \mid \mathfrak{F}_{t-1}]$; cf. also Amemiya (1977). The proof of Lemma 14.10, and the proof of the analogous result in Amemiya (1977), can now be motivated by the heuristic reasoning that the asymptotic variance covariance matrix C_n^{-1} of $\hat{\beta}_n$ should not be larger than the asymptotic variance covariance matrix Ψ_n^{-1} of the best N3SLS estimator, or equivalently that $C_n \geq \Psi_n$ should hold. If this can be established, then clearly condition (14.5), which implies that the smallest eigenvalues of Ψ_n are bounded away from zero, will also imply the same for the smallest eigenvalues of C_n. Hence the idea of the proof of Lemma 14.10 is to compare C_n with Ψ_n and to establish that indeed $C_n \geq \Psi_n$ holds.[22]

[22]As a byproduct, this discussion also shows that Lemma 14.10 establishes, under any set of regularity assumptions which imply the assumptions of the lemma as well as those of Theorem 11.5 (applied to the best N3SLS estimator), that the best N3SLS estimator can at most be asymptotically as efficient as $\hat{\beta}_n$.

15

CONCLUDING REMARKS

This book provides an asymptotic theory for M-estimators in the context of dynamic nonlinear models. To accommodate processes generated by dynamic nonlinear models the theory has to allow for temporal dependence and (possibly also) for temporal heterogeneity in the data generating process. This is achieved by employing weak dependence concepts like L_p-approximability or near epoch dependence, which are flexible enough to cover processes generated by dynamic nonlinear models, yet are strong enough to permit the derivation of laws of large numbers and central limit theorems for such processes.

Apart from providing consistency and asymptotic normality results for M-estimators in general, and for generalized method of moments and least mean distance estimators (including maximum likelihood estimators) in particular, the book provides a detailed discussion of the strategies used in consistency and asymptotic normality proofs, as well as of the statistical and probabilistic tools employed in this context.

The results in Chapter 3, together with the generic uniform laws of large numbers in Chapter 5 and the local laws of large numbers in Chapter 6, provide a set of basic modules which can be used to prove consistency of M-estimators in dynamic nonlinear models, as well as in other frameworks. Extensions of the results in Chapter 3 are discussed in Chapter 4. One of those extensions is a "generalized" consistency result for the case where the identifiable uniqueness condition fails.

Chapter 6 develops, based on the concept of L_p-approximability, an encompassing framework for the concepts of stochastic stability and near epoch dependence, which have been used in the literature on dynamic nonlinear models. Theorems 6.5 - 6.7 and Corollary 6.8 provide sets of conditions under which transformations of L_p-approximable or near epoch dependent processes have the corresponding property. Those results are important in various contexts, e.g., they can be used to derive local laws of large numbers based on the laws of large numbers given in Theorems 6.3 and 6.4. Of course, an important question that arises naturally in establishing asymptotic results for dynamic nonlinear models is: Under which conditions is the output process of such a model L_p-approximable or near epoch dependent, given the input process has the corresponding property? Theorems 6.10 - 6.12 provide sets of sufficient conditions. The last one of those theorems covers not only first order, but also higher order dynamic

B. M. Pötscher et al., *Dynamic Nonlinear Econometric Models*

models.

Exemplary catalogues of assumptions, which ensure consistency of least mean distance estimators and generalized method of moments estimators, are given in Chapter 7.

Chapter 8, together with the central limit theorems in Chapter 10 for L_p-approximable or near epoch dependent processes, provides basic modules for proving asymptotic normality of M-estimators for dynamic nonlinear systems. These results are used in Chapter 11 to provide exemplary catalogues of assumptions which ensure asymptotic normality of least mean distance estimators and generalized method of moments estimators.

Chapter 12 discusses consistency results for heteroskedasticity and autocorrelation robust variance covariance matrix estimators in case the data generating process is only assumed to be near epoch dependent on some mixing basis process. These results also provide rates of convergence of the variance covariance matrix estimators, which are essential for the optimal selection of the truncation lag parameter.

Theorems regarding the consistent estimation of the asymptotic variance covariance matrix of generalized method of moments and least mean distance estimators are given in Chapter 13.

In Chapter 14 the general asymptotic theory developed so far is applied to an important concrete example. In particular, this chapter provides consistency and asymptotic normality results for the (quasi) normal full information maximum likelihood estimator of a dynamic nonlinear simultaneous equation system. The results cover both the case of a correctly specified and a misspecified system. In case of a correctly specified system, the "weak dependence" property (L_p-approximability) of the process of endogenous variables is derived from "weak dependence" properties of the exogenous variables and disturbances based on Theorem 6.12, rather than being simply postulated as an assumption.

Although consistency and asymptotic normality results lay the foundation for inference in dynamic nonlinear models, this book has not covered a number of other important aspects of asymptotic inference in such models. In the following we discuss some of these aspects and point to the relevant literature.

(i) The results presented in this book focus primarily on parametric models. For example, the asymptotic normality results are based on the assumption that the parameter spaces are subsets of Euclidean spaces. The basic consistency results, however, do not explicitly rule out semiparametric or nonparametric situations, since for those results the parameter spaces are only assumed to be abstract metric spaces. We note, however, that the various compactness assumptions may be restrictive in an infinite dimensional setting. For surveys of semiparametric methods in econometrics see, e.g., Robinson (1988) and Powell (1994).

(ii) A further issue not considered in this book is the construction of test statistics. Standard tests of parametric hypotheses, like score tests, Wald tests or likelihood ratio type tests can, of course, be built in a quite straightforward manner on the basis of the asymptotic normality results presented in the book. For an array of such results within the context of dynamic nonlinear models see Gallant (1987a, Ch.7) and Gallant and White (1988, Ch.7). Model specification tests for nonlinear models can be found in, e.g., Bierens (1982b, 1984, 1990), Newey (1985a,b), Ruud (1984), Tauchen (1985), White (1987), and Wooldridge (1990, 1991), cf. also the books by Bierens (1994) and White (1994).

(iii) The book does not discuss one-step M-estimators. As is well-known, one-step M-estimators initialized by a $n^{1/2}$-consistent estimator of the parameter vector are frequently asymptotically normal, even if the corresponding M-estimators are not, and are asymptotically equivalent to the corresponding M-estimators, when those estimators are asymptotically normal. See, e.g., LeCam (1960) and Bickel (1975). For analogous results in the context of nonlinear and simultaneous equation models see, e.g., Pötscher and Prucha (1986a) and Prucha and Kelejian (1984). Clearly many of the techniques for establishing consistency and asymptotic normality of M-estimators discussed in this book are also applicable for establishing the corresponding properties of one-step M-estimators.

(iv) The book also does not address questions of efficiency. Certainly, efficiency questions play a central role in statistics and econometrics and have a long history in that literature; see, e.g., Beran (1996), Ghosh (1985), and Wong (1992) for recent reviews and a discussion of the history of the efficiency concept. Recent articles in the econometrics literature that treat questions of efficiency in a context more or less closely related to the context of the present book are Bates and White (1988), Chamberlain (1987), Hansen (1982, 1985, 1988), Hansen, Heaton and Ogaki (1988), and Hansen and Singleton (1991). For surveys see Newey (1990) and Newey and McFadden (1994). See also the recent review article by Jeganathan (1995).

A number of recent books may also be of interest for additional reading. Apart from Gallant (1987a) and Gallant and White (1988), Bierens (1994) and White (1994) are further econometrics texts that deal with asymptotic inference in nonlinear models; cf. also the survey article by Wooldridge (1994). The time series analysis perspective on inference in dynamic nonlinear models is well-represented in books by Tong (1990) and Guégan (1994).

Appendix A

PROOFS FOR CHAPTER 3

Lemma A1. *Let* (Φ, σ) *be a metric space and let* Λ *be a nonempty set. For sequences of functions* $g_n : \Phi \to \mathbf{R}$, $\bar{g}_n : \Phi \to \mathbf{R}$, $\varphi_n : \Lambda \to \Phi$, $\bar{\varphi}_n : \Lambda \to \Phi$ *consider the following conditions:*

(1) $\sup_{\lambda \in \Lambda} \sigma(\varphi_n(\lambda), \bar{\varphi}_n(\lambda)) \to 0$ *as* $n \to \infty$.

(2) *(i)* $\{\bar{g}_n : n \in \mathbf{N}\}$ *is uniformly equicontinuous on* Φ;

(ii) $\sup_{\varphi \in \Phi} |g_n(\varphi) - \bar{g}_n(\varphi)| \to 0$ *as* $n \to \infty$.

(2') *(i) There exists a compact set* $\Phi_* \subseteq \Phi$ *such that* $\{\bar{g}_n : n \in \mathbf{N}\}$ *is equicontinuous on the subset* Φ_* *of* Φ, *and* $\bar{\varphi}_n(\lambda) \in \Phi_*$ *for all* $\lambda \in \Lambda$ *and* $n \in \mathbf{N}$;

(ii) there exists an open set Φ_{**} *with* $\Phi_* \subseteq \Phi_{**} \subseteq \Phi$ *such that*

$$\sup_{\varphi \in \Phi_{**}} |g_n(\varphi) - \bar{g}_n(\varphi)| \to 0 \text{ as } n \to \infty.$$

(a) Then under Assumptions (1) and (2), or (1) and (2'),

$$\sup_{\lambda \in \Lambda} |g_n(\varphi_n(\lambda)) - \bar{g}_n(\bar{\varphi}_n(\lambda))| \to 0 \text{ as } n \to \infty. \tag{A.1}$$

(b) If Λ *is endowed with a metric* κ, *then under Assumptions (2)(i) or (2')(i) the family* $\{\bar{g}_n(\bar{\varphi}_n) : n \in \mathbf{N}\}$ *is equicontinuous [uniformly equicontinuous] on* Λ *if the family* $\{\bar{\varphi}_n : n \in \mathbf{N}\}$ *is equicontinuous [uniformly equicontinuous] on* Λ.

Proof. We first prove (A.1) under Assumptions (1) and (2). Note that (2)(i) implies that for any $\epsilon > 0$ there exists a $\delta > 0$ such that

$$\sup_k |\bar{g}_k(\varphi) - \bar{g}_k(\varphi^\bullet)| < \epsilon/2$$

whenever $\sigma(\varphi, \varphi^\bullet) < \delta$. Assumptions (1) and (2)(ii) imply that there exists an N such that

$$\sup_{\lambda \in \Lambda} \sigma(\varphi_n(\lambda), \bar{\varphi}_n(\lambda)) < \delta$$

and

$$\sup_{\varphi \in \Phi} |g_n(\varphi) - \bar{g}_n(\varphi)| < \epsilon/2$$

for all $n \geq N$. Hence for $n \geq N$ we have

$$\begin{aligned} & \sup_{\lambda\in\Lambda} |g_n(\varphi_n(\lambda)) - \bar{g}_n(\bar{\varphi}_n(\lambda))| \\ \leq & \sup_{\lambda\in\Lambda} |g_n(\varphi_n(\lambda)) - \bar{g}_n(\varphi_n(\lambda))| + \sup_{\lambda\in\Lambda} |\bar{g}_n(\varphi_n(\lambda)) - \bar{g}_n(\bar{\varphi}_n(\lambda))| \\ < & \ \epsilon/2 + \sup_{\lambda\in\Lambda}\sup_k |\bar{g}_k(\varphi_n(\lambda)) - \bar{g}_k(\bar{\varphi}_n(\lambda))| < \epsilon. \end{aligned}$$

Next we prove (A.1) under Assumptions (1) and (2'). Note that (2')(i) implies that for any $\epsilon > 0$ and every $\varphi^\bullet \in \Phi_*$ there exists a $\delta(\epsilon, \varphi^\bullet) > 0$ such that

$$\sup_k |\bar{g}_k(\varphi) - \bar{g}_k(\varphi^\bullet)| < \epsilon/2$$

for all $\varphi \in \Phi$ with $\sigma(\varphi, \varphi^\bullet) < \delta(\epsilon, \varphi^\bullet)$. Cover Φ_* by all open balls with centers at $\varphi^\bullet \in \Phi_*$ and radii $\delta(\epsilon, \varphi^\bullet)/2$. Since Φ_* is compact there exist finitely many $\varphi_1^\bullet, \ldots, \varphi_m^\bullet$ (which may depend on ϵ) such that the corresponding balls cover Φ_*. Let Φ^* be the union of these balls intersected with Φ_{**}. Then Φ^* is open and $\Phi_* \subseteq \Phi^* \subseteq \Phi_{**}$. Let

$$\delta_0 = \delta_0(\epsilon) = \min_{1\leq i\leq m} \delta(\epsilon, \varphi_i^\bullet) > 0.$$

It follows from Assumption (1) that there exists an integer N (depending on ϵ through Φ^* and δ_0) such that for all $n \geq N$ and all $\lambda \in \Lambda$ we have $\sigma(\varphi_n(\lambda), \bar{\varphi}_n(\lambda)) < \delta_0/2$ and $\varphi_n(\lambda) \in \Phi^* \subseteq \Phi_{**}$. The latter result follows from Lemma A2 below. As a consequence, for any $n \geq N$ and $\lambda \in \Lambda$ there exists an index i ($1 \leq i \leq m$ where i may depend on n and λ) such that

$$\sigma(\varphi_n(\lambda), \varphi_i^\bullet) < \delta(\epsilon, \varphi_i^\bullet)/2 \qquad (\#)$$

and hence also

$$\begin{aligned} \sigma(\bar{\varphi}_n(\lambda), \varphi_i^\bullet) & \leq \sigma(\bar{\varphi}_n(\lambda), \varphi_n(\lambda)) + \sigma(\varphi_n(\lambda), \varphi_i^\bullet) \qquad (\#\#) \\ & < \delta_0/2 + \delta(\epsilon, \varphi_i^\bullet)/2 \leq \delta(\epsilon, \varphi_i^\bullet). \end{aligned}$$

Consequently, for all $n \geq N$ and $\lambda \in \Lambda$:

$$\begin{aligned} & |g_n(\varphi_n(\lambda)) - \bar{g}_n(\bar{\varphi}_n(\lambda))| \\ \leq & \ |g_n(\varphi_n(\lambda)) - \bar{g}_n(\varphi_n(\lambda))| + |\bar{g}_n(\varphi_n(\lambda)) - \bar{g}_n(\bar{\varphi}_n(\lambda))| \\ \leq & \sup_{\varphi\in\Phi_{**}} |g_n(\varphi) - \bar{g}_n(\varphi)| + \sup_k |\bar{g}_k(\varphi_n(\lambda)) - \bar{g}_k(\varphi_i^\bullet)| \\ & + \sup_k |\bar{g}_k(\varphi_i^\bullet) - \bar{g}_k(\bar{\varphi}_n(\lambda))| \\ < & \sup_{\varphi\in\Phi_{**}} |g_n(\varphi) - \bar{g}_n(\varphi)| + \epsilon, \end{aligned}$$

where the last inequality follows from the construction of $\delta(\epsilon, \varphi_i^\bullet)$ and from (#) and (##). Because of (2')(ii) it follows that

$$\sup_{\lambda\in\Lambda} |g_n(\varphi_n(\lambda)) - \bar{g}_n(\bar{\varphi}_n(\lambda))| \leq 2\epsilon$$

for n sufficiently large.

Next we prove part (b) of the lemma under Assumption (2)(i). Since $\{\bar{g}_n : n \in \mathbf{N}\}$ is uniformly equicontinuous on Φ it follows that for every $\epsilon > 0$ there exists a $\delta = \delta(\epsilon) > 0$ such that

$$\sup_n |\bar{g}_n(\varphi) - \bar{g}_n(\varphi^\bullet)| < \epsilon$$

whenever $\sigma(\varphi, \varphi^\bullet) < \delta$. Given $\{\bar{\varphi}_n : n \in \mathbf{N}\}$ is equicontinuous it follows that for this $\delta > 0$ and any $\lambda^\bullet \in \Lambda$ there exists an $\eta = \eta(\delta(\epsilon), \lambda^\bullet) > 0$ such that

$$\sup_n \sigma\left(\bar{\varphi}_n(\lambda), \bar{\varphi}_n(\lambda^\bullet)\right) < \delta$$

whenever $\kappa(\lambda, \lambda^\bullet) < \eta$. Consequently, for every $\epsilon > 0$ and $\lambda^\bullet \in \Lambda$ we have for this $\eta = \eta(\delta(\epsilon), \lambda^\bullet)$ that

$$\sup_n |\bar{g}_n\left(\bar{\varphi}_n(\lambda)\right) - \bar{g}_n(\bar{\varphi}_n\left(\lambda^\bullet\right))| \leq \epsilon$$

whenever $\kappa(\lambda, \lambda^\bullet) < \eta$, i.e., $\{\bar{g}_n(\bar{\varphi}_n) : n \in \mathbf{N}\}$ is equicontinuous. The result for the case where equicontinuity is replaced by uniform equicontinuity follows by analogous argumentation. The proof of part (b) of the lemma under Assumption (2')(i) is completely analogous to that under Assumption (2)(i) observing that the family of restrictions of $\bar{g}_n$ to Φ_* is equicontinuous on Φ_* and hence uniformly equicontinuous on Φ_* (since Φ_* is compact), and observing that $\bar{\varphi}_n(\Lambda) \subseteq \Phi_*$. ■

Lemma A2. *Let (Φ, σ) be a metric space and let Φ_* and Φ^* be, respectively, a compact and an open subset of Φ with $\Phi_* \subseteq \Phi^*$. Let Λ be a nonempty set. Consider sequences of functions $\varphi_n : \Lambda \to \Phi$ and $\bar{\varphi}_n : \Lambda \to \Phi_*$. If*

$$\sup_{\lambda \in \Lambda} \sigma\left(\varphi_n(\lambda), \bar{\varphi}_n(\lambda)\right) \to 0 \text{ as } n \to \infty,$$

then there exists an N such that $\varphi_n(\lambda) \in \Phi^*$ for all $n \geq N$ and all $\lambda \in \Lambda$.

Proof. If $\Phi^* = \Phi$ the lemma is trivial. If $\Phi^* \neq \Phi$, observe that $\gamma = \inf\{\sigma(\varphi, \varphi^\bullet) : \varphi \in \Phi_*, \varphi^\bullet \in \Phi - \Phi^*\} > 0$; see Dieudonné (1960), Section 3.17. By assumption there exists an $N = N(\gamma)$ such that

$$\sup_{\lambda \in \Lambda} \sigma\left(\varphi_n(\lambda), \bar{\varphi}_n(\lambda)\right) < \gamma$$

for all $n \geq N$. Since $\bar{\varphi}_n(\lambda) \in \Phi_*$ it follows that $\varphi_n(\lambda) \in \Phi^*$ for all $\lambda \in \Lambda$ and $n \geq N$. ■

Proof of Lemma 3.1. Let $\Omega_0 \subseteq \Omega$ be a set of probability one on which (3.2) holds and (3.3) is satisfied for large n. Fix $\omega \in \Omega_0$. We first show that for every $\epsilon > 0$

$$\liminf_{n\to\infty}\left[\inf_{\rho_B(\beta,\bar{\beta}_n)\geq\varepsilon} R_n(\omega,\beta) - R_n(\omega,\bar{\beta}_n)\right] > 0 \tag{A.2}$$

holds. This follows since the l.h.s. of (A.2) is greater than or equal to

$$\begin{aligned}
& \liminf_{n\to\infty} \inf_{\rho_B(\beta,\bar{\beta}_n)\geq\varepsilon} \left[R_n(\omega,\beta) - \bar{R}_n(\beta)\right] \\
& + \liminf_{n\to\infty} \inf_{\rho_B(\beta,\bar{\beta}_n)\geq\varepsilon} \left[\bar{R}_n(\beta) - \bar{R}_n(\bar{\beta}_n)\right] \\
& + \liminf_{n\to\infty} \left[\bar{R}_n(\bar{\beta}_n) - R_n(\omega,\bar{\beta}_n)\right] \\
\geq\ & \liminf_{n\to\infty} \inf_{\rho_B(\beta,\bar{\beta}_n)\geq\varepsilon} \left[\bar{R}_n(\beta) - \bar{R}_n(\bar{\beta}_n)\right] \\
& - 2\limsup_{n\to\infty} \sup_{B} \left|\bar{R}_n(\beta) - R_n(\omega,\beta)\right| \\
>\ & 0,
\end{aligned}$$

making use of (3.1) and (3.2). Now (A.2) implies that there exists a $\delta = \delta(\epsilon,\omega) > 0$ such that

$$\inf_{\rho_B(\beta,\bar{\beta}_n)\geq\varepsilon} R_n(\omega,\beta) - R_n(\omega,\bar{\beta}_n) \geq \delta$$

for n sufficiently large. Since $R_n(\omega,\hat{\beta}_n) - R_n(\omega,\bar{\beta}_n) \leq 0$ for n sufficiently large in view of (3.3), it follows that $\rho_B(\hat{\beta}_n,\bar{\beta}_n) < \epsilon$ for large n. The convergence in probability version of the lemma follows from a standard subsequence argument. ∎

Proof of Lemma 3.2. To prove the a.s. version of part (a) let $\Omega_0 \subseteq \Omega$ be a set of probability one on which $\rho_T(\hat{\tau}_n,\bar{\tau}_n) \to 0$ as $n \to \infty$ and on which (3.4) holds. Fix $\omega \in \Omega_0$. We now make use of Lemma A1 and define $\Phi = T \times B$ with metric $\sigma = \rho_T + \rho_B$, $\Lambda = B$ with metric $\kappa = \rho_B$, $g_n(.) = Q_n(\mathbf{z}_1(\omega),\ldots,\mathbf{z}_n(\omega),.,.)$, $\bar{g}_n = \bar{Q}_n$, $\varphi_n(\lambda) = (\hat{\tau}_n(\omega),\beta)$, $\bar{\varphi}_n(\lambda) = (\bar{\tau}_n,\beta)$, with $\lambda = \beta \in \Lambda$. Then $\rho_T(\hat{\tau}_n(\omega),\bar{\tau}_n) \to 0$ as $n \to \infty$ translates into condition (1) of Lemma A1, the condition that $\{\bar{Q}_n : n \in \mathbf{N}\}$ is uniformly equicontinuous on $T \times B$ in the lemma translates into condition (2)(i) of Lemma A1, and condition (3.4) of the lemma translates into condition (2)(ii) of Lemma A1. The almost sure convergence version of part (a) of the lemma now follows directly from Lemma A1(a). The convergence in probability version can then be obtained from a standard subsequence argument. Part (b) of the lemma follows immediately from Lemma A1(b) in view of the fact that $\{\bar{\varphi}_n : n \in \mathbf{N}\}$ is trivially uniformly equicontinuous on Λ. ∎

Proof of Lemma 3.3. To prove the a.s. version of part (a) let $\Omega_0 \subseteq \Omega$ be a set of probability one on which (3.5) holds. Fix $\omega \in \Omega_0$. We again make use of Lemma A1 and define $\Phi = C \times T \times B$ with metric $\sigma = |.| + \rho_T + \rho_B$, $\Lambda = T \times B$ with metric $\kappa = \rho_T + \rho_B$, $g_n = \bar{g}_n = \vartheta_n$, $\varphi_n(\lambda) = (S_n(\omega, \tau, \beta), \tau, \beta)$, $\bar{\varphi}_n(\lambda) = (\bar{S}_n(\tau, \beta), \tau, \beta)$ with $\lambda = (\tau, \beta) \in \Lambda$. Condition (1) of Lemma A1 follows from (3.5). Conditions (I) and (II) of the lemma imply, respectively, conditions (2)(i) and (2')(i) of Lemma A1 with $\Phi_* = K \times T \times B$, and conditions (2)(ii) and (2')(ii) are trivially satisfied with $\Phi_{**} = C \times T \times B$. The almost sure convergence version of part (a) of the lemma now follows directly from Lemma A1(a). The convergence in probability version can then be obtained from a standard subsequence argument. Part (b) of the lemma follows from Lemma A1(b) since $\{\bar{\varphi}_n : n \in \mathbf{N}\}$ is clearly equicontinuous [uniformly equicontinuous] on Λ if $\{\bar{S}_n : n \in \mathbf{N}\}$ has the respective property. ■

Lemma A3. *Let $(S, \mathfrak{S})$ be a measurable space, let Θ be a compact metrizable space and let $u : S \times \Theta \to \mathbf{R}$ be a function that is $\mathfrak{S}$-$\mathfrak{B}(\mathbf{R})$-measurable in its first argument for each $\theta \in \Theta$ and that is continuous on Θ in its second argument for each $s \in S$. Then there exists an $\mathfrak{S}$-$\mathfrak{B}(\Theta)$-measurable function $\tilde{\theta} : S \to \Theta$ such that*

$$u(s, \tilde{\theta}(s)) = \inf_{\theta \in \Theta} u(s, \theta)$$

for each $s \in S$ holds.

Proof. The proof is an adaptation of the proof of Lemma 3.3 in Schmetterer (1966, Ch.5) to the case of a general compact and metrizable space Θ. We start with two preparatory remarks. First, there exists a countable dense subset Θ_0 of Θ. By continuity of $u(s, .)$ we have

$$\inf_{\theta \in \Theta} u(s, \theta) = \inf_{\theta \in \Theta_0} u(s, \theta)$$

and hence $\inf_{\theta \in \Theta} u(s, \theta)$ is an $\mathfrak{S}$-measurable real valued function. Consequently, subtracting $\inf_{\theta \in \Theta} u(s, \theta)$ from $u(s, \theta)$ allows us to assume w.l.o.g. that

$$\inf_{\theta \in \Theta} u(s, \theta) = 0$$

for all $s \in S$. Second, in view of Urysohn's metrization theorem (see, e.g., Willard (1970)), we may assume w.l.o.g. that Θ is a compact subset of the product of countably many copies of the real line. The elements $\theta \in \Theta$ then have a representation of the form $\theta = (\theta_1, \theta_2, \theta_3, \ldots)$ with $\theta_i \in \mathbf{R}$, and we may introduce the lexicographic order on Θ, i.e., $\theta < \theta^*$ iff $\theta \neq \theta^*$ and $\theta_i < \theta_i^*$ holds for the smallest index i with $\theta_i \neq \theta_i^*$.

For each $s \in S$ let $M(s)$ denote the set $\{\theta \in \Theta : u(s, \theta) = 0\}$, i.e., the set of minimizers of $u(s, .)$. Since Θ is compact and $u(s, .)$ is continuous the set $M(s)$ is nonempty and compact. We now show that $M(s)$ contains a largest element $\tilde{\theta}(s)$ w.r.t. the lexicographic order. For $i \geq 1$ define

$$M_i(s) = \{\theta \in M_{i-1}(s) : \theta_i = \sup\{\theta_i \in \mathbf{R} : \theta \in M_{i-1}(s)\}\}$$

with $M_0(s) = M(s)$. Clearly, M_i is nonempty, compact and satisfies $M_i \subseteq M_{i-1}$ for $i \geq 1$. Consequently, the intersection of all sets M_i must contain at least one element, $\tilde{\theta}(s)$ say. By construction $\tilde{\theta}(s) \in M(s)$ and any $\theta \in M(s)$, $\theta \neq \tilde{\theta}(s)$ satisfies $\theta < \tilde{\theta}(s)$. (It follows that $\tilde{\theta}(s)$ is in fact the only element in the intersection of all the sets M_i.) Next we show that $\tilde{\theta}(s)$ is $\mathfrak{S}$-$\mathfrak{B}(\Theta)$-measurable. It suffices to establish $\mathfrak{S}$-measurability of each component $\tilde{\theta}_i(s)$. To this end let α denote an arbitrary real number and consider the set $\{s : \tilde{\theta}_1(s) \geq \alpha\}$ which can also be expressed as

$$\bigcup_{\theta \in \Theta, \theta_1 \geq \alpha} \{s : u(s, \theta) = 0\}$$

in view of the definition of $\tilde{\theta}(s)$. Making use of compactness of Θ and of continuity of $u(s, .)$, this latter set can now be rewritten as

$$\bigcap_{n \in \mathbf{N}} \bigcup_{\theta \in \Theta, \theta_1 > \alpha - 1/n} \{s : u(s, \theta) < 1/n\}$$

as is easily seen. In turn this set can also be written as

$$\bigcap_{n \in \mathbf{N}} \bigcup_{\theta \in \Theta_0, \theta_1 > \alpha - 1/n} \{s : u(s, \theta) < 1/n\}$$

where the countable set Θ_0 was defined earlier in the proof. This establishes $\mathfrak{S}$-measurability of $\tilde{\theta}_1(s)$, since countable unions/intersections of $\mathfrak{S}$-measurable sets are $\mathfrak{S}$-measurable. Next consider $\tilde{\theta}_2(s)$. The set $\{s : \tilde{\theta}_2(s) \geq \alpha\}$ can be expressed as

$$\bigcup_{\theta \in \Theta, \theta_2 \geq \alpha} \{s : u(s, \theta) = 0,\ \theta_1 = \tilde{\theta}_1(s)\}$$

which can also be written as

$$\bigcup_{\theta \in \Theta, \theta_2 \geq \alpha} \{s : u(s, \theta) = 0,\ \tilde{\theta}_1(s) \leq \theta_1\}$$

in view of the definition of $\tilde{\theta}(s)$. Now, similarly as before, the latter set can be written as

$$\bigcap_{n \in \mathbf{N}} \bigcup_{\theta \in \Theta, \theta_2 > \alpha - 1/n} \{s : u(s, \theta) < 1/n,\ \tilde{\theta}_1(s) < \theta_1 + 1/n\}$$

which equals

$$\bigcap_{n\in\mathbf{N}} \bigcup_{\theta\in\Theta_0,\theta_2>\alpha-1/n} \{s : u(s,\theta) < 1/n,\ \tilde{\theta}_1(s) < \theta_1 + 1/n\}.$$

This proves $\mathfrak{S}$-measurability of $\tilde{\theta}_2(s)$, observing again that countable unions /intersections of $\mathfrak{S}$-measurable sets are $\mathfrak{S}$-measurable. Repeating this type of argument one establishes $\mathfrak{S}$-measurability of $\tilde{\theta}_i(s)$ for every $i \geq 1$. ■

Proof of Lemma 3.4. Lemma 3.4 follows immediately from Lemma A3. ■

Appendix B

PROOFS FOR CHAPTER 4

Proof of Lemma 4.1. Define

$$f(\beta) = \liminf_{n\to\infty} \left[\bar{R}_n(\beta) - \bar{R}_n(\bar{\beta}_n)\right],$$

then $0 \leq f(\beta) \leq \infty$. We first show that f is continuous: By assumption $\{\bar{R}_n : n \in \mathbf{N}\}$ is equicontinuous on B. Hence for every $\beta^\bullet \in B$ and every $\eta > 0$ there exists a $\delta > 0$ such that $\rho_B(\beta, \beta^\bullet) < \delta$ implies

$$\sup_n \left|\bar{R}_n(\beta) - \bar{R}_n(\beta^\bullet)\right| < \eta.$$

Consequently,

$$\bar{R}_n(\beta) - \eta < \bar{R}_n(\beta^\bullet) < \bar{R}_n(\beta) + \eta$$

for all $n \geq 1$ and all $\beta \in B$ with $\rho_B(\beta, \beta^\bullet) < \delta$ and hence

$$f(\beta) - \eta \leq f(\beta^\bullet) \leq f(\beta) + \eta$$

for all $\beta \in B$ with $\rho_B(\beta, \beta^\bullet) < \delta$, thus establishing continuity of f. For the proof of the "only if" part of the lemma fix $\beta^\bullet \in B$, $\beta^\bullet \neq \bar{\beta}$. Then choose $\epsilon = \rho_B(\beta^\bullet, \bar{\beta})/2$. For $n \geq N(\epsilon)$ we then have $\rho_B(\bar{\beta}_n, \bar{\beta}) < \epsilon$ and $\rho_B(\beta^\bullet, \bar{\beta}_n) > \epsilon$. Consequently

$$f(\beta^\bullet) \geq \liminf_{n\to\infty} \left[\inf_{\{\beta \in B : \rho_B(\beta, \bar{\beta}_n) \geq \epsilon\}} \bar{R}_n(\beta) - \bar{R}_n(\bar{\beta}_n)\right] > 0.$$

To prove the "if" part of the lemma fix $\epsilon > 0$. Then for all $n \geq N_0$ we have

$$\left\{\beta \in B : \rho_B(\beta, \bar{\beta}) < \epsilon/2\right\} \subseteq \left\{\beta \in B : \rho_B(\beta, \bar{\beta}_n) < \epsilon\right\}$$

since $\bar{\beta}_n \to \bar{\beta}$ as $n \to \infty$. Define

$$c = \inf_{\{\beta \in B : \rho_B(\beta, \bar{\beta}) \geq \epsilon/2\}} f(\beta),$$

then $c > 0$ because of compactness of B and continuity of f. (If $\{\beta \in B : \rho_B(\beta, \bar{\beta}) \geq \epsilon/2\} = \emptyset$ then $c = \infty$.) Choose a constant $c^\bullet$ with $0 < c^\bullet < c$. Now since

$$\liminf_{n\to\infty} \left[\inf_{\{\beta \in B : \rho_B(\beta, \bar{\beta}_n) \geq \epsilon\}} \bar{R}_n(\beta) - \bar{R}_n(\bar{\beta}_n)\right]$$

$$\geq \liminf_{n\to\infty} \left[\inf_{\{\beta \in B : \rho_B(\beta, \bar{\beta}) \geq \epsilon/2\}} \bar{R}_n(\beta) - \bar{R}_n(\bar{\beta}_n)\right]$$

it suffices to show that

$$\inf_{\{\beta\in B:\rho_B(\beta,\bar{\beta})\geq\epsilon/2\}} \bar{R}_n(\beta) - \bar{R}_n(\bar{\beta}_n) \geq c^\bullet$$

for n large enough. Suppose this is not the case. Then there exists a subsequence (n_i) and $\beta_{n_i} \in B$ with $\rho_B(\beta_{n_i},\bar{\beta}) \geq \epsilon/2$ such that

$$\bar{R}_{n_i}(\beta_{n_i}) - \bar{R}_{n_i}(\bar{\beta}_{n_i}) < c^\bullet.$$

Since B is compact we may assume that β_{n_i} converges to β^*, say, and then clearly $\rho_B(\beta^*,\bar{\beta}) \geq \epsilon/2$ holds. From equicontinuity we obtain that

$$\left|\bar{R}_{n_i}(\beta_{n_i}) - \bar{R}_{n_i}(\beta^*)\right| \to 0,$$

hence

$$\liminf_{i\to\infty} \left[\bar{R}_{n_i}(\beta^*) - \bar{R}_{n_i}(\bar{\beta}_{n_i})\right] \leq c^\bullet.$$

Consequently,

$$f(\beta^*) \leq \liminf_{i\to\infty} \left[\bar{R}_{n_i}(\beta^*) - \bar{R}_{n_i}(\bar{\beta}_{n_i})\right] \leq c^\bullet < c$$

which is a contradiction. The statement in parenthesis is now obvious in view of equicontinuity and $\bar{\beta}_n \to \bar{\beta}$ as $n \to \infty$. ■

Proof of Lemma 4.2. Let $\Omega_0 \subseteq \Omega$ be a set of probability one on which (4.14) holds and (4.15) is satisfied for large n. Fix $\omega \in \Omega_0$. We first show that for every $\epsilon > 0$

$$\liminf_{n\to\infty} \left[\inf_{\{\beta\in B:\rho_B(\beta,\bar{B}_n)\geq\epsilon\}} R_n(\omega,\beta) - \sup_{\beta\in\bar{B}_n} R_n(\omega,\beta)\right] > 0 \tag{B.1}$$

holds. (Note that the expression in brackets on the l.h.s. of (B.1) is well-defined even if $\{\beta \in B : \rho_B(\beta,\bar{B}_n) \geq \epsilon\} = \emptyset$, since $\sup_{\beta\in\bar{B}_n} R_n(\omega,\beta) < \infty$ for large n in view of (4.14) and since $\bar{R}_n(\beta) \leq \bar{c}_n < \infty$ for $\beta \in \bar{B}_n$.) The inequality (B.1) follows since the l.h.s. of (B.1) is equal to

$$\begin{aligned}
&\liminf_{n\to\infty} \left\{ \inf_{\{\beta\in B:\rho_B(\beta,\bar{B}_n)\geq\epsilon\}} \left[R_n(\omega,\beta) - \bar{R}_n(\beta) + \bar{R}_n(\beta)\right] \right.\\
&\left. + \inf_{\beta\in\bar{B}_n} \left[-R_n(\omega,\beta) + \bar{R}_n(\beta) - \bar{R}_n(\beta)\right] \right\}\\
\geq\ &\liminf_{n\to\infty} \left\{ \inf_{\{\beta\in B:\rho_B(\beta,\bar{B}_n)\geq\epsilon\}} \left[R_n(\omega,\beta) - \bar{R}_n(\beta)\right] \right.\\
&+ \inf_{\{\beta\in B:\rho_B(\beta,\bar{B}_n)\geq\epsilon\}} \bar{R}_n(\beta)
\end{aligned}$$

$$
\begin{aligned}
&+ \inf_{\beta \in \bar{B}_n} \left[\bar{R}_n(\beta) - R_n(\omega,\beta)\right] - \sup_{\beta \in \bar{B}_n} \bar{R}_n(\beta) \Big\} \\
\geq\ & \liminf_{n\to\infty} \left[\inf_{\{\beta \in B: \rho_B(\beta,\bar{B}_n)\geq \epsilon\}} \bar{R}_n(\beta) - \sup_{\beta \in \bar{B}_n} \bar{R}_n(\beta) \right] \\
&- 2 \limsup_{n\to\infty} \sup_{\beta \in B} \left|\bar{R}_n(\beta) - R_n(\omega,\beta)\right| \\
>\ & 0
\end{aligned}
$$

making use of (4.13) and (4.14). Now (B.1) implies that there exists a $\bar{\delta} = \bar{\delta}(\epsilon,\omega) > 0$ such that

$$\inf_{\{\beta \in B: \rho_B(\beta,\bar{B}_n)\geq \epsilon\}} R_n(\omega,\beta) - \sup_{\beta \in \bar{B}_n} R_n(\omega,\beta) \geq \bar{\delta} > 0$$

for n sufficiently large. Since

$$R_n(\omega,\hat{\beta}_n) - \sup_{\beta \in \bar{B}_n} R_n(\omega,\beta) \leq \delta_n$$

eventually in view of (4.15) and $\delta_n \to 0$ as $n \to \infty$ it follows that

$$\rho_B(\hat{\beta}_n, \bar{B}_n) < \epsilon$$

for n sufficiently large. The convergence in probability version of the lemma follows from a standard subsequence argument. ■

Proof of Corollary 4.3. We prove the a.s. part first. From Lemma 4.2 we have $\rho_B(\hat{\beta}_n, \bar{B}_n) \to 0$ a.s. Hence there exists an $\tilde{\beta}_n = \tilde{\beta}_n(\omega) \in \bar{B}_n$ with $\rho_B(\hat{\beta}_n, \tilde{\beta}_n) \to 0$ a.s. (Note that it is immaterial here whether $\tilde{\beta}_n$ is measurable or not.) Now uniform equicontinuity of ζ_n on $\bigcup\{\bar{B}_n : n \in \mathbf{N}\}$ implies

$$\zeta_n(\hat{\beta}_n) - \bar{\zeta}_n = \zeta_n(\hat{\beta}_n) - \zeta_n(\tilde{\beta}_n) \to 0 \text{ a.s. as } n \to \infty.$$

The i.p. part follows now from the a.s. result by a standard subsequence argument. ■

Appendix C

PROOFS FOR CHAPTER 5

Lemma C1. *Let $(\mathbf{w}_t)_{t\in\mathbf{Z}}$ be a stochastic process on $(\Omega, \mathfrak{A}, P)$ with values in the Borel set $W \subseteq \mathbf{R}^{p_w}$. Let H_t^w be the distribution of $\mathbf{w}_t$ and assume that*

$$\left\{\bar{H}_n^w = n^{-1}\sum_{t=1}^{n} H_t^w : n \in \mathbf{N}\right\}$$

is tight on W. Let $\mathbf{z}_t = (\mathbf{w}'_t, \mathbf{w}'_{t-1}, \ldots)'$ and assume that it takes its values in Z, a relatively closed subset of $\prod_{i=0}^{\infty} W$. Furthermore let H_t^z be the distribution of $\mathbf{z}_t$. Then

$$\left\{\bar{H}_n^z = n^{-1}\sum_{t=1}^{n} H_t^z : n \in \mathbf{N}\right\}$$

is tight on Z.

Proof. Choose a sequence $a_j > 0$ with $\sum_{j=0}^{\infty} j a_j = 1$ and put $\epsilon_{j,m} = (2m)^{-1} a_j$ for all $m \in \mathbf{N}$. Since W is a Borel subset of $\mathbf{R}^{p_w}$ it follows from Theorem 3.2 in Parthasarathy (1967, p.29) that each H_t^w is tight on W. Hence

$$\tilde{H}_j^w = \left(H_0^w + H_{-1}^w + \ldots + H_{-(j-1)}^w\right)/j$$

is tight on W for each $j \in \mathbf{N}$. Furthermore, since $\{\bar{H}_n^w : n \in \mathbf{N}\}$ is tight on W, there exist compact sets $K_{j,m} \subseteq W$ such that

$$\bar{H}_n^w(W - K_{j,m}) < \epsilon_{j,m}$$

for all $n \in \mathbf{N}$ and

$$\tilde{H}_j^w(W - K_{j,m}) < \epsilon_{j,m}.$$

Now define

$$K_m = \left(\prod_{j=0}^{\infty} K_{j,m}\right) \cap Z,$$

which is compact since Z is closed in $\prod_{i=0}^{\infty} W$. Then for each $n \in \mathbf{N}$:

$$\bar{H}_n^z(Z - K_m) = n^{-1}\sum_{t=1}^{n} H_t^z(Z - K_m) = n^{-1}\sum_{t=1}^{n} P(\mathbf{z}_t \notin K_m)$$

$$\begin{aligned}
&= n^{-1}\sum_{t=1}^{n} P\left(\bigcup_{j=0}^{\infty}\{\mathbf{w}_{t-j} \notin K_{j,m}\}\right) \leq n^{-1}\sum_{t=1}^{n}\sum_{j=0}^{\infty} P\left(\mathbf{w}_{t-j} \notin K_{j,m}\right) \\
&= n^{-1}\sum_{j=0}^{\infty}\left[\sum_{j<t\leq n} P\left(\mathbf{w}_{t-j} \notin K_{j,m}\right) + \sum_{1\leq t\leq \min(j,n)} P\left(\mathbf{w}_{t-j} \notin K_{j,m}\right)\right] \\
&\leq \sum_{j=0}^{\infty}\left[n^{-1}\sum_{t=1}^{n} P\left(\mathbf{w}_{t} \notin K_{j,m}\right) + n^{-1}\sum_{t=1}^{j} P\left(\mathbf{w}_{t-j} \notin K_{j,m}\right)\right] \\
&= \sum_{j=0}^{\infty}\left[\bar{H}_n^w\left(W - K_{j,m}\right) + (j/n)\tilde{H}_j^w\left(W - K_{j,m}\right)\right] \\
&\leq \sum_{j=0}^{\infty}\left[\epsilon_{j,m} + (j/n)\epsilon_{j,m}\right] \leq 1/m.
\end{aligned}$$

Hence $\sup_n \bar{H}_n^z(Z - K_m) \to 0$ as $m \to \infty$. ■

Lemma C2. *Sufficient conditions for* $\{\bar{H}_n^w : n \in \mathbf{N}\}$ *in Lemma C1 to be tight on* W *are that* W *is closed and*

$$\sup_n n^{-1}\sum_{t=1}^{n} E\left[s\left(|\mathbf{w}_t|\right)\right] < \infty,$$

where $s : [0,\infty) \to [0,\infty)$ *is a monotone function with* $s(x) \to \infty$ *as* $x \to \infty$.

Proof. For the compact sets $K_m = \{w \in W : |w| \leq m\}$ we have using Markov's inequality

$$\begin{aligned}
\sup_n \bar{H}_n^w(W - K_m) &= \sup_n n^{-1}\sum_{t=1}^{n} P\left(\mathbf{w}_t \notin K_m\right) \\
&= \sup_n n^{-1}\sum_{t=1}^{n} P\left(|\mathbf{w}_t| > m\right) \\
&= \sup_n n^{-1}\sum_{t=1}^{n} P\left(s\left(|\mathbf{w}_t|\right) \geq s(m)\right) \\
&\leq \sup_n n^{-1}\sum_{t=1}^{n} E\left[s\left(|\mathbf{w}_t|\right)\right]/s(m)
\end{aligned}$$

whenever m is large enough such that $s(m)$ is positive. The result now follows for $m \to \infty$. ■

Lemma C3. *Let $(\mathbf{w}_t)_{t\in\mathbf{Z}}$ be a stochastic process on $(\Omega, \mathfrak{A}, P)$ with values in the Borel set $W \subseteq \mathbf{R}^{p_w}$. Let $H_t^{w,k}$ be the distribution of $(\mathbf{w}_t', ..., \mathbf{w}_{t-k}')'$ and assume that*

$$\bar{H}_n^{w,k} = n^{-1} \sum_{t=1}^{n} H_t^{w,k}$$

converges weakly to some probability measure $\bar{H}^{w,k}$ on $\prod_{i=0}^{k} W$ for each $k \geq 0$. Let $\mathbf{z}_t = (\mathbf{w}_t', \mathbf{w}_{t-1}', \ldots)'$ and assume that it takes its values in Z, a relatively closed subset of $\prod_{i=0}^{\infty} W$. Furthermore, let H_t^z denote the distribution of $\mathbf{z}_t$, then

$$\bar{H}_n^z = n^{-1} \sum_{t=1}^{n} H_t^z$$

converges weakly to some probability measure $\bar{H}^z$ on Z. Furthermore, $\bar{H}^z$ and each $\bar{H}_n^z$ are tight on Z.

Proof. The assumptions imply that the finite dimensional marginal distributions $\bar{H}_n^{w,k}$ of $\bar{H}_n^z$ converge weakly to $\bar{H}^{w,k}$ on $\prod_{i=0}^{k} W$ for each $k \geq 0$. Clearly this implies that $\bar{H}_n^z$ (viewed as a probability measure on $\prod_{i=0}^{\infty} W$) converges weakly to some probability measure $\bar{H}^z$ on $\prod_{i=0}^{\infty} W$.[1] Since

$$\bar{H}_n^z(Z) = 1$$

and Z is closed it follows from Theorem 2.1 in Billingsley (1968, p.11,12) that

$$\bar{H}^z(Z) = 1.$$

Hence Lemma 3 in Billingsley (1968, p.39) implies that $\bar{H}_n^z$ converges weakly to $\bar{H}^z$ on Z. Since Z is clearly a Borel subset of $\mathbf{R}^\infty$ individual tightness of $\bar{H}^z$ and of each $\bar{H}_n^z$ follows from Theorem 3.2 in Parthasarathy (1967, p.29). ■

Proof of Theorem 5.3. Note that the assumptions of Theorem 5.2 are satisfied with $K = 1$ and $r_{kt} \equiv 1$, since Assumption D' implies Assumption D. Hence the conclusions of Theorem 5.2 hold. It follows furthermore from Lemma 31 in Royden (1968, p.178) that $q(z, \theta)$ is continuous on $Z \times \Theta$. Theorems 43.7 and 43.14 in Willard (1970) imply further that $q_t(z, \theta)$ converges to $q(z, \theta)$ uniformly on compact subsets of $Z \times \Theta$. As a consequence of Assumption D' there exists for every $\epsilon > 0$ a compact set $K_* \subseteq Z$ such that

$$\sup_n \bar{H}_n^z(Z - K_*) < \epsilon.$$

[1] This can be proved completely analogously to the case of $W = \mathbf{R}$, cf. Billingsley (1968, pp.19,30,38).

Then

$$\limsup_{n\to\infty} \sup_{\Theta} \left| n^{-1} \sum_{t=1}^{n} [Eq_t(\mathbf{z}_t,\theta) - Eq(\mathbf{z}_t,\theta)] \right| = 0, \qquad (\text{C.1})$$

since the l.h.s. is less than or equal to

$$\begin{aligned}
& \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} \sup_{K_*\times\Theta} |q_t(z,\theta) - q(z,\theta)| \\
& + \limsup_{n\to\infty} \sup_{\Theta} n^{-1} \sum_{t=1}^{n} E\left[|q_t(\mathbf{z}_t,\theta) - q(\mathbf{z}_t,\theta)|\, \mathbf{1}_{Z-K_*}(\mathbf{z}_t)\right] \\
\leq\ & \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} E\left[d_t(\mathbf{z}_t)\mathbf{1}_{Z-K_*}(\mathbf{z}_t)\right] \\
& + \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} E\left[d(\mathbf{z}_t)\mathbf{1}_{Z-K_*}(\mathbf{z}_t)\right] \\
\leq\ & \text{const} * \left(\epsilon^{\gamma/(1+\gamma)} + \epsilon^{\delta/(1+\delta)}\right)
\end{aligned}$$

and ϵ was arbitrary. The first inequality holds since even

$$\lim_{t\to\infty} \sup_{K_*\times\Theta} |q_t(z,\theta) - q(z,\theta)| = 0.$$

The last inequality holds since by applying Hölder's inequality twice and by Assumption C(i) we have

$$\begin{aligned}
& n^{-1} \sum_{t=1}^{n} E\left[d_t(\mathbf{z}_t)\mathbf{1}_{Z-K_*}(\mathbf{z}_t)\right] \\
\leq\ & \left\{ n^{-1} \sum_{t=1}^{n} E\left[d_t(\mathbf{z}_t)^{1+\gamma}\right] \right\}^{1/(1+\gamma)} \left\{ n^{-1} \sum_{t=1}^{n} H_t^z(Z - K_*) \right\}^{\gamma/(1+\gamma)} \\
\leq\ & \text{const} * \epsilon^{\gamma/(1+\gamma)}
\end{aligned}$$

for all n, and since analogously the second term is less than or equal to const$*\epsilon^{\delta/(1+\delta)}$. Next we show that

$$\lim_{n\to\infty} \sup_{\Theta} \left| n^{-1} \sum_{t=1}^{n} Eq(\mathbf{z}_t,\theta) - \int q(z,\theta) d\bar{H}^z \right| = 0. \qquad (\text{C.2})$$

Since $\bar{H}_n^z$ converges weakly to $\bar{H}^z$, since $d(z)$ is clearly continuous (as $q(z,\theta)$ is so and Θ is compact) and

$$\sup_{n} n^{-1} \sum_{t=1}^{n} E\left[d(\mathbf{z}_t)^{1+\delta}\right] < \infty$$

for some $\delta > 0$, it follows analogously as in Theorem 5.4 of Billingsley (1968) that $\int d(z)d\bar{H}^z$ and hence $\int q(z,\theta)d\bar{H}^z$ exist and are finite. Since the family $\{q(z,\theta) : \theta \in \Theta\}$ is clearly equicontinuous on Z, since $q(z,\theta)$ is continuous on $Z \times \Theta$ and Θ is compact, the assumptions of Theorem 9.2 in Parthasarathy (1967, p.204) are satisfied and hence (C.2) holds. Continuity of $n^{-1}\sum_{t=1}^{n} Eq(\mathbf{z}_t,\theta)$ on Θ follows, e.g., from (C.1) and Theorem 5.2(b), and continuity of $\int q(z,\theta)d\bar{H}^z$ follows then from (C.2). The convergence result (5.1) follows from (C.1), (C.2) and Theorem 5.2. ∎

Appendix D

PROOFS FOR CHAPTER 6

Proof of Theorem 6.1. (a) Since the case $p^\bullet = 0$ is trivial assume $p^\bullet > 0$. If $0 < p < p^\bullet$ then the claim follows immediately from Lyapunov's inequality. If $p = 0$ then using Markov's inequality and Lyapunov's inequality we have for $r = \min\{1, p^\bullet\}$:

$$\begin{aligned}
& \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} P\left(|\mathbf{v}_t - \mathbf{h}_t^m| > \delta\right) \\
\leq\ & \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} \|\mathbf{v}_t - \mathbf{h}_t^m\|_r^r / \delta^r \\
\leq\ & \delta^{-r} \left[\limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} \|\mathbf{v}_t - \mathbf{h}_t^m\|_r\right]^r \\
\leq\ & \delta^{-r} \left[\limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} \|\mathbf{v}_t - \mathbf{h}_t^m\|_{p^\bullet}\right]^r = 0.
\end{aligned}$$

Note that this proof also demonstrates that the $L_{p^\bullet}$-approximator $\mathbf{h}_t^m$ is also an L_p-approximator for $0 \leq p \leq p^\bullet$.

(b) Since the case $p = 0$ is trivial assume $0 < p < p^\bullet$. Let $\mathbf{h}_t^m$ be some L_0-approximator, then there exists a sequence $\delta_m > 0$, $\delta_m \to 0$ as $m \to \infty$, such that

$$\limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} P\left(|\mathbf{v}_t - \mathbf{h}_t^m| > \delta_m\right) \to 0 \text{ as } m \to \infty$$

(since the process $(\mathbf{v}_t)$ is L_0-approximable). Let r, s, $r^\bullet$, $s^\bullet$ be defined as follows: $r = p^\bullet/p$ and $1/r + 1/s = 1$, which implies $1 < s < \infty$ since $1 < r < \infty$; now choose $r^\bullet$ such that $1 < r^\bullet < 1 + \gamma$ (where $\gamma > 0$ is given in the theorem) and furthermore small enough such that $sp \leq s^\bullet$ where $1/r^\bullet + 1/s^\bullet = 1$. Fix an arbitrary $v_* \in \mathbf{R}^{p_v}$. Then choose a sequence M_m of real numbers with $M_m \geq |v_*|$, $M_m \to \infty$ as $m \to \infty$ satisfying

$$M_m^{s^\bullet} \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} P\left(|\mathbf{v}_t - \mathbf{h}_t^m| > \delta_m\right) \to 0 \text{ as } m \to \infty, \tag{D.1}$$

which is clearly possible. For notational convenience put

$$K = \left[\sup_n n^{-1} \sum_{t=1}^{n} \|\mathbf{v}_t\|_{p^\bullet}^{1+\gamma} \right]^{1/(1+\gamma)} .$$

Next define the measurable function $\tilde{h}_t^m : \mathbf{R}^{(2m+1)p_e} \to \mathbf{R}^{p_v}$ to be equal to h_t^m if $|h_t^m(.)| \leq M_m$ and to be equal to v_* if $|h_t^m(.)| > M_m$. Then define

$$\tilde{\mathbf{h}}_t^m = \tilde{h}_t^m(\mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m})$$

and observe that

$$\left|\tilde{\mathbf{h}}_t^m\right| \leq M_m.$$

We now show that $\tilde{\mathbf{h}}_t^m$ is an L_p-approximator for $(\mathbf{v}_t)$. Since

$$\left\|\mathbf{v}_t - \tilde{\mathbf{h}}_t^m\right\|_p \leq 3^{1/p}(A_{t,m}^1 + A_{t,m}^2 + A_{t,m}^3)$$

with

$$\begin{aligned} A_{t,m}^1 &= \left\|(\mathbf{v}_t - \tilde{\mathbf{h}}_t^m)\mathbf{1}_{\{|\mathbf{v}_t - \mathbf{h}_t^m| > \delta_m\}}\right\|_p, \\ A_{t,m}^2 &= \left\|(\mathbf{v}_t - \tilde{\mathbf{h}}_t^m)\mathbf{1}_{\{|\mathbf{v}_t - \mathbf{h}_t^m| \leq \delta_m, |\mathbf{h}_t^m| > M_m\}}\right\|_p, \\ A_{t,m}^3 &= \left\|(\mathbf{v}_t - \tilde{\mathbf{h}}_t^m)\mathbf{1}_{\{|\mathbf{v}_t - \mathbf{h}_t^m| \leq \delta_m, |\mathbf{h}_t^m| \leq M_m\}}\right\|_p, \end{aligned}$$

it suffices to show that

$$\limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} A_{t,m}^i \to 0 \text{ as } m \to \infty \quad \text{(i=1,2,3)}. \tag{D.2}$$

By applying Hölder's inequality first with (r, s), then with $(r^\bullet, s^\bullet)$, using Lyapunov's inequality and observing that $r^\bullet < 1 + \gamma$ and $p^\bullet = pr$ we get

$$\begin{aligned} &\limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} A_{t,m}^1 \\ \leq\ & \limsup_{n\to\infty} \left\{ \left[n^{-1} \sum_{t=1}^{n} \left\|\mathbf{v}_t - \tilde{\mathbf{h}}_t^m\right\|_{pr}^{r^\bullet} \right]^{1/r^\bullet} \right. \\ & \left. \left[n^{-1} \sum_{t=1}^{n} \left\|\mathbf{1}_{\{|\mathbf{v}_t - \mathbf{h}_t^m| > \delta_m\}}\right\|_{ps}^{s^\bullet} \right]^{1/s^\bullet} \right\} \\ \leq\ & \left[\limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} \left\|\mathbf{v}_t - \tilde{\mathbf{h}}_t^m\right\|_{p^\bullet}^{1+\gamma} \right]^{1/(1+\gamma)} \\ & \left[\limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} \left[P\left(|\mathbf{v}_t - \mathbf{h}_t^m| > \delta_m\right)\right]^{s^\bullet/(ps)} \right]^{1/s^\bullet} . \end{aligned}$$

Since

$$\begin{aligned}\left\|\mathbf{v}_t - \tilde{\mathbf{h}}_t^m\right\|_{p^\bullet}^{1+\gamma} &\leq 2^{(1+\gamma)/p^\bullet+\gamma}\left[\|\mathbf{v}_t\|_{p^\bullet}^{1+\gamma} + \left\|\tilde{\mathbf{h}}_t^m\right\|_{p^\bullet}^{1+\gamma}\right] \\ &\leq 2^{(1+\gamma)/p^\bullet+\gamma}\left[\|\mathbf{v}_t\|_{p^\bullet}^{1+\gamma} + M_m^{1+\gamma}\right]\end{aligned}$$

and since $s^\bullet/(ps) \geq 1$ we get

$$\begin{aligned}&\limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} A_{t,m}^1 \\ \leq\ & 2^{1/p^\bullet+\gamma/(1+\gamma)}(K+M_m)\left[\limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} P\left(|\mathbf{v}_t - \mathbf{h}_t^m| > \delta_m\right)\right]^{1/s^\bullet}.\end{aligned}$$

Now (D.2) with $i=1$ follows from $K<\infty$ and (D.1).

Analogously to the above argument and then using Markov's inequality with $\alpha = \min\{1, p^\bullet\}$ we get

$$\begin{aligned}&\limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} A_{t,m}^2 \\ =\ & \limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} \left\|(\mathbf{v}_t - v_*)\mathbf{1}_{\{|\mathbf{v}_t-\mathbf{h}_t^m|\leq\delta_m, |\mathbf{h}_t^m|>M_m\}}\right\|_p \\ \leq\ & 2^{1/p^\bullet+\gamma/(1+\gamma)}(K+|v_*|) \\ & \left[\limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} P\left(|\mathbf{v}_t - \mathbf{h}_t^m| \leq \delta_m, |\mathbf{h}_t^m| > M_m\right)\right]^{1/s^\bullet} \\ \leq\ & 2^{1/p^\bullet+\gamma/(1+\gamma)}(K+|v_*|)\left[\limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} P\left(|\mathbf{v}_t| \geq M_m - \delta_m\right)\right]^{1/s^\bullet} \\ \leq\ & 2^{1/p^\bullet+\gamma/(1+\gamma)}(K+|v_*|)\left[\limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} E\left|\mathbf{v}_t\right|^\alpha/(M_m-\delta_m)^\alpha\right]^{1/s^\bullet}.\end{aligned}$$

Now (D.2) with $i=2$ follows from $M_m \to \infty$ as $m \to \infty$ if we can show that

$$\limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} E\left|\mathbf{v}_t\right|^\alpha < \infty.$$

But by applying Lyapunov's inequality twice and observing that $\alpha = \min\{1, p^\bullet\} \leq 1+\gamma$ we have

$$\limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} E\left|\mathbf{v}_t\right|^\alpha \leq K^\alpha < \infty.$$

Condition (D.2) with $i = 3$ follows since $A^3_{t,m} \leq \delta_m$ keeping in mind the definition of $\tilde{h}^m_t$.

(c) For any $M > 0$ define $g_M(x) = 1$ if $|x| \leq M$ and $g_M(x) = M+1-|x|$ for $M < |x| < M+1$, and $g_M(x) = 0$ for $|x| \geq M+1$. We now show that

$$\mathbf{v}_{t,M} = g_M(|\mathbf{v}_t|)\,\mathbf{v}_t$$

is L_0-approximable with approximators

$$\mathbf{h}^m_{t,M} = g_M(|\mathbf{h}^m_t|)\,\mathbf{h}^m_t$$

(which of course is a measurable function of $\mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m}$). Since $g_M(|v|)v$ is uniformly continuous on $\mathbf{R}^{p_v}$ there exists for every $\delta > 0$ an $\eta > 0$ such that

$$|g_M(|v|)\,v - g_M(|v^\bullet|)\,v^\bullet| < \delta$$

whenever $|v - v^\bullet| \leq \eta$. Consequently,

$$\begin{aligned} & P\left(\left|\mathbf{v}_{t,M} - \mathbf{h}^m_{t,M}\right| > \delta\right) \\ \leq \;& P(|\mathbf{v}_t - \mathbf{h}^m_t| > \eta) + P\left(|\mathbf{v}_t - \mathbf{h}^m_t| \leq \eta, \left|\mathbf{v}_{t,M} - \mathbf{h}^m_{t,M}\right| > \delta\right) \\ = \;& P(|\mathbf{v}_t - \mathbf{h}^m_t| > \eta) \end{aligned}$$

for all t and m. Hence the L_0-approximability of $\mathbf{v}_{t,M}$ follows from the L_0-approximability of $\mathbf{v}_t$. Since $|\mathbf{v}_{t,M}| \leq M+1$ it follows from part (b) of the theorem that $\mathbf{v}_{t,M}$ is L_p-approximable for any $0 \leq p < \infty$. In particular for $p = 2$ it follows that the conditional mean $E(\mathbf{v}_{t,M} \mid \mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m})$ is an L_2-approximator for $\mathbf{v}_{t,M}$. Define

$$\mathbf{u}_{t,M} = \mathbf{v}_t - \mathbf{v}_{t,M},$$

then clearly

$$|\mathbf{u}_{t,M}| \leq |\mathbf{v}_t|\,\mathbf{1}_{\{|\mathbf{v}_t| \geq M\}}.$$

We now show that $E(\mathbf{v}_t \mid \mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m})$ is an L_p-approximator for $\mathbf{v}_t$ where $0 < p < p^\bullet$. Since

$$\|\mathbf{v}_t - E(\mathbf{v}_t \mid \mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m})\|_p \leq 2^{1/p}\left(B^1_{t,m,M} + B^2_{t,m,M}\right)$$

with

$$B^1_{t,m,M} = \|\mathbf{v}_{t,M} - E(\mathbf{v}_{t,M} \mid \mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m})\|_p$$

and

$$B^2_{t,m,M} = \|\mathbf{u}_{t,M} - E(\mathbf{u}_{t,M} \mid \mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m})\|_p$$

it suffices to show that

$$\limsup_{M\to\infty}\,\limsup_{m\to\infty}\,\limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} B^i_{t,m,M} = 0, \quad (i = 1, 2). \tag{D.3}$$

Note that for any random variable ξ with $|\xi| \leq c$ we have $\|\xi\|_p \leq \|\xi\|_2$ if $p \leq 2$ and $\|\xi\|_p \leq c^{(p-2)/p} \|\xi\|_2^{2/p}$ if $p > 2$. Since

$$|\mathbf{v}_{t,M} - E(\mathbf{v}_{t,M} \mid \mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m})| \leq 2(M+1)$$

it follows for $p \leq 2$ that

$$\begin{aligned} & \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} B^1_{t,m,M} \\ \leq\ & \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} \|\mathbf{v}_{t,M} - E(\mathbf{v}_{t,M} \mid \mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m})\|_2 \end{aligned}$$

and for $p > 2$ that

$$\limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} B^1_{t,m,M} \leq [2M+2]^{1-2/p}$$
$$\left[\limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} \|\mathbf{v}_{t,M} - E\left(\mathbf{v}_{t,M} \mid \mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m}\right)\|_2\right]^{2/p}$$

using Lyapunov's inequality and observing that $2/p < 1$. For $i = 1$ condition (D.3) now follows since $\mathbf{v}_{t,M}$ was shown to be L_2-approximable for any M. Next choose $q \geq \max\{1, p\}$ such that $q < p^\bullet$ and $p^\bullet/q \leq 1+\gamma$, which is clearly possible. Then using Lyapunov's inequality, the triangle inequality and the conditional Jensen inequality we get

$$\begin{aligned} B^2_{t,m,M} &\leq 2\|\mathbf{u}_{t,M}\|_q \leq 2\left\| |\mathbf{v}_t| \, \mathbf{1}_{\{|\mathbf{v}_t| \geq M\}} \right\|_q \\ &\leq 2\left\| |\mathbf{v}_t|^{1+(p^\bullet - q)/q} M^{(q-p^\bullet)/q} \right\|_q . \end{aligned}$$

Consequently

$$\begin{aligned} & \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} B^2_{t,m,M} \\ \leq\ & 2M^{(q-p^\bullet)/q} \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} \|\mathbf{v}_t\|_{p^\bullet}^{p^\bullet/q} \\ \leq\ & 2M^{(q-p^\bullet)/q} \left[\limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} \|\mathbf{v}_t\|_{p^\bullet}^{1+\gamma}\right]^{p^\bullet/[q(1+\gamma)]} \end{aligned}$$

by Lyapunov's inequality observing that $p^\bullet/q \leq 1+\gamma$. Condition (D.3) for $i = 2$ now follows since $K < \infty$ and observing that $q - p^\bullet < 0$. That the conditional mean is also an L_0-approximator follows now from part (a) of the theorem. ■

Lemma D1. *Let* $(\mathbf{v}_t)_{t\in\mathbf{N}}$ *and* $(\mathbf{v}_t^m)_{t\in\mathbf{N}}$ *for all* $m \in \mathbf{N}$ *be stochastic processes on* $(\Omega, \mathfrak{A}, P)$ *taking their values in* $\mathbf{R}^{p_v}$, *with* $E|\mathbf{v}_t| < \infty$, $E|\mathbf{v}_t^m| < \infty$. *Suppose that for every* $\delta > 0$

$$\limsup_{n\to\infty} P\left(\left|n^{-1}\sum_{t=1}^{n}(\mathbf{v}_t - \mathbf{v}_t^m)\right| > \delta\right) \to 0 \text{ as } m \to \infty, \tag{D.4}$$

and

$$\limsup_{n\to\infty} \left|n^{-1}\sum_{t=1}^{n} E(\mathbf{v}_t - \mathbf{v}_t^m)\right| \to 0 \text{ as } m \to \infty, \tag{D.5}$$

then given $(\mathbf{v}_t^m)$ *satisfies a weak LLN for each* $m \in \mathbf{N}$ *also* $(\mathbf{v}_t)$ *satisfies a weak LLN.*

Proof. First note that for a sequence of random variables (ξ_n) we have for any $\delta > 0$:

$$\begin{aligned}\limsup_{n\to\infty} P(|\xi_n| > \delta) &\leq \lim_{k\to\infty} P\left(\sup_{n\geq k}|\xi_n| > \delta\right) \\ &\leq \lim_{k\to\infty} P\left(\sup_{n\geq k}|\xi_n| \geq \delta\right) \\ &\leq P\left(\limsup_{n\to\infty}|\xi_n| \geq \delta\right),\end{aligned}$$

where the last inequality holds since the ω-sets on which $\sup_{n\geq k}|\xi_n| \geq \delta$ decrease monotonically to the ω-set where $\limsup_{n\to\infty}|\xi_n| \geq \delta$ as $k \to \infty$. Then clearly

$$\begin{aligned}&\limsup_{n\to\infty} P\left(\left|n^{-1}\sum_{t=1}^{n}(\mathbf{v}_t - E\mathbf{v}_t)\right| > \delta\right) \\ \leq\ &\limsup_{n\to\infty} P\left(\left|n^{-1}\sum_{t=1}^{n}(\mathbf{v}_t - \mathbf{v}_t^m)\right| > \delta/3\right) \\ &+ \limsup_{n\to\infty} P\left(\left|n^{-1}\sum_{t=1}^{n}E(\mathbf{v}_t - \mathbf{v}_t^m)\right| > \delta/3\right) \\ &+ \limsup_{n\to\infty} P\left(\left|n^{-1}\sum_{t=1}^{n}(\mathbf{v}_t^m - E\mathbf{v}_t^m)\right| > \delta/3\right) \\ \leq\ &\limsup_{n\to\infty} P\left(\left|n^{-1}\sum_{t=1}^{n}(\mathbf{v}_t - \mathbf{v}_t^m)\right| > \delta/3\right) \\ &+ P\left(\limsup_{n\to\infty}\left|n^{-1}\sum_{t=1}^{n}E(\mathbf{v}_t - \mathbf{v}_t^m)\right| \geq \delta/3\right)\end{aligned}$$

because the process $(\mathbf{v}_t^m)$ satisfies a LLN. The first term on the r.h.s. of the last of the above inequalities goes to zero as $m \to \infty$ because of (D.4) and the second term actually becomes zero because of (D.5). ∎

Lemma D2. *Let the process $(\mathbf{v}_t)_{t \in \mathbf{N}}$ be L_0-approximable by $(\mathbf{e}_t)_{t \in \mathbf{Z}}$. Suppose*

$$\sup_n n^{-1} \sum_{t=1}^{n} \|\mathbf{v}_t\|_{p^\bullet}^{1+\gamma} < \infty$$

for some $0 < p^\bullet < \infty$ and $\gamma > 0$. Then for each p, $0 \leq p < p^\bullet$, there exist L_p-approximators $\mathbf{h}_t^m$ that are bounded in absolute value uniformly in t, for each $m \in \mathbf{N}$ (that is, for some constants M_m we have $\sup_t |\mathbf{h}_t^m| \leq M_m < \infty$).

Proof. A specific construction of such L_p-approximators, $0 < p < p^\bullet$, was given in the proof of part (b) of Theorem 6.1. For $p = 0$ the lemma follows then from part (a) of Theorem 6.1. ∎

Remark. As a point of interest we note that if the moment condition in the above lemma is weakened to tightness of $\{\bar{H}_n^v = n^{-1} \sum_{t=1}^{n} H_t^v : n \in \mathbf{N}\}$ on $\mathbf{R}^{p_v}$ where H_t^v is the distribution of $\mathbf{v}_t$, then still the existence of uniformly bounded L_0-approximators can be established.

Proof of Theorem 6.2. To prove part (a) it suffices to show that conditions (D.4) and (D.5) of Lemma D1 with $\mathbf{v}_t^m = \mathbf{h}_t^m$ are satisfied. (D.4) follows from L_1-approximability via Markov's inequality. (D.5) is obviously implied by L_1-approximability. Also note that $E|\mathbf{h}_t^m| < \infty$ follows since $(\mathbf{h}_t^m)$ satisfies a LLN. Furthermore, since $(\mathbf{v}_t)$ is L_1-approximable with L_1-approximator $\mathbf{h}_t^m$, it follows that $E|\mathbf{v}_t| < \infty$. Part (b) follows from Lemma D2 (with $p^\bullet = 1 + \gamma = 1 + \epsilon$ and $p = 1$) since L_1-approximability implies L_0-approximability. ∎

Proof of Theorem 6.3. Because of Theorem 6.1 and the maintained moment condition $(\mathbf{v}_t)$ is L_1-approximable. Because of part (b) of Theorem 6.2 there exist L_1-approximators $\mathbf{h}_t^m = h_t^m(\mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m})$ which are bounded in absolute value by, say, M_m uniformly in t, for each $m \in \mathbf{N}$. Because of part (a) of Theorem 6.2 it suffices to show that the processes $(\mathbf{h}_t^m)$ satisfy a weak LLN for each $m \in \mathbf{N}$. Since $(\mathbf{e}_t)$ is α-mixing it follows that $(\mathbf{h}_t^m)$ is α-mixing, observing that $h_t^m(.)$ is measurable. Let $\mathbf{h}_{ti}^m$ denote the i-th component of $\mathbf{h}_t^m$ and let (for fixed i and m) α_j with $\alpha_j \to 0$ as $j \to \infty$ be the corresponding α-mixing coefficients. Then

$$\operatorname{var}\left(n^{-1} \sum_{t=1}^{n} \mathbf{h}_{ti}^m\right)$$

$$\begin{aligned}
&= n^{-2}\sum_{t=1}^{n}\operatorname{var}(\mathbf{h}_{ti}^{m}) + 2n^{-2}\sum_{1\le t<s\le n}\operatorname{cov}(\mathbf{h}_{ti}^{m},\mathbf{h}_{si}^{m}) \\
&\le 4M_m^2 n^{-1} + 8\left(2^{1/2}+1\right)M_m^2 n^{-2}\sum_{1\le t<s\le n}\alpha_{s-t}^{1/2} \\
&\le 24M_m^2 n^{-1}\left(1+\sum_{j=1}^{n}\alpha_j^{1/2}\right) \to 0
\end{aligned}$$

since $\alpha_j \to 0$ as $j \to \infty$; hence $(\mathbf{h}_{ti}^{m})$ satisfies a weak LLN. The first inequality in the above chain uses the inequality

$$\left|\operatorname{cov}(\mathbf{h}_{ti}^{m},\mathbf{h}_{si}^{m})\right| \le 4\left(2^{1/2}+1\right)M_m^2\alpha_{s-t}^{1/2};$$

cf. for example McLeish (1975a), Lemma 2.1. ■

Lemma D3. *Let $(\mathbf{v}_t)$ be a process defined on $(\Omega,\mathfrak{A},P)$ that takes its values in a Borel subset V of $\mathbf{R}^{p_v}$. Suppose $(\mathbf{v}_t)$ is L_p-approximable by $(\mathbf{e}_t)$ for some $0 \le p < \infty$, then there exists a sequence of L_p-approximators that take their values in V.*

Proof. Since V, as a subset of $\mathbf{R}^{p_v}$, is separable, there exists a countable dense subset of V, say $\{v^i : i \in \mathbf{N}\}$. By assumption there exist L_p-approximators $\mathbf{h}_t^m = h_t^m(\mathbf{e}_{t+m},\ldots,\mathbf{e}_{t-m})$ where $h_t^m(x)$ with $x \in \mathbf{R}^{(2m+1)p_e}$ takes its values in $\mathbf{R}^{p_v}$ and is Borel measurable. Choose some sequence $\delta_m > 0$, $\delta_m \to 0$ as $m \to \infty$. Define $\bar{h}_t^m(x) = h_t^m(x)$ for all x such that $h_t^m(x) \in V$; for any x such that $h_t^m(x) \notin V$ define $\bar{h}_t^m(x) = v^i$ where i is the first index satisfying

$$\left|v^i - h_t^m(x)\right| < \inf_j \left|v^j - h_t^m(x)\right| + \delta_m.$$

Clearly $\bar{h}_t^m$ takes its values in V and is Borel measurable. If $h_t^m(x) \notin V$, then

$$\begin{aligned}
\left|\bar{h}_t^m(x) - h_t^m(x)\right| &= \left|v^i - h_t^m(x)\right| \\
&< \inf_j \left|v^j - h_t^m(x)\right| + \delta_m \\
&= \inf_{v\in V}\left|v - h_t^m(x)\right| + \delta_m,
\end{aligned}$$

since $\{v^i : i \in \mathbf{N}\}$ is dense. If $h_t^m(x) \in V$, then $\left|\bar{h}_t^m(x) - h_t^m(x)\right| = 0$. Now let $\bar{\mathbf{h}}_t^m = \bar{h}_t^m(\mathbf{e}_{t+m},\ldots,\mathbf{e}_{t-m})$, then it follows that

$$\left|\mathbf{v}_t - \bar{\mathbf{h}}_t^m\right| \le \left|\mathbf{v}_t - \mathbf{h}_t^m\right| + \left|\mathbf{h}_t^m - \bar{\mathbf{h}}_t^m\right| \le 2\left|\mathbf{v}_t - \mathbf{h}_t^m\right| + \delta_m.$$

From this the result follows if we observe that

$$\left\|\mathbf{v}_t - \bar{\mathbf{h}}_t^m\right\|_p \leq 2^{1/p}\left(2\left\|\mathbf{v}_t - \mathbf{h}_t^m\right\|_p + \delta_m\right)$$

for $0 < p < \infty$ and

$$P\left(\left|\mathbf{v}_t - \bar{\mathbf{h}}_t^m\right| > \delta\right) \leq P\left(\left|\mathbf{v}_t - \mathbf{h}_t^m\right| > \delta/4\right)$$

for m large enough (since $\delta_m < \delta/2$ eventually). ■

Remark. Inspection of the proof of Lemma D3 shows further that if there exist L_p-approximators $\mathbf{h}_t^m$ $(p > 0)$ such that $\sup_t \left\|\mathbf{v}_t - \mathbf{h}_t^m\right\|_p$ is of size $-q$, then there exist L_p-approximators $\bar{\mathbf{h}}_t^m$ that take their values in V such that $\sup_t \left\|\mathbf{v}_t - \bar{\mathbf{h}}_t^m\right\|_p$ is also of size $-q$. This follows since the size of δ_m can be chosen arbitrarily.

In the following lemma let $g_t : V \to \mathbf{R}$ and let G denote the set of all points $v \in V$ at which $\{g_t : t \in \mathbf{N}\}$ is not equicontinuous. Observe that $G = \emptyset$ is equivalent to the condition that Assumption 6.1 holds.

Lemma D4. *Let $(\mathbf{v}_t)$ be a process defined on $(\Omega, \mathfrak{A}, P)$ that takes its values in a Borel subset V of $\mathbf{R}^{p_v}$. Suppose $(\mathbf{v}_t)$ is L_0-approximable by $(\mathbf{e}_t)$. Suppose further that*

(a) $\{\bar{H}_n^v : n \in \mathbf{N}\}$ is tight on V and there exists a sequence of open sets G_k with $G \subseteq G_k \subseteq V$ and

$$\lim_{k\to\infty} \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} P\left(\mathbf{v}_t \in G_k\right) = 0,$$

or

(b) Assumption 6.2 holds and

$$\limsup_{m\to\infty} \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} E\left[B_t(\mathbf{v}_t, \mathbf{h}_t^m)^\epsilon\right] < \infty$$

for some $\epsilon > 0$ and for some L_0-approximators $\mathbf{h}_t^m$ of $(\mathbf{v}_t)$ based on $(\mathbf{e}_t)$, where the approximators take their values in V.

Then $(g_t(\mathbf{v}_t))$ is L_0-approximable by $(\mathbf{e}_t)$.

Proof. (a) Fix $\delta > 0$ and choose $0 < \eta < \min\{1, \delta/3\}$. Because $\{\bar{H}_n^v : n \in \mathbf{N}\}$ is tight on V there exists a compact set $K_1 \subseteq V$ such that

$$\sup_n n^{-1} \sum_{t=1}^{n} P\left(\mathbf{v}_t \notin K_1\right) < \eta/2.$$

Furthermore, there exists a $k_0 \in \mathbf{N}$ such that

$$\limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} P\left(\mathbf{v}_t \in G_{k_0}\right) < \eta/2.$$

Put $K = K_1 - G_{k_o}$ which is nonempty since

$$P\left(\mathbf{v}_t \notin K\right) \leq P\left(\mathbf{v}_t \notin K_1\right) + P\left(\mathbf{v}_t \in G_{k_o}\right)$$

implies

$$\limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} P\left(\mathbf{v}_t \notin K\right) \leq \eta.$$

Clearly K is also compact. We next show that we can find a $\gamma > 0$ such that

$$\sup_t |g_t(v) - g_t(v^\bullet)| < \eta$$

whenever $v \in K$, $v^\bullet \in V$ and $|v - v^\bullet| \leq \gamma$. Since $K \cap G = \emptyset$ by construction it follows that for each $v \in K$ there exists a $\gamma(v) > 0$ such that for $v^\bullet \in V$ and $|v - v^\bullet| \leq \gamma(v)$ we have

$$\sup_t |g_t(v) - g_t(v^\bullet)| < \eta/2.$$

Now cover K by all balls

$$B\left(v, \gamma(v)/2\right) = \{v^\bullet \in V : |v - v^\bullet| \leq \gamma(v)/2\}$$

with $v \in K$. Since the interiors of these balls (relative to V) cover K and since K is compact we can find finitely many elements $v^1, \ldots, v^l$ of K such that

$$K \subseteq \bigcup_{i=1}^{l} B\left(v^i, \gamma(v^i)/2\right).$$

Now define $\gamma = \min\{\gamma(v^1), \ldots, \gamma(v^l)\}/2$. Let $v \in K$ and $v^\bullet \in V$ with $|v - v^\bullet| \leq \gamma$, then there exists an index i_0 such that $\left|v - v^{i_0}\right| \leq \gamma(v^{i_0})/2$ and hence $\left|v^\bullet - v^{i_0}\right| \leq \gamma + \gamma(v^{i_0})/2 \leq \gamma(v^{i_0})$. Hence by definition of $\gamma(v^{i_0})$ we have

$$\sup_t \left|g_t(v) - g_t(v^{i_0})\right| < \eta/2$$

and

$$\sup_t \left|g_t(v^\bullet) - g_t(v^{i_0})\right| < \eta/2,$$

which by the triangle inequality establishes the above claim.

Now let $\mathbf{h}_t^m$ be L_0-approximators which w.l.o.g. can be assumed to take their values in V because of Lemma D3. Decompose Ω into three disjoint sets:

$$\begin{aligned} A^1_{t,m} &= \{|\mathbf{v}_t - \mathbf{h}^m_t| > \gamma\} \cap \{\mathbf{v}_t \in K\}, \\ A^2_{t,m} &= \{|\mathbf{v}_t - \mathbf{h}^m_t| \leq \gamma\} \cap \{\mathbf{v}_t \in K\}, \\ A^3_{t,m} &= \{\mathbf{v}_t \notin K\}. \end{aligned}$$

Then

$$\begin{aligned} & P\left(|g_t(\mathbf{v}_t) - g_t(\mathbf{h}^m_t)| > \delta\right) \\ \leq\ & \sum_{i=1}^{3} P\left(|g_t(\mathbf{v}_t) - g_t(\mathbf{h}^m_t)|\, \mathbf{1}_{A^i_{t,m}} > \delta/3\right) \\ \leq\ & P\left(|\mathbf{v}_t - \mathbf{h}^m_t| > \gamma\right) + P(\emptyset) + P\left(\mathbf{v}_t \notin K\right). \end{aligned}$$

Consequently

$$\begin{aligned} & \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} P\left(|g_t(\mathbf{v}_t) - g_t(\mathbf{h}^m_t)| > \delta\right) \\ \leq\ & \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} P\left(|\mathbf{v}_t - \mathbf{h}^m_t| > \gamma\right) + \eta. \end{aligned}$$

The result follows since $(\mathbf{v}_t)$ is L_0-approximable and η can be chosen arbitrarily small.

(b) Fix $\delta > 0$ and $M > 0$. Then

$$\begin{aligned} & P\left(|g_t(\mathbf{v}_t) - g_t(\mathbf{h}^m_t)| > \delta\right) \\ \leq\ & P\left(B_t(\mathbf{v}_t, \mathbf{h}^m_t)\, |\mathbf{v}_t - \mathbf{h}^m_t| > \delta, B_t(\mathbf{v}_t, \mathbf{h}^m_t) > M\right) \\ & + P\left(B_t(\mathbf{v}_t, \mathbf{h}^m_t)\, |\mathbf{v}_t - \mathbf{h}^m_t| > \delta, B_t(\mathbf{v}_t, \mathbf{h}^m_t) \leq M\right) \\ \leq\ & E\left[B_t(\mathbf{v}_t, \mathbf{h}^m_t)^\epsilon\right]/M^\epsilon + P\left(|\mathbf{v}_t - \mathbf{h}^m_t| > \delta/M\right) \end{aligned}$$

using Markov's inequality. Now

$$\begin{aligned} & \limsup_{m\to\infty} \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} P\left(|g_t(\mathbf{v}_t) - g_t(\mathbf{h}^m_t)| > \delta\right) \\ \leq\ & \limsup_{m\to\infty} \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} E\left[B_t(\mathbf{v}_t, \mathbf{h}^m_t)^\epsilon\right]/M^\epsilon \end{aligned}$$

since the process $(\mathbf{v}_t)$ is L_0-approximable. The result follows since M can be made arbitrarily large. ■

Remark. Inspection of the proof of Lemma D4 reveals that the lemma holds if the tightness condition is dropped in (a) and the equicontinuity of $\{g_t : t \in \mathbf{N}\}$ on $V - G$ is strengthened to uniform equicontinuity on $V - G$.

We note that in this case G is open and hence we may put $G_k \equiv G$. (Recall the differences in the definitions of uniform equicontinuity on a subset and uniform equicontinuity of the restrictions to the subset.)

Proof of Theorem 6.5. The L_0-approximability of $(g_t(\mathbf{v}_t))$ follows directly from Lemma D4. The L_p-approximability then follows from part (b) of Theorem 6.1. ■

Proof of Theorem 6.6. We first prove part (a) of the theorem. The L_0-approximability of $(g_t(\mathbf{v}_t))$ follows directly from Lemma D4. The L_p-approximability then follows from part (b) of Theorem 6.1. To prove part (b) of the theorem observe that the L_p-approximators $\mathbf{h}_t^m$ can w.l.o.g. be assumed to take their values in V in view of Lemma D3. The L_p-approximability of $(g_t(\mathbf{v}_t))$ then follows immediately since $|g_t(\mathbf{v}_t) - g_t(\mathbf{h}_t^m)| \leq c\,|\mathbf{v}_t - \mathbf{h}_t^m|$. ■

Proof of Theorem 6.7. The proof of part (a) of the theorem is identical to that of Theorem 4.2 in Gallant and White (1988): Although the latter theorem only covers the case $V = \mathbf{R}^{p_v}$, its proof carries over to the case of general V under the additional assumption that the approximators $\mathbf{h}_t^m$ take their values in V, as is maintained here. Part (b) of the theorem follows immediately since $|g_t(\mathbf{v}_t) - g_t(\mathbf{h}_t^m)| \leq c\,|\mathbf{v}_t - \mathbf{h}_t^m|$. ■

Proof of Corollary 6.8. In view of the remark after Lemma D3 the near epoch dependence assumption implies the existence of approximators $\mathbf{h}_t^m$ that take their values in V and satisfy that $\sup_t \|\mathbf{v}_t - \mathbf{h}_t^m\|_2$ is of size $-q$. Using the minimum mean square error property of the conditional expectation we obtain

$$\begin{aligned} & \sup_t \|g_t(\mathbf{v}_t) - E\left(g_t(\mathbf{v}_t) \mid \mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m}\right)\|_2 \\ \leq \; & \sup_t \|g_t(\mathbf{v}_t) - g_t(\mathbf{h}_t^m)\|_2 \\ \leq \; & c \sup_t \|\mathbf{v}_t - \mathbf{h}_t^m\|_2 , \end{aligned}$$

from which the result follows immediately. ■

Proof of Lemma 6.9. The proofs of parts (a), (b) and (b') are obvious.

To prove part (a') observe that for $m \in \mathbf{N}$ and $t \in \mathbf{N}$

$$\begin{aligned}
\tilde{\nu}_{m,t} &= \left\| \begin{bmatrix} \xi_t \\ \cdot \\ \cdot \\ \cdot \\ \xi_{t-l} \\ \eta_t \\ \cdot \\ \cdot \\ \cdot \\ \eta_{t-k} \end{bmatrix} - \begin{bmatrix} E(\xi_t \mid \tilde{\mathbf{e}}^{\xi}_{t+m}, \tilde{\mathbf{e}}^{\eta}_{t+m}, \ldots, \tilde{\mathbf{e}}^{\xi}_{t-m}, \tilde{\mathbf{e}}^{\eta}_{t-m}) \\ \cdot \\ \cdot \\ \cdot \\ E(\xi_{t-l} \mid \tilde{\mathbf{e}}^{\xi}_{t+m}, \tilde{\mathbf{e}}^{\eta}_{t+m}, \ldots, \tilde{\mathbf{e}}^{\xi}_{t-m}, \tilde{\mathbf{e}}^{\eta}_{t-m}) \\ E(\eta_t \mid \tilde{\mathbf{e}}^{\xi}_{t+m}, \tilde{\mathbf{e}}^{\eta}_{t+m}, \ldots, \tilde{\mathbf{e}}^{\xi}_{t-m}, \tilde{\mathbf{e}}^{\eta}_{t-m}) \\ \cdot \\ \cdot \\ \cdot \\ E(\eta_{t-k} \mid \tilde{\mathbf{e}}^{\xi}_{t+m}, \tilde{\mathbf{e}}^{\eta}_{t+m}, \ldots, \tilde{\mathbf{e}}^{\xi}_{t-m}, \tilde{\mathbf{e}}^{\eta}_{t-m}) \end{bmatrix} \right\|_2 \\
&\leq \sum_{j=0}^{l} \left\| \xi_{t-j} - E\left(\xi_{t-j} \mid \tilde{\mathbf{e}}^{\xi}_{t+m}, \tilde{\mathbf{e}}^{\eta}_{t+m}, \ldots, \tilde{\mathbf{e}}^{\xi}_{t-m}, \tilde{\mathbf{e}}^{\eta}_{t-m}\right) \right\|_2 \\
&\quad + \sum_{j=0}^{k} \left\| \eta_{t-j} - E\left(\eta_{t-j} \mid \tilde{\mathbf{e}}^{\xi}_{t+m}, \tilde{\mathbf{e}}^{\eta}_{t+m}, \ldots, \tilde{\mathbf{e}}^{\xi}_{t-m}, \tilde{\mathbf{e}}^{\eta}_{t-m}\right) \right\|_2 \\
&\leq \sum_{j=0}^{l} \left\| \xi_{t-j} - E\left(\xi_{t-j} \mid \tilde{\mathbf{e}}^{\xi}_{t+m}, \ldots, \tilde{\mathbf{e}}^{\xi}_{t-m}\right) \right\|_2 \\
&\quad + \sum_{j=0}^{k} \left\| \eta_{t-j} - E\left(\eta_{t-j} \mid \tilde{\mathbf{e}}^{\eta}_{t+m}, \ldots, \tilde{\mathbf{e}}^{\eta}_{t-m}\right) \right\|_2 \\
&\leq \sum_{j=0}^{l} \left\| \xi_{t-j} - E\left(\xi_{t-j} \mid \tilde{\mathbf{e}}^{\xi}_{t-j+(m-j)}, \ldots, \tilde{\mathbf{e}}^{\xi}_{t-j-(m-j)}\right) \right\|_2 \\
&\quad + \sum_{j=0}^{k} \left\| \eta_{t-j} - E\left(\eta_{t-j} \mid \tilde{\mathbf{e}}^{\eta}_{t-j+(m-j)}, \ldots, \tilde{\mathbf{e}}^{\eta}_{t-j-(m-j)}\right) \right\|_2,
\end{aligned}$$

where the last inequalities hold since we condition on a smaller set of random variables. (If the list of conditioning variables is empty, i.e., if the index of the first conditioning variable is less than the index of the last conditioning variable in the list, which can only happen if $m < \max\{k, l\}$, the conditional expectation should be interpreted as the unconditional expectation.) Clearly,

$$\left\| \xi_{t-j} - E\left(\xi_{t-j} \mid \tilde{\mathbf{e}}^{\xi}_{t-j+(m-j)}, \ldots, \tilde{\mathbf{e}}^{\xi}_{t-j-(m-j)}\right) \right\|_2 \leq \nu^{\xi}_{m-j}$$

and

$$\left\| \eta_{t-j} - E\left(\eta_{t-j} \mid \tilde{\mathbf{e}}^{\eta}_{t-j+(m-j)}, \ldots, \tilde{\mathbf{e}}^{\eta}_{t-j-(m-j)}\right) \right\|_2 \leq \nu^{\eta}_{m-j}$$

for $t - j > 0$, where ν^{ξ}_m and ν^{η}_m are the approximation errors for ξ_t and η_t, respectively, in the definition of near epoch dependence on the basis processes $(\mathbf{e}_t^\xi)$ and $(\mathbf{e}_t^\eta)$, respectively. If $t-j \le 0$ and $m > 2\max\{k,l\}$

$$\left\|\xi_{t-j} - E\left(\xi_{t-j} \mid \tilde{\mathbf{e}}^\xi_{t-j+(m-j)}, \dots, \tilde{\mathbf{e}}^\xi_{t-j-(m-j)}\right)\right\|_2 = 0$$

and

$$\left\|\eta_{t-j} - E\left(\eta_{t-j} \mid \tilde{\mathbf{e}}^\eta_{t-j+(m-j)}, \dots, \tilde{\mathbf{e}}^\eta_{t-j-(m-j)}\right)\right\|_2 = 0,$$

and hence these expressions are also bounded from above by ν^ξ_{m-j} and ν^η_{m-j}, respectively. This leads to

$$\tilde{\nu}_{m,t} \le \sum_{j=0}^{l} \nu^\xi_{m-j} + \sum_{j=0}^{k} \nu^\eta_{m-j}$$

for $m > 2\max\{k,l\}$ and all $t \in \mathbf{N}$. Observing that ν^ξ_m and ν^η_m are of size $-q$ this completes the proof of the first claim in (a'). The second claim follows analogously with $\tilde{\mathbf{e}}^\xi_t$ and $\tilde{\mathbf{e}}^\eta_t$ replaced by $\mathbf{e}^\xi_t$ and $\mathbf{e}^\eta_t$, respectively, observing that for $t-j \le 0$

$$\left\|\xi_{t-j} \; - \; E\left(\xi_{t-j} \mid \mathbf{e}^\xi_{t-j+(m-j)}, \dots, \mathbf{e}^\xi_{t-j-(m-j)}\right)\right\|_2$$

and

$$\left\|\eta_{t-j} \; - \; E\left(\eta_{t-j} \mid \mathbf{e}^\eta_{t-j+(m-j)}, \dots, \mathbf{e}^\eta_{t-j-(m-j)}\right)\right\|_2$$

are then of size $-q$.

To prove part (c) let $h_t^{m,i}(\mathbf{e}^\xi_{t+m}, \dots, \mathbf{e}^\xi_{t-m})$ be L_p-approximators for (ξ^i_t), $m \in \mathbf{N}$. We now specify L_p-approximators $h_s^{r,\eta}$ for (η_s). Representing $s \ge 1$ as $s = (\tau-1)k + i + 1$ with $0 \le i \le k-1$ define

$$\begin{aligned} \mathbf{h}^{km,\eta}_{(\tau-1)k+i+1} &= h^{km,\eta}_{(\tau-1)k+i+1}\left(\mathbf{e}^\eta_{(\tau-1)k+i+1+km}, \dots, \mathbf{e}^\eta_{(\tau-1)k+i+1-km}\right) \\ &= h^{m,i}_\tau\left(\mathbf{e}^\xi_{\tau+m}, \dots, \mathbf{e}^\xi_{\tau-m}\right). \end{aligned}$$

Observe that this is well-defined since by definition of $(\mathbf{e}^\eta_s)$ the argument list $(\mathbf{e}^\eta_{(\tau-1)k+i+1+km}, \dots, \mathbf{e}^\eta_{(\tau-1)k+i+1-km})$ is obtained by listing the elements of $(\mathbf{e}^\xi_{\tau+m}, \dots, \mathbf{e}^\xi_{\tau-m})$ repeatedly. For arbitrary integers $r \in \mathbf{N}$ define

$$\begin{aligned} \mathbf{h}^{r,\eta}_{(\tau-1)k+i+1} &= h^{r,\eta}_{(\tau-1)k+i+1}\left(\mathbf{e}^\eta_{(\tau-1)k+i+1+r}, \dots, \mathbf{e}^\eta_{(\tau-1)k+i+1-r}\right) \\ &= h^{k[r/k],\eta}_{(\tau-1)k+i+1}\left(\mathbf{e}^\eta_{(\tau-1)k+i+1+k[r/k]}, \dots, \mathbf{e}^\eta_{(\tau-1)k+i+1-k[r/k]}\right), \end{aligned}$$

where $[x]$ denotes the largest integer not exceeding x. Clearly,

$$\left|\eta_{(\tau-1)k+i+1} \; - \; \mathbf{h}^{r,\eta}_{(\tau-1)k+i+1}\right| = \left|\xi^i_\tau - \mathbf{h}^{[r/k],i}_\tau\right|.$$

Hence for $p > 0$

$$\begin{aligned}
& n^{-1}\sum_{s=1}^{n}\|\eta_s - \mathbf{h}_s^{r,\eta}\|_p \\
\leq\ & n^{-1}\sum_{s=1}^{([n/k]+1)k}\|\eta_s - \mathbf{h}_s^{r,\eta}\|_p \\
=\ & \sum_{i=0}^{k-1} n^{-1}\sum_{\tau=1}^{[n/k]+1}\left\|\xi_\tau^i - \mathbf{h}_\tau^{[r/k],i}\right\|_p \\
\leq\ & 2\sum_{i=0}^{k-1}([n/k]+1)^{-1}\sum_{\tau=1}^{[n/k]+1}\left\|\xi_\tau^i - \mathbf{h}_\tau^{[r/k],i}\right\|_p ,
\end{aligned}$$

and for $p = 0$

$$\begin{aligned}
& n^{-1}\sum_{s=1}^{n} P\left(|\eta_s - \mathbf{h}_s^{r,\eta}| > \delta\right) \\
\leq\ & n^{-1}\sum_{s=1}^{([n/k]+1)k} P\left(|\eta_s - \mathbf{h}_s^{r,\eta}| > \delta\right) \\
=\ & \sum_{i=0}^{k-1} n^{-1}\sum_{\tau=1}^{[n/k]+1} P\left(\left|\xi_\tau^i - \mathbf{h}_\tau^{[r/k],i}\right| > \delta\right) \\
\leq\ & 2\sum_{i=0}^{k-1}([n/k]+1)^{-1}\sum_{\tau=1}^{[n/k]+1} P\left(\left|\xi_\tau^i - \mathbf{h}_\tau^{[r/k],i}\right| > \delta\right).
\end{aligned}$$

The result now follows, since (ξ_t^i), $i = 0,\ldots,k-1$, is L_p-approximable by assumption.

To prove part (c') we proceed as in the proof of part (c) except that we choose

$$h_t^{m,i}\left(\mathbf{e}_{t+m}^{\xi},\ldots,\mathbf{e}_{t-m}^{\xi}\right) = E\left(\xi_t^i \mid \mathbf{e}_{t+m}^{\xi},\ldots,\mathbf{e}_{t-m}^{\xi}\right).$$

It follows using the same notation as in the proof of part (c) from the minimum mean square error property of the conditional expectation that

$$\begin{aligned}
& \left\|\eta_{(\tau-1)k+i+1} - E\left(\eta_{(\tau-1)k+i+1} \mid \mathbf{e}_{(\tau-1)k+i+1+r}^{\eta},\ldots,\mathbf{e}_{(\tau-1)k+i+1-r}^{\eta}\right)\right\|_2 \\
\leq\ & \left\|\eta_{(\tau-1)k+i+1} - \mathbf{h}_{(\tau-1)k+i+1}^{r,\eta}\right\|_2 \\
=\ & \left\|\xi_\tau^i - \mathbf{h}_\tau^{[r/k],i}\right\|_2 \\
=\ & \left\|\xi_\tau^i - E\left(\xi_\tau^i \mid \mathbf{e}_{\tau+[r/k]}^{\xi},\ldots,\mathbf{e}_{\tau-[r/k]}^{\xi}\right)\right\|_2 .
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sup_{s\geq 1}\left\|\eta_s - E\left(\eta_s \mid \mathbf{e}^{\eta}_{s+r},\ldots,\mathbf{e}^{\eta}_{s-r}\right)\right\|_2 \\
\leq\; & \max_{0\leq i\leq k-1}\sup_{\tau\geq 1}\Big\|\eta_{(\tau-1)k+i+1} \\
& - E\left(\eta_{(\tau-1)k+i+1} \mid \mathbf{e}^{\eta}_{(\tau-1)k+i+1+r},\ldots,\mathbf{e}^{\eta}_{(\tau-1)k+i+1-r}\right)\Big\|_2 \\
\leq\; & \max_{0\leq i\leq k-1}\sup_{\tau\geq 1}\left\|\xi^i_\tau - E\left(\xi^i_\tau \mid \mathbf{e}^{\xi}_{\tau+[r/k]},\ldots,\mathbf{e}^{\xi}_{\tau-[r/k]}\right)\right\|_2 \\
\leq\; & \max_{0\leq i\leq k-1} \nu^i_{[r/k]}.
\end{aligned}$$

Here ν^i_m denotes the approximation error in the definition of near epoch dependence of ξ^i_t. This completes the proof of part (c') since the r.h.s. in the above inequality is of size $-q$.

To prove part (d) let $\mathbf{h}^m_t = h^m_t(\mathbf{e}^{\xi}_{t+m},\ldots,\mathbf{e}^{\xi}_{t-m})$ denote L_p-approximators for (ξ_t), $m \in \mathbf{N}$. We now specify approximators $h^{m,i}_t(\mathbf{e}^{\eta}_{t+m},\ldots,\mathbf{e}^{\eta}_{t-m})$ for (η^i_t) as

$$\begin{aligned}
\mathbf{h}^{m,i}_t &= h^{m,i}_t\left(\mathbf{e}^{\eta}_{t+m},\ldots,\mathbf{e}^{\eta}_{t-m}\right) \\
&= h^{km}_{(t-1)k+i+1}\left(\mathbf{e}^{\xi}_{(t-1)k+i+1+km},\ldots,\mathbf{e}^{\xi}_{(t-1)k+i+1-km}\right).
\end{aligned}$$

Observe that this is well-defined, since by definition of $\mathbf{e}^{\eta}_t$ the argument list on the r.h.s. is a subset of the argument list on the l.h.s. Consequently,

$$\left|\eta^i_t - \mathbf{h}^{m,i}_t\right| = \left|\xi_{(t-1)k+i+1} - \mathbf{h}^{km}_{(t-1)k+i+1}\right|,$$

and hence for $p > 0$

$$\begin{aligned}
& n^{-1}\sum_{t=1}^{n}\left\|\eta^i_t - \mathbf{h}^{m,i}_t\right\|_p \\
=\; & n^{-1}\sum_{t=1}^{n}\left\|\xi_{(t-1)k+i+1} - \mathbf{h}^{km}_{(t-1)k+i+1}\right\|_p \\
\leq\; & k(kn)^{-1}\sum_{t=1}^{kn}\left\|\xi_t - \mathbf{h}^{km}_t\right\|_p,
\end{aligned}$$

and for $p = 0$

$$\begin{aligned}
& n^{-1}\sum_{t=1}^{n}P\left(\left|\eta^i_t - \mathbf{h}^{m,i}_t\right| > \delta\right) \\
=\; & n^{-1}\sum_{t=1}^{n}P\left(\left|\xi_{(t-1)k+i+1} - \mathbf{h}^{km}_{(t-1)k+i+1}\right| > \delta\right) \\
\leq\; & k(kn)^{-1}\sum_{t=1}^{kn}P\left(\left|\xi_t - \mathbf{h}^{km}_t\right| > \delta\right).
\end{aligned}$$

The result now follows, since (ξ_t) is L_p-approximable by assumption.

To prove part (d') we proceed as in the proof of part (d) except that we choose

$$h_t^m\left(\mathbf{e}_{t+m}^{\xi},\ldots,\mathbf{e}_{t-m}^{\xi}\right)=E\left(\xi_t\mid\mathbf{e}_{t+m}^{\xi},\ldots,\mathbf{e}_{t-m}^{\xi}\right).$$

It follows from the minimum mean square error property of the conditional expectation – using the same notation as in the proof of part (d) and observing that $\mathbf{h}_t^{m,i}$ is clearly quadratically integrable – that

$$\begin{aligned}&\left\|\eta_t^i-E\left(\eta_t^i\mid\mathbf{e}_{t+m}^{\eta},\ldots,\mathbf{e}_{t-m}^{\eta}\right)\right\|_2\\ \le\;&\left\|\eta_t^i-\mathbf{h}_t^{m,i}\right\|_2\\ =\;&\left\|\xi_{(t-1)k+i+1}\right.\\ &\left.-E\left(\xi_{(t-1)k+i+1}\mid\mathbf{e}_{(t-1)k+i+1+km}^{\xi},\ldots,\mathbf{e}_{(t-1)k+i+1-km}^{\xi}\right)\right\|_2.\end{aligned}$$

Consequently,

$$\begin{aligned}&\sup_{t\ge1}\left\|\eta_t^i-E\left(\eta_t^i\mid\mathbf{e}_{t+m}^{\eta},\ldots,\mathbf{e}_{t-m}^{\eta}\right)\right\|_2\\ \le\;&\sup_{t\ge1}\left\|\xi_{(t-1)k+i+1}\right.\\ &\left.-E\left(\xi_{(t-1)k+i+1}\mid\mathbf{e}_{(t-1)k+i+1+km}^{\xi},\ldots,\mathbf{e}_{(t-1)k+i+1-km}^{\xi}\right)\right\|_2\\ \le\;&\sup_{s\ge1}\left\|\xi_s-E\left(\xi_s\mid\mathbf{e}_{s+km}^{\xi},\ldots,\mathbf{e}_{s-km}^{\xi}\right)\right\|_2.\end{aligned}$$

This completes the proof since the r.h.s. is of size $-q$ by assumption. ■

Lemma D5. *Let $(\mathbf{v}_t)_{t\in\mathbf{N}}$ and $(\mathbf{w}_t)_{t\in\mathbf{Z}}$ be stochastic processes taking their values in Borel subsets V and W of $\mathbf{R}^{p_v}$ and $\mathbf{R}^{p_w}$, respectively, and let $g_t: V\times W\to V$ be functions for $t\in\mathbf{N}$. Suppose that $(\mathbf{v}_t)$ is generated according to the dynamic system*

$$\mathbf{v}_t=g_t(\mathbf{v}_{t-1},\mathbf{w}_t),\quad t\in\mathbf{N},$$

where $\mathbf{v}_0$ is some initial random variable taking its values in V. Assume that for all $(v,v^\bullet)\in V\times V,(w,w^\bullet)\in W\times W$, and $t\in\mathbf{N}$

$$|g_t(v,w)-g_t(v^\bullet,w^\bullet)|\le d_v\,|v-v^\bullet|+d_{w,t}\,|w-w^\bullet|$$

holds where the global Lipschitz constants satisfy $0\le d_v<1$ and $0\le d_{w,t}<\infty$.

(a) If there exists an element $\bar{w}\in W$ such that $\left\|d_{w,t}\,|\mathbf{w}_t-\bar{w}|\,\right\|_r<\infty$ for $t\in\mathbf{N}$ and $\|\mathbf{v}_0\|_r<\infty$ for some $r\ge1$, then $\|\mathbf{v}_t\|_r<\infty$ for $t\in\mathbf{N}$. If additionally

$$\sup_{t\ge1}\left\|d_{w,t}\,|\mathbf{w}_t-\bar{w}|\,\right\|_r<\infty$$

and

$$\sup_{t\geq 1} |g_t(\bar{v}, \bar{w})| < \infty$$

holds for some element $\bar{v} \in V$*, then also*

$$\sup_{t\geq 1} \|\mathbf{v}_t\|_r < \infty.$$

(b) If there exists an element $\bar{w} \in W$ *such that*

$$\sup_{t\geq 1} \|d_{w,t}\,|\mathbf{w}_t - \bar{w}|\,\|_r < \infty$$

and $\|\mathbf{v}_0\|_r < \infty$ *for some* $r \geq 1$*, then* $(\mathbf{v}_t)$ *is* L_r*-approximable by* $(\mathbf{w}_t)$*. If furthermore* $(\mathbf{w}_t)_{t\in\mathbf{N}}$ *is also* L_r*-approximable by some process* $(\mathbf{e}_t)_{t\in\mathbf{Z}}$ *and if* $d_{w,t} \leq d_w < \infty$ *for all* $t \in \mathbf{N}$*, then* $(\mathbf{v}_t)$ *is* L_r*-approximable by* $(\mathbf{e}_t)$*.*

(c) If there exists an element $\bar{w} \in W$ *such that*

$$\sup_{t\geq 1} \|d_{w,t}\,|\mathbf{w}_t - \bar{w}|\,\|_2 < \infty$$

and $\|\mathbf{v}_0\|_2 < \infty$*, then* $(\mathbf{v}_t)$ *is near epoch dependent of size* $-q$ *on* $(\tilde{\mathbf{w}}_t)$ *for any* $q > 0$*, where* $\tilde{\mathbf{w}}_t = (\mathbf{w}_t', 0)'$ *for* $t \neq 1$ *and* $\tilde{\mathbf{w}}_1 = (\mathbf{w}_1', \mathbf{v}_0')'$*. If furthermore* $(\mathbf{w}_t)_{t\in\mathbf{N}}$ *is also near epoch dependent of size* $-q$ *on some process* $(\mathbf{e}_t)_{t\in\mathbf{Z}}$ *and if* $d_{w,t} \leq d_w < \infty$ *for all* $t \in \mathbf{N}$*, then* $(\mathbf{v}_t)$ *is near epoch dependent of size* $-q$ *on* $(\tilde{\mathbf{e}}_t)$*, where* $\tilde{\mathbf{e}}_t = (\mathbf{e}_t', 0)'$ *for* $t \neq 1$ *and* $\tilde{\mathbf{e}}_1 = (\mathbf{e}_1', \mathbf{v}_0')'$*. If additionally also* $\|\mathbf{v}_0 - E(\mathbf{v}_0 \mid \mathbf{e}_m, \ldots, \mathbf{e}_{-m})\|_2$ *is of size* $-q$*, then* $(\mathbf{v}_t)$ *is near epoch dependent of size* $-q$ *on* $(\mathbf{e}_t)$*.*

Remarks. (i) Note that $\tilde{\mathbf{w}}_t$ takes its values in $\mathbf{R}^{p_w+p_v}$, but not necessarily in $W \times V$, since 0 need not be an element of V. If a basis process $\tilde{\mathbf{w}}_t$ that takes its values in $W \times V$ is desired, one can replace 0 by any arbitrary fixed element of W in the definition of $\tilde{\mathbf{w}}_t$ for $t \neq 1$.

(ii) Inspection of the proof shows that with appropriate modifications analogous results in (c) also hold if an L_r-norm, $r \geq 1$, is used instead of the L_2-norm.

(iii) Inspection of the proof shows that the first claim in part (a) holds even if the functions g_t are only assumed to be Lipschitz, i.e.,

$$|g_t(v, w) - g_t(v^\bullet, w^\bullet)| \leq d_{v,t}\,|v - v^\bullet| + d_{w,t}\,|w - w^\bullet|$$

holds with $0 \leq d_{v,t} < \infty$ and $0 \leq d_{w,t} < \infty$.

Proof. Let $\bar{v}_0$ be an arbitrary element in V and define $\bar{v}_t = g_t(\bar{v}_{t-1}, \bar{w})$ for $t \in \mathbf{N}$ where $\bar{w}$ is as in the assumptions. We first prove part (a). Then we have for $t \in \mathbf{N}$

$$\begin{aligned} |\mathbf{v}_t - \bar{v}_t| &= |g_t(\mathbf{v}_{t-1}, \mathbf{w}_t) - g_t(\bar{v}_{t-1}, \bar{w})| \\ &\leq d_v |\mathbf{v}_{t-1} - \bar{v}_{t-1}| + d_{w,t} |\mathbf{w}_t - \bar{w}| \\ &= d_v |g_{t-1}(\mathbf{v}_{t-2}, \mathbf{w}_{t-1}) - g_{t-1}(\bar{v}_{t-2}, \bar{w})| + d_{w,t} |\mathbf{w}_t - \bar{w}| \\ &\leq d_v^2 |\mathbf{v}_{t-2} - \bar{v}_{t-2}| + d_v d_{w,t-1} |\mathbf{w}_{t-1} - \bar{w}| + d_{w,t} |\mathbf{w}_t - \bar{w}| \\ &\leq \ldots \quad \leq d_v^t |\mathbf{v}_0 - \bar{v}_0| + \sum_{i=0}^{t-1} d_v^i d_{w,t-i} |\mathbf{w}_{t-i} - \bar{w}|. \end{aligned}$$

Hence

$$\|\mathbf{v}_t - \bar{v}_t\|_r \leq d_v^t \|\mathbf{v}_0\|_r + d_v^t |\bar{v}_0| + \sum_{i=0}^{t-1} d_v^i \|d_{w,t-i} |\mathbf{w}_{t-i} - \bar{w}| \|_r < \infty$$

for $t \in \mathbf{N}$ by the assumptions. Since $|\bar{v}_t|$ is finite, it follows that $\|\mathbf{v}_t\|_r$ is finite for $t \in \mathbf{N}$. This proves the first claim of part (a). Now suppose that the additional assumptions for the second claim of part (a) hold. It then follows from the above inequality that

$$\sup_{t\geq 1} \|\mathbf{v}_t - \bar{v}_t\|_r \leq \|\mathbf{v}_0\|_r + |\bar{v}_0| + \sup_{t\geq 1} \|d_{w,t} |\mathbf{w}_t - \bar{w}| \|_r (1 - d_v)^{-1} < \infty$$

observing that $d_v < 1$. Now upon choosing $\bar{v}_0 = \bar{v}$ for the moment we also have for $t \in \mathbf{N}$

$$\begin{aligned} |\bar{v}_t - \bar{v}_0| &\leq |g_t(\bar{v}_{t-1}, \bar{w}) - g_t(\bar{v}_0, \bar{w})| + |g_t(\bar{v}_0, \bar{w}) - \bar{v}_0| \\ &\leq d_v |\bar{v}_{t-1} - \bar{v}_0| + M, \end{aligned}$$

where

$$M = \left(|\bar{v}_0| + \sup_{t\geq 1} |g_t(\bar{v}_0, \bar{w})|\right) < \infty.$$

Induction shows that

$$|\bar{v}_t - \bar{v}_0| \leq M(1 - d_v)^{-1}(1 - d_v^t) \leq M(1 - d_v)^{-1} < \infty.$$

Hence, $\sup_{t\geq 1} |\bar{v}_t| < \infty$ holds, and therefore

$$\sup_{t\geq 1} \|\mathbf{v}_t\|_r \leq \sup_{t\geq 1} \|\mathbf{v}_t - \bar{v}_t\|_r + \sup_{t\geq 1} |\bar{v}_t| < \infty.$$

This proves the second claim of part (a).

We next prove the first claim of part (b). Consider for every $t \geq 1$ and $m \geq 0$ the recursions

$$\xi_j^{t,m} = g_j(\xi_{j-1}^{t,m}, \mathbf{w}_j)$$

for $\max(1, t-m) \leq j \leq t$ initialized by $\xi_{\max(1,t-m)-1}^{t,m} = \bar{v}_{\max(1,t-m)-1}$. Define approximators $\mathbf{h}_t^m$ for $t \geq 1$ and $m \geq 0$ by setting $\mathbf{h}_t^m = \xi_t^{t,m}$. As is easily seen, $\mathbf{h}_t^m = h_t^m(\mathbf{w}_{t+m}, \ldots, \mathbf{w}_{t-m})$, that is $\mathbf{h}_t^m$ can be expressed as a measurable function of $\mathbf{w}_{t+m}, \ldots, \mathbf{w}_{t-m}$. In case $t \leq m+1$ we then have

$$\begin{aligned}
& |\mathbf{v}_t - \mathbf{h}_t^m| = \left|\mathbf{v}_t - \xi_t^{t,m}\right| = \left|g_t(\mathbf{v}_{t-1}, \mathbf{w}_t) - g_t(\xi_{t-1}^{t,m}, \mathbf{w}_t)\right| \\
\leq\ & d_v \left|\mathbf{v}_{t-1} - \xi_{t-1}^{t,m}\right| + d_{w,t} \left|\mathbf{w}_t - \mathbf{w}_t\right| \\
=\ & d_v \left|\mathbf{v}_{t-1} - \xi_{t-1}^{t,m}\right| \\
\leq\ & d_v^2 \left|\mathbf{v}_{t-2} - \xi_{t-2}^{t,m}\right| \leq \ldots \leq d_v^t \left|\mathbf{v}_0 - \bar{v}_0\right|.
\end{aligned}$$

Next consider the case where $t > m+1$:

$$\begin{aligned}
& |\mathbf{v}_t - \mathbf{h}_t^m| = \left|\mathbf{v}_t - \xi_t^{t,m}\right| = \left|g_t(\mathbf{v}_{t-1}, \mathbf{w}_t) - g_t(\xi_{t-1}^{t,m}, \mathbf{w}_t)\right| \\
\leq\ & d_v \left|\mathbf{v}_{t-1} - \xi_{t-1}^{t,m}\right| \leq \ldots \leq d_v^{m+1} \left|\mathbf{v}_{t-m-1} - \bar{v}_{t-m-1}\right| \\
\leq\ & d_v^{m+1} \left|g_{t-m-1}(\mathbf{v}_{t-m-2}, \mathbf{w}_{t-m-1}) - g_{t-m-1}(\bar{v}_{t-m-2}, \bar{w})\right| \\
\leq\ & d_v^{m+1} \left\{d_v \left|\mathbf{v}_{t-m-2} - \bar{v}_{t-m-2}\right| + d_{w,t-m-1} \left|\mathbf{w}_{t-m-1} - \bar{w}\right|\right\} \\
\leq\ & d_v^{m+1} \left\{d_v \left|g_{t-m-2}(\mathbf{v}_{t-m-3}, \mathbf{w}_{t-m-2}) - g_{t-m-2}(\bar{v}_{t-m-3}, \bar{w})\right| \right. \\
& \left. + d_{w,t-m-1} \left|\mathbf{w}_{t-m-1} - \bar{w}\right|\right\} \\
\leq\ & d_v^{m+1} \left\{d_v^2 \left|\mathbf{v}_{t-m-3} - \bar{v}_{t-m-3}\right| + d_v d_{w,t-m-2} \left|\mathbf{w}_{t-m-2} - \bar{w}\right| \right. \\
& \left. + d_{w,t-m-1} \left|\mathbf{w}_{t-m-1} - \bar{w}\right|\right\} \\
\leq\ & \ldots \leq d_v^t \left|\mathbf{v}_0 - \bar{v}_0\right| + d_v^{m+1} \sum_{i=1}^{t-m-1} d_v^{i-1} d_{w,t-m-i} \left|\mathbf{w}_{t-m-i} - \bar{w}\right|.
\end{aligned}$$

Hence, for all $t \geq 1$, $m \geq 0$ we have that

$$|\mathbf{v}_t - \mathbf{h}_t^m| \leq d_v^t \left|\mathbf{v}_0 - \bar{v}_0\right| + d_v^{m+1} \sum_{i=1}^{t-m-1} d_v^{i-1} d_{w,t-m-i} \left|\mathbf{w}_{t-m-i} - \bar{w}\right|$$

with the convention that a sum is zero if the range of summation is empty. This immediately implies

$$\begin{aligned}
\|\mathbf{v}_t - \mathbf{h}_t^m\|_r \leq\ & d_v^t \|\mathbf{v}_0 - \bar{v}_0\|_r \\
& + d_v^{m+1} \sum_{i=1}^{t-m-1} d_v^{i-1} \left\|d_{w,t-m-i} \left|\mathbf{w}_{t-m-i} - \bar{w}\right| \right\|_r \\
\leq\ & d_v^t (C + |\bar{v}_0|) + C d_v^{m+1} (1 - d_v)^{-1}
\end{aligned}$$

where C is a constant such that

$$\sup_{t \geq 1} \left\|d_{w,t} \left|\mathbf{w}_t - \bar{w}\right| \right\|_r \leq C$$

and $\|\mathbf{v}_0\|_r \leq C$ holds. Consequently,

$$0 \leq \limsup_{m \to \infty} \limsup_{n \to \infty} n^{-1} \sum_{t=1}^{n} \|\mathbf{v}_t - \mathbf{h}_t^m\|_r$$

$$\begin{aligned}&\leq \limsup_{m\to\infty}\limsup_{n\to\infty}\left[n^{-1}d_v(1-d_v)^{-1}(C+|\bar{v}_0|)+Cd_v^{m+1}(1-d_v)^{-1}\right]\\&= 0,\end{aligned}$$

which proves the first claim in part (b) of the theorem.

We next prove the first claim of part (c). To this end define $\mathbf{h}_t^m$ as in the proof of the first claim of part (b) except that the recursions

$$\xi_j^{t,m} = g_j(\xi_{j-1}^{t,m}, \mathbf{w}_j)$$

for $\max(1, t-m) \leq j \leq t$ are initialized by $\xi_0^{t,m} = \mathbf{v}_0$ for $t \leq m+1$ and by $\xi_{t-m-1}^{t,m} = \bar{v}_{t-m-1}$ for $t > m+1$. In other words, we set $\mathbf{h}_t^m = \mathbf{v}_t$ for $t \leq m+1$ and define $\mathbf{h}_t^m$ as above for $t > m+1$. Since $\mathbf{v}_t$ is a measurable function of $\mathbf{w}_t, \ldots, \mathbf{w}_1$ and $\mathbf{v}_0$, clearly $\mathbf{h}_t^m$ is for $t \leq m+1$ a measurable function of $\tilde{\mathbf{w}}_{t+m}, \ldots, \tilde{\mathbf{w}}_{t-m}$. For $t > m+1$ we can clearly also view $\mathbf{h}_t^m$ as a measurable function of $\tilde{\mathbf{w}}_{t+m}, \ldots, \tilde{\mathbf{w}}_{t-m}$. Given this definition of $\mathbf{h}_t^m$ we then have $|\mathbf{v}_t - \mathbf{h}_t^m| = 0$ for $t \leq m+1$ and

$$|\mathbf{v}_t - \mathbf{h}_t^m| \leq d_v^t\,|\mathbf{v}_0 - \bar{v}_0| + d_v^{m+1}\sum_{i=1}^{t-m-1} d_v^{i-1} d_{w,t-m-i}\,|\mathbf{w}_{t-m-i} - \bar{w}|$$

for $t > m+1$ as before. Hence for all $t \geq 1$ we obtain

$$\begin{aligned}&\|\mathbf{v}_t - \mathbf{h}_t^m\|_2\\ &\leq d_v^{m+1}(D+|\bar{v}_0|) + d_v^{m+1}\sum_{i=1}^{t-m-1} d_v^{i-1}\,\|d_{w,t-m-i}\,|\mathbf{w}_{t-m-i}-\bar{w}|\,\|_2\\ &\leq d_v^{m+1}(D+|\bar{v}_0|) + Dd_v^{m+1}(1-d_v)^{-1}\end{aligned}$$

where D is a constant such that

$$\sup_{t\geq 1}\|d_{w,t}\,|\mathbf{w}_t - \bar{w}|\,\|_2 \leq D$$

and $\|\mathbf{v}_0\|_2 \leq D$ holds. Therefore

$$0 \leq \sup_{t\geq 1}\|\mathbf{v}_t - \mathbf{h}_t^m\|_2 \leq c d_v^{m+1}$$

for some constant c and hence $\sup_{t\geq 1}\|\mathbf{v}_t - \mathbf{h}_t^m\|_2$ is of size $-q$ for any $q > 0$. The first claim in part (c) now follows since

$$0 \leq \|\mathbf{v}_t - E(\mathbf{v}_t \mid \tilde{\mathbf{w}}_{t+m}, \ldots, \tilde{\mathbf{w}}_{t-m})\|_2 \leq \|\mathbf{v}_t - \mathbf{h}_t^m\|_2$$

in view of the minimum mean square error property of the conditional expectation. (Note that the second moment of $\mathbf{v}_t$ exists according to part (a).)

We next prove the second claim in part (b). Let $\mathbf{w}_t^m$, for $t \geq 1$ and $m \geq 0$, denote an L_r-approximator for $\mathbf{w}_t$, i.e., $\mathbf{w}_t^m$ is a measurable function of $(\mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m})$ and satisfies

$$\limsup_{m\to\infty} \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} \|\mathbf{w}_t - \mathbf{w}_t^m\|_r = 0.$$

Because of Lemma D3 we can assume w.l.o.g. that $\mathbf{w}_t^m$ takes its values in W. (For notational convenience we also set $\mathbf{w}_t^m = \mathbf{w}_t$ if $t \leq 0$.) For every $t \geq 1$ and $m \geq 0$ consider recursions similar to the ones above but given now by

$$\xi_j^{t,m} = g_j(\xi_{j-1}^{t,m}, \mathbf{w}_j^m)$$

for $\max(1, t-m) \leq j \leq t$ and initialized by $\xi_{\max(1,t-m)-1}^{t,m} = \bar{v}_{\max(1,t-m)-1}$. Define approximators $\mathbf{h}_t^k$ for $\mathbf{v}_t$ for $t \geq 1$ and $k \geq 0$ by setting $\mathbf{h}_t^k = \xi_t^{t,k/2}$ if k is even and $\mathbf{h}_t^k = \mathbf{h}_t^{k-1}$ if k is odd. As is easily seen, $\mathbf{h}_t^k = h_t^k(\mathbf{e}_{t+k}, \ldots, \mathbf{e}_{t-k})$, that is $\mathbf{h}_t^k$ can be expressed as a measurable function of $\mathbf{e}_{t+k}, \ldots, \mathbf{e}_{t-k}$. Now assume first that k is even and let $m = k/2$. In case $t \leq m+1$ we then have similarly as before

$$\begin{aligned}
& \left|\mathbf{v}_t - \mathbf{h}_t^k\right| = \left|\mathbf{v}_t - \xi_t^{t,m}\right| = \left|g_t(\mathbf{v}_{t-1}, \mathbf{w}_t) - g_t(\xi_{t-1}^{t,m}, \mathbf{w}_t^m)\right| \\
\leq\ & d_v \left|\mathbf{v}_{t-1} - \xi_{t-1}^{t,m}\right| + d_{w,t}\left|\mathbf{w}_t - \mathbf{w}_t^m\right| \\
=\ & d_v \left|g_{t-1}(\mathbf{v}_{t-2}, \mathbf{w}_{t-1}) - g_{t-1}(\xi_{t-2}^{t,m}, \mathbf{w}_{t-1}^m)\right| + d_{w,t}\left|\mathbf{w}_t - \mathbf{w}_t^m\right| \\
\leq\ & d_v^2 \left|\mathbf{v}_{t-2} - \xi_{t-2}^{t,m}\right| + d_v d_{w,t-1}\left|\mathbf{w}_{t-1} - \mathbf{w}_{t-1}^m\right| + d_{w,t}\left|\mathbf{w}_t - \mathbf{w}_t^m\right| \\
\leq\ & \ldots \quad \leq d_v^t \left|\mathbf{v}_0 - \bar{v}_0\right| + \sum_{i=0}^{t-1} d_v^i d_{w,t-i}\left|\mathbf{w}_{t-i} - \mathbf{w}_{t-i}^m\right|.
\end{aligned}$$

Next consider the case where $t > m+1$:

$$\begin{aligned}
& \left|\mathbf{v}_t - \mathbf{h}_t^k\right| = \left|\mathbf{v}_t - \xi_t^{t,m}\right| = \left|g_t(\mathbf{v}_{t-1}, \mathbf{w}_t) - g_t(\xi_{t-1}^{t,m}, \mathbf{w}_t^m)\right| \\
\leq\ & d_v \left|\mathbf{v}_{t-1} - \xi_{t-1}^{t,m}\right| + d_{w,t}\left|\mathbf{w}_t - \mathbf{w}_t^m\right| \\
\leq\ & d_v^2 \left|\mathbf{v}_{t-2} - \xi_{t-2}^{t,m}\right| + d_v d_{w,t-1}\left|\mathbf{w}_{t-1} - \mathbf{w}_{t-1}^m\right| + d_{w,t}\left|\mathbf{w}_t - \mathbf{w}_t^m\right| \\
\leq\ & \ldots \quad \leq \sum_{i=0}^{m} d_v^i d_{w,t-i}\left|\mathbf{w}_{t-i} - \mathbf{w}_{t-i}^m\right| + d_v^{m+1}\left|\mathbf{v}_{t-m-1} - \bar{v}_{t-m-1}\right| \\
\leq\ & \sum_{i=0}^{m} d_v^i d_{w,t-i}\left|\mathbf{w}_{t-i} - \mathbf{w}_{t-i}^m\right| \\
& + d_v^{m+1}\left|g_{t-m-1}(\mathbf{v}_{t-m-2}, \mathbf{w}_{t-m-1}) - g_{t-m-1}(\bar{v}_{t-m-2}, \bar{w})\right| \\
\leq\ & \sum_{i=0}^{m} d_v^i d_{w,t-i}\left|\mathbf{w}_{t-i} - \mathbf{w}_{t-i}^m\right| \\
& + d_v^{m+1}\left\{d_v\left|\mathbf{v}_{t-m-2} - \bar{v}_{t-m-2}\right| + d_{w,t-m-1}\left|\mathbf{w}_{t-m-1} - \bar{w}\right|\right\}
\end{aligned}$$

$$
\begin{aligned}
\le\ & \sum_{i=0}^{m} d_v^i d_{w,t-i} \left|\mathbf{w}_{t-i} - \mathbf{w}_{t-i}^m\right| \\
& + d_v^{m+1}\{d_v \left|g_{t-m-2}(\mathbf{v}_{t-m-3}, \mathbf{w}_{t-m-2}) - g_{t-m-2}(\bar{v}_{t-m-3}, \bar{w})\right| \\
& + d_{w,t-m-1} \left|\mathbf{w}_{t-m-1} - \bar{w}\right|\} \\
\le\ & \sum_{i=0}^{m} d_v^i d_{w,t-i} \left|\mathbf{w}_{t-i} - \mathbf{w}_{t-i}^m\right| \\
& + d_v^{m+1}\{d_v^2 \left|\mathbf{v}_{t-m-3} - \bar{v}_{t-m-3}\right| + d_v d_{w,t-m-2} \left|\mathbf{w}_{t-m-2} - \bar{w}\right| \\
& + d_{w,t-m-1} \left|\mathbf{w}_{t-m-1} - \bar{w}\right|\} \\
\le\ & \dots \quad \le \sum_{i=0}^{m} d_v^i d_{w,t-i} \left|\mathbf{w}_{t-i} - \mathbf{w}_{t-i}^m\right| + d_v^t \left|\mathbf{v}_0 - \bar{v}_0\right| \\
& + d_v^{m+1} \sum_{i=1}^{t-m-1} d_v^{i-1} d_{w,t-m-i} \left|\mathbf{w}_{t-m-i} - \bar{w}\right|.
\end{aligned}
$$

Hence for all $t \ge 1$ we get in view of the notational convention adopted for $\mathbf{w}_t^m$ if $t \le 0$ and using the additional assumption $d_{w,t} \le d_w$

$$
\begin{aligned}
& \left\|\mathbf{v}_t - \mathbf{h}_t^k\right\|_r \\
\le\ & d_v^t \left\|\mathbf{v}_0 - \bar{v}_0\right\|_r + d_w \sum_{i=0}^{m} d_v^i \left\|\mathbf{w}_{t-i} - \mathbf{w}_{t-i}^m\right\|_r \\
& + d_v^{m+1} \sum_{i=1}^{t-m-1} d_v^{i-1} \left\|d_{w,t-m-i} \left|\mathbf{w}_{t-m-i} - \bar{w}\right|\right\|_r \\
\le\ & d_v^t (C + |\bar{v}_0|) + d_w \sum_{i=0}^{m} d_v^i \left\|\mathbf{w}_{t-i} - \mathbf{w}_{t-i}^m\right\|_r + C d_v^{m+1}(1 - d_v)^{-1},
\end{aligned}
$$

where C was defined above. If k is odd we arrive at the same inequality with $m = (k-1)/2$. Consequently,

$$
\begin{aligned}
0 \ \le\ & \limsup_{k\to\infty} \limsup_{n\to\infty} n^{-1} \sum_{t=1}^{n} \left\|\mathbf{v}_t - \mathbf{h}_t^k\right\|_r \\
\le\ & \limsup_{m\to\infty} \limsup_{n\to\infty} \left[n^{-1} d_v (1 - d_v)^{-1} (C + |\bar{v}_0|) \right. \\
& \left. + d_w \sum_{i=0}^{m} d_v^i n^{-1} \sum_{t=1}^{n} \left\|\mathbf{w}_{t-i} - \mathbf{w}_{t-i}^m\right\|_r + C d_v^{m+1}(1 - d_v)^{-1} \right] \\
\le\ & \limsup_{m\to\infty} \limsup_{n\to\infty} \left[d_w \sum_{i=0}^{m} d_v^i n^{-1} \sum_{t=1}^{n} \left\|\mathbf{w}_t - \mathbf{w}_t^m\right\|_r \right. \\
& \left. + C d_v^{m+1}(1 - d_v)^{-1} \right]
\end{aligned}
$$

$$\leq \limsup_{m\to\infty} \limsup_{n\to\infty} \left[d_w(1-d_v)^{-1} n^{-1} \sum_{t=1}^{n} \|\mathbf{w}_t - \mathbf{w}_t^m\|_r \right.$$
$$\left. + C d_v^{m+1}(1-d_v)^{-1} \right] = 0,$$

using the assumed L_r-approximability of $(\mathbf{w}_t)$ by $(\mathbf{e}_t)$. This proves the second claim in part (b).

Consider the second claim in part (c) next. Observe that because of the remark following Lemma D3 the approximators $E(\mathbf{w}_t \mid \mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m})$ appearing in the near epoch dependence assumption for $(\mathbf{w}_t)$ can be replaced by approximators $\mathbf{w}_t^m$ that take their values in W and have approximation errors $\sup_t \|\mathbf{w}_t - \mathbf{w}_t^m\|_2$ which are of the same size as

$$\sup_t \|\mathbf{w}_t - E(\mathbf{w}_t \mid \mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m})\|_2 .$$

We now define the approximators $\mathbf{h}_t^k$ for $\mathbf{v}_t$ similarly as in the proof of the second claim of part (b), except that now the recursions

$$\xi_j^{t,m} = g_j(\xi_{j-1}^{t,m}, \mathbf{w}_j^m)$$

for $\max(1, t-m) \leq j \leq t$ are initialized by $\xi_0^{t,m} = \mathbf{v}_0$ for $t \leq m+1$ and by $\xi_{t-m-1}^{t,m} = \bar{v}_{t-m-1}$ for $t > m+1$. As is easily seen, $\mathbf{h}_t^k$ is then a measurable function of $\tilde{\mathbf{e}}_{t+k}, \ldots, \tilde{\mathbf{e}}_{t-k}$. Now assume first that k is even and let $m = k/2$. In case $t \leq m+1$ we then have similarly as before

$$\begin{aligned} \left|\mathbf{v}_t - \mathbf{h}_t^k\right| &\leq d_v^t \left|\mathbf{v}_0 - \mathbf{v}_0\right| + \sum_{i=0}^{t-1} d_v^i d_{w,t-i} \left|\mathbf{w}_{t-i} - \mathbf{w}_{t-i}^m\right| \\ &= \sum_{i=0}^{t-1} d_v^i d_{w,t-i} \left|\mathbf{w}_{t-i} - \mathbf{w}_{t-i}^m\right| . \end{aligned}$$

In case $t > m+1$ we have exactly as before:

$$\begin{aligned} \left|\mathbf{v}_t - \mathbf{h}_t^k\right| &\leq \sum_{i=0}^{m} d_v^i d_{w,t-i} \left|\mathbf{w}_{t-i} - \mathbf{w}_{t-i}^m\right| + d_v^t \left|\mathbf{v}_0 - \bar{v}_0\right| \\ &\quad + d_v^{m+1} \sum_{i=1}^{t-m-1} d_v^{i-1} d_{w,t-m-i} \left|\mathbf{w}_{t-m-i} - \bar{w}\right| . \end{aligned}$$

Hence for all $t \geq 1$ we get in view of the notational convention adopted for $\mathbf{w}_t^m$ if $t \leq 0$ and using the additional assumption $d_{w,t} \leq d_w$

$$\left\|\mathbf{v}_t - \mathbf{h}_t^k\right\|_2 \leq d_v^{m+1} \left\|\mathbf{v}_0 - \bar{v}_0\right\|_2 + d_w \sum_{i=0}^{m} d_v^i \left\|\mathbf{w}_{t-i} - \mathbf{w}_{t-i}^m\right\|_2$$

$$
\begin{aligned}
&+ d_v^{m+1} \sum_{i=1}^{t-m-1} d_v^{i-1} \left\| d_{w,t-m-i} \left| \mathbf{w}_{t-m-i} - \bar{w} \right| \right\|_2 \\
\leq\ & d_v^{m+1}(D + |\bar{v}_0|) + d_w(1-d_v)^{-1} \sup_{t \geq 1} \|\mathbf{w}_t - \mathbf{w}_t^m\|_2 \\
&+ D d_v^{m+1}(1-d_v)^{-1},
\end{aligned}
$$

where D was defined above. If k is odd we arrive at the same inequality with $m = (k-1)/2$. Therefore clearly $\sup_{t \geq 1} \left\| \mathbf{v}_t - \mathbf{h}_t^k \right\|_2$ is of size $-q$. The second claim in part (c) now follows since

$$
0 \leq \|\mathbf{v}_t - E(\mathbf{v}_t \mid \tilde{\mathbf{e}}_{t+k}, \ldots, \tilde{\mathbf{e}}_{t-k})\|_2 \leq \left\| \mathbf{v}_t - \mathbf{h}_t^k \right\|_2
$$

in view of the minimum mean square error property of the conditional expectation. (Note that the second moment of $\mathbf{v}_t$ exists according to part (a).)

To prove the third claim in part (c) let $\mathbf{v}_0^m$ for $m \geq 0$ denote an approximator to $\mathbf{v}_0$ that is a measurable function of $\mathbf{e}_m, \ldots, \mathbf{e}_{-m}$, takes its values in V and whose approximation error $\|\mathbf{v}_0 - \mathbf{v}_0^m\|_2$ is of the same size as $\|\mathbf{v}_0 - E(\mathbf{v}_0 \mid \mathbf{e}_m, \ldots, \mathbf{e}_{-m})\|_2$. Such an approximator exists in view of the remark following Lemma D3. Define now $\mathbf{h}_t^k$ as above but where now the recursions

$$
\xi_j^{t,m} = g_j(\xi_{j-1}^{t,m}, \mathbf{w}_j^m)
$$

for $\max(1, t-m) \leq j \leq t$ are initialized by $\xi_0^{t,m} = \mathbf{v}_0^{m-1}$ for $t \leq m+1$ and by $\xi_{t-m-1}^{t,m} = \bar{v}_{t-m-1}$ if $t > m+1$, where again $m = k/2$ and where $\mathbf{v}_0^{m-1}$ is defined to be $\bar{v}_0$ if $m = 0$. As is easily seen, this makes $\mathbf{h}_t^k$ a measurable function of $\mathbf{e}_{t+k}, \ldots, \mathbf{e}_{t-k}$. Given this definition of $\mathbf{h}_t^k$ we then have by analogous arguments as before that for k even and $t \leq m+1$

$$
\begin{aligned}
\left| \mathbf{v}_t - \mathbf{h}_t^k \right| \ \leq\ & d_v^t \left| \mathbf{v}_0 - \mathbf{v}_0^{m-1} \right| + \sum_{i=0}^{t-1} d_v^i d_{w,t-i} \left| \mathbf{w}_{t-i} - \mathbf{w}_{t-i}^m \right| \\
\leq\ & d_v \left| \mathbf{v}_0 - \mathbf{v}_0^{m-1} \right| + \sum_{i=0}^{t-1} d_v^i d_{w,t-i} \left| \mathbf{w}_{t-i} - \mathbf{w}_{t-i}^m \right|,
\end{aligned}
$$

whereas for $t > m+1$

$$
\begin{aligned}
\left| \mathbf{v}_t - \mathbf{h}_t^k \right| \ \leq\ & \sum_{i=0}^{m} d_v^i d_{w,t-i} \left| \mathbf{w}_{t-i} - \mathbf{w}_{t-i}^m \right| + d_v^t \left| \mathbf{v}_0 - \bar{v}_0 \right| \\
&+ d_v^{m+1} \sum_{i=1}^{t-m-1} d_v^{i-1} d_{w,t-m-i} \left| \mathbf{w}_{t-m-i} - \bar{w} \right|.
\end{aligned}
$$

Hence using the convention $\mathbf{w}_t^m = \mathbf{w}_t$ for $t \leq 0$ we obtain for all $t \geq 1$

$$
\left\| \mathbf{v}_t - \mathbf{h}_t^k \right\|_2 \ \leq\ d_v \left\| \mathbf{v}_0 - \mathbf{v}_0^{m-1} \right\|_2 + d_v^{m+1}(D + |\bar{v}_0|)
$$

$$
\begin{aligned}
& + d_w \sum_{i=0}^{m} d_v^i \left\| \mathbf{w}_{t-i} - \mathbf{w}_{t-i}^m \right\|_2 \\
& + d_v^{m+1} \sum_{i=1}^{t-m-1} d_v^{i-1} \left\| d_{w,t-m-i} \left| \mathbf{w}_{t-m-i} - \bar{w} \right| \right\|_2 \\
\leq \; & d_v \left\| \mathbf{v}_0 - \mathbf{v}_0^{m-1} \right\|_2 + d_v^{m+1}(D + |\bar{v}_0|) \\
& + d_w(1-d_v)^{-1} \sup_{t \geq 1} \left\| \mathbf{w}_t - \mathbf{w}_t^m \right\|_2 + D d_v^{m+1}(1-d_v)^{-1}.
\end{aligned}
$$

For k odd we obtain the same upper bound where now $m = (k-1)/2$. Therefore clearly $\sup_{t \geq 1} \left\| \mathbf{v}_t - \mathbf{h}_t^k \right\|_2$ is of size $-q$. The third claim in part (c) now follows since

$$
0 \leq \left\| \mathbf{v}_t - E(\mathbf{v}_t \mid \mathbf{e}_{t+k}, \ldots, \mathbf{e}_{t-k}) \right\|_2 \leq \left\| \mathbf{v}_t - \mathbf{h}_t^k \right\|_2
$$

in view of the minimum mean square error property of the conditional expectation. (Note that the second moment of $\mathbf{v}_t$ exists according to part (a).) ■

Proof of Theorem 6.10. Follows as a special case of Lemma D5. ■

Proof of Theorem 6.11. First observe that for $t \geq 1$ and any σ-field $\mathfrak{F}$ we have $\left\| g_t(E(\mathbf{v}_{t-1} \mid \mathfrak{F}), \mathbf{w}_t) \right\|_2 < \infty$, since

$$
\begin{aligned}
\left| g_t(E(\mathbf{v}_{t-1} \mid \mathfrak{F}), \mathbf{w}_t) \right| \; & \leq \; \left| g_t(E(\mathbf{v}_{t-1} \mid \mathfrak{F}), \mathbf{w}_t) - g_t(\mathbf{v}_{t-1}, \mathbf{w}_t) \right| \\
& \quad + \left| g_t(\mathbf{v}_{t-1}, \mathbf{w}_t) \right| \\
& \leq \; d_v \left| E(\mathbf{v}_{t-1} \mid \mathfrak{F}) - \mathbf{v}_{t-1} \right| + |\mathbf{v}_t| \\
& \leq \; d_v \left| E(\mathbf{v}_{t-1} \mid \mathfrak{F}) \right| + d_v |\mathbf{v}_{t-1}| + |\mathbf{v}_t|
\end{aligned}
$$

and since $\|\mathbf{v}_t\|_2 < \infty$ for $t \geq 0$ is assumed. Note that $E(\mathbf{v}_{t-1} \mid \mathfrak{F})$ belongs to V since V is assumed to be convex. Now for $m \geq 2$ we obtain

$$
\begin{aligned}
\nu_m \; & = \; \sup_{t \geq 1} \left\| \mathbf{v}_t - E(\mathbf{v}_t \mid \mathbf{w}_{t+m}, \ldots, \mathbf{w}_{t-m}) \right\|_2 \\
& \leq \; \sup_{t \geq 1} \left\| \mathbf{v}_t - E(\mathbf{v}_t \mid \mathbf{w}_{t-2+m}, \ldots, \mathbf{w}_{t-m}) \right\|_2 \\
& = \; \sup_{t \geq 1} \left\| \mathbf{v}_t - E\left(g_t(\mathbf{v}_{t-1}, \mathbf{w}_t) \mid \mathbf{w}_{t-2+m}, \ldots, \mathbf{w}_{t-m}\right) \right\|_2 \\
& \leq \; \sup_{t \geq 1} \left\| \mathbf{v}_t - g_t\left(E(\mathbf{v}_{t-1} \mid \mathbf{w}_{t-2+m}, \ldots, \mathbf{w}_{t-m}), \mathbf{w}_t\right) \right\|_2 \\
& = \; \sup_{t \geq 1} \left\| g_t(\mathbf{v}_{t-1}, \mathbf{w}_t) - g_t\left(E(\mathbf{v}_{t-1} \mid \mathbf{w}_{t-2+m}, \ldots, \mathbf{w}_{t-m}), \mathbf{w}_t\right) \right\|_2 \\
& \leq \; d_v \sup_{t \geq 1} \left\| \mathbf{v}_{t-1} - E(\mathbf{v}_{t-1} \mid \mathbf{w}_{t-2+m}, \ldots, \mathbf{w}_{t-m}) \right\|_2 = d_v \nu_{m-1}.
\end{aligned}
$$

Observing that $\nu_m \leq 2 \sup_{t \geq 0} \left\| \mathbf{v}_t \right\|_2 < \infty$ for $m \geq 1$ it follows immediately that ν_m decays exponentially, which completes the proof. ■

Proof of Theorem 6.12. Define for $i = 0, \ldots, k^* - 1$ the processes $(\mathbf{v}_\tau^i)_{\tau \in \mathbf{N}}$ by

$$\mathbf{v}_\tau^i = \phi_\tau^i \left(\mathbf{v}_{\tau-1}^i, \mathbf{u}_\tau^i\right), \qquad \tau \in \mathbf{N},$$

where the iterations are initialized with $\mathbf{v}_0^i = \mathbf{v}_i$, and where

$$\phi_\tau^i = g_{(\tau-1)k^*+i+1}^{(k^*)},$$

and

$$\mathbf{u}_\tau^i = \begin{bmatrix} \mathbf{w}_{(\tau-1)k^*+i+1} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{w}_{(\tau-1)k^*+i+k^*} \end{bmatrix}.$$

Note that every $\mathbf{v}_t$, $t \geq 0$, can be expressed as $\mathbf{v}_{(\tau-1)k^*+i+k^*}$ with $\tau \geq 0$, $0 \leq i \leq k^*-1$, and that $\mathbf{v}_{(\tau-1)k^*+i+k^*} = \mathbf{v}_\tau^i$ holds. I.e., we can represent the elements of the process $(\mathbf{v}_t)$ in terms of the elements of the processes $(\mathbf{v}_\tau^i)$, $i = 0, \ldots, k^* - 1$. To prove the theorem we now verify that the dynamic systems

$$\mathbf{v}_\tau^i = \phi_\tau^i \left(\mathbf{v}_{\tau-1}^i, \mathbf{u}_\tau^i\right)$$

satisfy the assumptions of Theorem 6.10 for every $i = 0, \ldots, k^*-1$. Clearly, the functions ϕ_τ^i, $\tau \in \mathbf{N}$, satisfy the Lipschitz condition for the system functions postulated in Theorem 6.10 for every $i = 0, \ldots, k^* - 1$ in view of (6.6). Observe further that $\|\mathbf{w}_t\|_r < \infty$ for every $t \in \mathbf{N}$ implies $\left\|\mathbf{u}_\tau^i\right\|_r < \infty$ for every $\tau \in \mathbf{N}$ and every $i = 0, \ldots, k^* - 1$, and that

$$\sup_{t \geq 1} \|\mathbf{w}_t\|_r < \infty$$

implies

$$\sup_{\tau \geq 1} \left\|\mathbf{u}_\tau^i\right\|_r < \infty.$$

To prove part (a) note that by definition $\left\|\mathbf{v}_0^i\right\|_r = \|\mathbf{v}_i\|_r$ and that $\|\mathbf{v}_i\|_r < \infty$ for $i = 0, \ldots, k^*-1$ by assumption. It now follows from Theorem 6.10(a) that $\left\|\mathbf{v}_\tau^i\right\|_r < \infty$ for all $\tau \in \mathbf{N}$ and all $i = 0, \ldots, k^* - 1$. Consequently, $\|\mathbf{v}_t\|_r < \infty$ for all $t \in \mathbf{N}$. The boundedness assumption on $g_t^{(k^*)}$ in part (a) is equivalent to

$$\sup_{\tau \geq 1} \left|\phi_\tau^i(\bar{v}, \bar{u})\right| < \infty$$

for all $i = 0, \ldots, k^*-1$, where $\bar{u} = (\bar{w}_1', \ldots, \bar{w}_{k^*}')'$. Hence, under this boundedness condition it follows further from Theorem 6.10(a) that

$$\sup_{\tau \geq 1} \left\|\mathbf{v}_\tau^i\right\|_r < \infty$$

for all $i = 0, \ldots, k^* - 1$. This in turn implies

$$\sup_{t \geq 1} \|\mathbf{v}_t\|_r < \infty$$

observing that $\|\mathbf{v}_i\|_r < \infty$ for $i = 0, \ldots, k^* - 1$ by assumption.

We next prove part (b). It follows immediately from Lemma 6.9(a) that the process $(\mathbf{w}_t', \ldots, \mathbf{w}_{t+k^*}')'$ is L_r-approximable by $(\mathbf{e}_{t+k^*})$, and hence by $(\mathbf{e}_t)$. Lemma 6.9(d) implies further that the processes $(\mathbf{u}_t^i)$ are L_r-approximable by $(\mathbf{e}_t^+)$ with

$$\mathbf{e}_t^+ = \left(\mathbf{e}_{(t-1)k^*+1}', \ldots, \mathbf{e}_{(t-1)k^*+k^*}'\right)'.$$

Theorem 6.10(b) now shows that the processes $(\mathbf{v}_t^i)$, $i = 0, \ldots, k^* - 1$, are L_r-approximable by $(\mathbf{e}_t^+)$. Applying Lemma 6.9(c) we see that the process $(\mathbf{v}_t)_{t \geq k^*}$ is L_r-approximable by $(\mathbf{e}_t^{++})$, where

$$\mathbf{e}_{(\tau-1)k^*+i+1}^{++} = \mathbf{e}_\tau^+$$

for $\tau \in \mathbf{Z}$ and $i = 0, \ldots, k^* - 1$. Observing that the random variables appearing in the argument list $(\mathbf{e}_{t+m}^{++}, \ldots, \mathbf{e}_{t-m}^{++})$ of an L_r-approximator w.r.t. the basis process $(\mathbf{e}_t^{++})$ also appear in the list $(\mathbf{e}_{t+m+k^*}, \ldots, \mathbf{e}_{t-m-k^*})$, this approximator can also be viewed as an L_r-approximator w.r.t. the basis process $(\mathbf{e}_t)$. It follows that $(\mathbf{v}_t)_{t \geq k^*}$ is L_r-approximable by $(\mathbf{e}_t)$. Since $\|\mathbf{v}_t\|_r < \infty$ for $t = 1, \ldots, k^* - 1$ in light of part (a) of the theorem, clearly also $(\mathbf{v}_t)_{t \in \mathbf{N}}$ is L_r-approximable by $(\mathbf{e}_t)$.

To prove part (c) we apply Lemma 6.9(a') to the process $(\mathbf{w}_{t+k^*})$. Observing that $\|\mathbf{w}_i - E(\mathbf{w}_i \mid \mathbf{e}_{i+m}, \ldots, \mathbf{e}_{i-m})\|_2$, for $i = 1, \ldots, k^*$, is clearly of size $-q$ it follows from that lemma that the process $(\mathbf{w}_t', \ldots, \mathbf{w}_{t+k^*}')'$ is near epoch dependent of size $-q$ on $(\mathbf{e}_{t+k^*})$, and hence on $(\mathbf{e}_t)$. Lemma 6.9(d') implies further that the processes $(\mathbf{u}_t^i)$ are near epoch dependent of size $-q$ on $(\mathbf{e}_t^+)$ with

$$\mathbf{e}_t^+ = \left(\mathbf{e}_{(t-1)k^*+1}', \ldots, \mathbf{e}_{(t-1)k^*+k^*}'\right)'.$$

Recall that

$$\|\mathbf{v}_0^i\|_2 = \|\mathbf{v}_i\|_2 < \infty$$

for $i = 0, \ldots, k^* - 1$. Theorem 6.10(c) now shows that the processes $(\mathbf{v}_t^i)$, $i = 0, \ldots, k^* - 1$, are near epoch dependent of size $-q$ on $(\tilde{\mathbf{e}}_t^{+i})$, where $\tilde{\mathbf{e}}_t^{+i} = (\mathbf{e}_t^{+\prime}, 0)'$ if $t \neq 1$ and $\tilde{\mathbf{e}}_1^{+i} = (\mathbf{e}_1^{+\prime}, \mathbf{v}_0^{i\prime})'$. Hence, in light of the minimum mean square error property of the conditional expectation, the processes $(\mathbf{v}_t^i)$ are then also near epoch dependent of size $-q$ on the common basis process $(\tilde{\mathbf{e}}_t^+)$, where

$$\tilde{\mathbf{e}}_t^+ = (\mathbf{e}_t^{+\prime}, 0, \ldots, 0)'$$

if $t \neq 1$ and

$$\tilde{\mathbf{e}}_1^+ = \left(\mathbf{e}_1^{+\prime}, \mathbf{v}_0^{0\prime}, \ldots, \mathbf{v}_0^{k^*-1\prime}\right)' = \left(\mathbf{e}_1^{+\prime}, \mathbf{v}_0', \ldots, \mathbf{v}_{k^*-1}'\right)'.$$

Observing that $\|\mathbf{v}_t\|_2 < \infty$ in view of the already established part (a) of the theorem, applying Lemma 6.9(c') shows that the process $(\mathbf{v}_t)_{t \geq k^*}$ is near epoch dependent of size $-q$ on $(\tilde{\mathbf{e}}_t^{++})$, where

$$\tilde{\mathbf{e}}_{(\tau-1)k^*+i+1}^{++} = \tilde{\mathbf{e}}_\tau^+$$

for $\tau \in \mathbf{Z}$ and $i = 0, \ldots, k^* - 1$. Consider the approximation error

$$\left\|\mathbf{v}_t - E\left(\mathbf{v}_t \mid \tilde{\mathbf{e}}_{t+m}^{++}, \ldots, \tilde{\mathbf{e}}_{t-m}^{++}\right)\right\|_2.$$

Observing that the random variables appearing in the list of conditioning variables $(\tilde{\mathbf{e}}_{t+m}^{++}, \ldots, \tilde{\mathbf{e}}_{t-m}^{++})$ also appear in the list $(\tilde{\mathbf{e}}_{t+m+k^*}, \ldots, \tilde{\mathbf{e}}_{t-m-k^*})$, it follows again from the minimum mean square error property of the conditional expectation that $(\mathbf{v}_t)_{t \geq k^*}$ is near epoch dependent of size $-q$ on $(\tilde{\mathbf{e}}_t)$. Certainly,

$$\|\mathbf{v}_i - E\left(\mathbf{v}_i \mid \tilde{\mathbf{e}}_{i+m}, \ldots, \tilde{\mathbf{e}}_{i-m}\right)\|_2 = 0, \quad i = 0, \ldots, k^* - 1,$$

for $m \geq k^*$ in view of the definition of $\tilde{\mathbf{e}}_t$. Hence also $(\mathbf{v}_t)_{t \in \mathbf{N}}$ is near epoch dependent of size $-q$ on $(\tilde{\mathbf{e}}_t)$. This proves the first claim in part (c). The second claim follows analogously observing that in this case Theorem 6.10(c) implies that the processes $(\mathbf{v}_t^i)$, $i = 0, \ldots, k^* - 1$, are near epoch dependent of size $-q$ on $(\mathbf{e}_t^+)$. ■

Lemma D6. *Let X be a metrizable space and let Θ be a compact metrizable space. Let $\{f_t : t \in \mathbf{N}\}$ with $f_t : X \times \Theta \to \mathbf{R}$ be equicontinuous on $X \times \Theta$. Define*

$$f_t^A(x) = \sup_{\theta \in A} f_t(x, \theta)$$

where $\emptyset \neq A \subseteq \Theta$. Then $\{f_t^A : t \in \mathbf{N}\}$ is equicontinuous on X.

Proof. Since $f_t^A = f_t^{\bar{A}}$, where $\bar{A}$ is the closure of A, we can assume w.l.o.g. that A is compact. Suppose that $\{f_t^A : t \in \mathbf{N}\}$ is not equicontinuous on X, i.e., there exists an $\epsilon > 0$ and a sequence $x_n \to x_0$ such that

$$\epsilon < \sup_t \left|f_t^A(x_n) - f_t^A(x_0)\right|.$$

Since A is compact, $f_t^A(x)$ is finite and hence we have

$$\epsilon < \sup_t \left|f_t^A(x_n) - f_t^A(x_0)\right| \leq \sup_{\theta \in A} \sup_t \left|f_t(x_n, \theta) - f_t(x_0, \theta)\right|.$$

Hence there exist $\theta_n \in A$ such that $\epsilon < \sup_t |f_t(x_n, \theta_n) - f_t(x_0, \theta_n)|$. Since A is compact there exists a convergent subsequence $\theta_{n_j} \to \theta_0$ as $j \to \infty$ with $\theta_0 \in A$. Consequently,

$$0 < \epsilon < \sup_t \left| f_t(x_{n_j}, \theta_{n_j}) - f_t(x_0, \theta_0) \right| + \sup_t \left| f_t(x_0, \theta_0) - f_t(x_0, \theta_{n_j}) \right|.$$

Since the r.h.s. goes to zero by equicontinuity this yields a contradiction. ■

Lemma D7. *Let (X, d) be a metric space and let Θ be some set. Let $\{f_t : t \in \mathbf{N}\}$ with $f_t : X \times \Theta \to \mathbf{R}$ be a family such that for each $\theta \in \Theta$ we have*

$$|f_t(x, \theta) - f_t(x^\bullet, \theta)| \le B_t(x, x^\bullet) d(x, x^\bullet)$$

for $x, x^\bullet \in X$ and $B_t : X \times X \to [0, \infty)$. Let $A \subseteq \Theta$ be nonempty. Suppose

$$f_t^A(x) = \sup_{\theta \in A} f_t(x, \theta)$$

is real valued, then

$$\left| f_t^A(x) - f_t^A(x^\bullet) \right| \le B_t(x, x^\bullet) d(x, x^\bullet).$$

Proof. The result follows immediately observing that

$$\sup_{\theta \in \Theta} |f_t(x, \theta) - f_t(x^\bullet, \theta)| \ge \left| f_t^A(x) - f_t^A(x^\bullet) \right|.$$

■

Proof of Theorem 6.13. That $q_t^*(z, \theta; \eta)$ and $q_{t*}(z, \theta; \eta)$ are real valued and Borel measurable for any $\theta \in \Theta$ and any $\eta > 0$ follows immediately from compactness of Θ and continuity of $q_t(z, \theta)$. It follows further from Lemma D6 that $\{q_t^*(., \theta; \eta) : t \in \mathbf{N}\}$ is equicontinuous on Z. Theorem 6.5 then implies that $(q_t^*(\mathbf{z}_t, \theta; \eta))$ is L_0-approximable by $(\mathbf{e}_t)$. The weak LLN for $q_t^*(\mathbf{z}_t, \theta; \eta)$ then follows from Theorem 6.3 since the moment condition in Theorem 6.3 clearly follows from the moment condition on $d_t(\mathbf{z}_t)$ maintained in the present theorem. The proof of the weak LLN for $q_{t*}(\mathbf{z}_t, \theta; \eta)$ is completely analogous. ■

Proof of Theorem 6.14. Since $q_t^*(z, \theta; \eta)$ and $q_{t*}(z, \theta; \eta)$ are real valued (and Borel measurable) for any $\theta \in \Theta$ and $\eta > 0$ small enough, it follows from the assumptions and Lemma D7 that $\{q_t^*(., \theta; \eta) : t \in \mathbf{N}\}$ satisfies Assumption 6.2 with $V = Z$. Theorem 6.6(a) then implies that $(q_t^*(\mathbf{z}_t, \theta; \eta))$ is L_0-approximable by $(\mathbf{e}_t)$. The weak LLN for $q_t^*(\mathbf{z}_t, \theta; \eta)$ then follows from

Theorem 6.3 since the moment condition in Theorem 6.3 clearly follows from the moment condition on $d_t(\mathbf{z}_t)$ maintained in the present theorem. The proof of the weak LLN for $q_{t*}(\mathbf{z}_t, \theta; \eta)$ is completely analogous. ■

Proof of Theorem 6.15. The theorem follows from Theorems 6.4 and 6.7(a) by similar arguments as in the proof of Theorem 6.14. ■

Appendix E

PROOFS FOR CHAPTER 7

Proof of Theorem 7.1. First note that $\bar{R}_n$ is well-defined in view of Assumption 7.1(c). We next verify that the assumptions of Theorems 5.1 and 5.2 are satisfied for $q_{ti}(z, \tau, \beta)$ and $(\mathbf{z}_t)$ under Assumptions 7.1, 7.3 and 7.1, 7.2, respectively. All assumptions are, apart from the local LLNs, immediately seen to hold. The local LLNs for q_{ti} follow from Theorems 6.14 and 6.13, respectively. In applying Theorem 6.14 to q_{ti} under Assumptions 7.1, 7.3 it is necessary to verify the $\mathfrak{A}$-measurability of

$$d_{ti}(\mathbf{z}_t) = \sup_{T\times B} |q_{ti}(\mathbf{z}_t, \tau, \beta)| .$$

Given the compactness of $T \times B$ the measurability follows from Assumption 7.3(c), cf. Footnote 29 in Chapter 6. In applying Theorem 6.13 under Assumptions 7.1, 7.2 the measurability of $d_{ti}(\mathbf{z}_t)$ is automatically guaranteed in view of continuity of q_{ti} and compactness of $T \times B$. (Note furthermore that the $\mathfrak{A}$-measurability of $d_t(\mathbf{z}_t)$ implicit in Assumption 7.1(c) is now seen to automatically hold under both sets of assumptions, since $d_t(\mathbf{z}_t) = \max_{1\le i\le p_q} d_{ti}(\mathbf{z}_t)$.) Applying Theorems 5.1 and 5.2 we obtain

$$\sup_{T\times B} \left| n^{-1}\sum_{t=1}^{n} q_t(\mathbf{z}_t, \tau, \beta) - n^{-1}\sum_{t=1}^{n} Eq_t(\mathbf{z}_t, \tau, \beta) \right| \to 0 \text{ i.p. as } n \to \infty$$

and that

$$\left\{ n^{-1}\sum_{t=1}^{n} Eq_t(\mathbf{z}_t, \tau, \beta) : n \in \mathbf{N} \right\}$$

is equicontinuous. The measurability of the above supremum follows from Assumption 7.3(d) and Assumptions 7.1(a), 7.2, respectively.

Assumption 7.1(c) clearly implies for some constant $c \in \mathbf{R}$ that

$$\left| n^{-1}\sum_{t=1}^{n} Eq_t(\mathbf{z}_t, \tau, \beta) \right| \le c$$

for all $(\tau, \beta) \in T \times B$ and all $n \in \mathbf{N}$; hence there is a compact set $K \subseteq \mathbf{R}^{p_q}$ such that

$$n^{-1}\sum_{t=1}^{n} Eq_t(\mathbf{z}_t, \tau, \beta) \in K$$

for all $(\tau, \beta) \in T \times B$ and all $n \in \mathbf{N}$. Lemma 3.3 then implies that

$$\sup_{T \times B} \left| \vartheta_n \left(n^{-1} \sum_{t=1}^{n} q_t(\mathbf{z}_t, \tau, \beta), \tau, \beta \right) - \vartheta_n \left(n^{-1} \sum_{t=1}^{n} E q_t(\mathbf{z}_t, \tau, \beta), \tau, \beta \right) \right| \to 0$$

in probability as $n \to \infty$. The measurability of the above supremum follows from Assumption 7.3(d) and Assumptions 7.1(a),(b), 7.2, respectively. Lemma 3.3 also implies that

$$\left\{ \vartheta_n \left(n^{-1} \sum_{t=1}^{n} E q_t(\mathbf{z}_t, \tau, \beta), \tau, \beta \right) : n \in \mathbf{N} \right\}$$

is equicontinuous, and hence is also uniformly equicontinuous since $T \times B$ is compact. Therefore Lemma 3.2 implies the validity of (7.3) and the (uniform) equicontinuity of $\{\bar{R}_n : n \in \mathbf{N}\}$ on B. The consistency result now follows from Lemma 3.1. ■

Appendix F

PROOFS FOR CHAPTER 8

Lemma F1.[1] *Let A_n, B_n be real or complex square random matrices. Let B_n be non-singular eventually [with probability tending to 1], let*

$$A_n - B_n \to 0 \quad a.s. \ [i.p.] \ as \ n \to \infty,$$

and let the sequences B_n as well as B_n^+ be bounded normwise a.s. [i.p.]. Then the sequences A_n and A_n^+ are bounded normwise a.s. [i.p.], A_n is non-singular eventually [with probability tending to 1] and

$$A_n^+ - B_n^+ \to 0 \quad a.s. \ [i.p.] \ as \ n \to \infty.$$

Proof. To prove the almost sure part of the lemma it suffices to consider non-random matrices. Without loss of generality we may then also assume that the matrices B_n are non-singular. For any complex matrix the absolute value of any of its characteristic roots is bounded by the norm of the matrix, see, e.g., MacDuffee (1956, p.28). Hence norm boundedness of B_n^{-1} implies that the (possibly complex) characteristic roots of B_n^{-1} are bounded in absolute value uniformly in n. Consequently, the characteristic roots of B_n are bounded away from zero in absolute value. By the same token the characteristic roots of B_n are bounded in absolute value uniformly in n. Consequently, $c_1 < |\det(B_n)| < c_2$ for all n for some finite positive constants c_1, c_2. Since boundedness of A_n follows immediately from $A_n - B_n \to 0$ and the assumptions, we see that also the characteristic roots of A_n are bounded in absolute value uniformly in n. Since $A_n - B_n \to 0$ and since B_n^{-1} is bounded it follows that $B_n^{-1}A_n \to I$, and hence $\det(A_n)/\det(B_n) \to 1$. Hence $c_1^* < |\det(A_n)| < c_2^*$ for positive finite constants c_1^*, c_2^* and for large n. This shows that eventually $A_n^+ = A_n^{-1}$ holds. Since the absolutely largest characteristic root of A_n is bounded as shown above, it follows further that the characteristic roots of A_n are bounded away from zero in absolute value for large n. Since A_n is bounded so is $\mathrm{adj}(A_n)$. Hence $|\det(A_n)| > c_1^*$ implies that A_n^+ is bounded as $A_n^+ = A_n^{-1} = \mathrm{adj}(A_n)/\det(A_n)$ for large n. Finally for large n we have $A_n^+ - B_n^{-1} = A_n^{-1} - B_n^{-1} = A_n^{-1}(B_n - A_n)B_n^{-1} \to 0$ by

[1]The norm of a complex matrix A is defined as the square root of the largest eigenvalue of A^*A, where A^* is the conjugate transpose of A.

assumption and since A_n^{-1} and B_n^{-1} are bounded. The in probability part of the lemma can be proved along analogous lines. ■

Proof of Lemma 8.1.[2] Assumption 8.1(c) implies the existence of a sequence $\Omega_n \in \mathfrak{A}$ with $P(\Omega_n) \to 1$ as $n \to \infty$ such that on respective events Ω_n the line segments between $(\hat{\tau}_n, \hat{\beta}_n)$ and $(\bar{\tau}_n, \bar{\beta}_n)$ lie in the interior of $T \times B$. We also may assume that the sets Ω_n are disjoint from the exceptional null set in Assumption 8.1(b). For ease of notation set $Q_n^*(\omega, \tau, \beta) = nN_n^{+\prime} Q_n(\mathbf{z}_1, ..., \mathbf{z}_n, \tau, \beta)$. Then for $\omega \in \Omega_n$ we have by the mean value theorem applied to $\nabla_{\beta'} Q_n^*$

$$\begin{aligned} \nabla_{\beta'} Q_n^*(\omega, \hat{\tau}_n, \hat{\beta}_n) &= \nabla_{\beta'} Q_n^*(\omega, \bar{\tau}_n, \bar{\beta}_n) \\ &\quad + \nabla_{\beta\beta} Q_n^*(\omega, \{\tilde{\tau}_n^i\}, \{\tilde{\beta}_n^i\})(\hat{\beta}_n - \bar{\beta}_n) \\ &\quad + \nabla_{\beta\tau} Q_n^*(\omega, \{\tilde{\tau}_n^i\}, \{\tilde{\beta}_n^i\})(\hat{\tau}_n - \bar{\tau}_n). \end{aligned} \tag{F.1}$$

Here $\nabla_{\beta\beta} Q_n^*(\omega, \{\tilde{\tau}_n^i\}, \{\tilde{\beta}_n^i\})$ denotes the matrix whose j-th row is the j-th row of $\nabla_{\beta\beta} Q_n^*$ evaluated at $(\omega, \tilde{\tau}_n^j, \tilde{\beta}_n^j)$, where $(\tilde{\tau}_n^j, \tilde{\beta}_n^j)$ is the mean value arising from the application of the mean value theorem to the j-th component of $\nabla_{\beta'} Q_n^*(\omega, \hat{\tau}_n, \hat{\beta}_n)$. (Lemma 3 in Jennrich (1969) implies that the mean values actually can be chosen to be measurable.) It now follows from Assumption 8.1(d) that

$$\begin{aligned} o_p(1) &= \nabla_{\beta'} Q_n^*(\omega, \bar{\tau}_n, \bar{\beta}_n) \\ &\quad + \nabla_{\beta\beta} Q_n^*(\omega, \{\tilde{\tau}_n^i\}, \{\tilde{\beta}_n^i\})(\hat{\beta}_n - \bar{\beta}_n) \\ &\quad + \nabla_{\beta\tau} Q_n^*(\omega, \{\tilde{\tau}_n^i\}, \{\tilde{\beta}_n^i\})(\hat{\tau}_n - \bar{\tau}_n). \end{aligned} \tag{F.2}$$

Since M_n and N_n are non-singular with probability tending to one we have that

$$\nabla_{\beta\beta} Q_n^*(\omega, \{\tilde{\tau}_n^i\}, \{\tilde{\beta}_n^i\}) \left[I - N_n^+ N_n\right] (\hat{\beta}_n - \bar{\beta}_n)$$

and

$$\nabla_{\beta\tau} Q_n^*(\omega, \{\tilde{\tau}_n^i\}, \{\tilde{\beta}_n^i\}) \left[I - M_n^+ M_n\right] (\hat{\tau}_n - \bar{\tau}_n)$$

[2] The proof given here differs slightly from the proof given in Pötscher and Prucha (1991b), which is not entirely correct. The problem in the latter proof is subtle and pertains to the step where we need to establish that $nN_n^{+\prime} \nabla_{\beta\beta} Q_n N_n^+ - C_n$ and $nN_n^{+\prime} \nabla_{\beta\tau} Q_n M_n^+$ converge to zero. Since the matrices of second order derivatives are evaluated row-wise at different mean values, the required convergence does not follow directly from Assumptions 8.1(e),(f). Rather, as demonstrated below, a more subtle argumentation is needed. (The reason for this is that the rows of the corresponding matrices of second order derivatives appearing in these assumptions are evaluated at one and the same value.) However, as is easily seen, the proof given in Pötscher and Prucha (1991b) is correct in case N_n and M_n are scalar multiples of the identity matrix.

are zero on a sequence of ω-sets of probability tending to one. Hence

$$\begin{aligned} o_p(1) &= \nabla_{\beta'} Q_n^*(\omega, \bar{\tau}_n, \bar{\beta}_n) \\ &\quad + \nabla_{\beta\beta} Q_n^*(\omega, \{\tilde{\tau}_n^i\}, \{\tilde{\beta}_n^i\}) N_n^+ N_n(\hat{\beta}_n - \bar{\beta}_n) \\ &\quad + \nabla_{\beta\tau} Q_n^*(\omega, \{\tilde{\tau}_n^i\}, \{\tilde{\beta}_n^i\}) M_n^+ M_n(\hat{\tau}_n - \bar{\tau}_n). \end{aligned} \tag{F.2'}$$

Assumption 8.1(f) implies that

$$\nabla_{\beta\tau} Q_n^*(\omega, \tilde{\tau}_n^j, \tilde{\beta}_n^j) M_n^+ = n N_n^{+\prime} \nabla_{\beta\tau} Q_n(\omega, \tilde{\tau}_n^j, \tilde{\beta}_n^j) M_n^+ = o_p(1)$$

for $j = 1, \ldots, p_\beta$. Since the j-th row of $\nabla_{\beta\tau} Q_n^*(\omega, \{\tilde{\tau}_n^i\}, \{\tilde{\beta}_n^i\}) M_n^+$ coincides with the j-th row of $\nabla_{\beta\tau} Q_n^*(\omega, \tilde{\tau}_n^j, \tilde{\beta}_n^j) M_n^+$ it follows that

$$\nabla_{\beta\tau} Q_n^*(\omega, \{\tilde{\tau}_n^i\}, \{\tilde{\beta}_n^i\}) M_n^+ = o_p(1).$$

Combining this with (F.2') and using Assumption 8.1(c) we arrive at

$$A_n N_n(\hat{\beta}_n - \bar{\beta}_n) = -\nabla_{\beta'} Q_n^*(\omega, \bar{\tau}_n, \bar{\beta}_n) + o_p(1), \tag{F.3}$$

where $A_n = \nabla_{\beta\beta} Q_n^*(\omega, \{\tilde{\tau}_n^i\}, \{\tilde{\beta}_n^i\}) N_n^+$. Multiplying (F.3) by the Moore-Penrose inverse of A_n and rearranging terms we get

$$\begin{aligned} N_n(\hat{\beta}_n - \bar{\beta}_n) &= \left[I - A_n^+ A_n\right] N_n(\hat{\beta}_n - \bar{\beta}_n) \\ &\quad - A_n^+ \nabla_{\beta'} Q_n^*(\omega, \bar{\tau}_n, \bar{\beta}_n) + A_n^+ o_p(1). \end{aligned} \tag{F.4}$$

Assumption 8.1(e) implies that

$$\nabla_{\beta\beta} Q_n^*(\omega, \tilde{\tau}_n^j, \tilde{\beta}_n^j) N_n^+ - C_n = n N_n^{+\prime} \nabla_{\beta\beta} Q_n(\omega, \tilde{\tau}_n^j, \tilde{\beta}_n^j) N_n^+ - C_n = o_p(1)$$

for every $j = 1, \ldots, p_\beta$. Since the j-th row of $\nabla_{\beta\beta} Q_n^*(\omega, \{\tilde{\tau}_n^i\}, \{\tilde{\beta}_n^i\}) N_n^+$ coincides with the j-th row of $\nabla_{\beta\beta} Q_n^*(\omega, \tilde{\tau}_n^j, \tilde{\beta}_n^j) N_n^+$ it follows that

$$A_n - C_n = \nabla_{\beta\beta} Q_n^*(\omega, \{\tilde{\tau}_n^i\}, \{\tilde{\beta}_n^i\}) N_n^+ - C_n = o_p(1).$$

Since the matrices C_n are non-singular with probability tending to one, this together with Lemma F1 implies that the first term on the r.h.s. of (F.4) is zero on a sequence of ω-sets of probability tending to one, and furthermore that $|A_n^+| = O_p(1)$ and hence that the last term on the r.h.s. of (F.4) is $o_p(1)$. This gives

$$\begin{aligned} N_n(\hat{\beta}_n - \bar{\beta}_n) &= -A_n^+ n N_n^{+\prime} \nabla_{\beta'} Q_n(\omega, \bar{\tau}_n, \bar{\beta}_n) + o_p(1) \\ &= C_n^+ D_n \zeta_n + o_p(1) \end{aligned} \tag{F.5}$$

where $\nabla_{\beta'} Q_n(\omega, \bar{\tau}_n, \bar{\beta}_n)$ is shorthand for $\nabla_{\beta'} Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \bar{\tau}_n, \bar{\beta}_n)$. The last equality in (F.5) follows from Assumption 8.1(g), and since $|A_n^+| = O_p(1)$ and $A_n^+ - C_n^+ = o_p(1)$ by Lemma F1 and $A_n - C_n = o_p(1)$. Observe also that

$|D_n| = O_p(1)$ and $|\zeta_n| = O_p(1)$ hold. The remaining parts of the lemma are obvious. ∎

Lemma F2. *Let (η_n) and (ξ_n) be sequences of random vectors in $\mathbf{R}^p$ with $\eta_n = \xi_n + o_p(1)$, and let H_n^η and H_n^ξ denote the distributions of η_n and ξ_n, respectively. Assume that the sequence $(H_n^\xi : n \in \mathbf{N})$ is tight. Let $\mathcal{H}$ be the set of all accumulation points of this sequence in the topology of weak convergence. Then for every Borel subset C of $\mathbf{R}^p$ with $H(\partial C) = 0$ for all $H \in \mathcal{H}$ we have*

$$H_n^\eta(C) - H_n^\xi(C) \to 0 \text{ as } n \to \infty.$$

(Here ∂C denotes the boundary of C.)

Proof. Let C be as in the lemma. Assume $H_n^\eta(C) - H_n^\xi(C)$ does not converge to zero. Then there exists an $\epsilon > 0$ and a subsequence (n_i) such that

$$\left|H_{n_i}^\eta(C) - H_{n_i}^\xi(C)\right| > \epsilon$$

for all i. Because of tightness and Prokhorov's theorem there is a subsequence (n_{i_j}) of (n_i) with

$$H_{n_{i_j}}^\xi \xrightarrow{D} H \in \mathcal{H}.$$

Since $\eta_{n_{i_j}}$ and $\xi_{n_{i_j}}$ differ only by a term that is $o_p(1)$ we also get

$$H_{n_{i_j}}^\eta \xrightarrow{D} H.$$

Since $H(\partial C) = 0$ by assumption we obtain

$$H_{n_{i_j}}^\xi(C) \to H(C)$$

and

$$H_{n_{i_j}}^\eta(C) \to H(C),$$

which contradicts

$$\left|H_{n_i}^\eta(C) - H_{n_i}^\xi(C)\right| > \epsilon$$

for all i. ∎

Remark. (a) It is readily seen that under the assumptions of the above lemma also the sequence $(H_n^\eta : n \in \mathbf{N})$ is tight and that this sequence also has $\mathcal{H}$ as the set of its accumulation points.

(b) If $x \in \mathbf{R}^p$ is a point of continuity of every cumulative distribution function F corresponding to a $H \in \mathcal{H}$, then $F_n^\eta(x) - F_n^\xi(x) \to 0$ as $n \to \infty$, where F_n^η and F_n^ξ denote the cumulative distribution functions of η_n and

ξ_n, respectively. This follows from Lemma F2 by choosing $C = (-\infty, x]$ and observing that x is a continuity point of F if and only if $H(\partial C) = 0$.

Corollary F3. *Let (η_n) and (ζ_n) be sequences of random vectors in $\mathbf{R}^p$ and $\mathbf{R}^q$, respectively, and let (A_n) be a sequence of bounded (non-random) $p \times q$ matrices. Suppose*

$$\eta_n = A_n \zeta_n + o_p(1)$$

and

$$\zeta_n \xrightarrow{D} \zeta.$$

Let $\xi_n = A_n \zeta_n$ and $\psi_n = A_n \zeta$.

(a) Then $(H_n^\xi : n \in \mathbf{N})$ is tight, and hence the conclusions of Lemma F2 hold. Furthermore, the set $\mathcal{H}$ consists of all distributions of random variables of the form $A\zeta$, where A is an accumulation point of the sequence (A_n).

(b) If H_n^ψ denotes the distribution of ψ_n, then $(H_n^\psi : n \in \mathbf{N})$ is tight and its set of accumulation points coincides with $\mathcal{H}$. Furthermore,

$$H_n^\eta(C) - H_n^\psi(C) \to 0 \text{ as } n \to \infty$$

for every Borel set C of $\mathbf{R}^p$ which satisfies $H(\partial C) = 0$ for every $H \in \mathcal{H}$.

Proof. Tightness of the sequence H_n^ξ follows from boundedness of (A_n) and $\zeta_n \xrightarrow{D} \zeta$. Part (a) then follows immediately from Lemma F2. The first two claims in part (b) are readily verified. To prove the last claim in part(b) it suffices to show that $H_n^\xi(C) - H_n^\psi(C) \to 0$. This follows from a subsequence argument similar to the one in the proof of Lemma F2, observing that along subsequences (n_i) for which A_{n_i} converges to A, say, both ξ_{n_i} and ψ_{n_i} converge to $A\zeta$ in distribution. ■

Corollary F4. *Suppose the assumptions of Corollary F3 hold. Let F_n^η, F_n^ξ, and F_n^ψ be the cumulative distribution functions of η_n, $\xi_n = A_n\zeta_n$, and $\psi_n = A_n\zeta$.*

(a) Assume that $A\zeta$ has a continuous distribution for any accumulation point A of the sequence (A_n). Then

$$F_n^\eta(x) - F_n^\xi(x) \to 0 \text{ and } F_n^\eta(x) - F_n^\psi(x) \to 0 \text{ as } n \to \infty$$

for every $x \in \mathbf{R}^p$.

(b) Assume that ζ is distributed $N(\mu, \Sigma)$, with Σ positive definite and that $\liminf_{n\to\infty} \lambda_{\min}(A_n A_n') > 0$ holds. Then

$$F_n^\eta(x) - F_n^\xi(x) \to 0 \text{ and } F_n^\eta(x) - F_n^\psi(x) \to 0 \text{ as } n \to \infty$$

for every $x \in \mathbf{R}^p$. Of course, F_n^ψ is here the cumulative distribution function of a normal distribution with mean $A_n\mu$ and variance covariance matrix $A_n\Sigma A_n'$.

Proof. Part (a) is a consequence of Corollary F3 and Remark (b) after Lemma F2. Part (b) follows from part (a) since the condition on the smallest eigenvalues of A_nA_n' ensures that any accumulation point A has full row rank. Hence the distribution of $A\zeta$ is a nondegenerate normal distribution on $\mathbf{R}^p$ and hence is continuous. ■

Proof of Corollary 8.2. To prove part (a') of the corollary observe that

$$(D_n'D_n)^+D_n'C_nN_n(\hat{\beta}_n - \bar{\beta}_n) = (D_n'D_n)^+D_n'C_nC_n^+D_n\zeta_n + o_p(1),$$

since $|C_n| = O_p(1)$, $|D_n| = O_p(1)$ by Assumptions 8.1(e),(g), and since

$$\left|(D_n'D_n)^+\right| = O_p(1)$$

has been assumed in part (a') of the corollary. The result then follows observing that

$$\left[(D_n'D_n)^+D_n'C_nC_n^+D_n - I\right]\zeta_n = 0$$

on a sequence of ω-sets of probability tending to one. Part (a) follows as a special case. To prove part (b') observe from Lemma 8.1 that

$$G_nN_n(\hat{\beta}_n - \bar{\beta}_n) = G_nC_n^{-1}D_n\zeta_n + o_p(1)$$

with

$$G_n = U_n'\left[(C_n^{-1}D_n\Sigma D_n'C_n^{-1\prime})^+\right]^{1/2},$$

since G_n is bounded. The boundedness of G_n is seen as follows: Since $|U_n| = 1$, $\lambda_{\min}(\Sigma) > 0$,

$$C_n^{-1}D_n\Sigma D_n'C_n^{-1\prime} \geq \lambda_{\min}(\Sigma)C_n^{-1}D_nD_n'C_n^{-1\prime},$$

and since $C_n^{-1}D_n\Sigma D_n'C_n^{-1\prime}$ and $C_n^{-1}D_nD_n'C_n^{-1\prime}$ have the same rank d it suffices to show that the d non-zero eigenvalues of $C_n^{-1}D_nD_n'C_n^{-1\prime}$ are bounded away from zero. The non-zero eigenvalues of $C_n^{-1}D_nD_n'C_n^{-1\prime}$ coincide with the non-zero eigenvalues of $D_n'C_n^{-1\prime}C_n^{-1}D_n$. Since clearly

$$D_n'C_n^{-1\prime}C_n^{-1}D_n \geq \lambda_{\min}(C_n^{-1\prime}C_n^{-1})D_n'D_n = \lambda_{\max}^{-1}(C_nC_n')D_n'D_n,$$

since $\lambda_{\max}(C_nC_n') = \lambda_{\max}(C_n'C_n)$ and since the latter is bounded by Assumption 8.1(e), it suffices to show that the d non-zero eigenvalues of $D_n'D_n$, or equivalently of D_nD_n', are bounded away from zero. However, this is implied by the assumption that $|(D_n D_n')^+| = O(1)$. Given the boundedness of G_n has been established, it now suffices to show that

$$G_n C_n^{-1} D_n \zeta_n \xrightarrow{D} N\left(0, \operatorname{diag}(I_d, 0)\right).$$

Since $G_n C_n^{-1} D_n \zeta_n$ and $G_n C_n^{-1} D_n \tilde{\zeta}_n$ have the same distribution and since

$$G_n C_n^{-1} D_n \tilde{\zeta}_n = G_n C_n^{-1} D_n \tilde{\zeta} + o(1) \text{ a.s.},$$

where $\tilde{\zeta}_n$ and $\tilde{\zeta}$ are as in the discussion following Lemma 8.1, it suffices to show that $G_n C_n^{-1} D_n \tilde{\zeta}$ is distributed as $N(0,\operatorname{diag}(I_d, 0))$. Since $\tilde{\zeta}$ and ζ have the same distribution it follows that $G_n C_n^{-1} D_n \tilde{\zeta}$ is distributed $N(0, \Phi)$ with $\Phi = G_n C_n^{-1} D_n \Sigma (G_n C_n^{-1} D_n)'$. Substituting the expression for G_n into Φ, using the fact that U_n diagonalizes $C_n^{-1} D_n \Sigma D_n' C_n^{-1\prime}$ by definition, and observing that $\operatorname{rank}(D_n) = d$, it is readily seen that $\Phi = \operatorname{diag}(I_d, 0)$. Part (b) follows as a special case of part (b') after premultiplication of $G_n C_n^{-1} D_n \tilde{\zeta}$ by U_n. ■

Appendix G

PROOFS FOR CHAPTER 10

Proof of Theorem 10.1. (a) We apply the Cramér-Wold device. Let c be a $p_v \times 1$ vector with $c'c = 1$. Define

$$X_{nt} = c' V_n^{-1/2} \mathbf{v}_t.$$

Then X_{nt} is a square-integrable martingale difference array w.r.t. the filtration $(\mathfrak{F}_{nt})$ where $\mathfrak{F}_{nt} = \mathfrak{F}_t$. We verify the conditions (3.18) - (3.21) of Theorem 3.2 in Hall and Heyde (1980). Clearly condition (3.21) is satisfied. (Actually, within the context of Theorem 10.1 this condition is not required to hold, cf. Hall and Heyde (1980, p.59).) In light of the discussion in Hall and Heyde (1980, p.53), condition (3.18) is equivalent to the condition that

$$\sum_{t=1}^{n} X_{nt}^2 \mathbf{1}\left(|X_{nt}| > \epsilon\right) \to 0 \text{ i.p. as } n \to \infty$$

for every $\epsilon > 0$. We verify the latter condition by showing that even

$$E \sum_{t=1}^{n} X_{nt}^2 \mathbf{1}\left(|X_{nt}| > \epsilon\right) \to 0 \text{ as } n \to \infty.$$

This is the case since

$$\begin{aligned}
& E \sum_{t=1}^{n} X_{nt}^2 \mathbf{1}\left(|X_{nt}| > \epsilon\right) \\
= \; & E \sum_{t=1}^{n} |X_{nt}|^{2+\delta} \mathbf{1}\left(|X_{nt}| > \epsilon\right) / |X_{nt}|^{\delta} \\
\leq \; & n\epsilon^{-\delta} \left(n^{-1} \sum_{t=1}^{n} E\, |X_{nt}|^{2+\delta} \right) \\
= \; & n\epsilon^{-\delta} \left(n^{-1} \sum_{t=1}^{n} E \left| c' V_n^{-1/2} \mathbf{v}_t \right|^{2+\delta} \right) \\
\leq \; & n\epsilon^{-\delta} \left| V_n^{-1/2} \right|^{2+\delta} \left(n^{-1} \sum_{t=1}^{n} E\, |\mathbf{v}_t|^{2+\delta} \right) \\
= \; & n^{-\delta/2} \epsilon^{-\delta} \left(\lambda_{\min}(n^{-1} V_n) \right)^{-1-\delta/2} \left(n^{-1} \sum_{t=1}^{n} E\, |\mathbf{v}_t|^{2+\delta} \right) \\
= \; & o(1)
\end{aligned}$$

by the assumptions of the theorem and since

$$\left|V_n^{-1/2}\right| = (\lambda_{\min}(V_n))^{-1/2}$$

holds. We next verify condition (3.19). Since

$$n^{-1}\sum_{t=1}^{n} E\left|\mathbf{v}_t\right|^{2+\delta} < \infty$$

holds $(\mathbf{v}_t)$ is tight and hence $\mathbf{v}_t^i\mathbf{v}_t^j$ is L_0-approximable by $(\mathbf{e}_t)$ in view of Theorem 6.5 applied to the functions $g(v) = v^i v^j$, where $\mathbf{v}_t^i$ and v^i denote the i-th component of the vectors $\mathbf{v}_t$ and v. Consequently

$$n^{-1}\sum_{t=1}^{n}(\mathbf{v}_t\mathbf{v}_t' - E\mathbf{v}_t\mathbf{v}_t') \to 0$$

in probability as $n \to \infty$ by Theorem 6.3 applied to $\mathbf{v}_t^i\mathbf{v}_t^j$. Observing that

$$\left|n^{1/2}V_n^{-1/2}\right| = \left(\lambda_{\min}(n^{-1}V_n)\right)^{-1/2}$$

is bounded (for large n), it follows that

$$\sum_{t=1}^{n} X_{nt}^2 - 1 = nc'V_n^{-1/2}\left(n^{-1}\sum_{t=1}^{n}(\mathbf{v}_t\mathbf{v}_t' - E\mathbf{v}_t\mathbf{v}_t')\right)V_n^{-1/2}c \to 0$$

in probability as $n \to \infty$. Finally condition (3.20) follows since

$$E\left(\max_{1\le t\le n} X_{nt}^2\right) \le E\sum_{t=1}^{n} X_{nt}^2 = c'c = 1.$$

(b) Again using the Cramér-Wold device consider $\mathbf{u}_t = c'\mathbf{v}_t$, $c'c = 1$. Let

$$U_n = E\left[\left(\sum_{t=1}^{n}\mathbf{u}_t\right)^2\right] = c'V_n c.$$

Then $n^{-1}U_n \to U^* = c'V^*c$. Consider first the case where $U^* > 0$. Then $(\mathbf{u}_t)$ satisfies, as is easily seen, all assumptions of part (a) of the theorem. Hence

$$U_n^{-1/2}\sum_{t=1}^{n}\mathbf{u}_t \xrightarrow{D} N(0,1).$$

Since $n^{-1/2}U_n^{1/2} \to (U^*)^{1/2}$ we obtain

$$n^{-1/2}\sum_{t=1}^{n}\mathbf{u}_t \xrightarrow{D} N(0,U^*).$$

Next consider the case $U^* = 0$. But then

$$\left\| n^{-1/2} \sum_{t=1}^{n} \mathbf{u}_t \right\|_2 = n^{-1} U_n \to U^* = 0.$$

Since L_2-convergence implies convergence in distribution we have also in this case

$$n^{-1/2} \sum_{t=1}^{n} \mathbf{u}_t \xrightarrow{D} N(0, U^*)$$

with $U^* = 0$, where $N(0,0)$ denotes the degenerate normal distribution concentrated at zero. ■

Proof of Theorem 10.2. (a) Follows from Corollary 4.4 in Wooldridge (1986) or Theorem 5.3 in Gallant and White (1988) and the Cramér-Wold device. (b) Is proved analogously as part (b) of Theorem 10.1. ■

Appendix H

PROOFS FOR CHAPTER 11

Proof of Lemma 11.1. Let f_t be as in Assumption 11.2. To prove the lemma we verify the assumptions of Theorem 5.2 for f_t. Assumption 5.1 is satisfied since $T' \times B'$ is compact by Assumption 11.1(d). Assumptions B and C for $\{f_t : t \in \mathrm{N}\}$ are implied by Assumption 11.2. Assumption D follows from Assumption 11.1(f). Assumption 5.2 now follows from Theorem 6.13 in light of Assumption 11.1(e). Theorem 5.2(b) implies, in particular, continuity of $Ef_t(\mathbf{z}_t, \tau, \beta)$. This, together with continuity of f_t and the compactness of $T' \times B'$, implies measurability of the suprema in Lemma 11.1. Theorem 5.2(a) then completes the proof. ∎

Proof of Theorem 11.2. We first prove part (a) of the theorem. To prove the result we verify Assumption 8.1 maintained by Lemma 8.1 and Corollary 8.2. Assumptions 8.1(a)-(d) follow immediately from Assumptions 11.1(a)-(d) upon setting $M_n = n^{1/2}I$ and $N_n = n^{1/2}I$. From Lemma 11.1 we have that

$$\sup_{T' \times B'} |\nabla_{\beta\beta} Q_n - C_{1n}| \to 0 \text{ i.p. as } n \to \infty$$

and

$$\sup_{T' \times B'} |\nabla_{\beta\tau} Q_n - C_{2n}| \to 0 \text{ i.p. as } n \to \infty$$

and that C_{1n} and C_{2n} are equicontinuous on $T' \times B'$. For any sequence $(\tilde{\tau}_n, \tilde{\beta}_n)$ as in Assumption 8.1(e) we can in light of Assumption 11.1(d) assume w.l.o.g. that $(\tilde{\tau}_n, \tilde{\beta}_n) \in T' \times B'$ possibly after redefining $(\tilde{\tau}_n, \tilde{\beta}_n)$ on ω-sets, where the probability of those sets tends to zero as $n \to \infty$. Since

$$(\tilde{\tau}_n, \tilde{\beta}_n) - (\bar{\tau}_n, \bar{\beta}_n) \to 0 \text{ i.p. as } n \to \infty,$$

it follows then from Lemma 3.2 that

$$\nabla_{\beta\beta} Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \tilde{\tau}_n, \tilde{\beta}_n) - C_{1n}(\bar{\tau}_n, \bar{\beta}_n) \to 0 \text{ i.p. as } n \to \infty$$

and

$$\nabla_{\beta\tau} Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \tilde{\tau}_n, \tilde{\beta}_n) - C_{2n}(\bar{\tau}_n, \bar{\beta}_n) \to 0 \text{ i.p. as } n \to \infty.$$

Clearly, $C_{1n}(\bar{\tau}_n, \bar{\beta}_n) = C_n$, and $C_{2n}(\bar{\tau}_n, \bar{\beta}_n) = 0$ by Assumption 11.3(b). This implies Assumptions 8.1(e),(f), observing furthermore that $|C_n| =$

$O(1)$ by the dominance condition in Assumption 11.2 and that $|C_n^+| = |C_n^{-1}| = O(1)$ by Assumption 11.3(c). It remains that we verify Assumption 8.1(g) under either Assumption 11.4 or 11.5. Putting

$$V_n = n^2 E\left[\nabla_{\beta'}\underline{S}_n \nabla_{\beta}\underline{S}_n\right]$$

observe that

$$\liminf_{n\to\infty} \lambda_{\min}\left(n^{-1}V_n\right) > 0$$

by Assumption 11.3(d). Now let Assumption 11.4 hold and put $\mathbf{v}_t = \nabla_{\beta'}q_t(\mathbf{z}_t, \bar{\tau}, \bar{\beta})$. Then applying Theorem 6.5 to $\nabla_{\beta'}q_t(z, \bar{\tau}, \bar{\beta})$ and $(\mathbf{z}_t)$ and making use of Assumptions 11.1(e),(f) and Assumption 11.4, it follows that $(\mathbf{v}_t)$ is L_0-approximable by $(\mathbf{e}_t)$. From Assumption 11.4 and Theorem 10.1(a) it follows that

$$-n^{1/2}D_n^{-1}\nabla_{\beta'}\underline{S}_n \xrightarrow{D} N(0, I).$$

The boundedness of $n^{-1}V_n$ follows from the discussion after Theorem 10.2. Next suppose Assumption 11.5 holds and put

$$\mathbf{v}_{t,n} = \nabla_{\beta'}q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n) - E\nabla_{\beta'}q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n).$$

Then the same conclusion now follows from Theorem 10.2(a) observing that

$$E\left[\left(\sum_{t=1}^{n}\mathbf{v}_{t,n}\right)\left(\sum_{t=1}^{n}\mathbf{v}_{t,n}'\right)\right] = V_n = n^2 E\left[\nabla_{\beta'}\underline{S}_n \nabla_{\beta}\underline{S}_n\right]$$

and that $E\nabla_{\beta}\underline{S}_n = 0$ for all n. The boundedness of $n^{-1}V_n$ follows again from the discussion after Theorem 10.2. Since D_n is the square root of $n^{-1}V_n$, the boundedness of D_n now follows (under both sets of assumptions) from the boundedness of $n^{-1}V_n$. This establishes Assumption 8.1(g) with

$$\zeta_n = -n^{1/2}D_n^{-1}\nabla_{\beta'}\underline{S}_n.$$

Furthermore, D_n is nonsingular for large n and $|D_n^+| = |D_n^{-1}| = O(1)$ by Assumption 11.3(d). Part (a) of the theorem now follows from Lemma 8.1 and Corollary 8.2(a).

We next prove part (b) of the theorem. The proof is identical to that of part (a) up to the point of verification of Assumption 8.1(g) under either Assumption 11.4 or 11.5. Putting

$$V_n = n^2 E\left[\nabla_{\beta'}\underline{S}_n \nabla_{\beta}\underline{S}_n\right]$$

observe that $n^{-1}V_n \to \Lambda$ by assumption. Now let Assumption 11.4 hold and put $\mathbf{v}_t = \nabla_{\beta'}q_t(\mathbf{z}_t, \bar{\tau}, \bar{\beta})$. Then exactly as in the proof of part (a) it follows that $(\mathbf{v}_t)$ is L_0-approximable by $(\mathbf{e}_t)$. From Assumption 11.4 and Theorem 10.1(b) it follows that

$$-n^{1/2}\nabla_{\beta'}\underline{S}_n \xrightarrow{D} N(0, \Lambda).$$

Next suppose Assumption 11.5 holds and put

$$\mathbf{v}_{t,n} = \nabla_{\beta'} q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n) - E\nabla_{\beta'} q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n).$$

Then the same conclusion now follows from Theorem 10.2(b) observing that

$$E\left[\left(\sum_{t=1}^{n} \mathbf{v}_{t,n}\right)\left(\sum_{t=1}^{n} \mathbf{v}'_{t,n}\right)\right] = V_n = n^2 E\left[\nabla_{\beta'} \underline{S}_n \nabla_{\beta} \underline{S}_n\right]$$

and that $E\nabla_{\beta}\underline{S}_n = 0$ for all n. This establishes Assumption 8.1(g) with

$$\zeta_n = -n^{1/2}\nabla_{\beta'}\underline{S}_n$$

and $D_n = I$. Part (b) of the theorem now follows from Lemma 8.1 and the fact that $|C_n| = O(1)$. ■

Theorem H1. *Suppose the assumptions of Theorem 11.2(b) hold and let r denote the rank of Λ.*

(i) Let U_1 be an orthogonal matrix of eigenvectors of Λ, where the first r columns correspond to the nonzero eigenvalues. Then

$$(\Lambda^+)^{1/2} C_n n^{1/2}(\hat{\beta}_n - \bar{\beta}_n) \xrightarrow{D} N(0, \Xi_1),$$

where

$$\Xi_1 = U_1 \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} U_1'.$$

(ii) Suppose $C_n \to C$ and U_2 is an orthogonal matrix of eigenvectors of $C^{-1}\Lambda C^{-1\prime}$, where the first r columns correspond to the nonzero eigenvalues. Then

$$\begin{aligned} n^{1/2}(\hat{\beta}_n - \bar{\beta}_n) &\xrightarrow{D} N(0, C^{-1}\Lambda C^{-1\prime}), \\ C n^{1/2}(\hat{\beta}_n - \bar{\beta}_n) &\xrightarrow{D} N(0, \Lambda), \\ (\Lambda^+)^{1/2} C n^{1/2}(\hat{\beta}_n - \bar{\beta}_n) &\xrightarrow{D} N(0, \Xi_1), \end{aligned}$$

and

$$\left((C^{-1}\Lambda C^{-1\prime})^+\right)^{1/2} n^{1/2}(\hat{\beta}_n - \bar{\beta}_n) \xrightarrow{D} N(0, \Xi_2),$$

where

$$\Xi_2 = U_2 \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} U_2'.$$

(Of course, if Λ is nonsingular then Ξ_1 and Ξ_2 reduce to the identity matrix.)

Proof. By Theorem 11.2(b) we have

$$C_n n^{1/2}(\hat{\beta}_n - \bar{\beta}_n) \xrightarrow{D} N(0, \Lambda). \tag{H.1}$$

Part (i) now follows from $(\Lambda^+)^{1/2}\Lambda(\Lambda^+)^{1/2} = \Xi_1$ which is a well-known property of the Moore-Penrose inverse. The first two distributional convergence results in part (ii) follow immediately from (H.1) and $C_n \to C$, observing that C is nonsingular as a consequence of $|C_n^{-1}| = O(1)$. The last two distributional convergence results follow analogously to part (i) observing that

$$(\Lambda^+)^{1/2}\Lambda(\Lambda^+)^{1/2} = \Xi_1$$

and

$$\left((C^{-1}\Lambda C^{-1\prime})^+\right)^{1/2}(C^{-1}\Lambda C^{-1\prime})\left((C^{-1}\Lambda C^{-1\prime})^+\right)^{1/2} = \Xi_2.$$

■

Proof of Lemma 11.3. Let f_t be as in Assumption 11.6(c). Part (a) of the lemma follows analogously as Lemma 11.1. To prove part (b) of the lemma observe that because of Assumption 11.6(c) there exist compact subsets K_i of respective Euclidean spaces such that $ES_n \in K_1$, $E\nabla_\beta S_n \in K_2$, $E\nabla_\tau S_n \in K_3$, $E\nabla_{\beta\tau} S_n \in K_4$ and $E\nabla_{\beta\beta} S_n \in K_5$ for all $(\tau, \beta) \in T' \times B'$ and all n. Assumption 11.6(a) implies clearly that the restrictions of G_{1n} and G_{2n} to the Cartesian product of the appropriate Euclidean space with $T' \times B'$ are equicontinuous on the subset $K \times T' \times B'$ for any compact subset K of the appropriate Euclidean space. Putting $K = K_1 \times K_2 \times K_5$ for G_{1n} and $K = K_1 \times K_2 \times K_3 \times K_4$ for G_{2n}, part (b) follows from part (a) and Lemma 3.3. The measurability of the suprema in part (b) follows from the continuity of the functions involved and from compactness of $T' \times B'$. ■

Proof of Lemma 11.4. From a Taylor series expansion of

$$\nabla_\beta Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \bar{\tau}_n, \bar{\beta}_n) = \nabla_c \vartheta_n(\underline{S}_n, \bar{\tau}_n, \bar{\beta}_n)\nabla_\beta \underline{S}_n + \nabla_\beta \vartheta_n(\underline{S}_n, \bar{\tau}_n, \bar{\beta}_n)$$

around $E\underline{S}_n$ we obtain

$$\begin{aligned}
&\nabla_\beta Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \bar{\tau}_n, \bar{\beta}_n) - \nabla_\beta \bar{Q}_n(\bar{\tau}_n, \bar{\beta}_n) \\
= \; &\nabla_c \vartheta_n(E\underline{S}_n, \bar{\tau}_n, \bar{\beta}_n)(\nabla_\beta \underline{S}_n - E\nabla_\beta \underline{S}_n) \\
&+ (\underline{S}_n - E\underline{S}_n)' \left[\nabla_{cc}\vartheta_n(\{\xi_n^i\}, \bar{\tau}_n, \bar{\beta}_n)\nabla_\beta \underline{S}_n + \nabla_{c\beta}\vartheta_n(\{\chi_n^i\}, \bar{\tau}_n, \bar{\beta}_n)\right]
\end{aligned}$$

observing that $E\nabla_\beta \underline{S}_n = \nabla_\beta E\underline{S}_n$ in light of the dominance condition in Assumption 11.6(c). Here the matrix $\nabla_{cc}\vartheta_n(\{\xi_n^i\}, \bar{\tau}_n, \bar{\beta}_n)$ denotes the matrix whose j-th column coincides with the j-th column of the matrix $\nabla_{cc}\vartheta_n(\xi_n^j, \bar{\tau}_n, \bar{\beta}_n)$, where the ξ_n^j denote mean values. The matrix $\nabla_{c\beta}\vartheta_n(\{\chi_n^i\}, \bar{\tau}_n, \bar{\beta}_n)$ is defined analogously. (The mean values can be chosen measurably by Lemma 3 in Jennrich (1969).) Since $\underline{S}_n - E\underline{S}_n \to 0$ i.p. by Lemma 11.3 we have $\xi_n^j - E\underline{S}_n \to 0$ and $\chi_n^j - E\underline{S}_n \to 0$ i.p. Since $\nabla_{cc}\vartheta_n$ and $\nabla_{c\beta}\vartheta_n$ restricted to $K_1^* \times T' \times B'$ are equicontinuous for any compact set K_1^* it follows from Lemma 3.2(b) that $\nabla_{cc}\vartheta_n(., \bar{\tau}_n, \bar{\beta}_n)$ and

$\nabla_{c\beta}\vartheta_n(.,\bar{\tau}_n,\bar{\beta}_n)$ restricted to K_1^* are equicontinuous. Upon choosing K_1^* such that $\operatorname{int}(K_1^*) \supseteq K_1$ where K_1 is the compact set defined in the proof of Lemma 11.3, it follows that

$$\nabla_{cc}\vartheta_n(\{\xi_n^i\},\bar{\tau}_n,\bar{\beta}_n) = \nabla_{cc}\vartheta_n(E\underline{S}_n,\bar{\tau}_n,\bar{\beta}_n) + o_p(1)$$

and

$$\nabla_{c\beta}\vartheta_n(\{\chi_n^i\},\bar{\tau}_n,\bar{\beta}_n) = \nabla_{c\beta}\vartheta_n(E\underline{S}_n,\bar{\tau}_n,\bar{\beta}_n) + o_p(1).$$

By Lemma 11.3 we further have $\nabla_\beta \underline{S}_n = E\nabla_\beta \underline{S}_n + o_p(1)$. Since $E\nabla_\beta \underline{S}_n$ is bounded by Assumption 11.6(c) and since $\nabla_{cc}\vartheta_n(E\underline{S}_n,\bar{\tau}_n,\bar{\beta}_n)$ is bounded by Assumption 11.6(a) and the fact that $E\underline{S}_n \in K_1$ we have

$$\begin{aligned} &\nabla_{cc}\vartheta_n(\xi_n,\bar{\tau}_n,\bar{\beta}_n)\nabla_\beta \underline{S}_n + \nabla_{c\beta}\vartheta_n(\chi_n,\bar{\tau}_n,\bar{\beta}_n) \\ = \; &\nabla_{cc}\vartheta_n(E\underline{S}_n,\bar{\tau}_n,\bar{\beta}_n)E\nabla_\beta \underline{S}_n + \nabla_{c\beta}\vartheta_n(E\underline{S}_n,\bar{\tau}_n,\bar{\beta}_n) + o_p(1). \end{aligned}$$

The lemma now follows since $\underline{S}_n - E\underline{S}_n = O_p(n^{-1/2})$. ■

Proof of Theorem 11.5. We first prove part (a) of the theorem. To prove the result we verify Assumption 8.1 maintained by Lemma 8.1 and Corollary 8.2. Assumptions 8.1(a)-(d) follow immediately from Assumptions 11.1(a)-(d) and Assumption 11.6(a) upon setting $M_n = n^{1/2}I$ and $N_n = n^{1/2}I$. From Lemma 11.3 we have that

$$\sup_{T'\times B'} |\nabla_{\beta\beta}Q_n - C_{1n}| \to 0 \text{ i.p. as } n \to \infty$$

and

$$\sup_{T'\times B'} |\nabla_{\beta\tau}Q_n - C_{2n}| \to 0 \text{ i.p. as } n \to \infty$$

and that C_{1n} and C_{2n} are equicontinuous on $T' \times B'$. For any sequence $(\tilde{\tau}_n,\tilde{\beta}_n)$ as in Assumption 8.1(e) we can in light of Assumption 11.1(d) assume w.l.o.g. that $(\tilde{\tau}_n,\tilde{\beta}_n) \in T' \times B'$ possibly after redefining $(\tilde{\tau}_n,\tilde{\beta}_n)$ on ω-sets the probability of which tends to zero as $n \to \infty$. Since

$$(\tilde{\tau}_n,\tilde{\beta}_n) - (\bar{\tau}_n,\bar{\beta}_n) \to 0 \text{ i.p. as } n \to \infty,$$

it follows then from Lemma 3.2 that

$$\nabla_{\beta\beta}Q_n(\mathbf{z}_1,\ldots,\mathbf{z}_n,\tilde{\tau}_n,\tilde{\beta}_n) - C_{1n}(\bar{\tau}_n,\bar{\beta}_n) \to 0 \text{ i.p. as } n \to \infty$$

and

$$\nabla_{\beta\tau}Q_n(\mathbf{z}_1,\ldots,\mathbf{z}_n,\tilde{\tau}_n,\tilde{\beta}_n) - C_{2n}(\bar{\tau}_n,\bar{\beta}_n) \to 0 \text{ i.p. as } n \to \infty$$

where

$$C_{1n}(\bar{\tau}_n,\bar{\beta}_n) = G_{1n}(0,E\nabla_\beta \underline{S}_n,E\nabla_{\beta\beta}\underline{S}_n,\bar{\tau}_n,\bar{\beta}_n)$$

and

$$C_{2n}(\bar{\tau}_n, \bar{\beta}_n) = G_{2n}(0, E\nabla_\beta \underline{S}_n, 0, E\nabla_{\beta\tau}\underline{S}_n, \bar{\tau}_n, \bar{\beta}_n)$$

in view of Assumptions 11.7(a),(b). Observe that Assumption 11.6(b) implies also

$$\nabla_{c\beta}\vartheta_n(0, \tau, \beta) = 0$$

(and hence $\nabla_{\beta c}\vartheta_n(0, \tau, \beta) = 0$),

$$\nabla_{c\tau}\vartheta_n(0, \tau, \beta) = 0,$$

$$\nabla_{\beta\tau}\vartheta_n(0, \tau, \beta) = 0,$$

and

$$\nabla_{\beta\beta}\vartheta_n(0, \tau, \beta) = 0$$

for $(\tau, \beta) \in \operatorname{int}(T' \times B')$. Consequently,

$$C_{1n}(\bar{\tau}_n, \bar{\beta}_n) = E\nabla_{\beta'}\underline{S}_n \left[\nabla_{cc}\vartheta_n(0, \bar{\tau}_n, \bar{\beta}_n)\right] E\nabla_\beta \underline{S}_n = C_n$$

and

$$C_{2n}(\bar{\tau}_n, \bar{\beta}_n) = 0.$$

Because of Assumptions 11.6(a),(c) and since $(\bar{\tau}_n, \bar{\beta}_n) \in T' \times B'$ we have

$$\nabla_{cc}\vartheta_n(0, \bar{\tau}_n, \bar{\beta}_n) = O(1),$$

$$E\nabla_{\beta'}\underline{S}_n = O(1)$$

and hence $|C_n| = O(1)$. Furthermore $|C_n^+| = \left|C_n^{-1}\right| = O(1)$ because of Assumptions 11.7(c),(d). This verifies Assumptions 8.1(e),(f). It remains to verify Assumption 8.1(g) under either Assumption 11.8 or 11.9. Because of Assumptions 11.6 and 11.7(a) we have

$$\nabla_\beta \bar{Q}_n(\bar{\tau}_n, \bar{\beta}_n) = \nabla_c\vartheta_n(0, \bar{\tau}_n, \bar{\beta}_n) E\nabla_\beta \underline{S}_n + \nabla_\beta\vartheta_n(0, \bar{\tau}_n, \bar{\beta}_n) = 0.$$

Here we have made use of the fact that $\nabla_\beta ES_n$ exists and satisfies $\nabla_\beta ES_n = E\nabla_\beta S_n$ on $\operatorname{int}(T' \times B')$ as a consequence of Assumption 11.6(c). It now follows from Lemma 11.4 that

$$\begin{aligned} & n^{1/2}\nabla_{\beta'} Q_n(\mathbf{z}_1, \ldots, \mathbf{z}_n, \bar{\tau}_n, \bar{\beta}_n) \\ = \;& E\nabla_{\beta'}\underline{S}_n \nabla_{cc}\vartheta_n(0, \bar{\tau}_n, \bar{\beta}_n) n^{1/2}\underline{S}_n + o_p(1) \end{aligned} \tag{H.2}$$

observing that $E\underline{S}_n = 0$,

$$\nabla_{c\beta}\vartheta_n(0, \bar{\tau}_n, \bar{\beta}_n) = 0,$$

and

$$\nabla_c\vartheta_n(0, \bar{\tau}_n, \bar{\beta}_n) = 0.$$

The condition that $\underline{S}_n = O_p(n^{-1/2})$ maintained in Lemma 11.4 follows from the CLT and from the boundedness of $n^{-1}V_n$ established below. Putting $V_n = n^2 E[\underline{S}_n \underline{S}_n']$ observe that

$$\liminf_{n\to\infty} \lambda_{\min}(n^{-1}V_n) > 0$$

by Assumption 11.7(e). Given Assumption 11.8 holds, put $\mathbf{v}_t = q_t(\mathbf{z}_t, \bar{\tau}, \bar{\beta})$. Then applying Theorem 6.5 to $q_t(z, \bar{\tau}, \bar{\beta})$ and $(\mathbf{z}_t)$ and making use of Assumptions 11.1(e),(f) and Assumption 11.6(c), it follows that $(\mathbf{v}_t)$ is L_0-approximable by $(\mathbf{e}_t)$. From Assumption 11.8 and Theorem 10.1(a) it follows that

$$-V_n^{-1/2}(n\underline{S}_n) \xrightarrow{D} \zeta$$

where ζ is distributed as $N(0, I)$. The boundedness of $n^{-1}V_n$ follows from the discussion after Theorem 10.2. Next put

$$\mathbf{v}_{t,n} = q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n) - Eq_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n).$$

Then the same conclusion is true under Assumption 11.9 in view of Theorem 10.2(a) observing that

$$E\left[\left(\sum_{t=1}^{n} \mathbf{v}_{t,n}\right)\left(\sum_{t=1}^{n} \mathbf{v}_{t,n}'\right)\right] = V_n = n^2 E\left[\underline{S}_n \underline{S}_n'\right]$$

and that $E\underline{S}_n = 0$ for all n. The boundedness of $n^{-1}V_n$ follows again from the discussion after Theorem 10.2. Defining $\zeta_n = -V_n^{-1/2}(n\underline{S}_n)$ and given

$$D_n = E\nabla_{\beta'}\underline{S}_n \left[\nabla_{cc}\vartheta_n(0, \bar{\tau}_n, \bar{\beta}_n)\right] (nE\underline{S}_n\underline{S}_n')^{1/2},$$

Assumption 8.1(g) follows from (H.2) provided we can establish boundedness of D_n. As discussed above $\nabla_{cc}\vartheta_n(0, \bar{\tau}_n, \bar{\beta}_n) = O(1)$ and $E\nabla_{\beta'}\underline{S}_n = O(1)$. Since $(nE\underline{S}_n\underline{S}_n')^{1/2}$ is the square root of $n^{-1}V_n$, its boundedness follows from the boundedness of $n^{-1}V_n$. Hence $|D_n| = O(1)$. By Assumption 11.7(d) the matrix $E\nabla_{\beta'}\underline{S}_n$ and hence D_n has full row-rank for large n in light of Assumptions 11.7(c),(e). Furthermore, the sequence $\left|(D_nD_n')^{-1}\right|$ is bounded since clearly

$$\begin{aligned} & \lambda_{\min}(D_nD_n') \\ \geq\ & \lambda_{\min}(nE\underline{S}_n\underline{S}_n')\lambda_{\min}^2\left(\nabla_{cc}\vartheta_n(0, \bar{\tau}_n, \bar{\beta}_n)\right)\lambda_{\min}(E\nabla_{\beta'}\underline{S}_nE\nabla_{\beta}\underline{S}_n) \end{aligned}$$

is bounded away from zero by Assumptions 11.7(c)-(e). Part (a) of the theorem now follows from Lemma 8.1 and Corollary 8.2(b).

We next prove part (b) of the theorem. The proof is identical to that of part (a) up to the point of verification of Assumption 8.1(g) under either Assumption 11.8 or 11.9. Note that equation (H.2) again follows from Lemma 11.4. The condition that $\underline{S}_n = O_p(n^{-1/2})$ maintained in Lemma

11.4 follows from the CLT established below. Putting $V_n = n^2 E[\underline{S}_n \underline{S}_n']$ observe that $n^{-1}V_n \to \Lambda$ by assumption. Given Assumption 11.8 holds, put $\mathbf{v}_t = q_t(\mathbf{z}_t, \bar{\tau}, \bar{\beta})$. Then exactly as in the proof of part (a) it follows that $(\mathbf{v}_t)$ is L_0-approximable by $(\mathbf{e}_t)$. From Assumption 11.8 and Theorem 10.1(b) it follows that

$$-n^{1/2}\underline{S}_n \xrightarrow{D} N(0, \Lambda).$$

Next put

$$\mathbf{v}_{t,n} = q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n) - Eq_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n).$$

Then the same conclusion is true under Assumption 11.9 in view of Theorem 10.2(b) observing that

$$E\left[\left(\sum_{t=1}^{n} \mathbf{v}_{t,n}\right)\left(\sum_{t=1}^{n} \mathbf{v}_{t,n}'\right)\right] = V_n = n^2 E\left[\underline{S}_n \underline{S}_n'\right]$$

and that $E\underline{S}_n = 0$ for all n. Defining $\zeta_n = -n^{1/2}\underline{S}_n$ and given

$$D_n = E\nabla_{\beta'}\underline{S}_n \left[\nabla_{cc}\vartheta_n(0, \bar{\tau}_n, \bar{\beta}_n)\right],$$

Assumption 8.1(g) follows from (H.2) provided we can establish boundedness of D_n. As discussed above $\nabla_{cc}\vartheta_n(0, \bar{\tau}_n, \bar{\beta}_n) = O(1)$, $E\nabla_{\beta'}\underline{S}_n = O(1)$, and hence $|D_n| = O(1)$. By Assumption 11.7(d) the matrix $E\nabla_{\beta'}\underline{S}_n$ and hence D_n has full row-rank for large n in light of Assumption 11.7(c). Furthermore, the sequence $\left|(D_n D_n')^{-1}\right|$ is bounded since clearly

$$\lambda_{\min}(D_n D_n') \geq \lambda_{\min}^2\left(\nabla_{cc}\vartheta_n(0, \bar{\tau}_n, \bar{\beta}_n)\right)\lambda_{\min}(E\nabla_{\beta'}\underline{S}_n E\nabla_{\beta}\underline{S}_n)$$

is bounded away from zero by Assumptions 11.7(c),(d). Part (b) of the theorem now follows from Lemma 8.1. ■

Theorem H2. *Suppose the assumptions of Theorem 11.5(b) hold.*

(i) If Λ is nonsingular then

$$(C_n^{-1} D_n \Lambda D_n' C_n^{-1\prime})^{-1/2} n^{1/2}(\hat{\beta}_n - \bar{\beta}_n) \xrightarrow{D} N(0, I).$$

(ii) Suppose D_n is a square matrix and U_1 is an orthogonal matrix of eigenvectors of Λ, where the first r columns correspond to the r nonzero eigenvalues of Λ. Then

$$(\Lambda^+)^{1/2} D_n^{-1} C_n n^{1/2}(\hat{\beta}_n - \bar{\beta}_n) \xrightarrow{D} N(0, \Xi_1),$$

where

$$\Xi_1 = U_1 \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} U_1'.$$

(iii) Suppose $D_n \to D$ and U_2 is an orthogonal matrix of eigenvectors of $D\Lambda D'$, where the first r columns correspond to the r nonzero eigenvalues of $D\Lambda D'$. Then

$$C_n n^{1/2}(\hat{\beta}_n - \bar{\beta}_n) \xrightarrow{D} N(0, D\Lambda D'),$$

and hence

$$\left((D\Lambda D')^+\right)^{1/2} C_n n^{1/2}(\hat{\beta}_n - \bar{\beta}_n) \xrightarrow{D} N(0, \Xi_2)$$

where

$$\Xi_2 = U_2 \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} U_2'.$$

(iv) Suppose $C_n \to C$, $D_n \to D$ and U_3 is an orthogonal matrix of eigenvectors of $C^{-1}D\Lambda D'C^{-1\prime}$, where the first r columns correspond to the r nonzero eigenvalues of $C^{-1}D\Lambda D'C^{-1\prime}$. Then

$$n^{1/2}(\hat{\beta}_n - \bar{\beta}_n) \xrightarrow{D} N(0, C^{-1}D\Lambda D'C^{-1\prime}),$$

$$Cn^{1/2}(\hat{\beta}_n - \bar{\beta}_n) \xrightarrow{D} N(0, D\Lambda D'),$$

$$\left((D\Lambda D')^+\right)^{1/2} Cn^{1/2}(\hat{\beta}_n - \bar{\beta}_n) \xrightarrow{D} N(0, \Xi_2),$$

and

$$((C^{-1}D\Lambda D'C^{-1\prime})^+)^{1/2} n^{1/2}(\hat{\beta}_n - \bar{\beta}_n) \xrightarrow{D} N(0, \Xi_3),$$

where

$$\Xi_3 = U_3 \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} U_3'.$$

(Of course, if Λ is nonsingular, then Ξ_1 reduces to the identity matrix, and if $D\Lambda D'$ is nonsingular, then Ξ_2 and Ξ_3 reduce to the identity matrix. Observe also that D has full row-rank since $\left|(D_n D_n')^{-1}\right| = O(1)$ holds. Also $\operatorname{rank}(C^{-1}D\Lambda D'C^{-1\prime}) = \operatorname{rank}(D\Lambda D')$ holds.)

Proof. By Theorem 11.5(b) we have

$$n^{1/2}(\hat{\beta}_n - \bar{\beta}_n) = C_n^{-1} D_n \zeta_n + o_p(1)$$

with $\zeta_n \xrightarrow{D} N(0, \Lambda)$. Part (i) now follows from Corollary 8.2(b) observing that D_n has full row-rank and that $\left|(D_n D_n')^{-1}\right| = O(1)$, as was established in the proof of Theorem 11.5(b). Under the assumptions of part (ii) D_n is a square matrix and hence

$$D_n^{-1} C_n n^{1/2}(\hat{\beta}_n - \bar{\beta}_n) \xrightarrow{D} N(0, \Lambda),$$

since $\left|D_n^{-1}\right| = O(1)$ as a consequence of $\left|(D_n D_n')^{-1}\right| = O(1)$. The result in part (ii) now follows using standard properties of the Moore-Penrose inverse, analogously as in the proof of Theorem H1. By Theorem 11.5(b) we have that

$$C_n n^{1/2}(\hat{\beta}_n - \bar{\beta}_n) = D_n \zeta_n + o_p(1)$$

with $\zeta_n \xrightarrow{D} N(0, \Lambda)$. Parts (iii) and (iv) are now proved by analogous argumentation as in the proof of Theorem H1. ■

Remark. We note that in case Λ is singular and D_n is not a square matrix one can give examples where $((C_n^{-1} D_n \Lambda D_n' C_n^{-1\prime})^{+})^{1/2} n^{1/2}(\hat{\beta}_n - \bar{\beta}_n)$ does not converge in distribution. Hence in this case no analogue to part (i) of Theorem H2 exists.

Appendix I

PROOFS FOR CHAPTER 12

The following lemma follows immediately from Lemma 2.1 in McLeish (1975a) and Hölder's inequality observing that

$$|\text{cov}(X,Y)| = |E\{Y[E(X \mid \mathfrak{F}) - E(X)]\}|,$$

cf. also Hall and Heyde (1980), Appendix III.

Lemma I1. *Let X be a $\mathfrak{G}$-measurable and let Y be a $\mathfrak{F}$-measurable random variable such that for $1 \le p \le s \le \infty$ and $p^{-1}+q^{-1} = 1$ we have $\|X\|_s < \infty$, $\|Y\|_q < \infty$. Then*

$$\begin{aligned} |\text{cov}(X,Y)| &\le 2(2^{1/p}+1)\bar{\alpha}(\mathfrak{F},\mathfrak{G})^{1/p-1/s}\|X\|_s\|Y\|_q, \\ |\text{cov}(X,Y)| &\le 2\bar{\phi}(\mathfrak{F},\mathfrak{G})^{1-1/s}\|X\|_s\|Y\|_q, \end{aligned}$$

where

$$\bar{\alpha}(\mathfrak{F},\mathfrak{G}) = \sup\{|P(F\cap G) - P(F)P(G)| : F\in\mathfrak{F}, G\in\mathfrak{G}\}$$

and

$$\bar{\phi}(\mathfrak{F},\mathfrak{G}) = \sup\{|P(G|F) - P(G)| : F\in\mathfrak{F}, G\in\mathfrak{G}, P(F)>0\}.$$

Lemma I2. *Let $(\mathbf{w}_{t,n} : t\in\mathbf{N}, n\in\mathbf{N})$ be a real valued process with approximation errors*

$$\nu_m(\mathbf{w}) = \sup_n \sup_t \|\mathbf{w}_{t,n} - E(\mathbf{w}_{t,n}|\mathbf{e}_{t+m},\ldots,\mathbf{e}_{t-m})\|_2$$

with respect to the basis process $(\mathbf{e}_t)_{t\in\mathbf{Z}}$. Let $\alpha(k)$ and $\phi(k)$ denote the α-mixing and ϕ-mixing coefficients of $(\mathbf{e}_t)$ and let

$$c_w \equiv \sup_n \sup_t \|\mathbf{w}_{t,n}\|_r < \infty$$

for some $r\ge 2$, then for all $t\in\mathbf{N}$, $n\in\mathbf{N}$, $m\ge 0$, $\tau\ge 0$:

$$\begin{aligned} |\text{cov}(\mathbf{w}_{t,n},\mathbf{w}_{t+\tau,n})| &\le 4c_w\left[\nu_m(\mathbf{w}) + 2c_w(2^{1-1/r}+1)\alpha(\tau-2m)^{1-2/r}\right], \\ |\text{cov}(\mathbf{w}_{t,n},\mathbf{w}_{t+\tau,n})| &\le 4c_w\left[\nu_m(\mathbf{w}) + 2c_w\phi(\tau-2m)^{1-1/r}\right]. \end{aligned}$$

Proof. First consider the case where $E\mathbf{w}_{t,n} = 0$. Let

$$\mathbf{h}_{t,n}^m = E(\mathbf{w}_{t,n}|\mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m}),$$

then

$$\begin{aligned} & |\text{cov}(\mathbf{w}_{t,n}, \mathbf{w}_{t+\tau,n})| \\ \leq \; & \left|E\left[(\mathbf{w}_{t,n} - \mathbf{h}_{t,n}^m)\mathbf{w}_{t+\tau,n}\right]\right| + \left|E\left[\mathbf{h}_{t,n}^m(\mathbf{w}_{t+\tau,n} - \mathbf{h}_{t+\tau,n}^m)\right]\right| \\ & + \left|E[\mathbf{h}_{t,n}^m \mathbf{h}_{t+\tau,n}^m]\right| \\ \leq \; & \nu_m(\mathbf{w}) \left\|\mathbf{w}_{t+\tau,n}\right\|_2 + \nu_m(\mathbf{w}) \left\|\mathbf{h}_{t,n}^m\right\|_2 + \left|E[\mathbf{h}_{t,n}^m \mathbf{h}_{t+\tau,n}^m]\right| \end{aligned}$$

using the triangle and Cauchy-Schwarz inequality. By Jensen's and Lyapunov's inequality we have

$$\left\|\mathbf{h}_{t,n}^m\right\|_2 \leq \left\|\mathbf{w}_{t,n}\right\|_2 \leq \left\|\mathbf{w}_{t,n}\right\|_r \leq c_w.$$

Hence

$$|\text{cov}(\mathbf{w}_{t,n}, \mathbf{w}_{t+\tau,n})| \leq 2c_w \nu_m(\mathbf{w}) + \left|E\left[\mathbf{h}_{t,n}^m \mathbf{h}_{t+\tau,n}^m\right]\right|.$$

For $i \leq j$ let $\mathfrak{A}_i^j$ be the σ-field generated by $\mathbf{e}_i, \ldots, \mathbf{e}_j$. Since $\mathbf{h}_{t,n}^m$ is $\mathfrak{A}_{-\infty}^{t+m}$-measurable, $\mathbf{h}_{t+\tau,n}^m$ is $\mathfrak{A}_{t+\tau-m}^{\infty}$-measurable and $E\mathbf{h}_{t,n}^m = 0$, it follows from Lemma I1 upon choosing $X = \mathbf{h}_{t+\tau,n}^m$, $Y = \mathbf{h}_{t,n}^m$ and $q = s = r$ that

$$|\text{cov}(\mathbf{w}_{t,n}, \mathbf{w}_{t+\tau,n})| \leq 2c_w \nu_m(\mathbf{w}) + 2c_w^2(2^{1-1/r} + 1)\alpha(\tau - 2m)^{1-2/r}, \quad \text{(I.1)}$$

observing that $\bar{\alpha}(\mathfrak{A}_{-\infty}^{t+m}, \mathfrak{A}_{t+\tau-m}^{\infty}) \leq \alpha(\tau - 2m)$, cf. Definition 6.1, and since $\left\|\mathbf{h}_{t,n}^m\right\|_r \leq c_w$ by Jensen's inequality. The general case with $E\mathbf{w}_{t,n}$ possibly nonzero follows if we apply (I.1) to the centered process $\mathbf{w}_{t,n} - E\mathbf{w}_{t,n}$, observing that c_w has then to be replaced by $2c_w$ since

$$\left\|\mathbf{w}_{t,n} - E\mathbf{w}_{t,n}\right\|_r \leq 2\left\|\mathbf{w}_{t,n}\right\|_r.$$

The inequality with ϕ-mixing coefficients is proved analogously. ∎

Lemma I3. *Let $(\mathbf{v}_{t,n} : t \in \mathbf{N}, n \in \mathbf{N})$ be a real valued process with approximation errors $\nu_m(\mathbf{v})$ with respect to the basis process $(\mathbf{e}_t)_{t\in\mathbf{Z}}$. Let*

$$c \equiv \sup_n \sup_t \left\|\mathbf{v}_{t,n}^2\right\|_r < \infty$$

for some $r \geq 2$. Define

$$\xi_{t,n}^{(j)} = \mathbf{v}_{t,n}\mathbf{v}_{t+j,n}$$

for $j \geq 0$ and let $\nu_m(\xi^{(j)})$ denote the corresponding approximation errors w.r.t. $(\mathbf{e}_t)$. Then

$$\nu_m(\xi^{(j)}) \leq 2c$$

for all m, and furthermore

$$\nu_m(\xi^{(j)}) \leq A\left[\nu_{m-j}(\mathbf{v})\right]^{(r-2)/(2r-2)}$$

for $m \geq j$ *with* $A = (c^{(3r-2)/2}2^{5r-4})^{1/(2r-2)}$. *The same result holds if we define*

$$\xi_{t,n}^{(j)} = \mathbf{v}_{t,n}\mathbf{v}_{t+j,n} - E\mathbf{v}_{t,n}\mathbf{v}_{t+j,n}.$$

Proof. That $\nu_m(\xi^{(j)}) \leq 2c$ for all m follows easily from the Minkowski, Cauchy-Schwarz and Jensen inequalities. For $r = 2$ the second bound now holds trivially, hence assume $r > 2$. Let $m = k + j$ with $k \geq 0$, then

$$\begin{aligned}\nu_m(\xi^{(j)}) &= \nu_{k+j}(\xi^{(j)}) \\ &= \sup_n \sup_t \left\|\mathbf{v}_{t,n}\mathbf{v}_{t+j,n} - E(\mathbf{v}_{t,n}\mathbf{v}_{t+j,n}|\mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m})\right\|_2 \\ &\leq \sup_n \sup_t \left\|\mathbf{v}_{t,n}\mathbf{v}_{t+j,n} - \mathbf{h}_{t,n}^k\mathbf{h}_{t+j,n}^k\right\|_2\end{aligned}$$

where $\mathbf{h}_{t,n}^i = E(\mathbf{v}_{t,n}|\mathbf{e}_{t+i}, \ldots, \mathbf{e}_{t-i})$. The latter inequality follows since $\mathbf{h}_{t,n}^k\mathbf{h}_{t+j,n}^k$ is clearly measurable with respect to the σ-field generated by $\mathbf{e}_{t+m}, \ldots, \mathbf{e}_{t-m}$ and hence does not give a better L_2-approximation to $\mathbf{v}_{t,n}\mathbf{v}_{t+j,n}$ than the conditional mean. Consequently

$$\begin{aligned}\nu_m(\xi^{(j)}) \leq\ & \sup_n \sup_t \left\|(\mathbf{v}_{t,n} - \mathbf{h}_{t,n}^k)\mathbf{v}_{t+j,n}\right\|_2 \\ &+ \sup_n \sup_t \left\|\mathbf{h}_{t,n}^k(\mathbf{v}_{t+j,n} - \mathbf{h}_{t+j,n}^k)\right\|_2.\end{aligned}$$

Observe that for any pair of random variables X and Y and $s^\bullet > 2$, $s \geq 1$ the inequality

$$\|XY\|_2 \leq 2\left(\|Y\|_s^{(s^\bullet-2)}\|X\|_{s/(s-1)}^{(s^\bullet-2)}\|XY\|_{s^\bullet}^{s^\bullet}\right)^{1/(2s^\bullet-2)} \tag{I.2}$$

holds given the norms on the r.h.s. are finite. (This inequality is implicit in the proof of Theorem 4.1 in Gallant and White (1988).) Observe further that

$$\|\mathbf{v}_{t+j,n}\|_2 \leq c^{1/2},$$

$$\left\|\mathbf{h}_{t,n}^k\right\|_2 \leq \|\mathbf{v}_{t,n}\|_2 \leq c^{1/2},$$

and

$$\begin{aligned}\left\|(\mathbf{v}_{t,n} - \mathbf{h}_{t,n}^k)\mathbf{v}_{t+j,n}\right\|_r &\leq \|\mathbf{v}_{t,n}\mathbf{v}_{t+j,n}\|_r + \left\|\mathbf{h}_{t,n}^k\mathbf{v}_{t+j,n}\right\|_r \\ &\leq \|\mathbf{v}_{t,n}\|_{2r}\|\mathbf{v}_{t+j,n}\|_{2r} + \left\|\mathbf{h}_{t,n}^k\right\|_{2r}\|\mathbf{v}_{t+j,n}\|_{2r} \\ &\leq 2c\end{aligned}$$

using Hölder's and Jensen's inequalities. Similarly we obtain

$$\left\| \mathbf{h}_{t,n}^{k}(\mathbf{v}_{t+j,n} - \mathbf{h}_{t+j,n}^{k}) \right\|_r \leq 2c.$$

It then follows from inequality (I.2) upon choosing $s^{\bullet} = r$, $s = 2$ that

$$\nu_m(\xi^{(j)}) \leq A\,[\nu_k(\mathbf{v})]^{(r-2)/(2r-2)}$$

where A is given in the lemma. This establishes the result for $\xi_{t,n}^{(j)} = \mathbf{v}_{t,n}\mathbf{v}_{t+j,n}$. The result for the centered process follows immediately since $\mathbf{v}_{t,n}\mathbf{v}_{t+j,n}$ and the centered process $\mathbf{v}_{t,n}\mathbf{v}_{t+j,n} - E\mathbf{v}_{t,n}\mathbf{v}_{t+j,n}$ have the same approximation errors. ■

Lemma I4. *Let $(\mathbf{v}_{t,n} : t \in \mathbf{N}, n \in \mathbf{N})$ be a real valued process that is near epoch dependent on the basis process $(\mathbf{e}_t)_{t\in\mathbf{Z}}$ with approximation errors $\nu_m(\mathbf{v})$ of size $-2(r-1)/(r-2)$, and $(\mathbf{e}_t)$ is α-mixing with mixing coefficients of size $-r/(r-2)$ or ϕ-mixing with mixing coefficients of size $-r/(r-1)$, for some $r > 2$. Let*

$$c \equiv \sup_n \sup_t \left\| \mathbf{v}_{t,n}^2 \right\|_r < \infty.$$

Define

$$\xi_{t,n}^{(j)} = \mathbf{v}_{t,n}\mathbf{v}_{t+j,n} - E\mathbf{v}_{t,n}\mathbf{v}_{t+j,n}$$

for $j \geq 0$. Then for $0 \leq j \leq n$ we have

$$E\left| n^{-1}\sum_{t=1}^{n-j} \xi_{t,n}^{(j)} \right|^2 \leq C(j+1)/n,$$

where $C < \infty$ is a constant that does not depend on j or n (but only on c, r, $\nu_m(\mathbf{v})$ and $\alpha(k)$ or $\phi(k)$).

Proof. Clearly

$$E\left| n^{-1}\sum_{t=1}^{n-j} \xi_{t,n}^{(j)} \right|^2 \leq 2n^{-2} \sum_{\tau=0}^{n-j-1} \sum_{t=1}^{n-j-\tau} E\left[\xi_{t,n}^{(j)}\xi_{t+\tau,n}^{(j)}\right]$$

for each $j \geq 0$. Observing that

$$\sup_n \sup_t \left\| \xi_{t,n}^{(j)} \right\|_r \leq 2c,$$

we obtain from Lemma I2 applied to the process $(\xi_{t,n}^{(j)})$ that

$$E\left| n^{-1}\sum_{t=1}^{n-j} \xi_{t,n}^{(j)} \right|^2$$

$$
\begin{aligned}
&\leq 16cn^{-2}\sum_{\tau=0}^{n-j-1}(n-j-\tau)\Big[\nu_{m(\tau)}(\xi^{(j)}) \\
&\quad + 4c(2^{1-1/r}+1)\alpha(\tau-2m(\tau))^{1-2/r}\Big] \\
&\leq 16c(n-j)n^{-2}\sum_{\tau=0}^{n-j-1}\nu_{m(\tau)}(\xi^{(j)}) \\
&\quad + 64c^2(2^{1-1/r}+1)(n-j)n^{-2}\sum_{\tau=0}^{n-j-1}\alpha(\tau-2m(\tau))^{1-2/r}
\end{aligned}
$$

for any sequence $m(\tau) \geq 0$. Choosing in particular $m(\tau)$ as the integer part of $\tau/4$ it follows further from Lemma I3 that

$$
\begin{aligned}
&E\left|n^{-1}\sum_{t=1}^{n-j}\xi_{t,n}^{(j)}\right|^2 \\
&\leq 16c(n-j)n^{-2}\left\{\sum_{\tau=0}^{4j}2c+\sum_{\tau=4j}^{\infty}A\left[\nu_{m(\tau)-j}(\mathbf{v})\right]^{(r-2)/(2r-2)}\right\} \\
&\quad + 64c^2(2^{1-1/r}+1)(n-j)n^{-2}\sum_{\tau=0}^{\infty}\alpha\left(\tau-2m(\tau)\right)^{1-2/r} \\
&\leq 128c^2(j+1)/n + 64cAn^{-1}\sum_{m=0}^{\infty}\left[\nu_m(\mathbf{v})\right]^{(r-2)/(2r-2)} \\
&\quad + 128c^2(2^{1-1/r}+1)n^{-1}\sum_{k=0}^{\infty}\alpha(k)^{1-2/r} \\
&\leq C(j+1)/n
\end{aligned}
$$

where the constant C does not depend on j or n. Note that the infinite sums are finite by the size assumptions on $\nu_k(\mathbf{v})$ and $\alpha(k)$. The proof for the case where $(\mathbf{e}_t)$ is ϕ-mixing is analogous. ■

Proof of Lemma 12.1.[1] Clearly,

$$
\left|\hat{\Psi}_n - \Psi_n\right| \leq \left|\hat{\Psi}_n - \bar{\Psi}_n\right| + \left|\bar{\Psi}_n - E\bar{\Psi}_n\right| + \left|E\bar{\Psi}_n - \Psi_n\right|.
$$

Part (a) of the lemma now follows if we can show that

$$
\left|\hat{\Psi}_n - \bar{\Psi}_n\right| = O_p(\gamma_{1n}),
$$

[1]Because of Assumption 12.1 the space $\mathrm{int}(T \times B)$ is separable and $\mathbf{v}_t(\tau, \beta)$ is continuous a.s. on $\mathrm{int}(T \times B)$. Consequently, $\mathbf{v}_t(\hat{\tau}_n, \hat{\beta}_n)$ and Δ_t coincide with measurable functions on a set of probability one. Hence the various boundedness in probability statements involving $\hat{\Psi}_n$ and Δ_t are well-defined.

$$\left|\bar{\Psi}_n - E\bar{\Psi}_n\right| = O_p(\gamma_{2n}),$$

and

$$\left|E\bar{\Psi}_n - \Psi_n\right| = O_p(\gamma_{3n}).$$

Let θ_i, $\hat{\theta}_{n,i}$ and $\bar{\theta}_{n,i}$ denote the i-th component of (τ, β), $(\hat{\tau}_n, \hat{\beta}_n)$ and $(\bar{\tau}_n, \bar{\beta}_n)$, respectively, and let $p = p_\tau + p_\beta$. Analogously to the proof of Lemma 8.1 we can find sets $\Omega_n \in \mathfrak{A}$ with $P(\Omega_n) \to 1$ as $n \to \infty$ such that for all $\omega \in \Omega_n$ we have by the mean value theorem

$$\begin{aligned}
&\left|\hat{\Psi}_n - \bar{\Psi}_n\right| \\
=\ &\left| w(0,n)n^{-1}\sum_{t=1}^{n}\sum_{i=1}^{p}\left[\nabla_{\theta_i}\mathbf{v}_{t,n}\mathbf{v}_{t,n}' + \mathbf{v}_{t,n}\nabla_{\theta_i}\mathbf{v}_{t,n}'\right]_{|_{\theta=\tilde{\theta}}}(\hat{\theta}_{n,i} - \bar{\theta}_{n,i})\right. \\
&+ \sum_{j=1}^{n-1} w(j,n)\left[n^{-1}\sum_{t=1}^{n-j}\sum_{i=1}^{p}\left[\nabla_{\theta_i}\mathbf{v}_{t,n}\mathbf{v}_{t+j,n}' + \mathbf{v}_{t,n}\nabla_{\theta_i}\mathbf{v}_{t+j,n}'\right.\right. \\
&\left.\left.\left. + \mathbf{v}_{t+j,n}\nabla_{\theta_i}\mathbf{v}_{t,n}' + \nabla_{\theta_i}\mathbf{v}_{t+j,n}\mathbf{v}_{t,n}'\right]_{|_{\theta=\tilde{\theta}}}(\hat{\theta}_{n,i} - \bar{\theta}_{n,i})\right]\right|
\end{aligned}$$

where the mean values $\tilde{\theta}$ may be different in different entries. Note that each of the products involving a derivative w.r.t. θ_i is bounded normwise by $p_v^2\Delta_t^2$ and $p_v^2\Delta_t\Delta_{t+j}$, respectively. (The factor p_v^2 arises because of the possibility that each entry is evaluated at a different mean value.) Hence

$$\begin{aligned}
&\left|\hat{\Psi}_n - \bar{\Psi}_n\right| \\
\leq\ &\left\{2\left|w(0,n)\right|n^{-1}\sum_{t=1}^{n}\Delta_t^2\right. \\
&\left. + 4\sum_{j=1}^{n-1}\left|w(j,n)\right|n^{-1}\sum_{t=1}^{n-j}\Delta_t\Delta_{t+j}\right\}p_v^2\sum_{i=1}^{p}\left|\hat{\theta}_{n,i} - \bar{\theta}_{n,i}\right| \\
\leq\ & p_v^2 p^{1/2}\left|\hat{\theta}_n - \bar{\theta}_n\right|\left\{2\left|w(0,n)\right| + 4\sum_{j=1}^{n-1}\left|w(j,n)\right|\right\}n^{-1}\sum_{t=1}^{n}\Delta_t^2 \\
\leq\ & 4p_v^2 p^{1/2} n^{1/2}\left|\hat{\theta}_n - \bar{\theta}_n\right| n^{-1}\sum_{t=1}^{n}\Delta_t^2\left\{n^{-1/2}\sum_{j=0}^{n-1}\left|w(j,n)\right|\right\} \\
=\ & O_p(1)\gamma_{1n}.
\end{aligned}$$

Next consider

$$\begin{aligned}
&\left|\bar{\Psi}_n - E\bar{\Psi}_n\right| \\
=\ &\left| w(0,n)n^{-1}\sum_{t=1}^{n}\xi_{t,n}^{(0)} + \sum_{j=1}^{n-1}w(j,n)n^{-1}\sum_{t=1}^{n-j}\left[\xi_{t,n}^{(j)} + \xi_{t,n}^{(j)\prime}\right]\right|
\end{aligned}$$

where we use the abbreviation $\xi_{t,n}^{(j)} = \mathbf{v}_{t,n}\mathbf{v}_{t+j,n}' - E\mathbf{v}_{t,n}\mathbf{v}_{t+j,n}'$. By Chebychev's and the triangle inequality it follows for $M > 0$ that

$$\begin{aligned} P\left(\left|\bar{\Psi}_n - E\bar{\Psi}_n\right| > \gamma_{2n}M\right) &\leq \left(\left\|\left|\bar{\Psi}_n - E\bar{\Psi}_n\right|\right\|_2 / (\gamma_{2n}M)\right)^2 \\ &\leq \left(\sum_{j=0}^{n-1} |w(j,n)|\, a(j,n)/(\gamma_{2n}M)\right)^2 = 1/M^2 \end{aligned}$$

given $\gamma_{2n} > 0$. (The case $\gamma_{2n} = 0$ is trivial since $\left\|\left|\bar{\Psi}_n - E\bar{\Psi}_n\right|\right\|_2 \leq \gamma_{2n}$.) Since M is arbitrary, it follows that $\left|\bar{\Psi}_n - E\bar{\Psi}_n\right| = O_p(\gamma_{2n})$. Finally

$$\left|E\bar{\Psi}_n - \Psi_n\right| \leq \sum_{j=0}^{n-1} |w(j,n) - 1|\, \eta(j,n) = \gamma_{3n}. \tag{I.3}$$

This proves part (a) of the lemma. Parts (b) and (c) of the lemma are obvious. Since

$$\gamma_{3n} \leq \sum_{j=0}^{\infty} |w(j,n) - 1|\, \eta(j,n),$$

part (d) follows from the dominated convergence theorem (applied to counting measure). Part (e) follows easily from (I.3). ∎

Proof of Corollary 12.2. Clearly the general assumptions of Lemma 12.1 hold. We proceed by calculating under near epoch dependence upper bounds for γ_{1n}, γ_{2n} and γ_{3n}. The rate for γ_{1n} was already established in Lemma 12.1. Now let $\xi_{t,n}^{(j)} = \mathbf{v}_{t,n}\mathbf{v}_{t+j,n}' - E\mathbf{v}_{t,n}\mathbf{v}_{t+j,n}'$, then

$$a(0,n)^2 = E\left|n^{-1}\sum_{t=1}^{n} \xi_{t,n}^{(0)}\right|^2 \leq E\left|n^{-1}\sum_{t=1}^{n}\left[\xi_{t,n}^{(0)} + \xi_{t,n}^{(0)\prime}\right]\right|^2$$

and

$$a(j,n)^2 = E\left|n^{-1}\sum_{t=1}^{n-j}\left[\xi_{t,n}^{(j)} + \xi_{t,n}^{(j)\prime}\right]\right|^2$$

for $j \geq 1$. Observe that for any symmetric $p_v \times p_v$ matrix $D = (d_{ik})$ we have that

$$\begin{aligned} |D| &= \left(\lambda_{\max}(D'D)\right)^{1/2} \leq \sum_{i,k} |d_{ik}| \\ &\leq 2^{-1}\sum_{i,k}\left[|e_i'De_i| + |e_k'De_k| + |(e_i' + e_k')D(e_i + e_k)|\right] \\ &\leq 4^{-1}(p_v + 2)\sum_{i,k} |e_{ik}'De_{ik}| \end{aligned}$$

where $e_{ik} = e_i + e_k$ and e_i denotes the i-th element of the standard basis in $\mathbf{R}^{p_v}$. Hence the term $a(j,n)^2$ for $j \geq 0$ is bounded by

$$\begin{aligned} & E\left[4^{-1}(p_v+2)\sum_{i,k}\left|e_{ik}'\left(n^{-1}\sum_{t=1}^{n-j}\left[\xi_{t,n}^{(j)}+\xi_{t,n}^{(j)\prime}\right]\right)e_{ik}\right|\right]^2 \\ \leq\; & ((p_v+2)^2/16)\sum_{i,k}E\left|e_{ik}'\left(n^{-1}\sum_{t=1}^{n-j}\left[\xi_{t,n}^{(j)}+\xi_{t,n}^{(j)\prime}\right]\right)e_{ik}\right|^2. \end{aligned}$$

Applying Lemma I4 to the process $e_{ik}'\mathbf{v}_{t,n}$ for each (i,k) and observing that $\nu_m(e_{ik}'\mathbf{v}) \leq 2\nu_m(\mathbf{v})$ we obtain $a(j,n)^2 \leq C(j+1)/n$ where the constant $C < \infty$ does not depend on j and n. Hence

$$\begin{aligned} \gamma_{2n} \;&\leq\; C^{1/2}n^{-1/2}\sum_{j=0}^{n-1}|w(j,n)|\,(j+1)^{1/2} \\ &\leq\; \text{const} * n^{-1/2}\sum_{j=0}^{\ell_n-1}(j+1)^{1/2} \leq \text{const} * (\ell_n^3/n)^{1/2}. \end{aligned}$$

Next recall that

$$\eta(0,n) = \left|n^{-1}\sum_{t=1}^{n}E\mathbf{v}_{t,n}\mathbf{v}_{t,n}'\right| \leq \left|n^{-1}\sum_{t=1}^{n}\left[E\mathbf{v}_{t,n}\mathbf{v}_{t,n}' + E\mathbf{v}_{t,n}\mathbf{v}_{t,n}'\right]\right|$$

and

$$\eta(j,n) = \left|n^{-1}\sum_{t=1}^{n-j}\left[E\mathbf{v}_{t,n}\mathbf{v}_{t+j,n}' + E\mathbf{v}_{t+j,n}\mathbf{v}_{t,n}'\right]\right|$$

for $j \geq 1$. Using a similar reasoning as before and observing that $E\mathbf{v}_{t,n} = 0$ we obtain for $j \geq 0$ that

$$\begin{aligned} & \eta(j,n) \\ \leq\; & 4^{-1}(p_v+2)\sum_{i,k}\left|e_{ik}'\left(n^{-1}\sum_{t=1}^{n-j}\left[E\mathbf{v}_{t,n}\mathbf{v}_{t+j,n}' + E\mathbf{v}_{t+j,n}\mathbf{v}_{t,n}'\right]\right)e_{ik}\right| \\ =\; & 2^{-1}(p_v+2)\sum_{i,k}\left|n^{-1}\sum_{t=1}^{n-j}\operatorname{cov}(e_{ik}'\mathbf{v}_{t,n}, e_{ik}'\mathbf{v}_{t+j,n})\right|. \end{aligned}$$

Applying Lemma I2 to the process $e_{ik}'\mathbf{v}_{t,n}$ for each (i,k) and observing that $\nu_m(e_{ik}'\mathbf{v}) \leq 2\nu_m(\mathbf{v})$ yields for $n \geq j$:

$$\eta(j,n) \;\leq\; 2^{-1}(p_v+2)\sum_{i,k}n^{-1}\sum_{t=1}^{n-j}\Big[4d\Big\{\nu_m(e_{ik}'\mathbf{v})$$

$$+ 2d(2^{1-1/r} + 1)\alpha(j - 2m)^{1-2/r}\Big\}\Big]$$

$$\leq 2^{-1}(p_v + 2)p_v^2 n^{-1}(n - j)\Big[4d\Big\{2\nu_m(\mathbf{v}) + 2d(2^{1-1/r} + 1)\alpha(j - 2m)^{1-2/r}\Big\}\Big]$$

where

$$d = \sup_n \sup_t \|e'_{ik}\mathbf{v}_{t,n}\|_r \leq 2 \sup_n \sup_t \left\|\mathbf{v}'_{t,n}\mathbf{v}_{t,n}\right\|_r^{1/2} < \infty.$$

Choosing in particular m equal to $[j/4]$, the integer part of $j/4$, yields

$$\sup_{n\geq 1} \eta(j,n) \leq C\left(\nu_{[j/4]}(\mathbf{v}) + \alpha(j - 2[j/4])^{1-2/r}\right),$$

where C is a finite constant not depending on j. Since

$$\begin{aligned}\sum_{j=0}^{\infty} \sup_{n\geq 1} \eta(j,n) &\leq C\sum_{j=0}^{\infty}\left(\nu_{[j/4]}(\mathbf{v}) + \alpha(j - 2[j/4])^{1-2/r}\right) \\ &\leq 4C\sum_{j=0}^{\infty}\left(\nu_j(\mathbf{v}) + \alpha(j)^{1-2/r}\right) < \infty\end{aligned}$$

by the size requirements on $\nu_j(\mathbf{v})$ and $\alpha(j)$, it follows from Lemma 12.1(d) that $\gamma_{3n} \to 0$. Clearly, γ_{1n} and γ_{2n} converge to zero if $\ell_n = o(n^{1/3})$, which completes the proof for the case where $(\mathbf{e}_t)$ is α-mixing. The proof for the case where $(\mathbf{e}_t)$ is ϕ-mixing is analogous. ■

Proof of Corollary 12.3. Clearly all conditions on the process $(\mathbf{v}_{t,n})$ maintained in Corollary 12.2 as well as Assumption 12.1 are satisfied. The weights are clearly bounded. Hence the rates for γ_{1n} and γ_{2n} obtained in Corollary 12.2 apply, since the derivation of these rates did not make use of the condition $\lim_{n\to\infty} w(j,n) = 1$. Since $c_1\ell_n^* \leq \ell_n \leq c_2\ell_n^*$ for some positive finite constants (as discussed earlier) it follows that $\gamma_{1n} = O(\ell_n^*/n^{1/2})$ and $\gamma_{2n} = O((\ell_n^*)^{3/2}/n^{1/2})$. The rate for γ_{3n} follows from Lemma 12.1(e) and the discussion preceding Corollary 12.2 if we can establish that $\sum_{j=0}^{\infty} j^\rho \sup_{n\geq 1}\eta(j,n) < \infty$. Exactly as in the proof of Corollary 12.2 we get

$$\begin{aligned}\sum_{j=0}^{\infty} j^\rho \sup_{n\geq 1} \eta(j,n) &\leq C\sum_{j=0}^{\infty} j^\rho\left(\nu_{[j/4]}(\mathbf{v}) + \alpha(j - 2[j/4])^{1-2/r}\right) \\ &\leq 4^{\rho+1}C\sum_{j=0}^{\infty}(j+1)^\rho\left(\nu_j(\mathbf{v}) + \alpha(j)^{1-2/r}\right).\end{aligned}$$

Now, this sum is finite in view of the stronger size requirements maintained in the corollary. The consistency of $\hat{\Psi}_n$ and the rate for γ_n follow now immediately observing that γ_{1n} is dominated by γ_{2n}. The rate $\ell_n^* \sim n^{1/(2\rho+3)}$

implying the optimal rate for γ_n is obviously obtained by equating the orders of magnitude of γ_{2n} and γ_{3n}. The proof for the case where $(\mathbf{e}_t)$ is ϕ-mixing is analogous. ■

Appendix J

PROOFS FOR CHAPTER 13

Proof of Theorem 13.1. Similarly as in the proof of Theorem 11.2 we may assume w.l.o.g. that $(\hat{\tau}_n, \hat{\beta}_n) \in T' \times B'$. The proof that

$$\hat{C}_n - C_n \to 0 \text{ i.p. as } n \to \infty$$

follows from the proof of Theorem 11.2 (with $(\hat{\tau}_n, \hat{\beta}_n)$ replacing $(\tilde{\tau}_n, \tilde{\beta}_n)$) under both sets of assumptions of the theorem. We next prove that

$$\hat{\Phi}_n - D_n D_n' \to 0 \text{ i.p. as } n \to \infty$$

under the assumptions of part (a) of the theorem. Let $f_t(z, \tau, \beta)$ denote the restriction to $Z \times T' \times B'$ of any of the components of the matrix $\nabla_{\beta'} q_t(z, \tau, \beta) \nabla_\beta q_t(z, \tau, \beta)$. We show that

$$\sup_{T' \times B'} \left| n^{-1} \sum_{t=1}^{n} [f_t(\mathbf{z}_t, \tau, \beta) - E f_t(\mathbf{z}_t, \tau, \beta)] \right| \to 0 \text{ i.p. as } n \to \infty$$

and that $n^{-1} \sum_{t=1}^{n} E f_t(\mathbf{z}_t, \tau, \beta)$ is equicontinuous on $T' \times B'$ by verifying the assumptions of Theorem 5.2. Assumption 5.1 is satisfied since $T' \times B'$ is compact by Assumption 11.1(d). Assumptions B and C in Chapter 5 for $\{f_t : t \in \mathrm{N}\}$ are implied by Assumption 11.2*. Assumption D in Chapter 5 follows from Assumption 11.1(f). Assumption 5.2 now follows from Theorem 6.13 in light of Assumption 11.1(e). Thus all assumptions of Theorem 5.2 are satisfied. Consequently it follows from Lemma 3.2 that

$$n^{-1} \sum_{t=1}^{n} \left[f_t(\mathbf{z}_t, \hat{\tau}_n, \hat{\beta}_n) - E f_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n) \right] \to 0 \text{ i.p. as } n \to \infty,$$

and hence $\hat{\Phi}_n - D_n D_n' \to 0$ i.p. as $n \to \infty$, observing that

$$D_n D_n' = n^{-1} \sum_{t=1}^{n} E[\nabla_{\beta'} q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n) \nabla_\beta q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n)]$$

in view of Assumption 11.4. Part (a) of the theorem now follows using Lemma F1 and observing that $|C_n| = O(1)$, $|C_n^{-1}| = O(1)$ and $|D_n| = O(1)$ by Theorem 11.2(a). We next prove that

$$\hat{\Psi}_n - D_n D_n' \to 0 \text{ i.p. as } n \to \infty$$

under the assumptions of part (b) of the theorem by verifying the assumptions of Corollary 12.2 with

$$\mathbf{v}_t(\tau,\beta) = \nabla_{\beta'} q_t(\mathbf{z}_t,\tau,\beta).$$

Assumption 12.1 follows from Assumption 11.1 observing that $\hat{\beta}_n - \bar{\beta}_n = O_p(n^{-1/2})$ holds by Theorem 11.2(a). The condition

$$\sup_n \sup_t \left\| \mathbf{v}'_{t,n}\mathbf{v}_{t,n} \right\|_r < \infty$$

in Corollary 12.2 follows from Assumption 11.5*. The condition

$$n^{-1}\sum_{t=1}^{n} E\Delta_t^2 = O(1)$$

follows from Assumption 13.1. The remaining conditions on the process $(\mathbf{v}_{t,n})$ are satisfied in light of Assumptions 11.3*(a) and 11.5*. The assumptions on the weights in Corollary 12.2 are identical to those maintained in the theorem. Corollary 12.2 now yields

$$\hat{\Psi}_n - D_n D'_n \to 0 \text{ i.p. as } n \to \infty.$$

Part (b) of the theorem now follows using Lemma F1 observing that $|C_n| = O(1)$, $|C_n^{-1}| = O(1)$ and $|D_n| = O(1)$ by Theorem 11.2(a). ■

Proof of Theorem 13.2. Similarly as in the proof of Theorem 11.5 we may assume w.l.o.g. that $(\hat{\tau}_n, \hat{\beta}_n) \in T' \times B'$. We now show that

$$\hat{C}_n - C_n \to 0 \text{ i.p. as } n \to \infty$$

under both sets of assumptions of the theorem. It follows from Lemma 11.3 that

$$\sup_{T'\times B'} |\nabla_\beta S_n - E\nabla_\beta S_n| \to 0 \text{ i.p. as } n \to \infty,$$

and that the restrictions of $\{E\nabla_\beta S_n : n \in \mathbf{N}\}$ to $T' \times B'$ are equicontinuous on $T' \times B'$. The family of restrictions of $\nabla_{cc}\vartheta_n(0,\tau,\beta)$ to $T' \times B'$ is clearly equicontinuous in view of Assumption 11.6(a). It follows from Lemma 3.2 that

$$\nabla_\beta \hat{S}_n - E\nabla_\beta \underline{S}_n \to 0 \text{ i.p. as } n \to \infty$$

and

$$\nabla_{cc}\hat{\vartheta}_n - \nabla_{cc}\vartheta_n(0,\bar{\tau}_n,\bar{\beta}_n) \to 0 \text{ i.p. as } n \to \infty.$$

Furthermore, since $|E\nabla_\beta \underline{S}_n| = O(1)$ and $\left|\nabla_{cc}\vartheta_n(0,\bar{\tau}_n,\bar{\beta}_n)\right| = O(1)$ by Assumptions 11.1 and 11.6, it follows that $\hat{C}_n - C_n \to 0$ i.p. as $n \to \infty$. We next prove that

$$\hat{\Phi}_n - nE\underline{S}_n\underline{S}'_n \to 0 \text{ i.p. as } n \to \infty$$

under the assumptions of part (a) of the theorem. Let $f_t(z, \tau, \beta)$ denote the restriction to $Z \times T' \times B'$ of any of the components of $q_t(z, \tau, \beta)q_t(z, \tau, \beta)'$. We show that

$$\sup_{T' \times B'} \left| n^{-1} \sum_{t=1}^{n} [f_t(\mathbf{z}_t, \tau, \beta) - Ef_t(\mathbf{z}_t, \tau, \beta)] \right| \to 0 \text{ i.p. as } n \to \infty$$

and that $n^{-1} \sum_{t=1}^{n} Ef_t(\mathbf{z}_t, \tau, \beta)$ is equicontinuous on $T' \times B'$ by verifying the assumptions of Theorem 5.2. Assumption 5.1 is satisfied since $T' \times B'$ is compact by Assumption 11.1(d). Assumptions B and C in Chapter 5 for $\{f_t : t \in \mathrm{N}\}$ are implied by Assumption 11.6*. Assumption D in Chapter 5 follows from Assumption 11.1(f). Assumption 5.2 now follows from Theorem 6.13 in light of Assumption 11.1(e). Thus all assumptions of Theorem 5.2 are satisfied. Consequently it follows from Lemma 3.2 that

$$n^{-1} \sum_{t=1}^{n} \left[f_t(\mathbf{z}_t, \hat{\tau}_n, \hat{\beta}_n) - Ef_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n) \right] \to 0 \text{ i.p. as } n \to \infty,$$

and hence

$$\hat{\Phi}_n - nE\underline{S}_n\underline{S}_n' \to 0 \text{ i.p. as } n \to \infty,$$

since

$$nE\underline{S}_n\underline{S}_n' = n^{-1} \sum_{t=1}^{n} Eq_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n) q_t(\mathbf{z}_t, \bar{\tau}_n, \bar{\beta}_n)'$$

in view of Assumption 11.8. Note also that

$$\left| nE\underline{S}_n\underline{S}_n' \right| = O(1)$$

in view of the moment condition in Assumption 11.6*. Hence

$$\nabla_{\beta'} \hat{S}_n \nabla_{cc} \hat{\vartheta}_n \hat{\Phi}_n \nabla_{cc} \hat{\vartheta}_n \nabla_{\beta} \hat{S}_n - D_n D_n' \to 0 \text{ i.p. as } n \to \infty$$

and clearly $|D_n D_n'| = O(1)$. Part (a) now follows using Lemma F1, observing that $|C_n| = O(1)$ and $\left|C_n^{-1}\right| = O(1)$ by Theorem 11.5(a). We next prove that

$$\hat{\Psi}_n - nE\underline{S}_n\underline{S}_n' \to 0 \text{ i.p. as } n \to \infty$$

under the assumptions of part (b) of the theorem by verifying the assumptions of Corollary 12.2 with

$$\mathbf{v}_t(\tau, \beta) = q_t(\mathbf{z}_t, \tau, \beta).$$

Assumption 12.1 follows from Assumption 11.1 observing that $\hat{\beta}_n - \bar{\beta}_n = O_p(n^{-1/2})$ holds by Theorem 11.5(a). The condition

$$\sup_n \sup_t \left\| \mathbf{v}_{t,n}' \mathbf{v}_{t,n} \right\|_r < \infty$$

in Corollary 12.2 follows from Assumption 11.9*. The condition

$$n^{-1}\sum_{t=1}^{n} E\Delta_t^2 = O(1)$$

follows from Assumption 13.2. The remaining conditions on the process $(\mathbf{v}_{t,n})$ are satisfied in light of Assumptions 11.7*(a) and 11.9*. The assumptions on the weights in Corollary 12.2 are identical to those maintained in the theorem. Corollary 12.2 now yields

$$\hat{\Psi}_n - nE\underline{S}_n\underline{S}_n' \to 0 \text{ i.p. as } n \to \infty.$$

The boundedness of $nE\underline{S}_n\underline{S}_n'$ is established in the proof of Theorem 11.5(a). Part (b) of the theorem now follows analogously as part (a). ∎

Appendix K

PROOFS FOR CHAPTER 14

Proof of Theorem 14.1. To prove the theorem we verify Assumptions 7.1 and 7.2 of Theorem 7.1 with $p_q = 1$ and $\vartheta_n(c, \tau, \beta) \equiv c$.[1] Assumptions 7.1(a),(b),(e) are trivially satisfied. Tightness of $\{\bar{H}_n^z : n \in \mathbf{N}\}$ follows from Assumption 14.1(d),(e), cf. Lemmata C1 and C2. Assumption 7.1(d) follows from Assumption 14.2 and Lemma 6.9. Furthermore, we have

$$\sup_{A \times S} \left| f_t' \Sigma^{-1} f_t \right|^{1+\gamma} \leq a_1 \sup_A |f_t|^{2+2\gamma}$$

and

$$\sup_S |\ln \det(\Sigma)| \leq a_2$$

for some real constants a_1 and a_2 in view of Assumption 14.1(b). Hence

$$\begin{aligned} d_t(\mathbf{z}_t)^{1+\gamma} \quad \leq \quad & 3^\gamma \left[\sup_A |\ln |\det(\partial f_t / \partial y)||^{1+\gamma} \right. \\ & \left. + (a_2/2)^{1+\gamma} + (a_1/2)^{1+\gamma} \sup_A |f_t|^{2+2\gamma} \right]. \end{aligned}$$

Consequently, also Assumption 7.1(c) is satisfied in view of Assumption 14.1(c). Finally, equicontinuity of q_t follows from Assumption 14.1(f) if we can show that $f_t' \, \Sigma^{-1} f_t$ is equicontinuous on $Z \times A \times S$. This follows since in view of Assumptions 14.1(f),(g) $\{f_t : t \in \mathbf{N}\}$ is equicontinuous at any point (z_0, α_0) and $\sup_t |f_t(z, \alpha)|$ is bounded in a suitable small neighborhood of (z_0, α_0), and since $\Sigma^{-1}(\sigma)$ is continuous. ■

Lemma K1. *Let the Assumptions 14.1(a),(c)-(g) and 14.2 hold. Furthermore assume that S corresponds to the set of all symmetric and positive definite $p_e \times p_e$ matrices and that*

$$a_1 = \inf_n \inf_A \lambda_{\min} \left(n^{-1} \sum_{t=1}^n E f_t f_t' \right)$$

[1]Within the context of Theorem 14.1 no nuisance parameter τ is present. To incorporate this case into the framework of Theorem 7.1, we may view the objective function formally as a function on $T \times B$, where T can be chosen as an arbitrary compact subset of some Euclidean space. We then also set $\hat{\tau}_n = \bar{\tau}_n = \tau_0$, an arbitrary element of T.

is positive.

(a) Then

$$a_2 = \inf_n \inf_A \left(-n^{-1} \sum_{t=1}^{n} E \ln |\det(\nabla_y f_t)| \right)$$

is well-defined and finite. Furthermore,

$$\sup_n \sup_A \left| \bar{R}_n(\alpha, \sigma) \right| < \infty$$

for all $\sigma \in S$.

(b) For every $\delta \in \mathbf{R}$ *we can find constants* $0 < c_1 \leq c_2 < \infty$*, depending only on* δ*,* a_1*, and* a_2*, such that the level sets* $\{(\alpha, \sigma) \in A \times S : \bar{R}_n(\alpha, \sigma) < \delta\}$ *are contained in the compact set*

$$A \times \{\sigma \in S : c_1 \leq \lambda_{\min}(\Sigma(\sigma)), \lambda_{\max}(\Sigma(\sigma)) \leq c_2\}$$

for all $n \in \mathbf{N}$.

(c) For every $\delta \in \mathbf{R}$ *we can find constants* $0 < c_1^* \leq c_2^* < \infty$*, depending only on* δ*,* a_1*, and* a_2*, and a sequence* $\Omega_n \in \mathfrak{A}$ *with* $P(\Omega_n) \to 1$ *as* $n \to \infty$ *such that the level sets* $\{(\alpha, \sigma) \in A \times S : R_n(\omega, \alpha, \sigma) < \delta\}$ *are contained in the compact set*

$$A \times \{\sigma \in S : c_1^* \leq \lambda_{\min}(\Sigma(\sigma)), \lambda_{\max}(\Sigma(\sigma)) \leq c_2^*\}$$

for all $\omega \in \Omega_n$ *and all* $n \in \mathbf{N}$.

(d) There exists a $\delta^* \in \mathbf{R}$ *such that for* $\delta \geq \delta^*$ *the level sets* $\{(\alpha, \sigma) \in A \times S : \bar{R}_n(\alpha, \sigma) < \delta\}$ *and* $\{(\alpha, \sigma) \in A \times S : R_n(\omega, \alpha, \sigma) < \delta\}$ *are nonempty for* $\omega \in \Omega_n^\bullet$ *and all* $n \in \mathbf{N}$*, where* $\Omega_n^\bullet \in \mathfrak{A}$ *satisfies* $P(\Omega_n^\bullet) \to 1$ *as* $n \to \infty$.

Proof. We first prove part (a). The finiteness of a_2 follows immediately from Assumption 14.1(c). Since

$$\begin{aligned} \sup_A \left| \bar{R}_n(\alpha, \sigma) \right| \quad \leq \quad & n^{-1} \sum_{t=1}^{n} E \sup_A |\ln |\det(\nabla_y f_t)|| + |(1/2) \ln \det(\Sigma(\sigma))| \\ & + (2\lambda_{\min}(\Sigma(\sigma)))^{-1} n^{-1} \sum_{t=1}^{n} E \sup_A f_t' f_t, \end{aligned}$$

the second claim in part (a) follows from Assumption 14.1(c).

We next prove part (b). First note that a_1 is also finite because of Assumption 14.1(c). Clearly,

$$\bar{R}_n(\alpha, \sigma) \geq a_2 + (p_e/2) \ln (\lambda_{\min}(\Sigma(\sigma))) + (a_1/2) (\lambda_{\min}(\Sigma(\sigma)))^{-1},$$

where we have used the fact that for symmetric positive definite matrices C and D the inequality $\operatorname{tr}(CD) \geq \lambda_{\min}(C)\lambda_{\max}(D)$ holds. The function

$$\psi(x) = a_2 + (p_e/2) \ln x + (a_1/2) x^{-1}$$

satisfies $\lim_{x\to 0}\psi(x)=\infty$ and $\lim_{x\to\infty}\psi(x)=\infty$, since $a_1>0$. Furthermore, $\psi(x)$ has its unique minimizer at $x=a_1/p_e$, is strictly decreasing for $0<x<a_1/p_e$, and is strictly increasing for $a_1/p_e<x<\infty$. Consider first the case where $\delta>\inf\{\psi(x):x>0\}$. Clearly, $\psi(x)=\delta$ has exactly two solutions in the set $(0,\infty)$. Define c_1 as the smaller of the two solutions. It then follows that the set

$$\left\{\sigma\in S:\bar{R}_n(\alpha,\sigma)<\delta \text{ for some } \alpha\in A\right\}$$

is a subset of

$$\{\sigma\in S:c_1\leq\lambda_{\min}(\Sigma(\sigma))\}$$

and c_1 depends only on δ, a_1, and a_2. Next observe that

$$\bar{R}_n(\alpha,\sigma)\geq a_2+(1/2)\left[\ln\left(\lambda_{\max}\left(\Sigma(\sigma)\right)\right)+(p_e-1)\ln\left(\lambda_{\min}\left(\Sigma(\sigma)\right)\right)\right].$$

Utilizing what we have just shown, it follows further that for all $\sigma\in S$ with $\bar{R}_n(\alpha,\sigma)<\delta$ for some $\alpha\in A$ we have

$$\begin{aligned}\bar{R}_n(\alpha,\sigma) &\geq a_2+(1/2)\left[\ln\left(\lambda_{\max}\left(\Sigma(\sigma)\right)\right)+(p_e-1)\ln\left(\lambda_{\min}\left(\Sigma(\sigma)\right)\right)\right]\\ &\geq a_2+(1/2)\left[\ln\left(\lambda_{\max}\left(\Sigma(\sigma)\right)\right)+(p_e-1)\ln(c_1)\right].\end{aligned}$$

Hence, for all such σ we have $\lambda_{\max}(\Sigma(\sigma))\leq c_2$ where

$$c_2=c_1^{-(p_e-1)}\exp\left(2(\delta-a_2)\right),$$

which is finite and positive, and thus

$$\begin{aligned}&\left\{\sigma\in S:\bar{R}_n(\alpha,\sigma)<\delta \text{ for some } \alpha\in A\right\}\\ \subseteq\ &\{\sigma\in S:c_1\leq\lambda_{\min}\left(\Sigma(\sigma)\right),\lambda_{\max}\left(\Sigma(\sigma)\right)\leq c_2\}.\end{aligned}$$

This completes the proof for the case $\delta>\inf\{\psi(x):x>0\}$. For $\delta\leq\inf\{\psi(x):x>0\}$ the result holds also, because the sets $\{\sigma\in S:\bar{R}_n(\alpha,\sigma)<\delta$ for some $\alpha\in A\}$ are monotonically increasing with δ.

We next prove part (c). We first note that

$$\sup_A\left|n^{-1}\sum_{t=1}^n\ln\left|\det(\nabla_y f_t)\right|-n^{-1}\sum_{t=1}^n E\ln\left|\det(\nabla_y f_t)\right|\right|$$

and

$$\sup_A\left|n^{-1}\sum_{t=1}^n f_tf_t'-n^{-1}\sum_{t=1}^n Ef_tf_t'\right|$$

converge to zero in probability. This follows from Theorem 5.2 (utilizing Theorem 6.13 for the verification of the local LLNs postulated in Assumption 5.2). Observe that

$$\left\{n^{-1}\sum_{t=1}^n Ef_tf_t':n\in\mathbf{N}\right\}$$

is, in view of Assumption 14.1(c), contained in a compact set, say K, of symmetric nonnegative definite $p_e \times p_e$ matrices. Applying Lemma 3.3 with C equal to the set of all symmetric nonnegative definite $p_e \times p_e$ matrices and with $\vartheta_n(.) = \lambda_{\min}(.)$, which clearly is continuous, it follows further that

$$\sup_A \left| \lambda_{\min}\left(n^{-1}\sum_{t=1}^{n} f_t f_t'\right) - \lambda_{\min}\left(n^{-1}\sum_{t=1}^{n} E f_t f_t'\right)\right|$$

converges to zero in probability. Choose $\epsilon = a_1/2$ and let Ω_n be the set of all $\omega \in \Omega$ such that

$$\sup_A \left| n^{-1}\sum_{t=1}^{n} \ln|\det(\nabla_y f_t)| - n^{-1}\sum_{t=1}^{n} E \ln|\det(\nabla_y f_t)|\right| < \epsilon$$

and

$$\sup_A \left| \lambda_{\min}\left(n^{-1}\sum_{t=1}^{n} f_t f_t'\right) - \lambda_{\min}\left(n^{-1}\sum_{t=1}^{n} E f_t f_t'\right)\right| < \epsilon.$$

Then $P(\Omega_n) \to 1$ holds, because of the uniform convergence results just established. (In view of the continuity of the various expressions the suprema are measurable and thus $\Omega_n \in \mathfrak{A}$.) Hence for $\omega \in \Omega_n$, all $n \in \mathbf{N}$ and for all $\alpha \in A$, $\sigma \in S$ we have

$$\begin{aligned} R_n(\omega,\alpha,\sigma) \;\geq\; & -n^{-1}\sum_{t=1}^{n} E\ln|\det(\nabla_y f_t)| - \epsilon + (p_e/2)\ln(\lambda_{\min}(\Sigma(\sigma))) \\ & + (1/2)\left(\lambda_{\min}\left(n^{-1}\sum_{t=1}^{n} E f_t f_t'\right) - \epsilon\right)(\lambda_{\min}(\Sigma(\sigma)))^{-1} \\ \geq\; & (a_2 - a_1/2) + (p_e/2)\ln(\lambda_{\min}(\Sigma(\sigma))) \\ & + (a_1/4)(\lambda_{\min}(\Sigma(\sigma)))^{-1} \end{aligned}$$

and

$$\begin{aligned} R_n(\omega,\alpha,\sigma) \;\geq\; & -n^{-1}\sum_{t=1}^{n} E\ln|\det(\nabla_y f_t)| - \epsilon \\ & + (1/2)\left[\ln(\lambda_{\max}(\Sigma(\sigma))) + (p_e - 1)\ln(\lambda_{\min}(\Sigma(\sigma)))\right] \\ \geq\; & (a_2 - a_1/2) + (1/2)\left[\ln(\lambda_{\max}(\Sigma(\sigma)))\right. \\ & \left. + (p_e - 1)\ln(\lambda_{\min}(\Sigma(\sigma)))\right]. \end{aligned}$$

Exactly the same argument as in the proof of part (b) completes the proof of (c).

Finally we prove part (d). Choose some $\sigma^* \in S$ and some $\epsilon^\bullet > 0$. Define $\Omega_n^\bullet$ as the set of all $\omega \in \Omega$ such that

$$\sup_A \left| R_n(\omega,\alpha,\sigma^*) - \bar{R}_n(\alpha,\sigma^*)\right| < \epsilon^\bullet.$$

Then $\Omega_n^\bullet \in \mathfrak{A}$ and satisfies $P(\Omega_n^\bullet) \to 1$ as $n \to \infty$ since

$$\sup_A \left| R_n(\omega, \alpha, \sigma^*) - \bar{R}_n(\alpha, \sigma^*) \right|$$

converges to zero in probability. This uniform convergence follows similarly as the uniform convergence results established in the proof of part (c). Setting

$$\delta^* = \sup_n \sup_A \left| \bar{R}_n(\alpha, \sigma^*) \right| + \epsilon^\bullet$$

and observing that $\delta^* < \infty$ in view of Assumption 14.1(c) completes the proof. ■

Proof of Lemma 14.2. To prove the result we make use of part (b) of Theorem 6.12. As in the discussion before the lemma let

$$\begin{aligned} \mathbf{v}_t &= \left(\mathbf{y}_t', \ldots, \mathbf{y}_{t-l+1}'\right)', \\ \mathbf{w}_t &= (\mathbf{x}_t', \epsilon_t')' \end{aligned}$$

for $t \geq 1$ and set $\mathbf{w}_t = 0$ for $t \leq 0$. Clearly, given Assumption 14.5(b), the functions $\phi_t^{(k)}$ satisfy the Lipschitz-type conditions on the iterated transformation functions in Theorem 6.12. As a consequence of Assumptions 14.5(a) and 14.6 we have first for $i = 1$:

$$\begin{aligned} \|\mathbf{y}_i\|_1 &\leq \|\mathbf{y}_i - g_i(0, \ldots, 0, \alpha_0)\|_1 + |g_i(0, \ldots, 0, \alpha_0)| \\ &\leq c_1 \left(\|\mathbf{y}_{i-1}\|_1 + \ldots + \|\mathbf{y}_{i-l}\|_1\right) \\ &\quad + c_2 \left(\|\mathbf{x}_i\|_1 + \|\epsilon_i\|_1\right) + |g_i(0, \ldots, 0, \alpha_0)| < \infty. \end{aligned}$$

Applying the above inequality successively for $i = 2, 3, \ldots,$ we see that $\|\mathbf{y}_i\|_1 < \infty$ for $i = 1, \ldots, k^* - 1$ and in fact for all $i \geq 1$. Hence the assumption in Theorem 6.12 that $\|\mathbf{v}_i\|_1 < \infty$ for $i = 0, \ldots, k^* - 1$ is satisfied. Assumption 14.6 implies further that

$$\sup_{t \geq 1} \|\mathbf{w}_t\|_1 < \infty.$$

Since $\mathbf{w}_t = (\mathbf{x}_t', \epsilon_t')'$ is L_1-approximable by $(\mathbf{e}_t)$ by assumption, it then follows from part (b) of Theorem 6.12 that also $\mathbf{v}_t = (\mathbf{y}_t', \ldots, \mathbf{y}_{t-l+1}')'$ is L_1-approximable by $(\mathbf{e}_t)$. By Lemma 6.9 it follows further that $(\mathbf{y}_t', \mathbf{x}_t')'$ is L_1-approximable and hence, in light of Theorem 6.1, also L_0-approximable by $(\mathbf{e}_t)$. ■

Proof of Lemma 14.3. To prove the result we make use of part (a) of Theorem 6.12. Clearly, given Assumption 14.5(b), the functions $\phi_t^{(k^*)}$ satisfy the Lipschitz-type conditions on the iterated transformation functions in Theorem 6.12. Let $V = \mathbf{R}^{lp_y}$ and $W = \mathbf{R}^{p_x+p_e}$. We next verify the assumption in Theorem 6.12(a) that

$$\sup_{t\geq 1}\left|\phi_t^{(k^*)}(\bar{v},\bar{w}_1,\ldots,\bar{w}_{k^*})\right| < \infty$$

for some elements $\bar{v} = (\bar{y}'_{-1},\ldots,\bar{y}'_{-l})' \in V$ and $\bar{w}_i \in W$. To this end choose $\bar{v} = (\bar{y}',\ldots,\bar{y}')' \in V$ and $\bar{w}_i = \bar{w} = (\bar{x}',\bar{e}')' \in W$ where $\bar{y} \in \mathbf{R}^{p_y}$, $\bar{x} \in \mathbf{R}^{p_x}$ and $\bar{e} \in \mathbf{R}^{p_e}$ are arbitrary elements. Define

$$a_t = \phi_t^{(k^*)}(\bar{v},\bar{w},\ldots,\bar{w})$$

and

$$b_t = f_t(\bar{y},\bar{y},\ldots,\bar{y},\bar{x},\alpha_0).$$

Observe that then

$$\bar{y} = g_t(\bar{y},\ldots,\bar{y},\bar{x},b_t,\alpha_0)$$

and

$$\bar{v} = (\bar{y}',\ldots,\bar{y}')' = \phi_t^{(k^*)}(\bar{v},\bar{w}_{t,1},\ldots,\bar{w}_{t,k^*})$$

with $\bar{w}_{t,i+1} = (\bar{x}',b'_{t+i})'$, $i = 0,\ldots,k^*-1$, holds. By Assumption 14.1(g) the sequence b_t is bounded in absolute value. It now follows from Assumption 14.5(b) that

$$\begin{aligned}|a_t - \bar{v}| &= \left|\phi_t^{(k^*)}(\bar{v},\bar{w},\ldots,\bar{w}) - \phi_t^{(k^*)}(\bar{v},\bar{w}_{t,1},\ldots,\bar{w}_{t,k^*})\right| \\ &\leq d_2\left|\begin{bmatrix}\bar{e}-b_t\\ \cdot\\ \cdot\\ \cdot\\ \bar{e}-b_{t+k^*-1}\end{bmatrix}\right| \leq \text{const} < \infty,\end{aligned}$$

which verifies that

$$\sup_{t\geq 1}|a_t| = \sup_{t\geq 1}\left|\phi_t^{(k^*)}(\bar{v},\bar{w},\ldots,\bar{w})\right| < \infty.$$

Let $\mathbf{v}_t$ and $\mathbf{w}_t$ be defined as in the proof of Lemma 14.2. As demonstrated in the proof of Lemma 14.2 we have $\|\mathbf{v}_i\|_1 < \infty$ for $i = 0,\ldots,k^*-1$. Furthermore, observe that

$$\sup_{t\geq 1}\|\mathbf{w}_t\|_1 < \infty$$

by Assumption 14.6. It now follows from part (a) of Theorem 6.12 that

$$\sup_{t\geq 1}\|\mathbf{v}_t\|_1 = \sup_{t\geq 1}E|\mathbf{v}_t| < \infty.$$

Since

$$\sup_{n\geq 1} n^{-1}\sum_{t=1}^{n} E\left|\mathbf{y}_t\right| \leq \sup_{t\geq 1} E\left|\mathbf{y}_t\right| \leq \sup_{t\geq 1} E\left|\mathbf{v}_t\right|,$$

Assumption 14.1(d) is seen to hold with $\gamma = 1$. ■

Proof of Theorem 14.4. Given Assumptions 14.6* and 14.7 it follows from Theorem 6.1 that $(\mathbf{x}_t', \epsilon_t')'$ is L_1-approximable by the α-mixing basis process $(\mathbf{e}_t)$. Given Assumptions 14.3 - 14.5, 14.6* it now follows from Lemma 14.2 that $(\mathbf{y}_t', \mathbf{x}_t')'$ is L_0-approximable by the α-mixing basis process $(\mathbf{e}_t)$. Thus Assumption 14.2 holds. Assumption 14.1(e) follows from Assumption 14.6*(b). Assumption 14.1(d) follows from Lemma 14.3. Thus also all of Assumption 14.1 holds. Consequently, the uniform convergence result and equicontinuity of $\{\bar{R}_n : n \in \mathrm{N}\}$ follows from Theorem 14.1. Furthermore, under the assumptions of part (a) of the theorem, identifiable uniqueness of $\bar{\beta}_n \equiv \beta_0$ follows from Lemma 4.1. The claim in part (a) now follows from Theorem 14.1. To prove the claim in part (b) observe that in view of Ascoli-Arzelà's theorem the convergence of $\bar{R}_n$ to $\bar{R}$ is uniform over B and $\bar{R}$ is continuous, since $\{\bar{R}_n : n \in \mathrm{N}\}$ is equicontinuous. Hence R_n converges uniformly over B to $\bar{R}$ in probability. Clearly, β_0 is an identifiably unique minimizer of $\bar{R}$, since B is compact and since $\bar{R}$ is continuous. Result (b) then follows from Lemma 3.1 (applied to R_n and $\bar{R}$, and not to R_n and $\bar{R}_n$). ■

Proof of Lemma 14.5. The density of ϵ_t is given by

$$p(e, \sigma_0) = (2\pi)^{-p_e/2}\left[\det\left(\Sigma(\sigma_0)\right)\right]^{-1/2} \exp\left(-e'\Sigma(\sigma_0)^{-1}e/2\right).$$

Since ϵ_t is independent of $\{\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{1-l}, \mathbf{x}_t, \ldots, \mathbf{x}_1\}$ it follows that the conditional distribution of

$$\mathbf{y}_t = g_t(\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-l}, \mathbf{x}_t, \epsilon_t, \alpha_0)$$

given $\mathbf{y}_{t-1} = y_{t-1}, \ldots, \mathbf{y}_{1-l} = y_{1-l}, \mathbf{x}_t = x_t, \ldots, \mathbf{x}_1 = x_1$ is identical with the distribution of $g_t(y_{t-1}, \ldots, y_{t-l}, x_t, \epsilon_t, \alpha_0)$, cf. Theorem 5.3.22 in Gänssler and Stute (1977). Applying the transformation technique to $g_t(y_{t-1}, \ldots, y_{t-l}, x_t, \epsilon_t, \alpha_0)$ it follows that the density of $\mathbf{y}_t$ conditional on $\mathbf{y}_{t-1} = y_{t-1}, \ldots, \mathbf{y}_{1-l} = y_{1-l}, \mathbf{x}_t = x_t, \ldots, \mathbf{x}_1 = x_1$ is given by

$$\begin{aligned} & \pi_t^Y\left(y_t \mid y_{t-1}, \ldots, y_{1-l}, x_t, \ldots, x_1; \beta_0\right) \\ = \; & (2\pi)^{-p_e/2}\left|\det(\nabla_y f_t)\right| \left[\det\left(\Sigma(\sigma_0)\right)\right]^{-1/2} \exp\left(-f_t'\Sigma(\sigma_0)^{-1} f_t/2\right) \end{aligned}$$

where $f_t = f_t(y_t, \ldots, y_{t-l}, x_t, \alpha_0)$. An analogous argument shows that the conditional density $\pi_t^Y(y_t \mid y_{t-1}, \ldots, y_{t-l}, x_t; \beta_0)$ is given by the same formula. We next prove that β_0 minimizes the conditional expectation in part

(b) of the lemma. To show this observe that

$$\begin{aligned} & E\left[q_t\left(\mathbf{z}_t,\beta\right)-q_t\left(\mathbf{z}_t,\beta_0\right) \mid \mathbf{y}_{t-1}=y_{t-1},\ldots,\mathbf{y}_{t-l}=y_{t-l},\mathbf{x}_t=x_t\right] \\ = & \int\left[q_t\left(z_t,\beta\right)-q_t\left(z_t,\beta_0\right)\right]\pi_t^Y\left(y_t \mid y_{t-1},\ldots,y_{t-l},x_t;\beta_0\right)dy_t \end{aligned}$$

is equal to the Kullback-Leibler divergence between

$$\pi_t^Y\left(y_t \mid y_{t-1},\ldots,y_{t-l},x_t;\beta\right)$$

and

$$\pi_t^Y\left(y_t \mid y_{t-1},\ldots,y_{t-l},x_t;\beta_0\right).$$

The result then follows from a well-known property of the Kullback-Leibler divergence, see Wald (1949). The remaining claims are then simple consequences of this fact. ■

Proof of Theorem 14.6. Because of Assumption 14.8 clearly Assumption 14.6*(*c*) holds. Hence the uniform convergence result and equicontinuity of $\{\bar{R}_n : n \in \mathbf{N}\}$ follows from Theorem 14.4. Lemma 14.5 shows furthermore that β_0 minimizes $\bar{R}_n$ since $E\left|q_t(\mathbf{z}_t,\beta)\right| < \infty$ holds under the assumptions of the theorem. Consequently claim (a) follows from the corresponding claim in Theorem 14.4. As in the proof of Theorem 14.4 it follows that under (b) $\bar{R}_n$ converges uniformly to $\bar{R}$, and hence β_0 also minimizes $\bar{R}$. Therefore, also claim (b) follows from the corresponding claim in Theorem 14.4. ■

Lemma K2. *Suppose Assumptions 14.3, 14.4, and 14.8 hold, and suppose* $E\left|q_t(\mathbf{z}_t,\beta_0)\right| < \infty$ *for all* $t \geq 1$.

(a) The parameter β_0 *is identified in* B *at sample size* n, *in the sense that* $\beta_0 \neq \beta$ *implies* $R_n(\omega,\beta_0) \neq R_n(\omega,\beta)$ *with positive probability, if and only if* β_0 *is the unique minimizer of* $\bar{R}_n(\beta)$ *over* B.

(b) The parameter β_0 *is identified in* B *at sample size* n *if and only if* β_0 *is identified in* B *at sample size* m *for all* $m \geq n$.

Proof of Lemma K2. From Lemma 14.5 we know that β_0 minimizes $\bar{R}_n$ for all $n \geq 1$ and $Eq_t(\mathbf{z}_t,.)$ for every $t \geq 1$.

We first prove the sufficiency part of (a). Assume that there exists $\beta_1 \neq \beta_0$ that is also a minimizer of $\bar{R}_n$. Observe that $Eq_t(\mathbf{z}_t,\beta_0) \leq Eq_t(\mathbf{z}_t,\beta_1)$ and that $\bar{R}_n(\beta_0) = \bar{R}_n(\beta_1)$. Consequently, β_1 also minimizes $Eq_t(\mathbf{z}_t,.)$ for every $1 \leq t \leq n$, and hence $Eq_t(\mathbf{z}_t,\beta_0) = Eq_t(\mathbf{z}_t,\beta_1)$ for every $1 \leq t \leq n$. This implies

$$\begin{aligned} 0 &= Eq_t(\mathbf{z}_t,\beta_1) - Eq_t(\mathbf{z}_t,\beta_0) \\ &= \int\left\{\int\left[\ln \pi_t^Y\left(y_t \mid y_{t-1},\ldots,y_{t-l},x_t;\beta_0\right)\right.\right. \end{aligned}$$

$$- \ln \pi_t^Y (y_t \mid y_{t-1}, \ldots, y_{t-l}, x_t; \beta_1)]$$
$$\left. \pi_t^Y (y_t \mid y_{t-1}, \ldots, y_{t-l}, x_t; \beta_0)\, dy_t \right\} dH_t,$$

where H_t represents the joint distribution function of $\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-l}, \mathbf{x}_t$. The inner integral is the Kullback-Leibler divergence between the respective conditional densities and is therefore nonnegative. Consequently, the Kullback-Leibler divergence is zero H_t-a.s. This implies that the densities

$$\pi_t^Y (y_t \mid y_{t-1}, \ldots, y_{t-l}, x_t; \beta_0)$$

and

$$\pi_t^Y (y_t \mid y_{t-1}, \ldots, y_{t-l}, x_t; \beta_1)$$

coincide for all values $(y'_{t-1}, \ldots, y'_{t-l}, x'_t)' \in \mathbf{R}^{lp_y + p_x}$, except possibly for those in a set N_1 with $H_t(N_1) = 0$, and for all values of $y_t \in \mathbf{R}^{p_y}$, except possibly for those in a set $N_2(y_{t-1}, \ldots, y_{t-l}, x_t)$, which has Lebesque measure zero and hence has measure zero also under the conditional density

$$\pi_t^Y (y_t \mid y_{t-1}, \ldots, y_{t-l}, x_t; \beta_0).$$

This implies further that the functions

$$\pi_t^Y (y_t \mid y_{t-1}, \ldots, y_{t-l}, x_t; \beta_0)$$

and

$$\pi_t^Y (y_t \mid y_{t-1}, \ldots, y_{t-l}, x_t; \beta_1)$$

coincide, except possibly on a set which has measure zero under the distribution of $\mathbf{z}_t$.[2] Since both functions are strictly positive we may take logarithms to arrive at $q_t(\mathbf{z}_t, \beta_0) = q_t(\mathbf{z}_t, \beta_1)$ a.s. This contradicts the maintained identifiability assumption on β_0. The necessity part of (a) is trivial.

Given part (a) has already been established, we next prove the sufficiency part of (b) by showing that β_0 is the unique minimizer of $\bar{R}_m(\beta)$ for any $m \geq n$. Consider $\beta_1 \neq \beta_0$. In light of part (a) clearly

$$\sum_{t=1}^{n} E q_t(\mathbf{z}_t, \beta_0) = n\bar{R}_n(\beta_0) < n\bar{R}_n(\beta_1) = \sum_{t=1}^{n} E q_t(\mathbf{z}_t, \beta_1).$$

[2]Since both functions are clearly measurable, the set on which they differ, say N_3, is measurable. Clearly, $N_3 \subseteq (\mathbf{R}^{p_y} \times N_1) \cup (\bigcup \{N_2(y_{t-1}, \ldots, y_{t-l}, x_t) \times \{(y'_{t-1}, \ldots, y'_{t-l}, x'_t)'\} : (y'_{t-1}, \ldots, y'_{t-l}, x'_t)' \in \mathbf{R}^{lp_y + p_x} - N_1\})$. Observe that $\mathbf{R}^{p_y} \times N_1$ has measure zero under the distribution induced by $\mathbf{z}_t$, since $H_t(N_1) = 0$. For given $(y'_{t-1}, \ldots, y'_{t-l}, x'_t)'$ in the complement of N_1, the sections of the second set in the above union have measure zero under $\pi_t^Y (y_t \mid y_{t-1}, \ldots, y_{t-l}, x_t; \beta_0)$. Hence the second set has measure zero under the distribution induced by $\mathbf{z}_t$ in view of Fubini's theorem for regular conditional distributions.

Since $Eq_t(\mathbf{z}_t,\beta_0) \leq Eq_t(\mathbf{z}_t,\beta_1)$ for all $t \geq 1$, as noted above, it follows that

$$\begin{aligned} m\bar{R}_m(\beta_0) &= n\bar{R}_n(\beta_0) + \sum_{t=n+1}^{m} Eq_t(\mathbf{z}_t,\beta_0) \\ &< n\bar{R}_n(\beta_1) + \sum_{t=n+1}^{m} Eq_t(\mathbf{z}_t,\beta_1) \\ &= m\bar{R}_m(\beta_1), \end{aligned}$$

which establishes the sufficiency part. The necessity part of (b) is trivial. ■

Proof of Theorem 14.7. We first note that all assumptions of part (a) of Theorem 14.4 are satisfied and hence $\hat{\beta}_n$ is consistent for β_0. We now prove part (a) of the theorem by verifying the assumptions of Theorem 11.2(a).[3] Assumption 11.1(a) is satisfied because of Assumption 14.1(a),(b). The $\mathfrak{Z}$-measurability of $q_t(.,\beta)$ follows from the stronger assumption of equicontinuity expressed in Assumption 14.1(f). Twice continuous partial differentiability of $q_t(z,.)$ is a consequence of Assumption 14.10(a) and the assumed nonsingularity of $\nabla_y f_t(z,\alpha)$. Hence, Assumption 11.1(b) is satisfied. Since $\hat{\beta}_n$ is consistent for β_0, and β_0 is assumed to belong to the interior of B in Assumption 14.9, also $\hat{\beta}_n$ belongs to the interior for $\omega \in \Omega_n$ with $P(\Omega_n) \to 1$ as $n \to \infty$. Since $\hat{\beta}_n$ minimizes R_n, $\hat{\beta}_n$ satisfies the first order conditions and hence Assumption 11.1(c) holds. Because of consistency of $\hat{\beta}_n$ and because of Assumption 14.9 clearly also Assumption 11.1(d) holds upon choosing B' as a compact neighborhood of β_0 with $B' \subseteq \text{int}(B)$ (and T' as a compact neighborhood of τ_0 with $T' \subseteq \text{int}(T)$, cf. Footnote 3). Given Assumptions 14.3 - 14.5, 14.6* and 14.7, it follows from Theorem 6.1 that $(\mathbf{x}_t', \epsilon_t')'$ is L_1-approximable by the α-mixing basis process $(\mathbf{e}_t)$, and from Lemma 14.2 that $(\mathbf{y}_t', \mathbf{x}_t')'$ is L_0-approximable by the α-mixing basis process $(\mathbf{e}_t)$. Consequently, $(\mathbf{z}_t)$ is L_0-approximable by the α-mixing basis process $(\mathbf{e}_t)$ in view of Lemma 6.9. Hence Assumption 11.1(e) is satisfied. Assumptions 14.1(g), 14.3 - 14.6* together with Lemma 14.3 imply Assumptions 14.1(d),(e), which in turn implies Assumption 11.1(f), cf. Lemmata C1 and C2.

We next verify Assumption 11.2. First observe that $\Sigma(\sigma)$ is clearly (equi-) continuous and bounded on the compact set S, and hence on $\text{int}(S)$. Since the smallest eigenvalue of $\Sigma(\sigma)$ is bounded away from zero as σ varies in S, cf. Footnote 3 in Chapter 14, it follows that $\det(\Sigma(\sigma))$ is bounded away

[3]For the present problem no nuisance parameter τ is present. To incorporate this case into the framework of Theorem 11.2(a) we may view the objective function formally as a function on $T \times B$, where T can be chosen as an arbitrary subset of some Euclidean space with $\text{int}(T) \neq \emptyset$. We then also set $\hat{\tau}_n = \bar{\tau}_n = \tau_0$, an arbitrary element of $\text{int}(T)$.

from zero. Since the elements of Σ^{-1}, $\nabla_\sigma \text{vec}(\Sigma^{-1})$, and $\nabla_{\sigma\sigma}\text{vec}(\Sigma^{-1})$ are sums of products of the elements of Σ divided by (a power of) $\det(\Sigma)$, also Σ^{-1}, $\nabla_\sigma \text{vec}(\Sigma^{-1})$, and $\nabla_{\sigma\sigma}\text{vec}(\Sigma^{-1})$ are bounded continuous functions on S and hence on $\text{int}(S)$. Let c_0, c_1, c_2, and c_3, respectively, denote bounds for $|\Sigma(\sigma)|$, $|\Sigma^{-1}(\sigma)|$, $|\nabla_\sigma \text{vec}(\Sigma^{-1})|$ and $|\nabla_{\sigma\sigma}\text{vec}(\Sigma^{-1})|$ on S (and hence on $\text{int}(S)$). Equicontinuity of the components of $\nabla_{\alpha\alpha}q_t(z,\beta)$, $\nabla_{\sigma\alpha}q_t(z,\beta)$, $\nabla_{\alpha\sigma}q_t(z,\beta)$, $\nabla_{\sigma\sigma}q_t(z,\beta)$, and thus of $\nabla_{\beta\beta}q_t(z,\beta)$, on $Z\times\text{int}(B)$ now follows – in light of (14.4) – from Assumptions 14.1(f),(g), 14.10(b),(d), observing that sums of equicontinuous functions are equicontinuous and that products of equicontinuous functions are also equicontinuous, if the factors are bounded in t at each point (and hence in a suitably small neighborhood around each point). Next consider the following obvious inequalities:

$$
\begin{aligned}
|\nabla_{\alpha\alpha}q_t(z,\beta)| &\le |\nabla_{\alpha\alpha}[\ln|\det(\nabla_y f_t(z,\alpha))|]| + |\Sigma^{-1}|\,|\nabla_\alpha f_t(z,\alpha)|^2 \\
&\quad + \left|(f_t'(z,\alpha)\Sigma^{-1})\otimes I_{p_a}\right|\,|\nabla_{\alpha\alpha}f_t(z,\alpha)| \\
&\le |\nabla_{\alpha\alpha}[\ln|\det(\nabla_y f_t(z,\alpha))|]| + |\Sigma^{-1}|\,|\nabla_\alpha f_t(z,\alpha)|^2 \\
&\quad + |\Sigma^{-1}|\,|f_t(z,\alpha)|\,|\nabla_{\alpha\alpha}f_t(z,\alpha)| \\
&\le |\nabla_{\alpha\alpha}[\ln|\det(\nabla_y f_t(z,\alpha))|]| + c_1\,|\nabla_\alpha f_t(z,\alpha)|^2 \\
&\quad + c_1\,|f_t(z,\alpha)|\,|\nabla_{\alpha\alpha}f_t(z,\alpha)|,
\end{aligned}
$$

$$
\begin{aligned}
|\nabla_{\sigma\alpha}q_t(z,\beta)| &= |\nabla_{\alpha\sigma}q_t(z,\beta)| \\
&\le |f_t(z,\alpha)||\nabla_\alpha f_t(z,\alpha)|\,|\nabla_\sigma \text{vec}\,(\Sigma^{-1})| \\
&\le c_2\,|f_t(z,\alpha)|\,|\nabla_\alpha f_t(z,\alpha)|,
\end{aligned}
$$

$$
\begin{aligned}
|\nabla_{\sigma\sigma}q_t(z,\beta)| &\le (1/2)\,|\nabla_\sigma \text{vec}\,(\Sigma^{-1})|^2\,|\Sigma|^2 \\
&\quad + (1/2)p_e^{1/2}\left[|\Sigma| + |f_t(z,\alpha)|^2\right]|\nabla_{\sigma\sigma}\text{vec}\,(\Sigma^{-1})| \\
&\le (1/2)c_0^2c_2^2 + (1/2)p_e^{1/2}c_3\left[c_0 + |f_t(z,\alpha)|^2\right].
\end{aligned}
$$

In deriving the above inequalities we have used that $|A\otimes B| \le |A|\,|B|$. (This follows since $\lambda_{\max}(A'A\otimes B'B) \le \lambda_{\max}(A'A)\lambda_{\max}(B'B)$ holds.) Furthermore we have used that $|\text{vec}(A)| \le (\text{rank}(A'A))^{1/2}\,|A|$. Observing that

$$
(\sum_{i=1}^k |a_i|)^{1+\gamma} \le k^\gamma \sum_{i=1}^k |a_i|^{1+\gamma}
$$

holds for $\gamma \ge 0$ we obtain

$$
\sup_n n^{-1}\sum_{t=1}^n E\left[\sup_{\text{int}(B)} |\nabla_{\alpha\alpha}q_t(\mathbf{z}_t,\beta)|^{1+\gamma}\right] \tag{K.1a}
$$

$$
\begin{aligned}
\leq \; & 3^{\gamma}\left\{ \sup_n n^{-1} \sum_{t=1}^{n} E\left[\sup_{\mathrm{int}(A)} \left|\nabla_{\alpha\alpha}\left[\ln\left|\det\left(\nabla_y f_t(\mathbf{z}_t,\alpha)\right)\right|\right]\right|^{1+\gamma}\right]\right. \\
& + c_1^{1+\gamma} \sup_n n^{-1} \sum_{t=1}^{n} E\left[\sup_{\mathrm{int}(A)} \left|\nabla_{\alpha} f_t(\mathbf{z}_t,\alpha)\right|^{2+2\gamma}\right] \\
& \left. + c_1^{1+\gamma} \sup_n n^{-1} \sum_{t=1}^{n} E\left[\sup_{\mathrm{int}(A)} \left|f_t(\mathbf{z}_t,\alpha)\right|^{1+\gamma} \left|\nabla_{\alpha\alpha} f_t(\mathbf{z}_t,\alpha)\right|^{1+\gamma}\right]\right\},
\end{aligned}
$$

$$
\begin{aligned}
& \sup_n n^{-1} \sum_{t=1}^{n} E\left[\sup_{\mathrm{int}(B)} \left|\nabla_{\sigma\alpha} q_t(\mathbf{z}_t,\beta)\right|^{1+\gamma}\right] \qquad \text{(K.1b)} \\
\leq \; & c_2^{1+\gamma} \sup_n n^{-1} \sum_{t=1}^{n} E\left[\sup_{\mathrm{int}(A)} \left|f_t(\mathbf{z}_t,\alpha)\right|^{1+\gamma} \left|\nabla_{\alpha} f_t(\mathbf{z}_t,\alpha)\right|^{1+\gamma}\right],
\end{aligned}
$$

$$
\begin{aligned}
& \sup_n n^{-1} \sum_{t=1}^{n} E\left[\sup_{\mathrm{int}(B)} \left|\nabla_{\sigma\sigma} q_t(\mathbf{z}_t,\beta)\right|^{1+\gamma}\right] \qquad \text{(K.1c)} \\
\leq \; & (1/2)\left\{ \left(c_0^2 c_2^2 + p_e^{1/2} c_3 c_0\right)^{1+\gamma} \right. \\
& \left. + p_e^{(1+\gamma)/2} c_3^{1+\gamma} \sup_n n^{-1} \sum_{t=1}^{n} E\left[\sup_{\mathrm{int}(A)} \left|f_t(\mathbf{z}_t,\alpha)\right|^{2+2\gamma}\right]\right\}.
\end{aligned}
$$

Observe further that by the Cauchy-Schwarz inequality the r.h.s. of (K.1b) can be bounded from above by

$$
\begin{aligned}
c_2^{1+\gamma} & \left\{ \sup_n n^{-1} \sum_{t=1}^{n} E\left[\sup_{\mathrm{int}(A)} \left|f_t(\mathbf{z}_t,\alpha)\right|^{2+2\gamma}\right]\right. \\
& \left. \sup_n n^{-1} \sum_{t=1}^{n} E\left[\sup_{\mathrm{int}(A)} \left|\nabla_{\alpha} f_t(\mathbf{z}_t,\alpha)\right|^{2+2\gamma}\right]\right\}^{1/2}.
\end{aligned}
$$

Since we clearly can always find a $\gamma > 0$ such that Assumptions 14.1(c) and 14.11(a) hold simultaneously, Assumption 11.2 is seen to hold in light of these assumptions (and since the derivative w.r.t. τ is identically zero as q_t does not depend on τ).

Assumption 11.3(a) follows from the assumed martingale difference property of $\nabla_{\beta} q_t(\mathbf{z}_t, \beta_0)$, Assumption 11.3(b) holds trivially since q_t does not depend on τ, Assumption 11.3(c) coincides with Assumption 14.12(a). Assumption 11.3(d) follows from Assumption 14.12(b) observing that

$$\begin{aligned} & n^{-1}\sum_{t=1}^{n} E\left[\nabla_{\beta'} q_t(\mathbf{z}_t,\beta_0)\nabla_{\beta} q_t(\mathbf{z}_t,\beta_0)\right] \\ = \; & nE\left[\left(n^{-1}\sum_{t=1}^{n}\nabla_{\beta'} q_t(\mathbf{z}_t,\beta_0)\right)\left(n^{-1}\sum_{t=1}^{n}\nabla_{\beta} q_t(\mathbf{z}_t,\beta_0)\right)\right] \end{aligned}$$

in light of the maintained martingale difference property of $\nabla_\beta q_t(\mathbf{z}_t,\beta_0)$, provided the second moments exist, which will be demonstrated below. (This equality also shows that the formula for D_n given in the theorem coincides with the formula for D_n as given in Theorem 11.2.)

The martingale difference property postulated in Assumption 11.4 is, as already noted, one of the maintained assumptions of the theorem. From (14.4c) we obtain

$$\begin{aligned} & \left|\nabla_\alpha q_t(z,\beta)\right| \qquad \text{(K.2a)} \\ \leq \; & \left|\nabla_\alpha\left[\ln\left|\det\left(\nabla_y f_t(z,\alpha)\right)\right|\right]\right| + \left|\nabla_\alpha f_t(z,\alpha)\right|\left|\Sigma^{-1}\right|\left|f_t(z,\alpha)\right| \\ \leq \; & \left|\nabla_\alpha\left[\ln\left|\det\left(\nabla_y f_t(z,\alpha)\right)\right|\right]\right| + c_1\left|\nabla_\alpha f_t(z,\alpha)\right|\left|f_t(z,\alpha)\right| \end{aligned}$$

$$\begin{aligned} & \left|\nabla_\sigma q_t(z,\beta)\right| \qquad \text{(K.2b)} \\ \leq \; & (1/2)p_e^{1/2}\left|\nabla_\sigma \mathrm{vec}(\Sigma^{-1})\right|\left[|\Sigma| + \left|f_t(z,\alpha)\right|^2\right] \\ \leq \; & (1/2)p_e^{1/2}c_2\left[c_0 + \left|f_t(z,\alpha)\right|^2\right], \end{aligned}$$

and hence

$$\begin{aligned} & \sup_n n^{-1}\sum_{t=1}^{n} E\left|\nabla_\alpha q_t(\mathbf{z}_t,\beta_0)\right|^{2+\delta} \qquad \text{(K.3a)} \\ \leq \; & 2^{1+\delta}\left\{\sup_n n^{-1}\sum_{t=1}^{n} E\left|\nabla_\alpha\left[\ln\left|\det\left(\nabla_y f_t(\mathbf{z}_t,\alpha_0)\right)\right|\right]\right|^{2+\delta}\right. \\ & \left. + c_1^{2+\delta}\sup_n n^{-1}\sum_{t=1}^{n} E\left[\left|\nabla_\alpha f_t(\mathbf{z}_t,\alpha_0)\right|^{2+\delta}\left|\epsilon_t\right|^{2+\delta}\right]\right\}, \end{aligned}$$

$$\begin{aligned} & \sup_n n^{-1}\sum_{t=1}^{n} E\left|\nabla_\sigma q_t(\mathbf{z}_t,\beta_0)\right|^{2+\delta} \qquad \text{(K.3b)} \\ \leq \; & (1/2)\left(p_e^{1/2}c_2\right)^{2+\delta}\left\{c_0^{2+\delta} + \sup_n n^{-1}\sum_{t=1}^{n} E\left|\epsilon_t\right|^{4+2\delta}\right\}. \end{aligned}$$

Since clearly

$$\left|\nabla_\beta q_t(\mathbf{z}_t,\beta_0)\right|^{2+\delta} \leq 2^{1+\delta}\left[\left|\nabla_\alpha q_t(\mathbf{z}_t,\beta_0)\right|^{2+\delta} + \left|\nabla_\sigma q_t(\mathbf{z}_t,\beta_0)\right|^{2+\delta}\right]$$

holds, the moment condition in Assumption 11.4 follows from Assumptions 14.11(b),(d). Furthermore, from (14.4c) and Assumptions 14.1(f),(g), 14.10(b),(c),(d) it follows that $\nabla_\beta q_t(.,\beta_0)$ is equicontinuous on Z again observing that sums of equicontinuous functions are equicontinuous and that products of equicontinuous functions are also equicontinuous, if the factors are bounded in t. Thus Assumption 11.4 is satisfied. Part (a) of the theorem now follows from Theorem 11.2(a).

We prove part (b) of the theorem by verifying the assumptions of Theorem 13.1(a). To verify the assumptions of Theorem 13.1(a), it only remains to be shown that also Assumption 11.2* holds, since we have already shown above that Assumptions 11.1 - 11.4 hold. The conditions concerning $\nabla_{\beta\tau} q_t \equiv 0$ and $\nabla_{\beta\beta} q_t$ postulated in Assumption 11.2* are identical to those in Assumption 11.2. Hence only the verification of the conditions in Assumption 11.2* for $\nabla_{\beta'} q_t \nabla_\beta q_t$ remains. Analogously to the verification of Assumption 11.2 it is seen that – in light of (14.4) – the components of $\nabla_{\alpha'} q_t(z,\beta)\nabla_\alpha q_t(z,\beta)$, $\nabla_{\sigma'} q_t(z,\beta)\nabla_\alpha q_t(z,\beta)$, $\nabla_{\sigma'} q_t(z,\beta)\nabla_\sigma q_t(z,\beta)$, and thus of $\nabla_{\beta'} q_t(z,\beta)\nabla_\beta q_t(z,\beta)$, are equicontinuous on $Z\times\text{int}(B)$ because of Assumptions 14.1(f),(g), 14.10(b),(d), 14.14(a),(b).

Next observe that in light of (K.2) we have

$$\begin{aligned}
&\sup_n n^{-1}\sum_{t=1}^n E\left[\sup_{\text{int}(B)} |\nabla_\alpha q_t(\mathbf{z}_t,\beta)|^{2+2\gamma}\right] \\
\leq\; & 2^{1+2\gamma}\left\{\sup_n n^{-1}\sum_{t=1}^n E\left[\sup_{\text{int}(A)} |\nabla_\alpha [\ln|\det(\nabla_y f_t(\mathbf{z}_t,\alpha))|]|^{2+2\gamma}\right]\right. \\
&\left. + c_1^{2+2\gamma}\sup_n n^{-1}\sum_{t=1}^n E\left[\sup_{\text{int}(A)} |\nabla_\alpha f_t(\mathbf{z}_t,\alpha)|^{2+2\gamma}\,|f_t(\mathbf{z}_t,\alpha)|^{2+2\gamma}\right]\right\},
\end{aligned}$$

and

$$\begin{aligned}
&\sup_n n^{-1}\sum_{t=1}^n E\left[\sup_{\text{int}(B)} |\nabla_\sigma q_t(\mathbf{z}_t,\beta)|^{2+2\gamma}\right] \\
\leq\; & (1/2)p_e^{1+\gamma}c_2^{2+2\gamma}\left\{c_0^{2+2\gamma} + \sup_n n^{-1}\sum_{t=1}^n E\left[\sup_{\text{int}(A)} |f_t(\mathbf{z}_t,\alpha)|^{4+4\gamma}\right]\right\}.
\end{aligned}$$

Since

$$\begin{aligned}
&|\nabla_{\beta'} q_t(\mathbf{z}_t,\beta)\nabla_\beta q_t(\mathbf{z}_t,\beta)|^{1+\gamma} \\
\leq\; & |\nabla_\beta q_t(\mathbf{z}_t,\beta)|^{2+2\gamma} \\
\leq\; & 2^{1+2\gamma}\left[|\nabla_\alpha q_t(\mathbf{z}_t,\beta)|^{2+2\gamma} + |\nabla_\sigma q_t(\mathbf{z}_t,\beta)|^{2+2\gamma}\right]
\end{aligned}$$

the moment condition in Assumption 11.2* concerning the components of $\nabla_{\beta'} q_t \nabla_\beta q_t$ now follows from Assumption 14.14(c). This completes the proof of part (b) of the theorem. ∎

Proof of Theorem 14.8. We prove the theorem by reducing it to Theorem 14.7. From (K.2), Assumptions 14.1(c), 14.11(a),(c) and an application of the Cauchy-Schwarz inequality it follows that

$$E[\sup_{\text{int}(B)} |\nabla_\beta q_t(\mathbf{z}_t, \beta)|] < \infty$$

for all $t \geq 1$. Similarly, Assumptions 14.1(b),(c) imply

$$E[\sup_B |q_t(\mathbf{z}_t, \beta)|] < \infty$$

for all $t \geq 1$. Consequently,

$$\int \sup_{\text{int}(B)} |\nabla_\beta q_t(y_t, y_{t-1}, \ldots, y_{t-l}, x_t, \beta)| \, \pi_t^Y(y_t \mid y_{t-1}, \ldots, y_{t-l}, x_t; \beta_0) \, dy_t$$

$$= E\left[\sup_{\text{int}(B)} |\nabla_\beta q_t(\mathbf{z}_t, \beta)| \mid \mathbf{y}_{t-1} = y_{t-1}, \ldots, \mathbf{y}_{t-l} = y_{t-l}, \mathbf{x}_t = x_t\right] < \infty$$

and

$$\int \sup_B |q_t(y_t, y_{t-1}, \ldots, y_{t-l}, x_t, \beta)| \, \pi_t^Y(y_t \mid y_{t-1}, \ldots, y_{t-l}, x_t; \beta_0) \, dy_t$$

$$= E\left[\sup_B |q_t(\mathbf{z}_t, \beta)| \mid \mathbf{y}_{t-1} = y_{t-1}, \ldots, \mathbf{y}_{t-l} = y_{t-l}, \mathbf{x}_t = x_t\right] < \infty$$

for all $t \geq 1$ and all values $(y_{t-1}', \ldots, y_{t-l}', x_t')' \in \mathbf{R}^{lp_y + p_x}$, except possibly for values in a set N with $H_t(N) = 0$, where H_t is the distribution of $(\mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-l}', \mathbf{x}_t')'$. Hence, on the complement of N, we have for $\beta \in \text{int}(B)$

$$\begin{aligned}
& E\left[\nabla_\beta q_t(\mathbf{z}_t, \beta) \mid \mathbf{y}_{t-1} = y_{t-1}, \ldots, \mathbf{y}_{t-l} = y_{t-l}, \mathbf{x}_t = x_t\right] \\
= & \int \nabla_\beta q_t(y_t, y_{t-1}, \ldots, y_{t-l}, x_t, \beta) \, \pi_t^Y(y_t \mid y_{t-1}, \ldots, y_{t-l}, x_t; \beta_0) \, dy_t \\
= & \nabla_\beta \int q_t(y_t, y_{t-1}, \ldots, y_{t-l}, x_t, \beta) \, \pi_t^Y(y_t \mid y_{t-1}, \ldots, y_{t-l}, x_t; \beta_0) \, dy_t \\
= & \nabla_\beta E\left[q_t(\mathbf{z}_t, \beta) \mid \mathbf{y}_{t-1} = y_{t-1}, \ldots, \mathbf{y}_{t-l} = y_{t-l}, \mathbf{x}_t = x_t\right].
\end{aligned}$$

It follows furthermore from Lemma 14.5 that β_0 minimizes

$$E\left[q_t(\mathbf{z}_t, \beta) \mid \mathbf{y}_{t-1} = y_{t-1}, \ldots, \mathbf{y}_{t-l} = y_{t-l}, \mathbf{x}_t = x_t\right]$$

whenever $(y_{t-1}', \ldots, y_{t-l}', x_t')' \notin N$, since then

$$E\left[q_t(\mathbf{z}_t, \beta_0) \mid \mathbf{y}_{t-1} = y_{t-1}, \ldots, \mathbf{y}_{t-l} = y_{t-l}, \mathbf{x}_t = x_t\right] < \infty$$

holds. Since β_0 is an interior point of B in light of Assumption 14.9, it follows that the derivative of the conditional expectation is zero at $\beta = \beta_0$ and hence

$$E\left[\nabla_\beta q_t(\mathbf{z}_t, \beta_0) \mid \mathbf{y}_{t-1} = y_{t-1}, \dots, \mathbf{y}_{t-l} = y_{t-l}, \mathbf{x}_t = x_t\right] = 0$$

on the complement of N. Since

$$\pi_t^Y\left(y_t \mid y_{t-1}, \dots, y_{t-l}, x_t; \beta_0\right) = \pi_t^Y\left(y_t \mid y_{t-1}, \dots, y_{1-l}, x_t, \dots, x_1; \beta_0\right)$$

it follows further that, on the complement of N,

$$\begin{aligned} & E\left[\nabla_\beta q_t(\mathbf{z}_t, \beta_0) \mid \mathbf{y}_{t-1} = y_{t-1}, \dots, \mathbf{y}_{1-l} = y_{1-l}, \mathbf{x}_t = x_t, \dots, \mathbf{x}_1 = x_1\right] \\ = \; & E\left[\nabla_\beta q_t(\mathbf{z}_t, \beta_0) \mid \mathbf{y}_{t-1} = y_{t-1}, \dots, \mathbf{y}_{t-l} = y_{t-l}, \mathbf{x}_t = x_t\right] = 0. \end{aligned}$$

Consequently, $E[\nabla_\beta q_t(\mathbf{z}_t, \beta_0) \mid \mathfrak{F}_{t-1}] = 0$ a.s., which establishes the martingale difference property of $\nabla_\beta q_t(\mathbf{z}_t, \beta_0)$. Given the integrability of $\sup_B |q_t(\mathbf{z}_t, \beta)|$ established at the beginning of the proof, Lemma 14.5 also implies that β_0 minimizes $\bar{R}_n(\beta)$.

We next verify that Assumption 14.11(d) is satisfied. The second part of this assumption holds trivially in light of the normality of ϵ_t postulated in Assumption 14.8. Let $\gamma > 0$ be such that Assumption 14.11(a) holds. Choose $\delta > 0$ and $r > 1$ such that $(2+\delta)r \leq 2 + 2\gamma$, which is clearly possible. Then, using Hölder's inequality with $r^{-1} + s^{-1} = 1$ and Lyapunov's inequality we have

$$\begin{aligned} & \sup_n n^{-1} \sum_{t=1}^{n} E\left[\left|\nabla_\alpha f_t(\mathbf{z}_t, \alpha_0)\right|^{2+\delta} \left|\epsilon_t\right|^{2+\delta}\right] \\ \leq \; & \left\{\sup_n n^{-1} \sum_{t=1}^{n} E\left|\nabla_\alpha f_t(\mathbf{z}_t, \alpha_0)\right|^{(2+\delta)r}\right\}^{1/r} \left\{\sup_n n^{-1} \sum_{t=1}^{n} E\left|\epsilon_t\right|^{(2+\delta)s}\right\}^{1/s} \\ \leq \; & \left\{\sup_n n^{-1} \sum_{t=1}^{n} E\left|\nabla_\alpha f_t(\mathbf{z}_t, \alpha_0)\right|^{(2+2\gamma)}\right\}^{(2+\delta)/(2+2\gamma)} \left\{E\left|\epsilon_1\right|^{(2+\delta)s}\right\}^{1/s}. \end{aligned}$$

In view of Assumption 14.11(a) and since the disturbances ϵ_t are assumed to be i.i.d. normal the r.h.s. of the last inequality is finite. This shows that also the first part of Assumption 14.11(d) holds and thus completes the proof of the theorem. ■

Proof of Theorem 14.9. We first show that the equality

$$E\nabla_{\beta\beta} q_t(\mathbf{z}_t, \beta_0) = E\left[\nabla_{\beta'} q_t(\mathbf{z}_t, \beta_0) \nabla_\beta q_t(\mathbf{z}_t, \beta_0)\right]$$

holds, which implies in turn that $C_n = D_n^2$. Observe that

$$q_t(z_t, \beta) = -\ln\left[\pi_t^Y\left(y_t \mid y_{t-1}, \dots, y_{t-l}, x_t; \beta\right)\right] - (p_e/2)\ln(2\pi)$$

where $z_t = (y'_t, \ldots, y'_{t-l}, x'_t)'$, and hence

$$\begin{aligned}\nabla_{\beta\beta} q_t(z_t, \beta_0) &= \left(\pi_t^Y\right)^{-2} \nabla_{\beta'} \pi_t^Y \nabla_\beta \pi_t^Y - \left(\pi_t^Y\right)^{-1} \nabla_{\beta\beta} \pi_t^Y \\ &= \nabla_{\beta'} q_t(z_t, \beta_0) \nabla_\beta q_t(z_t, \beta_0) - (\pi_t^Y)^{-1} \nabla_{\beta\beta} \pi_t^Y,\end{aligned}$$

where the terms are evaluated at $\beta = \beta_0$. As was verified in the proof of Theorem 14.7, both $E\nabla_{\beta\beta} q_t(\mathbf{z}_t, \beta_0)$ and $E\nabla_{\beta'} q_t(\mathbf{z}_t, \beta_0) \nabla_\beta q_t(\mathbf{z}_t, \beta_0)$ exist and are finite.[4] Clearly then also the expected value of the expression $(\pi_t^Y)^{-1} \nabla_{\beta\beta} \pi_t^Y$ evaluated at $\mathbf{z}_t$ exists and is finite. Hence it remains to be shown that this expected value is zero. For this it clearly suffices to show that the integral of $(\pi_t^Y)^{-1} \nabla_{\beta\beta} \pi_t^Y$ w.r.t. $\pi_t^Y(y_t \mid y_{t-1}, \ldots, y_{t-l}, x_t; \beta_0)$ is zero for all values of $(y'_{t-1}, \ldots, y'_{t-l}, x'_t)'$ outside a set N with $H_t(N) = 0$, where H_t is the distribution of $(\mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-l}, \mathbf{x}'_t)'$. This integral clearly reduces to

$$\int \nabla_{\beta\beta} \pi_t^Y (y_t \mid y_{t-1}, \ldots, y_{t-l}, x_t; \beta_0) dy_t.$$

Since

$$\int \pi_t^Y (y_t \mid y_{t-1}, \ldots, y_{t-l}, x_t; \beta)\, dy_t = 1$$

for all β, we obtain the desired result if the operations of differentiation and integration can be interchanged. This interchange is permitted if

$$\int \sup_{\operatorname{int}(B)} \left| \nabla_\beta \pi_t^Y (y_t \mid y_{t-1}, \ldots, y_{t-l}, x_t; \beta) \right| dy_t < \infty$$

and

$$\int \sup_{\operatorname{int}(B)} \left| \nabla_{\beta\beta} \pi_t^Y (y_t \mid y_{t-1}, \ldots, y_{t-l}, x_t; \beta) \right| dy_t < \infty$$

hold. But these latter two conditions hold for all $(y'_{t-1}, \ldots, y'_{t-l}, x'_t)'$ outside a set N with $H_t(N) = 0$ in view of Assumption 14.13. To see this note that integrating the l.h.s. in the last two inequalities w.r.t. H_t gives precisely the l.h.s. of the expressions in Assumption 14.13. As a consequence, Assumption 14.12(b) now follows automatically from Assumption 14.12(a). The proof of part (a) of the theorem and the proof of $\hat{C}_n^{-1} \hat{\Phi}_n \hat{C}_n^{-1} - C_n^{-1} \to 0$ i.p. as $n \to \infty$, claimed in part (c) of the theorem, is now completed by appealing to Theorem 14.8. Now recall that the proof of $\hat{C}_n^{-1} \hat{\Phi}_n \hat{C}_n^{-1} - C_n^{-1} \to 0$ i.p. in Theorem 14.8 was based on the verification of the assumptions of Theorem 13.1(a) (via a verification of the assumptions of Theorem 14.7). Inspection of the proof of Theorem 13.1(a) shows that also $\hat{C}_n - C_n \to 0$ i.p. and $\hat{\Phi}_n - D_n^2 = \hat{\Phi}_n - C_n \to 0$ i.p. as $n \to \infty$. The proof for $\hat{C}_n - C_n \to 0$ i.p. in Theorem 13.1(a) only requires Assumption 11.2, but not the stronger

[4] We note that this step in the proof of Theorem 14.7 did not utilize Assumption 14.12.

Assumption 11.2*. Therefore the result $\hat{C}_n - C_n \to 0$ i.p. holds also without Assumption 14.14, whereas Assumption 11.2* and hence Assumption 14.14 is required for $\hat{\Phi}_n - C_n \to 0$ i.p. Since $|C_n|$ and $|C_n^{-1}|$ are bounded, as shown in part (a) of the theorem, $\hat{C}_n^{-1} - C_n^{-1} \to 0$ i.p. and $\hat{\Phi}_n^{-1} - C_n^{-1} \to 0$ i.p. as $n \to \infty$ follow now from Lemma F1. ■

Lemma K3. *Let M_n be a sequence of real symmetric nonnegative definite $m \times m$ matrices of the form*

$$M_n = \begin{bmatrix} \Gamma_n & \Delta_n \\ \Delta_n' & \Xi_n \end{bmatrix},$$

where Γ_n is of the dimension $m_* \times m_*$ and Ξ_n is of the dimension $(m - m_*) \times (m - m_*)$. Suppose $|\Delta_n| = O(1)$, $|\Xi_n| = O(1)$, and suppose

$$\liminf_{n\to\infty} \lambda_{\min}(\Xi_n) > 0$$

and

$$\liminf_{n\to\infty} \lambda_{\min}\left(\Gamma_n - \Delta_n \Xi_n^{-1} \Delta_n'\right) > 0,$$

then

$$\liminf_{n\to\infty} \lambda_{\min}(M_n) > 0.$$

Proof. Clearly, Ξ_n is positive definite for all n exceeding some n_0. In the following we only consider $n > n_0$. Now observe that M_n can be written as

$$M_n = P_n' \begin{bmatrix} \Gamma_n - \Delta_n \Xi_n^{-1} \Delta_n' & 0 \\ 0 & I \end{bmatrix} P_n \qquad \text{(K.4)}$$

with

$$P_n = \begin{bmatrix} I & 0 \\ \Xi_n^{-1/2} \Delta_n' & \Xi_n^{1/2} \end{bmatrix}.$$

Since M_n is nonnegative definite and since P_n is nonsingular, it follows that $\Gamma_n - \Delta_n \Xi_n^{-1} \Delta_n'$ is nonnegative definite. The smallest eigenvalue of the second matrix on the r.h.s. of (K.4) is clearly given by $\min\{1, \lambda_{\min}(\Gamma_n - \Delta_n \Xi_n^{-1} \Delta_n')\}$. Consequently

$$\lambda_{\min}(M_n) \geq \min\left\{1, \lambda_{\min}\left(\Gamma_n - \Delta_n \Xi_n^{-1} \Delta_n'\right)\right\} \lambda_{\min}\left(P_n' P_n\right).$$

Hence to complete the proof all that remains to be shown is that

$$\lambda_* = \liminf_{n\to\infty} \lambda_{\min}\left(P_n' P_n\right) > 0.$$

By assumption $|\Delta_n| = O(1)$, $|\Xi_n| = O(1)$. Furthermore,

$$\left|\Xi_n^{-1}\right| = (\lambda_{\min}(\Xi_n))^{-1} = O(1),$$

since the smallest eigenvalue of Ξ_n is assumed to be bounded away from zero. Hence

$$\left|\Xi_n^{1/2}\right| = [\lambda_{\max}(\Xi_n)]^{1/2} = |\Xi_n|^{1/2} = O(1)$$

and similarly

$$\left|\Xi_n^{-1/2}\right| = O(1).$$

Consequently also $|P_n| = O(1)$. Now let λ_n denote $\lambda_{\min}(P_n' P_n)$, and let v_n be a corresponding eigenvector of length one, i.e., $v_n' v_n = 1$. Clearly there exists a subsequence n_i of the natural numbers such that λ_{n_i} converges to λ_*. Furthermore, since v_n and P_n are bounded, there exists a subsequence $n_{i(j)}$ of n_i such that $v_{n_{i(j)}}$ converges to some element v_* with $v_*' v_* = 1$ and $P_{n_{i(j)}}$ converges to some matrix P_*. Note that P_* is necessarily of the form

$$P_* = \begin{bmatrix} I & 0 \\ P_{*21} & P_{*22} \end{bmatrix}.$$

Since P_{*22} is the limit of $\Xi_{n_{i(j)}}^{1/2}$ and since the smallest eigenvalues of Ξ_n are bounded away from zero it follows that P_{*22}, and hence P_*, are nonsingular. Consequently, $P_*' P_*$ is positive definite. Taking limits in the relation

$$P_{n_{i(j)}}' P_{n_{i(j)}} v_{n_{i(j)}} = \lambda_{n_{i(j)}} v_{n_{i(j)}}$$

we obtain

$$P_*' P_* v_* = \lambda_* v_*.$$

Since $v_* \neq 0$ and $P_*' P_*$ is positive definite, we have established that λ_* is positive. ■

Proof of Lemma 14.10. In the following let $\Sigma_0 = \Sigma(\sigma_0)$. As was shown in the proof of Theorem 14.9 we have

$$\begin{aligned} C_n &= n^{-1} \sum_{t=1}^{n} E \nabla_{\beta\beta} q_t(\mathbf{z}_t, \beta_0) \\ &= n^{-1} \sum_{t=1}^{n} E \left[\nabla_{\beta'} q_t(\mathbf{z}_t, \beta_0) \nabla_{\beta} q_t(\mathbf{z}_t, \beta_0) \right] = D_n^2. \end{aligned}$$

We emphasize that the above relationships have been established in the proof of Theorem 14.9 without the use of Assumption 14.12(a). We now partition C_n as

$$C_n = \begin{bmatrix} \Gamma_n & \Delta_n \\ \Delta_n' & \Xi_n \end{bmatrix}, \tag{K.5a}$$

with

$$\Gamma_n = n^{-1}\sum_{t=1}^{n} E\left[\nabla_{\alpha'} q_t(\mathbf{z}_t,\beta_0)\nabla_{\alpha} q_t(\mathbf{z}_t,\beta_0)\right], \tag{K.5b}$$

$$\begin{aligned}\Delta_n &= n^{-1}\sum_{t=1}^{n} E\left[\nabla_{\alpha'} q_t(\mathbf{z}_t,\beta_0)\nabla_{\sigma} q_t(\mathbf{z}_t,\beta_0)\right] \\ &= n^{-1}\sum_{t=1}^{n} E\nabla_{\alpha\sigma} q_t(\mathbf{z}_t,\beta_0)\end{aligned} \tag{K.5c}$$

$$\begin{aligned}\Xi_n &= n^{-1}\sum_{t=1}^{n} E\left[\nabla_{\sigma'} q_t(\mathbf{z}_t,\beta_0)\nabla_{\sigma} q_t(\mathbf{z}_t,\beta_0)\right] \\ &= n^{-1}\sum_{t=1}^{n} E\nabla_{\sigma\sigma} q_t(\mathbf{z}_t,\beta_0) \\ &= (1/2)\nabla_{\sigma'}\mathrm{vec}\left(\Sigma_0^{-1}\right)\left(\Sigma_0\otimes\Sigma_0\right)\nabla_{\sigma}\mathrm{vec}\left(\Sigma_0^{-1}\right),\end{aligned} \tag{K.5d}$$

where the last equality above follows from the expression for $\nabla_{\sigma\sigma} q_t$ given in (14.4c). Clearly $|\Xi_n| = O(1)$, since Ξ_n does not depend on n. Furthermore, $|\Delta_n| = O(1)$ in view of the expression for $\nabla_{\alpha\sigma} q_t$ given in (14.4c), the Cauchy-Schwarz inequality, and Assumptions 14.8(a) and 14.11(a). Additionally, observe that

$$\nabla_{\sigma}\mathrm{vec}\left(\Sigma_0^{-1}\right) = -\left(\Sigma_0^{-1}\otimes\Sigma_0^{-1}\right)\nabla_{\sigma}\mathrm{vec}\left(\Sigma_0\right),$$

and hence has full column rank since clearly $\nabla_{\sigma}\mathrm{vec}(\Sigma_0)$ has full column rank. Consequently, Ξ_n is nonsingular and, since it does not depend on n, its smallest eigenvalue is bounded away from zero. In view of Lemma K3 it follows that

$$\liminf_{n\to\infty} \lambda_{\min}(C_n) > 0,$$

if we can establish

$$\liminf_{n\to\infty} \lambda_{\min}\left(\Gamma_n - \Delta_n\Xi_n^{-1}\Delta_n'\right) > 0. \tag{K.6}$$

Thus, what remains to be shown in order to complete the proof is the validity of (K.6).

Observe that

$$\begin{aligned}\Gamma_n - \Delta_n\Xi_n^{-1}\Delta_n' = n^{-1}\sum_{t=1}^{n} E\Big\{&\left[\nabla_{\alpha'} q_t(\mathbf{z}_t,\beta_0) - \Delta_n\Xi_n^{-1}\nabla_{\sigma'} q_t(\mathbf{z}_t,\beta_0)\right] \\ &\left[\nabla_{\alpha'} q_t(\mathbf{z}_t,\beta_0) - \Delta_n\Xi_n^{-1}\nabla_{\sigma'} q_t(\mathbf{z}_t,\beta_0)\right]'\Big\}\end{aligned}$$

in view of (K.5). Now consider the following nonnegative definite matrix

$$\Omega_n = n^{-1}\sum_{t=1}^{n} E\Big\{\left[\nabla_{\alpha'} q_t(\mathbf{z}_t,\beta_0) - \Delta_n\Xi_n^{-1}\nabla_{\sigma'} q_t(\mathbf{z}_t,\beta_0)\right.$$

$$
\begin{aligned}
& - E\left(\nabla_{\alpha'} f_t(\mathbf{z}_t, \alpha_0) \mid \mathfrak{F}_{t-1}\right) \Sigma_0^{-1} \epsilon_t\big] \left[\nabla_{\alpha'} q_t(\mathbf{z}_t, \beta_0)\right. \\
& \left. - \Delta_n \Xi_n^{-1} \nabla_{\sigma'} q_t(\mathbf{z}_t, \beta_0) - E\left(\nabla_{\alpha'} f_t(\mathbf{z}_t, \alpha_0) \mid \mathfrak{F}_{t-1}\right) \Sigma_0^{-1} \epsilon_t\right]'\Big\} \\
= \; & \Gamma_n - \Delta_n \Xi_n^{-1} \Delta_n' - \Lambda_n - \Lambda_n' + \Psi_n,
\end{aligned}
$$

where

$$
\begin{aligned}
\Lambda_n \; = \; & n^{-1} \sum_{t=1}^{n} E\Big\{\left[\nabla_{\alpha'} q_t(\mathbf{z}_t, \beta_0) - \Delta_n \Xi_n^{-1} \nabla_{\sigma'} q_t(\mathbf{z}_t, \beta_0)\right] \\
& \left[E\left(\nabla_{\alpha'} f_t(\mathbf{z}_t, \alpha_0) \mid \mathfrak{F}_{t-1}\right) \Sigma_0^{-1} \epsilon_t\right]'\Big\},
\end{aligned}
$$

$$
\begin{aligned}
\Psi_n \; = \; & n^{-1} \sum_{t=1}^{n} E\Big\{\left[E\left(\nabla_{\alpha'} f_t(\mathbf{z}_t, \alpha_0) \mid \mathfrak{F}_{t-1}\right) \Sigma_0^{-1} \epsilon_t\right] \\
& \left[E\left(\nabla_{\alpha'} f_t(\mathbf{z}_t, \alpha_0) \mid \mathfrak{F}_{t-1}\right) \Sigma_0^{-1} \epsilon_t\right]'\Big\}.
\end{aligned}
$$

Note that the second moments of $E(\nabla_{\alpha'} f_t(\mathbf{z}_t, \alpha_0) \mid \mathfrak{F}_{t-1}) \Sigma_0^{-1} \epsilon_t$ exist in light of Assumptions 14.8(a) and 14.11(a). Hence Λ_n and Ψ_n are well defined. Since $E(\epsilon_t \epsilon_t' \mid \mathfrak{F}_{t-1}) = E(\epsilon_t \epsilon_t') = \Sigma_0$ in view of Assumption 14.8, it follows from the law of iterated expectations that

$$
\Psi_n = n^{-1} \sum_{t=1}^{n} E\left\{E\left(\nabla_{\alpha'} f_t(\mathbf{z}_t, \alpha_0) \mid \mathfrak{F}_{t-1}\right) \Sigma_0^{-1} E\left(\nabla_{\alpha} f_t(\mathbf{z}_t, \alpha_0) \mid \mathfrak{F}_{t-1}\right)\right\}.
$$

Observe that under the assumed persistent excitation condition (14.5) the smallest eigenvalue of Ψ_n is bounded away from zero, since $\lambda_{\min}(\Sigma_0^{-1}) > 0$. To complete the proof it therefore suffices to show that $\Lambda_n = \Psi_n$, since then $\Gamma_n - \Delta_n \Xi_n^{-1} \Delta_n' = \Omega_n + \Psi_n \geq \Psi_n$ will satisfy condition (K.6).

We now demonstrate that in fact $\Lambda_n = \Psi_n$. Using the expression for $\nabla_{\sigma'} q_t$ given in (14.4c), observe that

$$
\begin{aligned}
& n^{-1} \sum_{t=1}^{n} E\left\{\nabla_{\sigma'} q_t(\mathbf{z}_t, \beta_0) \left[E\left(\nabla_{\alpha'} f_t(\mathbf{z}_t, \alpha_0) \mid \mathfrak{F}_{t-1}\right) \Sigma_0^{-1} \epsilon_t\right]'\right\} \\
= \; & (1/2) \nabla_{\sigma'} \mathrm{vec}\left(\Sigma_0^{-1}\right) n^{-1} \sum_{t=1}^{n} E\{[-\mathrm{vec}(\Sigma_0) + \mathrm{vec}(\epsilon_t \epsilon_t')] \\
& \epsilon_t' \Sigma_0^{-1} E\left(\nabla_{\alpha} f_t(\mathbf{z}_t, \alpha_0) \mid \mathfrak{F}_{t-1}\right)\} \\
= \; & -(1/2) \nabla_{\sigma'} \mathrm{vec}\left(\Sigma_0^{-1}\right) \\
& n^{-1} \sum_{t=1}^{n} \left\{\mathrm{vec}(\Sigma_0) E\left[\epsilon_t' \Sigma_0^{-1} E\left(\nabla_{\alpha} f_t(\mathbf{z}_t, \alpha_0) \mid \mathfrak{F}_{t-1}\right)\right]\right. \\
& \left. - E\left[\mathrm{vec}(\epsilon_t \epsilon_t') \epsilon_t' \Sigma_0^{-1} E\left(\nabla_{\alpha} f_t(\mathbf{z}_t, \alpha_0) \mid \mathfrak{F}_{t-1}\right)\right]\right\} \\
= \; & 0.
\end{aligned}
$$

The last equality follows since both $E[E(\nabla_{\alpha'} f_t(\mathbf{z}_t, \alpha_0) \mid \mathfrak{F}_{t-1})\Sigma_0^{-1}\epsilon_t] = 0$ and $E(\epsilon_{ti}\epsilon_{tj}\epsilon_{tk} \mid \mathfrak{F}_{t-1}) = E(\epsilon_{ti}\epsilon_{tj}\epsilon_{tk}) = 0$ in view of Assumption 14.8. Hence Λ_n reduces to

$$\begin{aligned}\Lambda_n &= n^{-1}\sum_{t=1}^{n} E\left[\nabla_{\alpha'} q_t(\mathbf{z}_t, \beta_0)\epsilon_t'\Sigma_0^{-1} E\left(\nabla_\alpha f_t(\mathbf{z}_t, \alpha_0) \mid \mathfrak{F}_{t-1}\right)\right] \qquad (K.7)\\ &= n^{-1}\sum_{t=1}^{n} E\{\left[E\left(\nabla_{\alpha'} q_t(\mathbf{z}_t, \beta_0)\epsilon_t'\Sigma_0^{-1} \mid \mathfrak{F}_{t-1}\right)\right]\\ &\quad \left[E\left(\nabla_\alpha f_t(\mathbf{z}_t, \alpha_0) \mid \mathfrak{F}_{t-1}\right)\right]\}.\end{aligned}$$

Using the expression for $\nabla_{\alpha'} q_t$ given in (14.4c) we obtain

$$\begin{aligned}& E\left(\nabla_{\alpha'} q_t(\mathbf{z}_t, \beta_0)\epsilon_t'\Sigma_0^{-1} \mid \mathfrak{F}_{t-1}\right) \qquad (K.8)\\ =\; & -E\left\{\nabla_{\alpha'}\left[\ln\left|\det\left(\nabla_y f_t(\mathbf{z}_t, \alpha_0)\right)\right|\right]\epsilon_t'\Sigma_0^{-1} \mid \mathfrak{F}_{t-1}\right\}\\ & + E\left\{\nabla_{\alpha'} f_t(\mathbf{z}_t, \alpha_0)\Sigma_0^{-1}\epsilon_t\epsilon_t'\Sigma_0^{-1} \mid \mathfrak{F}_{t-1}\right\}.\end{aligned}$$

We now show that the second term on the r.h.s. of (K.8) can be expressed as

$$\begin{aligned}& E\left\{\nabla_{\alpha'} f_t(\mathbf{z}_t, \alpha_0)\Sigma_0^{-1}\epsilon_t\epsilon_t'\Sigma_0^{-1} \mid \mathfrak{F}_{t-1}\right\} \qquad (K.9)\\ =\; & E\left\{\nabla_{\alpha'}\left[\ln\left|\det\left(\nabla_y f_t(\mathbf{z}_t, \alpha_0)\right)\right|\right]\epsilon_t'\Sigma_0^{-1} \mid \mathfrak{F}_{t-1}\right\}\\ & + E\left(\nabla_{\alpha'} f_t(\mathbf{z}_t, \alpha_0) \mid \mathfrak{F}_{t-1}\right)\Sigma_0^{-1}.\end{aligned}$$

Clearly,

$$\begin{aligned}\nabla_{\alpha'} f_t(\mathbf{z}_t, \alpha_0)\Sigma_0^{-1}\epsilon_t\epsilon_t'\Sigma_0^{-1} &= \sum_{j=1}^{p_e} \nabla_{\alpha'} f_{tj}(\mathbf{z}_t, \alpha_0)\sigma_0^{j\cdot}\epsilon_t\epsilon_t'\Sigma_0^{-1} \qquad (K.10)\\ &= \sum_{j=1}^{p_e} \nabla_{\alpha'} f_{tj}(\mathbf{z}_t, \alpha_0)\epsilon_t'\Sigma_0^{-1}(\sigma_0^{j\cdot}\epsilon_t)\end{aligned}$$

where $f_{tj}(\mathbf{z}_t, \alpha_0)$ denotes the j-th element of $f_t(\mathbf{z}_t, \alpha_0)$. Here and in the following $a^{j\cdot}$, $a^{\cdot j}$, and a^{ij} stand, respectively, for the j-th row, j-th column and (i,j)-th element of the inverse, A^{-1}, of a matrix A. The (k,m)-th element of $\nabla_{\alpha'} f_{tj}(\mathbf{z}_t, \alpha_0)\epsilon_t'\Sigma_0^{-1}$ is then given by $f_{tjk}(\mathbf{z}_t, \alpha_0)\epsilon_t'\sigma_0^{\cdot m}$ where

$$f_{tjk}(\mathbf{z}_t, \alpha_0) = (\partial/\partial\alpha_k) f_{tj}(\mathbf{z}_t, \alpha_0).$$

Expressing $\mathbf{y}_t$ in terms of the reduced form we obtain

$$\begin{aligned}& f_{tjk}(\mathbf{z}_t, \alpha_0)\epsilon_t'\sigma_0^{\cdot m}\\ =\; & f_{tjk}\left(g_t(\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-l}, \mathbf{x}_t, \epsilon_t, \alpha_0), \mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-l}, \mathbf{x}_t, \alpha_0\right)\epsilon_t'\sigma_0^{\cdot m}.\end{aligned}$$

Now define for arbitrary $(y_{-1}', \ldots, y_{-l}', x')' \in \mathbf{R}^{lp_e + p_x}$ the functions $h_{j,km}(e)$ as

$$h_{j,km}(e) = f_{tjk}\left(g_t(y_{-1}, \ldots, y_{-l}, x, e, \alpha_0), y_{-1}, \ldots, y_{-l}, x, \alpha_0\right) e'\sigma_0^{\cdot m}.$$

In light of Assumption 14.8(c) the reduced form $g_t(y_{-1},\ldots,y_{-l},x,e,\alpha_0)$ is continuously partially differentiable w.r.t. e, since it is the inverse function of $f_t(y,y_{-1},\ldots,y_{-l},x,\alpha_0)$ (viewed as a function of y) with nonsingular derivative $\nabla_y f_t$, where the nonsingularity of $\nabla_y f_t$ has been assumed at the beginning of Chapter 14. Consequently, also

$$\begin{aligned}&\nabla_e g_t(y_{-1},\ldots,y_{-l},x,e,\alpha_0)\\ =\;&\left[\nabla_y f_t\left(g_t(y_{-1},\ldots,y_{-l},x,e,\alpha_0),y_{-1},\ldots,y_{-l},x,\alpha_0\right)\right]^{-1}.\end{aligned}$$

Using Assumption 14.15(a) we obtain for the derivative of $h_{j,km}$

$$\begin{aligned}&\nabla_e h_{j,km}(e) \qquad\qquad (K.11)\\ =\;&\left[\nabla_y f_{tjk}\left(g_t(y_{-1},\ldots,y_{-l},x,e,\alpha_0),y_{-1},\ldots,y_{-l},x,\alpha_0\right)\right]\\ &\left[\nabla_e g_t(y_{-1},\ldots,y_{-l},x,e,\alpha_0)\right]\left(e'\sigma_0^{\cdot m}\right)\\ &+f_{tjk}\left(g_t(y_{-1},\ldots,y_{-l},x,e,\alpha_0),y_{-1},\ldots,y_{-l},x,\alpha_0\right)\sigma_0^{m\cdot}\\ =\;&\left[\nabla_y f_{tjk}\left(g_t(y_{-1},\ldots,y_{-l},x,e,\alpha_0),y_{-1},\ldots,y_{-l},x,\alpha_0\right)\right]\\ &\left[\nabla_y f_t\left(g_t(y_{-1},\ldots,y_{-l},x,e,\alpha_0),y_{-1},\ldots,y_{-l},x,\alpha_0\right)\right]^{-1}\left(e'\sigma_0^{\cdot m}\right)\\ &+f_{tjk}\left(g_t(y_{-1},\ldots,y_{-l},x,e,\alpha_0),y_{-1},\ldots,y_{-l},x,\alpha_0\right)\sigma_0^{m\cdot}.\end{aligned}$$

We now verify that the conditions in the Lemma of Amemiya (1982) hold for $h_{j,km}$. Clearly, $\nabla_e h_{j,km}(e)$ is a continuous function of e in view of Assumption 14.15(a) and the assumed nonsingularity of $\nabla_y f_t(.,\alpha_0)$. This verifies condition (A) in Amemiya's Lemma. In view of Assumption 14.8 the vector of disturbances ϵ_t is independent of $(\mathbf{y}'_{t-1},\ldots,\mathbf{y}'_{t-l},\mathbf{x}'_t)'$. Hence Assumption 14.15(b) and Theorem 5.3.22 of Gänssler and Stute (1977) imply that

$$E\left\{\left|\left[\nabla_y f_{tjk}\left(g_t(y_{-1},\ldots,y_{-l},x,\epsilon_t,\alpha_0),y_{-1},\ldots,y_{-l},x,\alpha_0\right)\right]\right.\right.$$
$$\left.\left.\left[\nabla_y f_t\left(g_t(y_{-1},\ldots,y_{-l},x,\epsilon_t,\alpha_0),y_{-1},\ldots,y_{-l},x,\alpha_0\right)\right]^{-1}\right|^{1+\delta}\right\}<\infty$$

for H_t-almost all $(y'_{-1},\ldots,y'_{-l},x')'$. Here H_t denotes the distribution of $(\mathbf{y}'_{t-1},\ldots,\mathbf{y}'_{t-l},\mathbf{x}'_t)'$. In view of Hölder's inequality and observing that ϵ_t is normally distributed, and hence has moments of all orders, we thus obtain

$$E\left\{\left|\left[\nabla_y f_{tjk}\left(g_t(y_{-1},\ldots,y_{-l},x,\epsilon_t,\alpha_0),y_{-1},\ldots,y_{-l},x,\alpha_0\right)\right]\right.\right. \qquad (K.12)$$
$$\left.\left.\left[\nabla_y f_t\left(g_t(y_{-1},\ldots,y_{-l},x,\epsilon_t,\alpha_0),y_{-1},\ldots,y_{-l},x,\alpha_0\right)\right]^{-1}\left(\epsilon'_t\sigma_0^{\cdot m}\right)\right|\right\}<\infty.$$

Assumption 14.11(a) implies that

$$E\left|f_{tjk}\left(g_t(y_{-1},\ldots,y_{-l},x,\epsilon_t,\alpha_0),y_{-1},\ldots,y_{-l},x,\alpha_0\right)\sigma_0^{m\cdot}\right|<\infty \qquad (K.13)$$

for H_t-almost all $(y'_{-1},\ldots,y'_{-l},x')'$. Thus (K.12) and (K.13) imply that

$$E\left|\nabla_e h_{j,km}(\epsilon_t)\right|<\infty$$

for H_t-almost all $(y'_{-1}, \ldots, y'_{-l}, x')'$, and hence condition (B) in Amemiya's Lemma is satisfied. Analogous reasoning, using Assumption 14.11, the normality of ϵ_t and Hölder's inequality, shows that

$$E\left|h_{j,km}(\epsilon_t)\epsilon_t\right| < \infty$$

for H_t-almost all $(y'_{-1}, \ldots, y'_{-l}, x')'$. Hence also condition (C) in Amemiya's Lemma is satisfied.

Applying now Amemiya's Lemma to the functions $h_{j,km}$ and using (K.10) shows that for H_t-almost all $(y'_{-1}, \ldots, y'_{-l}, x')'$ the (k, m)-th element of

$$E\left\{\nabla_{\alpha'} f_t(\mathbf{z}_t, \alpha_0)\Sigma_0^{-1}\epsilon_t\epsilon_t'\Sigma_0^{-1} \mid (\mathbf{y}_{t-1}, \ldots, \mathbf{y}_{t-l}, \mathbf{x}_t) = (y_{-1}, \ldots, y_{-l}, x)\right\}$$

is given by

$$\sum_{j=1}^{p_e} E\left\{(\partial/\partial e_j)\, h_{j,km}(\epsilon_t)\right\}$$

where $(\partial/\partial e_j)h_{j,km}$ is the j-th element of $\nabla_e h_{j,km}$. Using the second equality in (K.11) we therefore obtain for the (k, m)-th element of

$$E\left\{\nabla_{\alpha'} f_t(\mathbf{z}_t, \alpha_0)\Sigma_0^{-1}\epsilon_t\epsilon_t'\Sigma_0^{-1} \mid \mathfrak{F}_{t-1}\right\}$$

the expression

$$\sum_{j=1}^{p_e} E\left\{\nabla_y f_{tjk}(\mathbf{z}_t, \alpha_0)\left[\nabla_y f_t(\mathbf{z}_t, \alpha_0)\right]^{\cdot j} \epsilon_t'\sigma_0^{\cdot m} + f_{tjk}(\mathbf{z}_t, \alpha_0)\sigma_0^{mj} \mid \mathfrak{F}_{t-1}\right\}. \tag{K.14}$$

The chain rule now shows that the k-th element of

$$\nabla_{\alpha'}\left[\ln\left|\det\left(\nabla_y f_t(\mathbf{z}_t, \alpha_0)\right)\right|\right]$$

equals

$$\sum_{j=1}^{p_e} \nabla_y f_{tjk}(\mathbf{z}_t, \alpha_0)\left[\nabla_y f_t(\mathbf{z}_t, \alpha_0)\right]^{\cdot j},$$

where we have also made use of the fact that we may interchange the order of differentiation w.r.t. α and y in view of Assumption 14.15(a). From this and (K.14) we obtain

$$\begin{aligned}
& E\left\{\nabla_{\alpha'} f_t(\mathbf{z}_t, \alpha_0)\Sigma_0^{-1}\epsilon_t\epsilon_t'\Sigma_0^{-1} \mid \mathfrak{F}_{t-1}\right\} \\
= \; & E\left\{\nabla_{\alpha'}\left[\ln\left|\det\left(\nabla_y f_t(\mathbf{z}_t, \alpha_0)\right)\right|\right]\epsilon_t'\Sigma_0^{-1} \mid \mathfrak{F}_{t-1}\right\} \\
& + \sum_{j=1}^{p_e} E\left\{\nabla_{\alpha'} f_{tj}(\mathbf{z}_t, \alpha_0)\sigma_0^{j\cdot} \mid \mathfrak{F}_{t-1}\right\} \\
= \; & E\left\{\nabla_{\alpha'}\left[\ln\left|\det\left(\nabla_y f_t(\mathbf{z}_t, \alpha_0)\right)\right|\right]\epsilon_t'\Sigma_0^{-1} \mid \mathfrak{F}_{t-1}\right\} \\
& + E\left\{\nabla_{\alpha'} f_t(\mathbf{z}_t, \alpha_0)\Sigma_0^{-1} \mid \mathfrak{F}_{t-1}\right\}
\end{aligned}$$

which shows that (K.9) holds indeed. From (K.7), (K.8) and (K.9) we then see immediately that $\Lambda_n = \Psi_n$, which completes the proof. ■

References

[1] Amemiya, T. Regression analysis when the dependent variable is truncated normal. *Econometrica* 41 (1973): 997-1016.

[2] Amemiya, T. The maximum likelihood and the nonlinear three-stage least squares estimator in the general nonlinear simultaneous equation model. *Econometrica* 45 (1977): 955-968.

[3] Amemiya, T. Correction to lemma. *Econometrica* 50 (1982): 1325-1328.

[4] Amemiya, T. Non-linear regression models. In Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics*, Vol. I, pp. 333-389. Amsterdam: North Holland, 1983.

[5] Amemiya, T. *Advanced Econometrics.* Cambridge, MA: Harvard University Press, 1985.

[6] Anderson, T.W. *The Statistical Analysis of Time Series.* New York: Wiley, 1971.

[7] Andrews, D.W.K. Non-strong mixing autoregressive processes. *Journal of Applied Probability* 21 (1984): 930-934.

[8] Andrews, D.W.K. Consistency in nonlinear econometric models: A generic uniform law of large numbers. *Econometrica* 55 (1987): 1465-1471.

[9] Andrews, D.W.K. Laws of large numbers for dependent non-identically distributed random variables. *Econometric Theory* 4 (1988): 458-467.

[10] Andrews, D.W.K. Asymptotics for semiparametric econometric models: I. Estimation. Cowles Foundation for Research in Economics Working Paper No. 908, 1989a.

[11] Andrews, D.W.K. Asymptotics for semiparametric econometric models: II. Stochastic equicontinuity. Cowles Foundation for Research in Economics Working Paper No. 909, 1989b.

[12] Andrews, D.W.K. Asymptotics for semiparametric econometric models: III. Testing and examples. Cowles Foundation for Research in Economics Working Paper No. 910, 1989c.

[13] Andrews, D.W.K. An empirical process central limit theorem for dependent non-identically distributed random variables. *Journal of Multivariate Analysis* 38 (1991a): 187-203.

[14] Andrews, D.W.K. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59 (1991b): 817-858.

[15] Andrews, D.W.K. Comment on "Basic structure of the asymptotic theory in dynamic nonlinear econometric models". *Econometric Reviews* 10 (1991c): 327-332.

[16] Andrews, D.W.K. Generic uniform convergence. *Econometric Theory* 8 (1992): 241-257.

[17] Andrews, D.W.K. An introduction to econometric applications to empirical process theory for dependent random variables. *Econometric Reviews* 12 (1993): 183-216.

[18] Andrews, D.W.K. & J.C. Monahan. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60 (1992): 953-966.

[19] Andrews, D.W.K. & D. Pollard. Stochastic equicontinuity with strong mixing random variables via bracketing. Cowles Foundation for Research in Economics. Unpublished manuscript, 1989.

[20] Andrews, D.W.K. & D. Pollard. An introduction to functional central limit theorems for dependent stochastic processes. *International Statistical Review* 62 (1994): 119-132.

[21] Arcones, M. & B. Yu. Central limit theorems for empirical and U-processes of stationary mixing sequences. *Journal of Theoretical Probability* 7 (1994): 47-71.

[22] Athreya, K.B. & S.G. Pantula. Mixing properties of Harris chains and AR processes. *Journal of Applied Probability* 23 (1986a).

[23] Athreya, K.B. & S.G. Pantula. A note on strong mixing of ARMA processes. *Statistics and Probability Letters* 4 (1986b): 187-190.

[24] Bahadur, R. Rates of convergence of estimates and test statistics. *Annals of Mathematical Statistics* 38 (1967): 303-324.

[25] Basawa, I.V. & H.L. Koul. Large-sample statistics based on quadratic dispersion. *International Statistical Review* 56 (1988): 199-219.

[26] Basawa, I.V. & D.J. Scott. *Asymptotic Optimal Inference for Nonergodic Models.* New York: Springer Verlag, 1983.

[27] Bates, C. Nonlinear parametric models with dependent observations. University of Rochester, Department of Economics PhD dissertation, 1984.

[28] Bates, C. & H. White. A unified theory of consistent estimation for parametric models. *Econometric Theory* 1 (1985): 151-178.

[29] Bates, C. & H. White. Efficient instrumental variables estimation of systems of implicit heterogeneous nonlinear dynamic equations with nonspherical errors. In W.A. Barnett, E.R. Berndt & H. White (eds.), *International Symposia in Economic Theory and Econometrics*, Vol.3. New York: Cambridge University Press, 1988.

[30] Beran, R. Fisher's Program. In S. Kotz (ed.), *Encyclopedia of Statistical Sciences*, Supplements Volume. New York: Wiley, 1996, forthcoming.

[31] Berk, K.N. Consistent autoregressive spectral estimates. *Annals of Statistics* 2 (1974): 489-502.

[32] Berk, R.H. Limiting behavior of posterior distributions when the model is incorrect. *Annals of Mathematical Statistics* 37 (1966): 51-58.

[33] Berk, R.H. Consistency a posteriori. *Annals of Mathematical Statistics* 41 (1970): 894-906.

[34] Bickel, P.J. One-step Huber estimates in the linear model. *Journal of the American Statistical Association* 70 (1975): 428-434.

[35] Bierens, H.J. *Robust Methods and Asymptotic Theory.* Lecture Notes in Economics and Mathematical Systems 192. Berlin: Springer Verlag, 1981.

[36] Bierens, H.J. A uniform weak law of large numbers under ϕ-mixing with applications to nonlinear least squares. *Statistica Neerlandica* 36 (1982a): 81-86.

[37] Bierens, H.J. Consistent model specification tests. *Journal of Econometrics* 20 (1982b): 105-134.

[38] Bierens, H.J. Uniform consistency of kernel estimators of a regression function under generalized conditions. *Journal of the American Statistical Association* 78 (1983): 699-707.

[39] Bierens, H.J. Model specification testing of time series regressions. *Journal of Econometrics* 26 (1984): 323-353.

[40] Bierens, H.J. ARMAX model specification testing, with an application to unemployment in the Netherlands. *Journal of Econometrics* 35 (1987): 161-190.

[41] Bierens, H.J. A consistent conditional moment test of functional form. *Econometrica* 58 (1990): 1443-1458.

[42] Bierens, H.J. *Topics in Advanced Econometrics: Estimation, Testing, and Specification of Cross-Section and Time Series Models.* Cambridge: Cambridge University Press, 1994.

[43] Billingsley, P. *Statistical Inference for Markov Processes.* Chicago: University of Chicago Press, 1961.

[44] Billingsley, P. *Convergence of Probability Measures.* New York: Wiley, 1968.

[45] Billingsley, P. *Probability and Measure.* New York: Wiley, 1979.

[46] Blum, J.R. & M. Rosenblatt. A class of stationary processes and a central limit theorem. *Duke Mathematical Journal* 24 (1957): 73-78.

[47] Boente, G. & R. Fraiman. On the asymptotic behavior of general maximum likelihood estimates for the nonregular case under nonstandard conditions. *Biometrika* 75 (1988): 45-56.

[48] Brockwell, P.J. & R.A. Davis. *Time Series: Theory and Methods.* New York: Springer Verlag, 1991.

[49] Brown, B.M. Martingale central limit theorems. *Annals of Mathematical Statistics* 42 (1971): 59-66.

[50] Brown, L.D. & R. Purves. Measurable selections of extrema. *Annals of Statistics* 1 (1973): 902-912.

[51] Bunke, H. & W.H. Schmidt. Asymptotic results on nonlinear approximation of regression functions and weighted least squares. *Mathematische Operationsforschung und Statistik, Serie Statistik* 11 (1980): 3-22.

[52] Burguete, J.F., A.R. Gallant & G. Souza. On unification of the asymptotic theory of nonlinear econometric models. *Econometric Reviews* 1 (1982): 151-190.

[53] Caines, P.E. Prediction error identification methods for stationary stochastic processes. *IEEE Transactions on Automatic Control* AC-21 (1976): 500-506.

[54] Caines, P.E. Stationary linear and non-linear system identification and predictor set completeness. *IEEE Transactions on Automatic Control* AC-23 (1978): 583-595.

[55] Caines, P.E. & L. Ljung. Asymptotic normality and accuracy of prediction error estimators. University of Toronto, Department of Electrical Engineering Research Report 7602, 1976.

[56] Calzolari, G. & L. Panattoni. Alternative estimators of FIML covariance matrix: A Monte Carlo study. *Econometrica* 56 (1988): 701-714.

[57] Chamberlain, G. Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics* 34 (1987): 305-334.

[58] Chanda, K.C. Strong mixing properties of linear stochastic processes. *Journal of Applied Probability* 11 (1974): 401-408.

[59] Chernick, M.R. A limit theorem for the maximum of autoregressive processes with uniform marginal distributions. *Annals of Probability* 9 (1981): 145-149.

[60] Cox, D.R. Partial likelihood. *Biometrika* 62 (1975): 269-276.

[61] Cramér, H. *Mathematical Methods of Statistics.* Princeton: Princeton University Press, 1946.

[62] Crowder, M.J. Maximum likelihood estimation for dependent observations. *Journal of the Royal Statistical Society,* Series B, 38 (1976): 45-53.

[63] Dahlhaus, R. & B.M. Pötscher. Convergence results for maximum likelihood type estimators in multivariable ARMA models II. *Journal of Multivariate Analysis* 30 (1989): 241-244.

[64] Daniels, H.E. The asymptotic efficiency of a maximum likelihood estimator. In *Proceedings of the Fourth Berkeley Symposium in Mathematical Statistics and Probability,* pp. 151-163 . Berkeley: University of California Press, 1961.

[65] Davidson, J. & R.M. De Jong. Strong laws of large numbers for dependent heterogeneous processes: A synthesis of new results. Unpublished manuscript, 1995.

[66] Deistler, M., W. Dunsmuir & E.J. Hannan. Vector linear time series models: Corrections and extensions. *Advances in Applied Probability* 10 (1978): 360-372.

[67] Deistler, M. & B.M. Pötscher. The behaviour of the likelihood function for ARMA models. *Advances in Applied Probability* 16 (1984): 843-865.

[68] De Jong, R.M. Laws of large numbers for dependent heterogeneous processes. *Econometric Theory* 11 (1995a): 347-358.

[69] De Jong, R.M. Central limit theorems for dependent heterogeneous random variables. Unpublished manuscript, 1995b.

[70] Den Haan, W.J. & A. Levin. Inferences from parametric and non-parametric covariance matrix estimation procedures. National Bureau of Economic Research, Technical Working Paper 195, 1996.

[71] Dieudonné, J. *Foundations of Modern Analysis.* New York: Academic Press, 1960.

[72] Domowitz, I. New directions in non-linear estimation with dependent observations. *Canadian Journal of Economics* 19 (1985): 1-19.

[73] Domowitz, I. & H. White. Misspecified models with dependent observations. *Journal of Econometrics* 20 (1982): 35-58.

[74] Donoho, D.L. & R.C. Liu. Pathologies of some minimum distance estimators. *Annals of Statistics* 16 (1988): 587-608.

[75] Doob, J.L. Probability and statistics. *Transactions of the American Mathematical Society* 36 (1934): 759-775.

[76] Doukhan, P. *Mixing: Properties and Examples.* Lecture Notes in Statistics 85. New York: Springer Verlag, 1994.

[77] Doukhan, P., P. Massart & E. Rio. The functional central limit theorem for strongly mixing processes. *Annales de l' Institut Henri Poincaré* 30 (1994): 63-82.

[78] Doukhan, P., P. Massart & E. Rio. Invariance principles for absolutely regular empirical processes. *Annales de l' Institut Henri Poincaré* 31 (1995): 393-427.

[79] Dunsmuir, W. & E.J. Hannan. Vector linear time series models. *Advances in Applied Probability* 8 (1976): 339-364.

[80] Dupačová, J. & R. Wets. Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *Annals of Statistics* 16 (1988): 1517-1549.

[81] Eberlein, E. & M.S. Taqqu (eds.). *Dependence in Probability and Statistics.* Basel: Birkhäuser, 1986.

[82] Epstein, L.G. & M. Denny. Endogenous capital utilization in a short-run production model: Theory and empirical application. *Journal of Econometrics* 12 (1980): 189-207.

[83] Foutz, R.V. & R.C. Srivastava. The performance of the likelihood ratio test when the model is incorrect. *Annals of Statistics* 5 (1977): 1183-1194.

[84] Foutz, R.V. & R.C. Srivastava. Statistical inference for Markov processes when the model is incorrect. *Advances in Applied Probability* 11 (1979): 737-749.

[85] Freedman, D.A. & P. Diaconis. On inconsistent M-estimators. *Annals of Statistics* 10 (1982): 454-461.

[86] Gallant, A.R. Three stage least squares estimation for a system of simultaneous nonlinear implicit equations. *Journal of Econometrics* 5 (1977): 71-88.

[87] Gallant, A.R. *Nonlinear Statistical Models.* New York: Wiley, 1987a.

[88] Gallant, A.R. Identification and consistency in seminonparametric regression. In T.F. Bewley (ed.), *Advances in Econometrics, Fifth World Congress*, Vol. 1, pp. 145-170. New York: Cambridge University Press, 1987b.

[89] Gallant, A.R. and A. Holly. Statistical inference in an implicit, nonlinear, simultaneous equation model in the context of maximum likelihood estimation. *Econometrica* 48 (1980): 697-720.

[90] Gallant, A.R. & H. White. *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models.* New York: Basil Blackwell, 1988.

[91] Gänssler, P. *Empirical Processes.* Hayward, California: IMS Lecture Notes-Monograph Series, Vol. 3, 1983.

[92] Gänssler, P. & W. Stute. *Wahrscheinlichkeitstheorie.* New York: Springer Verlag, 1977.

[93] Ghosh, J.K. Efficiency of estimates – Part I. *Sankhya*, Series A, 47 (1985): 310-325.

[94] Gorodetskii, V.V. On the strong-mixing property for linear sequences. *Theory of Probability and its Applications* 22 (1977): 411-413.

[95] Guégan, D. *Séries Chronologiques Nonlineaires à Temps Discret.* Paris: Economica, 1994.

[96] Haberman, S. Concavity and Estimation. *Annals of Statistics* 17 (1989): 1631-1661.

[97] Hall, P. & C.C. Heyde. *Martingale Limit Theory and its Application.* New York: Academic Press, 1980.

[98] Hannan, E.J. Nonlinear time series regression. *Journal of Applied Probability* 8 (1971): 767-780.

[99] Hannan, E.J. The asymptotic theory of linear time series models. *Journal of Applied Probability* 10 (1973): 130-145.

[100] Hannan, E.J. & M. Deistler. *The Statistical Theory of Linear Systems.* New York: Wiley, 1988.

[101] Hannan, E.J. & L. Kavalieris. Regression, autoregression models. *Journal of Time Series Analysis* 7 (1986): 27-49.

[102] Hansen, B.E. Strong laws for dependent heterogeneous processes. *Econometric Theory* 7 (1991): 213-221. Erratum. *Econometric Theory* 8 (1992): 421-422.

[103] Hansen, B.E. Consistent covariance matrix estimation for dependent heterogeneous processes. *Econometrica* 60 (1992): 967-972.

[104] Hansen, B.E. Stochastic equicontinuity for unbounded dependent heterogeneous arrays. *Econometric Theory* 12 (1996): 347-359.

[105] Hansen, L.P. Large sample properties of generalized method of moments estimators. *Econometrica* 50 (1982): 1029-1054.

[106] Hansen, L.P. A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators. *Journal of Econometrics* 30 (1985): 203-238.

[107] Hansen, L.P. Semi-parametric efficiency bounds for linear time series models. Unpublished manuscript, 1988.

[108] Hansen, L.P., J.C. Heaton & M. Ogaki. Efficiency bounds implied by multiperiod conditional moment restrictions. *Journal of the American Statistical Association* 83 (1988): 863-871.

[109] Hansen, L.P. & K.J. Singleton. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica* 50 (1982): 1269-1286.

[110] Hansen, L.P. & K.J. Singleton. Computing semi-parametric efficiency bounds for linear time series models. In W.A. Barnett, J. Powell & G.E. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, pp. 387-412. New York: Cambridge University Press, 1991.

[111] Heijmans, R.D.H. & J.R. Magnus. Consistent maximum likelihood estimation with dependent observations. *Journal of Econometrics* 32 (1986a): 253-285.

[112] Heijmans, R.D.H. & J.R. Magnus. On the first order efficiency and asymptotic normality of maximum likelihood estimators obtained from dependent observations. *Statistica Neerlandica* 40 (1986b).

[113] Heijmans, R.D.H. & J.R. Magnus. Corrigenda. *Journal of Econometrics* 35 (1987): 395.

[114] Herrndorf, N. The invariance principle for ϕ-mixing sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 63 (1983): 97-108.

[115] Herrndorf, N. A functional central limit theorem for weakly dependent sequences of random variables. *Annals of Probability* 12 (1984): 141-153.

[116] Hoadley, B. Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. *Annals of Mathematical Statistics* 42 (1971): 1977-1991.

[117] Horn, R.A. & C.R. Johnson. *Matrix Analysis*. Cambridge: Cambridge University Press, 1985.

[118] Huber, P. The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*, Vol. 1, pp. 221-233. Berkeley: University of California Press, 1967.

[119] Huber, P. *Robust Statistics*. New York: Wiley, 1981.

[120] Humak, K.M.S. *Statistische Methoden der Modellbildung II*. Berlin: Akademie Verlag, 1983.

[121] Ibragimov, I.A. Some limit theorems for stationary processes. *Theory of Probability and its Applications* 7 (1962): 349-382.

[122] Ibragimov, I.A. & Y.V. Linnik. *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters, Noordhoff Publishing, 1971.

[123] Jeganathan, P. Some aspects of asymptotic theory with applications to time series. *Econometric Theory* 11 (1995): 818-887.

[124] Jennrich, R.I. Asymptotic properties of nonlinear least squares estimators. *Annals of Mathematical Statistics* 40 (1969): 633-643.

[125] Kabaila, P. Parameter values of ARMA models minimizing the one step ahead prediction error when the true system is not in the model set. *Journal of Applied Probability* 20 (1983): 405-408.

[126] Kabaila, P. & G.C. Goodwin. On the estimation of the parameters of an optimal interpolator when the class of interpolators is restricted. *SIAM Journal of Control and Optimisation* 18 (1980): 121-144.

[127] Keener, R.W., J. Kmenta & N.C. Weber. Estimation of the covariance matrix of the least-squares regression coefficients when the disturbance covariance matrix is of unknown form. *Econometric Theory* 7 (1991): 22-45.

[128] Klimko, L.A. & P.I. Nelson. On conditional least squares estimation for stochastic processes. *Annals of Statistics* 6 (1978): 629-642.

[129] Kuan, C.-M. & H. White. Adaptive learning with nonlinear dynamics driven by dependent processes. *Econometrica* 62 (1994): 1087-1114.

[130] Kumar, K. Some recent developments in non-linear time series modelling. Monash University, Department of Econometrics, 1988.

[131] Lai, T.L. Asymptotic properties of nonlinear least squares estimates in stochastic regression models. *Annals of Statistics* 22 (1994): 1917-1930.

[132] Läuter, H. Note on the strong consistency of the least squares estimator in nonlinear regression. In O. Bunke & E. Jolivet (eds.), *Seminarbericht* No. 89, pp. 153-175. Berlin: Humboldt Universität, 1987.

[133] LeCam, L. On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *University of California Publications in Statistics*, Vol. 1, pp. 277-328. University of California Press, 1953.

[134] LeCam, L. Locally asymptotically normal families of distributions. *University of California Publications in Statistics*, Vol. 3, pp. 27-98. University of California Press, 1960.

[135] Lee, C.C. & P.C.B. Phillips. An ARMA prewhitened long-run variance estimator. Yale University, Department of Economics. Unpublished manuscript, 1994.

[136] Levine, D. A remark on serial correlation in maximum likelihood. *Journal of Econometrics* 23 (1983): 337-342.

[137] Liese, F. & I. Vajda. Necessary and sufficient conditions for consistency of generalized M-estimators. *Metrika* 42 (1995): 291-324.

[138] Ljung, L. On consistency and identifiability. In *Mathematical Programming Study*, No. 5, pp. 169-190. Amsterdam: North Holland, 1976a.

[139] Ljung, L. On the consistency of prediction error identification methods. In R.K. Mehra & D.G. Lainiotis (eds.), *System Identification, Advances and Case Studies*, pp. 121-164. New York: Academic Press, 1976b.

[140] Ljung, L. Convergence analysis of parametric identification methods. *IEEE Transactions on Automatic Control* AC-23 (1978): 770-783.

[141] Ljung, L. & P.E. Caines. Asymptotic normality of prediction error estimators for approximate models. *Stochastics* 3 (1979): 29-46.

[142] MacDuffee, C.C. *The Theory of Matrices.* New York: Chelsea, 1956.

[143] Malinvaud, E. The consistency of nonlinear regressions. *Annals of Mathematical Statistics* 41 (1970): 956-969.

[144] McLeish, D.L. Dependent central limit theorems and invariance principles. *Annals of Probability* 2 (1974): 620-628.

[145] McLeish, D.L. A maximal inequality and dependent strong laws. *Annals of Probability* 3 (1975a): 829-839.

[146] McLeish, D.L. Invariance principles for dependent variables. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 32 (1975b): 165-178.

[147] McLeish, D.L. On the invariance principle for nonstationary mixingales. *Annals of Probability* 5 (1977): 616-621.

[148] Mokkadem, A. Sur le mélange d'un processus ARMA vectoriel. *Comptes Rendus de l'Académie Sciences* 303, Série 1 (1986): 519-521.

[149] Mokkadem, A. Conditions suffisantes de mélange géométrique pour des autorégressifs polynomiaux; application aux processus ARMA et aux processus bilinéaires. *Comptes Rendus de l'Académie Sciences* 305, Série 1 (1987): 477-480.

[150] Newey, W.K. Generalized method of moments specification testing. *Journal of Econometrics* 29 (1985a): 229-256.

[151] Newey, W.K. Maximum likelihood specification testing and conditional moment tests. *Econometrica* 53 (1985b): 1047-1070.

[152] Newey, W.K. Expository notes on uniform laws of large numbers. Department of Economics, Princeton University, 1987.

[153] Newey, W.K. Asymptotic equivalence of closest moments and GMM estimators. *Econometric Theory* 4 (1988): 336-340.

[154] Newey, W.K. Semi-parametric efficiency bounds. *Journal of Applied Econometrics* 5 (1990): 99-135.

[155] Newey, W.K. Uniform convergence in probability and stochastic equicontinuity. *Econometrica* 59 (1991): 1161-1167.

[156] Newey, W.K. & D.L. McFadden. Large sample estimation and hypothesis testing. In R.F. Engle and D.L. McFadden (eds.), *Handbook of Econometrics*, Vol. IV, pp. 2113-2245. Amsterdam: Elsevier Science B.V., 1994.

[157] Newey, W.K. & K.D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55 (1987): 703-708.

[158] Newey, W.K. & K.D. West. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61 (1994): 631-653.

[159] Niemiro, W. Asymptotics for M-estimators defined by convex minimization. *Annals of Statistics* 20 (1992): 1514-1533.

[160] Norden, R.H. A survey of maximum likelihood estimation, part I. *International Statistical Review* 40 (1972): 329-354.

[161] Norden, R.H. A survey of maximum likelihood estimation, part II. *International Statistical Review* 41 (1973): 39-58.

[162] Pakes, A. & D. Pollard. Simulation and the asymptotics of optimization estimators. *Econometrica* 57 (1989): 1027-1057.

[163] Park, J.Y. & P.C.B. Phillips. Statistical inference in regressions with integrated processes: part 1. *Econometric Theory* 4 (1988): 468-498.

[164] Park, J.Y. & P.C.B. Phillips. Statistical inference in regressions with integrated processes: part 2. *Econometric Theory* 5 (1989): 95-131.

[165] Parthasarathy, K.R. *Probability Measures on Metric Spaces*. New York: Academic Press, 1967.

[166] Parzen, E. On consistent estimates of the spectrum of a stationary time series. *Annals of Mathematical Statistics* 28 (1957): 329-348.

[167] Parzen, E. Autoregressive spectral estimation. In D.R. Brillinger and P.R. Krishnaiah (eds.), *Handbook of Statistics*, Vol. 3, pp. 221-247. New York: North Holland, 1983.

[168] Peligrad, M. An invariance principle for dependent random variables. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 57 (1981): 495-507.

[169] Perlman, M.D. On the strong consistency of approximate maximum likelihood estimators. In *Proceedings of the Sixth Berkeley Symposium in Mathematical Statistics and Probability*, Vol. 1, pp. 263-281. Berkeley: University of California Press, 1972.

[170] Pfanzagl, J. On the measurability and consistency of minimum contrast estimators. *Metrika* 14 (1969): 249-272.

[171] Pham, D.T. & L.T. Tran. Some mixing properties of time series. *Stochastic Processes and their Applications* 19 (1985): 297-304.

[172] Phillips, P.C.B. On the consistency of the nonlinear FIML. *Econometrica* 50 (1982): 1307-1324.

[173] Phillips, P.C.B. Partially identified econometric models. *Econometric Theory* 5 (1989): 181-240.

[174] Ploberger, W. Slight misspecification of linear systems. In G. Feichtinger & P. Kall (eds.), *Operations Research in Progress*, pp. 413-424. Dordrecht: D. Reidel Publishing Company, 1982a.

[175] Ploberger, W. On the prediction error estimation of linear rational models I. University of Technology Vienna, Department of Econometrics and Operations Research. Research Report No.13, 1982b.

[176] Pollard, D. *Convergence of Stochastic Processes.* New York: Springer Verlag, 1984.

[177] Pollard, D. New ways to prove central limit theorems. *Econometric Theory* 1 (1985): 295-314.

[178] Pollard, D. *Empirical Processes: Theory and Applications.* NSF-CBMS Regional Conference Series, Vol. 2, 1990.

[179] Pollard, D. Bracketing methods in statistics and econometrics. In W.A. Barnett, J. Powell & G.E. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, pp. 337-355. New York: Cambridge University Press, 1991.

[180] Pötscher, B.M. Convergence results for maximum likelihood type estimators in multivariable ARMA models. *Journal of Multivariate Analysis* 21 (1987): 29-52.

[181] Pötscher, B.M. Noninvertibility and pseudo maximum likelihood estimation of misspecified ARMA models. *Econometric Theory* 7 (1991): 435-449. Corrigendum. *Econometric Theory* 10 (1994): 811.

[182] Pötscher, B.M. & I.R. Prucha. A class of partially adaptive one-step M-estimators for the nonlinear regression model with dependent observations. *Journal of Econometrics* 32 (1986a): 219-251.

[183] Pötscher, B.M. & I.R. Prucha. Consistency in nonlinear econometrics: A generic uniform law of large numbers and some comments on recent results. University of Maryland, Department of Economics. Working Paper 86-9, 1986b.

[184] Pötscher, B.M. & I.R. Prucha. A uniform law of large numbers for dependent and heterogeneous data processes. *Econometrica* 57 (1989): 675-683.

[185] Pötscher, B.M. & I.R. Prucha. Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part I: Consistency and approximation concepts. *Econometric Reviews* 10 (1991a): 125-216.

[186] Pötscher, B.M. & I.R. Prucha. Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part II: Asymptotic normality. *Econometric Reviews* 10 (1991b): 253-325.

[187] Pötscher, B.M. & I.R. Prucha. Generic uniform convergence and equicontinuity concepts for random functions: An exploration of the basic structure. *Journal of Econometrics* 60 (1994a): 23-63.

[188] Pötscher, B.M. & I.R. Prucha. On the formulation of uniform laws of large numbers: A truncation approach. *Statistics* 25 (1994b): 343-360.

[189] Powell, J.L. Estimation of semiparametric models. In R.F. Engle and D.L. McFadden (eds.), *Handbook of Econometrics*, Vol. IV, pp. 2444-2521. Amsterdam: Elsevier Science B.V., 1994.

[190] Press, H. & J.W. Tukey. *Power spectral methods of analysis and their application to problems in airplane dynamics.* Bell Systems Monograph No. 2606, 1956.

[191] Priestley, M.B. State-dependent models: A general approach to time series analysis. *Journal of Time Series Analysis* 1 (1980): 47-71.

[192] Priestley, M.B. *Spectral Analysis and Time Series*, Vol. 1. London: Academic Press, 1981.

[193] Priestley, M.B. *Non-linear and Non-stationary Time Series Analysis.* London: Academic Press, 1988.

[194] Prucha, I.R. On the estimation of the variance-covariance matrix of maximum likelihood estimators in nonlinear simultaneous equation systems: A Monte Carlo Study. University of Maryland, Department of Economics. Working Paper 84-14, 1984.

[195] Prucha, I.R. & H.H. Kelejian. The structure of simultaneous equation estimators: A generalization towards nonnormal disturbances. *Econometrica* 52 (1984): 721-736.

[196] Prucha, I.R. & M.I. Nadiri. On the computation of estimators in systems with implicitly defined variables. *Economics Letters* 26 (1988): 141-145.

[197] Prucha, I.R. & M.I. Nadiri. Endogenous capital utilization and productivity measurement in dynamic factor demand models: Theory and an application to the U.S. electrical machinery industry. *Journal of Econometrics* 71 (1996): 343-379.

[198] Quah, D. An improved rate for non-negative definite consistent covariance matrix estimation with heterogeneous dependent data. *Economics Letters* 33 (1990): 133-140.

[199] Ranga Rao, R. Relations between weak and uniform convergence of measures with applications. *Annals of Mathematical Statistics* 33 (1962): 659-680.

[200] Robinson, P.M. Non-linear regression for multiple time series. *Journal of Applied Probability* 9 (1972): 758-768.

[201] Robinson, P.M. Semiparametric econometrics: A survey. *Journal of Applied Econometrics* 3 (1988): 35-51.

[202] Roussas, G.G. Extension to Markov processes of results by A. Wald about the consistency of the maximum likelihood estimate. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 4 (1965): 69-73.

[203] Royden, H.L. *Real Analysis.* New York: Macmillan, 1968.

[204] Ruud, P.A. Tests of specification in econometrics. *Econometric Reviews* 3 (1984): 211-242.

[205] Schmetterer, L. *Mathematische Statistik*, 2nd edition. Wien: Springer Verlag, 1966.

[206] Schmidt, W.H. Asymptotic results for inadequate nonlinear regressions. In O. Bunke & E. Jolivet (eds.), *Seminarbericht* No. 89, pp. 103-129. Berlin: Humboldt Universität, 1987.

[207] Shapiro, A. Asymptotic properties of statistical estimators in stochastic programming. *Annals of Statistics* 17 (1989): 841-858.

[208] Silvey, S.D. The Lagrangian multiplier test. *Annals of Mathematical Statistics* 30 (1959): 389-407.

[209] Silvey, S.D. A note on maximum likelihood in the case of dependent random variables. *Journal of the Royal Statistical Society*, Series B, 23 (1961): 444-452.

[210] Slud, E.V. Partial likelihood for continuous-time stochastic processes. *Scandinavian Journal of Statistics* 19 (1992): 97-109.

[211] Tauchen, G. Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics* 30 (1985): 415-443.

[212] Tjøstheim, D. Estimation in nonlinear time series models. *Stochastic Processes and their Applications* 21 (1986): 251-273.

[213] Tjøstheim, D. Non-linear time series and Markov chains. *Advances in Applied Probability* 22 (1990): 587-611.

[214] Tong, H. *Threshold Models in Nonlinear Time Series Analysis.* Lecture Notes in Statistics 21. New York: Springer Verlag, 1983.

[215] Tong, H. *Non-linear Time Series: A Dynamical System Approach.* Oxford: Clarendon Press, 1990.

[216] Wald, A. Note on the consistency of the maximum likelihood estimate. *Annals of Mathematical Statistics* 20 (1949): 595-600.

[217] Weiss, L. Asymptotic properties of maximum likelihood estimators in some nonstandard cases I. *Journal of the American Statistical Association* 66 (1971): 345-350.

[218] Weiss, L. Asymptotic properties of maximum likelihood estimators in some nonstandard cases II. *Journal of the American Statistical Association* 68 (1973): 428-430.

[219] White, H. Nonlinear regression on cross-section data. *Econometrica* 48 (1980): 721-746.

[220] White, H. Consequences and detection of misspecified nonlinear regression models. *Journal of the American Statistical Association* 76 (1981): 419-433.

[221] White, H. Maximum likelihood estimation of misspecified models. *Econometrica* 50 (1982): 1-25.

[222] White, H. *Asymptotic Theory for Econometricians.* Orlando: Academic Press, 1984.

[223] White, H. Specification testing in dynamic models. In T. Bewley (ed.), *Advances in Econometrics - Fifth World Congress*, Vol. I, pp. 1-58. New York: Cambridge University Press, 1987.

[224] White, H. *Estimation, Inference and Specification Analysis.* Econometric Society Monographs No.22. Cambridge: Cambridge University Press, 1994.

[225] White, H. & I. Domowitz. Nonlinear regression with dependent observations. *Econometrica* 52 (1984): 143-161.

[226] Willard, S. *General Topology.* Reading, MA: Addison-Wesley, 1970.

[227] Withers, C.S. Conditions for linear processes to be strong-mixing. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 57 (1981a): 477-480.

[228] Withers, C.S. Central limit theorems for dependent variables. I. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 57 (1981b): 509-534.

[229] Withers, C.S. Corrigendum to 'Central limit theorems for dependent variables. I.'. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 63 (1983): 555.

[230] Wong, W.H. Theory of partial likelihood. *Annals of Statistics* 14 (1986): 88-123.

[231] Wong, W.H. On asymptotic efficiency in estimation theory. *Statistica Sinica* 2 (1992): 47-68.

[232] Wooldridge, J. Asymptotic properties of econometric estimators. University of California San Diego, Department of Economics. PhD dissertation, 1986.

[233] Wooldridge, J. A unified approach to robust, regression-based specification tests. *Econometric Theory* 6 (1990): 17-43.

[234] Wooldridge, J. On the application of robust, regression-based diagnostics to models of conditional means and conditional variances. *Journal of Econometrics* 47 (1991): 5-46.

[235] Wooldridge, J.M. Estimation and inference for dependent processes. In R.F. Engle and D.L. McFadden (eds.), *Handbook of Econometrics*, Vol. IV, pp. 2641-2738. Amsterdam: Elsevier Science B.V., 1994.

[236] Wooldridge, J. & H. White. Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory* 4 (1988): 210-230.

[237] Wu, C.F. Asymptotic theory of nonlinear least squares estimation. *Annals of Statistics* 9 (1981): 501-513.

[238] Zaman, A. Necessary and sufficient conditions for consistency of M-estimates. Columbia University, Department of Economics, 1985.

[239] Zaman, A. Consistency via type 2 inequalities: A generalization of Wu's theorem. *Econometric Theory* 5 (1989): 272-286.

[240] Zaman, A. Asymptotic suprema of averaged random functions. *Annals of Statistics* 19 (1991): 2145-2159.

Index

M

N

O

P

Y

Z