Applied Macroeconometrics

Carlo A. Favero

December 31, 1999

CONTENTS

1

APPLIED MACROECONOMETRICS. AN INTRODUCTION

1.1 Introduction

Once upon a time there was consensus both on the theoretical foundations of macroeconomics and on the correct approach to macroeconometric modelling (see, for example, Pesaran-Smith [7]). Such consensus, which was built around the "Cowles Commission" approach to model building, broke down dramatically at the beginning of the seventies when it was discovered that "...the models did not represent the data...did not represent the theory... were ineffective for practical purposes of forecasting and policy...". The breakdown of consensus has been rather spectacular, but, as Faust and Whiteman ([23]) put it "... even more impressive are the deep rifts that have emerged over the proper way to tease empirical facts from macroeconomic data..."

This book has the ambitious aim of discussing and illustrating the different approaches currently taken by the profession in doing applied macroeconometrics. We concentrate on the (large) subset of macroeconometrics dealing with time-series data. It is fair to say that the emergence of the deep rifts on the proper way to tease empirical facts from macroeconomic data has been paired with a deep awareness of the specificity of time-series data. We shall discuss the emergence of a plurality of approaches in macroeconomic modelling, within the framework provided by the statistical analysis of time series data. We begin our work with this introductory chapter, which reviews the basic in econometrics, describes the interaction between theory and data in applied work, and illustrates the importance of using time series instead of cross-section data in macroeconometrics.

1.2 From theory to data: the new-classical growth model.

Consider the Solow model of growth[1] This model takes as given the saving rate s, the rate of growth of population n, while technology, A, grows at a constant rate g. There are two inputs: capital, K,and labour, L, paid their respective marginal productivity. Output, Y, is determined by a Cobb-Douglas function with constant returns to scale:

$$Y_t = K_t^{\alpha} \left(A_t L_t \right)^{1-\alpha} \qquad 0 < \alpha < 1 \tag{1.1}$$

[1]The original reference is Solow ([9]). The data and the empirical analysis of this chapter replicate the results reported in Mankiw, Romer and Weil ([6]). For an excellent introduction to macroeconomic models of growth see Farmer ([1]).

$$L_t = L_{t-1}(1+n) \tag{1.2}$$

$$A_t = A_{t-1}(1+g) \tag{1.3}$$

Note that the number of effective unit of labour grows (approximately) at the rate $(n+g)$. The model is built by considering the production function together with two accounting identities and an ad hoc relation between savings and output. The two accounting identities are:

$$S_t = I_t \tag{1.4}$$

$$K_t = K_{t-1}(1-\delta) + I_t \tag{1.5}$$

where I, denotes investment, S,denotes savings and δ represents the rate of depreciation of the capital stock K. (1.4) makes immediately clear that we are considering a closed economy with no government sector.

The relationship between output and saving is determined by assuming a constant marginal propensity to save s:

$$\frac{S}{Y} = s \tag{1.6}$$

We define as k and y respectively the stock of capital per effective unit of labour (K/AL) and the level of output per effective unit of labour (Y/AL). By using all the equations in the model we have:

$$k_t(1+n)(1+g) = k_{t-1}(1-\delta) + sk_t^{\alpha} \tag{1.7}$$

Equation (1.7) determines the pattern over time of the stock of capital per effective unit of labour. From this relation we can pin down the steady state value of k, by setting $k^* = k_{t+i}$ for each i:

$$k^* = \left(\frac{s}{n+g+\delta}\right)^{\frac{1}{1-\alpha}} \tag{1.8}$$

(1.8) makes clear that the steady state k is positively related to the saving rate and negatively related to the rate of growth of population, the rate of technological progress and the rate of depreciation of capital.

By substituting (1.8) in the production function and taking logarithms we can derive the per capita steady state output as :

$$\ln\left(\frac{Y_t^*}{L_t}\right) = \ln A_0 + gt + \frac{\alpha}{1-\alpha}\ln(s) - \frac{\alpha}{1-\alpha}\ln(n+g+\delta) \qquad (1.9)$$

(1.8) makes very specific predictions on the impact on output of the saving rate and the rate of depreciation of capital, the rate of technological progress and the rate of depreciation of capital.

It is natural at this stage to raise a question on the empirical support to such well specified predictions. Mankiw, Romer and Weil ([6]) choose to test the model on data from a cross-section of countries. Such data are available in a database constructed by Summers-Heston (1988), which contains series on real output, private and government consumption, investment and population for virtually all countries in the world, excluding planned economies. The data are available at annual frequencies. Mankiw,Romer and Weil concentrate on the variables of interest for the period 1960-1985. The rate of growth of population, n, is measured by the average rate of growth of population in working age (15-64 years old). The rate of savings, s, is measured by the ratio of investment to GNP. n and s are averages for the period 1960-1985. y is measured by the log of GDP per working age person in 1985. $(g+\delta)$ is not directly observable and it is assumed constant at a value of 0.05. We concentrate on a sample of 75 countries labelled Intermediate by Mankiw, Romer and Weil and obtained considering non-oil producers countries with population higher than one million in 1960 and reliable data, thus excluding from the sample oil producers (as the bulk of GDP for such countries is not value added but extraction of existing resources), small countries and countries with low-quality data (receiving a grade of "D" from Summers and Heston). The data are contained, in EXCEL format, in the file MRW.XLS.

Now we have data and we have (1.9), which makes specific, theory-based predictions, on the relations between variables in our data set, the natural question is how we test empirically the Solow model?

The first point to note is that there is no stochastic structure in (1.9). Mankiw, Romer and Weil add a stochastic structure to the data by ignoring the difference between Y and Y^* and by concentrating on the term A. In fact A reflects not only the state of technology but also other factors, such as natural resources, climate, institutions, therefore the following specification is adopted for A :

$$\ln A_0 = a + \varepsilon_i$$

where a is a constant and ε_i represents a country-specific shock. (1.9) becomes now:

$$\ln y_i = \ln A_0 + gt + \frac{\alpha}{1-\alpha}\ln(s_i) - \frac{\alpha}{1-\alpha}\ln(n_i + g + \delta) + \varepsilon_i \quad (1.10)$$

which forms the basis for the empirical investigation.

1.3 The estimation problem: Ordinary Least Squares

The basis for the empirical test of the predictions of the Solow's growth model is the estimation of (1.10). Consider the estimation of the following model on our sample of 75 countries:

$$\ln y_i = \beta_0 + \beta_1 \ln(s_i) + \beta_2 \ln(n_i + g + \delta) + \varepsilon_i. \quad (1.11)$$

If the Solow model describes correctly the data, then the parameter β_0 should capture the term $\ln A_0 + gt$, which is a constant of the cross-section of data, while β_1 should be equal to $\frac{\alpha}{1-\alpha}$ and β_2 should instead take the value of $-\frac{\alpha}{1-\alpha}$. Therefore, independent information on factor shares could be used to assess the magnitude of the estimated coefficient: Mankiw,Romer and Weil claim that data on factor shares suggest one-third as a plausible value for α and therefore the elasticities of y_i with respect to s_i and $(n_i + g + \delta)$ should be respectively 0.5 and -0.5. Moreover, under the null of the validity of the Solow model, we have a testable restrictions on the parameters, namely $\beta_1 = -\beta_2$.

To illustrate how estimation can be performed, consider the following general representation of our model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$\mathbf{y} = \begin{pmatrix} y_1 \\ . \\ . \\ . \\ y_N \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & . & . & x_{1k} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ x_{N1} & x_{N2} & . & . & x_{Nk} \end{pmatrix}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ . \\ . \\ . \\ \beta_k \end{pmatrix}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \varepsilon_1 \\ . \\ . \\ . \\ \varepsilon_N \end{pmatrix}.$$

In our case $N = 75$, $k = 3$, the vector $\mathbf{y}$ contains 75 observations on per capita GDP while matrix $\mathbf{X}$ is (75×3). Note that the first column of $\mathbf{X}$ is made entirely of ones, the second column contains observations on

$\ln(s_i)$, while the third one contains observations on $\ln(n_i + g + \delta)$. The vector $\boldsymbol{\beta}$ contains three parameters: a constant and the two elasticities of interest in our economic problem.

The simplest method to derive estimates of the parameters of interest is the Ordinary Least Squares (OLS) method. Such method chooses values for the unknown parameter to minimize, in some sense, the magnitude of the non-observable components. Define the following quantity:

$$\mathbf{e}(\boldsymbol{\beta}) = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$$

where $\mathbf{e}(\boldsymbol{\beta})$ is a $(n \times 1)$ vector . If we treat $\mathbf{X}\boldsymbol{\beta}$, as a (conditional) prediction for $\mathbf{y}$, then we can consider $\mathbf{e}(\boldsymbol{\beta})$ as a forecasting error. The sum of the squared errors is then

$$\mathbf{S}(\boldsymbol{\beta}) = \mathbf{e}(\boldsymbol{\beta})' \mathbf{e}(\boldsymbol{\beta})$$

OLS produces an estimator of $\boldsymbol{\beta}$, $\widehat{\boldsymbol{\beta}}$, defined as follows:

$$\mathbf{S}\left(\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}\text{min}\mathbf{e}(\boldsymbol{\beta})' \mathbf{e}(\boldsymbol{\beta})$$

Given $\widehat{\boldsymbol{\beta}}$, we can define an associated vector of residual $\widehat{\epsilon}$ as
$\widehat{\epsilon} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}$. The OLS estimator can be derived by considering the necessary and sufficient conditions for $\widehat{\boldsymbol{\beta}}$ to be a unique minimum for $\mathbf{S}$:

i) $\mathbf{X}'\,\widehat{\epsilon} = 0$

ii) $\text{rank}(\mathbf{X}) = k$

Condition i) imposes orthogonality between the right-hand side variables on the OLS residuals, and ensures that the residual have an average of zero when a constant is included among the right-hand side variables (the regressors). Condition ii) requires that the columns of the $\mathbf{X}$ matrix are linearly independent: no variable in $\mathbf{X}$ can be expressed as a linear combination of the other variables in $\mathbf{X}$.

From i) we can derive an expression for the OLS estimates:

$$\mathbf{X}'\widehat{\epsilon} = \mathbf{X}'\left(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}\right) = \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} = 0$$
$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$

1.3.1 *Properties of the OLS estimates*

We have derived the OLS estimator without any assumption on the statistical structure of the data. In fact the statistical structure of the data is not needed to derive the estimator but to define its properties. To illustrate such properties we refer to the basic concepts of mean and variance of vector variables.

Given a generic vector of variables, $\mathbf{x}$

$$\mathbf{x} = \begin{pmatrix} x_1 \\ . \\ . \\ . \\ x_n \end{pmatrix}$$

we define the mean vector $E(\mathbf{x})$ and the mean matrix of outer products $E(\mathbf{xx}')$ as follows:

$$E(\mathbf{x}) = \begin{pmatrix} E(x_1) \\ . \\ . \\ . \\ E(x_n) \end{pmatrix}$$

$$\begin{aligned} E(\mathbf{xx}') &= E \begin{pmatrix} x_1^2 & x_1x_2 & . & . & x_1x_n \\ . & x_2^2 & . & . & x_2x_n \\ . & . & . & . & . \\ . & . & . & . & . \\ x_nx_1 & x_nx_2 & . & . & x_n^2 \end{pmatrix} \\ &= \begin{pmatrix} E(x_1^2) & E(x_1x_2) & . & . & E(x_1x_n) \\ . & E(x_2^2) & . & . & E(x_2x_n) \\ . & . & . & . & . \\ . & . & . & . & . \\ E(x_nx_1) & E(x_nx_2) & . & . & E(x_n^2) \end{pmatrix} \end{aligned}$$

The variance-covariance matrix of $\mathbf{x}$ is the defined as follows:

$$\begin{aligned} var(\mathbf{x}) &= E(\mathbf{x} - E(\mathbf{x})) E(\mathbf{x} - E(\mathbf{x}))' = \\ &= E(\mathbf{xx}') - E(\mathbf{x}) E(\mathbf{x})' \end{aligned}$$

Note that the variance-covariance matrix is symmetric and positive definite, by construction. In fact, given an arbitrary $\mathbf{A}$ vector of dimension n, we have :

$$var(\mathbf{A}'\mathbf{x}) = \mathbf{A}' var(\mathbf{x}) \mathbf{A}$$

The first relevant hypothesis for the derivation of the statistical properties of OLS regards the relationship between disturbances and regressors in the estimated equation. This hypothesis is constructed two parts: first it is assumed that

$E(\mathbf{y}_i \mid \mathbf{x}_i) = \mathbf{x}_i'\boldsymbol{\beta}$, this rules out the contemporaneous correlation between residuals and regressors (it is therefore valid if there are not omitted variables correlated with the regressors), second it is assumed that the components of the available sample are independently drawn. The second part of this assumption guarantees the equivalence between $E(\mathbf{y}_i \mid \mathbf{x}_i) = \mathbf{x}_i'\boldsymbol{\beta}$ and $\mathrm{E}(\mathbf{y}_i \mid \mathbf{x}_1, ...\mathbf{x}_i, ...\mathbf{x}_n) = \mathbf{x}_i'\boldsymbol{\beta}$. Using vector notation we have

$$E(\mathbf{y} \mid \mathbf{X}) = \mathbf{X}\boldsymbol{\beta}$$

which can be written equivalently as

$$E(\boldsymbol{\epsilon} \mid \mathbf{X}) = \mathbf{0} \tag{1.12}$$

Note that hypothesis (2.6) is very demanding. In fact, it implies that

$$E(\boldsymbol{\epsilon}_i \mid \mathbf{x}_1, ...\mathbf{x}_i, ...\mathbf{x}_n) = 0 \quad (i = 1, ...n)$$

The conditional mean is in general a non-linear function of $(\mathbf{x}_1, ...\mathbf{x}_i, ...\mathbf{x}_n)$, (2.6) requires that such function is a constant of zero. Note that (2.6) requires that each regressor is orthogonal not only to the error term associated to the same observation ($E(x_{ik}\varepsilon_i) = 0$ for all k) but also to the error tem associated to each other observations $(E(x_{jk}\varepsilon_i) = 0$ for all $j \neq k)$. This statement is proofed by using the properties of conditional expectations.

Given that $E(\boldsymbol{\epsilon} \mid \mathbf{X}) = \mathbf{0}$ implies, by the Law of Iterated Expectations, that $E(\boldsymbol{\epsilon}) = \mathbf{0}$, we have

$$E(\varepsilon_i \mid x_{jk}) = E[E(\varepsilon_i \mid \mathbf{x}) \mid x_{jk}] = 0 \tag{1.13}$$

Then

$$E(\varepsilon_i x_{jk}) = E[E(\varepsilon_i x_{jk} \mid x_{jk})] \tag{1.14}$$
$$= E[x_{jk} E(\varepsilon_i \mid x_{jk})] \tag{1.15}$$
$$= 0 \tag{1.16}$$

In the context of the Solow model (2.6) requires that s and n are independent from $\boldsymbol{\epsilon}$. Of course, such hypothesis is not going to hold in any time-series models when the time-series show some degree of persistence (in practice, always). Think of the simplest time-series model for a generic variable y :

$$y_t = a_0 + a_1 y_{t-1} + u_t.$$

It is clear that if $a_1 \neq 0$, then, although it is true that $\mathrm{E}(u_t \mid y_{t-1}) = 0$, $\mathrm{E}(u_{t-1} \mid y_{t-1}) \neq 0$ and (2.6)is destroyed, without any omitted variable problem.

This explains why we have used a cross-section example in our introductory chapter, we shall then complicate the framework to deal properly with time-series observations.

The second hypothesis defines constancy of the conditional variance of shocks:

$$E(\epsilon'\epsilon \mid \mathbf{X}) = \sigma^2 I \tag{1.17}$$

where σ^2 is a constant independent from $\mathbf{X}$.

The third hypothesis is the one, already introduced, which guarantees that the OLS estimator can be derived:

$$rank(\mathbf{X}) = k \tag{1.18}$$

Under hypotheses (2.6) – (1.18) we can derive the properties of the OLS estimator.

Property 1: unbiasedness

The conditional expectation (with respect to $\mathbf{X}$) of the OLS estimates is the vector of unknown parameters $\boldsymbol{\beta}$:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta}+\epsilon) \\
&= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon \\
E\left(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right) &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\epsilon \mid \mathbf{X}) \\
&= \boldsymbol{\beta}
\end{aligned}$$

by hypothesis (2.6) .

Property 2: variance of OLS

The conditional variance of the OLS estimator is $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$:

$$\begin{aligned}
var\left(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right) &= E\left(\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)' \mid \mathbf{X}\right) \\
&= E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon\epsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mid \mathbf{X}\right) \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\epsilon\epsilon' \mid \mathbf{X})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2 I\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}$$

Property 3: Gauss-Markov theorem

The OLS estimator is the most efficient in the class of linear unbiased estimators.

Consider the class of linear estimators:

$$\boldsymbol{\beta}_L = \mathbf{Ly}$$

This class is defined by the set of matrices (kxn) $\mathbf{L}$, which are fixed when conditioning upon $\mathbf{X}$. $\mathbf{L}$ does not depend on $\mathbf{y}$.Therefore we have:

$$\begin{aligned} E\left(\boldsymbol{\beta}_L \mid \mathbf{X}\right) &= E\left(\mathbf{LX}\boldsymbol{\beta} + \mathbf{L}\boldsymbol{\varepsilon} \mid \mathbf{X}\right) \\ &= \mathbf{LX}\boldsymbol{\beta} \end{aligned}$$

and $\mathbf{LX}\boldsymbol{\beta} = \boldsymbol{\beta}$ only if $\mathbf{LX} = \mathbf{I}_k$. Such condition is obviously satisfied by the OLS estimator, which is obtained by setting $\mathbf{L} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$. The variance of the general estimator in the class of linear unbiased estimators is readily obtained as:

$$\begin{aligned} var\left(\boldsymbol{\beta}_L \mid \mathbf{X}\right) &= E\left(\mathbf{L}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{L}' \mid \mathbf{X}\right) \\ &= \sigma^2\mathbf{LL}'. \end{aligned}$$

To show that the OLS estimator is the most efficient within this class we have to show that the variance of the OLS estimator differ from the variance of the generic estimator in the class by to a positive semi-definite matrix.

To this aim define $\mathbf{D} = \mathbf{L} - \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$; $\mathbf{LX} = \mathbf{I}$ requires $\mathbf{DX} = \mathbf{0}$.

$$\begin{aligned} \mathbf{LL}' &= \left(\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}' + \mathbf{D}\right)\left(\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} + \mathbf{D}'\right) \\ &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{D}' + \\ &\quad + \mathbf{DX}\left(\mathbf{X}'\mathbf{X}\right)^{-1} + \mathbf{DD}' \\ &= \left(\mathbf{X}'\mathbf{X}\right)^{-1} + \mathbf{DD}' \end{aligned}$$

from which we have that

$$var\left(\boldsymbol{\beta}_L \mid \mathbf{X}\right) = var\left(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right) + \sigma^2\mathbf{DD}'$$

which proves the point; in fact for any given matrix $\mathbf{D}$, not necessarily square, the symmetric matrix $\mathbf{DD}'$ is positive semidefinite.

1.4 OLS estimation of the Solow growth model

The results of the application of OLS to the Solow growth model are reported in Table 1. We report point estimates along with standard errors (square roots of the elements in the principal diagonal of the variance-covariance matrix of the OLS estimates). The Table is based on a regression run by using *E-Views* and exactly replicates the results in Table 1 of Mankiw, Romer and Weil (1992, p. 414).

TABLE 1: The estimation of the Solow model

Variable	Coefficient	Std. Error	t-ratio	Prob.
C	5.367698	1.540082	3.485333	0.0008
$\ln(s)$	1.325353	0.170611	7.768281	0.0000
$\ln(n+g+\delta)$	-2.013390	0.532830	-3.778672	0.0003
R-squared 0.601703	S.E. of regression 0.609456			
$\ln(s)$: log of savings rate (defined as LNS in MRW.XLS)				
$\ln(n+g+\delta)$:log of 0.05+ rate of growth of population (defined as LNNGD in MRW.XLS)				

Some consideration on these results are in order.

First, we have specified a model by deriving it directly from the theory and we have estimated it to derive empirical evidence on the validity of the model's prediction. In the light of the adoption of this specific strategy for research, the residual of the estimated model could be informative in that they reflect the impact of all variables omitted from the chosen specification. The analysis of residuals could be revealing on the mis-specification of the estimated model.

Second, the coefficients have the expected sign but the restriction implied by the theory are not exactly satisfied. In fact, the absolute values of the point estimates of the two elasticities are different, and their magnitude does not match available information on the capital-output ratio. MRW observe that the empirical observation of a capital-output ratio of about one third is consistent with an elasticity of the pro-capita output with respect to the saving rate of about 0.5 and elasticity of the pro-capita output with respect to $(n+g+\delta)$ of about -0.5. The natural question which raises at this point is related to the nature of estimated parameters. Given that they are random variable, it seems obvious to try and derive their distribution in order to test statistically hypothesis of economic interest. An interesting general hypothesis regards the significance of the estimated coefficients, while more specific hypothesis are of interest in testing the prediction of the theory (in our case $\beta_1 = -\beta_2, \beta_1 = 0.5, \beta_2 = -0.5$).

1.5 Residual analysis

Consider the following representation:

$$\begin{aligned}\widehat{\epsilon} &= \mathbf{y} - \mathbf{X}\widehat{\beta} \\ &= \mathbf{y} - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} = \mathbf{M}\mathbf{y}\end{aligned}$$

where $\mathbf{M} = \mathbf{I}_n - \mathbf{Q}$, and $\mathbf{Q} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$. The $(n \times n)$ matrices $\mathbf{M}$ and $\mathbf{Q}$, have the following properties:

i) they are symmetric: $\mathbf{M}' = \mathbf{M}, \mathbf{Q}' = \mathbf{Q}$;
ii) they are idempotent: $\mathbf{QQ} = \mathbf{Q}, \mathbf{MM} = \mathbf{M}$;
iii) $\mathbf{MX} = \mathbf{0}, \mathbf{MQ} = \mathbf{0}, \mathbf{QX} = \mathbf{X}$.

Note that the OLS projection for $\mathbf{y}$ can be written as $\widehat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{Qy}$, note also that $\widehat{\boldsymbol{\epsilon}} = \mathbf{My}$, from which we have the known result of orthogonality between the OLS residuals and regressors. We also have

$\mathbf{My} = \mathbf{MX}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\epsilon} = \mathbf{M}\widehat{\boldsymbol{\epsilon}}$, given that $\mathbf{MX} = \mathbf{0}$. Therefore we have a very well specified relation between the OLS residuals and the errors in the model $\widehat{\boldsymbol{\epsilon}} = \mathbf{M}\boldsymbol{\epsilon}$, which cannot be used to derive the errors given the residuals as the $\mathbf{M}$ matrix is not invertible.

We can re-write the sum of squared residuals as:

$$S\left(\widehat{\boldsymbol{\beta}}\right) = \widehat{\boldsymbol{\epsilon}}'\widehat{\boldsymbol{\epsilon}} = \boldsymbol{\epsilon}'\mathbf{M}'\mathbf{M}\boldsymbol{\epsilon} = \boldsymbol{\epsilon}'\mathbf{M}\boldsymbol{\epsilon}$$

$S\left(\widehat{\boldsymbol{\beta}}\right)$ is an obvious candidate for the construction of an estimate for σ^2. To derive an estimate of σ^2 from $S\left(\widehat{\boldsymbol{\beta}}\right)$ the concept of trace is useful. The trace of a square matrix is the sum of all elements on its principal diagonal. The following properties are relevant:

i) given any two square matrices $\mathbf{A}$ and $\mathbf{B}$,$tr\left(\mathbf{A} + \mathbf{B}\right) = tr\mathbf{A} + tr\mathbf{B}$;
ii) given any two matrices $\mathbf{A}$ and $\mathbf{B}$,$tr\left(\mathbf{AB}\right) = tr\left(\mathbf{BA}\right)$;
iii) the rank of an idempotent matrix is equal to its trace.

By using property ii) together with the fact that a scalar coincides with its trace we have :

$$\boldsymbol{\epsilon}'\mathbf{M}\boldsymbol{\epsilon} = tr\left(\boldsymbol{\epsilon}'\mathbf{M}\boldsymbol{\epsilon}\right) = tr\left(\mathbf{M}\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\right)$$

Now we can analyze the expected value of $S\left(\widehat{\boldsymbol{\beta}}\right)$,conditional upon $\mathbf{X}$:

$$\begin{aligned} E\left(S\left(\widehat{\boldsymbol{\beta}}\right) \mid \mathbf{X}\right) &= E\left(tr\mathbf{M}\boldsymbol{\epsilon}\boldsymbol{\epsilon}' \mid \mathbf{X}\right) \\ &= trE\left(\mathbf{M}\boldsymbol{\epsilon}\boldsymbol{\epsilon}' \mid \mathbf{X}\right) \\ &= tr\mathbf{M}\left(E\boldsymbol{\epsilon}\boldsymbol{\epsilon}' \mid \mathbf{X}\right) \\ &= \sigma^2 tr\mathbf{M} \end{aligned}$$

but, by using properties i) and ii), we have:

$$\begin{aligned} tr\mathbf{M} &= tr\mathbf{I}_n - tr\left(\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right) \\ &= n - tr\left(\mathbf{X}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right) \\ &= n - k \end{aligned}$$

Therefore an unbiased estimate of σ^2 is given by $S\left(\widehat{\boldsymbol{\beta}}\right) / \left(n - k\right)$.

This results allows the construction of the standard errors for the estimated OLS parameters reported in the second column of Table 1.

Using the result of orthogonality between the OLS projections and the OLS residuals we can write:

$$var(\mathbf{y}) = var(\widehat{\mathbf{y}}) + var(\widehat{\epsilon})$$

from which we can derive the following, residual based, indicator of goodness of fit:

$$R^2 = \frac{var(\widehat{\mathbf{y}})}{var(\mathbf{y})} = 1 - \frac{var(\widehat{\epsilon})}{var(\mathbf{y})}$$

The information contained in the R^2 is to be associated with the information contained in the standard error of the regression, which is the squared root of the estimated variance of OLS residuals. Note that, when a model is estimated in logarithms, the standard error of the regression does not depend on the unit of measures in which the variables are expressed. In fact, we have:

$$\begin{aligned} \widehat{\epsilon} &= \log(\mathbf{y}) - \log(\widehat{\mathbf{y}}) \\ &= \log\left(\frac{\mathbf{y}}{\widehat{\mathbf{y}}}\right) = \log\left(1 + \frac{\mathbf{y} - \widehat{\mathbf{y}}}{\widehat{\mathbf{y}}}\right) \simeq \frac{\mathbf{y} - \widehat{\mathbf{y}}}{\widehat{\mathbf{y}}} \end{aligned}$$

When the models are not specified in logs, standard errors are usually intepreted by dividing them by the mean of the dependent variable.

1.6 Elements of distribution theory

We consider the distribution of a generic n-dimensional vector $\mathbf{x}$, together with the derived distribution of the vector $\mathbf{y} = g(\mathbf{x})$ which admits inverse $\mathbf{x} = h(\mathbf{y})$,with $h = g^{-1}$. If $\text{prob}(x_1 < x < x_2) = \int_{x_1}^{x_2} f(x)\,dx$, and $\text{prob}(y_1 < y < y_2) = \int_{y_1}^{y_2} f^*(y)\,dy$, then

$$f^*(y) = f(h(\mathbf{y}))\,J$$

where $J = \begin{vmatrix} \frac{\partial h_1}{\partial y_1} & \dots & \frac{\partial h_n}{\partial y_1} \\ \dots & \dots & \dots \\ \frac{\partial h_1}{\partial y_n} & \dots & \frac{\partial h_n}{\partial y_n} \end{vmatrix} = \left|\frac{\partial \mathbf{h}}{\partial \mathbf{y}'}\right|$.

1.6.1 *The normal distribution*

The standardized normal univariate has the following distribution:

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right)$$
$$E(z) = 0, var(z) = 1$$

By considering the transformation $x = \sigma z + \mu$, we derive the distribution of the univariate normal as:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$
$$E(z) = \mu, var(z) = \sigma^2$$

Consider now the vector $\mathbf{z} = (z_1, z_2 ... z_n)$, such that

$$f(\mathbf{z}) = i = 1 \prod^{n} f(z_i) = (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\mathbf{z}'\mathbf{z}\right)$$

$\mathbf{z}$ is ,by construction, a vector of normal independent variables with zero mean and identity variance-covariance matrix. The conventional notation is $\mathbf{z} \sim \mathbf{N}(0, I_n)$.

Consider now the following linear transformation

$$\mathbf{x} = \mathbf{A}\mathbf{z} + \boldsymbol{\mu}$$

where $\mathbf{A}$ is an (nxn)invertible matrix. We consider the following transformation $\mathbf{z} = \mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ with Jacobian $J = \left|\mathbf{A}^{-1}\right| = \frac{1}{|\mathbf{A}|}$. By applying the formula for the transformation of variable, we have:

$$f(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} \left|\mathbf{A}^{-1}\right| \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \mathbf{A}^{-1\prime}\mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)$$

which, by defining the positive definite matrix $\Sigma = \mathbf{A}\mathbf{A}'$,can be re-written as follows:

$$f(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} \left|\Sigma^{-\frac{1}{2}}\right| \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \tag{1.19}$$

The conventional notation for the multivariate normal is $\mathbf{x} \sim \mathbf{N}(\boldsymbol{\mu}, \Sigma)$. An useful theorem is related to the multivariate normal:

Theorem 1.1 *For any* $\mathbf{x} \sim \mathbf{N}(\boldsymbol{\mu}, \Sigma)$ *given any* $(m \times n)$ $\mathbf{B}$ *matrix and any* $(m \times 1)$ *vector* $\mathbf{d}$,*if* $\mathbf{y} = \mathbf{B}\mathbf{x} + \mathbf{d}$, *then* $\mathbf{y} \sim \mathbf{N}\left(\mathbf{B}\boldsymbol{\mu} + \mathbf{d}, \mathbf{B}\Sigma\mathbf{B}'\right)$.

Consider partitioning a n-variate normal vector in two sub-vector of dimensions n_1 and $n - n_1$ as follows:

$$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right).$$

By applying the above theorem, we obtain the two following results:

i) $\mathbf{x}_1 \sim \mathbf{N}(\boldsymbol{\mu}_1, \Sigma_{11})$, which is obtained by applying the theorem in the case $\mathbf{d} = \mathbf{0}, \mathbf{B} = (I_{n_1} \ \ \mathbf{0})$;

ii) $(\mathbf{x}_1 \mid \mathbf{x}_2) \sim \mathbf{N}\left(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$, which is obtained by applying the theorem in the case $\mathbf{d} = \boldsymbol{\Sigma_{12}\Sigma_{22}^{-1}}\mathbf{x}_2, \mathbf{B} = \left(I_{n_1} \ \ -\Sigma_{12}\Sigma_{22}^{-1}\right)$.

Results ii) shows clearly that absence of correlation is equivalent to independence within the framework of a multivariate normal. This result is justified by the fact that the normal distribution is entirely described by its first two moments.

1.6.2 *Distributions derived from the normal*

Consider $\mathbf{z} \sim \mathbf{N}(0, I_n)$, an n-variate standard normal . The distribution of $\boldsymbol{\omega} = \mathbf{z}'\mathbf{z}$ is defined as a χ^2 with n degrees of freedom. Consider now two vectors $\mathbf{z}_1$ and $\mathbf{z}_2$ respectively of dimension n_1 and n_2 with the following distribution:

$$\begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} I_{n_1} & 0 \\ 0 & I_{n_2} \end{pmatrix} \right).$$

Then we have $\boldsymbol{\omega}_1 = \mathbf{z}_1'\mathbf{z}_1 \sim \chi^2(n_1)$, $\boldsymbol{\omega}_2 = \mathbf{z}_2'\mathbf{z}_2 \sim \chi^2(n_2)$, and $\boldsymbol{\omega}_1 + \boldsymbol{\omega}_2 = \mathbf{z}_1'\mathbf{z}_1 + \mathbf{z}_2'\mathbf{z}_2 \sim \chi^2(n_1 + n_2)$, in general the sum of two independent χ^2 is in itself distributed as χ^2 with a number of degrees of freedom equal to the sum of the degrees of freedom of the two χ^2.

From our discussion of the multivariate normal it follows that if $\mathbf{x} \sim \mathbf{N}(\boldsymbol{\mu}, \Sigma)$, then $(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \sim \chi^2(n)$.

A related result establish that if $\mathbf{z} \sim \mathbf{N}(0, I_n)$ and $\mathbf{M}$ is a symmetric idempotent $(n \times n)$ matrix of rank r, then $\mathbf{z}'\mathbf{M}\mathbf{z} \sim \chi^2(r)$.

Another distribution related to the normal is the F distribution. The F distribution is obtained as the ratio of two independent χ^2 divided by the respective degrees of freedom. Given $\boldsymbol{\omega}_1 \sim \chi^2(n_1)$, and $\boldsymbol{\omega}_2 \sim \chi^2(n_2)$, we have:

$$\frac{\boldsymbol{\omega}_1/n_1}{\boldsymbol{\omega}_2/n_2} \sim F(n_1, n_2).$$

The Student's t distribution is then defined as follows:

$$t_n = \sqrt{F(1, n)}.$$

Another useful result establish that two quadratic forms in the standard multivariate normal, $\mathbf{z}'\mathbf{M}\mathbf{z}$ and $\mathbf{z}'\mathbf{Q}\mathbf{z}$, are independent if $\mathbf{MQ} = \mathbf{0}$. We can finally state the following theorem, which is fundamental to the statistical inference in the linear model:

Theorem 1.2 *If* $\mathbf{z} \sim \mathbf{N}(0, I_n)$ *with* $\mathbf{M}$ *and* $\mathbf{Q}$ *symmetric and idempotent matrices of respectively of rank* r *and* s *and* $\mathbf{MQ} = \mathbf{0}$, *then we have:* $\frac{\mathbf{z}'\mathbf{Q}\mathbf{z}}{\mathbf{z}'\mathbf{M}\mathbf{z}} \frac{r}{s} \sim \mathbf{F}(s, r)$.

1.7 Inference in the linear regression model

In order to perform inference in the linear regression model a further hypothesis is needed to specify the distribution of $\mathbf{y}$ conditional upon $\mathbf{X}$:

$$y \mid \mathbf{X} \sim \mathbf{N}\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2 I\right) \tag{1.20}$$

or, equivalently

$$u \mid \mathbf{X} \sim \mathbf{N}\left(\mathbf{0}, \sigma^2 I\right) \tag{1.21}$$

given (1.20) we can immediately derive the distribution of $\left(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right)$ which, being a linear combination of a normal distribution, is also normal:

$$\left(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right) \sim \mathbf{N}\left(\boldsymbol{\beta}, \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right). \tag{1.22}$$

(1.22) constitutes the basis to construct confidence intervals and to perform hypothesis testing in the linear regression model. Consider the following expression:

$$\begin{aligned} \frac{\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)' \mathbf{X}'\mathbf{X}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)}{\sigma^2} &= \frac{\mathbf{u}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u}}{\sigma^2} \\ &= \frac{\mathbf{u}'\mathbf{Q}\mathbf{u}}{\sigma^2} \end{aligned}$$

and, by applying the results derived in the previous section, we know that :

$$\frac{\mathbf{u}'\mathbf{Q}\mathbf{u}}{\sigma^2} \mid \mathbf{X} \sim \chi^2(k) \tag{1.23}$$

(1.23) is not very useful in practice, as we do not know σ^2. However, we know that

$$\frac{S\left(\widehat{\boldsymbol{\beta}}\right) \mid \mathbf{X}}{\sigma^2} = \frac{\mathbf{u}'\mathbf{M}\mathbf{u}}{\sigma^2} \mid \mathbf{X} \sim \chi^2(T - k). \tag{1.24}$$

As $\mathbf{MQ} = \mathbf{0}$, we know the distribution of the ratio of (1.23) to (1.24), moreover by taking ratio we get rid of the unknown term σ^2 :

$$\frac{\frac{(\widehat{\beta}-\beta)'\mathbf{X}'\mathbf{X}(\widehat{\beta}-\beta)}{\sigma^2}}{\frac{s^2}{\sigma^2}} = \frac{\mathbf{u}'\mathbf{Q}\mathbf{u}}{\mathbf{u}'\mathbf{M}\mathbf{u}}(T-k) \sim kF(k,T-k). \tag{1.25}$$

Result (1.25) can be used by obtaining from the tables of the F distribution the critical value $F_{\alpha}^{*}(k,T-k)$ such that:

$$prob\left[F(k,T-k) > F_{\alpha}^{*}(k,T-k)\right] = \alpha \qquad 0 < \alpha < 1$$

for different values of α we are in the position of evaluating exactly inequality of the following form:

$$prob\left\{\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\mathbf{X}'\mathbf{X}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right) \leq ks^2F_{\alpha}^{*}(k,T-k)\right\} = 1-\alpha$$

which define confidence intervals for $\boldsymbol{\beta}$ centered upon $\widehat{\boldsymbol{\beta}}$. Hypothesis testing is strictly linked to the derivation of confidence interval. When we test hypothesis we aim at rejecting the validity of restrictions imposed on the model on the basis of the sample evidence. Within this framework our hypothesis (2.6) – (1.22) are the maintained hypothesis and the restricted version of the model is identified with the null hypothesis H_0. Following the Neyman-Pearson approach to hypothesis testing a statistic with known distribution under the null is derived. Then the probability of first type error (reject H_0 when it is true) is fixed at α. For example a test at the level α of the null hypothesis $\boldsymbol{\beta}=\boldsymbol{\beta}_0$, based on the F-statistic, is given when we do not reject the null H_0 if $\boldsymbol{\beta}_0$ lies within the confidence interval associated to the probability $1-\alpha$. However, in practice, this is not a useful way of proceeding, as very rarely the economic hypotheses of interest involve a number of restrictions equal to the number of estimated parameters. Reconsider the Solow's growth model: we have three estimated parameters but only one restriction.

The general case of interest to the economist is the one when we have r restrictions on the vector of parameters where $r < k$. If we limit our interest to the class of linear restrictions, we can express them as follows:

$$H_0 = \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$$

where $\mathbf{R}$ is a $(r \times k)$ matrix of parameters with rank k and $\mathbf{r}$ is and $(r \times 1)$ vector of parameters. To illustrate how $\mathbf{R}$ and $\mathbf{r}$ are constructed, consider the base line case of the Solow model: we want to impose the restriction $\beta_1 = -\beta_2$ on the following specification:

$$\ln y_i = \beta_0 + \beta_1 \ln(s_i) + \beta_2 \ln(n_i + g + \delta) + \varepsilon_i \tag{1.26}$$

$$\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$$

$$\begin{pmatrix} 0 \; 1 \; 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = (0).$$

The distribution of a known statistic under the null can be derived by applying known results.

If $\left(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right) \sim \mathbf{N}\left(\boldsymbol{\beta}, \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)$, then we have

$$\left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r} \mid \mathbf{X}\right) \sim \mathbf{N}\left(\mathbf{R}\boldsymbol{\beta} - \mathbf{r}, \sigma^2 \mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right) \tag{1.27}$$

The test is constructed by deriving the distribution of (1.27) under the null $\mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \mathbf{0}$.

Given that

$$\left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r} \mid \mathbf{X}\right) = \mathbf{R}\boldsymbol{\beta} - \mathbf{r} + \mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-\mathbf{1}}\mathbf{X}'\mathbf{u}$$

under H_0, we have

$$\left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}\right)' \left(\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right)^{-1} \left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}\right)$$

$$\begin{aligned} &= \mathbf{u}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-\mathbf{1}}\mathbf{R}'\left(\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right)^{-1}\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-\mathbf{1}}\mathbf{X}'\mathbf{u} \\ &= \mathbf{u}'\mathbf{P}\mathbf{u} \end{aligned}$$

where $\mathbf{P}$ is a symmetric idempotent matrix of rank r, orthogonal to $\mathbf{M}$.

We have then :

$$\frac{\left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}\right)' \left(\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right)^{-1} \left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}\right)}{s^2} \sim \mathbf{r}\mathbf{F}\left(r, T - k\right) \qquad \text{under } H_0.$$

Which can be used to test relevant hypothesis. We report the application of this methodology to our economic case of interest in Table 2.

TABLE 2: Testing linear restrictions on equation (1.11)

	$F(n_1, n_2)$	Probability
$\beta_1 = -\beta_2$	$F(1, 72) = 1.255627$	0.266204
$\beta_1 = 0.5$, $\beta_2 = -0.5$	$F(2, 72) = 23.48172$	0.000000

The null hypothesis $\beta_1 = -\beta_2$ cannot be rejected for values of α smaller than 0.2662, therefore such hypothesis cannot be rejected at the conventional five per cent. While the null hypothesis $\beta_1 = 0.5, \beta_2 = -0.5$ is rejected at the conventional five per cent, and also at the one per cent level. Note that, in Table 1, we have already reported the t-values on estimated coefficients, which did reject the null hypothesis of the coefficient being equal to zero at conventional critical levels. An interesting specific case of the test of the validity of restrictions on estimated coefficients are test for the significance of subset of coefficients, which we are going to discuss in the next section.

1.7.1 *Testing the significance of subset of coefficients*

In the general framework to test linear restrictions set $\mathbf{r} = \mathbf{0}, \mathbf{R} = \begin{bmatrix} I_r & 0 \end{bmatrix}$, and partition in a corresponding way $\boldsymbol{\beta}$ into $\begin{bmatrix} \boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 \end{bmatrix}$. In this case the restrictions $\mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \mathbf{0}$ are equivalent to $\boldsymbol{\beta}_1 = 0$ in the partitioned regression model:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{u}$$

in which partitioning creates two blocks of dimension r and $k - r$.

Before proceeding to the discussion of hypothesis testing it useful to derive the formula for the OLS estimator in the partitioned regression model. To obtain such results partition as follows the "normal equations" $\mathbf{X}'\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$:

$$\begin{pmatrix} \mathbf{X}_1' \\ \mathbf{X}_2' \end{pmatrix} \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\beta}}_1 \\ \widehat{\boldsymbol{\beta}}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1' \\ \mathbf{X}_2' \end{pmatrix} \mathbf{y}$$

or, equivalently

$$\begin{pmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\beta}}_1 \\ \widehat{\boldsymbol{\beta}}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1'\mathbf{y} \\ \mathbf{X}_2'\mathbf{y} \end{pmatrix} \tag{1.28}$$

system (1.28) can be resolved in two stages by deriving first an expression $\widehat{\boldsymbol{\beta}}_2$as follows:

$$\widehat{\boldsymbol{\beta}}_2 = \left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1} \left(\mathbf{X}_2'\mathbf{y} - \mathbf{X}_2'\mathbf{X}_1\widehat{\boldsymbol{\beta}}_1\right)$$

and then by substituting it in the first equation of (1.28) to obtain:

$$\mathbf{X}_1'\mathbf{X}_1\widehat{\boldsymbol{\beta}}_1 + \mathbf{X}_1'\mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\left(\mathbf{X}_2'\mathbf{y} - \mathbf{X}_2'\mathbf{X}_1\widehat{\boldsymbol{\beta}}_1\right) = \mathbf{X}_1'\mathbf{y}$$

from which we have[2]:

$$\widehat{\boldsymbol{\beta}}_1 = \left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{y}$$
$$\mathbf{M}_2 = \left(\mathbf{I} - \mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\right).$$

Note that, as $\mathbf{M}_2$ is idempotent, we can also write:

$$\widehat{\boldsymbol{\beta}}_1 = \left(\mathbf{X}_1'\mathbf{M}_2'\mathbf{M}_2\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{M}_2'\mathbf{M}_2\mathbf{y}$$

and $\widehat{\boldsymbol{\beta}}_1$can be interpreted as the vector of OLS coefficients of the regression of $\mathbf{y}$ on the matrix of residuals of the regression of $\mathbf{X}_1$ on $\mathbf{X}_2$. So an OLS regression on two regressors is equivalent to two OLS regressions on a single regressor (Frisch-Waugh theorem).

Finally, consider the residuals of the partitioned model:

$$\widehat{\mathbf{u}} = \mathbf{y} - \mathbf{X}_1\widehat{\boldsymbol{\beta}}_1 - \mathbf{X}_2\widehat{\boldsymbol{\beta}}_2$$

$$\widehat{\mathbf{u}} = \mathbf{y} - \mathbf{X}_1\widehat{\boldsymbol{\beta}} - \mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\left(\mathbf{X}_2'\mathbf{y} - \mathbf{X}_2'\mathbf{X}_1\widehat{\boldsymbol{\beta}}_1\right)$$

$$\begin{aligned}\widehat{\mathbf{u}} &= \mathbf{M}_2\mathbf{y} - \mathbf{M}_2\mathbf{X}_1\widehat{\boldsymbol{\beta}}_1 \\ &= \mathbf{M}_2\mathbf{y} - \mathbf{M}_2\mathbf{X}_1\left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{y} \\ &= \left(\mathbf{M}_2 - \mathbf{M}_2\mathbf{X}_1\left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{M}_2\right)\mathbf{y}\end{aligned}$$

but we already know that $\widehat{\mathbf{u}} = \mathbf{M}\mathbf{y}$,therefore it must be that

$$\mathbf{M} = \left(\mathbf{M}_2 - \mathbf{M}_2\mathbf{X}_1\left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{M}_2\right) \tag{1.29}$$

We are now in the position of reconsider testing for our null of interest. Under H_0 $\mathbf{X}_1$ has no additional explicatory power for $\mathbf{y}$ with respect to $\mathbf{X}_2$, therefore:

[2]Note that the expression for the estimator can be obtained by applying directly on the normal equations the formula of the partitioned inverse:
$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} E & -EBD^{-1} \\ -D^{-1}CE & D^{-1} + D^{-1}CEBD^{-1} \end{pmatrix}$
$E = \left(A - BD^{-1}C\right)^{-1}$

$H_0 : \mathbf{y} = \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{u}, \qquad (\mathbf{u} \mid \mathbf{X}_1, \mathbf{X}_2) \sim N\left(0, \sigma^2 I\right).$

Note that the statement

$\mathbf{y} = \mathbf{X}_2\boldsymbol{\gamma}_2 + \mathbf{u}, \qquad (\mathbf{u} \mid \mathbf{X}_2) \sim N\left(0, \sigma^2 I\right)$

it is always true within our maintained hypotheses. However, in general $\boldsymbol{\gamma}_2 \neq \boldsymbol{\beta}_2$.

In order to derive a statistic to test H_0 remember that the general matrix $\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'$ is the upper left block of $\left(\mathbf{X}'\mathbf{X}\right)^{-1}$, which we can now write as $\left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{-1}$. The statistic takes then the following form:

$$\frac{\widehat{\boldsymbol{\beta}}_1' \left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right) \widehat{\boldsymbol{\beta}}_1}{r s^2} =$$

$$= \frac{\mathbf{y}'\mathbf{M_2X}_1 \left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{-1} \mathbf{X}_1'\mathbf{M}_2\mathbf{y}}{\mathbf{y}'\mathbf{M}\mathbf{y}} \frac{T-k}{r} \sim F\left(T-k, r\right) \qquad (1.30)$$

given (1.29), (1.28) can be re-written as:

$$= \frac{\mathbf{y}'\mathbf{M_2y} - \mathbf{y}'\mathbf{My}}{\mathbf{y}'\mathbf{My}} \frac{T-k}{r} \sim F\left(T-k, r\right) \qquad (1.31)$$

where the denominator is the sum of squared residuals in the unconstrained model, while the numerator is the difference between the sum of residuals in the constrained model and the sum of residuals in the unconstrained model.

Consider now the limit case in $r = 1$ and β_1 is a scalar. In this case the F-statistic takes the following form:

$$\frac{\widehat{\boldsymbol{\beta}}_1^2}{s^2 \left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)} \sim F\left(T-k, r\right) \text{ under } H_0$$

where $\left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{-1}$ is now the element $(1,1)$ of the matrix $\left(\mathbf{X}'\mathbf{X}\right)^{-1}$.

Using the result on the relation between the F and the Student's t distribution we have :

$$\frac{\widehat{\boldsymbol{\beta}}_1}{s \left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{1/2}} \sim t\left(T-k\right) \text{ under } H_0.$$

Therefore an immediate test of significance of coefficient can be performed, as it is done in Table 1, by taking the ratio between each estimated coefficient and the associated standard error.

Let us now reconsider our results on the Solow growth model. We cannot reject the null of the validity of the model but the point estimates are rather far from the predicted coefficient on the basis of the theory.

Two questions naturally rise at this stage.

First what is the impact on our coefficient of having estimated the model without imposing the theoretical restrictions? Second, is it possible to explain the discrepancies between the estimated elasticities and the one predicted by the theory on the basis of the mis-specification of the model, i.e. of the omission from the estimated model of some variables relevant to explain $\mathbf{y}$?

In the following two sections we take these two questions in turn.

1.8 Estimation under linear constraints

In this section we analyze the impact on the OLS estimator of a kind of mis-specification deriving from ignoring the existence of constraints on estimated parameter. To analyze mis-specification we introduce the difference between the estimated model and the Data Generating Process (DGP).

The estimated model is the linear model analyzed up to now:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

while the DGP is instead:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ s.to \quad \mathbf{R}\boldsymbol{\beta} - \mathbf{r} &= \mathbf{0}. \end{aligned}$$

Where the constraints have been expressed using the so called implicit form. A very useful alternative way of expressing constraints, known as the "explicit form" has been expressed by Sargan (1988):

$$\boldsymbol{\beta} = \mathbf{S}\boldsymbol{\theta} + \mathbf{s}$$

where $\mathbf{S}$ is a $(k \times (k - r))$ matrix of rank $k - r$ and $\mathbf{s}$ is a $k \times 1$ vector.

To show how constraints are specified in the two alternatives let us reconsider the restrictions of the Solow growth model $\beta_1 = -\beta_2$ on the following specification:

$$\ln y_i = \beta_0 + \beta_1 \ln(s_i) + \beta_2 \ln(n_i + g + \delta) + \varepsilon_i \tag{1.32}$$

Using $\mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \mathbf{0}$ we have:

$$\begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = (0)$$

while using $\boldsymbol{\beta} = \mathbf{S}\boldsymbol{\theta} + \mathbf{s}$ we have

$$\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

In practice the constraints in explicit form are written by considering $\boldsymbol{\theta}$ as the vector of free parameters. Note that there is no unique way of expressing constraints in explicit form, in our case the same constraint could have been imposed as follows:

$$\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

As the two alternatives are indifferent $\mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \mathbf{0}$ is equivalently written as $\mathbf{RS}\boldsymbol{\theta} + \mathbf{Rs} - \mathbf{r} = \mathbf{0}$ which implies:
i) $\mathbf{RS} = \mathbf{0}$;
ii) $\mathbf{Rs} - \mathbf{r} = \mathbf{0}$.

We shall use the explicit form of imposing constraints to derive the Restricted Least Squares (RLS) estimators and to evaluate consistency and relative efficiency of OLS and RLS.

1.8.1 *The Restricted Least Squares (RLS) estimator*

To construct RLS substitute the constraint in the original model to obtain:

$$\mathbf{y} - \mathbf{Xs} = \mathbf{XS}\boldsymbol{\theta} + \mathbf{u} \tag{1.33}$$

equation (1.33)could be rewritten as :

$$\mathbf{y}^{*} = \mathbf{X}^{*}\boldsymbol{\theta} + \mathbf{u} \tag{1.34}$$

where $\mathbf{y}^{*} = \mathbf{y} - \mathbf{Xs}, \mathbf{X}^{*} = \mathbf{XS}$.

Note that the transformed model features the same residuals with the original model; therefore if hypotheses (2.6) – (1.20) hold for the original model, then they also hold for the transformed model. So we apply OLS to the transformed model to obtain:

$$\begin{aligned} \widehat{\boldsymbol{\theta}} &= \left(\mathbf{X}^{*\prime}\mathbf{X}^{*}\right)^{-1}\mathbf{X}^{*\prime}\mathbf{y}^{*} \\ &= \left(\mathbf{S}'\mathbf{X}'\mathbf{XS}\right)^{-1}\mathbf{S}'\mathbf{X}'\left(\mathbf{y} - \mathbf{Xs}\right) \end{aligned} \tag{1.35}$$

from (1.35) the RLS estimation is easily obtained by applying the tranformation $\widehat{\boldsymbol{\beta}}^{rls} = \mathbf{S}\widehat{\boldsymbol{\theta}} + \mathbf{s}$. Similarly the variance of the RLS estimator is easily obtained as follows:

$$var\left(\widehat{\boldsymbol{\theta}} \mid \mathbf{X}\right) = \sigma^2 \left(\mathbf{X}^{*\prime}\mathbf{X}^{*}\right)^{-1} = \sigma^2 \left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1}$$

$$\begin{aligned} var\left(\widehat{\boldsymbol{\beta}}^{rls} \mid \mathbf{X}\right) &= var\left(\mathbf{S}\widehat{\boldsymbol{\theta}} + \mathbf{s} \mid \mathbf{X}\right) \\ &= \mathbf{S}var\left(\widehat{\boldsymbol{\theta}} \mid \mathbf{X}\right)\mathbf{S}' \\ &= \sigma^2\mathbf{S}\left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1}\mathbf{S}'. \end{aligned}$$

We report in Table 3 the RLS estimator of the Solow growth model. By comparing Table 1 and Table 3 we note that estimated coefficients are very close but estimates in Table 3 are more precise. We also note that the hypothesis $\beta_1 = 0.5, \beta_2 = -0.5$ is still rejected, despite our imposition of the theory-based constraint on our estimated coefficients.

TABLE 3: The estimation of the constrained Solow model

Variable	Coefficient	Std. Error	t-ratio	Prob.
C	7.0857	0.1453	48.65	0.0000
lns-lnngd	1.43687	0.13882	10.35	0.0000
R-squared 0.594	S.E. of regression 0.61			

This observation leads naturally to the discussion of the properties of OLS and RLS in the case of a DGP with constraints.

unbiasedness

This is easy, under the assumed DGP they are both unbiased as such properties depend on the validity of hypotheses $(2.6) - (1.20)$, which is not affected by the imposition of constraints on parameters.

efficiency

On this issue we note immediately that if we interpret RLS ad the OLS estimator on the transformed model (1.35)we immediately derive the results that the RLS is the most efficient estimator as the hypotheses for the validity of the Gauss-Markov theorem are satisfied when OLS is applied to (1.35). Note that, by posing $\mathbf{L} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$ in the context of the transformed model, we do not in general obtain OLS but we obtain an estimator whose conditional variance with respect to $\mathbf{X}$, coincides with the conditional variance of the OLS estimator.

We support this intuition with a formal argument by showing that the difference between the variance of the OLS estimator and the variance of the RLS estimator is a positive semi-definite matrix.

$$var\left(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right) - var\left(\widehat{\boldsymbol{\beta}}^{rls} \mid \mathbf{X}\right) = \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1} - \sigma^2 \mathbf{S}\left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1}\mathbf{S}'$$

Define $\mathbf{A}$ as follows:

$$\mathbf{A} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} - \mathbf{S}\left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1}\mathbf{S}'$$

given that

$$\begin{aligned}
\mathbf{A}\mathbf{X}'\mathbf{X}\mathbf{A} &= \left(\left(\mathbf{X}'\mathbf{X}\right)^{-1} - \mathbf{S}\left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1}\mathbf{S}'\right)\mathbf{X}'\mathbf{X}\left(\left(\mathbf{X}'\mathbf{X}\right)^{-1} - \mathbf{S}\left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1}\mathbf{S}'\right) \\
&= \left(\mathbf{X}'\mathbf{X}\right)^{-1} - 2\mathbf{S}\left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1}\mathbf{S}' + \mathbf{S}\left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1}\mathbf{S}'\mathbf{S}\left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1}\mathbf{S}' \\
&= \left(\mathbf{X}'\mathbf{X}\right)^{-1} - \mathbf{S}\left(\mathbf{S}'\mathbf{X}'\mathbf{X}\mathbf{S}\right)^{-1}\mathbf{S}' = \mathbf{A}
\end{aligned}$$

we have that $\mathbf{A}$ is positive semi-definite, being the product of a matrix and its transpose.

The OLS estimator ignores available information and therefore it is less efficient than the RLS estimator. However, there is no difference between the two estimators in terms of unbiasedness.

So far we have evaluated the gains of imposing true restrictions, a related interesting exercise is the evaluation of the loss of imposing false restrictions.

1.9 The effects of mis-specification

We consider two general cases of mis-specification to evaluate empirically their importance within the Solow model. We take first the case of under-parameterization (the estimated model omits variables in the DGP) to move on to the case of over-parameterization (the estimated model includes more variables than the DGP). We evaluate the effects of mis-specification on the OLS estimators by using results from the partitioned regression model.

1.9.1 *Under-parameterisation*

Given the following DGP:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{u}, \tag{1.36}$$

for which hypotheses (2.6) – (1.20) hold, the following model is estimated :

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\nu}. \tag{1.37}$$

Therefore, the OLS estimates are given by the following expression

$$\widehat{\boldsymbol{\beta}}_1^{up} = \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{y} \tag{1.38}$$

while the OLS estimates which would have been obtained by estimation of the DGP would have been:

$$\widehat{\boldsymbol{\beta}}_1 = \left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{y} \tag{1.39}$$

The estimates in (1.39) are Best Linear Unbiased Estimators by construction, while the estimates in (1.38) are biased unless the correlation between $\mathbf{X}_1$and $\mathbf{X}_2$ is zero. To show this point consider that:

$$\widehat{\boldsymbol{\beta}}_1 = \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\left(\mathbf{X}_1'\mathbf{y} - \mathbf{X}_1'\mathbf{X}_2\widehat{\boldsymbol{\beta}}_2\right) \tag{1.40}$$

$$= \widehat{\boldsymbol{\beta}}_1^{up} + \widehat{\mathbf{D}}\widehat{\boldsymbol{\beta}}_2 \tag{1.41}$$

where $\widehat{\mathbf{D}}$ is the vector of coefficients in the regression of $\mathbf{X}_2$ on $\mathbf{X}_1$,and $\widehat{\boldsymbol{\beta}}_2$ is the OLS estimator obtained by fitting the DGP.

To provide further interpretation of these results note that if we have:

$$E\left(\mathbf{y} \mid \mathbf{X}_1, \mathbf{X}_2\right) = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2$$
$$E\left(\mathbf{X}_1 \mid \mathbf{X}_2\right) = \mathbf{X}_1\mathbf{D}$$

then

$$E\left(\mathbf{y} \mid \mathbf{X}_1\right) = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_1\mathbf{D}\boldsymbol{\beta}_2 = \mathbf{X}_1\boldsymbol{\alpha}.$$

Therefore the OLS estimator in the underparameterised model is a biased estimator of $\boldsymbol{\beta}_1$, but it is an unbiased estimator of $\boldsymbol{\alpha}$. Then, if the objective of the model is forecasting and $\mathbf{X}_1$ is more easily observed than $\mathbf{X}_2$, than the undeparameterised model can be safely used. On the other hand, if the objective of the model is to test specific predictions on parameters (as it is the case with the Solow's growth model), than the use of the under-parameterised model will deliver biased results. When we are interested in the effect of $\mathbf{X}_1$ on $\mathbf{y}$, independently from other factors, it is crucial to control for the effects of omitted variables.

1.9.2 *Over-parameterization*

Given the following DGP:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{u} \tag{1.42}$$

for which hypotheses (2.6) – (1.20)hold, the following model is estimated :

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{v}. \tag{1.43}$$

The OLS estimator of the over-parameterized model is

$$\widehat{\boldsymbol{\beta}}_1^{op} = \left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{y} \tag{1.44}$$

while, by estimating the DGP, we obtain:

$$\widehat{\boldsymbol{\beta}}_1 = \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{y}. \tag{1.45}$$

By substituting $\mathbf{y}$ from the DGP it is immediately shown that both estimators are unbiased. The difference is now made by the variance. In fact we have :

$$var\left(\widehat{\boldsymbol{\beta}}_1^{op} \mid \mathbf{X}_1, \mathbf{X}_2\right) = \sigma^2\left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{-1} \tag{1.46}$$

$$var\left(\widehat{\boldsymbol{\beta}}_1 \mid \mathbf{X}_1, \mathbf{X}_2\right) = \sigma^2\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1} \tag{1.47}$$

It can be shown that the estimator derived from the correct model is more efficient. In fact, the difference between the two variance-covariance matrices is a positive semi-definite matrix. To show this result remember that if two matrices $\mathbf{A}$ and $\mathbf{B}$ are positive definite and $\mathbf{A} - \mathbf{B}$ is positive semi-definite, then also the matrix $\mathbf{B}^{-1} - \mathbf{A}^{-1}$ is positive semi-definite. Then we have to show that $\mathbf{X}_1'\mathbf{X}_1 - \mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1$ is a positive semi-definite matrix. Such result is almost immediately shown:

$$\begin{aligned}\mathbf{X}_1'\mathbf{X}_1 - \mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1 &= \mathbf{X}_1'\left(\mathbf{I} - \mathbf{M}_2\right)\mathbf{X}_1 \\ &= \mathbf{X}_1'\mathbf{Q}_2\mathbf{X}_1 = \mathbf{X}_1'\mathbf{Q}_2\mathbf{Q}_2\mathbf{X}_1.\end{aligned}$$

We can then conclude that overparameterization impact on the efficiency of estimators and on the power of tests of hypotheses.

1.10 Human capital in the Solow's growth model

Let us reconsider the question of the estimated elasticities in the Solow growth model. We have seen that the theory implied restriction on the equality of elasticities cannot be rejected, but that imposing such constraint does not solve the problem of the implausibly high values for the point estimates of elasticities. Our discussion of the effect of omitted variables on OLS estimation illustrates a potential solution to the problem. MRW follow this lead and point out that human capital could be the relevant omitted variable. To illustrate the impact of human capital on the Solow growth model, let us augment our simple specification to consider three inputs: physical capital, K, human capital, H and labour, L. By keeping a constant-returns-to-scale Cobb-Douglas production function, we have:

$$Y_t = K_t^{\alpha} H_t^{\beta} \left(A_t L_t\right)^{1-\alpha-\beta} \tag{1.48}$$

Define as s_h the fraction of output invested in physical capital and as s_k the fraction of output invested in human capital. We maintain all the other original assumption in the Solow's model and we also assume that physical and human capital depreciate at the same speed. The evolution of the economy over time is now governed by the two following dynamic equations:

$$k_t \left(1+n\right)\left(1+g\right) = k_{t-1}\left(1-\delta\right) + s_k y_{t-1} \tag{1.49}$$

$$h_t \left(1+n\right)\left(1+g\right) = h_{t-1}\left(1-\delta\right) + s_h y_{t-1}. \tag{1.50}$$

By assuming $\alpha+\beta < 1$, we can derive the steady-state of the economy defined by the two following relationships:

$$k^* = \left(\frac{s_k^{1-\beta} s_h^{\beta}}{n+g+\delta}\right)^{\frac{1}{1-\alpha-\beta}} \tag{1.51}$$

$$h^* = \left(\frac{s_k^{1-\alpha} s_h^{\alpha}}{n+g+\delta}\right)^{\frac{1}{1-\alpha-\beta}}. \tag{1.52}$$

By substituting these two relationship in the production function and taking logs we have an expression for the pro-capita level of output in steady-state:

$$\ln\left(\frac{Y_t^*}{L_t}\right) = \ln A_0 + gt + \frac{\alpha}{1-\alpha-\beta}\ln\left(s_k\right) + \tag{1.53}$$
$$+\frac{\beta}{1-\alpha-\beta}\ln\left(s_h\right) - \frac{\alpha+\beta}{1-\alpha-\beta}\ln\left(n+g+\delta\right).$$

(1.52) shows how pro-capita output depends on the rate of growth of population, the rate of accumulation of human capital and the rate of accumulation of physical capital. (1.52) nests (1.8), and illustrate how direct estimation of (1.8) might deliver biased estimates of the parameters of interest as a consequence of under-parameterization. MRW construct a proxy for the rate of accumulation of human capital by merging two data-sets to obtain a measure of the percentage of working-age population that is in secondary. They call such variable SCHOOL and they include its logarithm in the regression. In this case even if SCHOOL is only proportional to s_h,it can be safely used in the estimation of the equation of interest as only the constant will be affected. On the other hand if SCHOOL is measured with error, the measurement error will deliver bias in the estimates only if it is correlated with the other regressors.

The results of the estimation of the augmented Solow model are reported in Table 4.

TABLE 4: The estimation of the augmented Solow model

Variable	Coefficient	Std. Error	t-ratio	Prob.
C	4.451	1.153	3.859	0.0002
lns	0.709	0.15	4.725	0.0000
lnngd	-1.497	0.402	-3.719	0.0004
lsch	0.728	0.095	7.666	0.0000
R-squared 0.782	S.E. of regression 0.45			

Note that now all the model-based restrictions on coefficients cannot be rejected and that estimated parameters are compatible with values of about 1/3 for α and β. Such values are deemed to be very reasonable by MRW who conclude that the estimation of the Solow model without human capital can be considered as a benchmark case to illustrate the effect of underparameterization.

1.11 The importance of time-series in macroeconomics

Results obtained thus far are based on a cross-sectional analysis of different countries, without using information on the time-series behaviuor of relevant variables. However, most of the interesting questions in macroeconomics are answered by analyzing the time-series behaviour of variables. The Solow's model predicts that each economies converges to its own steady-state. The obvious implication is that, over-time, differences in per capita output of countries featuring the same rate of capital accumulation and the same rate of growth of population should disappear. The empirical validity of such prediction has been heavily questioned. Recently an alternative theory of growth has developed: the endogenous growth theory (Lucas [5], Romer). Such theory basically modifies the new-classical growth model by introducing constant returns to scale in the production function for output in effective units of labour ($\alpha + \beta = 1$, instead of $\alpha + \beta < 1$). In such type of models the steady-state level of output is not defined[3]. Therefore, differences between countries can persist indefinitely even if

[3]The non-existence of equilibrium generates some problems for the construction of a theory of distribution. These problems are solved by introducing the idea of an aggregate production

the countries share the same rates of accumulation of capital and of growth of population.

If one considers time-series data, then it is very easy to discriminate between the two models on the basis of their different predictions. Consider the i-th country in our sample. If we have $t = 1, ...T$ time series observations on the relevant variables , then we can estimate the λ_i parameter in the following model:

$$\Delta \ln y_{i,t} = \lambda_i \left(\ln y_{i,t-1} - \ln y^*_{i,t-1}\right) + \varepsilon_{it} \quad (1.54)$$
$$\Delta \ln y_{i,t} = \left(\ln y_{i,t} - \ln y_{i,t-1}\right).$$

The new-classical growth model predicts $\lambda_i < 0$, while the endogenous growth model predicts $\lambda_i > 0$.

In fact $\lambda_i < 0$ warrants convergence of y to its steady-state (which it is time-varying in that the rate of accumulation of capital and the rate of population growth might be time-varying). While in the case $\lambda_i > 0$, we do not have convergence of y to its steady-state. The main complication in estimating and testing the model of interest within a time-series context is that the hypothesis $E(\mathbf{u} \mid \mathbf{X}) = 0$ does not hold and the derivation of properties of estimators and statistical distribution for hypotheses testing requires a new appropriate framework, which we will discuss in the following chapter.

To reinforce our point on the importance of time series in evaluating macroeconomic theories we evaluate the loss of information when λ_i is estimated by using the 75 cross-sectional observation used so far.

We can re-write(7.15) as follows:

$$\ln y_{i,t} = (1 + \lambda_i) \ln y_{i,t-1} + \lambda_i \ln y^*_{i,t-1} + \varepsilon_{it} \quad (1.55)$$

by recursively substituting in (1.55) ,we have:

$$\ln y_{i,t} - \ln y_{i,0} = -\left(1 - (1 + \lambda_i)^t\right) \ln y_{i,0} - \lambda_i \ln y^*_i j = 0\sum^{t-1} (1 + \lambda_i)^j + (1.56)$$
$$+ j = 0\sum^{t} (1 + \lambda_i)^j \varepsilon_{i,t-1}$$

which can be re-written as ,

$$\ln y_{i,t} - \ln y_{i,0} = -\left(1 - (1 + \lambda_i)^t\right) \ln y_{i,0} + \left(1 - (1 + \lambda_i)^t\right) \ln y^*_i + v_t (1.57)$$
$$v_t = j = 0\sum^{t} (1 + \lambda_i)^j \varepsilon_{i,t-1}. \quad (1.58)$$

function different from the production function faced by the specific firm, in fact productivity gains at firm levels are not expgenous as they depend on the general level of industrialization of the society (theory of the "learning by doing"). For a very clear discussion of this point see Farmer(1996).

Where in all derivation we have taken the steady-state to be constant. Now by adding to this assumption $\lambda_i = \lambda$, constant across countries. It is possible to estimate (1.57) on our cross section of 75 countries, by taking as dependent variable the difference between initial and final output. MRW do so by taking the difference between output in 1985 and output in 1960. Note that the error time in the cross-sectional model is much larger then the error term in the time-series model, being the cumulation of 25 time series residuals. We report MRW in Table 5. Note that we compare three models: the unconditional convergence model, and two conditional convergence model (Solow and augmented Solow). As expected the results on the estimation of λ change as the specification is changed. The richer model gives a point estimate for λ of -0.02, giving some support to the new-classical model (what is the standard-error associated to this point estimate?)

TABLE 5: Testing convergence (dep.var. lyl85-ly60)

Variable	Unconditional	Solow	Augmented Solow
Constant	0.568 (0.432)	2.26 (0.847)	2.48 (0.795)
lyl60	-0.002 (0.054)	-0.23 (0.056)	-0.36 (0.066)
lns	-	0.65 (0.103)	0.55 (0.101)
lnngd	-	-0.45 (0.304)	-0.54 (0.286)
lnsch	-	-	0.27 (0.079)
R^2	0.00002	0.38	0.47
σ	0.41	0.32	0.30
λ	-0.0078	-0.01	-0.018

1.12 Alternative strategies of research in macroeconometrics

Some final remarks on the research strategy behind the empirical work considered so far can be useful to set out the general framework for the organization of the material in this book. The starting point of the research strategy of Mankiw, Romer and Weil is a theoretical model, the Solow growth model. The estimated empirical relation is derived from the solution of the model. As the estimation of the relation explicitly derived from Solow growth model delivers disappointing results, a modification of the model is considered by introducing human capital in the original framework. Such modification generates satisfactory empirical results and it is capable of explaining the empirical failure of the original specification. At this point the authors have their message and are able to convey it to the profession.

Any empirical research strategy is based on the combination of theoretical analysis and work on the data to produce models of the economies. We have shown that time-series are the most natural empirical counterpart of varaiables in macroeconomic models. In the next chapter we discuss the statistical framework necessary to analyse time-series.

We shall then introduce identification; the crucial stage of research in applied macroeconometrics where theory and statistical analysis of the data are brought together. In fact, the different approaches currently adopted in applied empirical work in macroeconomics could be understood as different solutions to the identification problem. On the basis of the working knowledge of fundamentals built in these two chapters, we shall then consider the different approaches to applied macroeconometrics. We start from the Cowles Commission approach, by discussing a model of the monetary transmission mechanism built on the most famous "ad-hoc" framework (the IS-LM model augmented by some supply function) imposed on the data to ask the time honoured question "what does monetary policy do". Such model is designed to identify the impact of policy variables on macroeconomic quantities. The objective of the exercise is to determine the value to be assigned to the monetary instruments to achieve a given target for the macroeconomic variables. Exogeneity of the policy variables is assumed on the ground that these are the instruments controlled by the policymaker.We illustrate how the model is used by estimating a small odel of the US economy and replicate the empirical failure of the generation of macroeconometric models in achieving the objective of their simulation.

Such failure has been rationalized in different ways, leading to different approaches to replace the Cowles Commission research programme.

The LSE ([3])approach explains the failure of the Cowles Commission methodology by attributing it to the lack of attention for the statistical model underlying the particular econometric structure adopted to analyse the effect of alternative monetary policies. The LSE methodology considers econometric policy evaluation an interesting and feasible exercise. However, the way in which the Cowles Commission approach deals with a legitimate question is not seen as correct. The lack of sufficient interest for the statistical model is interpreted as the root of the failure of the Cowles Commission approach to provide at acceptable answer to an interesting question. The diagnosis is careful diagnostic checking on the specification adopted. By applying the LSE approach to the same problem faced by the Cowles Commission model we shall show merits and limits. On the positive side, we evaluate the improvements on the econometric specification, while, on the negative side, we show why such methodology has not been universally adopted as a unique substitute for the Cowles Commission approach.

Differently from the LSE explanation of traditional structural modelling the two most famous and demolishing critiques, due to Lucas ([36]) and Sims ([48]), concentrate on the weak theoretical basis for the Cowles Commission models. The Lucas critique explains the failure of structural models when the coefficent describing the impact of monetary policy on the macroeconomic variables of interest depend on the monetary policy regimes; in this case no model estimated under a specific regime can be used to simulate the effects of a different monetary policy regime. Such situation is naturally generated when agents behaviour is determined by intertemporal optimization. The Sims critique attacks identification from a different perspective, pointing out that the restrictions needed to support exogeneity in structural Cowles Commision-type models are "incredible" in an environment where agents optimise intertemporally.

The natural outcome of these two critiques is that policy simulation should not be undertaken on the basis of structural econometric models but rather on the basis of simulation of model economies based on microeconomic foundations.

However, econometrics play still an important role for the selection of the appropriate model economy and for the estimation of the deep parameters describing taste and technology and independent from expectations.

The research programme initiated by Sims lead to the estimation of VAR models in empirical macroeconomics. VAR models of the transmission mechanism are not estimated to yield advice on the best monetary policy; they are rather estimated to provide empirical evidence on the response of macroeconomic variables to monetary policy impulses in order to discriminate between alternative theoretical models of the economy. Monetary policy actions should be identified using theory-free restrictions, taking into account the potential endogeneity of policy instruments.

The Generalised Method of Moments is the econometric methodology naturally applied to the first order conditions for the solution of intertemporal optimisation problems to derive estimates of the deep parameters in the economy.

Once deep parameters of interest are estimated, the micro-founded model can be calibrated and the effect of relevant economic policies can then be assessed.

We shall devote three chapters to VAR models, GMM estimation and calibration to illustrate the strategy of empirical research in macroeconomics consistent with the view that policy advice should be based on the simulation of theoretical models considering explicitly the intertemporal optimisation problem of agents.

REFERENCES

Farmer R.(1998) Macroeconomics, South-Western Publishing

Faust J. and C.H.Whiteman(1997) "General-to-specific procedures for fitting a data-admissible, theory-inspired, congruent, parsimonious, encompassing, weakly-exogenous, identified, structural model to the DGP: a translation and a critique." Board of Governors of the Federal Reserve System, International Finance Discussion Papers, 576

Hendry D.H.F.(1996) "Dynamic Econometrics" Oxford University Press, Oxford

Lucas R.E. Jr. (1976) "Econometric Policy Evaluation: A Critique", in K. Brunner and A. Meltzer (eds.) *The Phillips curve and labor markets,* Amsterdam, North-Holland

Lucas R.E.(1988) "On the Mechanics of Economic Development" Journal of Monetary Economics, 22, 3-42

Mankiw G., Romer D., Weil D.(1992) "A Contribution to the Empirics of Economic Growth" , Quarterly Journal of Economics, 408-438

Pesaran M.H. and R.Smith (1995) "The role of theory in econometrics" Journal of Econometrics

Sims C.A. (1980) "Macroeconomics and Reality", *Econometrica*, 48, 1-48

Solow R.(1956) "A Contribution to the Theory of Economic Growth", Quarterly Journal of Economics, 70, 65-94

2

THE PROBABILISTIC STRUCTURE OF TIME SERIES DATA

2.1 Introduction: what is a time-series ?

In the previous chapter we have introduced time-series to show that one of the fundamental properties necessary to perform valid estimation and inference in the linear model is generally violated by time-series. In this chapter we shall discuss this issue at greater depth and length by defining precisely time series and the fundamental concepts to analyze them, by illustrating how the problem introduced can be resolved in the context of stationary time-series, to finally extend our discussion to non-stationarity and cointegration.

We write a time-series as

$$\{x_1, x_2, ...x_T\} \ or \ \{x_t\}, \ t = 1, ...T$$

where t is an index denoting the period in time in which x occurs. We shall treat x_t as a random variables, hence a time-series is a sequence of random variables ordered in time. Such sequence is known as a stochastic process. The probability structure of a sequence of random variables is determined by the joint distribution of a stochastic process.

A possible probability model for such a joint distribution is :

$$x_t = \epsilon_t, \ \epsilon_t \sim n.i.d. \left(0, \sigma_\epsilon^2\right)$$

i.e. x_t is normally independently distributed over time with constant variance and zero mean. In other words x_t is a *white-noise* process. A white-noise process is not a proper model for most macroeconomic time-series because it does not feature their most common characteristic, namely persistence. To show the point consider the data-set USUK.XLS which contains, in EXCEL format, quarterly time series data for nominal and real personal disposable income and consumption in the UK and the US over the sample 1959:1-1998:1. The data-set, retrieved from DATASTREAM, contains nine variables:

TABLE 1: Dataset USUK.XLS

ukpdispid	personal disposable income in the UK at constant 1992 prices
uspdispid	personal disposable income in the US at constant 1992 prices
uscndurb	consumption of durable goods in the US at current prices
uscndurb	consumption of durable goods in the US at constant 1992 prices
uscnnondb	consumption of non-durable goods in the US at current prices
uscnnondd	consumption of non-durable goods in the US at constant 1992 prices
uscnservb	consumption of services in the US at current prices
uscnservd	consumption of services in the US at constant 1992 prices

All series are adjusted for seasonality. To assess the behaviour of an typical economic time series against the benchmark of the white-noise process, we have imported all series in an E-Views workfile and run the following routine:

```
smpl 1959:1 1998:1
genr lyus=log(uspdispid)
genr WN= 8.03+0.36*nrnd
plot WN lyus
```

The routine generates the log of US real disposable income and an artificial series defined as a constant (8.03) plus a normal random variable with zero mean and standard deviation of 0.36, where 8.03 and 0.36 are respectively the sample mean and the sample standard deviation of lyus. Having generated the series the program plots them to obtain the following result:

Figure 2.1 clearly shows that the white noise model does not capture the interesting property of persistence that motivates the study of time series. In order to construct more realistic models combinations of ϵ_t. We shall concentrate on a class of models created by taking linear combinations of white noise, the ARMA models:

$$
\begin{aligned}
AR(1): &\quad x_t = \rho x_{t-1} + \epsilon_t \\
MA(1): &\quad x_t = \epsilon_t + \theta \epsilon_{t-1} \\
AR(p): &\quad x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + ... + \rho_p x_{t-p} + \epsilon_t \\
MA(q): &\quad x_t = \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q} \\
ARMA(p,q): &\quad x_t = \rho_1 x_{t-1} + ... + \rho_p x_{t-p} + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q}
\end{aligned}
$$

In case it is not already clear, we shall show why ARMA models are obtained by taking linear combinations of white noise in the next section, where we discuss the strictly necessary fundamentals to analyze time series.

Note that each of the above models can be easily put to action to generate the equivalent time-series by modifying appropriately and running the following programme in Eviews, which generates an AR(1) series:

```
smpl 1 1
```

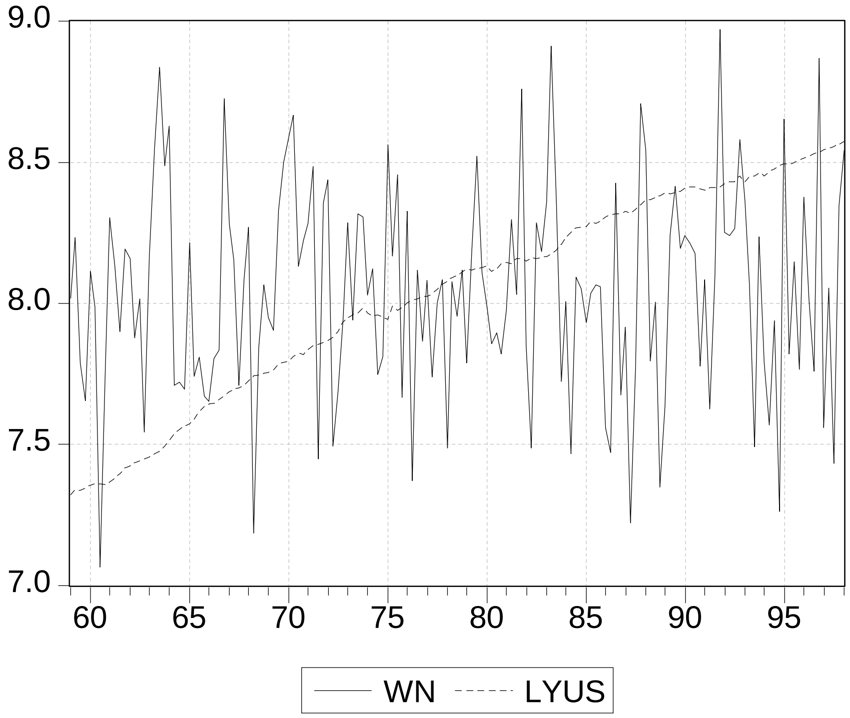


FIG. 2.1. A *white-noise* process and the log of US real disposable income

```
genr X=0
smpl 2 200
series x=0.5*x(-1) +NRND
```

The programme above generates a sample of 200 observations from an AR(1) model with $\rho = 0.5$. The series is first initialized for the first observations, the command series then generates the series for the specified process, each observation is 0.5 time the previous observation plus a random disturbance drawn from a serially independent standard normal distribution.

The time series behaviour of the generated X is plotted in Figure 2.2.

The following modified version of the programme will generate an ARMA(1,1) series:

```
smpl 1 1
genr X=0
smpl 1 200
genr u=NRND
smpl 2 200
```

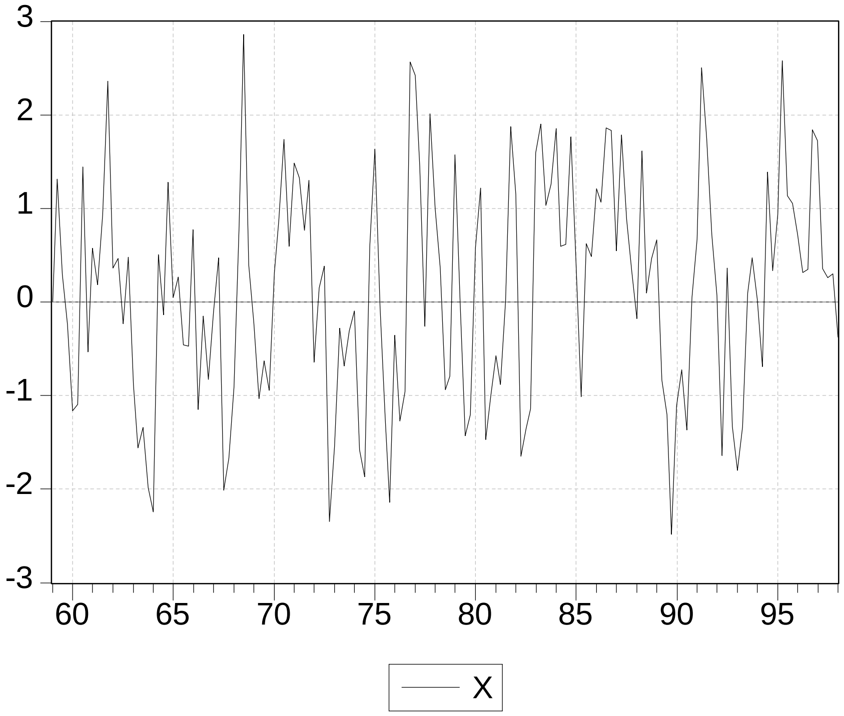


FIG. 2.2. A stationary ARMA(1,1) process

```
series x=0.5*x(-1) +u +0.4*u(-1)
```

2.2 Analyzing time-series: the fundamentals.

To illustrate empirically all the fundamentals we consider an interesting member of the the ARMA family: the AR model with drift :

$$x_t = \rho_0 + \rho_1 x_{t-1} + \epsilon_t \tag{2.1}$$
$$\epsilon_t \sim n.i.d. \left(0, \sigma_\epsilon^2\right)$$

Given that each realization of our stochastic process is a random variable, the first relevant fundamental is the density of each observations. In particular, we distinguish between conditional and unconditional densities. Having introduced these two concepts we shall define and discuss stationarity, we then generalize form our specific member to the whole family of ARMA models, to end this section with a discussion of deterministic and stochastic trends and de-trending methods. Note that at this introductory stage we concentrate almost exclusively on univariate models. We do so just for the sake of exposition. After the completion of our introductory tour, we shall concentrate on multivariate models, which are the focus of this book.

2.2.1 *Conditional and unconditional densities*

We distinguish between conditional and unconditional density of a time-series. The unconditional density is obtained under the hypothesis that no observation on the time series is available, while conditional densities are based on the observation of some realization of the random variables. In the case of the time series we derive unconditional by putting ideally ourselves at the moment in time preceeding the observation of any realization of the time series. At the moment the information set is given only by the knowledge of the process generating the observations. As observations become available conditional densities can be computed. As distributions are summarized by their moments, let us illustrate the difference between conditional and unconditional densities by looking at our AR(1) model.

The moments of the density of x_t conditional upon x_{t-1} are immediately obtained from (8.18)as follows:

$$
\begin{aligned}
E\left(x_t \mid x_{t-1}\right) &= \rho_0 + \rho_1 x_{t-1} \\
Var\left(x_t \mid x_{t-1}\right) &= \sigma_\epsilon^2 \\
Cov\left[\left(x_t \mid x_{t-1}\right), \left(x_{t-j} \mid x_{t-j-1}\right)\right] &= 0 \ \ for\ each\ j
\end{aligned}
$$

To derive the moments of the density of x_t conditional upon x_{t-2}, we need to substitute for x_{t-1} in terms of x_{t-2} from (8.18) to obtain:

$$
\begin{aligned}
E\left(x_t \mid x_{t-2}\right) &= \rho_0 + \rho_0 \rho_1 + \rho_1^2 x_{t-2} \\
Var\left(x_t \mid x_{t-2}\right) &= \sigma_\epsilon^2 \left(1 + \rho_1^2\right) \\
Cov\left[\left(x_t \mid x_{t-2}\right), \left(x_{t-j} \mid x_{t-j-2}\right)\right] &= \rho_1 \sigma_\epsilon^2 \quad for\ j = 1 \\
Cov\left[\left(x_t \mid x_{t-2}\right), \left(x_{t-j} \mid x_{t-j-2}\right)\right] &= 0 \quad for\ j > 1
\end{aligned}
$$

Finally, unconditional moments are derived by substituing recursively from (8.18) to express x_t as a function of information available at time time t_0, the moment before we start observing realizations of our process.

$$
\begin{aligned}
E\left(x_t\right) &= \rho_0 \left(1 + \rho_1 + \rho_1^2 + ...\rho_1^{t-1}\right) + \rho_1^t x_0 \\
Var\left(x_t\right) &= \sigma_\epsilon^2 \left(1 + \rho_1^2 + \rho_1^4 + ...\rho_1^{2t-2}\right) \\
\gamma\left(j\right) &= Cov\left(x_t, x_{t-j}\right) = \rho_1^j \, Var\left(x_t\right) \\
\rho\left(j\right) &= \frac{Cov\left(x_t, x_{t-j}\right)}{\sqrt{Var\left(x_t\right) Var\left(x_{t-1}\right)}} = \frac{\rho_1^j \, Var\left(x_t\right)}{\sqrt{Var\left(x_t\right) Var\left(x_{t-1}\right)}}
\end{aligned}
$$

Note that $\gamma(j)$ and $\rho(j)$ are function of j, known respectively as the autocovariance function and the autocorrelation function.

2.2.2 *Stationarity*

A stochastic process is said to be stricly stationary if its joint density function does not depend on time. More formally a stochastic process is stationary if ,for each $j_1, j_{2,...} j_n$, the joint distribution,

$$f\left(x_t, x_{t+j_1}, x_{t+j_2}, x_{t+j_n}\right)$$

does not depend on t.

A stochastic process is said to be covariance stationary if its two first undiconditional moments do not depend on time, i.e. if the following relations are satisfied for each h,i,j:

$$\begin{aligned} E\left(x_t\right) &= E\left(x_{t+h}\right) = \mu \\ E\left(x_t^2\right) &= E\left(x_{t+h}^2\right) = \mu_2 \\ E\left(x_{t+i} x_{t+j}\right) &= \mu_{ij} \end{aligned}$$

In the case of our AR(1) process the condition for stationarity is that $|\rho_1| < 1$. In fact, when such condition is satisfied we have:

$$\begin{aligned} E\left(x_t\right) &= E\left(x_{t+h}\right) = \frac{\rho_0}{1-\rho_1} \\ Var\left(x_t\right) &= Var\left(x_{t+h}\right) = \frac{\sigma_\epsilon^2}{1-\rho_1^2} \\ Cov\left(x_t, x_{t-j}\right) &= \rho_1^j \, Var\left(x_t\right) \end{aligned}$$

on the other hand it easily shown that, when $|\rho_1| = 1$, the process is non stationary.

In fact we have:

$$\begin{aligned} E\left(x_t\right) &= \rho_0 t + x_0 \\ Var\left(x_t\right) &= \sigma_\epsilon^2 t \\ Cov\left(x_t, x_{t-j}\right) &= \sigma_\epsilon^2 \, (t-j) \end{aligned}$$

To illustrate graphically the properties of different AR process we generate, using the programme in E-views described above, we generate three AR process with ρ_1 set to 0.6 (series X1), 0.8 (series X2), and 1 (series X3) respectively. To allow direct comparison we do not include a drift in all process so for all of them

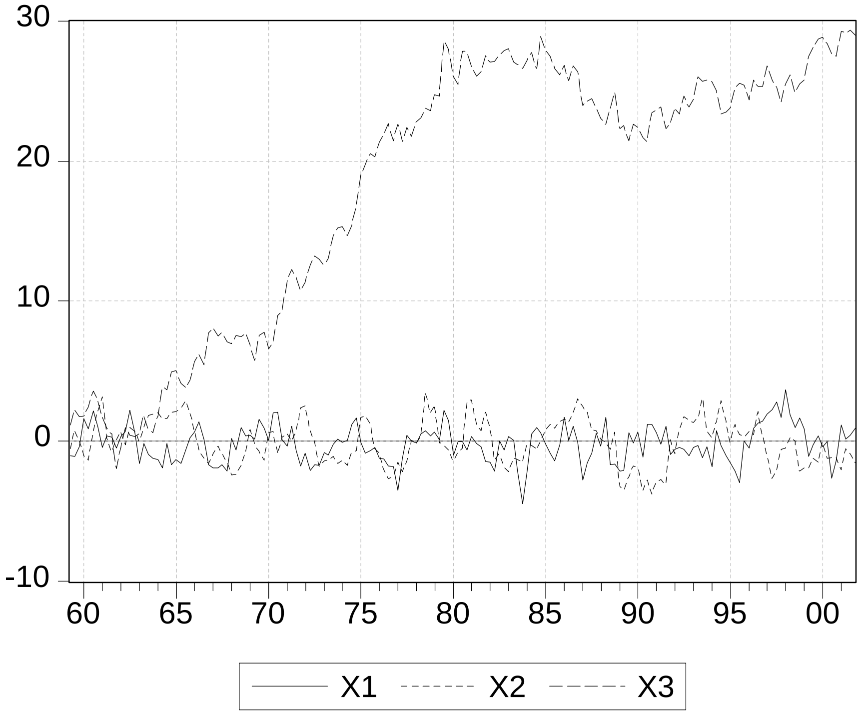


FIG. 2.3. First order autoregressive processes with $\rho_1 = 0.6$ (X1), $\rho_1 = 0.8$ (X2), $\rho_1 = 1$ (X3)

we have $\rho_0 = 0$. The time-series behaviour of the three-processes is reported in Figure 2.3.

Note that X1 and X2 tend to revert towards their unconditional mean rather quickly. The unconditional mean of X3 is also zero but X3 does not show any tendency for reverting towards its mean, in fact, as the sample size grows, the variance of X3 increases without any bound.

2.2.3 *ARMA processes*

Before introducing the fundamentals of time-series we have asserted that white-noise processes were too simplistic to describe economic time series and that a closer fit could be obtained by considering combination of white-noises. we have then introduced ARMA models and discussed the fundamentals to understand their properties, but we have not yet shown that ARMA models can be considered as combination of white-noise processes. The point is shown by considering a time-series as a polynomial distributed lag of a white-noise process:

$$
\begin{aligned}
x_t &= u_t + b_1 u_{t-1} + b_2 u_{t-2} + \ldots b_n u_{t-n} \\
&= \left(1 + b_1 L + b_2 L^2 + \ldots + b_n L^n\right) u_t \\
&= b(L) u_t
\end{aligned}
$$

where L is the lag operator. The Wald-decomposition theorem, which states that any stationary stochastic process could be expressed as the sum of a deterministic component and of a stochastic moving average component warrant generality of our representation. However in general, to describe successfully a time-series, a very high order in the polynomial b(L) is required. This feature can be problematic for estimation, given the usual limitations for sample sizes. This potential problem if the polynomial b(L) can be represented as the ratio of two polynomial of lower order. In this case we have:

$$
\begin{aligned}
x_t &= b(L) u_t \\
&= \frac{a(L)}{c(L)} u_t \\
c(L) x_t &= a(L) u_t
\end{aligned}
\tag{2.2}
$$

(2.2) is an ARMA process. The process is stationary when the roots of $c(L) = 0$ lie outside the unit circle. The MA component is said to be invertible when the roots of $a(L) = 0$ lie outside the unit circle. Invertibility of the MA components allow to represent it as an autoregressive process.

To illustrate how the autocovariance and the autocorrelation functions of an ARMA model are derived, we consider the simplest case: the ARMA(1,1) process:

$$
\begin{aligned}
x_t &= c_1 x_{t-1} + \epsilon_t + a_1 \epsilon_{t-1} \qquad (2.3) \\
(1 - c_1 L) x_t &= (1 + a_1 L) \epsilon_t
\end{aligned}
$$

(2.3) can be re-written as:

$$
\begin{aligned}
x_t &= \frac{1 + a_1 L}{1 - c_1 L} \epsilon_t \\
&= (1 + a_1 L)\left(1 + c_1 L + (c_1 L)^2 + \ldots\right) \epsilon_t \\
&= \left[1 + (a_1 + c_1) L + c_1 (a_1 + c_1) L^2 + c_1^2 (a_1 + c_1) L^3 + \ldots\right] \epsilon_t
\end{aligned}
$$

Then we have

$$
\begin{aligned}
Var(x_t) &= \left[1 + (a_1 + c_1)^2 + c_1^2 (a_1 + c_1)^2 + \ldots\right] \sigma_\epsilon^2 \\
&= \left[1 + \frac{(a_1 + c_1)^2}{1 - c_1^2}\right] \sigma_\epsilon^2
\end{aligned}
$$

$$
\begin{aligned}
Cov\left(x_t, x_{t-1}\right) &= \left[\left(a_1+c_1\right)+c_1\left(a_1+c_1\right)+c_1^2\left(a_1+c_1\right)+\ldots\right]\sigma_\epsilon^2 \\
&= \left[\left(a_1+c_1\right)+\frac{c_1\left(a_1+c_1\right)^2}{1-c_1^2}\right]\sigma_\epsilon^2
\end{aligned}
$$

Hence

$$
\begin{aligned}
\rho(1) &= \frac{Cov\left(x_t, x_{t-1}\right)}{Var\left(x_t\right)} \\
&= \frac{\left(1+a_1 c_1\right)\left(a_1+c_1\right)}{1+c_1^2+2a_1 c_1}
\end{aligned}
$$

Successive values for $\rho(j)$ are obtained from the recurrence relation $\rho(j) = c_1\rho(j-1)$ for $j \geq 2$.

To illustrate the difference between an AR and an ARMA, we have generated an AR(0.7) process and an ARMA (0.7, 0.4) process in E-Views. The two autocorrelation functions (for lags up to 10) are reported in Table 2.

TABLE 2: Autocorrelation functions

AR (0.7)	ARMA (0.7,0.4)
0.712	0.836
0.561	0.639
0.437	0.491
0.304	0.364
0.254	0.305
0.270	0.305
0.270	0.313
0.298	0.326
0.279	0.323
0.296	0.316

Note that the autocorrelation of the ARMA(1,1) process is higher than the autocorrelation of the AR(1) process, this is because $a_1 > 0$.

2.2.4 *Deterministic and Stochastic Trends*

Figure 2.1 at the beginning of this chapter shows that macroeconomic time series, beside being persistent, feature (generally) upwarding trends. Non-stationarity of time-series is a possible manifestation of a trend. Consider for example the random walk with drift:

$$x_t = a_0 + x_{t-1} + \epsilon_t$$
$$\epsilon_t \sim n.i.d.\left(0, \sigma_\epsilon^2\right)$$

In this case recursive substitution yields:

$$x_t = x_0 + a_0 t + i = 0 \sum^{t-1} \epsilon_{t-i} \tag{2.4}$$

which shows that the non-stationary series contains both a deterministic $(a_0 t)$ and a stochastic$\left(i = 0 \sum^{t-1} \epsilon_{t-i} \right)$ trend.

One of the easiest way to make a non-stationary series stationary is by differencing it:

$$\Delta x_t = x_t - x_{t-1} = (1 - L)\, x_t = a_0 + \epsilon_t$$

In general if a time series needs to be differenced k times to be stationary, then that series is said to be integrated of order k or $I(k)$. Our random walk is I(1).When the $d - th$ difference of a time-series x, $\Delta^d x_t$, can be represented by an ARMA(p, q) model we say that x_t is an integrated moving-average process of order p, d, q and we denote it as ARIMA(p, d, q).

It interesting to compare the behaviour of integrated process with that of trend stationary process. Trend stationary processes feature only a deterministic trend:

$$z_t = \alpha + \beta t + \epsilon_t \tag{2.5}$$

The z_t process is non-stationary, but the non-stationary is removed just by regressing z_t on a deterministic trend. This is not the case for integrated processes like (5.26) where the removal of the deterministic trend does not deliver a stationary time-series. Deterministic trend have no memory while integrated variables have infinite memory. Both integrated variable and deterministic trend exhibits systematic variations, but in one case the variation is predictable in the other case it is not. This point is easily seen in Figure 2.4 where we report three series for a sample of 200 observations. The series are generated in E-views by running the following programme:

```
smpl 1 1
genr ST1=0
genr ST2=0
smpl 2 200
```

```
series ST1= 0.1+ST1(-1) +nrnd
series ST2=0.1+ST2(-1)+nrnd
series DT= 0.1*@trend +nrnd
```

We have a deterministic trend (DT) generated by simulating equation (2.5) with $\alpha = 0, \beta = 0.1$, and a white-noise independently distributed as a standard normal (nrnd), and two integrated series (ST1 and ST2), which are random walks with a drift of 0.1. The only difference between ST1 and ST2 is in the realizations from the error terms, which are different drawings from the same serially independent standard normal distribution.

FIG. 2.4. Deterministic (DT) and stochastic (ST1 and ST2) trends

2.3 Persistence. A Monte-Carlo experiment

Persistence of time-series destroys one of the crucial properties to implement valid estimation and inference in the linear model. We have already seen that in the context of the linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

The following property is required to implement valid estimation aand inference

$$E(\epsilon \mid \mathbf{X}) = \mathbf{0} \tag{2.6}$$

Hypothesis (2.6) implies that

$$E(\epsilon_i \mid \mathbf{x}_1, ...\mathbf{x}_i, ...\mathbf{x}_n) = 0 \quad (i = 1, ...n)$$

Think of the simplest time-series model for a generic variable y :

$$y_t = a_0 + a_1 y_{t-1} + \epsilon_t$$

It is clear that if $a_1 \neq 0$, then, although it is true that $E(\epsilon_t \mid y_{t-1}) = 0$, $E(\epsilon_{t-1} \mid y_{t-1}) \neq 0$ and (2.6)is destroyed.

The question is how serious is the problem. To assess intuitively the consequence of persistence we construct a small Monte-Carlo simulation on the short sample properties of the OLS estimator of the parameters in an AR(1) process.

A Monte-Carlo simulation is based on the generation of a sample from a known Data Generating Process(DGP). A set of random numbers from a given distribution is generated first (a normally independent white-noise disturbance in our case) for a sample size of interest (in our case 200 observations) and then the process of interest is constructed (in our case an AR(1) process). When a sample of observations on the process of interest is available, then the relevant parameters can be estimated and their fitted value can be compared with the known true value. For this reason the Monte-Carlo simulation is a sort of controlled experiment. The only potential problem with this procedure is that the set of random numbers drawn is just one possible outcome and the estimates are dependent on the sequence of simulated white-noise residuals. To overcome this problem in a Monte-Carlo study the DGP is replicated many times. For each replication a set of estimates is obtained and then averages across replications of the estimated parameters are computed to be assessed against the known true values.

Our Monte-Carlo simulation is performed by running the following programme in E-Views:

```
genr a1sum=0
for !i=1 to 500
smpl 1 1
genr y{!i}=10
smpl 2 200
series y{!i}=1+0.9*y{!i}(-1) +nrnd
equation eq.ls y{!i}= c(1)+c(2)*y{!i}(-1)
eq.rls(c,s)
```

```
genr a1sum=a1sum+R_c2
next
genr a1mean=a1sum/500
```

The first line of the programme generate a series to store the values of the estimated a_1 in each replication. In the next step we set a counter to keep track of the replications (in the specific case we have 500 of them).The loop for the five hundred replications is then set. In each replications a sample of two hundred observations from an AR(1) is generated and then the autoregressive parameters is estimated. Note that such estimation is performed recursively starting with a sample of five observations and then by adding one observation at the time until the last one. The series of these estimates is stored at each replications with the command eq.rls(c,s). At the end of all replications we have 500 hundred series each containing a series of 195 estimated parameters (the first being the parameter estimated on the sample 1-5, the second being the parameter estimated on the sample 1-6, the last one being the parameter estimated on the full sample). We report the average across replications in Figure 2.5.

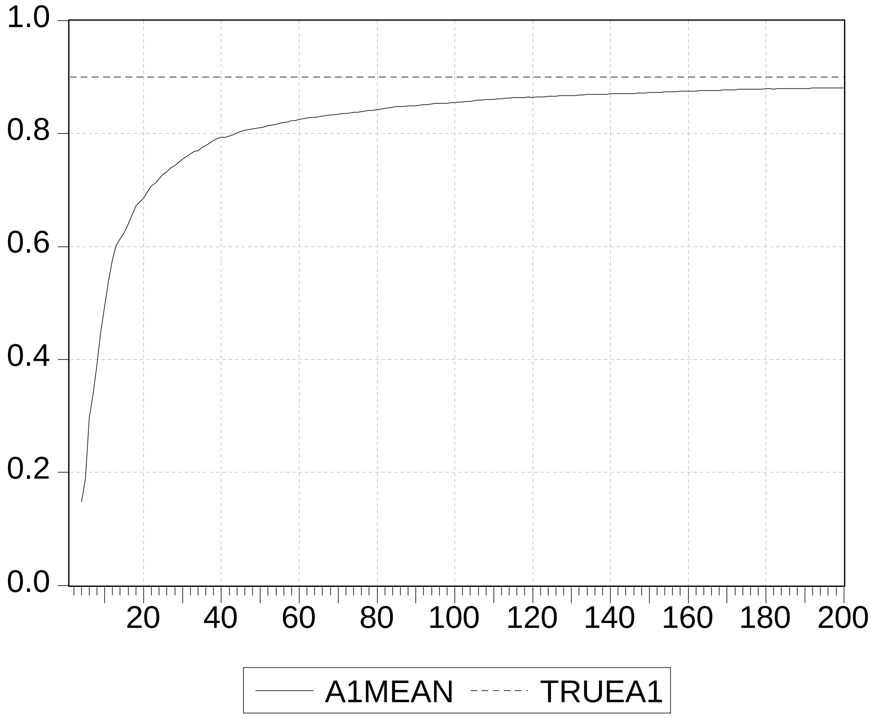


FIG. 2.5. Small sample bias

From the Figure 2.5 we note that the estimate of a_1 is heavily biased in small samples, but the bias is reduced as the sample gets larger to eventually disappear. In fact, it can be shown analytically that the average of the OLS estimate of a_1 is $a_1\left(1-\frac{2}{T}\right)$. This is an interesting result, which could be generalized. For stationary time-series, the correlation, which destroys the orthogonality between residuals and regressors in the linear regression model, tends to disappear as the distance between observations increases. Therefore, as we shall show in the next section, the finite sample results can be extended to time-series by considering large samples. Such aim is obtained by introducing asymptotic theory.

2.4 The traditional solution: asymptotic theory

Stationary time-series feature time-independent distributions, as a consequence the effect of any specific innovation disappear as time elapses. We shall show in this section of the intuition given by the simple Monte-Carlo simultaion can be extended and asymtptotic theory can be used to perform valid estimation and inference when modelling *stationary* time-series.

2.4.1 *Basic elements of asymptotic theory*

In this section we shall introduce the elements of asymptotyc theory necessary to illustrate how all the results in estimation and inference for the linear model applied to cross-sectional data in Chapter 1 can be extended to time-series models[4].

Consider a sequence $\{X_T\}$ of random variables with the associated sequence of distribution functions $\{F_T\} = F_1, ..., F_T$, we give the following definitions of convergence for X_T

2.4.1.1 *Convergence in distribution* Given a random variable X with distribution function F, X_T converges in distribution to X if the following equality is satisfied:

$$T \rightarrow \infty \lim pr\left\{X_T < x_0\right\} = pr\left\{X < x_0\right\}$$

for all x_0,where the function $F(x)$ is continuous.

2.4.1.2 *Convergence in probability* Given a random variable X with distribution function F, X_T converges in probability to X if , for each $\epsilon > 0$,the following relation holds:

$$T \rightarrow \infty \lim pr\left\{|X_T - X| < \epsilon\right\} = 1$$

Note that convegence in probability implies convergence in distribution.

[4]For a formal treatment of all these topics see White([60])

2.4.1.3 *Central limit theorem (formulation of Lindeberg-Levy)* Given a sequence $\{X_T\}$ of identically and independently distributed random variables with mean μ and finite variance σ^2, defining

$$\bar{X} = \frac{1}{T} i = 1 \sum^{T} X_i$$

$$\omega = \sqrt{T} \frac{\left(\bar{X} - \mu\right)}{\sigma}$$

ω converges in distribution to a standard normal.

2.4.1.4 *Slutsky's Theorem* For any random variable X_T such that $plim X_T = a$, where a is a constant, given a function $g(\cdot)$ continuous in a, we have that $p \lim g\left(X_T\right) = g\left(a\right).$

2.4.1.5 *Cramer's Theorem* Given two random variables X_T and Y_T such that Y_T converges in distribution to Y and X_T converges in probability to a constant a, the two following relationships hold:

- $X_T + Y_T$ converges in distribution to $(a + Y)$
- Y_T / a_T converges in distribution to (Y/a)
- $Y_T \cdot a_T$ converges in distribution to $(Y \cdot a)$

Note that all theorems introduced so far are extended to vectors of random variables.

2.4.1.6 *Mann-Wald Theorem* Consider a vector $\mathbf{z}_t$ $(kx1)$ of random variables which satisfies the following property:

$$p \lim T^{-1} i = 1 \sum^{T} \mathbf{z}_t \mathbf{z}_t' = \mathbf{Q}$$

where $\mathbf{Q}$ is a positive definite matrix. Consider also a sequence ϵ_t of random variables identically and independently distributed with zero mean and finite variance σ^2, for which finite moments of each order are defined. If $E\left(\mathbf{z}_t \epsilon_t\right) = 0$,then we have

$$p \lim T^{-1} i = 1 \sum^{T} \mathbf{z}_t \epsilon_t = \mathbf{0}, \sqrt{\frac{1}{T}} i = 1 \sum^{T} \mathbf{z}_t \epsilon_t \xrightarrow{d} N\left(0, \sigma^2 \mathbf{Q}\right)$$

2.4.2 *Application to models for stationary time-series*

Consider the following time-series model:

$$y_t = \alpha y_{t-1} + \beta x_t + u_t$$

where x_t is a stationary variable and $|\alpha| < 1$. As already shown $E\left(y_t u_{t-i}\right) \neq 0$ and the OLS estimator of α is biased.

Re-write the model as :

$$\begin{aligned} y_t &= \mathbf{z}_t \boldsymbol{\gamma} + u_t \\ \mathbf{z}_t &= \begin{bmatrix} y_{t-1} & x_t \end{bmatrix} \\ \boldsymbol{\gamma} &= \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \end{aligned}$$

By applying the Mann-Wald results we can derive the asymptotic distribution of the OLS estimator of $\gamma, \widehat{\gamma}$:

$$\widehat{\gamma} \xrightarrow{d} N\left[\gamma, \sigma^2 \mathbf{Q}^{-1}\right]$$

and all the finite sample results available for cross-section can be extended to stationary time-series just by considering large-sample theory.

2.5 Stochastic-trends and spurious regressions

From what we have discussed so far it should be clear that most econometric analysis is based on the variance and covariance among variables. In the case of indepedent sampling (cross-section) we can use finite sample moments for estimation and inference, in the case of stationary time-series the consideration of moments in large samples can solve the problems peculiar to time-series in small samples. Within this framework it should be immediately clear that non-stationary causes problems. In fact, we know unconditional moments are not defined for non-stationary time-series. Consider, for the sake of illustration, an OLS regression of an I(0) variable y_t on an I(1) variable x_t. The OLS estimator of the regression y_t on x_t converges to zero as the sample size increases, in fact the variance of x_t, being divergent, dominates the covariance between the two variables. In general asymptotic theory is not applicable to non-stationary time-series (see, for example, Hatanaka([23]) and Maddala-Kim([39]) . So, unless all the trends observed in time-series are deterministic, the solution of reverting to asymptotic theory is not directly accessible.

To give an intuition of the importance of non-stationarity in time-series and to illustrate the problems related to non-stationarity, consider the results of a "crazy" regression, obtained by relating the log of consumption in the US to the log of personal disposable income in the UK :

TABLE 3: Regressing US consumption on UK disposable income
Sample : 1959 : 1 1998 : 1 , *Dependent Variable LCUS*

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-5.612676	0.160374	-34.99740	0.0000
LYUK	1.208592	0.014419	83.81657	0.0000

R-squared 0.978413, S.E. of regression 0.052291, DW-stat 0.140469

Note that the regression features a very high R^2 and the UK disposable income is very significant in explaining US consumption. We have a case of a spurious regression, which witnesses the relevance of non-stationarity in economic time-series. To elaborate on this point consider the two following simple univariate time-series models for LYUS and LYUK.

TABLE 4: Univariate Time-series models for US consumption and UK disposable income

	Coefficient	Std. Error	t-Statistic	Prob.
Dependent variable LCUS				
C	0.039	0.008	4.91	0.0000
LCUS(-1)	0.996	0.001	964.9	0.0000
R-squared 0.999835, S.E. of regr 0.004537, DW stat 1.397403.				
Dependent variable LCUS				
C	0.050	0.049	1.00	0.3185
LYUK(-1)	0.996	0.004	222	0.0000
R-squared 0.999835, S.E. of regr 0.004537, DW stat 1.397403.				

despite the simplicity of the two time-series models for LYUS and LYUK, we note that they can both be approximated by random walk models:

$$
\begin{aligned}
LCUS_t &= a_0 + LCUS_{t-1} + \epsilon_{1t} \\
LYUK_t &= b_0 + LYUK_{t-1} + \epsilon_{2t} \\
\epsilon_{1t} &\sim n.i.d.\left(0, \sigma_{\epsilon_1}^2\right) \\
\epsilon_{2t} &\sim n.i.d.\left(0, \sigma_{\epsilon_2}^2\right)
\end{aligned}
$$

As we already know, recursive substitution yields:

$$LCUS_t = LCUS_0 + a_0 t + i = 0\sum^{t-1} \epsilon_{1t-i}$$
$$LYUK_t = LYUK_0 + b_0 t + i = 0\sum^{t-1} \epsilon_{2t-i}$$

When the following model is estimated

$$LCUS_t = \widehat{\alpha} + \widehat{\beta}\ LYUK_t + \widehat{u}_t,$$

the coefficient $\widehat{\beta}$ is significant as both series have a deterministic trend. However, in order to have a non-spurious relation we would need that the regression removes also the stochastic trend from the dependent variables, leaving stationary residuals. If this does not happen, then the correlation we observe can be labelled as spurious. We report in Figure 2.6 the residuals from the OLS regression of LCUS on LYUK,

visual impression confirms the intuition that the regression has delivered a spurious relation, having not removed the stochastic trend form the non-stationary dependent variable. The reported DW statistic of 0.14 gives a more formal background to the visual impression. In fact the Durbin-Watson statistic, originally designed to test for the presence of first order autocorrelation in the residuals, can be re-calibrated to test for stationarity. We have

$$DW = \frac{i = 2\sum^{T} \left(\widehat{u}_t - \widehat{u}_{t-1}\right)^2}{i = 2\sum^{T} \widehat{u}_t} \simeq 2\left(1 - \widehat{\rho}\right)$$

where $\widehat{\rho}$ is the OLS coefficient from the regression of $\widehat{u}_t$ on $\widehat{u}_{tt-1}$. The test was originally tabulated to test the hypothesis $\mathrm{H}_0 : \rho = 0$, but critical values for the null of non-stationarity $\mathrm{H}_0 : \rho = 1$, have been provided by Sargan-Bhargava([51]). According to such critical values the null of non-stationarity cannot be rejected by an observed value of 0.14 for the DW statistic.

In conclusion we note that non-stationarity of time-series is problematic in that it might generate spurious regression and it does not allow the use of standard large-sample theory for valid estimation and inference in the linear model. Before considering the solutions to the problem we shall in the section clarify it further by re-illustrting it form a different perspective.

2.5.1 *Non-stationarity and the likelihood function.*

Consider a vector $\mathbf{x}_t$ containing observations an time series variables at time t. A sample of T time series observations on all the variables can be represented as follows:

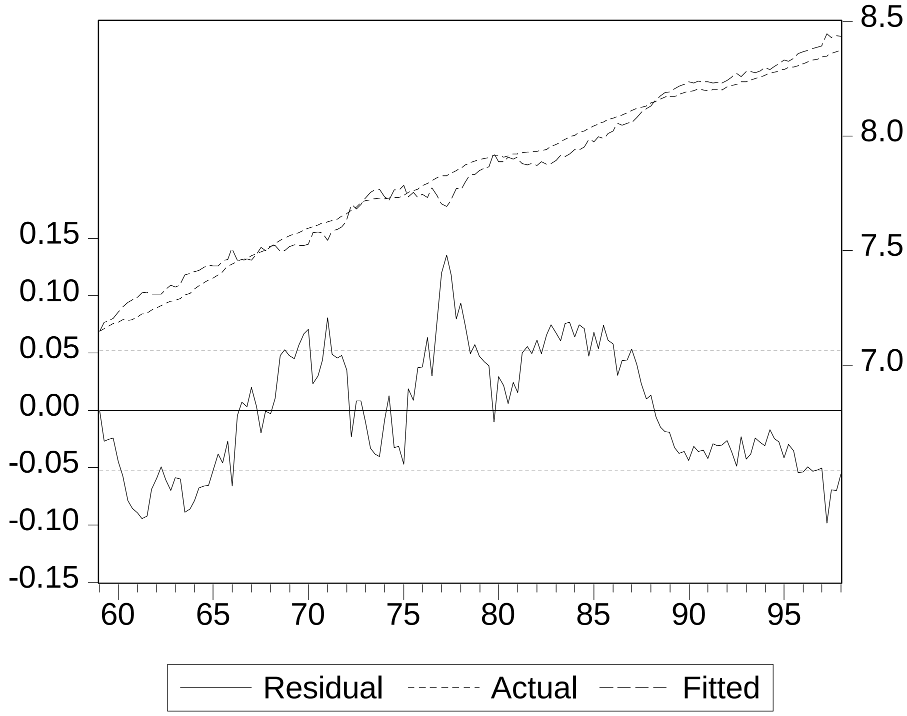


FIG. 2.6. A spurious regression

$$\mathbf{X}_T^1 = \begin{bmatrix} \mathbf{x}_1 \\ . \\ . \\ . \\ \mathbf{x}_T \end{bmatrix}.$$

In general, estimation is performed by considering the joint sample density function, known also as the likelilihood function, which we can express as $\mathrm{D}\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right)$. The likelihood function is defined on the parameters space Θ, given the observation of the observed sample $\mathbf{X}_T^1$ and of a set of initial conditions $\mathbf{X}_{0.}$ Such initial conditions can be interpreted as the pre-sample observations on the relevant variables (which are usually not available). In case of independent observations the likelihood function can be written as the product of the density functions for each observation. However this is not the relevant case for time-series, as time-series observations are in general sequentially correlated. In the case of time-series the sample density is then constructed using the concept of sequential conditioning.

The likelihood function, conditioned with respect to initial conditions, can always be written as the product of a marginal density and a conditional density as follows:

$$D\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) = D\left(\mathbf{x}_1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) D\left(\mathbf{X}_T^2 \mid \mathbf{X}_1, \boldsymbol{\theta}\right).$$

Obviously we also have

$$D\left(\mathbf{X}_T^2 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) = D\left(\mathbf{x}_2 \mid \mathbf{X}_1, \boldsymbol{\theta}\right) D\left(\mathbf{X}_T^3 \mid \mathbf{X}_2, \boldsymbol{\theta}\right)$$

and, by recursive substitution, we eventually obtain :

$$D\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) = t = 1\prod^{T} D\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\theta}\right).$$

Having obtained $D\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right)$we can in theory derive $D\left(\mathbf{X}_T^1, \boldsymbol{\theta}\right)$ by integrating with respect to X_0 the density conditional on pre-sample observations. In practice this could be not tractable analitically as $D\left(X_0\right)$ is not known. The hypothesis of stationarity becomes crucial at this stage, as stationarity restricts the memory of time series and limits to the first observations in the sample the effects of pre-sample observations. This is the reason why, in the case of stationary processes, initial conditions can be simply ignored. Clearly the larger the sample, the better, as the weight of the information lost becomes smaller. Moreover note also that, even by omitting initial conditions we have:

$$D\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) = D\left(\mathbf{x}_1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) t = 2\prod^{T} D\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\theta}\right).$$

Therefore the likelihood function is separated in the product on $T-1$ conditional distribution and one unconditional distribution. In the case of non-stationarity the unconditional distribution is not defined. On the other hand, in the case of stationarity the DGP is completely described by the conditional density function $D\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\theta}\right)$.

2.5.1.1 *An illustration: the first order autoregressive process* To give more empirical content to our case, let us consider again the case of the univariate first order autoregressive process.

$$x_t \mid \mathbf{X}_{t-1} \sim N\left(\lambda x_{t-1}, \sigma^2\right) \tag{2.7}$$

$$D\left(\mathbf{X}_T^1 \mid \lambda, \sigma^2\right) = D\left(x_1 \mid \lambda, \sigma^2\right) t = 2\prod^{T} D\left(x_t \mid \mathbf{X}_{t-1}, \lambda, \sigma^2\right). \tag{2.8}$$

From (2.8) it is clear that the likelihood function involves $T-1$ conditional densities and one unconditional densities. The conditional densities are given by

(2.7) , the unconditional density can be derived only in the case of stationarity. In fact given :

$$x_t = \lambda x_{t-1} + u_t$$
$$u_t \sim N.I.D\left(0, \sigma^2\right),$$

we can obtain by recursive substitution:

$$x_t = u_t + \lambda u_{t-1} + ... + \lambda^{n-1} u_1 + \lambda^n x_0.$$

And only if $|\lambda| < 1$, the effect of the initial condition disappear and we can write the unconditional density of x_t as:

$$D\left(x_t \mid \lambda, \sigma^2\right) = N\left(0, \frac{\sigma^2}{1-\lambda^2}\right).$$

There under stationarity we can write down the exact likelihood function as:

$$D\left(\mathbf{X}_T^1 \mid \lambda, \sigma^2\right) = (2\pi)^{-\frac{T}{2}} \sigma^{-T} \left(1-\lambda^2\right)^{\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2}\left(\left(1-\lambda^2\right) x_1^2 + t = 2\sum^{T} \left(x_t - \lambda x_{t-1}\right)^2\right)\right] \tag{2.9}$$

and estimates of the parameters of interest are derived by maximizing this function. Note that $\widehat{\lambda}$ cannot be derived by analytical methods using the exact likelihood function,but it requires conditioning the likelihood and operating a grid search. Note also that the idea of using in large sample the approximate likelihood function by dropping the first observation works only under the hypothesis of stationarity in a large samples. When the first observation is dropped and the approximate likelihood fnction is considered, it can be shown analytically that the ML estimate of λ coincides with the OLS estimate.

2.6 Univariate decompositions of time-series

The general solution proposed to the problem introduced in the previous section is the search for a stationary representation of non-stationary time-series. This has been done both in an univariate and in a multivariate framework. As an introduction we shall briefly discuss methodologies used in a uni-variate framework to move swiftly to decompositions in a multivariate framework, which are at the heart of our discussion of modern macroeconometrics.

Beveridge-Nelson (1981) provide an elegant way of decomposing a non-stationary time-series into a permanent component and a temporary, cyclical,component by applying ARIMA methods. For any non-stationary time-series x_t integrated of the first order the Wold decomposition theorem could be applied to its first difference, to deliver the following representation:

$$\Delta x_t = \mu + C(L)\,\epsilon_t$$
$$\epsilon_t \sim n.i.d.\left(0, \sigma_\epsilon^2\right)$$

where $C(L)$ is a polynomial of order q in the lag operator. Consider now the polynomial $D(L)$ defined as follows:

$$D(L) = C(L) - C(1) \tag{2.10}$$

given that $C(1)$ is a constant, also $D(L)$ will be of order q. It can immediately be seen that

$$D(1) = 0$$

therefore 1 is a root of $D(L)$, and we can write

$$D(L) = C^*(L)(1 - L) \tag{2.11}$$

where $C^*(L)$ is a polynomial of order $q - 1$.
By equating (2.10) to (2.11) we have:

$$C(L) = C^*(L)(1 - L) + C(1)$$

and

$$\Delta x_t = \mu + C^*(L)\,\Delta\epsilon_t + C(1)\,\epsilon_t \tag{2.12}$$

by integrating (2.12) we finally have:

$$x_t = C^*(L)\,\epsilon_t + \mu t + C(1)\,z_t$$
$$= C_t + TR_t$$

where z_t is a process for which we have $\Delta z_t = \epsilon_t$. C_t is the cyclical component and TR_t is the trend component made of a deterministic trend and a stochastic trend. Note that the trend component can be represented as follows:

$$TR_t = TR_{t-1} + \mu + C(1)\,\epsilon_t.$$

2.6.1 *Beveridge-Nelson decomposition of an IMA(1,1) process*
Consider the process:

$$\Delta x_t = \epsilon_t + \theta \epsilon_{t-1}, \quad 0 < \theta < 1.$$

In this case we have:

$$C(L) = 1 + \theta L$$

$$C(1) = 1 + \theta$$

$$\begin{aligned} C^*(L) &= \frac{C(L) - C(1)}{1 - L} \\ &= -\theta. \end{aligned}$$

The BN decomposition gives the following result:

$$\begin{aligned} x_t &= C_t + TR_t \\ &= -\theta \epsilon_t + (1 + \theta) z_t. \end{aligned}$$

2.6.2 *Beveridge-Nelson decomposition of an ARIMA(1,1) process*
Consider the process:

$$\Delta x_t = \rho \Delta x_{t-1} + \epsilon_t + \theta \epsilon_{t-1}$$

In this case we have

$$\begin{aligned} C(L) &= \frac{1 + \theta L}{1 - \rho L} \\ C(1) &= \frac{1 + \theta}{1 - \rho} \\ C^*(L) &= \frac{C(L) - C(1)}{1 - L} \\ &= -\frac{\theta + \rho}{(1 - \rho)(1 - \rho L)} \end{aligned}$$

and the BN decomposition gives the following result:

$$\begin{aligned} x_t &= C_t + TR_t \\ &= -\frac{\theta + \rho}{(1 - \rho)(1 - \rho L)} \epsilon_t + \frac{1 + \theta}{1 - \rho} z_t \end{aligned}$$

2.6.3 *Deriving the Beveridge-Nelson decomposition in practice*

The practical derivation of a BN decomposition for any ARIMA process is easily derived by applying a methodology suggested by Cuddington and Winters([6]). For any I(1) process, we have seen that the stochastic trend can be represented as follows:

$$TR_t = TR_{t-1} + \mu + C(1)\,\epsilon_t \tag{2.13}$$

The decomposition can then be applied by the following steps:

- identify the appropriate ARIMA model and estimate ϵ_t and all the parameters in μ and $C(1)$ and
- given an initial values for TR_0 use (2.13) to generate the permanent component of the time-series
- generate the cyclical component as the difference between the observed value in each period and the permanent component

The above procedure will give the permanent component up to constant, if the precision of this procedure isnot satisfactory, one can use further conditions to identify more precisely the decomposition. For example one can impose the condition that the sample mean of the cyclical component is zero to pin down the constant in the permanent component.

To illustrate how the procedure works in practice we have simulated an ARIMA(1,1,1) in E-Views for a sample of 200 observations, by running the following programme:

```
smpl 1 2
genr x=0
smpl 1 200
genr u=nrnd
smpl 3 200
series x= x(-1) +0.6*x(-1)-0.6*x(-2) +u+0.5*u(-1)
```

From the previous section we know the exact BN decomposition of our x_t :

$$\begin{aligned} x_t &= C_t + TR_t \\ &= -\frac{1.1}{(1-0.6)(1-0.6L)}\epsilon_t + \frac{1.5}{0.4}z_t \\ TR_t &= TR_{t-1} + \frac{1.5}{0.4}\epsilon_t \end{aligned}$$

we can therefore generate the permanent component of X and the transitory component as follows:

```
smpl 1 2
```

```
genr p=0
smpl 3 200
series TR= TR(-1)+(1.5/0.4)*u
genr CYCLE=X-TR
```

The series X, TR and CYCLE are reported in Figure 2.7.

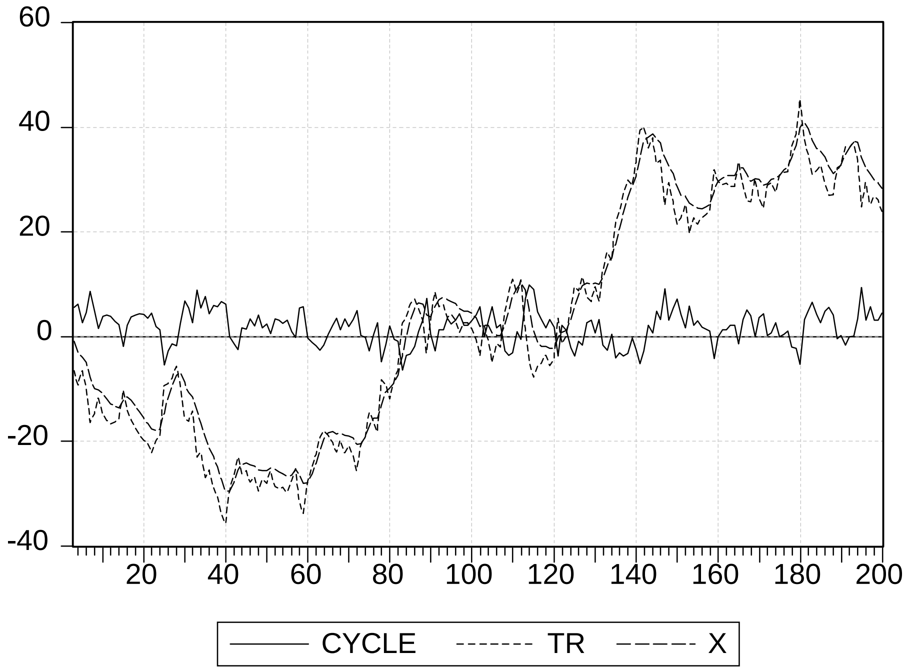


FIG. 2.7. A Beveridge-Nelson decomposition of an ARIMA(1,1,1) process

This is exactly the procedure that we follow in practice except that we estimate parameters rather than impute them from the known DGP.

2.6.4 *Assessing the Beveridge-Nelson decomposition*

The properties of the permanent and temporary component of an integrated time-series delivered by the BN decomposition are worth some comments. The innovations in the permanent and the transitory components are perfectly negatively correlated, moreover the trend component is more volatile than the actual time series as the negative correlation between the permanent and the transitory component acts to smooth the original time-series. These results are easily seen for the simplest case we have already discussed. For example in the case of the

IMA(1,1) process the correlation between the innovations in the permanent and transitory component is -$E_t(1.5\epsilon_t 0.5\epsilon_t) = 1$, the variance of the innovation in trend component is $(1.5)^2 \sigma_\epsilon^2 > \sigma_\epsilon^2$. Note that in general the variance of innovation might have economic interpretation and economic theory might suggests different pattern of correlations between innovations from a perfect negative correlation. As we shall see in one of the next chapters, an interesting pattern could be the absence of correlation between the innovation in the cycle and trend component of an integrated time-series. In general, different restrictions on the correlation between the trend and cycle components lead to the identification of different stochastic trends for integrated time-series. As a consequence the Beveridge-Nelson decomposition is not unique. In general uni-variate decompositions are not unique. To see this point more explicitly we can compare the BN trend with the trend extracted using an alternative technique which has been recently very successful in time-series analysis: The Hodrick-Prescott filter.

Hodrick and Prescott proposed their method to analyze postwar U.S. business cycles in a working paper circulated in the early 1980s and published in 1997([27]). The Hodrick-Prescott (HP) filter computes the permanent component TR_t of a series x_t by minimizing the variance of x_t around TR_t, subject to a penalty that constrains the second difference of TR_t . That is, the HP filter is derived by minimizing the following expression:

$$t = 1\sum^{T} (x_t - TR_t)^2 + \lambda\, t = 2\sum^{T-1} \left[(TR_{t+1} - TR_t)^2 - (TR_t - TR_{t-1})^2 \right].$$

The penalty parameter λ controls the smoothness of the series, by controlling the ratio of the variance of the cyclical component to the variance of the series. The larger the λ, the smoother the TR_t approaches a linear trend as λ goes to infinite. In practical applications λ is set to 100 for annual data, 1600 for quarterly data and 14400 for monthly data.

In the following Figure we report the BN trend and the HP trend (with $\lambda = 100$) for the data generated in the previous section.

Note that the BN trend is more volatile than the HP trend. It is possible to increase the volatility of the HP trend by reducing the parameter λ, however the HP filter can reach at most the volatility of the actual time series which, as we already know, is smaller than the volatility of the BN trend.

The HP filter has the advantage of removing the same trend from all time series; this might be desirable as some theoretical models, as for example real business cycle models, indicate that macroeconomic variables share the same stochastic trend. However, it has been shown by Harvey and Jaeger([22])that the use of such filter can lead to the identification of spurious cyclical behaviour. In fact the two authors above predicate a different approach to modelling time-series, known as structural time series modelling, which, we do not consider in our analysis as it is not related to macroeconomic models, but certainly merits

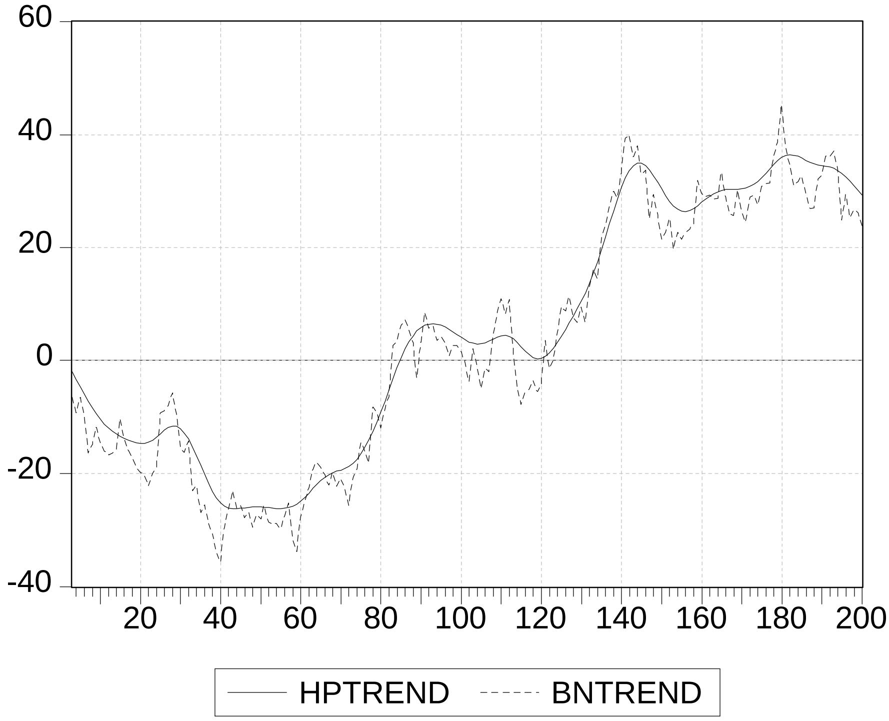


FIG. 2.8. Trend components: Hodrick-Prescott versus Beveridge-Nelson

some attention.[5]

The comparison between the HP and the BN trend reinforces the argument of non-uniqueness of univariate decomposition made before, moreover we are left with the problem of how to use the filtered series in applied macroeconometrics and how to relate them to theoretical models. The empirical counterparts of theoretical macroeconomic models are multivariate time-series. Theoretical models often predict that different time-series share the same stochastic trend. The natural question at this point is if it is possible that the problem of non-stationarity in time-series could be resolved by considering multivariate models. In this context, stationarity is obtained by considering combination of non-stationary time series sharing the same stochastic trends. If such results could be achieved, it would be in principle possible to justify the identification of trends by relating them to macroeconomic theory. We shall consider this possibility in the next sections.

[5]We refer the interested reader to the work by Andrew Harvey and Augustin Maravall.

2.7 Multivariate decompositions and dynamic models

Let us reconsider our spurious regression for US consumption in the context of a dynamic model. We do so by augmenting the static regression to consider consumption and income lagged up to one year, i.e. we consider four lags of each variables. Results shown over Table 4, witness that the spurious regression result disappears: i.e contemporaneous and lagged US disposable income is significant in explaining US consumption, while contemporaneous and lagged UK disposable income is not.

TABLE 4: A dynamic model for US consumption

	Dependent variable $LCUS_t$, regression by OLS, 1960:1-1998:1			
	Model with US income		*Model with UK income*	
	Coefficient	S.E.	Coefficient	S.E
c	0.367	0.106	0.333	0.150
$LCUS_{t-1}$	0.987	0.087	1.197	0.083
$LCUS_{t-2}$	-0.006	0.120	-0.156	0.131
$LCUS_{t-3}$	0.012	0.121	0.142	0.130
$LCUS_{t-4}$	-0.172	0.085	-0.196	0.082
LYUS	0.258	0.037		
$LYUS_{t-1}$	-0.126	0.049		
$LYUS_{t-2}$	-0.068	0.050		
$LYUS_{t-3}$	0.021	0.049		
$LYUS_{t-4}$	0.034	0.042		
LYUK			0.009	0.020
$LYUK_{t-1}$			0.018	0.028
$LYUK_{t-2}$			-0.034	0.028
$LYUK_{t-3}$			-0.0163	0.028
$LYUK_{t-4}$			0.0015	0.0229
Trend	0.00039	0.0001	0.00023	0.0001
R^2	0.99		0.99	
S.E.	0.0037		0.0042	
F-test on income	F(5,155)=10.324		F(5,155)=1.239	

.

This is an interesting result which leads to think that, in the case also the problems related to non-stationarity could be solved, dynamic multivariate time-series models are the right foundation for macroeconometrics.

2.7.1 *Cointegration and Error Correction Models*

To explain why the spurious results disappear when dynamic models are estimated let us consider a simplified version of the dynamic specification estimated for consumption:

$$c_t = a_0 + a_1 c_{t-1} + a_2 y_t + a_3 y_{t-1} + u_t \tag{2.14}$$

This specification has some interesting dynamic properties which are worth discussing. First note that the short-run elasticity of consumption with respect to income is different from the long-run elasticity. In fact the short-run elasticity is a_2 while the long-run elasticity is $\frac{a_2+a_3}{1-a_1}$. The long-run elasticity is found by setting all variables in the dynamic model (2.14)to their steady state value $c_{t+i} = \overline{c}$, $y_{t+i} = \overline{c}$. To see immediately this point consider the following re-parameterisation of (2.14):

$$\Delta c_t = a_0 + a_2 \Delta y_t - \alpha \left(c_{t-1} - \beta_1 y_{t-1} \right) + u_t \tag{2.15}$$

$$\alpha = (1 - a_1), \ \beta_1 = \frac{a_2 + a_3}{1 - a_1} \tag{2.16}$$

The estimated dynamic model includes both first differences and levels. The presence of the level variables generates a long-run solution, derived by setting all first differences either to zero (steady-state with no deterministic trend) or to a constant(steady-state). Note now the role of the terms in level: we can interpret $\beta_1 y_{t-1}$ as the long-run equilibrium level c^* for the log of real consumption c. When $\alpha < 0$ consumption increases at time t whenever $c_{t-1} < c^*_{t-1}$, and decreases whenever $c_{t-1} > c^*_{t-1}$.The system equilibrates in presence of dis-equilibrium (i.e. a discrepancy between c and c^*) such error correction features guarantees that in the long-run the consumption will converge to its equilibrium value. For this reason the specification (7.15) ,with $\alpha < 0$, is termed Error Correction Model. Note that, in the case of an ECM representation, the difference between c and c^* is a stationary series. This defines co-integration. We say that two non stationary variables integrated of order q are cointegrated of order p if there exist a linear combination of them which is integrated of order $p - q$. The case $p = 1$, $q = 1$, is interesting in that co-integration implies an ECM representation, which allows to re-write a model in levels, which involves non-stationary time-series, as a model which involves only stationary variables. Such variables are stationary either because they are first diferences of non-stationary variables or because they are stationary linear combination of non-stationary variables (cointegrating vectors).

The inclusion of both differences and levels in the estimated relationship is the key factor to the solution of the problems related to non-stationarity of the level of variables included in the specification. This solution to the non-stationarity problem has also the feature of revealing immediately to the economist the long-run properties of the estimated model. To see this point practically we can use E-Views to simulate the following bivariate model:

$$\Delta c_t = 0.25 \Delta y_t - 0.2 \left(c_{t-1} - y_{t-1} \right) + 0.003 u_{1t} \tag{2.17}$$

$$\Delta y_t = 0.02 + 0.009 u_{2t}$$

where u_{1t} and u_{2t} are independently distributed standard normal, the parameters are calibrated to reflect the long-run properties of the consumption function reported in Table 4. The volatility of the innovations are again calibrated to estimated processes on real data for the US economy; income is more volatile than consumption.

To show the properties of the model, we first generate samples for the two innovation process, then we generate artificial data for consumption and income by constructing the above model and solving it dynamically. We do so for a sample of 100 observations, the simulated series are plotted in Figures 2.9 and 2.10.

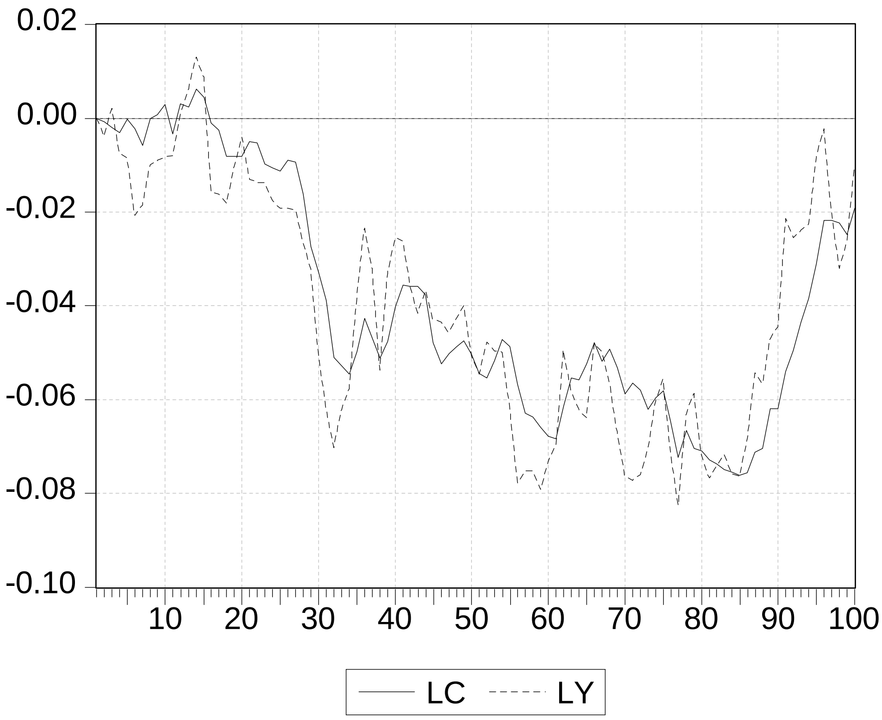


FIG. 2.9. Two cointegrated series

Note that the the levels of LC and LY, share a stochastic trend, which disappears from the series (LC-LY). The parameter α in the ECM spefication determines the speed of adjustment in presence of disequilibrium. To illustrate the role of this parameter we report the two series (LC-LY) generated by taking the same innovations for the sample 1 200. The innovations are drawn for normal independent for all observations, except for observation 101 where the residuals in the income process are augmented by 0.036. We have then a shock four standard deviation away from the mean of the distribution, we can then visually

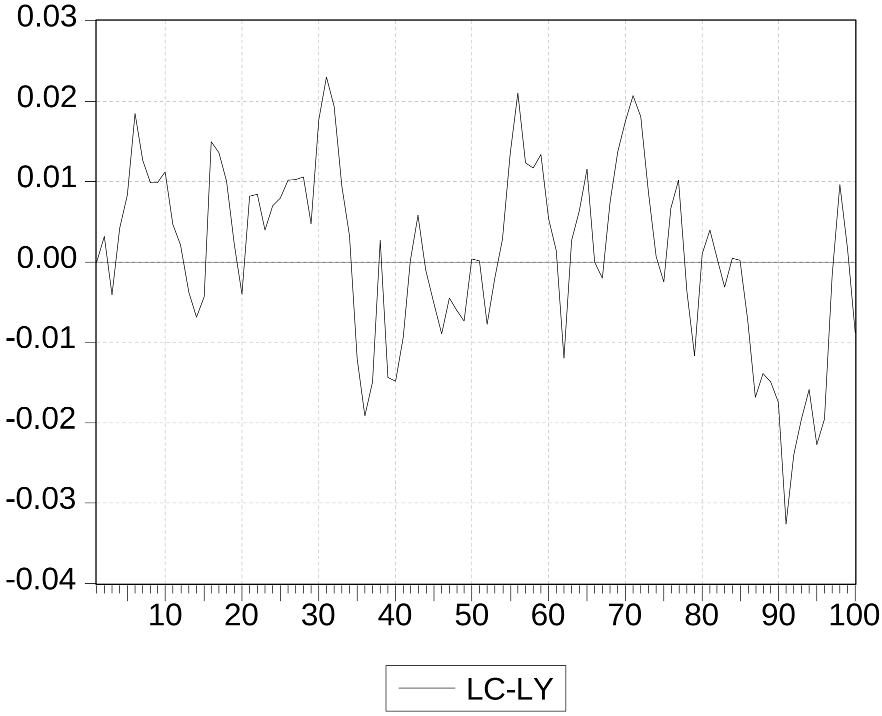


FIG. 2.10. Disequilibrium

inspect the behaviour of the simulated series in the presence of an outlier. The process (2.17) is used to generate the first time-series of disequilibria, while the second time-series is generated using keeping all parameters unchanged with the exception of α, which is trebled to 0.6 from 0.2. The resulting observations for disequilibria are reported in the Figure 2.11.

The disequilibria from the second simulation run are less persitent to witness that the second system feature a fastest speed of adjustment in presence of disequilibrium. All the simulated series are contained in the E-Views workfile ECM.WF1, with which the reader can experiment to convince herself of the properties of Error Correction Models.

As an application of further interest let us reconsider the static regression in the light of our discussion of dynamic models.

Given the following DGP:

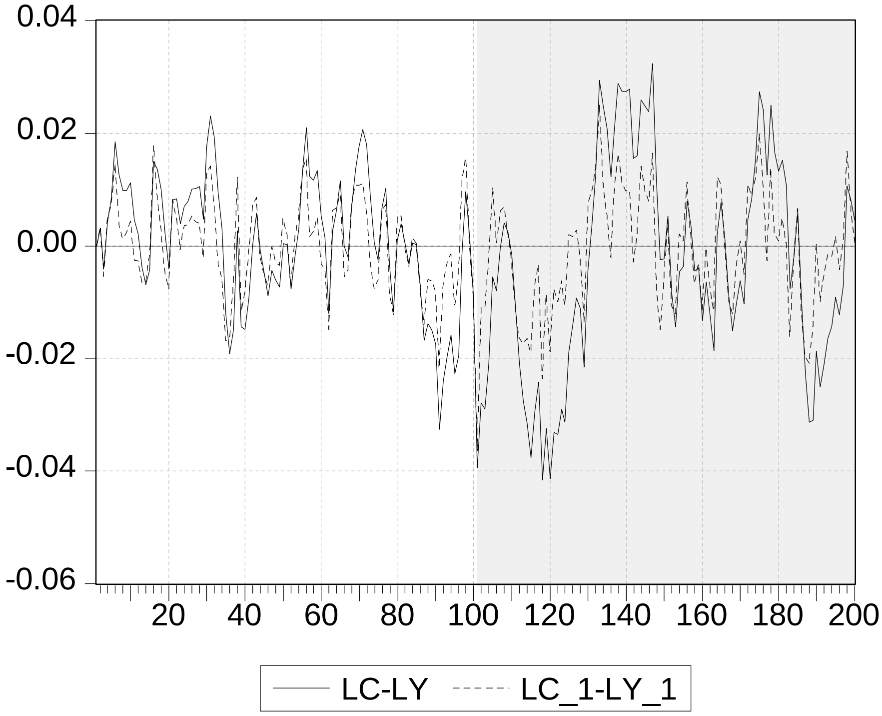


FIG. 2.11. Speed of adjustments and disequilibrium

$$
\begin{aligned}
y_t &= a_1 y_{t-1} + a_2 x_t + a_3 x_{t-1} + u_{1t} \\
x_t &= b_1 x_{t-1} + u_{2t}
\end{aligned}
\tag{2.18}
$$

$$\begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \sim N.I.D.\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix} \right)$$

a static model is estimated by OLS:

$$y_t = \gamma x_t + \varepsilon_t$$

$$\widehat{\gamma} = \frac{\sum x_t y_t}{\sum x_t^2}$$

Assess the results of running the static model by taking $p \lim \hat{\gamma}$:

$$p\lim \widehat{\gamma} = p\lim \left[a_1 \frac{\sum x_t y_{t-1}/T}{\sum x_t^2/T} + a_2 + a_3 \frac{\sum x_t x_{t-1}/T}{\sum x_t^2/T} + \frac{\sum x_t u_t/T}{\sum x_t^2/T} \right]$$

under the hypothesis that (2.18) is stationary ($b_1 < 1$) we can substitute for x_t in terms of x_{t-1} and u_{2t} and apply Slutsky's and Cramer's theorem to derive the following result:

$$p\lim \widehat{\gamma} = \frac{a_2 + a_3 b_1}{1 - a_1 b_1}$$
$$a_2 \leq p\lim \widehat{\gamma} \leq \frac{a_2 + a_3}{1 - a_1}$$

Note that as b_1 approaches 0 the elasticy of y with respect to x delivered by the static regression goes asymptotically to the true short-run elasticity, while as b_1 approaches to 1 such elasticity converges to the long-run elasticity. Technically speaking we cannot show what happens when b_1 is one because this violates the stationarity conditions which we have used to derive the asymptotic behaviour of the OLS estimator. However, confirming the above intuition, Stock([55]) has shown that the OLS estimator of the parameters determining the long-run relationship non-stationary cointegrated series is super-consistent. In fact it converges towards the true value at speed $\left(\frac{1}{T}\right)$,higher than the speed of much $\left(\frac{1}{\sqrt{T}}\right)$,with which the OLS estimator converges to its true value in regression between stationary time series. This result has given some background to a two-step research strategy, according to which the cointegrating relations is estimated first in static model and the used to estimate a dynamic ECM model, involving only stationary variables. This strategy is less efficient than the simultaneous estimation of short-run and long-run dynamics. In fact the static regression delivers super-consitent estimates of the cointegrating parameter despite being mis-specified, because the omitted variables are the stationary variables determining the short-run dynamics, which, in large-samples, should not affect the estimation of cointegrating parameters. It has been shown through Monte-Carlo simulation that the dimension of the samples required to appeal to the super-consitency theorem are much higher than the dimension of the samples usually available for time-series modelling (see, for example, [2],[3], [4]). Moreover the empirical counterpart of macroeconomic models are usually dynamic multivariate time-series models. Therefore, there must be a price to be paid in considering static uni-variate models as a basis for empirical work. We shall devote more attention to this issue in the next section.

2.7.2 *Cointegration in a multivariate framework*

So far we have stressed the importance of the magnitude of the adjustment parameter α as the relevant discriminant to decide on cointegration, but we have not yet provided a statistical framework to test for such an hypothesis.We also mentioned the importance of dimensionality of the system to be considered in empirical work. In this section we shall elaborate on these points and illustrate the Johansen's([30], [34]) approach to cointegration in a multivariate framework.

So far we have considered cointegration in a bi-variate context. Things differ when we consider a multivariate context. In fact , in general between n non-stationary series we can have up to $n-1$ cointegrating vector and the single equation dynamic modelling can cause serious troubles when there are multiple cointegrating vectors. To illustrate the problem let us consider the case of an investigator who uses cointegration techniques to investigate money demand. Following the standard economic background to empirical investigations of money demand (see, for example, Hendry and Ericsson,[26]) the chosen data set includes money, m, a price index, p, real income, y, the own interest rate on money, R^m, and the opportunity cost of holding money, R^b. All variables are in logarithms, with the exception of interest rates. The investigator specifies a dynamic single-equation model for real money, towards the identification of a money demand equation, which takes the following shape:

$$\begin{aligned}(m-p)_t &= a_0 + a_1 (m-p)_{t-1} + a_2 y_{t-1} + a_3 y_{t-2} + \\ &+ a_4 R^m_{t-1} + a_5 R^m_{t-2} + a_6 R^b_{t-1} + a_7 R^b_{t-2} + u_t\end{aligned} \tag{2.19}$$

This statistical model fits the data well. As it is found that $a_1 < 1$, the investigation leads to the identification of a long-run equilibrium money demand, which results clearly form the ECM re-parametrization of the dynamic model (2.19) :

$$\begin{aligned}\Delta (m-p)_t &= a_0 - a_3 \Delta y_{t-1} - a_5 \Delta R^m_{t-1} - a_6 \Delta R^b_{t-1} + \\ & (a_1 - 1) \left[(m-p)_{t-1} - (m-p)^*_{t-1} \right] + u_t \\ (m-p)^*_{t-1} &= \frac{a_2 + a_3}{1-a_1} y_{t-1} + \frac{a_4 + a_5}{1-a_1} R^m_{t-1} + \frac{a_6 + a_7}{1-a_1} R^b_{t-1}\end{aligned} \tag{2.20}$$

However, the good fit of the statistical model might be combined with an incorrect identification of the long-run solution. Think, for example, of the case in which the non-stationary vector containing the five variables of interests admits two cointegration relationships:$(m-p-y)$ and $\left(R^m - \beta_{22} R^b\right)$. Where the first one is generated by the stationarity of the velocity of circulation of money and the second one by the behaviour of the banking sector, which sets the interest rate on money as a mark-down on the opportunity cost of holding money. In the short-run money reacts to disequilibria with respect to both long-run solutions, hence money demand is correctly parametrised as follows:

$$\begin{aligned}\Delta (m-p)_t &= \pi_0 + \pi_1 \Delta y_{t-1} + \pi_2 \Delta R^m_{t-1} + \pi_3 \Delta R^b_{t-1} \\ & - \alpha_1 (m_{t-1} - p_{t-1} - y_{t-1}) + \alpha_2 \left(R^m_{t-1} - \beta_{22} R^b_{t-1} \right) + u_t\end{aligned} \tag{2.21}$$

Note that the statistical specification of (7.15) and (2.21) is identical, in fact the residuals u_t are the same, however identification is very different. In fact when (2.21) represents the correct specification, (7.15) identifies as long-run elasticities what in fact are mixtures of cointegrating parameters and parameters determining the speed of adjustment with respect to disequilibria in the true model. Single-equation approach leads to believe that the long-run elasticity of money demand with respect to the opportunity cost of holding money is $\frac{a_6+a_7}{1-a_1}$,while in fact such estimated coefficient is a convolution of the parameter α_2, determining the speed with which money demand reacts to a disalignment of interest rates with respect to their equilibrium value, and the parameter c, determining the mark-down of the own interest rate on money with respect to the interest rate on the opportunity cost of holding money. This identification has serious consequences in the intepretation of estimated parameters. In fact when the above problem occurs a structural instability in the short-term adjustment parameter β_{22} would mis-lead the researcher into the belief that long-run money demand is unstable.

The solution of this identification problem requires a framework to allow the researcher to find the number of cointegrating vectors among a set of variables and to identify them. The procedure proposed by Johansen([30], [32]) within the framework of the Vector Autoregressive Model achieves both results.

2.7.3 *The Johansen Procedure*

To illustrate the procedure proposed by Johansen consider the multivariate generalisation of the single-equation dynamic model discussed so far, i.e. a Vector Autoregressive Model (VAR) for the vector of, possibly non-stationary, m-variables $\mathbf{y}$:

$$\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} + ... + \mathbf{A}_n\mathbf{y}_{t-n} + \mathbf{u}_t \tag{2.22}$$

by proceeding in the same way we did for the simple single-equation dynamic model, we can reparameterise the VAR in levels as a model involving levels and first-differences of variables.

Start by subtracting $\mathbf{y}_{t-1}$ from both sides of the VAR to obtain:

$$\Delta\mathbf{y}_t = (\mathbf{A}_1 - \mathbf{I})\,\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} + ... + \mathbf{A}_n\mathbf{y}_{t-n} + \mathbf{u}_t \tag{2.23}$$

now subtract $(\mathbf{A}_1 - \mathbf{I})\,\mathbf{y}_{t-2}$ from both sides to obtain:

$$\Delta\mathbf{y}_t = (\mathbf{A}_1 - \mathbf{I})\,\Delta\mathbf{y}_{t-1} + (\mathbf{A}_1 + \mathbf{A}_2 - \mathbf{I})\,\mathbf{y}_{t-2} + ... + \mathbf{A}_n\mathbf{y}_{t-n} + \mathbf{u}_t \tag{2.24}$$

By iterating this procedure until n-1, we end up with the following specification:

$$\Delta \mathbf{y}_t = \Pi_1 \Delta \mathbf{y}_{t-1} + \Pi_1 \Delta \mathbf{y}_{t-2} + ... + \Pi \mathbf{y}_{t-n} + \mathbf{u}_t \quad (2.25)$$

$$= i = 1 \sum^{n-1} \Pi_i \Delta \mathbf{y}_{t-i} + \Pi \mathbf{y}_{t-n} + \mathbf{u}_t \quad (2.26)$$

where :

$$\Pi_i = -\left(I - j = 1 \sum^{i} \mathbf{A}_j \right)$$

$$\Pi = -\left(I - i = 1 \sum^{n} \mathbf{A}_i \right)$$

Clearly the long-run properties of the system are described by the properties of the matrix Π. There are three cases of interest:

- rank $(\Pi) = 0$. The system is non-stationary, with no cointegration between the variables considered. This is the only case in which non-stationarity is correctly removed just by taking first difference of the variables considered
- rank $(\Pi) = m$, $full$. The system is stationary.
- rank $(\Pi) = k < m$. The system is non stationary but there are k cointegrating relationships among the considered variables. In this case we have $\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$, where α is an $(m \times k)$ matrix of weights and $\boldsymbol{\beta}$ is an $(m \times k)$ matrix of parameters determining the cointegrating relationships.

Therefore, the rank of Π is crucial in determining the number of cointegrating vectors. The Johansen procedure is based on the fact that the rank of a matrix is equal to the number of its characteristic roots that differ from zero. Here is the intuition on how the tests can be constructed. Having obtained estimates for the parameters in the Π matrix, we associate to them estimates for the m characteristic roots and we order them as follows $\lambda_1 > \lambda_2 > ... > \lambda_m$. If the variables are not cointegrated, then the rank of Π is zero and all the characteristic roots will be zero. In this case each of the expression $\ln(1-\lambda_i)$ will be zero. If instead the rank of Π is one, and $0 < \lambda_1 < 1$, then $\ln(1-\lambda_1)$ will be negative and $\ln(1-\lambda_2)$ $=\ln(1-\lambda_3) = ... = \ln(1-\lambda_m) = 0$. Johansen derives a test on the number of characteristic roots that are different from zero by considering the two following statistics:

$$\lambda_{trace}(k) = -Ti = k + 1 \sum^{m} \ln\left(1 - \widehat{\lambda}_i\right)$$

$$\lambda_{\max}(k, k+1) = -T \ln\left(1 - \widehat{\lambda}_{k+1}\right)$$

where T is the number of observations used to estimate the VAR. The first statistic test the null of at most k cointegrating vectors against a generic alternative. The test should be run in sequence starting from the null of at most zero cointegrating vectors up to the case of at most m cointegrating vectors. The second statistic tests the null of at most k cointegrating vectors against the alternative of at most $k+1$ coitegrating vectors. Both statistics will be small under the null hypothesis. Critical values are tabulated by Johansen and they depend on the number of non-stationary component under the null and on the specification of the deterministic component of the VAR. Johansen has shown in the past ([33]) some preference for the trace test on the argument that the maximum eigenvalue test does not give rise to a coherent testing strategy.

To illustrate briefly the intuition behind the procedure, consider the VAR representation of our simple dynamic model (2.18), introduced in one of the previous sections, for the two variables x and y:

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \tag{2.27}$$

(2.27) can be reparameterised as follows in terms of the VECM representation:

$$\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} a_{11}-1 & a_{12} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \tag{2.28}$$

from which it is clear that

$$\Pi = \begin{pmatrix} a_{11}-1 & a_{12} \\ 0 & 0 \end{pmatrix}, \alpha = \begin{pmatrix} a_{11}-1 \\ 0 \end{pmatrix}, \beta' = \begin{pmatrix} 1 & -\frac{a_{12}}{1-a_{11}} \end{pmatrix}$$

To expand on this intuition let us reconsider our example on money demand from the previous section.

The baseline VAR could be specified as follows:

$$\begin{bmatrix} (m-p)_t \\ y_t \\ R_t^m \\ R_t^b \end{bmatrix} = A_0 + A_1 \begin{bmatrix} (m-p)_{t-1} \\ y_{t-1} \\ R_{t-1}^m \\ R_{t-1}^b \end{bmatrix} + A_2 \begin{bmatrix} (m-p)_{t-2} \\ y_{t-2} \\ R_{t-2}^m \\ R_{t-2}^b \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ u_{4t} \end{bmatrix}$$

Which could then be reparameterised in VECM form:

$$\begin{bmatrix} \Delta (m-p)_t \\ \Delta y_t \\ \Delta R_t^m \\ \Delta R_t^b \end{bmatrix} = \Pi_0 + \Pi \begin{bmatrix} (m-p)_{t-1} \\ y_{t-1} \\ R_{t-1}^m \\ R_{t-1}^b \end{bmatrix} + \Pi_1 \begin{bmatrix} \Delta (m-p)_{t-1} \\ \Delta y_{t-1} \\ \Delta R_{t-1}^m \\ \Delta R_{t-1}^b \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ u_{4t} \end{bmatrix}$$

Given that we know that there are two cointegrating vectors, we have:

$$\Pi = \alpha\beta'$$
$$rank\,\Pi = 2$$
$$\beta' = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -\beta_{22} \end{bmatrix}$$

As we have analysed only one equation in our previous discussion of the system, the only constraints we have on the specification for α are $\alpha_{11} < 0, \alpha_{12} > 0$. A possible specification for α would then be:

$$\alpha = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ 0 & 0 \\ 0 & \alpha_{32} \\ 0 & 0 \end{bmatrix}$$

with the above specification for the loadings, money demand adjusts both in presence of misalignements of velocity with respect to the equilibrium velocity and of misalignments of interest rates with respect to their equilibrium spread. In particular money demand increases when velocity is "too high" and the opportunity cost of holding money is "too low". In case of disequilibrium in interest rates it is the interest on money which adjusts, while the dynamics of interest rates on the alternative of money in agents'portfolio does not react to disequilibria.

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} \\ 0 & 0 \\ 0 & \alpha_{32} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -\beta_{22} \end{bmatrix} = \begin{bmatrix} \alpha_{11} & -\alpha_{11} & \alpha_{12} & -\alpha_{12}\beta_{22} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha_{32} & -\alpha_{32}\beta_{22} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

2.7.4 *Identification of multiple cointegrating vectors*

The Johansen procedure allows to identify the number of cointegrating vectors. However, in the case of existence of multiple cointegrating vectors, an interesting identification problem arises. In fact, α and β, are only determined up to the space spanned by them and, for any non-singular matrix ξ conformable by product, we have:

$$\Pi = \alpha\beta' = \alpha\xi^{-1}\xi\beta'$$

In other words β and $\beta'\xi$ are two observationally equivalent basis of the cointegrating space. The obvious implication is that, before solving such identification problem, no meaningful economic interpretation of coefficients in cointegrating vectors can be proposed. The solution to such problem is achieved by imposing a sufficient number of restrictions on parameters such that the matrix satisfying such constraints in the cointegrating space is unique. Such criterion is derived in Johansen(1992) and discussed in the work of Johansen-Juselius, Giannini([15])

and Hamilton([20]). Given the matrix of cointegrating vectors $\boldsymbol{\beta}$ we can formulate linear constraints on the different cointegrating vectors using the R_i matrices of dimensions $r_i \times n$. Let us consider the columns of $\boldsymbol{\beta}$, i.e. the parameters in each cointegrating vectors, ignoring the normalisation constraint to1 of one variable in each cointegrating vector. Any structure of linear contraints can be represented as follows:

$$\mathbf{R}_i \boldsymbol{\beta}_i = 0$$

$R_i\left(r_i \times n\right)$, $\boldsymbol{\beta}_i(n \times 1)$,rank $R_i = r_i$.

The same constraints can be expressed in explicit forms as follows:

$$\boldsymbol{\beta}_i = \mathbf{S}_i \theta_i$$

$S_i\left(n \times \left(n - r_i\right)\right)$, $\boldsymbol{\beta}_i(n \times 1), \theta_i\left(\left(n - r_i\right) \times 1\right)$, rank $S_i = n - r_i, \mathbf{R}_i \mathbf{S}_i = 0$.

A necessary and sufficient condition for identification of parameters in the $i - th$ cointegrating vectors is the following:

$$rank\left(\mathbf{R}_i \boldsymbol{\beta}\right) = r - 1 \tag{2.29}$$

when (6.5) is satisfied it is not possible to replicate the cointegrating vector $i - th$ by taking linear combinations of the parameters in the other cointegrating vectors. In this case the matrix obtained by applying to the cointegrating space the restrictions of the $i - th$ cointegrating vectors has rank $r - 1$.

A necessary condition for identification is immediately derived in that $\mathbf{R}_i \boldsymbol{\beta}$ must have enough rows to satisfy condition (6.5) , therefore a necessary condition for identification is that each cointegrating vectors has at least $r - 1$ restrictions.

A sufficient condition for identification is provided by Johansen by considering the implicit and explicit form of expressing constraints:

Theorem 2.1 *The i-th cointegrating vector is identified by the constraints* $\mathbf{S}_1, \mathbf{S}_2, ...\mathbf{S}_r$ *if for each k=1,...,r-1 and for each set of indices* $1 < j_1 < ... < j_k < r$, *not containing i, we have that :* $rank\left[R_i S_{j_1}, ...R_i S_{j_k}\right] > k$

Given identification of the system we can distinguish the case of just-identification and over-identification. In case of over-identification, the over-identifying restrictions are testable.

2.7.4.1 *An illustrative example* Let us reconsider our example on money demand.Considering the following vectorial representation of the series $\left(m - p \; y \; R^m \; R^b\right)'$, and leaving aside normalizations, the matrix $\boldsymbol{\beta}$ can be represented as follows:

$$\begin{pmatrix} \beta_{11} & 0 \\ -\beta_{11} & 0 \\ 0 & \beta_{32} \\ 0 & -\beta_{42} \end{pmatrix}$$

given the following general representation of the matrix $\boldsymbol{\beta}$:

$$\begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \\ \beta_{41} & \beta_{42} \end{pmatrix}$$

our constraints imply the following specification for the matrices R_i ans S_i :

$$R_1 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, S_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}$$

$$R_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, S_2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

The necessary conditions for identification are obviously satisfied, while the sufficient conditions for identification requires: $\text{rank}(R_1 S_2) \geq 1$, $\text{rank}(R_2 S_1) \geq 1$. They are also satisfied, in fact:

$$R_1 S_2 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, R_2 S_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

2.7.5 *Hypothesis testing with multiple cointegrating vectors*

The Johansen procedure allows for testing the validity of restricted forms of cointegrating vectors. More precisely, the validity of restrictions in additions(over-identifying restrictions) to those necessary to identify the long-run equilibria could be tested. The intuition behind the construction of all tests is that when there are r cointegrating vectors only these r linear combination of variables are stationary, therefore the test statistics involve comparing the number of cointegrating vectors under the null and the alternative hypothesis. Following this intuition, we understand immediately why only over-identifying restrictions can be tested, in fact just-identified model feature the same long-run matrix Π, and therefore the same eigenvalues of Π. Consider the case of testing restrictions on a set r of identified cointegrating vectors stacked in the matrix $\boldsymbol{\beta}$. The test statistic involves comparing the number of cointegrating vectors under the null and the alternative hypothesis. Let $\widehat{\lambda}_1, \widehat{\lambda}_2, ..., \widehat{\lambda}_r,$ the ordered eigenvalues of the Π matrix in the unrestricted model, and $\widehat{\lambda}_1^*, \widehat{\lambda}_2^*, ..., \widehat{\lambda}_r^*$the ordered eigenvalues of Π matrix in the restricted model, restricitions on $\boldsymbol{\beta}$ are testable by forming the following test statistic:

$$Ti = 1\sum^{r}\left[\ln\left(1-\widehat{\lambda}_i^*\right) - \ln\left(1-\widehat{\lambda}_i\right)\right] \tag{2.30}$$

Johansen ([32]) shows that the statistic (2.30) takes a χ^2 distribution with degrees of freedom equal to the number of over-identifying restrictions. Note that small values of $\widehat{\lambda}_i^*$ with respect to $\widehat{\lambda}_i$ imply a reduction of rank of Π when the restrictions are imposed and hence the rejection of the null hypothesis. This testing procedure can be extended to tests on restrictions on the matrix of weigths $\boldsymbol{\alpha}$ or on the deterministic components (constant and trends) of the cointegrating vectors.

2.7.6 *Cointegration and Common Stochastic Trends*

Having discussed the VECM representation for a vector of m non-stationary variables admitting k cointegrating relationships, it is interesting to compare it with the multivariate extension of the Beveridge-Nelson decomposition. Consider the simple case of an I(1) vector $\mathbf{y}_t$ featuring first order dynamics and no deterministic components:

$$\Delta\mathbf{y}_t = \boldsymbol{\alpha\beta}'\mathbf{y}_{t-1} + \mathbf{u}_t \tag{2.31}$$

where α is the $(m \times k)$ matrix of loadings and $\boldsymbol{\beta}$ is the $(m \times k)$ matrix of parameters in the cointegrating relationships. When As $\mathbf{y}_t$ is I(1), we can apply the Wold decomposition theorem to $\Delta\mathbf{y}_t$ to obtain the following representation:

$$\Delta\mathbf{y}_t = \mathbf{C}(L)\,\mathbf{u}_t$$

From which, by applying the algebra illustrated in our discussion of the univariate Beveridge-Nelson decomposition, we can derive the following stochastic trends representation:

$$\mathbf{y}_t = \mathbf{C}^*(L)\,\mathbf{u}_t + \mathbf{C}(1)\,\mathbf{z}_t$$

where $\mathbf{z}_t$ is a process for which we have $\Delta\mathbf{z}_t = \mathbf{u}_t$. The existence of cointegration imposes some restrictions on the C matrices, in fact the stochastic trends must cancel out when the k stationary linear combinations of the variables in $\mathbf{y}_t$ are considered in other words we must have:

$$\boldsymbol{\beta}'\mathbf{C}(1) = 0$$

By investigating further the relation between the VECM and the stochastic trends representations we can give a more precise parameterisation of the matrix $\mathbf{C}(1)$.

Note first that equation (2.21) can be re-written as :

$$\mathbf{y}_t = \left(I_m + \boldsymbol{\alpha\beta}'\right)\mathbf{y}_{t-1} + \mathbf{u}_t \tag{2.32}$$

Premultiplying this system by $\boldsymbol{\beta}'$ yields:

$$\begin{aligned}\boldsymbol{\beta}'\mathbf{y}_t &= \boldsymbol{\beta}'\left(I_m + \boldsymbol{\alpha\beta}'\right)\mathbf{y}_{t-1} + \boldsymbol{\beta}'\mathbf{u}_t \\ &= \left(I_k + \boldsymbol{\alpha\beta}'\right)\boldsymbol{\beta}'\mathbf{y}_{t-1} + \boldsymbol{\beta}'\mathbf{u}_t\end{aligned}$$

Solving this model recursively, we obtain the MA representation for the k cointegrating relationships:

$$\boldsymbol{\beta}'\mathbf{y}_t = i = 0\overset{\infty}{\sum}\left(I_k + \boldsymbol{\alpha\beta}'\right)^i\boldsymbol{\beta}'\mathbf{u}_{t-i} \tag{2.33}$$

By substituting (2.33) in (2.21) we have the MA representation for $\Delta\mathbf{y}_t$:

$$\Delta\mathbf{y}_t = i = 1\overset{\infty}{\sum}\boldsymbol{\alpha}\left(I_k + \boldsymbol{\alpha\beta}'\right)^{i-1}\boldsymbol{\beta}'\mathbf{u}_{t-i} + \mathbf{u}_t$$

from which we have:

$$\mathbf{C}(1) = I_n - \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\boldsymbol{\alpha}\right)^{-1}\boldsymbol{\beta}' \tag{2.34}$$

Now note the beatiful[6] relation

$$I_n = \beta_\perp\left(\alpha'_\perp\beta_\perp\right)^{-1}\alpha'_\perp + \alpha\left(\beta'\alpha\right)^{-1}\beta' \tag{2.35}$$

where $\beta_\perp, \alpha_\perp$ are $((m \times (m-k))$ matrices of rank $m-k$ such that $\alpha'_\perp\alpha = 0, \beta'_\perp\beta = 0$.

By using (2.35) in (2.34) , we have

$$\mathbf{C}(1) = \beta_\perp\left(\alpha'_\perp\beta_\perp\right)^{-1}\alpha'_\perp$$

and

$$\mathbf{y}_t = \mathbf{C}^*(L)\mathbf{u}_t + \beta_\perp\left(\alpha'_\perp\beta_\perp\right)^{-1}\left(\alpha'_\perp\mathbf{z}_t\right)$$

which shows that a system of m variables with k cointegrating relationships features $(m-k)$ linearly independent common trends (**TR**). The common trends

[6]See Johansen([34]) ,page 40

are given by $(\alpha'_{\perp}\mathbf{z}_t)$, while the coefficients on these trends are $\beta_{\perp}(\alpha'_{\perp}\beta_{\perp})^{-1}$. Note also that stochastic trends depend on a set of initial conditions and on cumulated disturbances in fact

$$\mathbf{TR}_t = \mathbf{TR}_{t-1} + C(1)\,\mathbf{u}_t$$

Our brief discussion should have made clear that the VECM model and the MA model are complementary. As a consequence the identification problem relevant for the vector of parameters in the cointegrating vectors β is also relevant for the vector of parameters determining the stochastic trends $\boldsymbol{\alpha}_{\perp}$. However, there is one aspect in which the two concepts are different. In theory, identified cointegrating relationships on a given set of variables should be robust to augmentation of the information set by adding new variables, which should have a zero coefficient in the cointegrating vectors of the VECM representation of the larger information set. This is not true for the stochastic trends. Consider the case of augmenting an information set consiting of m variables admitting k cointegrating vectors to $m+n$ variables, the number of cointegrating vectors is constant while the number of stochastic trends increases by n, moreover an unanticipated shock in a small system need not to be unanticipated in a larger system. Note that we have added in theory to our statement, this is because, in practice, given the size of available samples application of the procedure to analyze cointegration in a larger set of variables might lead to identify different cointegrating relationships from those obtained on a smaller set of variables.

2.7.7 *VECM and common trends representations*

The joint behaviour of consumption and income under the Permanent Income Hypothesis (PIH) is a good empirical example to illustrate VECM and common trends representations. Let y_t, y_t^p and c_t denote respectively the logarithms of aggregate disposable income, permanent income and consumption. Under PIH the joint distribution of consumption and income can be characterised as follows:

$$\begin{aligned} y_t &= y_t^p + v_t \\ y_t^p &= \mu_y + y_{t-1}^p + u_t \\ c_t &= y_t^p \end{aligned}$$

permanent income is the stochastic trend of income, which is made of the permanent component and of a transitory component, v_t and u_t are the shocks to the transitory and the permanent component of income, it is natural to think of them as orthogonal shocks normally and independently distributed. Consumption and income are cointegrated, in fact they share the single unobservable common stochastic trend in this system.

By eliminating the unobservable stochastic trend from the system, we have a bi-variate structural representation:

$$\begin{aligned} y_t &= c_t + v_t \\ c_t &= \mu_y + c_{t-1} + u_t \end{aligned} \tag{2.36}$$

We obtain the VAR(1) representation by substituting for c_t in the first equation from the second equation of (2.36)

$$\begin{aligned} \begin{pmatrix} y_t \\ c_t \end{pmatrix} &= \begin{pmatrix} \mu_y \\ \mu_y \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ c_{t-1} \end{pmatrix} + \begin{pmatrix} w_t \\ u_t \end{pmatrix} \\ w_t &= u_t + v_t \end{aligned}$$

From which we immediately obtain the VECM representation

$$\begin{pmatrix} \Delta y_t \\ \Delta c_t \end{pmatrix} = \begin{pmatrix} \mu_y \\ \mu_y \end{pmatrix} + \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ c_{t-1} \end{pmatrix} + \begin{pmatrix} w_t \\ u_t \end{pmatrix}$$

where

$$\begin{aligned} \Pi &= \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} -1 & 1 \end{pmatrix} \\ &= \alpha\beta' \end{aligned}$$

The derivation of the common trends representation is derived by considering that, as $y_t - c_t = v_t$, the MA representation for consumption and income growth is then

$$\begin{pmatrix} \Delta y_t \\ \Delta c_t \end{pmatrix} = \begin{pmatrix} \mu_y \\ \mu_y \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} w_t \\ u_t \end{pmatrix} + \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} w_{t-1} \\ u_{t-1} \end{pmatrix}$$

from which we have

$$\begin{pmatrix} y_t \\ c_t \end{pmatrix} = \begin{pmatrix} \mu_y \\ \mu_y \end{pmatrix} t + \mathbf{C}^* (L) \begin{pmatrix} w_t \\ u_t \end{pmatrix} + C(1)\mathbf{z}_t$$

and

$$\begin{aligned} C(1) &= \beta_\perp \left(\alpha'_\perp \beta_\perp\right)^{-1} \alpha'_\perp \\ \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} \left[\begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right]^{-1} \begin{pmatrix} 0 & 1 \end{pmatrix} \end{aligned}$$

Given that in this application $\left(\alpha'_\perp \beta_\perp\right)^{-1} = 1$, it follows that consumption and income have a single common stochastic trend. Such trend can be represented as

$\alpha'_{\perp}\left(\begin{pmatrix}\mu_y\\ \mu_y\end{pmatrix}t+\begin{pmatrix}i=1\sum^{t}w_t\\ i=1\sum^{t}u_t\end{pmatrix}\right)$, and only shocks to the permanent component of income enter the trend.

2.8 Multivariate cointegration: an application to US data

To illustrate empirically how cointegration analysis is performed let us consider monthly data from the US economy for the variables considered in basic macroeconomic models: the log of the real M2 $(m-p)$,annual, seasonally adjusted, CPI inflation (π) , the log of monthly real GDP (y), the nominal own return on M2 (R^m), the nominal opportunity cost of holding money as measured by the interest rate on three-month Treasury Bills $\left(R^b\right)$. All series except R^m are those used in Leeper-Sims-Zha([38]) , R^m has been retrieved by the St.Louis FED Website at *http://www.stls.frb.org/fred/.* They are available in the file LSZUSA.XLS. We shall perform cointegration analysis using the package PC-FIML by Doornik and Hendry ([5]), alternative menu-driven packages are available in RATS (see [41], [21]) , E-Views does not allow to perform all the steps of the analysis in that specification and testing of the long-run restrictions are not (yet) available.

2.8.1 *Specification of the VAR*

The first step of the empirical analysis is the specification of the VAR. The specification of the VAR requires the consideration of two issues pertaining respectively to the set of variables included in the VAR and to the lag length of the VAR. These are important issues in that mis-specification of the VAR leads to misleading inference. In general the set of variables to be included in the VAR is determined by the economic problem at hand, however this criterion does not rule out the possibility of mis-specification. Consider the case of the set of variables chosen for our example, they include all the variables used in a simple IS-LM model of a closed economy, but nothing guarantees that the US economy is correctly described by such model. Suppose that the central bank targets expected inflation by using short term interest rates as an instrument. The model is mis-specified if it omits any leading indicator for inflation monitored by Central Bank. An obvious candidate is the commodity price index but there might be more, such as long-term interest rates or other asset prices. In absence of an obvious baseline model, the behaviour of residuals is taken as an indicator of mis-specification. In a correctly specified model residuals should be random normal variables with zero mean and constant variance-covariance matrix, departure of fitted residuals from those hypotheses could be taken as an indicator of mis-specification. However, even when all the relevant variables have been included, the model could be still mis-specified because of omitted relevant dynamics. The selection of the order of the VAR is an important step in the specification. Sims([48]) suggests a statistics to test the validity of restrictions imposed on a general model:

$$(T-k)\left[\log|\Sigma_r| - \log|\Sigma_{unr}|\right]$$

where T is the sample size, k is the number of parameters estimated in each equation of the VAR, $|\Sigma_r|$ is the determinant of the variance-covariance matrix of the residuals of the fitted restricted model, $|\Sigma_{unr}|$ is the determinant of the variance-covariance matrix of the residuals of the fitted general unrestricted model. The statistic has a χ^2 distribution with degrees of freedom equal to the number of restrictions in the system. The term $(T-k)$ includes a small sample correction, in fact as T becomes larger the correction for small sample $(T-k)/T$ converges towards unity. Obviously, the selection of variables and the selection of the lag length are not independent processes, in fact a longer lag length might be the consequence of omission of one relevant variable from the VAR. In practice we shall start from a baseline VAR including the set of variables suggested by the theory and a sufficiently long lag, check the behaviour of residuals. When well-behaved residuals are obtained, we proceed to reduction of the lag length by testing the validity of the implied restrictions.

Our general baseline model is a VAR estimated over the sample 1960:1-1979:6, including fifteen lags of each of the five variables, a constant and a trend, so we have:

$$\begin{pmatrix} y_t \\ \pi_t \\ (m-p)_t \\ R_t^m \\ R_t^b \end{pmatrix} = \mathbf{a}_o + \mathbf{a}_1 t + i = 1\sum^{14} A_i L^i \begin{pmatrix} y_t \\ \pi_t \\ (m-p)_t \\ R_t^m \\ R_t^b \end{pmatrix} + \mathbf{u}_t. \tag{2.37}$$

We have chosen to end our estimation in 1979 because, from the second part of 1979 to 1982, the Fed has changed is operating procedure moving from an interest rate targeting regime to a reserves targeting regime. As a consequences paramaters in the Fed's reaction function must have changed. It is very important to estimate cointegrating models using data from a single regime. In fact, structural instability might be very dangerous when cointegrated models are used. The intuition is very simple, in the presence of parameters instability a cointegrated model is very likely to push the system towards the "wrong" long-term equilibria with very serious consequences for forecasting and policy simulation. Checking residual behavious is very important also with this respect, as pathologic behaviour of residuals is a clear symptom of parameters' instability.

The estimation of our base-line model delivers the set of residuals reported in Figure 2.12.

The residual are normalized, hence residual with absolute value higher than 1.96 occur with a probability of one percent under the null of normality. We note many outliers, in fact when a formal test of normality of residuals is performed the null is strongly rejected[7]. This is worrying in that non-normality

[7] We shall discuss tests for normality at a later stage of the book

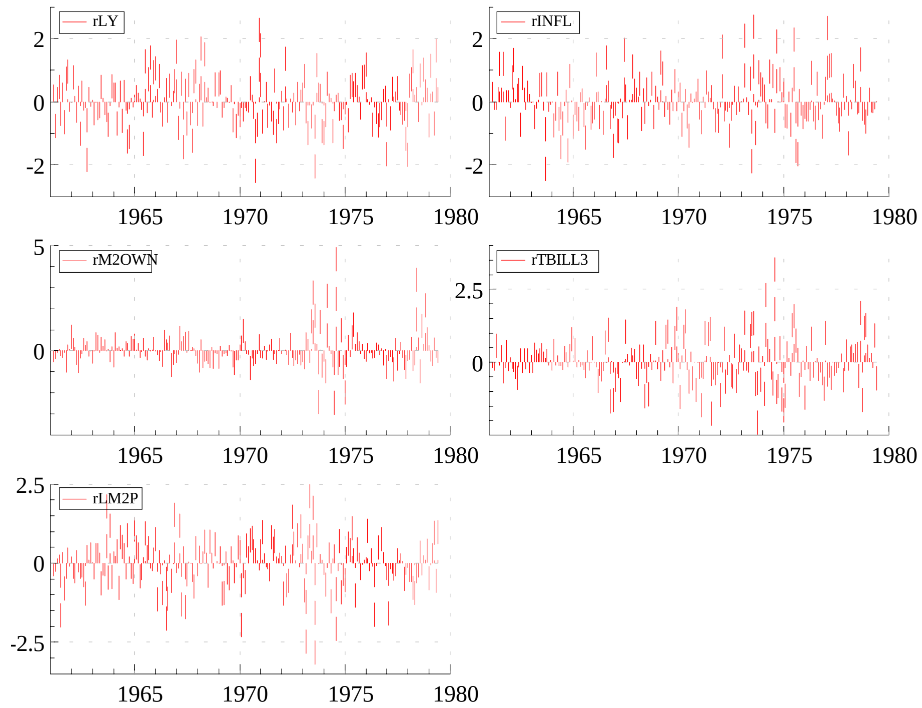


FIG. 2.12. VAR residuals with outliers

might signal mis-specification but also in that departure from normality might induce misleading inference in the application of the Johansen procedure. Interestingly, most outliers occur on occasion of the oil price crises. So, probably a commodity price index is the relevant omitted variables causing non-normality in the residuals. However, the inclusion of a commodity price index as a further endogenous variable in our system would simply shift the outlier problem from our equation for interest rates to the commodity price index. In fact no variable included in this system has an high explanatory potential for a commodity price index. We have then included in the system contemporaneous and lagged (up to the sixth lag) commodity price inflation. We consider commodity price inflation as a stationary exogenous variable, this choice shall be discussed later on. We have also included dummies for exceptional periods during the oil price crises as exogenous variables in our system. In general, dumMMYY is a variable taking value 1 in the MM month of the year YY and zero anywhere else. We include the following dummy variables: dum7306,dum7307, dum7308, dum7310, dum7311, dum7312, dum7402, dum7403, dum7407, dum7408,dum7409, dum7501, dum7505, dum7806, dum7808, dum7811, dum7904. Note that the inclusion of dummies and exogenous variables in the specification modifies the deterministic nucleus of our model and appropriate critical values for the tests should be re-computed(see**??**,**??**). We do not take this step and report the critical values automatically indicated by version 9 of Pc-Fiml in Table 7.

The inclusion of the dummies in the system delivers a new set of residuals, reported in Figure 2.13, which are virtually free from outliers and do not show any departure from the hypothesis of normality.

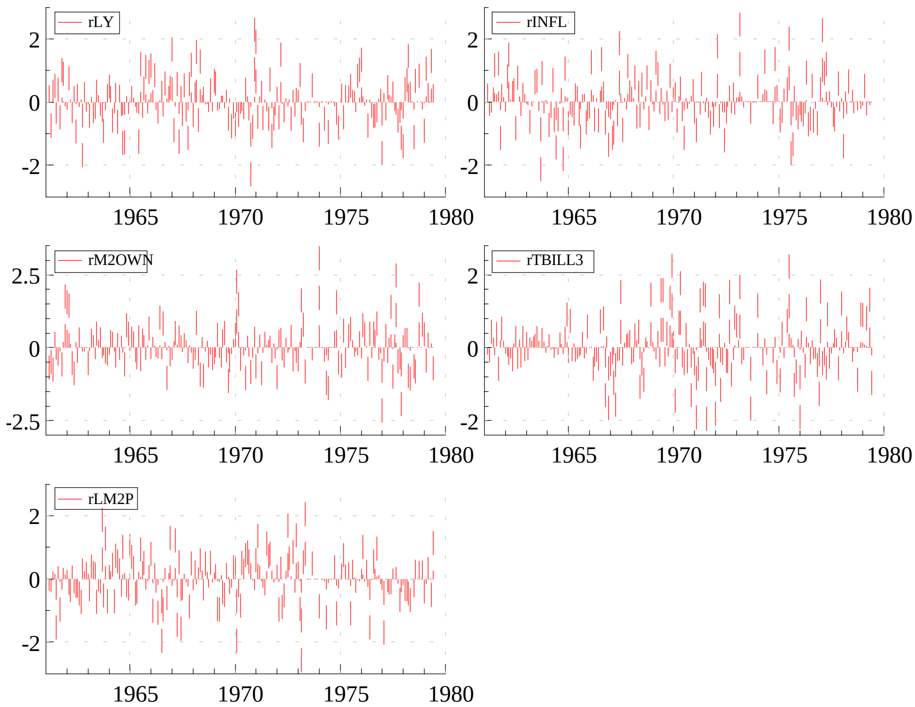


FIG. 2.13. VAR residuals without outliers

We proceed then to assess the possibility of simplification of the system. The progressive simplification strategy, based on the likelihood ratio tests discussed above, leads us to a specification with 6 lags.

2.8.2 *Selection of the deterministic components in the VECM specification*

The choice of the determistic components in the VAR is not trivial, given that it affects the distribution of the relevant statistics to perform cointegration analysis. Given the following general VECM model:

$$\Delta \mathbf{y}_t = \mu_0 + \mu_1 t + \Pi_1 \Delta \mathbf{y}_{t-1} + \Pi_1 \Delta \mathbf{y}_{t-2} + ... + \Pi \mathbf{y}_{t-n} + \mathbf{u}_t$$

five possible specifications for the deterministic components have been considered in the literature:

(i) $\mu_0 = 0, \mu_1 = 0$ this would determine a zero mean in the I(0) components and a non-zero mean in the I(1) components

(ii) $\mu_0 = \alpha\beta_1, \mu_1 = 0$ this would restrict the constant to belong to the cointegrating space inducing a non-zero mean both in the I(0) and the I(1) components

(iii) $\mu_0 = unrestricted, \mu_1 = 0$ this would generate a zero mean in the I(0) components and a linear trend in the I(1) component

(iv) $\mu_0 = unrestricted, \mu_1 = \alpha\beta_1$ this would generate a linear trend both in the I(0) and the I(1) components

(v) $\mu_0 = unrestricted, \mu_1 = unrestricted,$ this would generate a linear trend in the I(0) components and a quadratic trend in the I(1) components

Different critical values have been tabulated for each specification ([**?**]) and are now automatically available in all packages peforming the Johansen procedure. Note that the inclusion of intervention dummies also modifies the deterministic components of the VAR and this requires in principle in ad-hoc tabulation of the relevant critical values([11]).

In our application we choose specification (iv) as some of our series show trends in levels and, as already stated, we ignore the modification of the relevant distribution generated by the inclusion of dummies.

2.8.3 *Test for the rank of* Π

Having specified the VAR and chosen the specification of th deterministic component , we can estimate the Π matrix and start our analysis of the long-run properties of the system.We apply the Johansen procedure to identify the rank of the matrix Π in the following re-parameterisation of our model:

$$\begin{bmatrix} \Delta y_t \\ \Delta \pi_t \\ \Delta R_t^m \\ \Delta R_t^b \\ \Delta (m-p)_t \end{bmatrix} = \begin{bmatrix} d_{0,11} \\ d_{0,21} \\ d_{0,31} \\ d_{0,41} \end{bmatrix} + i = 0 \sum^{6} F_i \Delta_{12} LPCM_{t-i} + \mathbf{g}'\mathbf{DUM}_t +$$

$$i = 1 \sum^{5} D_i \begin{bmatrix} \Delta y_{t-i} \\ \Delta \pi_{t-i} \\ \Delta R_{t-i}^m \\ \Delta R_{t-i}^b \\ \Delta (m-p)_{t-i} \end{bmatrix} + \Pi \begin{bmatrix} y_{t-6} \\ \pi_{t-6} \\ R_{t-6}^m \\ R_{t-6}^b \\ (m-p)_{t-6} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ u_{4t} \\ u_{5t} \end{bmatrix}$$

The results of the Johansen procedure are reported in Table 6:

TABLE 6: Analysis of the Π matrix in the estimated VAR model

Eigenvalue	$H_0 : rank = p$	Eigenval	C.Eigenval	95%	Trace	C.Trace	95%
0.185	$p = 0$	45.74	39.61	37.5	134.1	116.1	87.3
0.148	$p \leq 1$	35.93	31.11	31.5	88.36	76.53	63
0.133	$p \leq 2$	32.14	27.84	25.5	52.44	45.41	42.4
0.083	$p \leq 3$	19.56	16.94	19	20.3	17.58	25.3
0.0033	$p \leq 4$	0.74	0.64	12.3	0.74	0.64	12.3

where Eigenvalis the max eigenvalue test, C.Eigenvalue is the max eigenvalue test corrected for small sample, i.e. using $T - k$ instead of T, Trace is the trace test, C.Trace is the trace test corrected for small sample amd 95% are the critical values tabulated for our specification of the deterministic components. Table 7 poses an interesting problem to the applied reseracher in that the trace statistics and the maximum eigenvalue statistic deliver different results, with more relevant differences in the case of the adoption of small sample correction for the statistics. We opt for rejecting the null of at most one cointegating vector and do not reject the null of at most two. Of course, such choice is debatable.

Note that, before any identifying restrictions are introduced, most available cointegrating packages do deliver some point estimates of the α and β matrices as follows:

TABLE 7: Cointegrating vectors: the Johansen interpretation.

		Standardised β'			
y	π	$m-p$	R^m	R^b	$Trend$
1	0.078	-0.40	3.55	-3.96	-0.18
1.08	1	-0.61	-1.20	-1.08	-0.15
		Standardised α			
y	-0.02	-0.013			
π	0.02	-0.005			
$m-p$	0.047	-0.018			
R^m	-0.00008	-0.002			
R^b	0.03	0.01			

These estimates are obtained by imposing a default identification which delivers cointegrating vectors orthogonal to each other ([36]). In some context, for example a demand and supply system, this assumption might be the economic case of interest. However, this is not the case in general and in our specific example. In the next section we shall evaluate the potential of different identification of economic interest by checking the validity of over-identifying restrictions.

2.8.4 *Specification and testing of the long-run restrictions*

We consider two different proposals. In the first one we identify a traditional money demand and a relation beween the own interest rate on money and the opportunity cost of holding money. As an alternative, we identify an interest rate reaction funcion in which the nominal interest rate responds to inflation, output and a linear trend, alongwith a relation between interest rates and inflation. We have selected these two specifications because, as we shall see, they form the basis for two alternative targeting strategies: inflation targeting via the control of money growth and inflation targeting via the control of interest rates.

We can parameterise the restrictions implied the first identification scheme as follows:

$$\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$$

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \\ \alpha_{41} & \alpha_{42} \\ \alpha_{51} & \alpha_{52} \\ \alpha_{61} & \alpha_{62} \end{pmatrix}$$

$$\boldsymbol{\beta}' = \begin{pmatrix} \beta_{11} & 0 & 1 & \beta_{41} & \beta_{51} & \beta_{61} \\ 0 & \beta_{22} & 0 & 1 & \beta_{52} & 0 \end{pmatrix}$$

The results, reported in Table 9, show that the two over-identifying restrictions are not rejected. The first cointegrating vector is consistent with a money demand function as far as the semi-elasticities with respect to interest rates are concerned, the elasticity with respect to income is somewhat high, although it is compensated by a deterministic trend with the opposite sign. However, looking at the weights on the cointegrating vectors we note that real money reacts very little to disequilibrium in the first cointegrating relationship. In fact the only strongly significant weight is the one describing the reaction of real income to disequilibrium in the second cointegrating relationship.

TABLE 9: A scheme of overidentified cointegrating vectors

Standardised β'					
y	π	$m-p$	R^m	R^b	*Trend*
(0.17)−2.20	0	1	(2.16)−7.29	(0.96)7.51	(0.06)0.38
0	(0.22)1.08	0	(0.59)−3.16	1	0

Standardised α		
y	(0.015)0.064	(0.036)−0.17
π	(0.009)−0.0016	(0.021)−0.034
$m-p$	(0.009)−0.019	(0.023)−0.014
R^m	(0.0001)−0.0006	(0.003)0.002
R^b	(0.008)−0.023	(0.02)0.03
LR-test, rank=2:$\chi^2(2) = 1.03$ [0.59]		

We then consider the second alternative and parameterise restrictions as follows:

$$\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$$

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \\ \alpha_{41} & \alpha_{42} \\ \alpha_{51} & \alpha_{52} \\ \alpha_{61} & \alpha_{62} \end{pmatrix}$$

$$\boldsymbol{\beta}' = \begin{pmatrix} \beta_{11} & -1 & 0 & 0 & 1 & \beta_{51} \\ 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix}$$

The results, reported in Table 9, show the plausibility of the interpretation of the first cointegrating relation as a reaction function for the monetary policy maker. Policy rates react to inflation, with a coefficient which can be restricted to one, and to deviation of output form a deterministic trend (a non-stationary variable in our specification). The estimated weights strongly support the identification of this relationship as an equilibrium for the policy rates. The second cointegrating vector does not differ significantly from the one obtained within the first identification scheme.

TABLE 9: A scheme of overidentified cointegrating vectors

Standardised β'					
y	π	$m-p$	R^m	R^b	$Trend$
(0.03)−0.22	−1	0	0	(0.68)1	(0.009)0.08
0	(0.17)−0.96	0	(0.47)2.75	1	0

Standardised α		
y	(0.069)0.13213	(0.039)−0.002
π	(0.038)0.047	(0.022)−0.017
$m-p$	(0.042)−0.09	(0.034)−0.09
R^m	(0.005)0.007	(0.0029)0.003
R^b	(0.036)−0.13	(0.021)−0.062

LR-test, rank=2: $\chi^2(4) = 6.1$ [0.19]

We conclude our analysis of these two alternative identification schemes by stressing that statistical criteria do not lead to an unequivocal identification, then the choice between the two alternatives is very likely to rely on economic criteria.

2.9 Multivariate decompositions: some considerations

The purpose of our illustration of the Johansen procedure in the previous section was to show that the identification of cointegrating vectors requires involves a multi-step process. The outcome of many of these steps is not so clear-cut and therefore the final product might be differ across researchers. The presence of structural breaks paired with the specification issues and size of available samples have an important impact on the empirical application of the procedure. Alternative methods to the this procedure have been proposed in the literature, see, for example, Horvath and Watson([28]), Phillips ([45]) , Reimers([47]) and Saikkonen([49]) . However, it is important to remember that the specification of a dynamic model in levels has proved sufficient to remove the spurious regression problem and that the VECM representation of a VAR model in level is just a reparameterisation, before the rank reduction restrictions are imposed. Sims,Stock and Watson([51]) , argue that a VAR model in levels in the presence of cointegration is over-parameterised and therefore leads to inefficient but consistent estimates of the parameters of interest. The loss of efficiency has to be weighted against the risk of inconsistency of estimates which occurs when the "wrong" cointegrating restrictions are imposed. Imposing the "wrong" cointegrating parameters will make the system converge to the "wrong" long-run equilibria but it will also bias the short-run dynamics as the system is pulled in the wrong direction. For this reason the recent research has taken a defined line and VAR in levels rather than co-integrated VARs are used when the issue of economic interest is not related to the short-run rather than to the long-run. A standard example is the analysis of the monetary transmission mechanism. Of course there is more in Sims,Stock and Watson([51]) than these considerations.

In fact they show that standard distribution can be applied when doing inference in a VAR model which involves variables admitting stationary linear combinations, reverting to non-standard distributions is necessary only when the subsets of variables on which inference is performed do not admit any stationary linear combination.

As a matter of fact cointegrated VARs are mainly data-driven specifications. The macro-model for the relevant DGP is not fully specified, as it is clearly the case with the example discussed in this chapter where we started off our investigation by a model centered on money and we end up specifying a long-run structure where the quantity of money, being fully demand determined, plays no role in the monetary transmission mechanism. It is not easy to interpret the results from a simultaneous model, when we have (loose) theories generating only a subset of the equations. Moreover, there are difficulties with an approach aimed discriminating between theories on the basis of the outcome of test statistics, based on a number of joint hypothesis, some of which are clearly independent from the theories tested. There is also an issue with the critical values for the testing procedures in the Johansen framework. First, they depend crucially on the specification of the deterministic nucleus of the VAR , so the inclusion of dummies for outliers introduces modifications in the relevant critical values. A solution to this problem is available, see Johansen-Nielsen[11]. Second, recent work by Johansen[35], has shown that it is important to implement small sample corrections for the asymptotic critical values, when applicable. Taking these two aspects together, it is likely that a re-assessment of all the empirical evidence proposed in the nineties without implementing the appropriate corrections is necessary. So what do we make of all the sentences issued on theories using the wrong critical values?

Note also that cointegration analysis based on a multi-step framework: specification of the VAR and its deterministic component, identification of the number of cointegrating vectors, identification of the parameters in cointegrating vectors, tests on the speed of adjustment with respect to disequilibria. The results of the final test depend on the outcome of the previous stages in the empirical analysis, but the outcome of each step is not so easily and uniquely established empirically.

Of course there is something to be said for a methodology aimed at exploiting cointegration to deliver a stationary representation of a non-stationary vector autoregressive process in which short-run and long-run dynamics are naturally separated and sound statistical inference can be applied. However, the practical implementation of such methodology requires theresearcher to deal with specification and identification problems which are not easily, and above all not uniquely, solved.

REFERENCES

Amisano G., Giannini C.(1997) Topics in Structural VAR Econometrics, second edition, Springer Verlag, New York

Banerjee A., Dolado J.J., Hendry D.F. and Smith G.W. (1986) Exploring Equilibrium Relationships in Econometrics through Static Models: some Monte Carlo Evidence, Oxford Bulletin of Economics and Statistics, August

Banerjee A., Dolado J.J., Galbraith J.W. and Hendry D.F.(1993) Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data, Oxford University Press, August

Banerjee A., Hendry D.F.(1992) "Testing Integration and Cointegration", Oxford Bulletin of Economics and Statistics, 54, Special Issue

Beveridge S. and C.-Nelson(1981) A New Approach to the Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to the Measurement of Business Cycle Journal of Monetary Economics, 151-174

Cuddington J.T. and L.A.Winters(l987) "The Beveridge-Nelson decomposition of economic time-series. A quick computational method" Journal of Monetary Economics,19, 125-127.

Bhargava A.(1986) "On the Theory of Testing for Unit Roots in Observed Time Series", Review of Economic Studies,53 pp. 369-384

DeJong D.N. and C.H. Whiteman (1991) Reconsidering "Trends and random walks in macroeconomic time-series" , Journal of Monetary Economics, 28, 221-254

Dickey D.A.,Fuller W.A.(1979) "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", Econometrica,49,pp.1057-1072

Doornik J. and Hendry D.F.(1994)"PcFiml 8.0. Interactive Econometric Modelling of Dynamic Systems", International Thomson Publishing, London

Enders W.(1995) Applied Econometric Time Series, Wiley Series in Probability and Mathematical Statistics, New York

Engle R.F.,Granger C.W.J.(1987) "Cointegration and Error Correction : Representation, Estimation and Testing" Econometrica,55,pp.251-276

Engle R.F.,Granger C.W.J.,Hallman J.J.(1987) "Combining Short and Long Run Forecasts: an Application of Seasonal Cointegration to Monthly Electricity Sales Forecasting",Unpublished paper, University of California San Diego

Fuller W.A. (1976) Introduction to Statistical Time Series, New York,J.Wiley ans Sons

Giannini C. (1992) Topics in Structural VAR Econometrics, Lecture Notes in Economics and Mathematical Systems, Springer-Verlag

Giannini C., Lanzarotti S. and M. Seghelini (1994) "A Traditional Interpretation of Macroeconomi Fluctuations: the Case of Italy.", European Journal of Political Economy

Granger C.W.J.(1986) "Developments in the Study of Cointegrated Economic Variables",Oxford Bulletin of Economic and Statistics,48,pp.213-228

Gonzalo J. (1994) "Five Alternative Methods of Estimating Long-run Equilibrium Relationships", Journal of Econometrics, 60, 203-233

Granger C.W.J.,Newbold P.(1974) "Spurious Regressions in Econometrics", Journal of Econometrics,2, pp.111-120

Hamilton J. (1994) Time-Series Analysis, Princeton, Princeton University Press,

Hansen H. and K.Juselius(1995) "CATS in RATS. Cointegration Analysis of Time Series", Estima, Evanston, Illinois

Harvey,A.C. and Jaeger A.(1993) 'Detrending, Stylized Facts and the Business Cycles' Journal of Applied Econometrics

Hatanaka M.(1996) Time-Series-based-econometrics, Oxford university Press, Oxford

Hendry D.F.(1987) "Econometric Modelling with Cointegrated Variables", Oxford Bulletin of Economic and
Statistics,48,pp.201-212

Hendry D.F. (1995) Dynamic Econometrics Oxford University Press

Hendry D.F. and N.Ericsson(1991) Modeling the demand for narrow money in the United Kingdom and the United States, European Economic Review, 35, 833-886

Hodrick, R.J. and E.C. Prescott (1997) Postwar U.S. Business Cycles: An Empirical Investigation, Journal of Money, Credit, and Banking, 29, 116.

Horvath M.T.K., and M.W. Watson(1995) "Testing for cointegration when some of the cointegrating vectors are known" Econometric Theory, 11, 984-1014

Kiwiatkowski D., P.C.B. Phillips, P.Schmidt and Y. Shin (1992) Testing the Null Hypothesis of Stationarity against the alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root ?, Journal of Econometrics, 54, 159-178

Johansen S. (1988), Statistical Analysis of Cointegration Vectors, "Journal of Economics Dynamics and Control, March.

Johansen S. (1989) Likelihood-based Inference on Cointegration: Theory and Applications, mimeo, University of Copenhagen

Johansen S.(1992) Identifying Restrictions of Linear Equations, University of Copenaghen, Institute of Mathematical Statistics

Johansen S.(1994) The role of the Constant and Linear Terms in Cointegration Analysis of Nonstationary Variables, Econometric Reviews,13, 205-229

Johansen S. (1995) Likelihood Based Inference on Cointegration in the Vector Autoregressive Model, Oxford University Press, Oxford

Johansen, S. (1999), "A Bartlett Correction Factor for Tests on the Cointegrating Relationships", downloadable from http://www.iue.it/Personal/Johansen/Welcome.html

Johansen S., Juselius, K. (1990), Maximum Likelihood Estimation and Inference on Cointegration with Applications to the Demand for Money, "Oxford Bulletin of Economics and Statistics", May.

Johansen S., Nielsen B.G.(1993) Asymptotics for the cointegration rank test tests in the presence of intervantion dummies, Manual for the simulation program DisCo, manual and program are available at the URL http://www.nuff.ox.ac.uk/users/nielsen/d

Leeper E.M., C.A. Sims and T. Zha (1997) "What does monetary policy do?", Brookings Paper on Economic Activity, also available at ftp://ftp.econ.yale.edu/pub/sims/mpolicy

Maddala G.S., and In-Moo Kim(1998) Unit roots, Cointegration and Structural Change, Cambridge University Press, Cambridge

Mellander E., Vredin A. e A.Warne (1993) "Stochastic Trends and Economic Fluctuations in a Small Open Economy" Journal of Applied Econometrics

Mosconi R.(1998) "Malcolm. The theory and practice of cointegration analysis in RATS", Cafoscarina, Venezia

Nelson, C.R. e C.I. Plosser(1982) Trends versus Random Walks in economic time-series. Some evidence and implications, Journal of Monetary Economics, 10, 139-162

Osterwald-Lenum M.(1992) "A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics", Oxford Bulletin of Economics and Statistics, 54, 461-472

Perron P.(1987) "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis", Research Paper, Department of Economics, University of Montreal,Quebec, Canada

Phillips P.C.B. (1991) Optimal Inference in Cointegrated Systems, Econometrica, 59, 283-306

Phillips P.C.B. e P. Perron (1988) Testing for a unit root in time-series regression, Biometrika, 75, 335-346.

Reimers H.E. (1992) "Comparison of tests for multivariate cointegration", Statistical Papers, 33, 335-359

Salmon M.(1982) "Error Correction Mechanisms",The Economic Journal,92,pp.615-629.

Saikkonen P.(1993) Estimation of cointegration Vectors with Linear Restrictions, Econometric Theory, 9, 19-35

Sargan D. (1964) Wages and Prices in the United Kingdom: A Study in Econometric Methodology, in P.E. Hart, G.Mills and J.K. Whitaker (eds.), Econometric Analysis of National Economic Planning, London, Butterworth, 25-63

Sargan J.D.,Barghava A.(1983) "Testing Residuals from Least Squares Regression for Being Generated by Gaussian Random Walk",Econometrica,51,pp.153-174

Sims C.(1980) "Macroeconomics and Reality" Econometrica

Sims C., Stock J. e Watson M.(1990) Inference in Linear Time Series Models with Some Unit Roots Econometrica, 58, 1, 113-144

Spanos A.(1986)"Statistical Foundations of Econometric Modelling",Cambridge University Press,Cambridge

Stock J.H.(1987)"Asymptotic Properties of the Least Squares Estimator of Cointegrating Vectors." Econometrica, 55,pp. 1035-1056

Stock J.H.,West K.D.(1987) "Integrated Regressors and the Test of the Permanent Income Hypothesis",Princeton University,Woodrow Wilson School D.P. 127

Zivot, E. e D.W.K. Andrews(1992) Further Evidence on the Great Crash, the Oil-Price Shock and the Unit Root Hypothesis", Journal of Business Economics and Statistics,10, 251.270

Mellander E., A.Vredin and A.Warne(1992) Stochastic trends and economic fluctuations in a small open economy, Journal of Applied Econometrics, 7, 369-394

King R.G., C.I.Plosser, J.H.Stock,M.W.Watson(1991) "Stochastic trends and economic fluctuations" the American Economic Review, 81,4, 820-840

White H. (1984) "Asymptotic theory for econometricians", Academic Press, London

3

THE IDENTIFICATION PROBLEM IN MACROECONOMETRICS

3.1 Introduction

VAR in levels and VECM representations specify the probabilistic structure of the data. Consider the case of an empirical investigation of the monetary transmission mechanism gear to evaluate the impact of monetary policy on macroeconomic variables, and partition the vector of n variables of interest into two subsets: $\mathbf{Y}$,which represesents the vector of macroeconomic variables of interest and $\mathbf{M}$, the vector of monetary variables determined by the interaction between the monetary policy maker and the economy. As we have seen that the VECM is obtained by imposing restrictions on the VAR, consider the general unrestricted system:

$$\begin{pmatrix} \mathbf{Y}_t \\ \mathbf{M}_t \end{pmatrix} = D_1(L) \begin{pmatrix} \mathbf{Y}_{t-1} \\ \mathbf{M}_{t-1} \end{pmatrix} + \mathbf{u}_t \tag{3.1}$$

$$\mathbf{u}_t = \begin{pmatrix} \mathbf{u}_t^Y \\ \mathbf{u}_t^M \end{pmatrix}$$

$$u_t \mid I_{t-1} \sim n.i.d. \left(\mathbf{0}, \sum\right)$$

$$\begin{pmatrix} \mathbf{Y}_t \\ \mathbf{M}_t \end{pmatrix} \mid I_{t-1} \Bigg) \sim \left(D_1(L) \begin{pmatrix} \mathbf{Y}_{t-1} \\ \mathbf{M}_{t-1} \end{pmatrix}, \sum \right)$$

This system specifies the statistical distribution for the vector of variables of interest conditional upon the information set available at $t-1$. In the case of a VECM specification, after the solution of the identification problems of cointegrating vectors, the information set available at $t-1$ contains n lagged endogenous variables and r cointegrating vectors. We face an identification problem in that there is more than one structure of economic interest which can give rise to the same statistical model for our vector of variables.

In fact for any given structure,

$$\mathbf{A} \begin{pmatrix} \mathbf{Y}_t \\ \mathbf{M}_t \end{pmatrix} = \mathbf{C}_1(L) \begin{pmatrix} \mathbf{Y}_{t-1} \\ \mathbf{M}_{t-1} \end{pmatrix} + \mathbf{B} \begin{pmatrix} \boldsymbol{\nu}_t^Y \\ \boldsymbol{\nu}_t^M \end{pmatrix} \tag{3.2}$$

$$\left(\begin{matrix} \boldsymbol{\nu}_t^Y \\ \boldsymbol{\nu}_t^M \end{matrix} \mid I_{t-1} \right) \sim (\mathbf{0}, \mathbf{I}) \tag{3.3}$$

which give rise to the observed reduced form (4.11) when the following restrictions are satisfied:

$$\mathbf{A}^{-1}\mathbf{C}_1(L) = D_1(L), \; \mathbf{A}\begin{pmatrix}\mathbf{u}_t^Y \\ \mathbf{u}_t^M\end{pmatrix} = \mathbf{B}\begin{pmatrix}\boldsymbol{\nu}_t^Y \\ \boldsymbol{\nu}_t^M\end{pmatrix}$$

there exists a whole class of structures which give rise to the same statistical model (4.11) under the same class of restrictions :

$$F\mathbf{A}\begin{pmatrix}\mathbf{Y}_t \\ \mathbf{M}_t\end{pmatrix} = F\mathbf{C}_1(L)\begin{pmatrix}\mathbf{Y}_{t-1} \\ \mathbf{M}_{t-1}\end{pmatrix} + F\mathbf{B}\begin{pmatrix}\boldsymbol{\nu}_t^Y \\ \boldsymbol{\nu}_t^M\end{pmatrix} \tag{3.4}$$

where F is an admissible matrix in the sense it is conformable by product with $\mathbf{A}$, $\mathbf{C}_1(L)$, $\mathbf{B}$ and $F\mathbf{A}$, $F\mathbf{C}_1(L), F\mathbf{B}$ feature the same restriction with $\mathbf{A}$, $\mathbf{C}_1(L)$, $\mathbf{B}$.

3.1.1 *Identifiability*

A model is identifiable if all its possible structure are identifiable, i.e. each structure is associated to a different distribution, this happens when the only admissible F matrix is identity.

Let us show the point by considering identification of the first equation. In order to achieve identification some restrictions must be imposed, as the number of parameters in the reduced form (4.11) is smaller than the number of parameters in the structure (3.2) . For the sake of exposition we begin by considering zero restrictions on the matrices $\mathbf{A}$ and $\mathbf{C}_1$, determining the first moment of the conditional distributions of the vector of variables of interest and concentrate on a first order autoregressive representation:

$$\mathbf{A}\begin{pmatrix}\mathbf{Y}_t \\ \mathbf{M}_t\end{pmatrix} = \mathbf{C}_1\begin{pmatrix}\mathbf{Y}_{t-1} \\ \mathbf{M}_{t-1}\end{pmatrix} + \begin{pmatrix}\boldsymbol{\epsilon}_t^Y \\ \boldsymbol{\epsilon}_t^M\end{pmatrix} \tag{3.5}$$
$$\begin{pmatrix}\boldsymbol{\epsilon}_t^Y \\ \boldsymbol{\epsilon}_t^M\end{pmatrix} = \mathbf{B}\begin{pmatrix}\boldsymbol{\nu}_t^Y \\ \boldsymbol{\nu}_t^M\end{pmatrix}$$

Zero restrictions on the first equations can be represented as follows:

$$\begin{matrix}1 \\ n-1\end{matrix}\begin{bmatrix}1, a_1' & 0 \\ A_1^* & A_1\end{bmatrix}, \begin{bmatrix}c_1' & 0 \\ C_1^* & C_1\end{bmatrix}$$

so $(n - n_1)$ endogenous variables and $(n - k_1)$ exogenous variables are restricted to zero. A_1^* is a $((n-1) \times n_1)$ matrix containing the coefficients with which the $n_1$variables entering contemporaneously the first equation enters the remaining $n-1$ equations in the system, A_1 is a $((n-1)(n-n_1))$ matrix containing the coefficients with which the $n-n_1$variables entering contemporaneously the first equation enters the remaining $n-1$ equations in the system. Analogously, C_1^* is a $((n-1) \times k_1)$ matrix containing the coefficients with which the

k_1 variables entering with a lag in the first equation enters the remaining $n-1$ equations in the system, C_1 is a $((n-1)(n-k_1))$ matrix containing the coefficients with which the $n-n_1$ variables not entering with a lag in the first equation enters as lagged variables the remaining $n-1$ equations in the system. Represent the first row of F as $\begin{pmatrix} 1 & f' \end{pmatrix}$. Admissibility implies

$$f'(A_1 \mid C_1) = 0$$

in fact only when these conditions are satisfied the first row of $F\mathbf{A}$, $F\mathbf{C}_1$ feature the same restriction with $\mathbf{A}$, $\mathbf{C}_1$.

Identification implies that the only solution is $f\prime = 0$, as the first rows of an $(n \times n)$ identity matrix has unity as the first element and zeroes as the remaining $(n-1)$ elements. Therefore the condition for identification is

$$rank(A_1 \mid C_1) = n-1.$$

This is a necessary and sufficient condition for identification. As $(A_1 \mid B_1)$ has $(n-1)$ rows and $(n-n_1)+(n-k_1)$ columns, a necessary condition for identification is that the number of colums is sufficiently large to let the rank be equal to $n-1$:

$$n-k_1 \geq n_1 - 1.$$

Therefore, in order for the first equation to be identified, we need that the number of omitted lagged variables must be greater than the number of included contemporaneous variables minus one (the one variable with respect to which the equation is normalized). At this point it is natural to state that the mode is not identified when $n-k_1 < n_1-1$, the model is just-identified when $n-k_1 = n_1-1$, the model is over-identified, with $n+1-(n_1+k_1) >$ over-idenifying restrictions, when $n-k_1 > n_1-1$.

This discussion of identifiability can be generalised to non-zero restrictions by considering the following representation:

$$\begin{pmatrix} \mathbf{Y}_t \\ \mathbf{M}_t \\ \mathbf{Y}_{t-1} \\ \mathbf{M}_{t-1} \end{pmatrix} = \mathbf{w}_t, (\mathbf{A} \mid -\mathbf{C}_1) = \mathbf{D}$$

$$\epsilon_t = \begin{pmatrix} \epsilon_t^Y \\ \epsilon_t^M \end{pmatrix}$$

the system becomes

$$\mathbf{D}\mathbf{w}_t = \epsilon_t \tag{3.6}$$

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_1' \\ . \\ \mathbf{D}_n' \end{pmatrix}$$

General restriction on the i-th equation can be represented as:

$$\mathbf{R}_i \mathbf{D}_i = 0,$$

where $\mathbf{R}_i$ is the $(k_i \times (n+n))$ matrix imposing k_i restrictions on the $2n$ elements of i-th equation in the system.

Necessary and sufficient condition for identification are then

$$rank \ \ \mathbf{R}_i \left(\mathbf{D}_1 \mid \mathbf{D}_2 ... \mid \mathbf{D}_n \right) = n - 1$$

which shows the equivalences between the conditions for short-run identification and the conditions required to achieve long-run identification of cointegrating parameters. We end this section by noting the short-run and long-run identification in a VECM are two completely separated problems ([16]) . Consider the simplest VECM representation of a first order VAR:

$$\begin{aligned} \Delta \mathbf{y}_t &= \left(\mathbf{A}_1 - \mathbf{I} \right) \mathbf{y}_{t-1} + \mathbf{u}_t \\ &= \Pi \mathbf{y}_{t-1} + \mathbf{u}_t . \end{aligned}$$

When the long-run identification problem is solved we decompose Π into $\boldsymbol{\alpha\beta}'$ and we can re-write the VECM as follows:

$$\begin{aligned} \Delta \mathbf{y}_t &= \boldsymbol{\alpha} \mathbf{z}_{t-1} + \mathbf{u}_t \\ \mathbf{z}_{t-1} &= \boldsymbol{\beta}' \mathbf{y}_{t-1} \end{aligned}$$

which makes clear that the identification of parameters in the structural form:

$$\mathbf{A} \Delta \mathbf{y}_t = \mathbf{A} \boldsymbol{\alpha} \mathbf{z}_{t-1} + \mathbf{A} \mathbf{u}_t$$

is independent from the identification of parameters in the matrix $\boldsymbol{\beta}$.

3.2 Identification in the "Cowles Commission" approach

The traditional, usually referred to as the Cowles Commission, approach to econometric modelling of the monetary transmission mechanism is aimed at the quantitative evaluation of the effects of modification in the exogenous variables in the system on the endogenous variables in the system. Variables controlled by the monetary policy maker (the instruments of monetary policy) are taken as exogenous, while macroeconomic variables, which represents the final goals of the policy maker, are taken as endogenous. The policy experiment of interest consists usually in modifying such exogenous variables to assess the impact on the macroeconomic variables of interest. Leaving aside the deterministic component, the Cowles Commission specification modifies the general dynamic model of the previous section , as follows:

$$\mathbf{A}\begin{pmatrix}\mathbf{Y}_{t-1}\\ \mathbf{M}_{t-1}\end{pmatrix} = \mathbf{C}_1(L)\begin{pmatrix}\mathbf{Y}_{t-1}\\ \mathbf{M}_{t-1}\end{pmatrix} + \mathbf{C}_2(L)\left(\bar{M}_t\right) + \begin{pmatrix}\epsilon_t^Y\\ \epsilon_t^M\end{pmatrix} \tag{3.7}$$

where $\mathbf{Y}$ represents the vector of macroeconomic variables of interest, while $\mathbf{M}$ is the vector of monetary variables determined by the interaction between the monetary policy maker and the economy and $\bar{M}$ represents a sub-vector of the monetary variables, assumed exogenous because directly and fully controlled by the monetary policy maker. The process generating these variables does not contain any interaction with the other variables in the system. $\mathbf{C}_1(L)$, $\mathbf{C}_2(L)$ are polynomials in the lag operator L, taking the general form $\mathbf{C}_i(L) = c_0 + c_1 L + c_2 L^2 + ... c_n L^n$.

The general conditions for identification derived in the previous section are applicable to the above specification. In fact the consideration of some variables as exogenous aids identification in that exogenous variables are treated, from the point of view of identification, as the lagged endogenous variables. Considering the general conditions for identification

$$rank \;\; \mathbf{R}_i\left(\mathbf{D}_1 \mid \mathbf{D}_2 ... \mid \mathbf{D}_n\right) = n - 1$$

note that the inclusion of exogenous variables increase the columns of the matrix $\mathbf{R}_i\left(\mathbf{D}_1 \mid \mathbf{D}_2 ... \mid \mathbf{D}_n\right)$ and therefore the chances for the model to be identified.

3.2.1 *An illustration: identifying the IS-LM-AD-AS model*

Let us consider the simplest possible macroeconomic model for a closed economy to illustrate how conditions for identification can be checked. The model consists of four equations:

$$y_t = c_{11} + y_t^* - a_{13}(R_t - \pi_t^e) + \epsilon_{1t} \tag{3.8}$$
$$\pi_t = \pi_t^e + a_{21}(y_t - y_t^*) + \epsilon_{2t} \tag{3.9}$$
$$m_t - p_t = c_{13} + a_{31}y_t - a_{33}R_t + \epsilon_{3t} \tag{3.10}$$
$$\pi_t^e = \beta\pi_{t-1} + (1-\beta)\pi_t + \epsilon_{4t}. \tag{3.11}$$

Equation (3.10) describes an LM equation, which, for $a_1 = 1$, relates the nominal interest rate to (the log of) money velocity circulation. Equation (7.57) describes an IS curve in close economy and shows immediately that monetary policy authority can influence the level of activity only if, by controlling the nominal interest rate, it manages to influence the real interest rate. Equation (7.58) is an expectations augmented Phillips curve, according to which actual inflation is determined by expected inflation and deviations of output from its potential level y^*, which we take as a deterministic trend. Equation (3.11) describes the mechanism with which expectations are formed. The extreme case of price-stickines is obtained by posing $\beta = 1$, while the case of rational expectations-perfect price flexibility is obtained by posing $\beta = 0$.

Note that no equation for money is included in the model . In fact money supply is not modelled as it is considered exogenous, i.e. fully controlled by the monethary authority. The econometrician's task is the estimation of the unknown parameters to simulate the impact of different path for the exogenous variable controlled by the monetary authority. The model uses four equations to determine four endogenous variables, π_t, π_t^e, R_t ed y_t, for given values of the two exogenous variables y_t^* and m_t. The exogeneity status is attributed to y_t^* and m_t, either because they describe the available technology and demography or because they are fully controlled by the policy-maker.

Consider the extreme case of price-stickiness, given by $\beta = 1$, and use equation (3.11) to eliminate expected inflation from the model, the IS-LM-AD-AS model can be specified as a special case of our general specification (3.7) :

$$\mathbf{A}\begin{bmatrix} y_t \\ p_t \\ R_t \\ m_t \\ y_t^* \end{bmatrix} = \mathbf{C}_0\begin{bmatrix} 1 \\ t \end{bmatrix} + \mathbf{C}_1\begin{bmatrix} y_{t-1} \\ p_{t-1} \\ m_{t-1} \end{bmatrix} + \mathbf{C}_2\begin{bmatrix} y_{t-2} \\ p_{t-2} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \\ \epsilon_{4t} \\ \epsilon_{5t} \end{bmatrix} \tag{3.12}$$

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & a_{13} & 0 & -1 \\ -a_{21} & 1 & 0 & 0 & a_{21} \\ -\frac{a_{31}}{a_{33}} & -\frac{1}{a_{33}} & 1 & \frac{1}{a_{33}} & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{C}_0 = \begin{bmatrix} c_{0,11} & 0 \\ c_{0,21} & 0 \\ c_{0,31} & 0 \\ c_{0,41} & 0 \\ c_{0,51} & c_{0,52} \end{bmatrix}, \quad \mathbf{C}_1 = \begin{bmatrix} 0 & a_{13} & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{C}_2 = \begin{bmatrix} -a_{13} & 0 \\ 0 & -1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Note that y_t^* is a deterministic trend and money is exogenous in that its rate of growth is fixed to c_{41}, the effect of monetary policy on macroeconomic variables is evaluated by assessing the impact on the system of modifications in this parameter. To apply to this specific case our general discussion of identifiability consider that, using the representation (3.6) , we have:

$$\mathbf{w}_t = \begin{bmatrix} y_t \\ p_t \\ R_t \\ m_t \\ y_t^* \\ 1 \\ t \\ y_{t-1} \\ p_{t-1} \\ m_{t-1} \\ y_{t-2} \\ p_{t-2} \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_1' \\ \mathbf{D}_2' \\ \mathbf{D}_3' \\ \mathbf{D}_4' \\ \mathbf{D}_5' \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 0 & a_{13} & 0 & -1 & c_{0,11} & 0 & 0 & a_{13} & 0 & -a_{13} & 0 \\ -a_{21} & 1 & 0 & 0 & a_{21} & c_{0,21} & 0 & 0 & 2 & 0 & 0 & -1 \\ -\frac{a_{31}}{a_{33}} & -\frac{1}{a_{33}} & 1 & \frac{1}{a_{33}} & 0 & c_{0,31} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & c_{0,41} & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & c_{0,51} & c_{0,52} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The restrictions in the first equations are imposed by the following matrix:

$$R_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The first equation is then over-identified with five (nine-four) over-identifying restrictions when the following rank condition is satisfied:

$$rank\ \ \mathbf{R}_1 \left(\mathbf{D}_1 \mid \mathbf{D}_2 ... \mid \mathbf{D}_5 \right) = 4$$

$$rank \begin{bmatrix} -a_{21} & 1 & 0 & 0 & a_{21} & c_{0,21} & 0 & 0 & 2 & 0 & 0 & -1 \\ -\frac{a_{31}}{a_{33}} & -\frac{1}{a_{33}} & 1 & \frac{1}{a_{33}} & 0 & c_{0,31} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & c_{0,41} & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & c_{0,51} & c_{0,52} & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = 4.$$

By applying the procedure to all the equations in the system it can be shown that the second equation is over-identified with five over-identifying restrictions, the third equation is over-identified with three over-identifying restrictions, the fourth equation is over-identified with six over-identifying restrictions, the fifth equation is over-identified with five over-identifying restrictions. We conclude that the model is identified and imposes a total twenty-four over-identifying restrictions.

3.3 The great critiques

Cowles Commission approach to identification of structural econometric models broke down in the seventies when it was discovered that this type of models

> "...did not represent the data, ... did not represent the theory... were ineffective for practical purposes of forecasting and policy evaluation..." ([17]).

Different explanations of the failure of Cowles Commission approach gave rise to the different prominent methods of empirical research: the LSE (London School of Economics) approach, the VAR approach, and the intertemporal optimization-Real Business Cycle approach. We shall discuss and illustrate the empirical research strategy of these three alternative approaches by interpreting them as different proposals to solve problems observed in the Cowles Commission approach.

The LSE approach was initiated by Denis Sargan but owes its diffusion to a number of Sargan's students and it extremely well described in the book by

David Hendry ([9]) . This approach to macroeconometric modelling explains the ineffectiveness of the Cowles Commission models for practical purposes of forecasting and policy in terms of their incapability of representing the data. The root of the failure of the traditional approach lies in the little attention paid to the statistical model implicit in the estimated structure. Consider our simple example of the IS-LM-AD-AS model, the identified structure is estimated without checking that the implicit statistical model is an accurate description of the data. Spanos ([20]) considers tha case of a simple demand and supply models to show how reduced form are ignored in the traditional approach, in fact most of the widely used estimators allow to derive numerical values for the strucutral parameters without even seeing the statistical models represented by the reduced form. There are several possible causes for the inadeguacy of statistical models implicit in structural econometric models: omission of relevant variables, omission of the relevant dynamics for the included variables (note, for example, that the estimated money demand in our simple example relation is a simple, static equation), invalid assumptions of exogeneity. The LSE solution to the specification problem is the theory of reduction. Any econometric model is interpreted as a simplified representation of the unobservable Data Generating Process (DGP). For the representation to be valid or "congruent", to use Hendry's own terminology, the information lost in moving from the DGP to the its representation, given by the adopted specification, must be irrelevant to the problem at hand. Adequacy of the statistical model is evaluated by analyzing the reduced form. Therefore, the prominence of structural model with respect to reduced form representation in the Cowles Commission approach to identification and specification is reversed. The LSE approach starts its specification and identification procedure by specifying a general dynamic reduced form model. The congruency of such model cannot be directly assessed against the true DGP, which is not observable. However, a series of diagnostic tests are proposed as criteria for evaluating the congruency of the baseline model. The general principle guiding the application of such criteria is that congruent models should feature true random residuals, hence any departure of the vector of residuals from a random normal multivariate distribution should be taken as a symptom of mis-specification. Once the base-line model has been validated, the reduction process begins by simplifying the dynamics and by reducing the dimensionality of the model by omitting to include equtions for those variables for which the null hypothesis of exogeneity is not rejected. In fact the concept of exogeneity is refined within the LSE approach and is broken down in different categories, determined by the purpose of the estimation of the econometric model. A further stage in the simplification process could be the imposition of the rank reduction restrictions in the matrix determining the long-run equilibria of the system and the identification of cointegrating vectors. The product of this stage is a statistical model for the data, possibly discriminating between short-run dynamics and long-run equilibria. Only after this validation procedure structural model are identified and estimated. Just-identified specification do not require any further testing, as their implicit reduced form does not impose any further restrictions on the base-line statistical model. The validity of over-identified specification is instead tested by evaluating the validity of the restrictions implicitly imposed on the general reduced form. After this last diagnostics for the validity of the reduction process have been performed, structural models are used for the practical purposes of forecasting and policy evaluation.

If the LSE approach finds its explanation of the failure of Cowles Commission models in their incapability of representing the data, different approaches, initiated by two famous critiques by Lucas ([36]) and Sims ([48]) , relate their explanations of the failure to the incapability of Cowles Commission models to represent the theory. The general class of theoretical models of reference for these two critiques are forward-looking intertemporal optimization models. Lucas attacks the identification scheme proposed by the Cowles Commission by pointing out that these model do not take explicitly into account expectations and therefore identified parameters within the Cowles Commission approach are in fact mixture of "deep parameters" describing preference and technology into the economy and expectational parameters which, by their nature, are not stable across different policy regimes. The main consequence of such instability across different regime is that traditional structural macro-model are useless for the purpose of policy simulation. To show the point let us reconsider the case of our simple model of the monetary transmission mechanism estimated for simulating the impact of different monetary policies on macroeconomic variables.

Assume the following DGP, in which expected monetary policy matters for the determination of macroeconomic variables in the economy:

$$\begin{pmatrix} \mathbf{Y}_t \\ \mathbf{M}_t \end{pmatrix} = D_1(L) \begin{pmatrix} \mathbf{Y}_{t-1} \\ \mathbf{M}_{t-1} \end{pmatrix} + D_2 \left(\bar{M}^e_{t+1} \right) + \begin{pmatrix} \mathbf{u}^Y_t \\ \mathbf{u}^M_t \end{pmatrix} \tag{3.13}$$

A Cowles Commission model is estimated without explicitly including expectations over a sample featuring the following money supply rule:

$$M_{t+1} = a_0 + M_t. \tag{3.14}$$

The fitted model will therefore have the following specification:

$$\begin{pmatrix} \mathbf{Y}_t \\ \mathbf{M}_t \end{pmatrix} = D_1(L) \begin{pmatrix} \mathbf{Y}_{t-1} \\ \mathbf{M}_{t-1} \end{pmatrix} + D_2 a_0 + D_2 \left(M_t \right) + \begin{pmatrix} \mathbf{u}^Y_t \\ \mathbf{u}^M_t \end{pmatrix}. \tag{3.15}$$

(3.15) is not the correct model to simulate the impact of any rule different from (3.14) . Consider the case of :

$$M_{t+1} = a_1 + M_t$$

The correct model, in terms of observable variables, for simulating the effect of the new policy would be:

$$\begin{pmatrix} \mathbf{Y}_t \\ \mathbf{M}_t \end{pmatrix} = D_1(L) \begin{pmatrix} \mathbf{Y}_{t-1} \\ \mathbf{M}_{t-1} \end{pmatrix} + D_2 a_1 + D_2 (M_t) + \begin{pmatrix} \mathbf{u}_t^Y \\ \mathbf{u}_t^M \end{pmatrix} \tag{3.16}$$

and simulation based on (3.15) would give the wrong prediction on the effect of monetary policy.

The Sims ([48]) critique runs parallel to the Lucas critique but concentrate on the statust of exogeneity arbitrarily attributed to some variables to achieve identification within structural Cowles Commission models. Sims argues that no variable can be deemed as exogenous in a world of forward-looking agents whose behaviour depends on the solution of an intertemporal optimization model. Again with reference to our example, reconsider the case for exogeneity of money supply. If the monetary authority uses money supply as an instrument to achieve given targets for the macroeconomic variables, then it would be very "natural" for money supply to react not only the output and inflation but also to leading indicators for these variables.Assuming money supply as exogenous, the estimated model omits completely a very relevant feedback and looses an important feature of the data. Moreover, by assuming incorrectly exogeneity, the model might induce a spurious statistic efficacy of monetary policy in the determination of macroeconomic variables. The endogeneity of money does generate a correlations between macroeconomic variables and monetary variables, which, by assuming invalidly, money as exogenous could be interpreted as a causal relation running from money to the macroeconomic variables.

We shall devote to some deeper illustration of the different approaches to identification in response to the problems with the Cowles Commission approach, we shall then devote the rest of the book to the illustration of how such different approaches are put to work to constrcut macroeconometric models.

3.4 Identification in the LSE methodology

To illustrate the LSE approach to identification re-consider the Cowles Commission specification of the IS-LM-AS model. The Cowles Commission's strategy direct specifies of an identified structure of interest such as (5.1) . The LSE research strategy begins form the specification of a general statistical model, i.e. a reduced form. There are no specific rule in the choice of the baseline specification, the only constraint being that such specification should be sufficently general to deliver a congruent representation of the underlying unknown Data Generating Process. A possible baseline specification could be the following:

$$\begin{bmatrix} y_t \\ p_t \\ R_t \\ m_t \\ y_t^* \end{bmatrix} = \begin{bmatrix} d_{0,11} & d_{0,12} \\ d_{0,21} & d_{0,22} \\ d_{0,31} & d_{0,32} \\ d_{0,41} & d_{0,42} \\ d_{0,51} & d_{0,52} \end{bmatrix} \begin{bmatrix} 1 \\ t \end{bmatrix} + \tag{3.17}$$

$$+ \begin{bmatrix} d_{1,11} & d_{1,12} & d_{1,13} & d_{1,14} & d_{1,15} \\ d_{1,21} & d_{1,22} & d_{1,23} & d_{1,24} & d_{1,25} \\ d_{1,31} & d_{1,32} & d_{1,33} & d_{1,34} & d_{1,35} \\ d_{1,41} & d_{1,42} & d_{1,43} & d_{1,44} & d_{1,45} \\ d_{1,51} & d_{1,52} & d_{1,53} & d_{1,54} & d_{1,55} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ p_{t-1} \\ R_{t-1} \\ m_{t-1} \\ y_{t-1}^* \end{bmatrix} +$$

$$\begin{bmatrix} d_{2,11} & d_{2,12} & d_{2,13} & d_{2,14} & d_{2,15} \\ d_{2,21} & d_{2,22} & d_{2,23} & d_{2,24} & d_{2,25} \\ d_{2,31} & d_{2,32} & d_{2,33} & d_{2,34} & d_{2,35} \\ d_{2,41} & d_{2,42} & d_{2,43} & d_{2,44} & d_{2,45} \\ d_{2,51} & d_{2,52} & d_{2,53} & d_{2,54} & d_{2,55} \end{bmatrix} \begin{bmatrix} y_{t-2} \\ p_{t-2} \\ R_{t-2} \\ m_{t-2} \\ y_{t-2}^* \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ u_{4t} \\ u_{5t} \end{bmatrix}.$$

Note that this model is much more general than the Cowles Commission specification as far as the dynamics of all variables is concerned, moreover no a-priori restriction on the nature of the trend is imposed. The first step of the model validation procedure consists in the evaluation of the lag truncation: is the chosen length of the polynomial in the lag operator ($L = 2$, in our case) long enough to capture the dynamics in the data ? If the answer is yes, then the next step of the specification strategy can be taken, to identify the long-run structure and re-specify the reduced form as a VECM. As we have already pointed out this step can be skipped at the only risk of loss in efficiency. To keep the LSE specification more directly comparable with the Cowles Commission IS-LM-AS model, we do run this risk and keep the reduced form in levels. The last step of the LSE identification strategy is the specification of structural models. Just-identified model do not impose any further restriction on validated reduced form, while over-identified structure do impose testable restrictions on the reduced form. Testing such restrictions is the final model evaluation criteria. We have seen that in our specific example we have twenty five over-identifying restrictions. The validity of the over-identifying restrictions can be checked by comparing the reduced form implicit in the structural model (5.1) with the general reduced form (3.17) . The reduced form implicit in the structural model (5.1) is found by pre-multiplying the model by:

$$\begin{bmatrix} 1 & 0 & a_{13} & 0 & -1 \\ -a_{21} & 1 & 0 & 0 & a_{21} \\ -\frac{a_{31}}{a_{33}} & -\frac{1}{a_{33}} & 1 & \frac{1}{a_{33}} & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}^{-1} = \mathbf{A}^{-1}$$

:

$$= k \begin{bmatrix} a_{33} & -a_{13} & -a_{13}a_{33} & a_{13} & (a_{33}+a_{21}a_{13}) \\ a_{21}a_{33} & (a_{33}+a_{31}a_{13}) & -a_{21}a_{13}a_{33} & a_{21}a_{13} & -a_{31}a_{21}a_{13} \\ a_{21}+a_{31} & 1 & a_{33} & -1 & a_{31} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$k = \frac{1}{a_{33}+a_{21}a_{13}+a_{31}a_{13}}$$

$$\begin{bmatrix} y_t \\ p_t \\ R_t \\ m_t \\ y_t^* \end{bmatrix} = \mathbf{A}^{-1} \begin{bmatrix} c_{0,11} & 0 \\ c_{0,21} & 0 \\ c_{0,31} & 0 \\ c_{0,41} & 0 \\ c_{0,51} & c_{0,52} \end{bmatrix} \begin{bmatrix} 1 \\ t \end{bmatrix} + \tag{3.18}$$

$$+\mathbf{A}^{-1} \begin{bmatrix} 0 & a_{13} & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ p_{t-1} \\ R_{t-1} \\ m_{t-1} \\ y_{t-1}^* \end{bmatrix} +$$

$$\mathbf{A}^{-1} \begin{bmatrix} -a_{13} & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{t-2} \\ p_{t-2} \\ R_{t-2} \\ m_{t-2} \\ y_{t-2}^* \end{bmatrix} + \mathbf{A}^{-1} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \\ \epsilon_{4t} \\ \epsilon_{5t} \end{bmatrix}$$

note that (3.18) imposes more than twenty-five restrictions on (3.17), in fact it imposes twenty-five over-identifying restrictions on (3.17) in addition to those necessary to determine the chosen specification for m_t and y_t^*. The LSE methodology finds the roots of the failure of Cowles Commission models in the choice of specification rather than general baseline specifications.

3.5 Identification in the VAR methodology

We have seen that the LSE methodology has intepreted the failure of the traditional Cowles Commission approach as the result of a specification strategy leading to mis-specified and ill-identified model. The LSE methodology however does not question the potential of macroeconometric modelling for simulation and econometric policy evaluation. The VAR approach share with the LSE approach the diagnosis of the problem of Cowles Commission models but also questions the potential of macroeconometric modelling for policy simulation and econometric policy evaluation. VAR models differ from structural LSE models as to the purpose of their specification and estimation. In the traditional approach the typical question asked within a macroeconometric framework is "What is the optimal response by the monetary authority to movement in macroeconomic variables in order to achieve given targets for the same variables?". After the Lucas' critique questions like "How should a Central Bank respond to shocks in macroeconomic variables?" are to be answered within the framework of quantitative monetary general equilibrium models of the business cycle. So the answer has to be based on a theoretical model rather than on an empirical ad-hoc macroeconometric model. Within this framework there is a new role for empirical analysis, i.e. to provide the evidence on the stylized facts to be included in the theoretical model adopted for policy analysis and to decide between competing general equilibrium monetary models. The operationalization of this research programme is very well described in a recent paper by Christiano, Eichenbaum and Evans ([14]). Three are the relevant steps:

- monetary policy shocks are identified in actual economies;
- the response of relevant economic variables to monetary shocks is then described;
- the same experiment is then performed in the model economies to compare actual and model-based responses as an evaluation tool and a selection criterion for theoretical models.

To pin down more precisely the symmetries and differences between LSE-type structural models and VAR models consider again the case of the monetary transmission mechanism (MTM). The two type of models have a common structure which we have represented as follows:

$$\mathbf{A}\begin{pmatrix}\mathbf{Y}_t\\ \mathbf{M}_t\end{pmatrix} = \mathbf{C}(L)\begin{pmatrix}\mathbf{Y}_{t-1}\\ \mathbf{M}_{t-1}\end{pmatrix} + \mathbf{B}\begin{pmatrix}\boldsymbol{\nu}_t^Y\\ \boldsymbol{\nu}_t^M\end{pmatrix} \tag{3.19}$$

The main difference between the two approaches lies in the aim for which models are estimated.

Traditional Cowles Commission structural models are designed to identify the impact of policy variables on macroeconomic quantities in order to determine the value to be assigned to the monetary instruments ($\mathbf{M}$) to achieve a given target for the macroeconomic variables ($\mathbf{Y}$), assuming exogeneity of the policy variables in $\mathbf{M}$ on the ground that these are the instruments controlled by the policymaker. Identification in traditional structural models is obtained without assuming the orthogonality of structural disturbances: remember that we have labelled as structural disturbance in traditional models and LSE models the vector $\boldsymbol{\epsilon}$, where

$$\begin{pmatrix}\boldsymbol{\epsilon}_t^Y\\ \boldsymbol{\epsilon}_t^M\end{pmatrix} = \mathbf{B}\begin{pmatrix}\boldsymbol{\nu}_t^Y\\ \boldsymbol{\nu}_t^M\end{pmatrix}$$

As we shall see the impact of monetary policy is described by dynamic multipliers, which descibe the response of macroeconomic variables to a modification in the exogenous monetary instruments controlled by the policy maker. Dynamic multipliers are traditionally computed without separating changes in the monetary variable into the expected and unexpected components.

The assumed exogeneity of the monetary variables makes the model invalid for policy analysis if monetary policy reacts endogenously to the macroeconomic variables. LSE methodology would recognise the problem of the invalid exogeneity assumption for $\mathbf{M}$, it would then proceed to the identification of an alternative enlarged model (presumably such identification will be obtained through the imposition on a-priori restrictions on the dynamics of the lagged variables). However, the new model would be still used for simulation and econometric policy evaluation, whenever the appropriate concept of exogeneity (respectively, as we shall see, strong-exogeneity and super-exogeneity) where satisfied by the adopted specification.

VAR modelling would reject the Cowles Commision identifying restrictions as "incredible" for reason not very different from the ones pinned down by the LSE approach, however VAR models of the transmission mechanism are not estimated to yield advice on the best monetary policy; they are rather estimated to provide empirical evidence on the response of macroeconomic variables to monetary policy impulses in order to discriminate between alternative theoretical models of the economy. Monetary policy actions should be identified using restrictions independent from the competing models of the transmission mechanism under empirical investigation, taking into account the potential endogeneity of policy instruments.

In a series of papers, Christiano, Eichenbaum and Evans ([12], [13]) apply the VAR approach to derive "stylized facts" on the effect of a contractionary policy shock, and conclude that plausible models of the monetary transmission mechanism should be consistent at least with the following evidence on price, output and interest rates:

(i) the aggregate price level initially responds very little;

(ii) interest rates initially rise;

(iii) aggregate output initially falls, with a j-shaped response, with a zero long-run effect of the monetary impulse.

Such evidence leads to the dismissal of traditional real business cycle model, which are not compatible with the liquidity effect of monetary policy on interest rates, and of the Lucas ([35]) model of money, in which the effect of monetary policy on output depends on "price misperceptions". The evidence seems to be more in line with alternative intepretations of the monetary transmission mechanism based on sticky prices models (Goodfriend and King [11]), limited participation models (Christiano and Eichenbaum [11]) or models with indeterminacy-sunspot equilibria (Farmer [21]).

Having stated the objective of VAR models we are now in the position of assessing how the technical opportunities for identification, estimation and simulation are exploited to analyse the MTM. VAR models concentrate on shocks.

First the relevant shocks are identified, the response of the system to shocks is described by analyzing impulse responses (the propagation mechanism of the shocks), forecasting error variance decomposition, and historical decomposition.

Following Amisano and Giannini ([1]), we represent the general structural model of interest as follows:

$$\mathbf{A}\begin{pmatrix}\mathbf{Y}_t\\ \mathbf{M}_t\end{pmatrix}=\mathbf{C}(L)\begin{pmatrix}\mathbf{Y}_{t-1}\\ \mathbf{M}_{t-1}\end{pmatrix}+\mathbf{B}\begin{pmatrix}\boldsymbol{\nu}_t^Y\\ \boldsymbol{\nu}_t^M\end{pmatrix} \tag{3.20}$$

where $\mathbf{Y}$ and $\mathbf{M}$ are vectors of macroeconomic (non-policy) variables (e.g. output and prices) and variables controlled by the monetary policymaker (e.g. interest rates and monetary aggregates containing information on monetary policy actions) respectively. Matrix $\mathbf{A}$ describes the contemporaneous relations among the variables and $\mathbf{C}(L)$ is a matrix finite-order lag polinomial. $\boldsymbol{\nu}\equiv\begin{pmatrix}\boldsymbol{\nu}^Y\\ \boldsymbol{\nu}^M\end{pmatrix}$ is a vector of structural disturbances to the non-policy and policy variables normally independently distributed with identity variance-covariance matrix; non-zero off-diagonal elements of $\mathbf{B}$ allow some shocks to affect directly more than one endogenous variable in the system.

The structural model (6.2) is not directly observable, however a VAR can be estimated as the reduced form of the underlying structural model :

$$\begin{pmatrix}\mathbf{Y}_t\\ \mathbf{M}_t\end{pmatrix}=\mathbf{A}^{-1}\mathbf{C}(L)\begin{pmatrix}\mathbf{Y}_{t-1}\\ \mathbf{M}_{t-1}\end{pmatrix}+\begin{pmatrix}\mathbf{u}_t^Y\\ \mathbf{u}_t^M\end{pmatrix} \tag{3.21}$$

where $\mathbf{u}$ denotes the VAR residual vector, normally independently distributed with full variance-covariance matrix Σ. The relation between the VAR residuals in $\mathbf{u}$ and the structural disturbances in $\boldsymbol{\nu}$ is therefore:

$$\mathbf{A}\begin{pmatrix}\mathbf{u}_t^Y\\ \mathbf{u}_t^M\end{pmatrix}=\mathbf{B}\begin{pmatrix}\boldsymbol{\nu}_t^Y\\ \boldsymbol{\nu}_t^M\end{pmatrix} \tag{3.22}$$

undoing the partitioning we have

$$\mathbf{u}_t=\mathbf{A}^{-1}\mathbf{B}\boldsymbol{\nu}_t$$

from which we can derive the relation between the variance-covariance matrix of $\mathbf{u}_t$ (observed) and the variance-covariance matrix of $\boldsymbol{\nu}_t$ (not observed) as follows:

$$E\left(\mathbf{u}_t\mathbf{u}_t'\right)=\mathbf{A}^{-1}\mathbf{B}E\left(\boldsymbol{v}_t\boldsymbol{v}_t'\right)\mathbf{B}'\mathbf{A}^{-1}.$$

Substituting population moments with sample moments we have:

$$\widehat{\Sigma}=\widehat{\mathbf{A}}^{-1}\widehat{\mathbf{B}}\mathbf{I}\widehat{\mathbf{B}}'\widehat{\mathbf{A}}^{-1}$$

$\widehat{\Sigma}$ contains $\frac{n(n+1)}{2}$ different elements, this is the maximum number of identifiable parameters in matrices $\mathbf{A}$ and $\mathbf{B}$, terefore identifying restrictions are imposed on these matrices. We shall analyze the different type of identifying restrictions in one of the next chapters, devoted to VAR models. Once shock have been identified, the dynamic properties of the system can be described by analyzing the response of all variables in the system to such shocks. Note that VAR models do not include explicitly expectations and might therefore be subject to the Lucas critique. The general defense of VAR modellers against the Lucas critiques relies upon the fact that the variables shocked are the shocks and therefore the estimated parameters are not modified for simulation purposes.

3.6 Identification in intertemporally optimized models

The natural outcome of the Lucas critique are intertemporally optimized models in which deep parameters, independent from a particular policy regime, are identified separately from expectational parameters, specific to policy regimes. The intertemporal optimisation approach to macroeconomics leads naturally to a framework for identification and estimation of the deep parameters of interest. In fact the first order condition for the solution of intertemporal optimization problems are orthogonality conditions which can be exploited for identification and estimation of the structural parameters of interest. To illustrate the point consider the simplest possible version of the inflation targeting problem, see Svensson ([21]). The central bank faces the following intertemporal optimisation problem:

$$Minimize \quad E_t i = 0 \sum^{\infty} \delta^i L_{t+i} \tag{3.23}$$

where:

$$L = \frac{1}{2}\left[(\pi_t - \pi^*)^2 + \lambda x_t^2\right] \tag{3.24}$$

where E_t denotes expectations conditional upon the information set available at time t, δ is the relevant discount factor, L is the loss function of the central bank, π_t is inflation at time t, π^* is the target level of inflation, x represents deviations of output from its natural level, λ is a parameter which determines the degree of flexibility in inflation targeting. When $\lambda = 0$ the central bank is defined as a strict inflation targeter. As the monetary instrument is the policy rate, i_t , the structure of the economy must be described to obtain an explicit form for the policy rule. We consider the following specification for aggregate supply and demand in a closed economy[8]:

$$x_{t+1} = \beta_x x_t - \beta_r \left(i_t - E_t \pi_{t+1} - \bar{r}\right) + u_{t+1}^d \tag{3.25}$$

[8] As we shall see in one of the next chapter these two function are the outcome of the solutions of intertemporal optimisation problems by agents in the private sector.

$$\pi_{t+1} = \pi_t + \alpha_x x_t + u^s_{t+1} \tag{3.26}$$

Note that macroeconomic variables do not react contemporaneously to the instrument of monetary policy, this is a first identifying restriction for the relevant parameters in the model. As shown in Svensson ([21]), the first order conditions for optimality may be written as follows:

$$\frac{dL}{di_t} = (E_t\pi_{t+2} - \pi^*) = -\frac{\lambda}{\delta\alpha_x k}E_t x_{t+1} \tag{3.27}$$

$$k = 1 + \frac{\delta\lambda k}{\lambda + \delta\alpha_x^2 k} \tag{3.28}$$

(7.59) are orthogonality conditions involving all the deep parameters describing the structure of preferences of the central banker, π^*, δ, λ and just one parameter coming from the structure of the economy, α_x. By using (7.58) in (7.57) we obtain:

$$E_t\pi_{t+2} = E_t\pi_{t+1} + \alpha_x[\beta_x x_t - \beta_r(i_t - E_t\pi_{t+1} - \overline{r})] \tag{3.29}$$

and by substituting (7.61) in (7.59) we derive an interest rate rule:

$$i_t = \overline{r} + \pi^* + \left(\frac{1+\alpha_x\beta_r}{\alpha_x\beta_r}\right)(E_t\pi_{t+1} - \pi^*) + \tag{3.30}$$
$$+\frac{\beta_x}{\beta_r}x_t + \frac{\lambda}{\delta\alpha_x k}\frac{1}{\alpha_x\beta_r}E_t x_{t+1}.$$

The parameters in the interest rate rule (7.62) are convolutions of the parameters describing central banks preferences (π^*, λ, δ) and of those describing the structure of the economy $\left(\alpha_x, \beta_r, \beta_x, \overline{r}\right)$. It is then impossible to assess from the estimation of the rule if the responses of central banks to output and inflation are consistent with the parameters describing the impact of the policy instrument on these variables. Note, for example, that the estimation of an interest rate rule relating the policy rate to the output gap and to the deviation of expected inflation from target does not help to distinguish a strict inflation targeter ($\lambda = 0$, in the terminology of Svensson), from a flexible inflation targeter $(\lambda > 0)$.

In fact, there is only one empirical implication of the rule which can be confronted with the data independently from the identification of the parameters of interest, namely whether the parameter describing the reaction of policy rates to a gap between expected and target inflation is larger than one. A monetary policy which accommodates changes in inflation, $\frac{\partial i_t}{\partial E_t\pi_{t+1}} \leq 1$,will not in general converge to the target rate π^*. This empirical prediction is the one which has attracted most of the discussion on estimated monetary policy rules (see Clarida,Gali and Gertler, [4], [7], [8]).

By comparing the first order conditions for optimality, known as Euler equations, with the explicit interest rate rule we note that the deep parameters of interest are much more easily identifiable from (7.59) . In fact, while in our specific example (7.59) depend mainly on deep parameters describing taste and technology, there are macroecnomic applications in which the Euler equations depend only on these parameters. The identification and estimation strategy naturally consistent with the intertemporal optimization approach, is then to derive first the Euler equation and use them to pin down the deep parameters of interest. This step can be achieved by applying an estimation method directly based on orthogonality conditions, the Generalised Method of Moments. Numerical values to the remaining parameters in the model are then attributed, not necessarily by estimation. Then models are simulated and evaluated by comparing actual data with simulated data.

In the next chapters of the book we shall consider more deeply all the different approaches to macroeconometric modelling by considering a common macroeconomic issue: the analysis of the monetary transmission mechanism.

REFERENCES

Amisano G., Giannini C.(1997) Topics in Structural VAR Econometrics, second edition, Springer Verlag, New York

Christiano L.J. and M. Eichenbaum (1992) “Liquidity Effects and the Monetary Transmission Mechanism”, *American Economic Review*, 82, 2, 346-353

Christiano L.J., M. Eichenbaum and C.L. Evans (1996a) “The Effects of Monetary Policy Shocks: Evidence from the Flow of Funds”, *Review of Economics and Statistics*, 78, 16-34

Christiano L.J., M. Eichenbaum and C.L. Evans (1996b) “Monetary Policy Shocks and their Consequences: Theory and Evidence”, paper presented at ISOM 1996

Christano L.J., M.Eichenbaum and C.Evans (1998) ”Monetary Policy shocks: what have we learned and to what end?” NBER working paper 6400

Clarida R., J. Gali and M. Gertler (1997) ”The science of monetary policy”, mimeo, New York University

Clarida R., J. Gali and M. Gertler (1998) ”Monetary policy rules in practice:some international evidence” *European Economic Review*, 42,

Clarida R., J.Gali and M.Gertler (1998)”Monetary policy rules and macroeconomic stability: evidence and some theory”, NBER working paper No. 6442

Farmer R.E.A. (1997) “Money in a Real Business Cycle Model”, mimeo, Dept. of Economics, UCLA

Giannini C. (1992) Topics in Structural VAR Econometrics, Lecture Notes in Economics and Mathematical Systems, Springer-Verlag

Goodfriend M . and R.King (1997) ”The new neoclassical synthesis and the role of monetary policy”, N*BER Macroeconomics Annual 1997*, MIT Press

Hamilton J. (1994) Time-Series Analysis, Princeton, Princeton University Press,

Hendry D.F. (1995) Dynamic Econometrics Oxford University Press

Lucas R.E. Jr. (1972) “Expectations and the Neutrality of Money”, *Journal of Economic Theory*, 4, April, 103-124

Lucas R.E. Jr. (1976) “Econometric Policy Evaluation: A Critique”, in K. Brunner and A. Meltzer (eds.) *The Phillips curve and labor markets,* Amsterdam, North-Holland

Johansen S., Juselius K. (1994), Identification of the Long-run and the Short-Run Structure. An application to the IS-LM model. Journal of Econometrics,63, 7-36

Pesaran M.H., Smith R.(1994) Theory and Evidence in Economics, Journal of Econometrics

Sargan D. (1964) Wages and Prices in the United Kingdom: A Study in Econometric Methodology, in P.E. Hart, G.Mills and J.K. Whitaker (eds.), Econo-

metric Analysis of National Economic Planning, London, Butterworth, 25-63
Sims C.A. (1980) “Macroeconomics and Reality”, *Econometrica*, 48, 1-48
Spanos A.(1986?)”Identification.... A reconsideration” Journal of Econometrics
Svensson L.E.O. (1997) ”Inflation Forecast Targeting: Implementing and Monitoring Inflation Targets”, *European Economic Review*, 41, 1111-1146

4

THE COWLES COMMISSION'S APPROACH

4.1 Introduction

The traditional, usually referred to the Cowles Commission, approach to econometric modelling of the monetary transmission mechanism is aimed at the quantitative evaluation of the effects of modification in the variables controlled by the monetary policy maker (the instruments of monetary policy) on the macroeconomic variables which represents the final goals of the policy maker. We can identify three stages in the traditional approach:

- *specification and identification* of the theoretical model;
- *estimation* of the relevant parameters, and assessment of the dynamic properties of the model, with particular emphasys for the long-run properties;
- *simulation* of the effects of monetary policies.

We have already discussed the Cowles Commission approach to specification and identification in the previous chapters.

Before illustrating the approach at work we shall devote sections to the discussion of estimation, simulation and policy evaluation.

4.2 Estimation in the Cowles Commission Approach

The crucial features in the identification-specifcation stage, which are well shown by our IS-LM-AD-AS example, is that the specified empirical model is usually loosely related to theoretical models and that identification is achieved by imposing many a-priori restrictions delivering exogeneity status to a number of variables. As a consequence, identification is easily achieved within Cowles Commission models, usually with a large number of over-identifying restrictions. We have also seen that criticisms of this approach attributes the roots of its failure in the imposition of too many restrictions and in their incapability of recovering the structural deep parameters of economic interest, describing preference of the agents and the satus of technology.

However it is interesting to note that traditional modelling was in a sense aware of the presence of some mis-specification in the estimated equations. Such presence of mis-specification resulted in departure from the conditions which warrant that OLS estimators are BLUE. The solution proposed was not re-specification but rather modification of the estimation techniques. This is well reflected in the structure of the traditional textbooks, see for example Goldberger ([5]), Johnston, where the OLS estimator is introduced first and then different estimators are considered as solutions to different pathologies in the model residuals. Pathologies are identified as departures from the assumptions which guarantee that OLS are BLUE. I think that it is by now very well established that correcting the estimator is a strategy clearly inferior to improving the specification, i.e. correcting the model. Nevertheless, we devote some space to the discussion of alternative estimators in that they could be a last resort, to be used when models could not be improved for lack of the necessary informations.

4.2.1 *Heteroscedasticity, autocorrelation and the GLS estimator*

Let us reconsider the single equation model of Chapter 1, to generalize it to the case in which the hypotheses of diagonality and constance of the conditional variances-covariance matrix of the residuals do not hold:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon} &\sim n.i.d.\left(\mathbf{0}, \boldsymbol{\sigma}^2\Omega\right) \end{aligned} \tag{4.1}$$

where the vector $\mathbf{y}$ contains T observations on the dependent variables, $\mathbf{X}$ contains $(T \times K)$ observations on the K explanatory variables exogenous for the estimation of the vector $(K \times 1)$ $\boldsymbol{\beta}$, and Ω is a $(T \times T)$ symmetric and positive definite matrix. When the OLS estimator it is applied to model (4.1) it delivers estimators which are consistent but not efficient, moreover the traditional formula for the variance-covariance matrix of the OLS estimators, $\boldsymbol{\sigma}^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}$, is wrong and it leads to incorrect inference. In fact, by using the standard algebra of Chapter 1 it can be shown that the correct formula for the variance-covariance matrix of the OLS estimator is:

$$\boldsymbol{\sigma}^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\Omega\mathbf{X} \left(\mathbf{X}'\mathbf{X}\right)^{-1}.$$

A general solution to this problem is found in general by remembering that the inverse of a symmetric definite positive matrix is also symmetric and definite positive and that for a given matrix Ω, symmetric and definite positive, it always exists a $(T \times T)$ non-singular matrix $\mathbf{K}$ such that $\mathbf{K}'\mathbf{K} = \Omega^{-1}$ and $\mathbf{K}\Omega\mathbf{K}' = \mathbf{I}_T$.

To find how the solution is implemented consider the regression model obtained by pre-multiplying both the right hand side and the left hand side of (4.1) by $\mathbf{K}$:

$$\begin{aligned} \mathbf{K}\mathbf{y} &= \mathbf{K}\mathbf{X}\boldsymbol{\beta} + \mathbf{K}\boldsymbol{\epsilon} \\ \mathbf{K}\boldsymbol{\epsilon} &\sim n.i.d.\left(\mathbf{0}, \boldsymbol{\sigma}^2\mathbf{I}_T\right). \end{aligned} \tag{4.2}$$

The OLS estimator of the parameters of transformed model (4.2) satisfies all the conditions for the applications of the Gauss-Markov theorem, therefore the estimator

$$\begin{aligned}\widehat{\beta}_{GLS} &= \left(\mathbf{X}'\mathbf{K}'\mathbf{K}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{K}'\mathbf{K}\mathbf{y} \\ &= \left(\mathbf{X}'\Omega^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\Omega^{-1}\mathbf{y}\end{aligned}$$

known as the Generalised Least Squares estimator, is BLUE. The variance of the GLS estimator, conditional upon $\mathbf{X}$, becomes

$$Var\left(\hat{\beta}_{GLS}|\,\mathbf{X}\right) = \sigma^2\left(\mathbf{X}'\Omega^{-1}\mathbf{X}\right)^{-1}$$

Note that, from the applicability of the Gauss-Markov theorem, it follows immediately that the variance of the GLS estimator is equal to the sum of the variance of any other linear estimator and a positive semi-definite matrix. Consider for example the variance of the OLS and of the GLS estimators. Using the fact that if $\mathbf{A}$ and $\mathbf{B}$ are positive definite and $\mathbf{A}-\mathbf{B}$ is positive semi-definite, then $\mathbf{B}^{-1}-\mathbf{A}^{-1}$ is also positive semi-definite, we have:

$$\begin{aligned} & \left(\mathbf{X}'\Omega^{-1}\mathbf{X}\right) - \left(\mathbf{X}'\mathbf{X}\right)\left(\mathbf{X}'\Omega\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{X}\right) \\ &= \mathbf{X}'\mathbf{K}'\mathbf{K}\mathbf{X} - \left(\mathbf{X}'\mathbf{X}\right)\left(\mathbf{X}'\mathbf{K}^{-1}\left(\mathbf{K}'\right)^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{X}\right) \\ &= \mathbf{X}'\mathbf{K}'\left(\mathbf{I} - \left(\mathbf{K}'\right)^{-1}\mathbf{X}\left(\mathbf{X}'\mathbf{K}^{-1}\left(\mathbf{K}'\right)^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{K}^{-1}\right)\mathbf{K}\mathbf{X} \\ &= \mathbf{X}'\mathbf{K}'\mathbf{M}_W'\mathbf{M}_W\mathbf{K}\mathbf{X} \\ \mathbf{W} &= \left(\mathbf{K}'\right)^{-1}\mathbf{X} \\ \mathbf{M}_W &= \left(\mathbf{I} - \mathbf{W}\left(\mathbf{W}'\mathbf{W}\right)^{-1}\mathbf{W}'\right)\end{aligned}$$

The applicability of the GLS estimator requires an empirical specification for the matrix $\mathbf{K}$. We consider here two specific applications, where appropriate choice of the such matrix leads to fix problems in the OLS estimator generated respectively by the presence of first order serial correlation and of heteroscedasticity in the residuals.

Consider first the case of first order serial correlation in the residuals, we have the following model:

$$\begin{aligned} y_t &= \mathbf{x}_t'\boldsymbol{\beta} + u_t \\ u_t &= \rho u_{t-1} + \epsilon_t \\ \epsilon_t &\sim n.i.d.\left(0, \sigma_\epsilon^2\right)\end{aligned}$$

which, using our general notation, can be re-written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \epsilon \tag{4.3}$$
$$\epsilon \sim n.i.d.\left(\mathbf{0}, \boldsymbol{\sigma}^2\Omega\right)$$
$$\boldsymbol{\sigma}^2 = \frac{\sigma_\epsilon^2}{1-\rho^2} \tag{4.4}$$

$$\Omega = \begin{bmatrix} 1 & \rho & \rho^2 & . & . & \rho^{T-1} \\ \rho & 1 & \rho & . & . & \rho^{T-2} \\ \rho^2 & . & 1 & . & . & . \\ . & . & . & . & . & . \\ \rho^{T-2} & . & . & \rho & 1 & \rho \\ \rho^{T-1} & \rho^{T-2} & . & . & \rho & 1 \end{bmatrix}$$

In this case the knowledge of the parameter ρ allows the empirical implementation of the GLS estimator. An intuitive procedure to implement the GLS estimator, could then be the following:

- estimate the vector $\boldsymbol{\beta}$ by OLS and save the vector of residuals $\widehat{u}_t$
- regress $\widehat{u}_t$ on $\widehat{u}_{t-1}$ to obtain an estimate $\widehat{\rho}$ of ρ
- construct the transformed model and regress $(y_t - \widehat{\rho}y_{t-1})$ on $(\mathbf{x}_t - \widehat{\rho}\mathbf{x}_{t-1})$ to obtain the GLS estimator of the vector of parameters of interest.

Note that the above procedure, known as the Cochrance-Orcutt procedure, could be iterated until convergence.

In the case of heteroscedasticity our general model becomes

$$\mathbf{y} = \mathbf{X}\beta + \epsilon \tag{4.5}$$
$$\epsilon \sim n.i.d.\left(\mathbf{0}, \Omega\right) \tag{4.6}$$

$$\Omega = \begin{bmatrix} \sigma_1^2 & 0 & 0 & . & . & 0 \\ 0 & \sigma_2^2 & 0 & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & . & . & 0 & \sigma_{T-1}^2 & 0 \\ 0 & 0 & . & . & 0 & \sigma_T^2 \end{bmatrix}$$

In this case, in order to construct the GLS estimator, we need to model heteroscedascity choosing appropriately the $\mathbf{K}$ matrix. White ([16]) proposes a specification based on the consideration that in the case of heteroscedasticity the variance-covariance matrix of the OLS estimator takes the form:

$$\boldsymbol{\sigma}^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\Omega\mathbf{X} \left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

which could be used for inference, once an estimator for Ω is available. The following unbiased estimator of Ω is proposed

$$\hat{\Omega} = \begin{bmatrix} \widehat{u}_1^2 & 0 & 0 & . & . & 0 \\ 0 & \widehat{u}_2^2 & 0 & . & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & . & . & 0 & \widehat{u}_{T-1}^2 & 0 \\ 0 & 0 & . & . & 0 & \widehat{u}_T^2 \end{bmatrix}$$

Alternative models for heteroscedasticity, known as ARCH (Autoregressive Conditional Heteroscedasticity) processes, useful for high-frequency financial series, and based upon simultaneous modelling of the first two moments of time-series processes have been proposed by Engle([4]) and Bollerslev([1]) .

4.2.2 *Endogeneity*

The estimation of simultaneous system needs the solution of a problem, independent of mis-specification, which has prompted most of the advances in estimation theory within the Cowles Commission approach: simultaneity.

To discuss simultaneity we consider the following representation of a model of interest:

$$\begin{aligned} \mathbf{B}\mathbf{y}_t + \Gamma\mathbf{z}_t &= \mathbf{u}_t \\ \mathbf{u}_t &\sim n.i.d. \left(\mathbf{0}, \Sigma\right) \end{aligned} \tag{4.7}$$

where $\mathbf{y}_t$ is a $(G \times 1)$ vector of endogenous variables, $\mathbf{z}_t$ is a $(M \times 1)$ vector of exogenous variables, these variables are considered exogenous in that they are orthogonal to residuals. Therefore, in the case of a dynamic specification, it contains all contemporaneous variables considered orthogonal to residuals, their lags, and the lags of the variables $\mathbf{y}_t$. $\mathbf{B}$ and Γ are matrices of parameters, respectively $(G \times G)$ and $(G \times M)$. Using matrix notation we can represent (4.7) alternatively as follows:

$$\mathbf{B}\mathbf{y}' + \Gamma\mathbf{z}' = \mathbf{u}' \tag{4.8}$$

where $\mathbf{y}$ is a $(T \times G)$ matrix, $\mathbf{z}$ is a $(T \times M)$ matrix and $\mathbf{u}$ is a $(T \times G)$ matrix.

To illustrate the problems with the OLS estimator generated by endogeneity consider the first equation of model, which we write as

$$\mathbf{y}_1 = \mathbf{x}_1 \boldsymbol{\delta}_1 + \mathbf{u}_1 \tag{4.9}$$

where $\mathbf{y}_1$ is a $(T \times 1)$ vector containing all the observations on the first endogenous variable in the model, $\mathbf{x}_1$ is a $(T \times (G_1 + M_1 - 1))$ containing all observations on the M_1 exogenous variables included in the first equation and on the $G_1 - 1$ contemporaneous endogenous variables included in the first equation. Given that the matrix $\mathbf{x}_1$ contains some endogenous variables, in general we have:

$$p \lim \frac{1}{T} \mathbf{x}_1' \mathbf{u}_1 \neq 0 \tag{4.10}$$

and the OLS estimator of the parameters of interest is not consistent. Condition (4.10) is immediately understood by referring to the reduced form of the system (4.7)

$$\begin{aligned} \mathbf{y}_t &= \mathbf{B}^{-1} \Gamma \mathbf{z}_t + \mathbf{B}^{-1} \mathbf{u}_t \\ \mathbf{u}_t &\sim n.i.d. \, (\mathbf{0}, \Sigma) \end{aligned} \tag{4.11}$$

which shows that, with the exception of special configurations for the matrices $\mathbf{B}$ and Σ, all endogenous variables are correlated with all residuals in $\mathbf{u}_1$.

4.2.3 *GIVE estimators*

The Generalized Instrumental Variables ($GIVE$) estimator is derived by considering that, in the simultaneous model, condition (4.10) holds but we have also

$$p \lim \frac{1}{T} \mathbf{z}' \mathbf{u}_1 = 0 \tag{4.12}$$

therefore a consistent estimator of the parameters of interest can be derived by solving the following system of equations:

$$\begin{aligned} \mathbf{z}' \widehat{u}_1 &= 0 \\ \mathbf{z}' \left(\mathbf{y}_1 - \mathbf{x}_1 \widehat{\boldsymbol{\delta}}_1 \right) &= 0 \end{aligned} \tag{4.13}$$

System (4.13) contains a number of equations equal to the number of variables in $\mathbf{z}_1, M$, the number of unknowns is equal to the number of parameters in the vector $\boldsymbol{\delta}_1$, $K_1 = G_1 + M_1 - 1$. We have then three cases of interest:

(i) $M < K_1$: the number of unknowns is larger than the number of equations and no estimators of the parameters of interest can be derived. Not surprisingly, in fact in this case the equation is not identified.

(ii) $M = K_1$: the number of uknowns is exactly equal to the number of equations, the system is just identified and the solution to (4.13) has a unique solution and delivers an estimator of the parameters of interest:

$$\widehat{\boldsymbol{\delta}}_1 = \left(\mathbf{z}'\mathbf{x}_1\right)^{-1}\mathbf{z}'\mathbf{y}_1$$

(iii) $M > K_1$: the number of equations is larger than the number of unknowns, the equation is over-identified and the estimator of parameters of interest is not unequivocally determined by the orthogonality condition (4.13)

An intuitive solution for the over-identification case is obtained by taking K linear combinations of the M_1 orthogonality conditions. Define a matrix $\mathbf{L}$ of dimensions $(K_1 \times M)$. Pre-multiplying the system (4.13) by $\mathbf{L}$, we have:

$$\begin{aligned} \mathbf{L}\mathbf{z}'\widehat{\mathbf{u}}_1 &= 0 \\ \mathbf{L}\mathbf{z}'\left(\mathbf{y}_1 - \mathbf{x}_1\widehat{\boldsymbol{\delta}}_1\right) &= 0 \end{aligned} \tag{4.14}$$

from (7.27) we derive the following estimator:

$$\widehat{\boldsymbol{\delta}}_1 = \left(\mathbf{L}\mathbf{z}'\mathbf{x}_1\right)^{-1}\mathbf{L}\mathbf{z}'\mathbf{y}_1 \tag{4.15}$$

From (4.15) it follows that

$$\widehat{\boldsymbol{\delta}}_1 - \boldsymbol{\delta}_1 = \left(\mathbf{L}\mathbf{z}'\mathbf{x}_1\right)^{-1}\mathbf{L}\mathbf{z}'\mathbf{u}_1 \tag{4.16}$$

and

$$\sqrt{T}\left(\widehat{\boldsymbol{\delta}}_1 - \boldsymbol{\delta}_1\right) = \left(\frac{1}{T}\mathbf{L}\mathbf{z}'\mathbf{x}_1\right)^{-1}\mathbf{L}\sqrt{T}\mathbf{z}_1'\mathbf{u}_1 \tag{4.17}$$

Given that consistency of the estimator is guaranteed by the hypothesis (4.12), assuming that

$$p\lim\frac{1}{T}\mathbf{L}\mathbf{z}'\mathbf{x}_1 = \mathbf{L}\mathbf{M}_{zx_1},\ \sqrt{T}\mathbf{z}'\mathbf{u}_1 \xrightarrow{L} N\left(0, \sigma_{11}\mathbf{M}_{zz}\right)$$

we can apply the Cramer's theorem to conclude that:

$$\sqrt{T}\left(\widehat{\boldsymbol{\delta}}_1 - \boldsymbol{\delta}_1\right) \sim N\left(0, \sigma_{11}\left(\mathbf{L}\mathbf{M}_{zx_1}\right)^{-1}\mathbf{L}\mathbf{M}_{zz}\mathbf{L}'\left(\mathbf{M}_{x_1z}\mathbf{L}'\right)^{-1}\right) \tag{4.18}$$

(4.18) characterizes completely the properties of the estimator, however empirical implementation requires the knowledge of the matrix $\mathbf{L}$. Note that the variance of the estimator depends on $\mathbf{L}$, hence a natural criterion for choosing this matrix is the maximization of the efficiency of the estimator. Sargan([12]) shows that the variance of the estimator is minimized when $\mathbf{L} = \mathbf{M}_{x_1 z}\mathbf{M}_{zz}^{-1}$, in which case we have

$$\sqrt{T}\left(\widehat{\boldsymbol{\delta}}_1 - \boldsymbol{\delta}_1\right) \sim N\left(0, \sigma_{11}\left(\mathbf{M}_{x_1 z}\mathbf{M}_{zz}^{-1}\mathbf{M}_{z x_1}\right)^{-1}\right) \tag{4.19}$$

The choice of $\mathbf{L}$ defines the following estimator:

$$\widehat{\boldsymbol{\delta}}_1 = \left(\mathbf{x}_1'\mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}'\mathbf{x}_1\right)^{-1}\mathbf{x}_1'\mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}'\mathbf{y}_1 \tag{4.20}$$

whose variance-covariance matrix can be estimated as follows:

$$s_1^2\left(\mathbf{x}_1'\mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}'\mathbf{x}_1\right)^{-1}$$
$$s_1^2 = \frac{1}{T}\left(\mathbf{y}_1 - \mathbf{x}_1\widehat{\boldsymbol{\delta}}_1\right)'\left(\mathbf{y}_1 - \mathbf{x}_1\widehat{\boldsymbol{\delta}}_1\right)$$

(7.27) defines the Genaralized Instrumental Variables Estimator $(GIVE)$. It is easy to see that in the case of exact identification GIVE simplifies to $\left(\mathbf{z}'\mathbf{x}_1\right)^{-1}\mathbf{z}'\mathbf{y}_1$.

An equivalent estimator to GIVE is derived by the following two-step procedure:

- regress by OLS $\mathbf{x}_1$ on $\mathbf{z}$ and construct fitted values $\widehat{\mathbf{x}}_1 = \mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}'\mathbf{x}_1 = \mathbf{z}\widehat{\mathbf{Q}}_1$
- regress by OLS $\mathbf{y}_1$ on $\widehat{\mathbf{x}}_1$ to obtain

$$\begin{aligned}\widehat{\boldsymbol{\delta}}_1 &= \left(\widehat{\mathbf{x}}_1'\widehat{\mathbf{x}}_1\right)^{-1}\widehat{\mathbf{x}}_1'\mathbf{y}_1 \\ &= \left(\mathbf{x}_1'\mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}'\mathbf{x}_1\right)^{-1}\mathbf{x}_1'\mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}'\mathbf{y}_1\end{aligned}$$

Which is known as the Two-Stage Least Squares (TSLS) estimator. Note that, in order to obtain a TSLS estimator as efficient as the GIVE estimator, it is important to avoid generating an estimate of the variance of the estimator using the residuals from the second stage. In fact we have:

$$\begin{aligned}\widehat{\mathbf{u}}_{1,TSLS} &= \mathbf{y}_1 - \widehat{\mathbf{x}}_1\widehat{\boldsymbol{\delta}}_1 \\ &= \mathbf{y}_1 - \mathbf{x}_1\widehat{\boldsymbol{\delta}}_1 - \left(\widehat{\mathbf{x}}_1 - \mathbf{x}_1\right)\widehat{\boldsymbol{\delta}}_1 \\ &= \widehat{\mathbf{u}}_{1,GIVE} - \mathbf{x}_1\widehat{\boldsymbol{\delta}}_1 - \left(\widehat{\mathbf{x}}_1 - \mathbf{x}_1\right)\widehat{\boldsymbol{\delta}}_1\end{aligned}$$

which would result in an upward biased estimator of the variance. The problem is easily solved, and in virtually all econometric packages available there is no difference between the TSLS and the GIVE estimators.

More interesting problems are generated by mis-specification. In general we have mis-specification when the instruments are correlated with residuals. The classical form of mis-specification is omitted variables. Consider the following case:

The Data Generating Process can be represented as follows:

$$\mathbf{y}_1 = \mathbf{x}_1 \boldsymbol{\delta}_1 + \mathbf{x}_1^* \boldsymbol{\delta}_1^* + \mathbf{u}_1$$

The following model is estimated:

$$\mathbf{y}_1 = \mathbf{x}_1 \boldsymbol{\delta}_1 + \mathbf{v}_1$$

The GIVE-TSLS estimator of $\boldsymbol{\delta}_1$ is:

$$\begin{aligned}
\widehat{\boldsymbol{\delta}}_1 &= \left(\mathbf{x}_1' \mathbf{z} \left(\mathbf{z}'\mathbf{z}\right)^{-1} \mathbf{z}'\mathbf{x}_1\right)^{-1} \mathbf{x}_1' \mathbf{z} \left(\mathbf{z}'\mathbf{z}\right)^{-1} \mathbf{z}'\mathbf{y}_1 \\
&= \boldsymbol{\delta}_1 + \left(\mathbf{x}_1' \mathbf{z} \left(\mathbf{z}'\mathbf{z}\right)^{-1} \mathbf{z}'\mathbf{x}_1\right)^{-1} \mathbf{x}_1' \mathbf{z} \left(\mathbf{z}'\mathbf{z}\right)^{-1} \mathbf{z}' \left(\mathbf{x}_1^* \boldsymbol{\delta}_1^* + \mathbf{u}_1\right)
\end{aligned}$$

which is not consistent whenever $\mathbf{x}_1^*$ is correlated with $\mathbf{z}$. In this case we have

$$p \lim \frac{1}{T} \mathbf{z}'\mathbf{v}_1 \neq 0$$

and the instruments in $\mathbf{z}$ cannot be considered as valid.

Sargan ([12]) derived a statistic to test the null hypothesis of validity of instruments by showing that the quantity:

$$\begin{aligned}
C &= \frac{\widehat{\mathbf{u}}_1' \mathbf{z} \left(\mathbf{z}'\mathbf{z}\right)^{-1} \mathbf{z}' \widehat{\mathbf{u}}_1}{s_1^2} \\
s_1^2 &= \frac{1}{T} \left(\mathbf{y}_1 - \mathbf{x}_1 \widehat{\boldsymbol{\delta}}_1\right)' \left(\mathbf{y}_1 - \mathbf{x}_1 \widehat{\boldsymbol{\delta}}_1\right)
\end{aligned}$$

is distributed as a χ^2 with $M - K_1$ degrees of freedom under the null hypothesis of validity of instruments.

4.2.4 *Three-stage least squares (3SLS) and Seemingly Unrelated Regressions (SURE) estimators*

The estimators we have considered so far solve the problems generated by simultaneity without reverting to the specification of the full structural model. For this reason GIVE-TSLS estimators are known as limited information estimators. To analyze full-information estimators we need to introduce some new definitions.

Far any two matrices $\mathbf{A}\,(m \times n)$ and $\mathbf{B}\,(p \times q)$ define as the *Kronecker product* $\mathbf{A} \otimes \mathbf{B}$ the matrix $(mp \times nq)$ obtained by multiplying each element of $\mathbf{A}$ by $\mathbf{B}$. The following properties are related to the Kronecker product:

- $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$, whenever the matrices $\mathbf{AC}$ and $\mathbf{BD}$ are defined
- $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$
- $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$, whenever the matrices $\mathbf{A}^{-1}$and $\mathbf{B}^{-1}$ are defined

Define vec$(\mathbf{A})$, the *vectorization* of the $\mathbf{A}$, as the vector $(mn \times 1)$ obtained by stacking the m transposed rows of $\mathbf{A}$:

$$m \times n\mathbf{A} = \begin{pmatrix} a_{11} & . & . & a_{1n} \\ . & . & . & . \\ . & . & . & . \\ a_{n1} & . & . & a_{nn} \end{pmatrix}$$

$$vec(\mathbf{A}) = \begin{pmatrix} a_{11} \\ . \\ . \\ a_{1n} \\ a_{21} \\ . \\ . \\ a_{2n} \\ . \\ . \\ . \\ a_{nn} \end{pmatrix}$$

Vectorization and Kronecker product are linked by the following property:

$$vec(\mathbf{ABC}) = (\mathbf{A} \otimes \mathbf{C})' vec(\mathbf{B})$$

To discuss full-information estimation, consider that the i-th equations of our model can be represented as

$$\mathbf{y}_i = \mathbf{x}_i \boldsymbol{\delta}_i + \mathbf{u}_i$$

where $\mathbf{y}_i$ is a $(T \times 1)$ vector containing T observations on i-th endogenous variables, $\mathbf{x}_i$ is a $(T \times K_i)$ matrix, with $K_i = (G_i + M_i - 1)$ containing all observations on the $G_i - 1$ endogenous and on the M_i exogenous variables included in the i-th equation. We can give the following compact representation of the model:

$$\mathbf{y}^{+} = \mathbf{x}^{+}\boldsymbol{\delta}^{+} + \mathbf{u}^{+} \tag{4.21}$$

where $\mathbf{y}^{+}$is a $(GT \times 1)$ vector, $\mathbf{x}^{+}$ is a $\left(GT \times i = 1\sum^{G} K_i\right)$ matrix, $\boldsymbol{\delta}^{+}$ is a $\left(i = 1\sum^{G} K_i \times 1\right)$, $\mathbf{u}^{+}$ is a $(GT \times 1)$ vector:

$$\mathbf{y}^{+} = \begin{pmatrix} \mathbf{y}_1 \\ . \\ . \\ \mathbf{y}_G \end{pmatrix}, \mathbf{x}^{+} = \begin{pmatrix} \mathbf{x}_1 & 0 & 0. & 0 \\ 0 & \mathbf{x}_2 & . & . \\ 0 & . & . & . \\ 0 & . & . & \mathbf{x}_G \end{pmatrix}$$

$$\boldsymbol{\delta}^{+} = \begin{pmatrix} \boldsymbol{\delta}_1 \\ . \\ . \\ \boldsymbol{\delta}_G \end{pmatrix}, \mathbf{u}^{+} = \begin{pmatrix} \mathbf{u}_1 \\ . \\ . \\ \mathbf{u}_G \end{pmatrix}$$

The following properties hold for $\mathbf{u}^{+}$

$$E\left(\mathbf{u}^{+}\right) = 0$$

$$E\left(\mathbf{u}^{+}\mathbf{u}^{+\prime}\right) = \begin{pmatrix} E\left(\mathbf{u}_1\mathbf{u}_1'\right) & E\left(\mathbf{u}_1\mathbf{u}_2'\right) & . & . & E\left(\mathbf{u}_1\mathbf{u}_G'\right) \\ E\left(\mathbf{u}_2\mathbf{u}_1'\right) & E\left(\mathbf{u}_2\mathbf{u}_2'\right) & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ E\left(\mathbf{u}_G\mathbf{u}_1'\right) & . & . & . & E\left(\mathbf{u}_G\mathbf{u}_G'\right) \end{pmatrix}$$

where each block of the above matrix is $(T \times T)$.

Assuming that all residuals are contemporaneously correlated but not serially correlated, with non singular variance-covariance matrix Σ,[9] we have:

$$E\left(\mathbf{u}^{+}\mathbf{u}^{+\prime}\right) = \Sigma \otimes I_T$$

$$= \begin{pmatrix} \sigma_{11}I_T & \sigma_{12}I_T & . & . & \sigma_{1G}I_T \\ \sigma_{21}I_T & \sigma_{22}I_T & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ \sigma_{G1}I_T & . & . & . & \sigma_{GG}I_T \end{pmatrix}$$

The problem to be solved is the estimation of parameters in (4.21) taking into account simultaneity and the strucutre of correlations in Σ. These problems are solved in turn by the Three-Stage Least Squares (3SLS) estimator.

[9]the last assumption requires that all identities are excluded from the model

4.2.4.1 *First stage: the diagonalization of* Σ Consider the following decomposition for Σ^{-1}

$$\Sigma^{-1} = \mathbf{H}\mathbf{H}' \tag{4.22}$$

which always exists. From(4.22) we have

$$\mathbf{H}\Sigma\mathbf{H}' = I_G$$

By pre-multiplying (4.21) by $\mathbf{H}' \otimes I_T$, we obtain

$$(\mathbf{H}' \otimes I_T)\,\mathbf{y}^+ = (\mathbf{H}' \otimes I_T)\,\mathbf{x}^+\boldsymbol{\delta}^+ + (\mathbf{H}' \otimes I_T)\,\mathbf{u}^+ \tag{4.23}$$

where residuals of (4.23) feature a diagonal variance-covariance matrix:

$$\begin{aligned} E\left((\mathbf{H}' \otimes I_T)\,\mathbf{u}^+\mathbf{u}^{+\prime}\,(\mathbf{H}' \otimes I_T)'\right) &= (\mathbf{H}' \otimes I_T)\,(\Sigma \otimes I_T)\,(\mathbf{H}' \otimes I_T)' \\ &= (\mathbf{H}'\Sigma \otimes I_T)\,(\mathbf{H} \otimes I_T) \\ &= I_G \otimes I_T = I_{GT} \end{aligned}$$

The completion of the first stage has left us with the following transformed model:

$$(\mathbf{H}' \otimes I_T)\,\mathbf{y}^+ = (\mathbf{H}' \otimes I_T)\,\mathbf{x}^+\boldsymbol{\delta}^+ + (\mathbf{H}' \otimes I_T)\,\mathbf{u}^+ \tag{4.24}$$

$$\mathbf{y}^* = \mathbf{x}^*\boldsymbol{\delta}^* + \mathbf{u}^* \tag{4.25}$$

$$E\,(\mathbf{u}^*) = 0,\; E\,(\mathbf{u}^*\mathbf{u}^{*\prime}) = I_{GT} \tag{4.26}$$

in which, the variance-covariance matrix is diagonal, but we still have simultaneity, in fact

$$p\lim \frac{1}{T}\mathbf{x}^{*\prime}\mathbf{u}^* \neq 0$$

4.2.4.2 *The second stage: choice of instruments* To select instruments, remember that the reduced form of our original system can be represented as follows:

$$\mathbf{y}' = \mathbf{B}^{-1}\Gamma\mathbf{z}' + \mathbf{B}^{-1}\mathbf{u}' \tag{4.27}$$

Vectorisation of (4.27) delivers:

$$vec\left(\mathbf{y}'\right) = vec\left(\mathbf{B}^{-1}\Gamma\mathbf{z}'\right) + vec\left(\mathbf{B}^{-1}\mathbf{u}'\right) \tag{4.28}$$

from which it follows that

$$\mathbf{y}^{+} = vec\left(I_G\mathbf{B}^{-1}\Gamma\mathbf{z}'\right) + vec\left(\mathbf{B}^{-1}\mathbf{u}'\right) \tag{4.29}$$

$$= \left(I_G \otimes \mathbf{z}\right) vec\left(\mathbf{B}^{-1}\Gamma\mathbf{z}'\right) + vec\left(\mathbf{B}^{-1}\mathbf{u}'\right) \tag{4.30}$$

Then, the natural choice of instruments is $(I_G \otimes \mathbf{z})$

4.2.4.3 *The third stage: applying the GIVE principle* By applying the GIVE principle to (4.25) ,choosing $\mathbf{z}^{*}= (I_G \otimes \mathbf{z})$ as instruments, we have

$$\widehat{\boldsymbol{\delta}}_1^{+} = \left(\mathbf{x}^{*\prime}\mathbf{z}^{*}\left(\mathbf{z}^{*\prime}\mathbf{z}^{*}\right)^{-1}\mathbf{z}^{*\prime}\mathbf{x}^{*}\right)^{-1}\mathbf{x}^{*\prime}\mathbf{z}^{*}\left(\mathbf{z}^{*\prime}\mathbf{z}^{*}\right)^{-1}\mathbf{z}^{*\prime}\mathbf{y}^{*}$$

At this stage, remembering that

$$\mathbf{x}^{*\prime}\mathbf{z}^{*} = \mathbf{x}^{+\prime}\left(\mathbf{H} \otimes I_T\right)\left(I_G \otimes \mathbf{z}\right)$$
$$\mathbf{z}^{*\prime}\mathbf{z}^{*} = \left(I_G \otimes \mathbf{z}'\right)\left(I_G \otimes \mathbf{z}\right) = I_G \otimes \mathbf{z}'\mathbf{z}$$
$$\mathbf{z}^{*\prime}\mathbf{y}^{*} = \left(I_G \otimes \mathbf{z}'\right)\left(\mathbf{H} \otimes I_T\right)\mathbf{y}^{+} = \left(\mathbf{H}' \otimes \mathbf{z}'\right)\mathbf{y}^{+}$$

we can show the following results:

$$\left(\mathbf{x}^{*\prime}\mathbf{z}^{*}\left(\mathbf{z}^{*\prime}\mathbf{z}^{*}\right)^{-1}\mathbf{z}^{*\prime}\mathbf{x}^{*}\right) = \mathbf{x}^{+\prime}\left(\mathbf{H} \otimes \mathbf{z}\right)\left(I_G \otimes \mathbf{z}'\mathbf{z}\right)^{-1}\left(\mathbf{H}' \otimes \mathbf{z}'\right)\mathbf{x}^{+}$$
$$= \mathbf{x}^{+\prime}\left(\Sigma^{-1} \otimes \mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}'\right)\mathbf{x}^{+}$$

$$\left(\mathbf{x}^{*\prime}\mathbf{z}^{*}\left(\mathbf{z}^{*\prime}\mathbf{z}^{*}\right)^{-1}\mathbf{z}^{*\prime}\mathbf{y}^{*}\right) = \mathbf{x}^{+\prime}\left(\mathbf{H} \otimes \mathbf{z}\right)\left(I_G \otimes \mathbf{z}'\mathbf{z}\right)^{-1}\left(\mathbf{H}' \otimes \mathbf{z}'\right)\mathbf{y}^{+}$$
$$= \mathbf{x}^{+\prime}\left(\Sigma^{-1} \otimes \mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}'\right)\mathbf{y}^{+}$$

and, finally, we have an expression for the 3SLS estimator

$$\widehat{\boldsymbol{\delta}}_1^{+} = \left(\mathbf{x}^{+\prime}\left(\Sigma^{-1} \otimes \mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}'\right)\mathbf{x}^{+}\right)^{-1}\mathbf{x}^{+\prime}\left(\Sigma^{-1} \otimes \mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}'\right)\mathbf{y}^{+}$$

The asymptotic distribution of the 3SLS can be written as follows:

$$\widehat{\boldsymbol{\delta}}_1^+ \sim N\left(\boldsymbol{\delta}^+, \left(\mathbf{x}^{+\prime}\left(\Sigma^{-1} \otimes \mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}'\right)\mathbf{x}^+\right)^{-1}\right)$$

To make the estimator operational an estimate of Σ is needed. This can be obtained by using the sample correlations of the residuals from 2SLS estimation.

To analyze the estimator more closely we can re-write in a more extensive format. In fact, we have

$$\mathbf{x}^{+\prime}\left(\Sigma^{-1} \otimes \mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}'\right)\mathbf{x}^+ =$$

$$\begin{pmatrix} \mathbf{x}_1' & 0 & 0 & 0 \\ 0 & \mathbf{x}_2' & . & . \\ 0 & . & . & . \\ 0 & . & . & \mathbf{x}_G' \end{pmatrix} \begin{pmatrix} \sigma_{11}\mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}' & .\ . & \sigma_{11}\mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}' \\ . & .\ . & . \\ . & .\ . & . \\ \sigma_{G1}\mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}' & .\ . & \sigma_{GG}\mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}' \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 & 0 & 0 & 0 \\ 0 & \mathbf{x}_2 & . & . \\ 0 & . & . & . \\ 0 & . & . & \mathbf{x}_G \end{pmatrix}$$

$$\mathbf{x}^{+\prime}\left(\Sigma^{-1} \otimes \mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}'\right)\mathbf{y}^+ =$$

$$\begin{pmatrix} j = 1\sum^{G}\sigma_{1j}\mathbf{x}_1'\mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}'\mathbf{y}_j \\ . \\ . \\ j = 1\sum^{G}\sigma_{Gj}\mathbf{x}_G'\mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}'\mathbf{y}_j \end{pmatrix}$$

where σ_{ij} represents the generic element i,j of the matrix Σ^{-1}. We are now in the position of considering some specific cases of the 3SLS estimator.

Note first that the 3SLS estimator coincides with the 2SLS when the matrix Σ is diagonal. In this case we have:

$$\begin{aligned} \hat{\boldsymbol{\delta}}^+ &= \left(\mathbf{x}^{+\prime}\left(\Sigma^{-1} \otimes \mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}'\right)\mathbf{x}^+\right)^{-1}\mathbf{x}^{+\prime}\left(\Sigma^{-1} \otimes \mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}'\right)\mathbf{y}^+ \\ &= \begin{pmatrix} \left(\mathbf{x}_1'\mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}'\mathbf{x}_1\right)^{-1}\mathbf{x}_1'\mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}'\mathbf{y}_1 \\ . \\ . \\ \left(\mathbf{x}_G'\mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}'\mathbf{x}_G\right)^{-1}\mathbf{x}_G'\mathbf{z}\left(\mathbf{z}'\mathbf{z}\right)^{-1}\mathbf{z}'\mathbf{y}_G \end{pmatrix} \end{aligned}$$

This equivalence result holds also when all the equations in the system are exactly identified.

Another interesting case arises when the matrix $\mathbf{B}$ in our structural model (4.7) is diagonal, when we have:

$$\mathbf{y}^+ = \mathbf{x}^+\boldsymbol{\delta}^+ + \mathbf{u}^+ \tag{4.31}$$
$$E\left(\mathbf{u}^+\right) = 0 \tag{4.32}$$
$$E\left(\mathbf{u}^+\mathbf{u}^{+\prime}\right) = \Sigma \otimes I_T \tag{4.33}$$
$$p\lim \frac{1}{T}\mathbf{x}^{+\prime}\mathbf{u}^+ = 0 \tag{4.34}$$

The particular structure of $\mathbf{B}$ implies that all the simultaneity in the system comes from the correlation of residuals, therefore after the implemetation of the first stage of the 3SLS, the diagonalization of the variance-covariance matrix, a consistent estimator is derived by applying OLS to the transformed model. The relevant estimator is then:

$$\begin{aligned}\hat{\boldsymbol{\delta}}^+ &= \left(\mathbf{x}^{*\prime}\mathbf{x}^*\right)^{-1}\mathbf{x}^{*\prime}\mathbf{y}^* \\ &\left(\mathbf{x}^{+\prime}\left(\Sigma^{-1}\otimes \mathbf{I}_T\right)\mathbf{x}^+\right)^{-1}\mathbf{x}^{+\prime}\left(\Sigma^{-1}\otimes \mathbf{I}_T\right)\mathbf{y}^+\end{aligned}$$

which is known as the Seemingly Unrelated Regression Equations ($SURE$) or Zellner's estimator.

A further interesting specific case of the SURE estimator is obtained when each equation of the system contains the same set of regressors:

$$\begin{aligned}\mathbf{x}^+ &= \begin{pmatrix} \mathbf{x}_1 & 0 & 0. & 0 \\ 0 & \mathbf{x}_2 & . & . \\ 0 & . & . & . \\ 0 & . & . & \mathbf{x}_G \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{x} & 0 & 0. & 0 \\ 0 & \mathbf{x} & . & . \\ 0 & . & . & . \\ 0 & . & . & \mathbf{x} \end{pmatrix} = I \otimes \mathbf{x}\end{aligned}$$

by substituting for $\mathbf{x}^+$ in the expression for the Zellner estimator, we obtain:

$$\begin{aligned}\widehat{\boldsymbol{\delta}}_1^+ &= \left((I\otimes \mathbf{x})'\left(\Sigma^{-1}\otimes \mathbf{I}_T\right)(I\otimes \mathbf{x})\right)^{-1}(I\otimes \mathbf{x})'\left(\Sigma^{-1}\otimes \mathbf{I}_T\right)\mathbf{y}^+ \\ &= \left(I\Sigma^{-1}I\otimes \mathbf{x}'I_T\mathbf{x}\right)^{-1}\left(I\Sigma^{-1}\otimes \mathbf{x}'\mathbf{I}_T\right)\mathbf{y}^+ \\ &= \left(\Sigma\otimes \left(\mathbf{x}'\mathbf{x}\right)^{-1}\right)\left(\Sigma^{-1}\otimes \mathbf{x}'\right)\mathbf{y}^+ \\ &= \left(I_G\otimes \left(\mathbf{x}'\mathbf{x}\right)^{-1}\mathbf{x}'\right)\mathbf{y}^+\end{aligned}$$

which gives a compact representation of the OLS estimators applied equation by equation.

4.2.5 *FIML estimator*

Lastly, we give a brief description of the most general full-information esimator: the Full Information Maximum Likelihood ($FIML$) estimator. Considering the reduced form of our model (4.11) and taking logarithms, we can write the joint distribution of $\mathbf{y}_1, ..., \mathbf{y}_T$, as follows:

$$\log L = \frac{-GT}{2} \log(2\pi) - \frac{T}{2} \log \left| \mathbf{B}^{-1} \Sigma \mathbf{B}'^{-1} \right| + \quad (4.35)$$
$$-\frac{1}{2} t = 1 \sum^{T} \left(\mathbf{y}_t - \mathbf{B}^{-1} \Gamma \mathbf{x}_t \right)' \mathbf{B}' \Sigma^{-1} \mathbf{B} \left(\mathbf{y}_t - \mathbf{B}^{-1} \Gamma \mathbf{x}_t \right)$$

Note now that

$$\left(\mathbf{y}_t - \mathbf{B}^{-1} \Gamma \mathbf{x}_t \right)' \mathbf{B}' = (\mathbf{B}\mathbf{y}_t - \Gamma \mathbf{x}_t)'$$

and that, from a standard results on determinants, it follows that:

$$-\frac{T}{2} \log \left| \mathbf{B}^{-1} \Sigma \mathbf{B}'^{-1} \right| = T \left| \log \mathbf{B} \right| - \frac{T}{2} \log |\Sigma|$$

So we can re-write our log-likelihood function as:

$$\log L = \frac{-GT}{2} \log(2\pi) + T \left| \log \mathbf{B} \right| - \frac{T}{2} \log |\Sigma| + \quad (4.36)$$
$$-\frac{1}{2} t = 1 \sum^{T} (\mathbf{B}\mathbf{y}_t - \Gamma \mathbf{x}_t)' \Sigma^{-1} (\mathbf{B}\mathbf{y}_t - \Gamma \mathbf{x}_t)$$

FIML estimator are derived by maximizing (4.36) with respect to $\mathbf{B}, \Gamma, \Sigma$. A number of technical issue arise as the problem is non-linear. For a good discussion of these problems, and solutions, see Hendry ([7]). The FIML estimator is the most general system estimator in that all other estimators can be derived as its special cases, for a detailed derivation see Hendry([6]) .

4.3 Simulation

Having identified the model and estimated the parameters of interest it is possible to proceed to the simulation. For given values of the parameters and of the exogenous variables, values for the endogenous variables are found by finding the dynamic solution of the model. To illustrate how this result is accomplished, consider the following general representation of a model including n endogenous variables $\mathbf{y}$=(y_1,y_2...y_n) and in k exogenous variables $\mathbf{x}$=(x_1,x_2...x_k) ,:

$$y_{1t} = f_1\left(y_{1t}, ...y_{nt}, A_1(L)\mathbf{y}_{t-1}, \mathbf{x}_t, B_1(L)\,\mathbf{x}_{t-1}\right)$$

$$y_{2t} = f_2\left(y_{1t}, ...y_{nt}, A_2(L)\mathbf{y}_{t-1}, \mathbf{x}_t, B_2(L)\,\mathbf{x}_{t-1}\right)$$

$$...$$

$$y_{kt} = f_k\left(y_{1t}, ...y_{nt}, A_k(L)\mathbf{y}_{t-1}, \mathbf{x}_t, B_k(L)\,\mathbf{x}_{t-1}\right)$$

$$...$$

$$y_{nt} = f_n\left(y_{1t}, ...y_{nt}, A_n(L)\mathbf{y}_{t-1}, \mathbf{x}_t, B_n(L)\,\mathbf{x}_{t-1}\right)$$

In the specifications discussed so far the functions f_i are linear, but more general specification could be accomodated within this framework.

Solving the model amounts to find a fixed-point such that:

$$\mathbf{y}_t = f\left(\mathbf{y}_t, \mathbf{A}(L)\,\mathbf{y}_{t-1}, \mathbf{x}_t, \mathbf{B}(L)\,\mathbf{x}_{t-1}\right).$$

A popular numerical method, implemented in many widely available packages, such as E-Views, is the Gauss-Seidel method. Guass-Seidel method finds the fixed-point by iteration using the updating rule:

$$\mathbf{y}_t^{i+1} = f\left(\mathbf{y}_t^{i}, \mathbf{A}(L)\,\mathbf{y}_{t-1}^{i+1}, \mathbf{x}_t, \mathbf{B}(L)\,\mathbf{x}_{t-1}\right)$$

Gauss-Seidel solves the equations in the order that they appear in the model. So if an endogenous variables that has already been solved for appears later in some other equation, Gauss-Seidel uses the value as solved in that iteration. To illustrate matters, the k-th variable in the i-th iteration is solved by:

$$y_{kt}^{i} = f_k\left(y_{1t}^{i}, ..., y_{k-1t}^{i}, y_{kt}^{i}, y_{k+1t}^{i-1}, ...y_{nt}^{i-1}, A_k(L)\mathbf{y}_{t-1}, \mathbf{x}_t, B_k(L)\,\mathbf{x}_{t-1}\right)$$

As a consequence the ordering of the variables matters and equations with relative few right-hand side endogenous variables should be listed early in the model. As the model is solved by setting disturbances to zero we have a deterministic solution, a stochastic solution can easily be generated solving the model by adding drawings from random variables and taking expected values afterwards.

4.4 Policy Evaluation

Dynamic simulation can be used to evaluate the effect of different policies, defined by specifying different patterns for the exogenous variables. Policy evaluation is implemented by examining how the predicted values of the endogenous variables change after an (some) exogenous variables is (are) modified. Policy evaluation implies simulating the model twice: first a baseline, *control*, simulation is run. Such simulation can be run within sample, in which case observed data are available for the exogenous variables, or outside the available sample, in which case values are assigned to the exogenous variables. In the case of out-of-sample simulation, which is equivalent to forecasting the endogenous variables for a given scenario for the exogenous variables, it is useful to assign values to the exogenous variables such that the baseline simulation path exhibits standard historical patterns for the endogenous variables. The results of such baseline simulation are then compared with those obtained from an alternative, *disturbed*, simulation, based on the modification of the relevant exogenous variables. Policy evaluation is usually based on *dynamic multipliers*.

Consider the case of the simulation of a model over a sample of size T, and index by t the generic observation in that sample. Denote by x_t^b the series of values attributed to the exogenous variable x in the baseline simulation, and by $x_t^d = x_t^b + \delta$ the series of alternative values attributed to the same variables in the disturbed simulation. Similarly, denote by y_{nt}^b the solved value for the endogenous variable y_n at time t in the baseline simulation and by y_{nt}^d the solved value for the endogenous variable y_n at time t in the disturbed simulation.

The dynamic multiplier is the defined as follows

$$DM_t = \frac{\left(y_{nt}^d - y_{nt}^b\right)}{\left(x_t^d - x_t^b\right)} = \frac{\left(y_{nt}^d - y_{nt}^b\right)}{\delta} \tag{4.37}$$

When model are stable, long-run multipliers, obtained for large t, converge to fixed numbers. Note that in linear systems long-run multipliers can be also obtained by giving a temporary (one period) impulse to the exogenous variable and by then computing the **cumulative** response of the endogenous variables.

To illustrate matters assume that the estimation over a given sample, say 1960:1-1998:1, of a simple dynamic model for consumption and income, has delivered the following results, similar to those obtained in the dynamic model of US consumption discussed in Chapter 2,.

$$\begin{aligned} \Delta c_t &= 0.25 * \Delta y_t - 0.15 * (c_{t-1} - y_{t-1}) \\ \Delta y_t &= 0.008 \end{aligned}$$

We aim at deriving the dynamic multiplier, describing the response of consumption to a one per cent increase in income by simulating the model over the period 1998:2-2020:4.

The following E-Views programme (run after having opened the file USUK.WF1) achieves the result:

```
SMPL 1998:1 1998:1
LCUS=LYUS
SMPL 1998:2 2020:4
model consinc
consinc.append LCUS =LCUS(-1)-0.15*LCUS(-1)+0.15*LYUS(-1) +0.250*(LYUS-LYUS(-1))
consinc.append LYUS =0.008 +LYUS(-1)
COPY CONSINC M_TEMP
M_TEMP.APPEND ASSIGN @ALL _BL
M_TEMP.SOLVE
delete M_TEMP
SMPL 1998:1 1998:1
LYUS=LCUS+0.01
SMPL 1998:2 2020:4
COPY CONSINC M_TEMP
M_TEMP.APPEND ASSIGN @ALL _DS
M_TEMP.SOLVE
delete M_TEMP
SMPL 1998:1 2020:4
genr DM=100*(LCUS_DS-LCUS_BL)
plot DM
SMPL 1998:1 1998:1
LYUS=LCUS-0.01
```

The programme begins by setting all variables at their long-run solution. Then the relevant model is constructed by defining it as CONSINC and by including the specification for the two equations. The Model CONSINC is then copied to a temporaray model, which is solved dynamically for the sample 1998:2 2020:4, and the suffix _BL, for baseline, is attributed to the variables generated by the solution.In the following step the disturbed solution is generated by adding a one per cent shock for one period (1998:1) to LYUS. Note that, as LYUS has a unit root, the one-period shock has permanent effect. The disturbed solution is then computed and the suffix _DS is attached to the generated variables. Lastly the dynamic multiplier is computed by applying formula (4.37) , we report it in the following Figure :

Having illustrated the basics with this simple case we move to discuss a more articulated model Cowles Commission model of the monetary tranmsission mechanism, by taking all steps from specification to simulation.

4.5 A model of the monetary transmisssion mechanism

4.5.1 *Specification of the theoretical model*

We consider the close-economy IS-LM specification with autoregressive expectations:

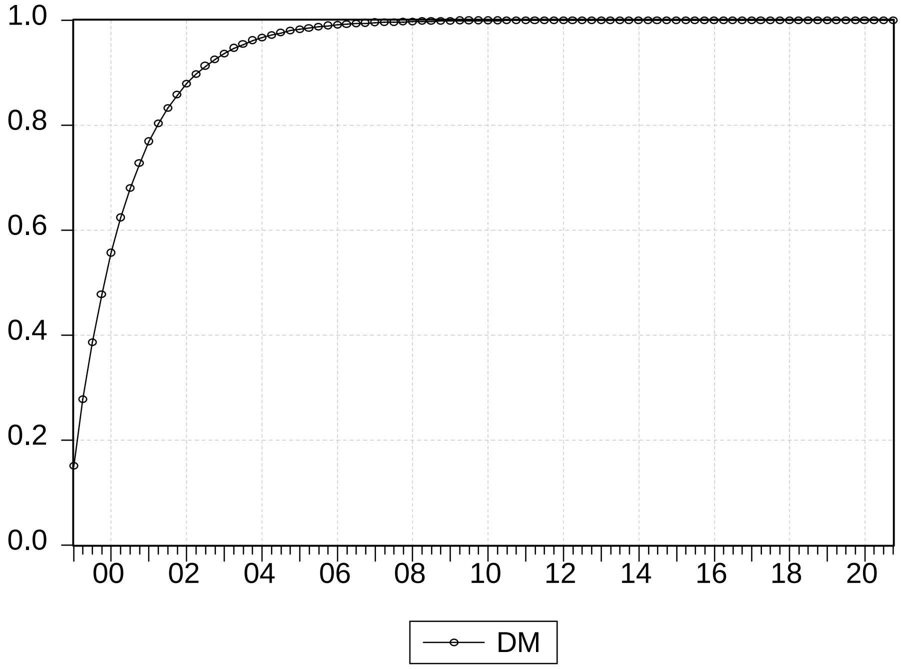


FIG. 4.1.

$$y_t = c_{0,11} + y_t^* - a_{13}(R_t - \pi_t^e) + \epsilon_{1t} \quad (4.38)$$

$$\pi_t = \pi_t^e + a_{21}\left(y_t - y_t^*\right) + \epsilon_{2t} \quad (4.39)$$

$$\pi_t^e = \pi_{t-1} \quad (4.40)$$

$$m_t - p_t = c_{0,31} + a_{31}y_t - a_{33}R_t + \epsilon_{3t} \quad (4.41)$$

$$m_t = c_{0,41} + m_{t-1} \quad (4.42)$$

$$y_t^* = c_{0,51} + c_{0,52}t + \epsilon_{4t} \quad (4.43)$$

Note that money supply is not stochastic as it is considered as fully controlled by the monethary authority. The econometrician's task is the estimation of the unknown parameters to simulate the impact of different path for the exogenous variable controled by the monetary authority. The model uses four equations to determine four endogenous variables, π_t, π_t^e, R_t ed y_t, for given values of the two exogenous variables y_t^* and m_t.The exogeneity status is attributed to y_t^* and m_t, either because they describe the available technology and demography or because they are fully controlled by the policy-maker. Note that, under the hypothesis of dynamic stability of the estimated model, the estimated values for the parameters will only determine the short-run dynamics of output and inflation, as the long-run equilibrium solutions are determined almost independently (totally independently in the case $a_1 = 1$) from the estimated parameters ($\pi = \Delta m - a_1 \Delta y^*$, $y = y^*$).

4.5.2 *Estimation of the parameters of interest*

We consider a monthly data-set for the US economy (which we take as a close economy) to construct, estimate and simulate a version of the macroeconomic model. The data-set, available in EXCEL format as LSZUSA.XLS contains the following variables for the sample 1959:7-1996:3 (for a complete description of the data-set see Leeper-Sims-Zha (1997))

```
CPISA: Consumer price index adjusted for seasonality
M1SA: M1 stock adjusted for seasonality
M2SA: M2 stock adjusted for seasonality
PCM : IMF index of commodity price in dollars
RGDP : real US GDP at quarterly frequencies
RGDPMON : real US GDP at quarterly frequencies (quarterly data
 interpolated by Chow-Lin procedure)
TBILL3 : annually compounded nominal return on three-month TBills
TBOND10: annually compounded redemption yield on 10-year TBonds
```

Having imported the data in EXCEL format into EViews the following transformation are performed using the programme files LSZ.PRG . The program is listed as follows:

```
genr lp=100*log(cpisa)
genr ly=100*log(rgdpmon)
genr infl=(lp-lp(-12))
genr rr=tbill3-infl(-1)
genr lm1=100*log(m1sa)
genr lm2=100*log(m2sa)
genr d12lm2=(lm2-lm2(-12))
genr d12lm1=(lm1-lm1(-12))
genr lyst=773.27+0.275*@TREND(1959:1)
genr vel=lp+ly-lm2
```

We initially deal with non-observable variables. We solve the problem of expected inflation by setting (arbitrarily) $\beta = 1$ in equation (3.11)and substituting lagged inflation for expected inflation. We the obtain an observable proxy for potential output fitting a simple deterministic trend for output.

TABLE 1: LYST=C(1)+C(2)*@TREND(1959:1)

Coefficient	Estimate	Std. Error	t-Statistic	Prob.
C(1)	773.2748	0.466437	1657.833	0.0000
C(2)	0.274909	0.002476	111.0279	0.0000
R-squared 0.975006		Mean dependent var 818.4973		
Adjusted R-squared 0.974927		S.D. dependent var 25.59787		
S.E. of regression 4.053267		Akaike info criterion 2.805316		
Sum squared resid 5191.555		Schwarz criterion 2.828976		
Log likelihood -895.2677		F-statistic 12327.19		
Durbin-Watson stat 0.018264		Prob(F-statistic) 0.000000		

Note that from the estimated parameter values we have that potential ouptut grows at annual rate of $(1+0.0027)^{12}$ -1 = 0. 0329 per cent.

We proceed now to the estimation of all the structural relations included in the model. We begin by money demand, which we simplify to a linear relation between the log of velocity of circulation of money and the short-term interest rate, which we take as a proxy of the opportunity cost of holding money.

TABLE 2: VEL =C(1)+C(2)*TBILL3

Coefficient	Estimate	Std. Error	t-Statistic	Prob.
C(1)	527.9095	0.431272	1224.075	0.0000
C(2)	1.781791	0.061783	28.83933	0.0000
R-squared 0.724668		Mean dependent var 539.0975		
Adjusted R-squared 0.723797		S.D. dependent var 6.392876		
S.E. of regression 3.359777		Akaike info criterion 2.430019		
Sum squared resid 3567.040		Schwarz criterion 2.453679		
Log likelihood -835.5954		F-statistic 831.7067		
Durbin-Watson stat 0.117228		Prob(F-statistic) 0.000000		

Note that the semi-elasticity of the velocity circulation with respect to interest rate is 1.78, implying that increase of hundred basis point in short-term rates is paired with a 178 points increase in velocity circulation. Note that this is not the elasticity, in fact the elasticity $\eta_{r,VEL} = \frac{\partial(p+y-m)}{\partial \log(R)}$,while the semi-elasticity $s\eta_{r,VEL} = \frac{\partial(p+y-m)}{\partial(R)}$, therefore $\eta_{r,VEL} = s\eta_{r,VEL} * R$,as $d\log(R) = \frac{dR}{R}$. Therefore by specifying the money demand with the log of real money as a function of the level of the nominal interest rate has the important implication of making the elasticity of money demand with respect to the opportunity cost of holding money function of the level of interest rates. This is important in that it is not desirable to impose that the elasticity of money demand to interest rate is constant.

The third relation we estimate is an aggregate demand curve:

TABLE 3: LY= C(1)+ LYST +C(2)*(TBILL3-INFL(-1))

Coefficient	Estimate	Std. Error	t-Statistic	Prob.
C(1)	1.470880	0.274338	5.361572	0.0000
C(2)	-0.383906	0.102396	-3.749208	0.0002
R-squared 0.970221		Mean dependent var 820.4813		
Adjusted R-squared 0.970123		S.D. dependent var 24.21989		
S.E. of regression 4.186429		Akaike info criterion 2.870232		
Sum squared resid 5310.435		Schwarz criterion 2.894627		
Log likelihood -868.4866		F-statistic 9871.903		
Durbin-Watson stat 0.021376		Prob(F-statistic) 0.000000		

Note that LYST is included in the fitted relation with a coefficient constrained to unity. As a consequence we can easily compute the level of long-run equilibrium real interest (as the real interest rate obtained by setting $y = y^*$), such level is 3.82=(1.47/0.38).

The fourth estimated relation is the aggregate supply function:

TABLE 4: INFL=C(1)*INFL(-1)+C(2)*(LY-LYST)

Coefficient	Estimate	Std. Error	t-Statistic	Prob.
C(1)	0.9996	0.0032	319.10	0.0000
C(2)	0.027	0.0047	5.89	0.0000
R-squared 0.99		Mean dependent var 5.100013		
Adjusted R-squared 0.99		S.D. dependent var 3.296546		
S.E. of regression 0.32		Akaike info criterion 0.60		
Sum squared resid 32.14056		Schwarz criterion 0.63		
Log likelihood -89.62183		Durbin-Watson stat 1.54		

Note that the estimated values for the parameters are extremely close to the case of maximum price stickiness and the adjustment of inflation with respect to the gap between output and potential output is significant but extremely slow.

4.5.3 *Simulating the effect of monetary policy*

Having estimated the model, we are now in the position to proceed to simulating it by considering the estimated equations as a system of differential equations, which can be solved after the specification of a money supply function.This procedure allows the construction of a baseline, which can be used to evaluate the effect of monetary policy by specifying an alternative rule for monetary policy and by computing multipliers. The E-VIEWS programme SOLVED1.PRG allows the computation of the dynamic multipliers generated by an one per cent increase in money supply. The programme contains the following statements:

```
' This a program to compute dynamic multipliers
if @isobject(''m_temp'')=1 then
```

```
delete m_temp
endif
if @isobject(''dfbase'')=1 then
delete dfbase
endif
if @isobject(''dfshock'')=1 then
delete dfshock
endif
'baseline simulation
smpl 1986:01 2001:12
'define growth rate of money
genr x=6
model dfbase
'define exogenous variables
dfbase.append lm2=lm2(-12) +x
dfbase.append lyst=773.27+0.275*@TREND(1959:01)
'loadind endogenous variables
dfbase.merge df
copy dfbase m_temp
m_temp.append assign @all _bl
m_temp.solve
delete m_temp
group exog_bl d12lm2_bl d12lyst_bl
group endog_bl tbill3_bl infl_bl d12ly_bl
'disturbed simulation
smpl 1986:01 2001:12
'define shock to the growth rate of money
genr y=1
model dfshock
'exogenous variables
dfshock.append lm2=lm2(-12) +(x+y)
dfshock.append lyst=773.27+0.275*@TREND(1959:01)
'loading endogenous variables
dfshock.merge df
copy dfshock m_temp
m_temp.append assign @all _ds
m_temp.solve
delete m_temp
group exog_ds d12lm2_ds d12lyst_ds
group endog_ds tbill3_ds infl_ds d12ly_ds
plot tbill3_bl tbill3_ds
plot infl_bl infl_ds
plot d12ly_bl d12ly_ds
'computing dynamic multipliers
```

```
genr dm_tbill3=(tbill3_ds-tbill3_bl)/(x+y)
genr dm_infl=(infl_ds-infl_bl)/(x+y)
genr dm_d12ly=(d12ly_ds-d12ly_bl)/(x+y)
group dm dm_tbill3 dm_infl dm_d12ly
plot dm
```

The first block of the programme defines objects that will contain the baseline model (dfbase) and the disturbed model (dfshock); it also defines a temporary object (m_temp) which will contain the model to be simulated in each round:

```
if @isobject(''m_temp'')=1 then
delete m_temp
endif
if @isobject(''dfbase'')=1 then
delete dfbase
endif
if @isobject(''dfshock'')=1 then
delete dfshock
endif
```

A baseline simulation is then created. The simulation sample is first chosen; as the model has been estimated over the sample 1959:07-1985:12 we proceed to simulate in from 1986:1 onwards. In fact the sample for the simulation is purely artificial as all series are model generated when computing dynamic multipliers. Chosing a specific sample make sense only when historical values are considered for some of the variables. Having chosen the sample we set the rate of growth of money x at six per cent, the exogenous policy controlled variable. Then all the estimated equations in the previous section are included into the model. The exogenous variables are included using an append statement, while the endogenous variables are included by importing directly into model dfbase the model df, containing all the estimated equations. Then the model is solved dynamically by using Gauss-Seidel and the extension _bl is appended to all generated variables. The variables are then grouped according to their status into exogenous and endogenous.

```
'baseline simulation
smpl 1986:01 2001:12
'define growth rate of money
genr x=6
model dfbase
'define exogenous variables
dfbase.append lm2=lm2(-12) +x
dfbase.append lyst=773.27+0.275*@TREND(1959:01)
'loadind endogenous variables
dfbase.merge df
copy dfbase m_temp
m_temp.append assign @all _bl
```

```
m_temp.solve
delete m_temp
group exog_bl d12lm2_bl d12lyst_bl
group endog_bl tbill3_bl infl_bl d12ly_bl
```

A disturbed sumulation is then created following the same steps

```
'disturbed simulation
smpl 1986:01 2001:12
'define shock to the growth rate of money
genr y=1
model dfshock
'exogenous variables
dfshock.append lm2=lm2(-12) +(x+y)
dfshock.append lyst=773.27+0.275*@TREND(1959:01)
'loading endogenous variables
dfshock.merge df
copy dfshock m_temp
m_temp.append assign @all _ds
m_temp.solve
delete m_temp
group exog_ds d12lm2_ds d12lyst_ds
group endog_ds tbill3_ds infl_ds d12ly_ds
plot tbill3_bl tbill3_ds
plot infl_bl infl_ds
plot d12ly_bl d12ly_ds
```

At the end of the block simulated values for the endogenous variables are plotted. Finally dynamic multiplier are computed , grouped into dm and plotted

```
'computing dynamic multipliers
genr dm_tbill3=(tbill3_ds-tbill3_bl)/(x+y)
genr dm_infl=(infl_ds-infl_bl)/(x+y)
genr dm_d12ly=(d12ly_ds-d12ly_bl)/(x+y)
group dm dm_tbill3 dm_infl dm_d12ly
plot dm
```

We report in the computed dynamic multipliers in Figure 2:

The one per cent increase in money supply has a one-to one impact on inflation in the long-run and a zero impact on deviation on GDP growth. Money is neutral in the long-eun but it does have a short-run impact on the output cycle as prices are sticky. As a consequence of price stickiness we also observe a short-run liquidity effect on interest rates while in the long run the Fisher relationship applies and monetary policy does not have any impact on real interest rates. Note that there is some cyclicality in the interest rate multiplier, this is due to the cycle in nominal interes rates generated by the model. In fact the cycle of output is not matched by any cycle in money supply, which has just a trend. As a consequence nominal interest rates reflect to some extent fluctuations in

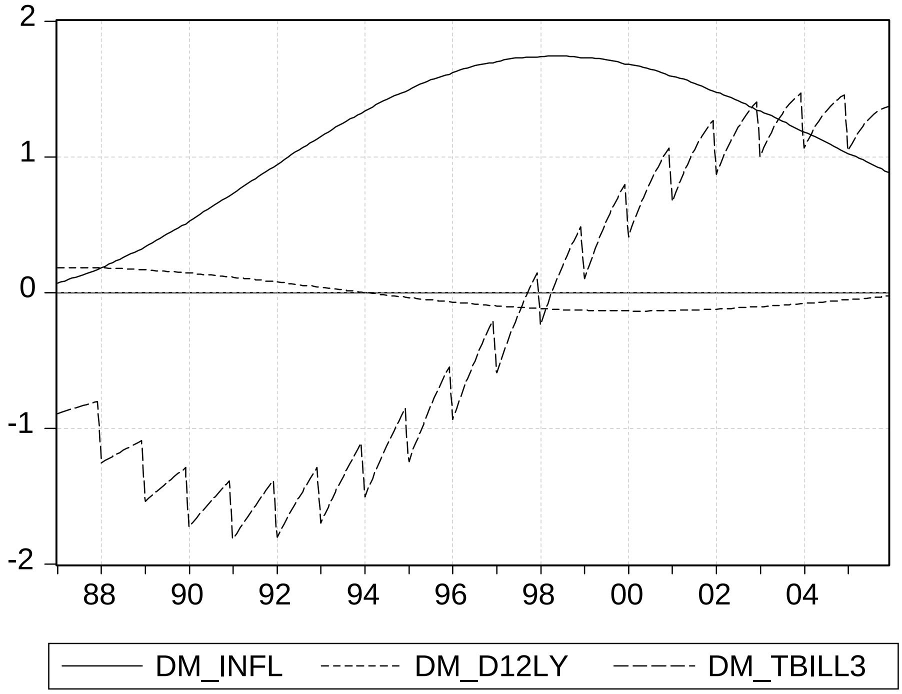


FIG. 4.2. Dynamic Multipliers

output. Setting money supply as completely exogenous generates artifical series incapable of replicating some feature of the observed data. We shall re-address this point later.

At this stage we feel it is more important to concentrate on the reliability of the description of the response of the economy to monetary poicy derived from the dynamic multipliers.

4.6 Assessing econometric evaluation of monetary policy.

To have a first assessment of the reliability of our simulations we the following approach: we assume that the monetary authority has followed rule which delivered the observed data on the money stock and, using such variable as exogenous, we endogenously generated the relevant macroeconomic time series foe a sample covering both the estimation (up to 1985:12) and the simulation (from 1986:1 to 1996:3) period. Such result is obtained by solving the following version of the baseline model:

```
assign @all _fal
lyst=773.27+0.275*@TREND(1959:01)
```

```
'lm2=lm2(-12)+10
tbill3=(-1/1.758218)*(lm2-lp-ly) -(527.8596/1.758218)
ly=lyst+1.47-0.383*(tbill3-infl(-1))
(lp-lp(-12))=0.1 +0.975*(lp(-1)-lp(-13))+0.029*(ly-lyst)
INFL=(lp-lp(-12))
d12lm2=lm2-lm2(-12)
d12lyst=lyst-lyst(-12)
d12ly=ly-ly(-12)
cycle=ly-lyst
```

Note that now the equation for lm2 is commented out: the model will now be solved by taking lm2 as exogenous ad using the historical values for this variable.

We report in Figure 3 and 4 the simulated (defined with suffix _fal) and observed relevant macroeconomic variables: cycle (defined as the deviation of output from trend output) and inflation.

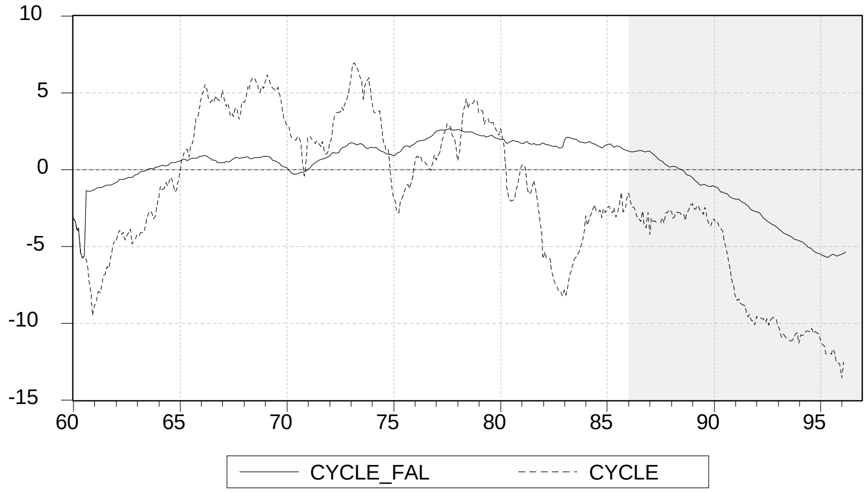


FIG. 4.3. Observed and simulated cycle

Shaded areas distinguish the simulation sample from the estimation sample.The analysis of Figures 3-4 reveals two problems. Over the estimation period the simulated series do no have a sufficiently rich dynamics match the observed time series. However, there is no tendency for a systematic deviation of simulated series from observed series: the difference between model generated and observed time series has a long-memory but there is a pattern for a reversion toward the zero mean. When we revert to the simulation period the first problem persist and, in addition, we start observing a systematic pattern in the divergence between simulated and observed time-series. Such evidence probably justifies some

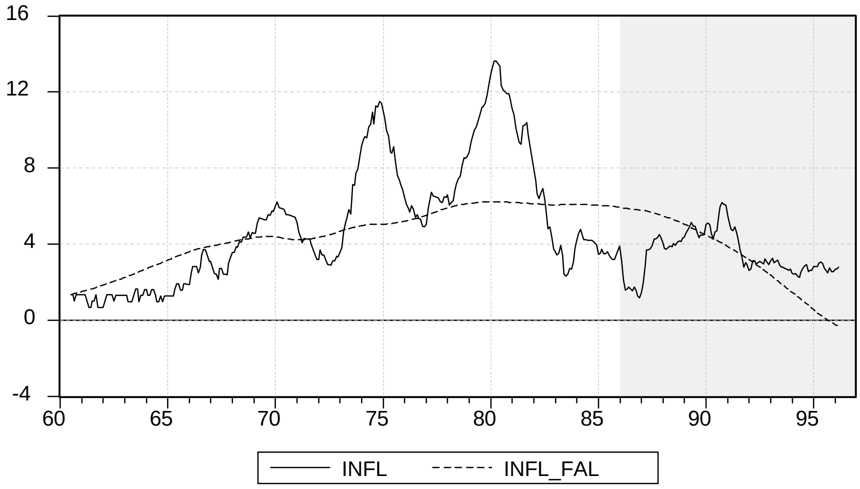


FIG. 4.4. Observed and simulated inflation

skepticism towards econometric evaluation of monetary policy. To further elaborate on this point we consider some diagnosis of the causes behind the problem to discuss the solutions proposed by recent developments in econometric modelling of the monetary transmission mechanism.

4.7 What is wrong with econometric policy evaluation?

The two problems with the our application of the Cowles Commission approach seems to be serious enough to warrant some discussion. We shall organize our discussion by dividing explanation in two classes: those explanations concentrating on modifications in the estimation technique and those suggesting some modifications in he modelling strategy.

The small structural model we have considered is estimated using single equation OLS method. This method is clearly not appropriate, even by taking the "a priori"exogeneity assumptions on money supply and trend output as valid. Consider the money demand equation, alongwith the aggregate demand and supply schedules. These relations establish a simultaneous feedback between output, prices and the interest rate which make not appropriate OLS as the estimation method. In the velocity equation, for example, the nominal interest rate, the only stochastic regressor, should be correlated with the residual and therefore the OLS estimate of the semi-elasticity of money demand with respect to interest rate should be biased. Biased estimates of the parameters of interest could obviously explain the disappointing performance of the model under simulation. However this potential explanation does not seem to be the relevant one. We report in the following table the results of the estimation of the velocity equation by GIVE using valid instruments.

TABLE 5: VEL =C(1)+C(2)*TBILL3

GIVE Estimation				
Coefficient	Estimate	Std. Error	t-Statistic	Prob.
C(1)	527.41	0.439	1212	0.0000
C(2)	1.85	0.06	29.46	0.0000

Instruments:C TBILL3(-1) LY(-1) LP(-1) LP(-2) LM2(-1)	
R-squared 0.72	Mean dependent var 539.0975
Adjusted R-squared 0.72	S.D. dependent var 6.392876
S.E. of regression 3.34	Akaike info criterion 2.42
Sum squared resid 3567.040	Schwarz criterion 2.44
Log likelihood -835.5954	F-statistic 868
Durbin-Watson stat 0.125	Prob(F-statistic) 0.000000

Note that the GIVE estimates are not very different from the OLS. This result is robust to the consideration of full information estimation methods[10]. The observed problems under simulations do not seem to be explained by the chosen estimation method but rather by problems in identification and specification. These are the two issues closely addressed by modern aproaches to econometric modelling. The first problem, i.e. the incapability of the estimated model to capture the observed dynamics of the variables of interest, could be explained by the following considerations:

- The statistical model implicit in the estimated structure is "too restrictive". There are two interpretation of the excessive simplicity in the specification: omission of relevant variables, omission of the relevant dynamics for the included variables (note, for example, that the estimated money demand relation is a simple, static equation)
- the identifying restrictions, altough necessary from to make the estimation meaningful, deliver a structure which cannot adequately describe reality. Think of money supply in the estimated model: if the monetary authority uses money supply as an instrument to achieve given targets for the macroeconomic variables, then it would be very "natural" for money supply to react not only the output and inflation but also to leading indicators for these variables.Assuming money supply as exogenous, the estimated model omits completely a very relevant feedback and looses an important feature of the data. Moreover, by assuming incorrectly exogeneity, the model might induce a spurious statistic efficiacy of monetary policy in the determination of macroeconomic variables. The endogeneity of money does generate a correlations between macroeconomic variables and monetary variables, which, by assuming invalidly, money as exogenous could be interpreted as a

[10] A useful exercise

causal relation running from money to the macroeconomic variables (Sims critique).

The worsening of the model's performance under simulation could instead be explained by the following considerations:

- incorrect specification. Omitted variables have an effect which, not detected when the model is estimated (possibly because the omitted variables were "silent") becomes relevant in explaining parameters' instability of the estimated equations in the simulation period. Incorrectly specified dynamic models feature parameters' instability in out-of-sample simulations.
- Model simulation implies considering alternative monetary policy regimes. A change in regime might imply a structural shift in the parameters of the estimated equations, therefore, the model estimated under the "baseline" regime cannot be used to evaluate the effect of the "control" policy. In other words the "Lucas critique" applies.

In the next chapters we shall consider in turn all these explanations by discussing all the alternative modern approaches to applied macroeconometrics.

REFERENCES

Bollerslev T.(1986) "Generalized Autoregressive Conditional Heteroscedasticity", Journal of Econometrics, 31, 307-327

Chow G. and Lin

Doornik J. and D.F. Hendry (1996) *PcFIML: Interactive econometric modelling of dynamic system*, International Thompson Publishing, London

Engle R.F.(1982) "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation"

Goldberger A.S.(1991) "A course in econometrics", Harvard University Press, Cambridge, Massachussets

Hendry D.F.(1976) "The structure of simultaneous equation estimators" Journal of Econometrics, 4, 51-88

Hendry D.F. (1996) *Dynamic Econometrics*, Oxford University Press, Oxford

Klein L.(1983) *Lectures in Econometrics*, North-Holland, Amsterdam

Johnston (1972) *Econometric Methods,* 2^{nd} edition, New-York:McGraw-Hill

Leeper E.M., C.A. Sims and T. Zha (1996) "What does monetary policy do?", available at ftp://ftp.econ.yale.edu/pub/sims/mpolicy

Lucas R.E. Jr. (1976) "Econometric Policy Evaluation: A Critique", in K. Brunner and A. Meltzer (eds.) *The Phillips curve and labor markets,* Amsterdam, North-Holland

Sargan D.(1988) "Lectures on advanced econometric theory", Basil Blackwell, Oxford

Sims C.A. (1980) "Macroeconomics and Reality", *Econometrica*, 48, 1-48

Sims C.A. (1992) "Interpreting the Macroeconomic Time-Series Facts: the Effects of Monetary Policy", *European Economic Review*, 36, 975-1011

Spanos A. (1990) "The simultaneous-equations model revisited. Statistical adequacy and identification", *Journal of Econometrics*, 44, 87-105

White A.(1980) "A heteroskedastic consistent covariance matrix estimator and a direct test for heteroscedasticity" Econometrica, 48, 817-838

5

THE LSE APPROACH

5.1 Introduction

The LSE approach explains the failure of the Cowles Commission methodology by attributing it to the lack of attention for the statistical model underlying the particular econometric structure adopted to analyse the effect of alternative monetary policies. The LSE methodology considers econometric policy evaluation an interesting and feasible exercise. However the way in which the Cowles Commission approach deals with a legitimate question is not seen as correct. The lack of sufficient interest for the statistical model is at the root of the failure of the Cowles Commission approach to provide at acceptable answer to an interesting question. As it can be seen form the application discussed in the previous chapter, the econometric analyses within the Cowles Commission tradition begin from the idea that the structural form of the process generating the data is known qualitatively, reduced form are then derived from such structures. Within such framework the validity of the reduced form is not tested. The LSE approach views this lack of validation of the reduced form as undermining the credibility of the structural parameter estimates. The LSE approach recognizes that economic theory suggests the general specification of the relevant form, but the precise representation of the Data Generating Process is almost never known in advance. Thus modelling procedure are required to determine the credibility of the estimated models. The reduced form takes a central role within this approach in that it represents the crucial probabilistic structure of the data (see [15],[10]). The traditional logic of the Cowles Commission, according to which the reduced form is derived given the structural model, is turned upside down within the LSE approach. The reduced form is defined first, by defining a system via the set of variables considered, their classification into modelled and non-modelled variables (endogenous and exogenous in the traditional terminology) and the specification of the lag polynomials. The system is then validated by applying the three basic principles of econometrics :" test, test and test". The null hypothesis of interest here being the absence of symptoms of mis-specification, such as residual non-normality, autocorrelation, heteroscedasticity, parameters non-constancy. If the null is not rejected and the system can be considered as a congruent representation of the unknown Data Generating Process, then non-stationarity can be dealt with and the long-run properties of the sytem can be identified by implementing cointegration analysis. Note again that cointegration analysis is fully implemented on the reduced form and the identification of the structural long-run relationships is a totally separated problem from the identification of the structural short-run simultaneous relationships. In the last step a structural model is identified and estimated. No further validation is possible for just-identified model as they impose no restrictions on the system, while the validity of over-identified models is testable by testing the validity of the over-identifying restrictions implicitly imposed on the redeuced form. Finally, policy simulation can be performed after testing that the necessary requirement for the model to be robust to the Lucas critique, i.e. superexogeneity of the relevant variables for the estimation of the parameters of interest, is satisfied.

5.2 The LSE diagnosis.

The LSE diagnosis of the problems displayed by Cowles Commission models is simple

"...the statistical properties attributed to the structural estimators and related tests are in general invalid unless the probabilistic structure imposed on the data via the reduced form is invalid. A glance at the empirical literature confirms that not only are the statistical assumptions underlying the reduced form not tested, but the reduced form is rarely estimated explicitly. Indeed, the most popular estimation methods for the structural parameters are limited-information instrumental-variable methods such as two-stage least squares which do not even specify the implied reduced form..." $([15], p.90)$.

Consider the structural model used to illustrate the Cowles Commission approach in the previous chapter :

$$\mathbf{A}\begin{bmatrix} y_t \\ p_t \\ R_t \\ m_t \\ y_t^* \end{bmatrix} = \mathbf{C}_0\begin{bmatrix} 1 \\ t \end{bmatrix} + \mathbf{C}_1\begin{bmatrix} y_{t-1} \\ p_{t-1} \\ R_{t-1} \\ m_{t-1} \\ y_{t-1}^* \end{bmatrix} + \mathbf{C}_2\begin{bmatrix} y_{t-2} \\ p_{t-2} \\ R_{t-2} \\ m_{t-2} \\ y_{t-2}^* \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \\ \epsilon_{4t} \\ \epsilon_{5t} \end{bmatrix} \tag{5.1}$$

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & a_{13} & 0 & -1 \\ -a_{21} & 1 & 0 & 0 & a_{21} \\ -\frac{a_{31}}{a_{33}} & -\frac{1}{a_{33}} & 1 & \frac{1}{a_{33}} & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{C}_0 = \begin{bmatrix} c_{0,11} & 0 \\ c_{0,21} & 0 \\ c_{0,31} & 0 \\ c_{0,41} & 0 \\ c_{0,51} & c_{0,52} \end{bmatrix}, \quad \mathbf{C}_1 = \begin{bmatrix} 0 & a_{13} & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{C}_2 = \begin{bmatrix} -a_{13} & 0 \\ 0 & -1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

only parameters different from zero or one are to be considered "free" and are then estimated to describe the economic properties of the adopted structure.

From this representation we immediately note the the exogeneity assumptions implies a remarkable number of the restrictions on the set of the parameters of interest. Another substantial set of restrictions derives from the very limited dynamics adopted in the specification of the model. The implied reduced form features a remarkable number of restrictions, as easily checked by pre-multiplying the model by $\mathbf{A}^{-1}$:

$$\begin{bmatrix} p_t \\ y_t \\ R_t \\ m_t \\ y_t^* \end{bmatrix} = \begin{bmatrix} d_{0,11} & d_{0,12} \\ d_{0,21} & d_{0,22} \\ d_{0,31} & d_{0,32} \\ d_{0,41} & 0 \\ d_{0,51} & d_{0,52} \end{bmatrix} \begin{bmatrix} 1 \\ TREND \end{bmatrix} + \begin{bmatrix} d_{1,11} & 0 & 0 & d_{1,13} & 0 \\ d_{1,21} & 0 & 0 & d_{1,23} & 0 \\ d_{1,31} & 0 & 0 & d_{1,23} & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} p_{t-1} \\ y_{t-1} \\ R_{t-1} \\ m_{t-1} \\ y_{t-1}^* \end{bmatrix}$$
$$\begin{bmatrix} d_{2,11} & 0 & 0 & 0 & 0 \\ d_{2,21} & 0 & 0 & 0 & 0 \\ d_{2,31} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} p_{t-2} \\ y_{t-2} \\ R_{t-2} \\ m_{t-2} \\ y_{t-2}^* \end{bmatrix} + \mathbf{A}^{-1} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \\ \epsilon_{4t} \\ \epsilon_{5t} \end{bmatrix}$$

According to the LSE criticism the validity of such reduced form is not properly addressed within the Cowles Commission tradition. Structural inference with based on an improper statistical model is the LSE diagnosis for the failure of the Cowles Commission approach model by evaluating the properties of the residuals.

5.3 The reduction process

Econometric modelling is formalized within the LSE camp as the result of a reduction process. The starting point of the reduction process is a long way up: think of a vector $\mathbf{x}_t$ containing observations an all economic variables at time t. A sample of T time series observations on all the variables can be represented as follows:

$$\mathbf{X}_T^1 = \begin{bmatrix} \mathbf{x}_1 \\ . \\ . \\ . \\ \mathbf{x}_t \end{bmatrix}$$

The starting point of the reduction process is a model for the Data Generating Process(DGP). The DGP is described by the joint density function $D\left(\mathbf{X}_T^1, \boldsymbol{\theta}\right)$ where $\mathbf{X}_{t-1}$ is the matrix incuding observations on all variables in $\mathbf{x}$ from time 1 to time $t-1$, and $\boldsymbol{\theta}$ is a set of parameters.

Model specification amounts to choosing a particular functional form for the density. Having chosen the model, a structure for the model is pinned down by identifying parameters and estimating them. In general estimation is performed by considering the joint sample density function, known also as the likelilihood function, which we can express as $D\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right)$.The likelihood function is defined on the parameters space Θ, given the observation of the observed sample $\mathbf{X}_T^1$ and of a set of initial conditions $\mathbf{X}_{0.}$ Such initial conditions can be interpreted as the pre-sample observations on the relevant variables (which are usually not available). In case of independent observations the likelihood function can be written as the product of the density functions for each observation. However this is not the relevant case for time-series, as time-series observations are in general sequentially correlated. In the case of time-series the sample density is then constructed using the concept of sequential conditioning. The likelihood function, conditioned with respect to initial conditions, can always be written as the product of a marginal density and a conditional density as follows:

$$D\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) = D\left(\mathbf{x}_1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) D\left(\mathbf{X}_T^2 \mid \mathbf{X}_1, \boldsymbol{\theta}\right)$$

Obviously we also have

$$D\left(\mathbf{X}_T^2 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) = D\left(\mathbf{x}_2 \mid \mathbf{X}_1, \boldsymbol{\theta}\right) D\left(\mathbf{X}_T^3 \mid \mathbf{X}_2, \boldsymbol{\theta}\right)$$

and, by recursive substitution, we eventually obtain :

$$D\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) = t = 1\prod^T D\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\theta}\right)$$

Having obtained $D\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right)$we can in theory derive $D\left(\mathbf{X}_T^1, \boldsymbol{\theta}\right)$ by integrating with respect to X_0 the density conditional on pre-sample observations. In practice this could be not tractable analitically as $D\left(X_0\right)$ is not known. The hypothesis of stationarity becomes crucial at this stage, as stationarity restricts the memory of time series and limits to the first observations in the sample the effects of pre-sample observations. This is the reason why, in the case of stationary processes, initial conditions can be simply ignored. Clearly the larger the sample, the better as the weights of the information lost becomes smaller. Moreover note also that, even by omitting initial conditions we have:

$$D\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) = D\left(\mathbf{x}_1 \mid \mathbf{X}_0, \boldsymbol{\theta}\right) t = 2\prod^T D\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\theta}\right)$$

therefore the likelihood function is separated in the product on $T-1$ conditional distribution and one unconditiona distribution. In the case of non-stationarity the unconditional distribution is not defined. On the other hand, in the case of stationarity the DGP is completely described by the conditional density function $D\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\theta}\right)$.

The base line of the reduction process, the DGP or the Haavelmo distribution, is then completely described by $D(\mathbf{x}_t \mid \mathbf{X}_{t-1}, \boldsymbol{\theta})$. The first step of the reduction process can be understood by partitoning $\mathbf{x}$ in three types of variables:

$$\mathbf{x}_t = (\mathbf{w}_t, \mathbf{y}_t, \mathbf{z}_t)$$

$\mathbf{w}_t$ identifies variables which are not observables or are not relevant to the problem investigated by the econometrician. In practice these variables are ignored, in theory such result is obtained by factorising the joint density and integrating it with respect to $\mathbf{w}_t$:

$$D(\mathbf{y}_t, \mathbf{z}_t \mid \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta}) = \iint D(\mathbf{y}_t, \mathbf{z}_t, \mathbf{w}_t \mid \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \mathbf{W}_{t-1}, \boldsymbol{\theta}) \\ D(\mathbf{W}_{t-1} \mid \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\theta})\, d\mathbf{W}_{t-1} d\mathbf{w}_t$$

In this case we have a potential information loss which becomes real when the variables judged irrelevant for the problem at hand are not so. In formal terms, we do not have any information loss only if:

$$D(\mathbf{y}_t, \mathbf{z}_t \mid \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta}) = D(\mathbf{y}_t, \mathbf{z}_t, \mathbf{w}_t \mid \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \mathbf{W}_{t-1}, \boldsymbol{\theta})$$

This is the statistical description of the model considered by the econometrician, it is in other words the reduced form of the structure of interest to the economy. In general, at the empirical level this, is the earliest stage of the reduction process, in fact a reduced form for all the variables of interest (a VAR) is the most general model we fitted to the data. However, such general model viable for empirical estimation does not certainly coincide with the Haavelmo distribution for all economic variables! How can we be sure that no loss of relevant information occurred in moving from the Haavelmo distribution to the estimated empirical model ? By applying the three fundamental rules of LSE econometrics "test, test and test" to our reduced form. In fact $D(\mathbf{y}_t, \mathbf{z}_t \mid \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta})$ is empirically constructed by parameterising $E(\mathbf{y}_t, \mathbf{z}_t \mid \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta})$ as follows:

$$E(\mathbf{y}_t, \mathbf{z}_t \mid \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta}) = \begin{pmatrix} \boldsymbol{\beta}_{11}(L) & \boldsymbol{\beta}_{12}(L) \\ \boldsymbol{\beta}_{21}(L) & \boldsymbol{\beta}_{22}(L) \end{pmatrix} \begin{pmatrix} \mathbf{y}_{t-1} \\ \mathbf{z}_{t-1} \end{pmatrix}$$

From the specification of conditional means the vector of innovations $\mathbf{u}_t$ is derived :

$$\mathbf{u}_t = \begin{pmatrix} \mathbf{y}_t \\ \mathbf{z}_t \end{pmatrix} - \begin{pmatrix} \boldsymbol{\beta}_{11}(L) & \boldsymbol{\beta}_{12}(L) \\ \boldsymbol{\beta}_{21}(L) & \boldsymbol{\beta}_{22}(L) \end{pmatrix} \begin{pmatrix} \mathbf{y}_{t-1} \\ \mathbf{z}_{t-1} \end{pmatrix}$$

Going back to our application to the monetary transmission mechanism, the baseline of the investigation is a reduced form of the type:

$$\begin{pmatrix} \mathbf{Y}_t \\ \mathbf{M}_t \end{pmatrix} = D_1(L) \begin{pmatrix} \mathbf{Y}_{t-1} \\ \mathbf{M}_{t-1} \end{pmatrix} + D_2 a_1 + \begin{pmatrix} \mathbf{u}_t^Y \\ \mathbf{u}_t^M \end{pmatrix} \tag{5.2}$$

Any empirical model is in itself the product of some step in the reduction process. So the starting point of the empirical analysis is the implementation of a battery of diagnostic tests, where the null hypothesis of interest is the validity of the baseline model as a simplified representation of the unknown DGP

5.4 Test, test and test

Given that the GDP is unknown, the validity of reduction can be checked by ensuring that the vector of innovations $\mathbf{u}_t$ possesses all the features of true statistical innovations: absence of correlation, heteroscedasticity, non-normality. Any pattern of this type or any instability in the $\boldsymbol{\beta}$ parameters can then be interpreted as a signal of a loss information occurred in the hidden reduction from the DGP to the particular estimated form adopted. The three fundamental principles of the LSE methodology are "test, test, and test" because only by implementing diagnostic checks we can discard invalid structural models. Testing usually concentrates on residuals because any non-randomness in residual behaviour could be interpreted as a signal of incorrect specification of the underlying model. The residuals of a statistical model are generated by the specification adopted by the econometrician and are a by-product of omitted variables (both in the sense of omitted important variables and of omitted lags of included variables), and errors-in-included-variables of several type (measurement errors, expectational errors). We illustrate how the relevant tests can be constructed with reference to the statistical model (5.2).

5.4.1 *Testing autocorrelated residuals*

Residual autocorrelation is usually tested via a Lagrange Multiplier test [4], which uses the following formulation:

$$\begin{pmatrix} \hat{\mathbf{u}}_t^Y \\ \hat{\mathbf{u}}_t^M \end{pmatrix} = i = 1\sum^n \boldsymbol{\delta}_i' \begin{pmatrix} \hat{\mathbf{u}}_{t-i}^Y \\ \hat{\mathbf{u}}_{t-i}^M \end{pmatrix} + D_1(L) \begin{pmatrix} \mathbf{Y}_{t-1} \\ \mathbf{M}_{t-1} \end{pmatrix} + D_2 a_1 + \begin{pmatrix} \mathbf{e}_t^Y \\ \mathbf{e}_t^M \end{pmatrix} \tag{5.3}$$

So residuals autocorrelation of the n-th order is checked by testing if the components of lagged fitted residuals not explained by the regressors in the original model are significant in explaining contemporaneous fitted residuals. A test against the null of absence of serial correlation of order n is implemented by consider lags up to the n-th of fitted residuals. The null hypothesis of interest is:

$$H_0 : \boldsymbol{\delta}_i = 0$$

The test, based on the R^2 of the auxiliary system, is asymptotically distributed as a χ^2 with nm^2 degrees of freedom, where m is the number of the variables entering the reduced form[8]. An F-approximation with small sample corrections is also available[12]. The intuition of this procedure of model evaluation by variable addition is understood by considering that, nder the null hypothesis, the component of lagged residuals not explained by the regressors in the model is not significant in explaining current residuals[14].

5.4.2 *Testing heteroscedastic residuals*

To illustrate tests for the null of homoscedasticity, consider the simple case where we have a system including two variables, one monetary variable and one macroeconomic variable. After estimation of (5.2) , a tests can be performed by running the following auxiliary model:

$$\begin{pmatrix} \left(\hat{u}_t^{Y}\right)^2 \\ \left(\hat{u}_t^{M}\right)^2 \\ \left(\hat{u}_t^{Y}\hat{u}_t^{M}\right) \end{pmatrix} = \boldsymbol{\delta}_0 + D^*\left(L\right)\begin{pmatrix} Y_{t-1}^2 \\ M_{t-1}^2 \\ Y_{t-1}M_{t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \end{pmatrix} \tag{5.4}$$

Under the null hypothesis the variance-covariance of the system residuals is constant. Hence, we have

$$H_0 : D^*\left(L\right) = 0$$

The test is easily generalized to systems of m variables, with the proviso that, as m gets large, the limitation in degrees of freedom might make it not feasible. The procedure is best interpreted as an extension of the heteroscedasticity tests proposed by White[17] in the context of single-equation models. Of course, whenever the degrees of freedom problem is binding, a White test can be run on all the equations separately. Not rejecting the null in this case would satisfy a necessary condition for homoscedasticity of the sytem residuals. The condition is not sufficient because it does not provide a test for constancy of covariances. At the single equation level, ARCH[6] type of tests could also be run, by specifying the following models:

$$\begin{pmatrix} \left(\hat{u}_t^{Y}\right)^2 \\ \left(\hat{u}_t^{M}\right)^2 \end{pmatrix} = \boldsymbol{\delta}_0 + i = 1\sum^{n}\boldsymbol{\delta}_i' \begin{pmatrix} \left(\hat{u}_{t-i}^{Y}\right)^2 \\ \left(\hat{u}_{t-i}^{M}\right)^2 \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \tag{5.5}$$

Where the null of interest is:

$$H_0 : \boldsymbol{\delta}_i = 0$$

Note that all the test for heteroscedasticy here presented take some estimate of the variance-covariance matrix of the system residual and check its constancy over time. The difference between different tests lies in the specification of the alternative, i.e. of the variables used to capture the fluctuations over time of the moments under the alternative distribution.

5.4.3 *Testing residuals normality*

Normality of residuals is a crucial property in that all the statistical framework used to "test, test and test" is based on this assumption. A vector normality tests has been proposed by Doornik and Hansen [5].

The test is constructed by first standardising the residuals $\begin{pmatrix} \hat{\mathbf{u}}_t^{Y} & \hat{\mathbf{u}}_t^{M} \end{pmatrix}$. Define the vector of standardised residuals$(\mathbf{r}_{1,...,}\mathbf{r}_T)$ as $\mathbf{R}$. So $\mathbf{C} = T^{-1}\mathbf{R}'\mathbf{R}$, is the correlation matrix. The standardised residuals, normally distributed under the null with zero mean and variance-covariance matrix $\mathbf{C}$, can be transformed into independent standard normals:

$$\mathbf{e}_t = \mathbf{E}\Lambda^{-\frac{1}{2}}\mathbf{E}'\mathbf{r}_t$$

where Λ is a diagonal matrix with the eigenvalues of $\mathbf{C}$ on the principal diagonal and the columns of $\mathbf{E}$ are the correspondent eigenvectors, such that $\mathbf{E}'\mathbf{E} = \mathbf{I}$, and $\Lambda = \mathbf{E}'\mathbf{C}\mathbf{E}$.

The test is performed by computing univariate skewness and kurtosis of each transformed residuals and comparing them with those of the normal distribution. Define $\mathbf{b}_1' = (b_{11},...,b_{1m})$, $\mathbf{b}_2' = (b_{21},...,b_{2m})$, as the vectors containing the sample estimates of the skewness and kurtosis of the transformed residuals of the m equations included in the model we have that the test statistic:

$$\frac{T\mathbf{b}_1'\mathbf{b}_1}{6} + \frac{T\left(\mathbf{b}_2 - 3\mathbf{i}\right)'\left(\mathbf{b}_2 - 3\mathbf{i}\right)}{24} \overset{asy}{\sim} \chi^2\left(2m\right)$$

where $\mathbf{i}$ is the unit vector. As the above requires large samples, corrected versions are proposed and implemented in the PC-FIML package.

5.4.4 *Testing parameters stability*

Within the LSE methodology variable parameters is an oxymoron. In fact "...Models which have no set of constancies will be useless for forecasting the future, analysing economic policy, or test economic theories, since they lack entities on which to base those activities..."[7]. Testing parameter constancy is therefore an important aspect of the diagnostic checking procedure. This is usually done within the LSE tradition by estimating models recursively and applying Chow tests for parameters stability[**?**].

Single equation Chow tests include 1-step F-tests, break-point F-tests and forecast F-tests.

1-step forecasts tests are $F(1, t-k-1)$ under the null of constant parameters, for $t = N, ...T$ and k included regressors. A typical statistic is calculated as:

$$\frac{(RSS_t - RSS_{t-1})(t-k-1)}{RSS_{t-1}}$$

Where RSS_t is the residual sum of squares computed from the estimation on t observations. And they are computed by PC-GIVE and PC-FIML for all possible break points after initialization of the estimation.

Break-point F-tests are $F(T-t+1, t-k-1)$ for $t = N, ..T$. The null of interest is the stability of parameters when model is estimated on the sample 1 to t against an alternative which allows any form of change over $t+1$ to T. A typical statistic is calculated as

$$\frac{(RSS_T - RSS_{t-1})(t-k-1)}{RSS_{t-1}(T-k-1)}$$

Forecast F-test are $F(T-N+1, M-k-1)$ for $t = N, ..T$, they test stability of the model estimated on the sample 1 to $(N-1)$ against an alternative which allows any form of change over N to T. A typical statistic is calculated as

$$\frac{(RSS_T - RSS_{N-1})(N-k-1)}{RSS_{N-1}(T-N-1)}$$

All these tests can be extended to systems by defining F-approximations to likelihood ratios statistics[4].

Chow tests are tests for instability generated by a single-break point, occurring at a known date within the sample. Refinements of the testing procedure have been proposed to deal with breaks occuring at uncertain dates and with multiple breaks. Andrews [1] proposes to deal with uncertainty by using trimming points to define a subsample in which the break has likely occurred , by then computing all possible Chow tests (in χ^2 form) for every breakpoint. The largest statistic so obtained provides a stability test ("maximum Chow" test) for an unknown break point. The article provides the underlying distributional theory and critical values, which are function of degres of fredom and trimming points.

5.5 Testing the Cowles Commission model

We consider as a baseline modelling the following generalization of the statistical model underlying the simple Cowles Commission specification:

$$\begin{bmatrix} \pi_t \\ y_t \\ R_t \\ (m-p)_t \end{bmatrix} = \begin{bmatrix} d_{0,11} & d_{0,12} \\ d_{0,21} & d_{0,22} \\ d_{0,31} & d_{0,32} \\ d_{0,41} & d_{0,42} \end{bmatrix} \begin{bmatrix} 1 \\ TREND \end{bmatrix} + i = 1 \sum^{6} D_i \begin{bmatrix} \pi_{t-i} \\ y_{t-i} \\ R_{t-i} \\ (m-p)_{t-i} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ u_{4t} \end{bmatrix}$$

We estimate the abover specification by OLS, using PC-FIML, over the sample 1959:7-1985:12. The residuals for the four equation in the system are reported in Figure (**??**), while diagnostic tests are reported in Table 1.

TABLE 1: Diagnostic tests

	$AR1-7F(7,267)$	$Normality\chi^2(2)$	$ARCH7F(7,260)$	$Xi\hat{}2F(50,223)$
LY	0.7691 [0.6137]	4.9172 [0.0856]	0.72095 [0.6543]	0.60953 [0.9807]
INFL	2.671 [0.0109] *	9.9844 [0.0068] **	0.922 [0.4898]	1.5344 [0.0196] *
LM2P	1.9475 [0.0625]	21.096 [0.0000] **	2.4126 [0.0208] *	1.1241 [0.2807]
TBILL3	2.1913 [0.0353] *	162.11 [0.0000] **	20.494 [0.0000] **	4.7914 [0.0000] **
Vector	2.4529 [0.0000] **	214.5 [0.0000] **	-	1.3827 [0.0000] **

The plot of the residuals and the results of the diagnostic tests reported in Table 1 make a point: the adopted specification does not deliver an acceptable statistical model. Given that this model is more general than the simple specification used to illustrate the Cowles commission approach, the results is valid *a fortiori* for such model.

5.6 Searching for a congruent specification

In the previous section we have illustrated the diagnosis of the problems of the Cowles commission models proposed by LSE. We consider now the prognosis: begin the search of the final specification starting from an appropriate statistical model for the data. Looking at the behaviour of the residuals from the previous estimated model we note that the equation for the interest rate shows a substantial degree of instability over the period 1979-1982. In fact in this period a different monetary regime has been adopted by the Fed who abandoned a strategy aimed at controlling interest rates to embrace a non-borrowed reserves targeting regime. As a consequence the volatility of short-term interest rates changes dramatically over the period 1979-1982. Such volatility goes back to

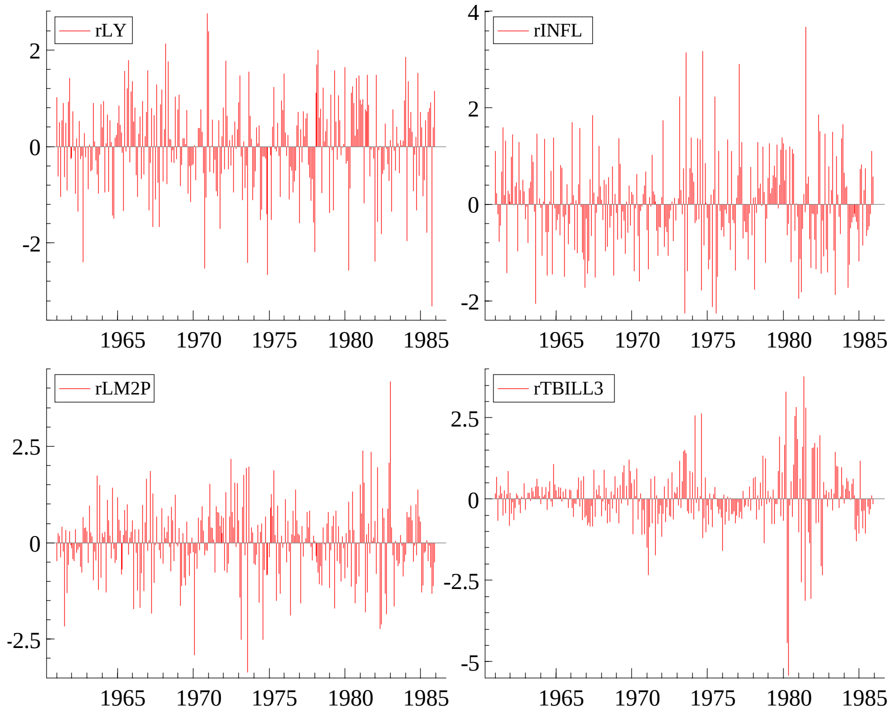


FIG. 5.1.

pre-1979 levels only when non-borrowed reserves targeting is abandoned at the end of 1982 (see Walsh[?]) motivation to be introduced here. Mixing two different policy regimes is a recipe for parameters instability, therefore we concentrate on a single regime and shorten the sample to end estimation in 1979:10. A second problem is detected by the diagnostics in the equation for inflation. Several outliers here are generated by the oil price shocks of 1973 and 1979. To fix this problem we include among the variables in the system an index of commodity prices. Such variable could also be important in modelling the monetary policy maker behaviour, if it palys a role as a leading indicator for inflation. Lastly to model properly money demand it seems necessary to consider explicitly the own return on money in the construction of the opportunity cost of holding this asset. A time series for this variable is made available by the Fed at the internet site http//www.??.??. . we extend our baseline system to include such new variable. We re-estimate the system over the shortened sample with the new endogenous and exogenous variables. As the residuals show some persistent sign of non-normality we include a set of dummies to remove outliers (observations generating observed residuals of a magnitude exceeding, in absolute value, three times the standard deviation of fitted residuals). We then choose the following as our baseline model:

$$\begin{bmatrix} y_t \\ \pi_t \\ R_t^m \\ R_t^b \\ (m-p)_t \end{bmatrix} = \begin{bmatrix} d_{0,11} & d_{0,12} \\ d_{0,21} & d_{0,22} \\ d_{0,31} & d_{0,32} \\ d_{0,41} & d_{0,42} \end{bmatrix} \begin{bmatrix} 1 \\ TREND \end{bmatrix} + i = 1 \sum^{6} D_i \begin{bmatrix} y_{t-i} \\ \pi_{t-i} \\ R_{t-i}^m \\ R_{t-i}^b \\ (m-p)_{t-i} \end{bmatrix} + \quad (5.6)$$

$$+ i = 0 \sum^{6} F_i \Delta_{12} LPCM_{t-i} + \mathbf{g}'\mathbf{DUM}_t + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ u_{4t} \end{bmatrix}.$$

$\mathbf{DUM}_t$ is a vector of dummy variables containing: dum7306,dum7307, dum7308, dum7310, dum7311, dum7312, dum7402, dum7403, dum7407, dum7408,dum7409, dum7501, dum7505, dum7806, dum7808, dum7811, dum7904. In general, dumMMYY is a variable taking value 1 in the MM month of the year YY and zero anywhere else.

We plot the residuals in Figure (5.2) and report the diagnostic tests in Table 2.

TABLE 2:

	$AR1-7F(7,267)$	$Normality\chi^2(2)$	$ARCH7F(7,260)$	$Xi\hat{}2F(50,223)$
LY	0.92159 [0.4913]	1.577 [0.4545]	1.0196 [0.4196]	0.82561 [0.7929]
INFL	1.4775 [0.1787]	3.1857 [0.2033]	0.38832 [0.9081]	0.59005 [0.9874]
LM2P	2.0011 [0.0580]	3.04 [0.2187]	0.49178 [0.8395]	0.69226 [0.9418]
M2OWN	5.2877 [0.0000] **	12.04 [0.0024] **	0.70325 [0.6693]	1.0182 [0.4606]
TBILL3	1.0123 [0.4246]	6.6683 [0.0356] *	2.8439 [0.0081] **	0.89271 [0.6835]
Vector	1.4779 [0.0004] **	24.745 [0.0058] **		0.66171 [1.0000]

The situation looks much improved now, although the equation for the own rate on money shows still some problem of autocorrelation and non-normality, signalled both by the single-equation and the system diagnostics. We attribute this problems to the very peculiar time-series behaviour of this series (see Figure X) and decide to proceed further in our analysis by considering 5.6 as a congruent representation for the unknown Data Generating Process.

5.7 Cointegration Analysis

The next step in the specification strategy is the identification of the long-run equilibria in our model. The number of cointegrating vectors can be detected by applying the Johansen procedure to identify the rank of the matrix Π in the following re-parameterisation of our model:

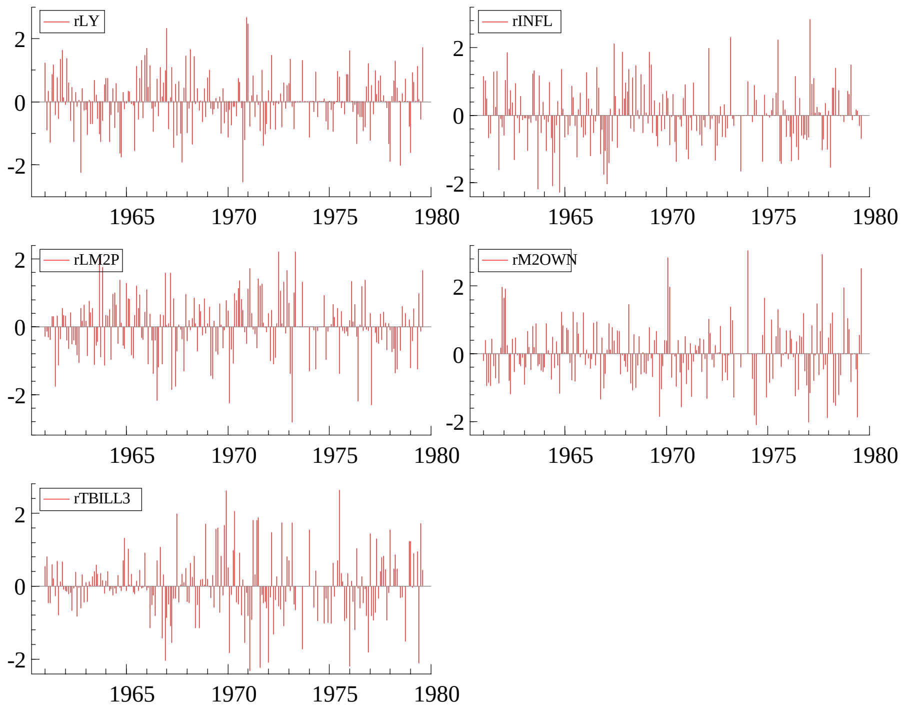


FIG. 5.2.

$$
\begin{bmatrix} \Delta y_t \\ \Delta \pi_t \\ \Delta R_t^m \\ \Delta R_t^b \\ \Delta (m-p)_t \end{bmatrix} = \begin{bmatrix} d_{0,11} \\ d_{0,21} \\ d_{0,31} \\ d_{0,41} \end{bmatrix} + i = 0 \sum^{6} F_i \Delta_{12} LPCM_{t-i} + \mathbf{g}' \mathbf{DUM}_t +
$$

$$
i = 1 \sum^{5} D_i \begin{bmatrix} \Delta y_{t-i} \\ \Delta \pi_{t-i} \\ \Delta R_{t-i}^m \\ \Delta R_{t-i}^b \\ \Delta (m-p)_{t-i} \end{bmatrix} + \Pi \begin{bmatrix} y_{t-6} \\ \pi_{t-6} \\ R_{t-6}^m \\ R_{t-6}^b \\ (m-p)_{t-6} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ u_{4t} \\ u_{5t} \end{bmatrix}
$$

We have already seen in Chapter 2 that two identification schemes deliver alternative representations of the long-run equilibria based on over-identifying restrictions not rejected by the data. The first scheme is centered upon a money demand relation, while the second is centered upon an interest rate reaction function. We have also shown that the analysis of the adjustment parameters makes the second scheme preferable. On the basis of these results, we opt for the second identification scheme and proceed to specify a structural model for the policy rates, output and inflation. Within such scheme, real money is completely determined by the demand side and looses any interest for the analysis of monetary policy. When the researcher looses interest in real money the return on money becomes also uninteresting. Our economic interpretation of the results of the cointegration analysis makes our baseline reduced form unnecessarily complicated. The natural question at this point regards the legitimacy of a simplification of the model in moving from the reduced form to the structural model of interest.

5.8 Specifying the structural model

Having validated the reduced form, the econometrician is left with the problem of identifying the appropriate structure. Moreover, we have seen that the reduced form might constitute in itself a model unnecessarily complicated for the problem at hand. It is then important to identify the cases in which further simplification, obtained by reducing the dimension of the estimated system, is viable with no loss of relevant information for the purposes of analysis.

5.8.1 *Exogeneity*

Suppose that the relevant problem is inference on subset $\boldsymbol{\beta}_1$ of the parameters determiningthe joint density of

$\mathbf{y}_t$, and $\mathbf{z}_t$.In general it is always possible to re-write $D(\mathbf{y}_t, \mathbf{z}_t \mid \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta})$ as follows:

$$D(\mathbf{y}_t, \mathbf{z}_t \mid \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta}) = D(\mathbf{y}_t \mid \mathbf{z}_t, \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)\, D(\mathbf{z}_t \mid \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \tag{5.7}$$

The general case admit as a specific case the existence of a "sequential cut", which we represent as follows:

$$D(\mathbf{y}_t, \mathbf{z}_t \mid \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta}) = D(\mathbf{y}_t \mid \mathbf{z}_t, \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta}_1)\, D(\mathbf{z}_t \mid \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta}_2) \tag{5.8}$$

If this is the case and if the set on which the parameters $\boldsymbol{\beta}_1$are defined is totally independent from the set on which the parameters $\boldsymbol{\beta}_2$ are defined ($\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are *variation free*) then inference on $\boldsymbol{\beta}_1$could be performed by concentrating only on the conditional density for $\mathbf{y}_t$, without explicitly treating the marginal density for $\mathbf{z}_t$. To have an intuition of this argument think of the problem of deriving an estimator of $\boldsymbol{\beta}_1$ by using the(5.8) as the likelihood function. Taking logs of (5.8) we have:

$$\log D(\mathbf{y}_t, \mathbf{z}_t \mid \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta}) = \log D(\mathbf{y}_t \mid \mathbf{z}_t, \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta}_1) + \log D(\mathbf{z}_t \mid \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta}_2) \tag{5.9}$$

from which it is clear that the log of the joint process is equal to the sum of two factors. The second factor is a constant with respect to $\boldsymbol{\beta}_1$, it does not affect the maximum likelihood estimator of $\boldsymbol{\beta}_1$ and can be ignored when the main interest of research is inference on $\boldsymbol{\beta}_1$. When the sequential cut can be operated and $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are "variation free", $\mathbf{z}_t$ is said to be weakly exogenous for the estimation of $\boldsymbol{\beta}_1$. Weak exogeneity can be confronted with Granger non-causality (Granger(1987)). $\mathbf{z}_t$ Granger-causes $\mathbf{y}_t$ if the knowledge of $\mathbf{z}_t$ helps the prediction of $\mathbf{y}_{t+j}$,j>0. Granger- causality is independent from the choice of the parameters of interest, while weak exogeneity obviously is. As a consequence it is perfectly admissible that $\mathbf{z}_t$ is not Granger-caused by $\mathbf{y}$ but these variables are not weakly exogenous for the estimation of the parameters of interest. Think of the following case:

$$D(\mathbf{y}_t, \mathbf{z}_t \mid \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta}) = D(\mathbf{y}_t \mid \mathbf{z}_t, \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2) D(\mathbf{z}_t \mid \mathbf{Z}_{t-1}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \tag{5.10}$$

The link between Granger causality and weak-exogeneity is established by the concept of strong-exogeneity, which is defined as the intersection of the two concepts, therefore we have strong-exogeneity when the joint density can be factorised as follows:

$$D(\mathbf{y}_t, \mathbf{z}_t \mid \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta}) = D(\mathbf{y}_t \mid \mathbf{z}_t, \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta}_1) D(\mathbf{z}_t \mid \mathbf{Z}_{t-1}, \boldsymbol{\beta}_2) \tag{5.11}$$

Weak exogeneity constitute the basis for the definition of a third concept of exogeneity: super-exogeneity. Superexogeneity requires weak exogeneity and that the conditional model $D(\mathbf{y}_t \mid \mathbf{z}_t, \mathbf{Y}_{t-1}, \mathbf{Z}_{t-1}, \boldsymbol{\beta}_1)$is structurally invariant, i.e. changes in the ditribution of the marginal model for $\mathbf{z}_t$ do not affect the $\boldsymbol{\beta}_1$ parameters.

These three concept are useful to define the validity of the reduction from the data congruent reduced form and the adopted structural model.

If the objective of the analysis is inference on the $\boldsymbol{\beta}_1$ parameters, then the joint-density can be reduced to a conditional model if $\mathbf{z}_t$ is weakly exogenous for the estimation of the parameters of interest.

If the objective of the analysis is dynamic simulation, then the joint-density can be reduced to a conditional model if $\mathbf{z}_t$ satisfies the conditions for strong exogeneity

If the objective of the analysis is econometric policy evaluation, then the joint-density can be reduced to a conditional model if $\mathbf{z}_t$ satisfies the conditions for super-exogeneity.

Tests for the validity of all these three concepts have been devoleped to sustain the validity of the last stage of the reduction processes.

5.8.2 *Exogeneity in ECM representations*

To illustrate how the concepts of exogeneity are applied to linear dynamic models, consider the following DGP:

$$\begin{aligned} y_t &= a0_{12} z_t + \varepsilon_{1t} \\ \varepsilon_{1t} &= \rho \varepsilon_{1t-1} + u_{1t} \quad 0 < \rho < 1 \\ a0_{21} y_t + a0_{22} z_t &= a1_{21} y_{t-1} + a1_{22} z_{t-1} + \varepsilon_{2t} \\ \varepsilon_{2t} &= \varepsilon_{2t-1} + u_{2t} \end{aligned} \tag{5.12}$$

$$\begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} N.I.D. \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix} \right]$$

This is a non-stationary process, integrated of the first order, admitting one cointegrating vector. The non-stationarity of the process stems from the presence of a unit root in ε_{1t},while the cointegrating vector is defined by $y_t - a_{12} z_t$, as ε_{1t} is stationary. The system([?]) can be re-parameterised as follows:

$$\mathbf{A}_0 \begin{pmatrix} \Delta y_t \\ \Delta z_t \end{pmatrix} = \mathbf{A}_1 \begin{pmatrix} \Delta y_{t-1} \\ \Delta z_{t-1} \end{pmatrix} + \mathbf{C} \begin{pmatrix} y_{t-1} \\ z_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \tag{5.13}$$

$$\begin{aligned} \mathbf{A}_0 &= \begin{pmatrix} 1 & -a0_{12} \\ a0_{21} & a0_{22} \end{pmatrix}, \mathbf{A}_1 = \begin{pmatrix} 0 & 0 \\ a1_{21} & a1_{22} \end{pmatrix}, \\ \mathbf{C} &= \begin{pmatrix} -(1-\rho) & a0_{12}(1-\rho) \\ 0 & 0 \end{pmatrix} \\ \mathbf{C} &= \mathbf{A}_0 \boldsymbol{\alpha} \boldsymbol{\beta}' \end{aligned}$$

Note that (6.11)can be considered a congruent representation of the DGP, as it features well-behaved residuals. This is not true of (5.12) which features autocorrelated residuals. Autocorrelation is generated by the omitted first order dynamics in the static equation. The omitted dynamics admit specific restrictions known as COMFAC (common factor restriction). [11]

To analyse the different concepts of exogeneity, consider the probabilistic structure of the data underlying model (6.11) .In other words, let us derive the reduced form associated with (6.11) :

[11] The common factor restriction is singular in that the effects of the omitted dynamic can be cured by a Cochrane-Orcutt estimator (static model+autocorrelated error terms).

$$\begin{pmatrix} \Delta y_t \\ \Delta z_t \end{pmatrix} = \begin{pmatrix} \frac{a0_{12}a1_{21}}{k} & \frac{a0_{12}a1_{22}}{k} \\ \frac{a1_{21}}{k} & \frac{a1_{22}}{k} \end{pmatrix} \begin{pmatrix} \Delta y_{t-1} \\ \Delta z_{t-1} \end{pmatrix} + \tag{5.14}$$

$$+ \begin{pmatrix} -\frac{a0_{22}(1-\rho)}{k} \\ \frac{a0_{21}(1-\rho)}{k} \end{pmatrix} \begin{pmatrix} 1 & -a0_{12} \end{pmatrix} \begin{pmatrix} y_{t-1} \\ z_{t-1} \end{pmatrix} + \tag{5.15}$$

$$+ \begin{pmatrix} \frac{a0_{22}}{k} & \frac{a0_{12}}{k} \\ \frac{-a0_{21}}{k} & \frac{1}{k} \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \tag{5.16}$$

$$K = \frac{1}{a0_{22} - a0_{21}a0_{12}}$$

From the reduced form we have that the conditional joint density of y_t and z_t can be written as follows:

$$\begin{pmatrix} \Delta y_t \\ \Delta z_t \end{pmatrix} \mid I_{t-1} \Big) \, N.I.D. \left[\begin{pmatrix} \mu_y \\ \mu_z \end{pmatrix}, \begin{pmatrix} \sigma_{yy} & \sigma_{yz} \\ \sigma_{yz} & \sigma_{zz} \end{pmatrix} \right] \tag{5.17}$$

where

$$\mu_y = \frac{a0_{12}a1_{21}}{k}\Delta y_{t-1} + \frac{a0_{12}a1_{22}}{k}\Delta z_{t-1} - \frac{a0_{22}(1-\rho)}{k}(y_{t-1} - a0_{12}z_{t-1})$$

$$\mu_z = \frac{a1_{21}}{k}\Delta y_{t-1} + \frac{a1_{22}}{k}\Delta z_{t-1} + \frac{a0_{21}(1-\rho)}{k}(y_{t-1} - a0_{12}z_{t-1})$$

$$\sigma_{yy} = \left(\frac{a0_{22}}{k}\right)^2 \sigma_{11} + \left(\frac{a0_{12}}{k}\right)^2 \sigma_{22}$$

$$\sigma_{zz} = \left(\frac{a0_{21}}{k}\right)^2 \sigma_{11} + \left(\frac{1}{k}\right)^2 \sigma_{22}$$

$$\sigma_{yz} = -\left(\frac{a0_{22}}{k}\right)\left(\frac{a0_{21}}{k}\right)\sigma_{11} + \left(\frac{a0_{12}}{k}\right)\left(\frac{1}{k}\right)\sigma_{22}$$

By applying the known properties of the multivariate normal, we derive from the statistical representation of the data the conditional mean of Δy_t with respect to Δz_t and I_{t-1} as follows:

$$E(\Delta y_t \mid \Delta z_t, I_{t-1}) = \mu_y + \frac{\sigma_{yz}}{\sigma_{zz}}(z - \mu_z) \tag{5.18}$$

z_t is said to be weakly exogenous for the estimation of the parameters of interest if the conditional mean for Δy_t derived from (5.18)coincides with the conditional mean for Δy_t derived from the first equation of model (6.11) . As the conditional mean from the first eqaution of (6.11) is :

$$E\left(\Delta y_t \mid \Delta z_t,\, I_{t-1}\right) = a0_{12}\Delta z_t - (1-\rho)\left(y_{t-1} - a0_{12}z_{t-1}\right) \tag{5.19}$$

Weak exogeneity of Δz_t for the estimation of the parameter $a0_{12}$ is obtained when $a0_{21} = 0$.

Strong exogeneity requires Granger non-causality in addition to weak exogeneity, strong exogeneity is satisfied when $a0_{21} = 0$, $a1_{21} = 0$.

Super-exogeneity requires weak-exogeneity and independence of the parameters of interest from the distribution of Δz_t. In our example, whenever the conditions for weak-exogeneity are satisfied, super-exogeneity also holds. To show a case in which this does not happen, consider the following modification of our DGP:

$$\begin{aligned} y_t &= a0_{12}E\left(z_{t+1} \mid I_t\right) + \varepsilon_{1t} \\ z_t &= a1_{22}z_{t-1} + \varepsilon_{2t} \end{aligned} \tag{5.20}$$

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} N.I.D. \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix}\right]$$

In this case the conditional mean $E\left(y_t \mid z_t,\, I_{t-1}\right)$ is given by the following expression:

$$E\left(y_t \mid z_t,\, I_{t-1}\right) = a0_{12}a1_{22}z_t \tag{5.21}$$

and it depends on $a1_{22}$, the parameter determining the conditional mean of z_t.

We conclude with the following remarks:

- exogeneity is defined independently form the parameters defining the cointegrating vectors, but it is related to the weigths. Weak exogeneity has a precise relation with the direction of adjustment in presence of disequilibria
- this is a special case, as we take a diagonal variance-covariance matrix. In general weak exogeneity requires a lower triangular structure and absence of correlation in the variance covariance-matrix.
- note the impossibility of reverse regression. The condition of weak exogeneity of Δz_t for the estimation of $a0_{12}$ are mutually exclusive with the condition of weak exogeneity of Δy_t for the estimation of $\frac{1}{a0_{12}}$.

5.8.3 *Testing exogeneity*

The preceeding example shows how weak-exogeneity can be tested for within the framework of cointegration. To provide a more general introduction to the issue of testing exogeneity consider a bivariate process for two generic variables y_t and z_t conditioned with respect to the information available, which includes all past history for the process:

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} \mid I_{t-1} \Big) N.I.D. \left[\begin{pmatrix} \mu_t^y \\ \mu_t^z \end{pmatrix}, \begin{pmatrix} \sigma_t^{yy} & \sigma_t^{yz} \\ \sigma_t^{yz} & \sigma_t^{zz} \end{pmatrix} \right] \tag{5.22}$$

the conditional model for y_t can be written as:

$$(y_t \mid z_t, I_{t-1})\, N.I.D \left[\frac{\sigma_t^{yz}}{\sigma_t^{zz}} (z_t - \mu_t^z) + \mu_t^y,\, \sigma_t^{yy} - \left(\frac{\sigma_t^{yz}}{\sigma_t^{zz}} \right) \right] \tag{5.23}$$

and the marginal model for z_t is instead:

$$(z_t \mid I_{t-1})\, N.I.D\, (\mu_t^z, \sigma_t^{zz}) \tag{5.24}$$

The parameters of interest feature the following relationship:

$$\mu_t^y = \beta \mu_t^z + \mathbf{w}_t' \boldsymbol{\delta} \tag{5.25}$$

of which, for the sake of exposition, we consider the special case:

$$y_t = \beta z_t + \mathbf{w}_{t-1}' \boldsymbol{\delta}_t + u_t \tag{5.26}$$

where $\mathbf{w}_{t-1}$is included in I_{t-1}.

Weak exogeneity of z_t for the estimation of β implies that this parameter could be estimated directly from (5.26)without any loss of relevent information. For this to happen, it is necessary that we have a sequential cut and that the conditional model does not depend on $\mu_t^z, \sigma_t^{yz}, \sigma_t^{zz}$. To pin down formally conditions for weak exogeneity substitute (5.25) in (5.23) obtaining:

$$(y_t \mid z_t, I_{t-1})\, N.I.D \left[\beta z_t + \mathbf{w}_{t-1}' \boldsymbol{\delta} + \left(\frac{\sigma_t^{yz}}{\sigma_t^{zz}} - \beta \right) (z_t - \mu_t^z),\, \sigma_t^{yy} - \left(\frac{\sigma_t^{yz}}{\sigma_t^{zz}} \right) \right] \tag{5.27}$$

Therefore we have weak exogeneity of z for the estimation of β, if the following condition is satisfied:

$$\frac{\sigma_t^{yz}}{\sigma_t^{zz}} = \beta$$

From this condition we can easily understand test for exogeneity available in the literature (Hausman(1978), Wu(1973)) and based on two-stages procedures.

In the first stage μ_t^z is parameterised by fitting a conditional model for z_t of the following type:

$$z_t = \mathbf{s}_t' \boldsymbol{\pi} + u_t \tag{5.28}$$

Where vector $\mathbf{s}$ includes all variables necessary to obtain a satisfactory specification for z_t. In the second stage the significance of residuals from (5.28) in equation (5.26) and the null of weak exogeneity coincide with the null of non-significance of such constructed variable. The argument here can be extended to test the null of super-exogeneity (Engle-Hendry(1993), Favero-Hendry(1992)). The alternative hypothesis is now complicated as follows:

$$\beta\left(\mu_t^z, \sigma_t^{zz}\right) = \beta_0 + \beta_1 \mu_t^z + \beta_2 \sigma_t^{zz} + \beta_3 \frac{\sigma_t^{zz}}{\mu_t^z} \tag{5.29}$$

and the null of interest is weak exogeneity augmented by $\beta_1 = \beta_2 = \beta_3 = 0$
To see ho the test is derived, substitute from (5.29) (5.25) in (5.23) obtaining:

$$\left(y_t \mid z_t,\, I_{t-1}\right) N.I.D\ \left(\mu_t, \Omega_t\right) \tag{5.30}$$

$\mu_t = \beta_0 z_t + \mathbf{w}_{t-1}' \boldsymbol{\delta} + \left(\frac{\sigma_t^{yz}}{\sigma_t^{zz}} - \beta_0\right)\left(z_t - \mu_t^z\right) + \beta_1 \left(\mu_t^z\right)^2 + \beta_2 \sigma_t^{zz} \mu_t^z + \beta_3 \sigma_t^{zz}$
$\Omega_t = \sigma_t^{yy} - \left(\frac{\sigma_t^{yz}}{\sigma_t^{zz}}\right)$
by using the first-order expansion:
$\frac{\sigma_t^{yz}}{\sigma_t^{zz}} = \delta_0 + \delta_1 \sigma_t^{zz}$
we reach the following estimable relation:
$y_t = \beta_0 z_t + \mathbf{w}_{t-1}' \boldsymbol{\delta} + \left(\delta_0 - \beta_0\right)\left(z_t - \mu_t^z\right) + \delta_1 \sigma_t^{zz}\left(z_t - \mu_t^z\right) + \beta_1 \left(\mu_t^z\right)^2 + \beta_2 \sigma_t^{zz} \mu_t^z + \beta_3 \sigma_t^{zz}$
where the null hypothesis of interest can now be empirically tested by parameterising the first two moments of the conditional model for z_t.

Hendry[?] provides an alternative assessment of superexogeneity by analysing the encompassing implications of feedback versus feedforward models.

This procedure is based on the explicit consideration of two alternative specifications for the DGP.

The feedback model, denoted H_b,is:

$$\begin{aligned} y_t &= \boldsymbol{\beta}' \mathbf{z}_t + v_t \\ E_b\left(\mathbf{z}_t v_t\right) &= 0 \end{aligned} \tag{5.31}$$

The feedforward model, denoted H_f,is:

$$
\begin{aligned}
y_t &= \boldsymbol{\delta}' E\left(\mathbf{z}_{t+1} \mid I_t\right) + \epsilon_t && (5.32)\\
\mathbf{z}_t &= \boldsymbol{\gamma}_t \mathbf{z}_{t-1} + \mathbf{u}_t && (5.33)\\
E_f\left(\mathbf{z}_t \epsilon_t\right) &= 0\\
\mathbf{u}_t &\sim i.d.\left(0, \Omega_t\right)
\end{aligned}
$$

Note that in H_f the parameters of the marginal model for $\mathbf{z}_t$ are function of time through γ_t and Ω_t. Moreover, we restrict ourselves to the case in which the only relevant information in I_t to predict $\mathbf{z}_{t+1}$are the realizations at time t of $\mathbf{z}$.

We can now explore the encompassing predictions of each model for the other. We do so by evaluating the performance of each model when the congruent representation of the DGP is the alternative model.

When a (5.31) and (5.28) are a congruent representation of the DGP, the following implications hold.

(1) When γ_t and Ω_t are non-constant, also the projection of y_t on $\mathbf{z}_{t-1}$ is non-constant, in fact

$$E_b\left(y_t \mid \mathbf{z}_{t-1}\right) = \boldsymbol{\beta}' \boldsymbol{\gamma}_t \mathbf{z}_{t-1}$$

(2) the error-variance is also non constant:

$$
\begin{aligned}
y_t - E_b\left(y_t \mid \mathbf{z}_{t-1}\right) &= \boldsymbol{\beta}' \mathbf{u}_t + v_t = \varphi_t\\
E_b\left(\varphi_t^2\right) &= \sigma_v^2 + \boldsymbol{\beta}' \Omega_t \boldsymbol{\beta}
\end{aligned}
$$

(3) the projection of y_t on $\mathbf{z}_{t-1}$should fit worse than the behavioural model [?], in fact

$$E_b\left(\varphi_t^2\right) = \sigma_v^2 + \boldsymbol{\beta}' \Omega_t \boldsymbol{\beta} > \sigma_v^2$$

(4) the behavioural model ([?]) should feature constant parameters

When instead (5.32) and (5.28) are a congruent representation of the DGP, the following implications hold.

(1) the conditional model cannot be constant when the marginal model for $\mathbf{z}_t$ is sufficiently variable since:

$$E_f\left(y_t \mid \mathbf{z}_t\right) = \boldsymbol{\delta}' \boldsymbol{\gamma}_t \mathbf{z}_{t-1}$$

(2) the projection of y_t on $\mathbf{z}_{t-1}$ is non constant but with parameter vector $\boldsymbol{\delta}' \boldsymbol{\gamma}_t \boldsymbol{\gamma}_{t+1}$

(3) no variance ranking is possible as:

$$y_t - E_f\left(y_t \mid \mathbf{z}_t\right) = \epsilon_t$$

as in (5.32).

The analysis of the encompassing implications of the two cases reveals that when the feedback model is stable and the marginal process is not stable, then the feedforward specification cannot be a congruent representation of the DGP. As a consequence, the relevance of the Lucas critique could be analysed by assessing simultaneously the stability of the feedback structural model and the stability of the marginal models for the regressors in the feedback model.

This procedure deserves some discussion.

A first observations is related to the power of tests for structural stability, in fact for the procedure to work it is essential that the marginal model is sufficiently variable and that such variability is detectable through tests for parameters' stability. As we have already seen, the issue is not trivial in that multiple breaks at unknown points are not so easily detected.

Setting aside the power of the tests, there is a logical issue related to the reduction procedure. In fact, if parameters stability is taken as one of the criteria for congruency, then congruent reduced forms should never feature parameters instability. In practice, as we have seen in our application, congruent specifications often need the inclusion of dummies. Therefore, the significance of the same dummies in different equations of the adopted model could be exploited to apply the procedure for the evaluation of the relevance of the Lucas critique.

A related question refers to the power of the procedure in the case of limited information, i.e. the case in which the parameters instability is generated by omitted variables in the marginal models. Hendry[**?**] considers explicitly such case, by adopting the following alternative specification for the marginal model:

$$\mathbf{z}_t = \boldsymbol{\gamma}_1 \mathbf{z}_{t-1} + \boldsymbol{\gamma}_2 \mathbf{s}_{t-1} + \mathbf{u}_{2t} \tag{5.34}$$

$$\mathbf{u}_{2t} \sim i.d.\left(0, \Omega_2\right) \tag{5.35}$$

when (5.34) is a congruent representation of the DGP, (5.28) features instability because of a limited information problem: the omission of $\mathbf{s}_{t-1}$ from the relevant information set. However, if(5.28) is observed, then the relation between $\mathbf{z}_t$ and $\mathbf{s}_t$ cannot be constant, in fact it must be the case that:

$$\mathbf{x}_t = \boldsymbol{\mu}_t \mathbf{z}_t + \boldsymbol{\xi}_t$$
$$\boldsymbol{\xi}_t \sim i.d.\left(0, \Xi_t\right)$$

and then

$$\boldsymbol{\gamma}_t = \boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_2 \boldsymbol{\mu}_t$$
$$\Omega_t = \Omega_2 + \boldsymbol{\gamma}_2 \Xi_t \boldsymbol{\gamma}_2'$$

and the result of stability of the feedback model paired with the instability of the (mis-specified) marginal model can still rule out the congruency of the feedforward model.

5.9 A model of the monetary transmission mechanism

To illustrate the specification of a structural model for the monetary transmission mechanism we consider as a baseline the cointegrated reduced form discussed in one of the previous sections:

$$\begin{bmatrix} \Delta y_t \\ \Delta \pi_t \\ \Delta R_t^b \end{bmatrix} = \begin{bmatrix} d_{0,11} \\ d_{0,21} \\ d_{0,31} \end{bmatrix} + i = 0\sum^{6} F_i \Delta_{12} LPCM_{t-i} + \mathbf{g}'\mathbf{DUM}_t + \tag{5.36}$$

$$i = 1\sum^{5} D_i \begin{bmatrix} \Delta y_{t-i} \\ \Delta \pi_{t-i} \\ \Delta R_{t-i}^b \end{bmatrix} + \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \\ \alpha_{31} \end{bmatrix} \left(R_{t-1}^b - R_{t-1}^{b*} \right) + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}$$

$$R_{t-1}^{b*} = \pi_{t-1} + 0.22 y_{t-1} - 0.08t$$

Note that (5.36) is the result of the reduction of the baseline model, which contains five equations. The original model delivered two cointegrating relationships, we have identified the first one as an interest rate reaction function and the second one as a rule for determining the interest rate on bank deposits. To describe the monetary transmission mechanism under interest rate targeting we need to supplement the interest rate reaction function with equations for the target variables, inflation and output. Real money, being demand determined, looses interest and so does the opportunity cost of holding money. Therefore we have omitted from the original model the two equations determining real money and the interest rate on bank deposits together with the equilibrium relationships for the interest rate on bank deposits. The validity of this step in the reduction process is testable. Congruency of our selected specification requires that the weights on the second cointegrating vector can be constrained to zero in our three maintained equations, that the weights on the first cointegrating vector can be constrained to zero in the equations for real money and the interest rate on bank deposits, and finally that lagged value of real money and the interest rate on bank deposits do not enter significantly system (5.36). Having asserted the validity of this further step in reduction, we proceed to the specification of the following structural model:

$$\Delta y_t = (0.03)0.23 + (0.06)0.33 \Delta y_{t-1} - (0.09)0.21 \Delta R_{t-3}^b - (0.46)1.19 DUM7312 - (0.46)1.11 DUM7308 \tag{5.37}$$

$$\begin{aligned}\Delta\pi_t = {} & (0.02){-0.05} + (0.03)0.08\Delta y_{t-1} + (0.03)0.06\Delta y_{t-5} + (0.05)0.2\Delta\pi_{t-5} + (0.05)0.14\Delta\pi_{t-6} \\ & +(0.005)0.03\Delta_{12}LPCM_{t-1} - (0.006)0.02\Delta_{12}LPCM_{t-2} - (0.27)0.92DUM7307 \\ & +(0.26)0.87DUM7308 + (0.26)0.69DUM7407 - (0.25)0.56DUM7408 + (0.25)1.04DUM7409 \\ & -(0.25)0.75DUM7505\end{aligned} \quad (5.38)$$

$$\begin{aligned}\Delta R_t^b = {} & -(3.08)8.34 + (0.13)0.31\Delta\pi_t + (0.03)0.08\Delta y_{t-3} + (0.06)0.36\Delta R_{t-1}^b + (0.006)0.01\Delta_{12}LPCM_t \\ & -(0.005)0.013\Delta_{12}LPCM_{t-1} - (0.018)0.04ECM_{t-1} + (0.24)0.60DUM7306 + (0.28)0.71DUM73 \\ & -(0.24)0.90DUM7310 + (0.25)1.06DUM7311 - (0.24)0.76DUM7312 - (0.24)0.86DUM7402 \\ & +(0.24)1.13DUM7403 + (0.25)1.64DUM7408 - (0.28)1.72DUM7409 - (0.24)0.72DUM7501 \\ & +(0.23)0.49DUM7808 + (0.28)0.57DUM7811\end{aligned} \quad (5.3$$

LR test of over-identifying restrictions: $\chi^2(89) = 95.3354\ [0.3037]$.

The model is estimated over the sample 1961:2 1979:8 by FIML. The 89 over-identifying restrictions imposed by the reduced form implicit in our structure on the unconstrained reduced form are not rejected. The first equation can be interpreted as an aggregate demand equation along which the output gap (deviation of output form a stochastic trend) depends on lagged change in nominal interest rates. The second equation is stylised aggregate supply which determines inflation as a function of past inflation, the commodity price inflation and the output gaps. Finally, the third equation is an interest rate reaction function which describes short-run dynamics around a long-run solution determined by response of interest rates to inflation and output. Note that, because of the dynamic specification, the response of the monetary instruments to fluctuations in the target variables is different in the short.run and in the long-run. To illustrate the within sample performance of the model we report actual and fitted values in Figure5.3

5.9.1 *Simulating monetary policy*

We are now in the position of simulating monetary policy. We simulate the impact of an hundred basis point exogenous monetary policy shock by computing dynamic multipliers. The baseline model is obtained by simulating dynamically , for given values of the exogenous variables, the three endogenous variables are generated by equations (5.38) , (5.37) , (5.39) over the sample considered for estimation. The perturbed solution is obtained by adding an exogenous one-off 100 basis point shock to equation (5.39) in the first period of the simulation. Obviously a one-off hundred basis point shock to the first difference of the policy rates is a permanent one-hundred basis points shock to the level of policy rates. Dynamic multipliers are then computed by adapting the E-Views procedures already discussed in chapter four. All computations are available in the file LSE.WF1. We report dynamic multipliers in Figure 5.4

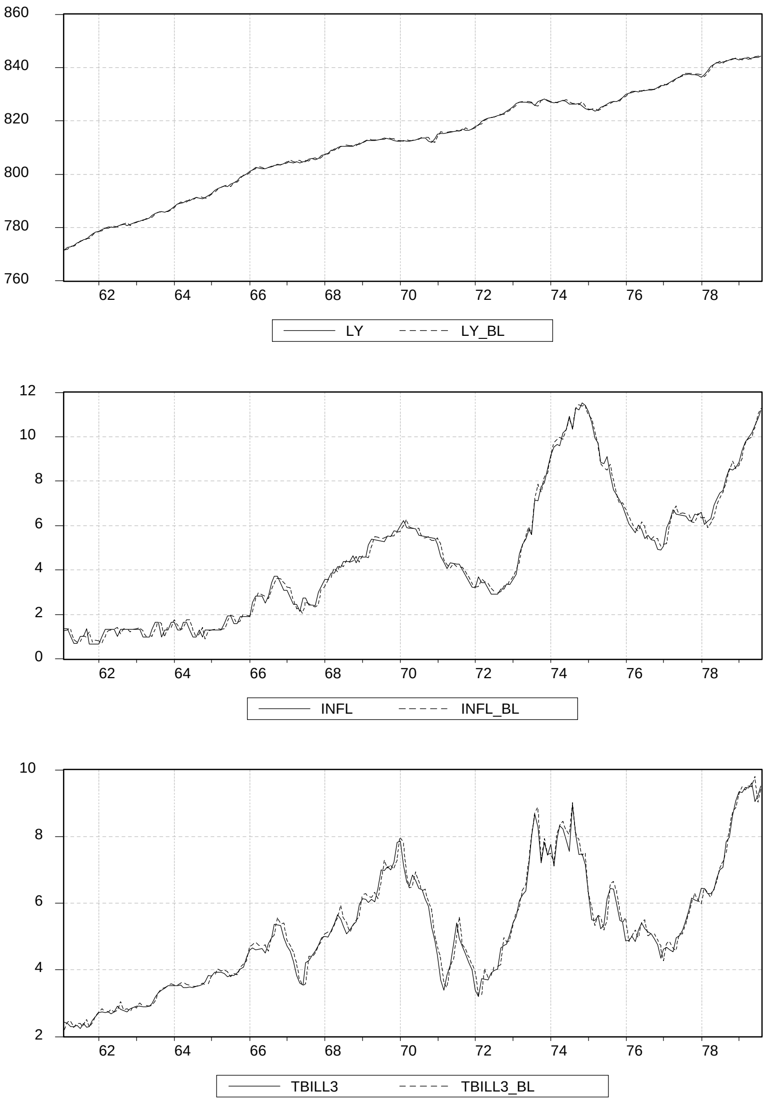


FIG. 5.3. Actual and fitted values from the structural model

Dynamic multipliers confirm the stability of the model and reveal a much stronger impact of monetary policy on outptut fluctuations than on inflation. Note also that the pattern of multipliers is much smoother than the corresponding pattern for the model used to illustrate the Cowles Commission strategy. Such smoothness is a consequence of the better dynamic specification of the

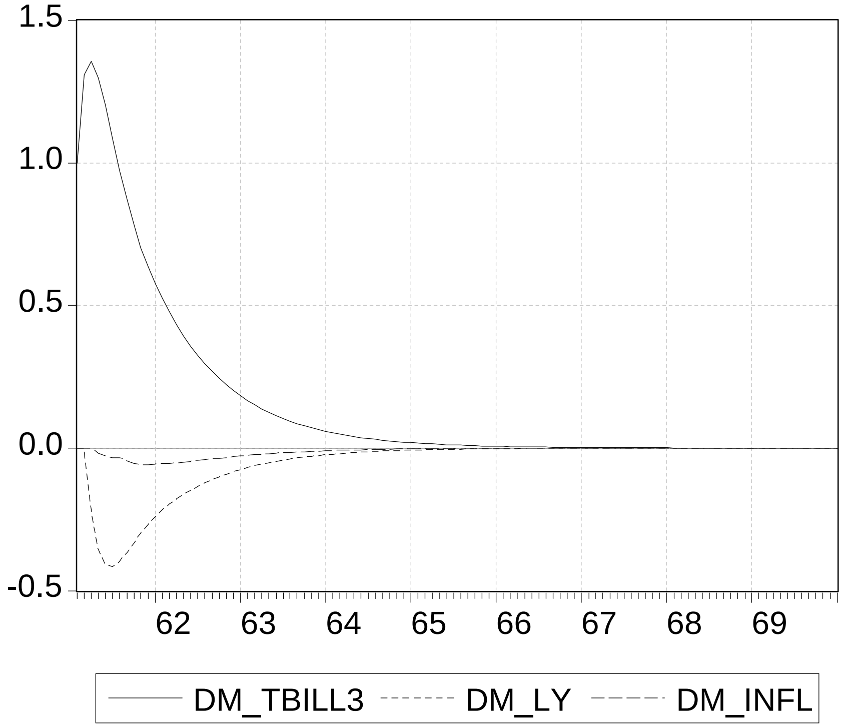


FIG. 5.4. Dynamic multipliers

LSE model.

5.9.2 *Model Evaluation*

To complete our comparative evaluation of the LSE and the Cowles commission specifications we have still considered out-of -sample evaluation, where the performance of the Cowles commission specification was at its worst. We simulate the LSE model dynamically over the period 1985:01- 1996:03. In doing so we deliberately skip the period of non-borrowed reserves targeting where our specification is clearly not appropriate. The dynamically simulated series are reported with the actual series in Figure 5.5.

Figure 5.5 shows an improved performance with respect to the Cowles commission specification but clearly there are problems in the out-of-sample simulation. The implementation of diagnostic tests guarantees the quality of the within sample results but cannot ensure against structural shifts in parameters: congruent models within sample might perform very poorly in out-of-sample simulations.

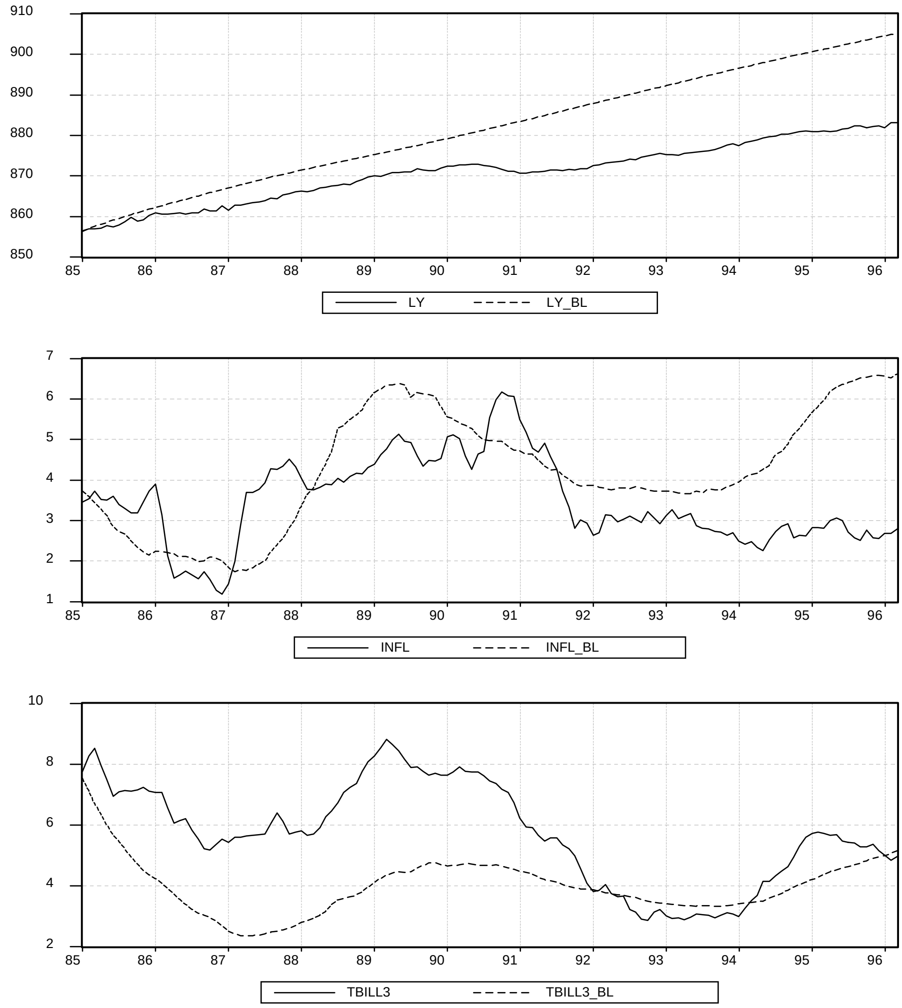


FIG. 5.5. Out-of-sample dynamic simulation

5.9.3 *Testing the Lucas critique*

Our simple model of the monetary transmission mechanism offers an opportunity to implement empirically tests for super-exogeneity.

In our discussion of identification in chapter 3 we have shown that when a central bank faces the following intertemporal optimisation problem:

$$Minimize \quad E_t i = 0 \sum^{\infty} \delta^i L_{t+i} \tag{5.40}$$

where:

$$L = \frac{1}{2}\left[(\pi_t - \pi^*)^2 + \lambda x_t^2\right] \tag{5.41}$$

under the constraints of the following specification for aggregate supply and demand in a closed economy:

$$x_{t+1} = \beta_x x_t - \beta_r \left(i_t - E_t \pi_{t+1} - \bar{r}\right) + u_{t+1}^d \tag{5.42}$$

$$\pi_{t+1} = \pi_t + \alpha_x x_t + u_{t+1}^s \tag{5.43}$$

the optimal interest rate rule can be written as:

$$\begin{aligned} i_t = \bar{r} + \pi^* + \left(\frac{1+\alpha_x\beta_r}{\alpha_x\beta_r}\right)(E_t\pi_{t+1} - \pi^*) + \\ + \frac{\beta_x}{\beta_r} x_t + \frac{\lambda}{\delta\alpha_x k}\frac{1}{\alpha_x\beta_r} E_t x_{t+1}. \end{aligned} \tag{5.44}$$

where k is a combination of parameters describing the structure of the economy and the preferences of the central banker. We have then an intertemporal optimization framework which offer a feedforward monetary rule, in which output gap and inflation are not superexogenous for the estimation of the parameters of interest. Contrast this specification with equation (5.39) in our model. Our estimated eqaution is a feedback specification, which does not include explicitly expectations and whose parameters are estimated independently from those in the aggregate demand and supply schedules. Therefore we have a natural candidate to test the validity of the Lucas critique. Tests of feedforward versus feedback model are difficult to apply in that we have designed our model to pass diagnostic tests, however note that the reaction function and the aggregate supply and demand equation contain a common set of dummies. This a clear indication of common outliers in the three equations which does not refute the hypothesis of validity of the feedforward interpretation. The presence of dummies shall also impact on the Engle-Hendry superexogeneity tests. This test is applicable by exploiting the specification of supply and demand equations to derive proxies for the first two moments in the conditional model for these two variables and then by adding them to the interest rate reaction function. The impact of the dummies on the test is determined by the fact that they capture some portion of the variability in the additional regressors on which joint significant is tested.

5.10 What have we learned?

In our opinion, the major strengths of the LSE methodology are related to a careful diagnosis of the problems of the Cowles Commission approach and to the attept of giving "scientific dignity" to the specification of dynamic econometric models. The concept of cointegration fits naturally in the context of dynamic specification of ECM models. Such research strategy is based on a multi-step framework: specification of the VAR and its deterministic component, identification of the number of cointegrating vectors, identification of the parameters in cointegrating vectors, tests on the speed of adjustment with respect to disequilibria. The results of the final test depend on the outcome of the previous stages in the empirical analysis, but the outcome of each step is not so easily and uniquely established empirically. The reduction process has been criticised by macroeconomist for its tendency to deliver preferred specification "...a bit over-cooked..." and to loosen considerably the link between econometric model and economic theory. Consider the following money demand specification, taken from Baba, Hendry, Starr[**?**], as a typical LSE model:

$$\begin{aligned}\Delta(m-p)_t = {} & [0.097]{-0.334}\Delta_4(m-p)_{t-1} - [0.039]0.156\Delta^2(m-p)_{t-4} - [0.015]0.249\left(m-p-\frac{1}{2}y\right)_{t-2} (5.45)\\ & -[0.046]0.33\Delta\hat{p} - [0.132]1.097\Delta_4 p_{t-1} + [0.079]0.859V_t + [1.49]11.68\Delta SV_{t-1}\\ & -[0.104]1.409AS_t - [0.063]0.973AR_{1t} - [0.049]0.255\Delta R_{ma,t} + [0.055]0.435R_{nsa,t}\\ & +[0.07]0.395\Delta Ay_t + [0.003]0.013D_t + [0.02]0.352 + \hat{u}_t\end{aligned}$$

where heteroscedasticity consistent estimators are reported in brackets.The BHS specification for U.S. money demand, is estimated on quarterly data covering the period 1960-1988. m is the log of M1; y is the log of real GNP using 1982 as base year; p is the log of the deflator; Δ^2 is the square of the difference operator Δ; $\Delta\hat{p} = \Delta(1+\Delta)p_t$; $\Delta_4(m-p)_{t-1} = 0.25\left((m-p)_{t-1} - (m-p)_{t-5}\right)$; V_t is a nine-quarter moving-average of quarterly averages of twelve-month moving standard deviations of 20-year bond yields; $SV_t = \max(0, S_t) * V_t$ where S is the spread between the 20-year Treasury bond yield and the coupon equivalent yield on a one-month TBill; $AS_t = 0.5(S_t + S_{t-1})$; AR_{1t} is a two-quarter moving-average of the one-month T-bill yield; $R_{ma,t}$ is the maximum of a passbook savings rate, an weighted certificate of deposit rate and a weighted money market mutual fund rate; $R_{nsa,t}$ is the average of weighted NOW and SuperNow rates; $Ay_t = 0.5(y_t + y_{t-1})$and D_t is a credit control dummy which is -1 in 1980(2), 1 in 1980(3), and zero everywhere else. BHS report 11 diagnostics, all passed.

The achievement of data congruency implies some evident cost in terms of parsimony of the specification and economic interpretability of the results.equation **??** also illustrates why the LSE methodology is not easily applied to system of equations, even of very limited dimensions. General-to-specific methodology is usually applied in single-equation specification (money demand and consumption of non-durables functions are the preferred application of the LSE approach), applications to system with few variables are reported in the literature but it becomes very hard and very rare to apply such methodology when the dimension of the system exceeds a small dimension (say five equtions). Moreover, we have seen that the applicability of the concept of cointegration is very rapidly complicated as n increases in n-variate systems.

Faust and Whiteman[23] note that **??**is a much richer specification than the one implicitly contained in the standard VAR approach, including moving averages and moving standard deviations of interest rates. I am somewhat skeptical that such specification could be produced by a VAR approach to cointegration, or by any VAR analysis. Criticism of the use of this generated variables are mainly based on the argument that, by constuction, they capture within- sample fluctuations in the data and their peformance out-of-sample worsens considerably. Moreover such transformations, being data instigated, are usually related to theory with some difficulty. Many applied macroeconometricians feel not at ease in using variables which perform well empirically but whose links with theory are not so clear. Of course, for the LSE methodology, this is a problem with the profession rather than with the econometric methodology. In fact, this is probably the centre of the debate.

REFERENCES

Andrews D.W.(1993) “Tests for parameter instability and structural change with unknown change point”, *Econometrica*, 61, 821-853

Baba Y., Hendry D.F. and Starr R.M.(1992) ”The demand for M1 in the USA,1960-1988” Review of Economic Studies, 59, 25-61

Chow G.C.(1960) ”Tests of equality between sets of coeffcients in two linear regressions”, Econometrica, 28, 591-605

Doornik J. and Hendry D.F.(1997)”PcFIML 9.0. Interactive econometric modelling of dynamic systems”,Thomson Publishing, Oxford

Doornik J., and Hansen H.(1994) ”A practical test of multivariate normality”, Nuffield College, Oxford

Engle R.F.(1982) ”Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflations” Econometrica, 50, 987-1007

Faust J. and C.H.Whiteman(1997) ”General-to-specific procedures for fitting a data-admissible, theory-inspired, congruent, parsimonious, encompassing, weakly-exogenous, identified, structural model to the DGP: a translation and a critique.” Board of Governors of the Federal Reserve System, International Finance Discussion Papers, 576

Godfrey L.G.(1988) Misspecification tests in Econometrics,Cambridge, Cambridge University Press

Hendry D.F. (1995) Dynamic Econometrics, Oxford University Press

Hendry D.F., Neale A. and F.Srba(1988) ”Econometric analysis of small linear systems using PC-FIML” Journal of Econometrics, 38, 203-226

Johansen S., Nielsen B.G.(1993) Asymptotics for the cointegration rank test tests in the presence of intervantion dummies, Manual for the simulation program DisCo, manual and program are available at the URL http://www.nuff.ox.ac.uk/users/nielsen/d

Kiviet J.F.(1986) ”On the rigour of some mis-specification tests for modelling dynamic relationships, Review of Economic Studies, 53, 241-261

Osterwald-Lenum M.(1992) ”A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics”, Oxford Bulletin of Economics and Statistics, 54, 461-472

Pagan A.(1984) ”Model evaluation by variable addition”,Ch.5 of Hendry D.F. and K.F.Wallis (eds.) Econometrics and Quantitative Economics, Basil Blackwell, Oxford

Spanos A.(1990) ”The simultaneous-equations model revisited. Statistical adequacy and identification” Journal of Econometrics, 44, 87-105

Walsh C.(1998) ”Monetary theaory and policy”,

White H.(1980) ”A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity”, Econometrica, 48, 817-838

6

THE VAR APPROACH

6.1 Introduction: why VAR models ?

The LSE methodology has intepreted the failure of the traditional Cowles Commission approach, heralded by the critiques due to Lucas [36] and Sims [48], as the result of the use of mis-specified and ill-identified models. The LSE methodology however does not question the potential of macroeconometric modelling for simulation and econometric policy evaluation. In fact, at the stage of simulation and policy evaluation, there no difference between the traditional Cowles commission approach and the LSE approach. The LSE solution to the problems of traditional macroeconometric modelling is concentrated on the stages of identification and specification. The importance of estimation is de-emphasized, in that congruency of the specification is considered as a much higher priority than the choice of the most appropriate estimator. No innovation is proposed at the stage of the simulation and policy evaluation: the traditional methods are applied, after having tested, tested, and tested.

The VAR approach shares with the LSE approach the diagnosis of the problem of Cowles Commission models but also questions the potential of traditional macroeconometric modelling for policy simulation and econometric policy evaluation. VAR models of the monetary transmission mechanism differ from structural LSE models as to the purpose of their specification and estimation. In the traditional approach the typical question asked within a macroeconometric framework is "What is the optimal response by the monetary authority to movement in macroeconomic variables in order to achieve given targets for the same variables?". The VAR approach recognizes fully the potential of the Lucas'critique and acknowledges that questions like "How should a central bank respond to shocks in macroeconomic variables?" are to be answered within the framework of quantitative monetary general equilibrium models of the business cycle. So the answer has to be based on a theoretical model rather than on an empirical ad-hoc macroeconometric model. Within this framework there is a new role for empirical analysis, i.e. to provide evidence on the stylized facts to be included in the theoretical model adopted for policy analysis and to decide between competing general equilibrium monetary models. The operationalization of this research programme is very well described in a recent paper by Christiano, Eichenbaum and Evans [14]. Three are the relevant steps:

- monetary policy shocks are identified in actual economies

- the response of relevant economic variables to monetary shocks is then described
- the same experiment is then performed in the model economies to compare actual and model-based responses as an evaluation tool and a selection criterion for theoretical models.

LSE-type structural models and VAR models of the monetary transmission mechanism have a common structure which, using the notation of Chapter 3, can be represented as follows:

$$\mathbf{A}\begin{pmatrix}\mathbf{Y}_t\\ \mathbf{M}_t\end{pmatrix} = \mathbf{C}(L)\begin{pmatrix}\mathbf{Y}_{t-1}\\ \mathbf{M}_{t-1}\end{pmatrix} + \mathbf{B}\begin{pmatrix}\boldsymbol{\nu}_t^Y\\ \boldsymbol{\nu}_t^M\end{pmatrix} \tag{6.1}$$

where $\mathbf{Y}$ and $\mathbf{M}$ are vectors of macroeconomic (non-policy) variables (e.g. output and prices) and variables controlled by the monetary policymaker (e.g. interest rates and monetary aggregates containing information on monetary policy actions) respectively. Matrix $\mathbf{A}$ describes the contemporaneous relations among the variables and $\mathbf{C}(L)$ is a matrix finite-order lag polinomial. $\boldsymbol{\nu} \equiv \begin{pmatrix}\boldsymbol{\nu}^Y\\ \boldsymbol{\nu}^M\end{pmatrix}$ is a vector of structural disturbances to the non-policy and policy variables; non-zero off-diagonal elements of $\mathbf{B}$ allow some shocks to affect directly more than one endogenous variable in the system. The main difference between the two approaches lies in the aim for which models are estimated.

Traditional Cowles Commission structural models are designed to identify the impact of policy variables on macroeconomic quantities in order to determine the value to be assigned to the monetary instruments ($\mathbf{M}$) to achieve a given target for the macroeconomic variables ($\mathbf{Y}$), assuming exogeneity of the policy variables in $\mathbf{M}$ on the ground that these are the instruments controlled by the policymaker. Identification in traditional structural models is obtained without assuming the orthogonality of structural disturbances. Dynamic multipliers are used to describe the impact of monetary policy variables on macroeconomic quantities. In the computation of dynamic multipliers the responses of macroeconomic variables to monetary policy can be obtained without decomposing monetary policy into its endogenous and exogenous components, and, in fact, in most traditional empirical applications such decomposition is not implemented.

The assumed exogeneity of the monetary variables in the traditional approach makes the model invalid for policy analysis if monetary policy reacts endogenously to macroeconomic variables. The LSE methodology would recognise the problem of the invalid exogeneity assumption for $\mathbf{M}$,it would then proceed to the identification of an alternative enlarged model (presumably such identification will be obtained through the imposition on a-priori restrictions on the dynamics of the lagged variables). However, the new model would be still used for simulation and econometric policy evaluation, whenever the appropriate concept of exogeneity (respectively strong and super) where satisfied by the adopted specification.

VAR modelling would reject the Cowles Commision identifying restrictions as "incredible" for reason not very different from the ones pinned down by the LSE approach, however VAR models of the transmission mechanism are not estimated to yield advice on the best monetary policy. They are rather estimated to provide empirical evidence on the response of macroeconomic variables to monetary policy impulses in order to discriminate between alternative theoretical models of the economy. It becomes then crucial to identify monetary policy actions using restrictions independent from the competing models of the transmission mechanism under empirical investigation, taking into account the potential endogeneity of policy instruments.

In a series of recent papers, Christiano, Eichenbaum and Evans [12], [13] apply the VAR approach to derive "stylized facts" on the effect of a contractionary policy shock, and conclude that plausible models of the monetary transmission mechanism should be consistent at least with the following evidence on price, output and interest rates: *(i)* the aggregate price level initially responds very little; *(ii)* interest rates initially rise, and *(iii)* aggregate output initially falls, with a j-shaped response, with a zero long-run effect of the monetary impulse. Such evidence leads to the dismissal of traditional real business cycle model, which are not compatible with the liquidity effect of monetary policy on interest rates, and of the Lucas [35] model of money, in which the effect of monetary policy on output depends on price misperceptions. The evidence seems to be more in line with alternative intepretations of the monetary transmission mechanism based on sticky prices models (Goodfriend and King [26]), limited participation models (Christiano and Eichenbaum [11]) or models with indeterminacy-sunspot equilibria (Farmer [21]).

Having stated the objective of VAR models we are now in the position of assessing how identification, estimation and simulation are implemented to analyse the monetary transmission mechanism.

VAR models concentrate on shocks.

First the relevant shocks are identified, the response of the system to shocks is described by analyzing impulse responses(the propagation mechanism of the shocks), forecasting error variance decomposition, and historical decomposition.

6.2 Identification and Estimation

We have introduced the identification problem for VAR in Chapter 3. Given the representation of the general structural model of interest :

$$\mathbf{A}\begin{pmatrix}\mathbf{Y}_t\\ \mathbf{M}_t\end{pmatrix}=\mathbf{C}(L)\begin{pmatrix}\mathbf{Y}_{t-1}\\ \mathbf{M}_{t-1}\end{pmatrix}+\mathbf{B}\begin{pmatrix}\boldsymbol{\nu}_t^Y\\ \boldsymbol{\nu}_t^M\end{pmatrix} \tag{6.2}$$

The structural model (6.2) is not directly observable, however a VAR can be estimated as the reduced form of the underlying structural model :

$$\begin{pmatrix} \mathbf{Y}_t \\ \mathbf{M}_t \end{pmatrix} = \mathbf{A}^{-1}\mathbf{C}(L) \begin{pmatrix} \mathbf{Y}_{t-1} \\ \mathbf{M}_{t-1} \end{pmatrix} + \begin{pmatrix} \mathbf{u}_t^Y \\ \mathbf{u}_t^M \end{pmatrix} \tag{6.3}$$

where $\mathbf{u}$ denotes the VAR residual vector, normally independently distributed with full variance-covariance matrix Σ. The relation between the VAR residuals in $\mathbf{u}$ and the structural disturbances in $\boldsymbol{\nu}$ is therefore:

$$\mathbf{A} \begin{pmatrix} \mathbf{u}_t^Y \\ \mathbf{u}_t^M \end{pmatrix} = \mathbf{B} \begin{pmatrix} \boldsymbol{\nu}_t^Y \\ \boldsymbol{\nu}_t^M \end{pmatrix} \tag{6.4}$$

undoing the partitioning we have

$$\mathbf{u}_{t=}\mathbf{A}^{-1}\mathbf{B}\boldsymbol{v}_t$$

from which we can derive the relation between the variance-covariance matrix of $\mathbf{u}_t$ (observed) and the variance-covariance matrix of $\boldsymbol{\nu}_t$ (not observed) as follows:

$$E\left(\mathbf{u}_t\mathbf{u}_t'\right) = \mathbf{A}^{-1}\mathbf{B}E\left(\boldsymbol{v}_t\boldsymbol{v}_t'\right)\mathbf{B}'\mathbf{A}^{-1}$$

Substituting population moments with sample moments we have:

$$\hat{\sum} = \hat{\mathbf{A}}^{-1} \mathbf{BI} \, \hat{\mathbf{B}}' \hat{\mathbf{A}}^{-1} \tag{6.5}$$

$\hat{\sum}$ contains $\frac{n(n+1)}{2}$ different elements, this is the maximum number of identifiable parameters in matrices $\mathbf{A}$ and $\mathbf{B}$. Therefore, a necessary condition for identification is that the maximum number of parameters contained in the two matrices is $\frac{n(n+1)}{2}$, such condition makes the number of equations equal to the number of unknowns in system 6.5. As usual, for such condition to be also sufficient for identification it also needed that no equations in 6.5 is a linear combinations of any of the other equations in the system, Amisano-Giannini[1], Hamilton[30]. As for traditional models we have the three possible cases of under-identification, just-identification and over-identification. As for traditional models, the validity of over-identifying restrictions can be tested via a statistic distributed as a χ^2 with a number of degrees of freedom equal to the number of over-identifying restrictions Amisano-Giannini[1]. Once identification has been achieved, the estimation problem is solved by applying Generalised Method of Moments estimation. We shall describe this class of estimators in the next chapter.

In practice , identification requires the imposition of some restrictions on the parameters of the $\mathbf{A}$ and $\mathbf{B}$. This step has been historically implemented in a number of different ways, we concentrate on the most widely used strategies in the next subsections

6.2.1 *Choleski Decomposition*

In the famous article which introduced VAR methodology to the profession, Sims[48] proposed the following identification strategy, based on the Choleski decomposition of matrices:

$$\mathbf{A} = \begin{pmatrix} 1 & 0\,0 & & 0 \\ a_{21} & 1\,0 & & 0 \\ . & .\ 1 & & . \\ a_{n1} & . & a_{nn-1} & 1 \end{pmatrix}, \mathbf{B} \begin{pmatrix} b_{11} & 0 & 0 & 0 \\ 0 & b_{22} & 0 & 0 \\ . & . & b_{ii}. & . \\ 0 & 0 & 0 & b_{nn} \end{pmatrix} \tag{6.6}$$

This is obviously a just-identification scheme, where the identification of structural shocks depends on ordering of variables. It corresponds to a recursive economic structure, with the most endogenous variable ordered last

6.2.2 *Structural models with contemporaneous restrictions*

In this identification scheme some a-priori information is used to impose restrictions on the elements of matrices **A** and **B,** different from the Choleski ordering**.** If the objective of VAR is provide evidence to choose between competing models, the identifying restrictions should be independent from the theoretical predictions of those models. The recent literature on the monetary transmission mechanism, (Strongin[57], Bernanke-Mihov[5], Christiano, Eichenbaum and Evans[12], Leeper,SimsandZha[33]), offers good examples on how this kind of restrictions can be derived. VARs of the monetary transmission mechanism are specified on six variables, with the vector of macroeconomic non-policy variables including gross domestic product (GDP), the consumer price index (P) and the commodity price level (Pcm), the vector of policy variables includes the federal funds rate (FF), the quantity of total bank reserves (TR) and the amount of nonborrowed reserves (NBR). Given the estimation of the reduced form VAR for the six macro and monetary variables, a structural model is identified by: (i) assuming orthogonality of the structural disturbances; (ii) imposing that macroeconomic variables do not simultaneously react to monetary variables, while the simultaneous feedback in the other direction is allowed, and (iii) imposing restrictions on the monetary block of the model reflecting the operational procedures implemented by the monetary policy maker. All identifyng restrictions satisfy the criterion of independence from specific theoretical models, in fact ,within the class of models estimated on monthly data, restrictions (ii) are consistent with a wide spectrum of alternative theoretical structures and imply a minimal assumption on the lag of the impact of monetary policy on macroeconomic variables, whereas restrictions (iii) are based on institutional analysis.

Restrictions (ii) are made operational by setting to zero an appropriate block of elements of the **A** matrix.

The contemporaneous relations among the Fed funds rate and the reserve aggregates are derived, as in Bernanke and Mihov [5], from a specific model of the reserve market.

$$u^{TR} = -\alpha\, u^{FF} + \nu^D \tag{6.7}$$
$$u^{BR} = \beta\, u^{FF} + \nu^B \tag{6.8}$$
$$u^{NBR} = \phi^D \nu^D + \phi^B \nu^B + \nu^S \tag{6.9}$$

Equation (6.7) and (6.8) describe banks' demand equations (expressed in innovation -i.e. VAR residual- form) for total and borrowed reserves BR (time subscripts are omitted): the federal funds rate affects negatively the demand for total reserves (6.7) and positively the demand for borrowed reserves[12]. ν^D and ν^B are disturbances to total and borrowed reserves respectively. The supply of nonborrowed reserves in (6.9) reflects the behaviour of the Federal Reserve. In particular, by means of open-market operations, the Fed can change the amount of NBR supplied to the banking system in response to (readily observed) disturbances to total and borrowed reserve demand. Moreover, variations in nonborrowed reserves may be due to monetary policy shocks unrelated to reserve demand behaviour. In (6.9) the coefficients ϕ^D and ϕ^B measure the reaction of the Fed to total and borrowed reserve demand movements respectively, and ν^S represents the monetary policy shock to be empirically identified. The market for reserves featuring the assumed simultaneous relations is described by the following figure :

Combining the market for reserves with the macroeconomic variables, we can explicitly rewrite (6.4) as follows:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ a_{21} & 1 & 0 & 0 & 0 & 0 \\ a_{31} & a_{32} & 1 & 0 & 0 & 0 \\ a_{41} & a_{42} & a_{43} & 1 & -\frac{1}{\beta} & \frac{1}{\beta} \\ a_{51} & a_{52} & a_{53} & \alpha & 1 & 0 \\ a_{61} & a_{62} & a_{63} & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_t^{GDP} \\ u_t^{P} \\ u_t^{Pcm} \\ u_t^{FF} \\ u_t^{TR} \\ u_t^{NBR} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{\beta} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \phi^B & \phi^D & 1 \end{pmatrix} \begin{pmatrix} \nu_{1t}^{NP} \\ \nu_{2t}^{NP} \\ \nu_{3t}^{NP} \\ \nu_t^{B} \\ \nu_t^{D} \\ \nu_t^{S} \end{pmatrix} \tag{6.10}$$

Several features of (6.10) must be noted. First, VAR residuals from the first three equations, describing the non-policy part of the system, are orthogonalized simply by assuming a recursive (Choleski) structure for the corresponding block of the **A** matrix. This procedure yields orthogonal disturbances to which we do not attach a specific "structural" interpretation, labelling them simply as ν_i^{NP} $(i = 1, 2, 3)$, where NP denotes a non-policy shock.

Second, as shown by Bernanke and Mihov [5], the general formulation in (6.10), is still not identified, but identification can be completed by a careful analysis of the operational procedures followed by the Central Bank.

- Case 1: Federal funds targeting

[12]We assume from the start that movements in the discount rate, which would enter (6.8) with a negative sign, are completely anticipated, so that the innovation in the Fed funds-discount rate differential is entirely attributable to the former rate.

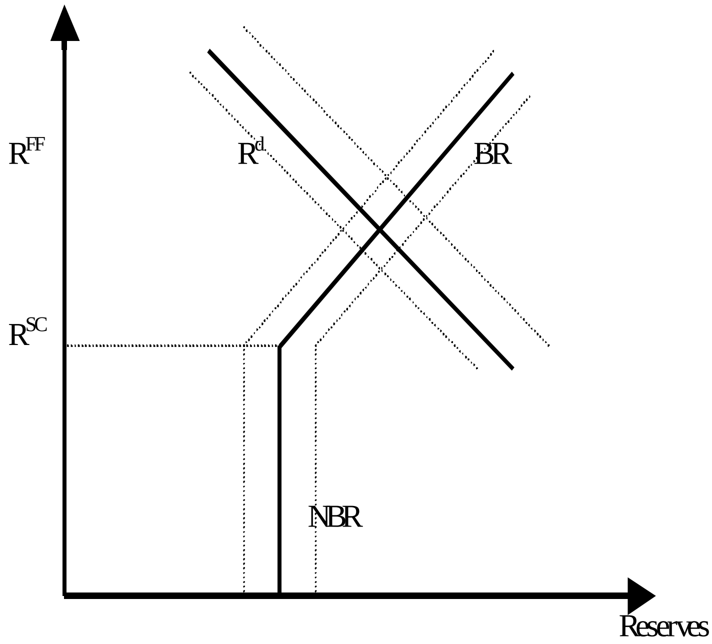


FIG. 6.1. The U.S. market for bank reserves

In this case we have $\varphi^d = 1, \varphi^b = -1$. Central Banks uses NBR to neutralize shocks coming from banks and households behaviour . We then have for the monetary block identification:

$$\begin{bmatrix} u_t^{TR} \\ u_t^{NBR} \\ u_t^{FF} \end{bmatrix} = \begin{bmatrix} 1 & \frac{\alpha}{\beta+\alpha} & 0 \\ 1 & 1 & -1 \\ 0 & -\frac{1}{\beta+\alpha} & 0 \end{bmatrix} \begin{bmatrix} \nu_t^D \\ \nu_t^S \\ \nu_t^B \end{bmatrix}$$

The model is now over-identified . Choleski plus additional restrictions

- Case II: targeting NBR.

$\varphi^d = 0, \varphi^b = 0.$

$$\begin{bmatrix} u_t^{TR} \\ u_t^{NBR} \\ u_t^{FF} \end{bmatrix} = \begin{bmatrix} \frac{\beta}{\beta+\alpha} & \frac{\alpha}{\beta+\alpha} & \frac{\alpha}{\beta+\alpha} \\ 0 & 1 & 0 \\ \frac{1}{\beta+\alpha} & -\frac{1}{\beta+\alpha} & -\frac{1}{\beta+\alpha} \end{bmatrix} \begin{bmatrix} \nu_t^D \\ \nu_t^S \\ \nu_t^B \end{bmatrix}$$

NBR is now informative on monetary policy shocks

- Case III: Strongin identification (1994)

Shocks to reserves are demand shocks which the CB has to accomodate. Therefore monetary policy shocks are the shocks to NBR orthogonal to shocks to TR. Moreover CB does not react to Borrowed Reserves. $\alpha = 0, \varphi^b = 0$

$$\begin{bmatrix} u_t^{TR} \\ u_t^{NBR} \\ u_t^{FF} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \varphi^d & 1 & 0 \\ -\frac{\varphi^d - 1}{\beta} & -\frac{1}{\beta} & -\frac{1}{\beta} \end{bmatrix} \begin{bmatrix} \nu_t^D \\ \nu_t^S \\ \nu_t^B \end{bmatrix}$$

NBR is now informative

- Case IV: controlling Borrowed Reserves

In this case TR-NBR is only function of shocks ν^s. $\varphi^d = 1, \varphi^b = \frac{\alpha}{\beta}$. In questo caso si ha:

$$\begin{bmatrix} u_t^{TR} \\ u_t^{NBR} \\ u_t^{FF} \end{bmatrix} = \begin{bmatrix} 1 & \frac{\alpha}{\beta+\alpha} & \frac{\alpha}{\beta} \\ 1 & 1 & \frac{\alpha}{\beta} \\ 0 & -\frac{1}{\beta+\alpha} & -\frac{1}{\beta} \end{bmatrix} \begin{bmatrix} \nu_t^D \\ \nu_t^S \\ \nu_t^B \end{bmatrix}$$

It is easily seen that alternative regime would imply identification which are technically not far from Choleski's triangularization, with different ordering of the monetary variables. In a Fed Fund targeting regime FF does not react contempèoraneously to the other monetary variables while in a Non-Borrowed Reserves targeting regime it is NBR that does not react contemporaneously to the other two monetary shocks. Moreover, information on the operating procedure by the FED are important in determining the appropriate identification scheme and, more importantly, VAR models of the MTM should be estimated within a single policy regime. Bagliano and Favero [2], provide evidence on the structural instability of VAR of the MTM estimated across different monetary policy regimes.

6.2.3 *Structural model with long-run restrictions*

Often long-run behaviour of shocks provide restrictions acceptable within a wide range of theoretical model. A typical restriction compatible with virtually all macroeconomic models is that in the long-run demand shocks have zero impact on output. Blanchard-Quah [8] show how these restrictions can be used to identify VARs.

The structural model of interest is specified by posing $\mathbf{A}$ equal to the identity matrix and by not imposing any zero restriction on the $\mathbf{B}$ matrix. We then have for a generic vector of variables $\mathbf{y}_t$ the following specification:

$$\mathbf{y}_t = i = 1\sum^{p} \mathbf{A}_i \mathbf{y}_{t-i} + \mathbf{B}\mathbf{v}_t$$

from which it is possible to derive the matrix which describes the long-run effect of the structural shocks on the variables of interest as follows:

$$\left(\mathbf{I} - i = 1\sum^{p} \mathbf{A}_i \right)^{-1} \mathbf{B}\mathbf{v}_t = -\Pi^{-1}\mathbf{B}\mathbf{v}_t$$

Coefficients in Π are obtained from the reduced form, therefore we are able to impose long-run restrictions given the estimation of the reduced form

Two points are worth noting

- $(I - A_1)$ is -Π,for this matrix to be invertible the VAR must be specified on stationary variables
- the long-run restrictions are restrictions on the cumulative impulse response function

Let us now consider the Blanchard-Quah[8] data-set. The authors aim at separating demand shocks from supply shocks, they consider a VAR on two variables, the unemployment rate, UN, and the quarterly rate of growth of GDP, ΔLY. The original sample contains quarterly data from 1951:2 to 1987:4, we have retrieved the two series from Datastream and they are available only for the sample 1951:2-1987:4. The series are available in the file BQ.WKS. The VAR is specified with 8 lags a constant and a deterministic trend (in the original paper a break in the constant for ΔLY is also allowed but we do not allow it) as follows.

$$\begin{pmatrix} \Delta LY_t \\ UN_t \end{pmatrix} = A_1 \begin{pmatrix} \Delta LY_{t-1} \\ UN_{t-1} \end{pmatrix} + ...A_8 \begin{pmatrix} \Delta LY_{t-8} \\ UN_{t-8} \end{pmatrix} + A_9 \begin{pmatrix} 1 \\ TREND \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

The structure of interest is the following:

$$\begin{pmatrix} \Delta LY_t \\ UN_t \end{pmatrix} = A_1 \begin{pmatrix} \Delta LY_{t-1} \\ UN_{t-1} \end{pmatrix} + ...A_8 \begin{pmatrix} \Delta LY_{t-8} \\ UN_{t-8} \end{pmatrix} + A_9 \begin{pmatrix} 1 \\ TREND \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}$$

To obtain the identifying restrictions consider that

$$\begin{pmatrix} DLY_t \\ UN_t \end{pmatrix} = \left(\mathbf{I} - i = 1 \sum^{p} \mathbf{A}_i \right)^{-1} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} =$$
$$= \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}$$

demand shocks are identified by imposing that their long-run impact on the level of output is zero:

$$k_{11} b_{11} + k_{12} b_{21} = 0$$

Note that by imposing the restriction that the cumulative impulse response of the rate of outptut growth to a demand shock is zero we impose the restriction that the impulse response of the level of outptut to a demand shock is zero in the long-run. As the variables are stationary the long-run response of ΔLY and UN to all shocks is zero by definition.

We implement the procedure on the data by using MALCOLM[41], a package written for RATS.

We from the estimation of the VAR, we then implement Johansen on this VAR, we make sure that the null of stationarity is not rejected. We then retrieve the Π matrix. $\Pi = \begin{bmatrix} 0.1451 & 0.2168 \\ -0.5741 & -0.0693 \end{bmatrix}$,

Then we specify $(-\Pi)^{-1} = \begin{bmatrix} -0.1451 & -0.2168 \\ 0.5741 & 0.0693 \end{bmatrix}^{-1} = \begin{bmatrix} .60572 & 1.8949 \\ -5.0179 & -1.2683 \end{bmatrix}$
and our long-run identifying restriction is

$$.60572 b_{11} + 1.8949 b_{21} = 0$$

Note the difference between this methodology and the Cholesky decomposition, which would simply restrict b_{21} to zero.

6.2.4 *Identification in cointegrated VARs*

Let us consider now how the identification problem changes when we have a cointegrated VAR. Considering, for simplicity, only first order dynamics, the cointegrated reduced form is :

$$\Delta \mathbf{y}_t = \Pi \mathbf{y}_{t-1} + \mathbf{v}_t$$

where $\Pi = \boldsymbol{\alpha}\boldsymbol{\beta}'$. As we know, identification of the cointegrating vectors is a problem totally separated from identification of the structural shocks of interest. Therefore, having solved the identification of the cointegrating relationships, we have still to deal with the problem of posing appropriate restrictions on the parameters of the $\mathbf{B}$ matrix in order to pin down the shocks $\mathbf{u}_t$

$$\Delta \mathbf{y}_t = \Pi \mathbf{y}_{t-1} + \mathbf{B}\mathbf{u}_t$$

In the context of cointegration, the identification problem can be solved in a very natural way. Consider, for simplicity, the case of a bivariate model $\mathbf{y}_t = (y_t, x_t)$, in which variables are non stationary I(1) but cointegrated with cointegrating vector$(1, -1)$, so the rank of the Π matrix is 1 and we use the following representation of the stationary reduced form:

$$\begin{pmatrix} \Delta y_t \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} \alpha_{11} \\ \alpha_{21} \end{pmatrix} \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} \tag{6.11}$$

model (6.11) can be re-written as follows[40]

$$\begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1-L & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} (y_t - x_t) \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} \alpha_{11} & 0 \\ \alpha_{21} & 0 \end{pmatrix} \begin{pmatrix} (y_{t-1} - x_{t-1}) \\ \Delta x_{t-1} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} \tag{6.12}$$

The two representation are absolute identical (same residuals). The second representation has been widely use in research based on present value models.

The cointegrating properties of the system suggests the presence of two types of shocks: a permanent one(to be related to the single common trend shared by the two variables) and a transitory one (to be related to the cointegrating relation). It seems therefore natural to identify one shock as permanent the other as transitory. Given that we have a stationary system, the identification of shocks is obtained by deriving long-run responses of the variables of interest to relevant shocks. From (6.12) we have:

$$\left(\begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1-L & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \alpha_{11}L & 0 \\ \alpha_{21}L & 0 \end{pmatrix}\right)\begin{pmatrix} (y_t - x_t) \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}\begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} \tag{6.13}$$

From which long-run responses are obtained by setting L=1 and by inverting the matrix pre-multiplying variables in the stationary representation of VAR

$$\begin{pmatrix} (y_t - x_t) \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} -\alpha_{11} & 1 \\ -\alpha_{21} & 1 \end{pmatrix}^{-1}\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}\begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} \tag{6.14}$$

$$\begin{pmatrix} (y_t - x_t) \\ \Delta x_t \end{pmatrix} = \begin{pmatrix} \frac{-b_{11}+b_{21}}{\alpha_{11}-\alpha_{21}} & -\frac{b_{12}-b_{22}}{\alpha_{11}-\alpha_{21}} \\ \frac{-\alpha_{21}b_{11}+\alpha_{11}b_{21}}{\alpha_{11}-\alpha_{21}} & \frac{-\alpha_{21}b_{12}+\alpha_{11}b_{22}}{\alpha_{11}-\alpha_{21}} \end{pmatrix}\begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} \tag{6.15}$$

so v_{2t} can be identified as the transitory shock by imposing the following restriction:

$$-\alpha_{21}b_{12} + \alpha_{11}b_{22} = 0$$

which, given knowledge of the α parameters from cointegration analysis, provides the just-identifying restriction for the parameters in $\mathbf{B}$. Interestingly, there is one case in which this identification is equivalent to the Choleski ordering, the case in which $\alpha_{11} = 0$.Note that this is the case in which Δy_t is weakly exogenous for the estimation of b_{21}. An application of this identifying scheme is provided in Favero, Giavazzi, Spaventa,[24] where the procedure is implemented to separate international from local factors in the determination of interest rates fluctuations

6.3 Why shocks ?

Having identified the "monetary rule" by proposing an explicit solution to the problem of the endogeneity of money, the VAR approach concentrates on *deviations* from the rule. Deviations from the rule can be obtained either by changing the systematic component of monetary policy or by considering exogenous shocks, which leave systematic monetary policy unaltered. In the former case the deviation from the rule is obtained by changin some parameters in the **A** matrix describing the simultaneous relations among variables, while in the latte case the parameters in the matrices **A** and **B** are not altered. Consider for example the case of federal fund targeting. The first type of deviations is obtained by modifying the response of the federal fund targeting to macroeconomic conditions, i.e.fluctuations in output, commodity prices and the consumer price index, while the second type of deviations is obtained by considering an exogenous shock which does not alter the response of the monetary policy maker to macroeconomic conditions.

VAR modellers have exclusively concentrated on simulating shocks, leaving the sytematic component of monetary policy unaltered.

The VAR approach to the monetary transmission mechanism has been criticised on the basis that it views central banks as "random number generators". This does not seem to be correct: in fact, monetary policy rules are explicitly estimated in structural VAR models. However, the focus is not on rules but on deviations from rules, since only when central banks deviate from their rules it becomes possible to collect interesting information on the response of macroeconomic variables to monetary policy impulses, to be compared with the predictions of the alternative theoretical models. In fact, deviations from monetary policy rules provide researchers with the best opportunity to detect the response of macroeconomic variables to monetary impulses that are not expected by the market. The first chain of most models of the monetary transmission mechanism links the policy rates to the term structure of the interest rates and the most popular model of the term structure, the expectational model, predicts that the term structure does not generally react to expected monetary impulses. The monetary impulses relevant to the transmission analysis are therefore structural shocks in (6.2).

Recently, McCallum[**?**] has criticised the choice of VAR modellers of concentrating on shocks which leave the systematic component of monetary policy unaltered. It is argued that the emphasis on the shock component is misplaced because the unsystematic portion of policy-instrument variability is very small in relation to the variability of the sytematic component. "... Indeed, it is conceivable that the policy behaviour could be virtually devoid of any unsystematic component. In the limit, that is, the variance of the shock component could approach zero. But this would not imply that monetary policy is unimportant for price level behaviour, central bank's main responsibility...." ([**?**], p.5).

The simulation of systematic monetary policy requires, for robustness to the Lucas'critique, the specification of a forward-looking model in which "deep parameters" are identified independently from nuisance parameters describing expectations formation and dependent on the policy regime. This is what McCallum effectively does in a series of papers[**?**],[39] where the impact of modifications in the monetary policy maker reaction function is dynamically simulated.

However, it is important to note that McCallum work is not aimed at model selection but rather at model simulation. The question of using the empirical evidence to judge between different theoretical model is not addressed in McCallum work, based on a specific model.

If VAR models are instead used to describe the empirical evidence relevant to the choice between alternative theoretical models, then there is a possible defense of the choice of concentrating on shocks rather than on the systematic components of monetay policy. Such defense is related to the Lucas critique.

Consider the following Data Generating Process:

$$
\begin{aligned}
y_t &= a_1 m_{t+1}^e + a_2 y_{t-1} + u_{1t} \\
m_t &= b_0 + b_1 y_{t-1} + b_2 m_{t-1} + u_{2t}
\end{aligned}
$$

Where y is the macroeconomic variable and m is the monetary policy variable.

The DGP is the relevant theoretical model, which is unknown to the empirical researcher. The empirical researcher tries instead to describe the empirical relation between the monetary instruments and the macroeconomic variables by specifying the following structural VAR:

$$
\begin{aligned}
y_t &= c_o + c_1 m_t + c_2 y_{t-1} + v_{1t} \qquad (6.16) \\
m_t &= b_0 + b_1 y_{t-1} + b_2 m_{t-1} + v_{2t}
\end{aligned}
$$

Where the following restrictions hold: $c_o = a_1 b_0$, $c_1 = a_1 b_2, c_2 = a_2 + a_1 b_1$.

6.16 is not viable for econometric evaluation of sytematic monetary policy, in that the parameters in the equation for y cannot be kept constant when the sytematic component of the monetary policy rule is altered. However, the simulation of the dynamic impact of a monetary policy shocks identified a-la-Choleski ordering m first is still viable in that it is perfomed while keeping all parameters constant.

Note that this small example reiterates the importance of estimating parameters in the Structural VAR models by concentranting on a single policy regime, in fact regime shifts require different parameterizations.

6.4 Description of VAR models

After the identification of structural shocks of interest, the properties of VAR models are described using impulse response analysis, variance decomposition and historical decomposition.

Consider a strucutral VAR model for a generic vector $\mathbf{y}_t$, containing m variables:

$$
\mathbf{A}_0 \mathbf{y}_t = i = 1 \sum^{p} \mathbf{A}_i \mathbf{y}_{t-i} + \mathbf{B} \mathbf{v}_t
$$

which we can rewrite as :

$$[\mathbf{A}_0 - \mathbf{A}(L)]\,\mathbf{y}_t = \mathbf{B}\mathbf{v}_t$$
$$\mathbf{A}(L) = i = 1\sum^{p}\mathbf{A}_i L^i$$

now by inverting $[\mathbf{A}_0 - \mathbf{A}(L)]$ (under the assumption of invertibility of this polynomial) we obtain the moving average representation for our VAR process:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{C}(L)\,\mathbf{v}_t \\ \mathbf{y}_t &= \mathbf{C}_0\mathbf{v}_t + \mathbf{C}_1\mathbf{v}_{t-1} + ... + \mathbf{C}_s\mathbf{v}_{t-s} \\ \mathbf{C}(L) &= [\mathbf{A}_0 - \mathbf{A}(L)]^{-1} \\ \mathbf{C}_0 &= \mathbf{A}_0^{-1}\mathbf{B} \end{aligned} \tag{6.17}$$

To illustrate the concept of an *impulse response function,* we interpret the generic matrix $\mathbf{C}_s$within the moving average representation as follows :

$$\mathbf{C}_s = \frac{\partial \mathbf{y}_{t+s}}{\partial \mathbf{v}_t}$$

in other word the generic element $\{i, j\}$ of matrix $\mathbf{C}_s$ represent the impact of a shock hitting the j-th variable of the system at time t on the i-th variable of the system at time t+s. As s varies we have a function describing the response of variable i to an impulse in variable j. For this function of partial derivative to be meaningful we must allow that a shocks in variable j occurs while all other shocks are kept to zero. Of course this is allowed for structural shocks, as they are identified by imposing they are orthogonal to each other. Note howeve that the concept of an impulse response function is not applicable to reduced form VAR innovations, which, in general, are correlated to each other.

Hystorical decomposition is obtained by using the structural MA representation to separate series in the components(orthogonal to each other) attributable to the different structural shocks.

Finally *Forecasting Error Variance Decomposition* are obtained from(6.17) by deriving the error in forecasting $\mathbf{y}$ s period in the future as:

$$(\mathbf{y}_{t+s} - E_t\mathbf{y}_{t+s}) = \mathbf{C}_0\mathbf{v}_t + \mathbf{C}_1\mathbf{v}_{t-1} + ... + \mathbf{C}_s\mathbf{v}_{t-s}$$

from which we can construct the variance of such forecasting error as:

$$Var(\mathbf{y}_{t+s} - E_t\mathbf{y}_{t+s}) = \mathbf{C}_0 I\mathbf{C}_0' + \mathbf{C}_1 I\mathbf{C}_1' + ... + \mathbf{C}_s I\mathbf{C}_s'$$

from which we can compute the share of the total variance attributable to the variance of each structural shocks. Note again that such composition make sense only if shocks are orthogonal to each other. In fact it is only in this case that we can write the variance of the total forecasting error as a sum of varinces of the single shocks (as the covariance terms are zero following the orthogonality property of structural shocks).

To illustrate the three concepts consider the following bivariate VAR, in which structural parameters have been identified and estimated via a Choleski decomposition:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix}$$

the MA representation is

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} v_{1t} \\ v_{2t} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} v_{1t-1} \\ v_{2t-1} \end{pmatrix} + \\ + ... + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{s} \begin{pmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} v_{1t-s} \\ v_{2t-s} \end{pmatrix}$$

From which impulse response functions, historical decomposition and Forecasting Error Variance Decomposition are immediately obtained.

6.5 Monetary policy in closed economies

Cumulative work on the analysis of the monetary transmission mechanism in the U.S. led to the specification of a VAR system which has by now become the standard reference model. We have already seen and discussed this benchmark specification which contains six variables: gross domestic product (GDP), the consumer price index (P) and the commodity price level (Pcm) together with the federal funds rate (FF), the quantity of total bank reserves (TR) and the amount of nonborrowed reserves (NBR).

It is interesting to see how the specification of the benchmark model has developed over-time.

Initially models were estimated on rather limited set of variables, i.e. prices, money and output, and identified imposing a diagonal form to the matrix B and a lower triangular form to the matrix A with money coming last in the ordering of the variables included in the VAR (Choleski identification). This first type of models are discussed in Leeper,Sims e Zha [33], we replicate their results on the data-set LSZUSA.WF1. The underlying structural model is specified as follows:

$$\mathbf{A}_0\mathbf{y}_t = i = 1\sum^{k} \mathbf{A}_i\mathbf{y}_{t-i} + \mathbf{B}\boldsymbol{v}_t \tag{6.18}$$

$$\mathbf{y}_t = \begin{bmatrix} p_t \\ y_t \\ m_t \end{bmatrix}, \ \mathbf{A}_0 = \begin{bmatrix} 1 & 0 & 0 \\ a_{21} & 1 & 0 \\ a_{31} & a_{32} & 1 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} b_{11} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & b_{33} \end{bmatrix}, \boldsymbol{\nu}_t \sim N.I.D.\,(0, I)$$

All variables are expressed in logs. Identification is Choleski with money ordered last. This is a model geared to deliver monetary policy shocks, so the identification of shocks to LP and LY does not matter. As [33],we have estimated the model on the sample 1960:1-1996:3, including six lags of each variable. The following impulse responses are obtained:

Prices slowly react to monetary policy, output responds in the short-run, in the long-run (from two years after the shock onwards) price start adjusting and the significant effect on output vanishes. There is no strong evidence for the endogeneity of money. This is easily checked by looking at the estimated parameters in A_0 and by analysing FEVD

Macroeconomic variables play a very limited role in explaining the variance of the forecasting error of money, while money plays instead an important role in explaining fluctuations of both the macroeconomic variables.

Sims [**?**] extended tha VAR to include the interest rate on Federal Funds ordered just before money as a penultimate variables in the Choleski identification. The idea is to see the robustness of the above results after identifying the part of money which is endogenously to the interest rate. Impulse response functions are modified as follows:

while FEVD is modified as follows:

Impulse response function and FEVD raise a number of issues:

- though little of the variation in money is predictable from past output and prices, a considerable amount becomes predictable when past short-term interest rates are included in the information set.
- it is difficult to interpret the behaviour of money as driven by money supply shocks. The response to money innovations gives rise to the "liquidity puzzle": the interest rate declines very slightly contemporaneously in response to a money shock to start increasing afterwards.
- There are diffiuclties also with interpreting shocks to interest rates as monetary policy shocks. The response of prices to an innovation in interest rates gives rise to the "price puzzle": prices increase significantly after an interest rate hike. An accepted intepretation of the liquidity puzzle relies on the argument that the money stock is dominated by demand rather than by supply shock.Moreover the interpretation of money as demnd shocks driven is consistent with the impulse response of money to interest rates.

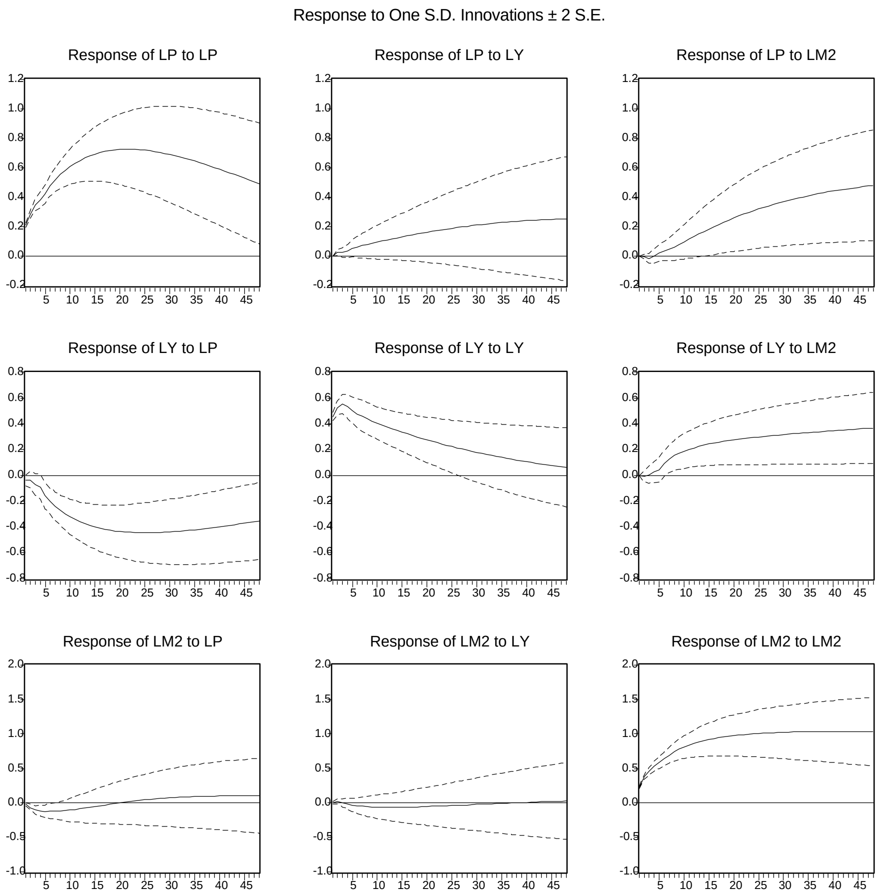


FIG. 6.2. Impulse Response functions in a three-variables VAR of the MTM

Note also that, even if the money stock were to be dominated by supply shocks, it would be reflecting both the behaviour of central banks and of the banking system. For both these reasons the broad monetary aggregate has been substituted by narrower aggregates, bank reserves, on which is easier to identify shocks mainly driven by the behaviour of the monetary policy maker. The "price puzzle" has been attributed to misspecification of the four-variables VAR used by Sims. Suppose that there exists a leading indicator for inflation to which the FED reacts. If such a leading indicator is omitted from the VAR, we have then an omitted variable positively correlated with inflation and interest rates which makes the VAR misspecified and explains the positive relation between prices and interest rates observed in the impulse response functions. It has been ob-

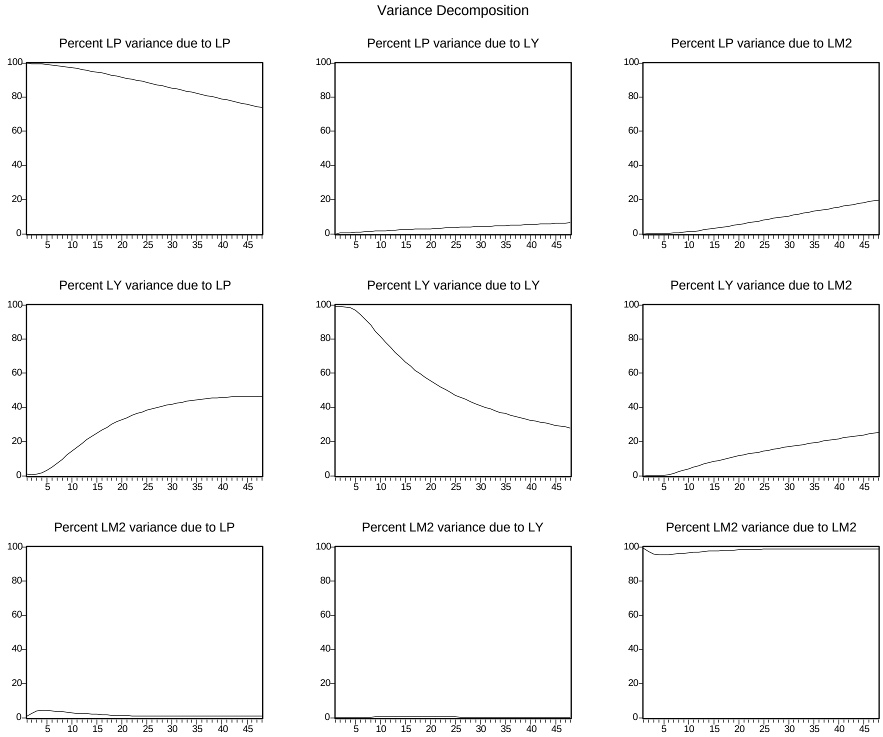


FIG. 6.3. FEVD in a three-variables VAR of the MTM

served Christiano, Eichenbaum and Evans[12],Sims[53] that the inclusion of a Commodity Price Index in the VAR solves the "price puzzle". Our brief historical record of the empirical analysis of closed economy VAR of the MTM has brought us to the justification of the six-variables included in what is by now known as the benchmark VAR model. We have already discussed its identification, let us now examine impulse response function derive by using the FED fund targeting identifying restrctions.

The evidence reported in the IRF represents the relevant fact to be included in theoretical models of the MTM. It is this kind of evidence that has established that plausible models of the monetary transmission mechanism should be consistent at least with the following evidence on price, output and interest rates: *(i)* the aggregate price level initially responds very little; *(ii)* interest rates initially rise, and *(iii)* aggregate output initially falls, with a j-shaped response, with a zero long-run effect of the monetary impulse.

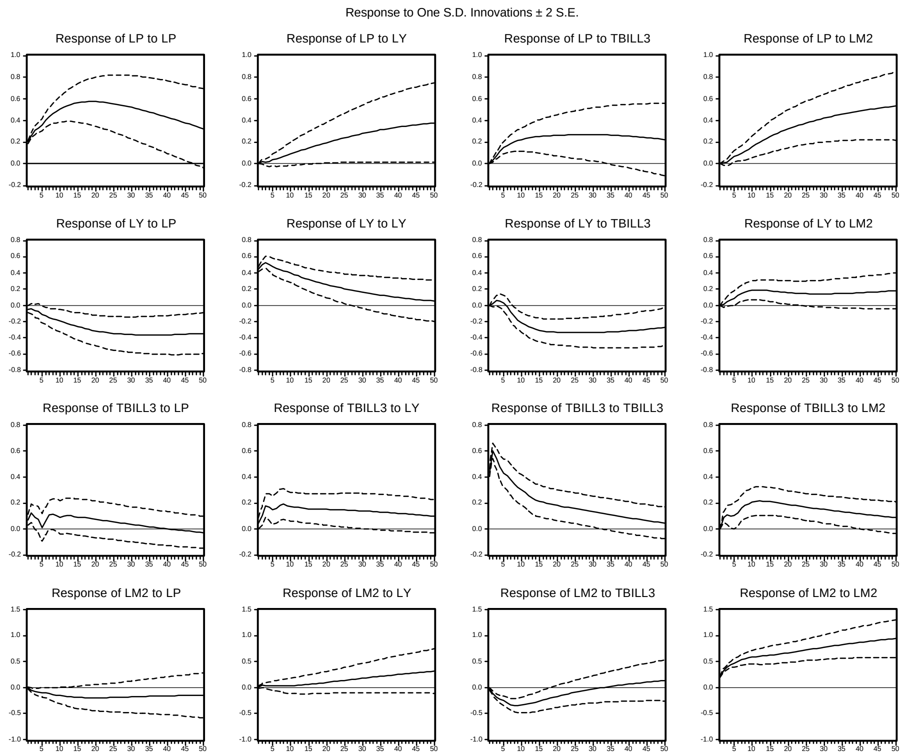


FIG. 6.4. Impulse responses in a four-variables VAR of the MTM

6.6 Monetary policy in open economies

Various papers have examined the effects of monetary shocks in open economies, but this strand of literature has been distinctly less successful in providing accepted empirical evidence than the VAR approach in closed economies.

The first results have been provided by Eichenbaum and Evans (1995), using an open-economy VAR with the following structure:

$$\mathbf{A}_0\mathbf{y}_t = i = 1\sum^{k}\mathbf{A}_i\mathbf{y}_{t-i} + \mathbf{B}\boldsymbol{v}_t \tag{6.19}$$

where $\mathbf{y}_t = \left[Y_t^{US}\ P_t^{US}\ NBRX_t^{US}(FF_t)\ Y_t^{FOR}\ P_t^{FOR}\ R_t^{FOR}\ e_t(q_t)\right]'$

Y^{US}, P^{US} are logs of US output and price, $NBRX^{US}$ is the ratio of non-borrowed to total reserves (the appropriate variable from which extract monetary policy shocks under a regime of non-borrowed reserves targeting). FF is the Federal Funds rate, which is considered in alternative to $NBRX^{US}$, and it is the informative variable for the extraction of monetary policy shocks under

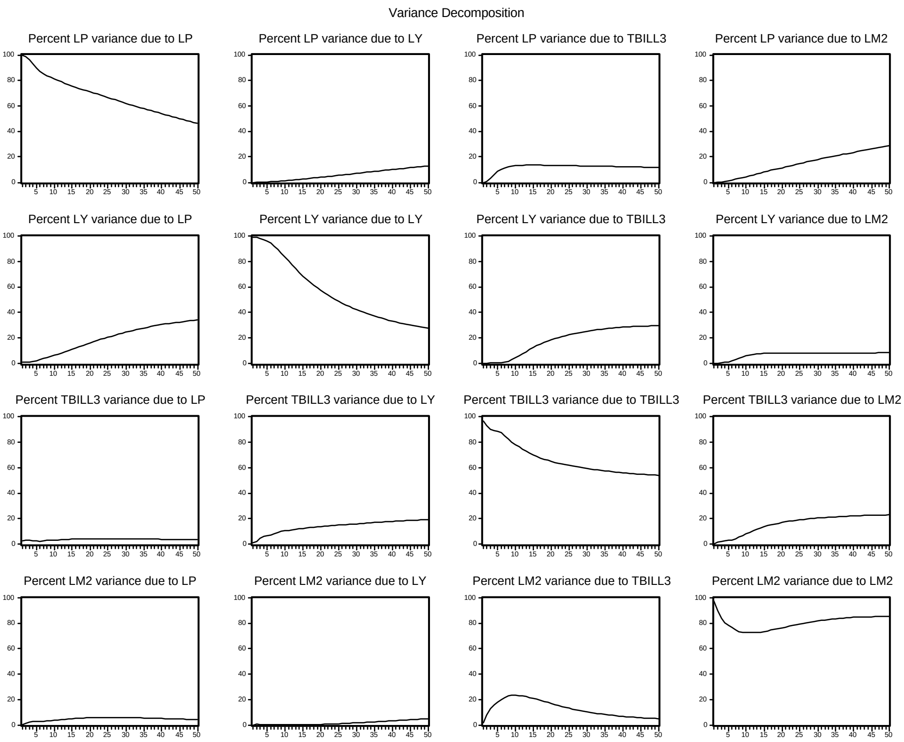


FIG. 6.5. FEVD in a four-variable VAR of the MTM

the regime of interest rate targeting; Y^{FOR}, P^{FOR}, and R^{FOR} are respectively the logs of output, prices, and the level of short-term interest rate in the foreign country; e is the nominal bilateral exchange rate, while q is the real bilateral exchange rate. The matrix $\mathbf{B}$ is diagonal and $\mathbf{A}$ is lower-triangular. The empirical analysis is implemented by considering in turn as a foreign country each of the G7 countries on a sample of monthly data from 1974:1 to 1990:5. The following evidence emerges: (i) a restrictive US monetary policy shock generates a significant and persistent appreciation of the US dollar; (ii) a restrictive US monetary policy shock generates a significant and persistently larger effect on the domestic interest rate with respect to the foreign rate; (i) and (ii) imply a sharp deviation from the uncovered interest parity condition in favour of US dollar-denominated investments (the "forward-discount puzzle"); (iii) identified US monetary policy shocks are not different from the shocks derived within closed-economy VARs (iv) the closed-economy response of US prices and output to monetary policy shocks is robust to the extension of the VAR to the open economy; (v) a restrictive foreign monetary policy shock generates an appreciation of the US dollar (the "exchange-rate puzzle"); and (vi) the response of the real exchange rate to

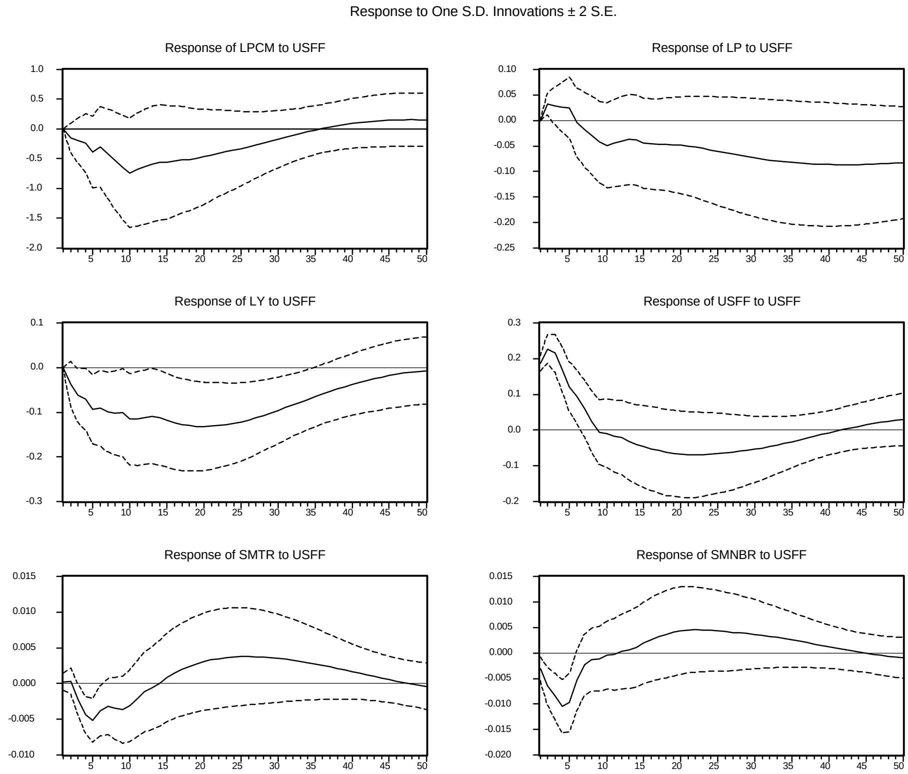


FIG. 6.6. Impulse response functions in a six variables VAR of the MTM

the US and foreign monetary policy shocks does not differ significantly from the response of the nominal exchange rate. Such evidence is substantially confirmed by the the work of Schlagenhauf and Wrase (1995), who consider a very similar specification for the G-5 countries over the sample 1972:2-1990:2, using quarterly data.

Some considerations are in order to help the interpretation of the above results.

First, the empirical models are estimated over samples including shifts in US and foreign monetary policy regimes: therefore, parameter instability is a potential problem.

Second, the extension to the open economy features the omission from the VAR of the commodity price index and of the monetary variables not relevant to the extraction of the policy shocks. While the simplification of the monetary block is sustainable in the light of the absence of contemporaneous feedback between the informative variables and the other monetary variables under the chosen identification schemes, the omission of the commodity price index is not justifiable as it leads to the same misspecification as in the closed economy model for US monetary policy shocks. Moreover, such omission might well also bias the identification of the foreign monetary policy shocks if the commodity price index is regarded as a leading indicator of inflation by the foreign policymaker. Therefore, it could be argued that the observed puzzles might depend on the incorrect specification of the VAR generated by the omission of the commodity price index.

Third, on the identification scheme. While some rationale can be provided for a quasi-recursive scheme in closed economies, a similar justification is much harder to provide in open economies. In fact, the recursive identification scheme with the exchange rate ordered last implies that neither the US nor the foreign monetary authority react contemporaneously to exchange rate fluctuations. This assumption seems to be sustainable for the US (the FED benign neglect for the dollar) but it is certainly heavily questionable when the foreign countries are considered, as they are much more open economies than the US. The failure of the recursive identification scheme could also be responsible for the observation of the puzzles. In fact, most of the recent empirical work is aimed at breaking such recursive structure in the identification scheme.

Kim and Roubini (1997) obtain such aim by introducing a structural identification by the explicit consideration of a money demand and supply functions. Their specification can be described as follows for the generic non-US country:

$$\mathbf{A}_0 \mathbf{y}_t = i = 1 \sum^{k} \mathbf{A}_i \mathbf{y}_{t-i} + \mathbf{B} \boldsymbol{v}_t \tag{6.20}$$

$$\mathbf{y}_t = \begin{bmatrix} OPW_t \\ FF_t \\ Y_t^{FOR} \\ P_t^{FOR} \\ M_t^{FOR} \\ R_t^{FOR} \\ e_t \end{bmatrix}, \quad \mathbf{A}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ a_{21} & 1 & 0 & 0 & 0 & 0 & 0 \\ a_{31} & 0 & 1 & 0 & 0 & 0 & 0 \\ a_{41} & 0 & a_{43} & 1 & 0 & 0 & 0 \\ 0 & 0 & a_{53} & a_{54} & 1 & a_{56} & 0 \\ a_{61} & 0 & 0 & 0 & a_{65} & 1 & a_{67} \\ a_{71} & a_{72} & a_{73} & a_{74} & a_{75} & a_{76} & 1 \end{bmatrix} \tag{6.21}$$

$$\boldsymbol{\nu}_t \sim N.I.D.\,(0, I) \tag{6.22}$$

with R^{FOR} denoting the short-term non-US policy rate, M^{FOR} a monetary aggregate (M0 or M1), P^{FOR} the log of consumer price index, Y^{FOR} the log of industrial production, OPW the world index of oil price in dollars, FF the Federal Funds rate, and e the nominal exchange rate against the dollar. **B** is a diagonal matrix. The model is estimated over the sample 1974:7-1992:5, on monthly data. The main differences between the proposed strucutral identification and the recursive identification scheme can be understood by analysing the A_0 matrix. Some elements under the principal diagonal are set to zero to allow the introduction of simultaneous feedbacks between demand and supply for money and central bank behaviour and exchange rates. The estimated model is over-identified in that 23 parameters are estimated in the A_0 and B matrix, out of a possible maximum of 28. The over-identifying restrictions are tested and not rejected. The identifying restrictions are rather standard.US economy is taken as exogenous and the exchange rate does not enter in the FED reaction function, US output and prices are not included in the VAR, while a simultaneous feedback is allowed between money demand and supply (the central bank rule). According to this rule, contemporaneous US interest rate movements are relevant to the foreign central bank only if they affect the exchange rate. Only the exchange rate is allowed to contemporaneously react to news in all the other variables.

Unfortunately the coefficients in the A_0 matrix are estimated rather imprecisely. If we consider the case US-Germany the only significant parameters in the matrix are a_{53} and a_{72}. The first parameter is difficult to intepret, given that the identification scheme does not address explicitly macroeconomic shocks, while the point estimate of second parameter implies an appreciation of the dollar against the D-mark in response to a US restrictive monetary policy. The potential simultaneous feedback between foreign monetary policy and the exchange rate does not seem to be empirically relevant. However, all the puzzles disappears and the empirical results for the impulse response functions seem to be broadly in line with results from the US closed economy model. Given that this VAR included some proxy for commodity price index the evidence cannot be decisive on the source of the "puzzles", although the fact that the simultaneous feedback between foreign interest rates and the exchange rate is not significant is consistent with attributing a substantial role to the omisson of commodity prices.

Also in this case the sample considered spans different regime, moreover this methodology brings back into the specification broad monetary aggregates. Interestingly money is now used to extract demand rather than supply shocks, however the specification of money demand implicit in the VAR might not be rich enough to capture the dynamic in the data. As pointed out by Faust and Whiteman(97), single equation work by Hendry and colleagues on money demand has clearly shown the importance of including in the model the opportunity cost of holding money, which is often a spread between the interest rates. Interest spreads capturing the opportunity cost of holding money are never included in VAR models of the MTM. An identification similar to the one adopted by Kim-Roubini is the one proposed for the Canadian case by Cushman-Zha(97), who aid the strucutral identification by introducing explicitly the trade sector into the model.

An interesting alternative approach to the identification of the simultaneous feedback between non US interest rates and exchange rates is proposed by Smets [54],[55]. Smets considers the following structural model for non-US countries:

$$\mathbf{A}_0\mathbf{y}_t = i = 1\sum^{k} \mathbf{A}_i\mathbf{y}_{t-i} + \mathbf{B}\boldsymbol{v}_t \tag{6.23}$$

$$\mathbf{y}_t = \begin{bmatrix} \Delta y_t \\ \Delta p_t \\ R_t \\ \Delta e_t \end{bmatrix}, \ \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & 0 & 0 \\ \frac{-(k_{11}b_{11}+k_{13}b_{31}+k_{14}b_{41})}{k_{12}} & b_{22} & 0 & 0 \\ b_{31} & b_{32} & 1 & 0 \\ b_{41} & b_{42} & b_{43} & b_{44} \end{bmatrix}$$

$$\mathbf{A}_0 = \begin{bmatrix} 1\,0\,0 & 0 \\ 0\,1\,0 & 0 \\ 0\,0\,1-\omega & \omega \\ 0\,0\,0 & 1 \end{bmatrix}$$

where Δy_t is output growth, Δp_t is inflation, R is a short term interest rate and Δe_t is the exchange rate appreciation. No US variables are introduced, and the commodity price index is also excluded. However Smets is more ambitious than average aiming at identifying both macroeconomis and monetary shocks. Three type of restrictions are imposed. first the semi-structural restrictions, macro variables do no react contemporaneously to monetary variables. Second, macroeconomic supply shocks are identified for macroeconomic demand shocks by assuming that the long-run effect of demand shocks on outptut is zero. Third, monetary policy shocks are identified from exchange rate shocks by assuming that the Central Bank reacts proportionally to interest rate and exchange-rate developments (short-term MCI). Macroeconomic shocks are separated into demand and supply shocks by noting that the long-run response of output to a demand shock is given by the element (1,1) of the matrix $\left(A_0 - i = 1\sum^{k} A_i\right)^{-1} B$. Given the block-recursiveness assumption on A_0,the elements of $\left(A_0 - i = 1\sum^{k} A_i\right)^{-1}$ relevant to determine the element (1,1) of $\left(A_0 - i = 1\sum^{k} A_i\right)^{-1} B$ will be not be function of the A_0 matrix and therefore from the estimation of the reduced-form VAR one can retrieve all the elements in the A_i matrix a generate an identifying restriction for the structural parameters in the B matrix by setting the element (1,1) $\left(A_0 - i = 1\sum^{k} A_i\right)^{-1} B$ to zero. In practice, given that $\left(A_0 - i = 1\sum^{k} A_i\right)^{-1} = \begin{bmatrix} k_{11} & k_{12} & k_{13} & k_{14} \\ k_{21} & k_{22} & k_{23} & k_{24} \\ k_{31} & k_{32} & k_{33} & k_{34} \\ k_{41} & k_{42} & k_{43} & k_{44} \end{bmatrix}$,it can be easily shown that $\mathrm{k}_{11}, k_{12}, k_{13}, k_{14}$ are determined independently form the parameters in A_0, therefore restricting to zero the long-run effect of demand shock on output, $\mathrm{k}_{11}b_{11} + k_{12}b_{21} + k_{13}b_{31} + k_{14}b_{41}$, we have $\mathrm{b}_{21} = \frac{-(k_{11}b_{11}+k_{13}b_{31}+k_{14}b_{41}.)}{k_{12}}$.

Lastly in the monetary block, monetary policy shock are identified from exchange rate shocks by assuming that the appropriate indicator of exogenous monetary stance is a short-term MCI where exchange rate and interest rate are appropriately weighted. The weights can be estimated or imposed given the knowledge of the relative weights in several Central Banks MCI's. This approach encompasses as special case the pure interest rate targeting and the pure exchange rate targeting. The main empirical problem with this procedure are the instability of the estimated ω and the potentially disruptive implications of misspecification for the identification of aggregate demand and supply shocks,(see Faust and Leeper[22]on this point).

6.6.1 *Replicating the empirical evidence*

The data-set BERLIN.WF1 contains the relevant variables to replicate the discussed so far on open-economy VAR models.

We estimate first a benchmark open economy model for the US and the German economy without including the Commodity Price Index. The model is estimated on monthly data over the sample 1983:1 1997:12. The VAR is specified by including six lags of US industrial production, US consumer price index, Federal Funds rate, German industrial production,German consumer price index, German call money rate, and the US-dollar/Deutschemark nominal exchange rate (unit of DM for oneUS dollar). The choice of the sample is motivated by the need of having a single monetary policy regime for the US, featuring Fed funds targeting, Bagliano-Favero([2]) . Impulse responses for all variables to US and German monetary policy shocks are reported in the following figures.

Figure 6.7 Impulse responses to US monetary policy shock in the benchmark VAR open-economy model
(dashed lines: 68% confidence interval bands)

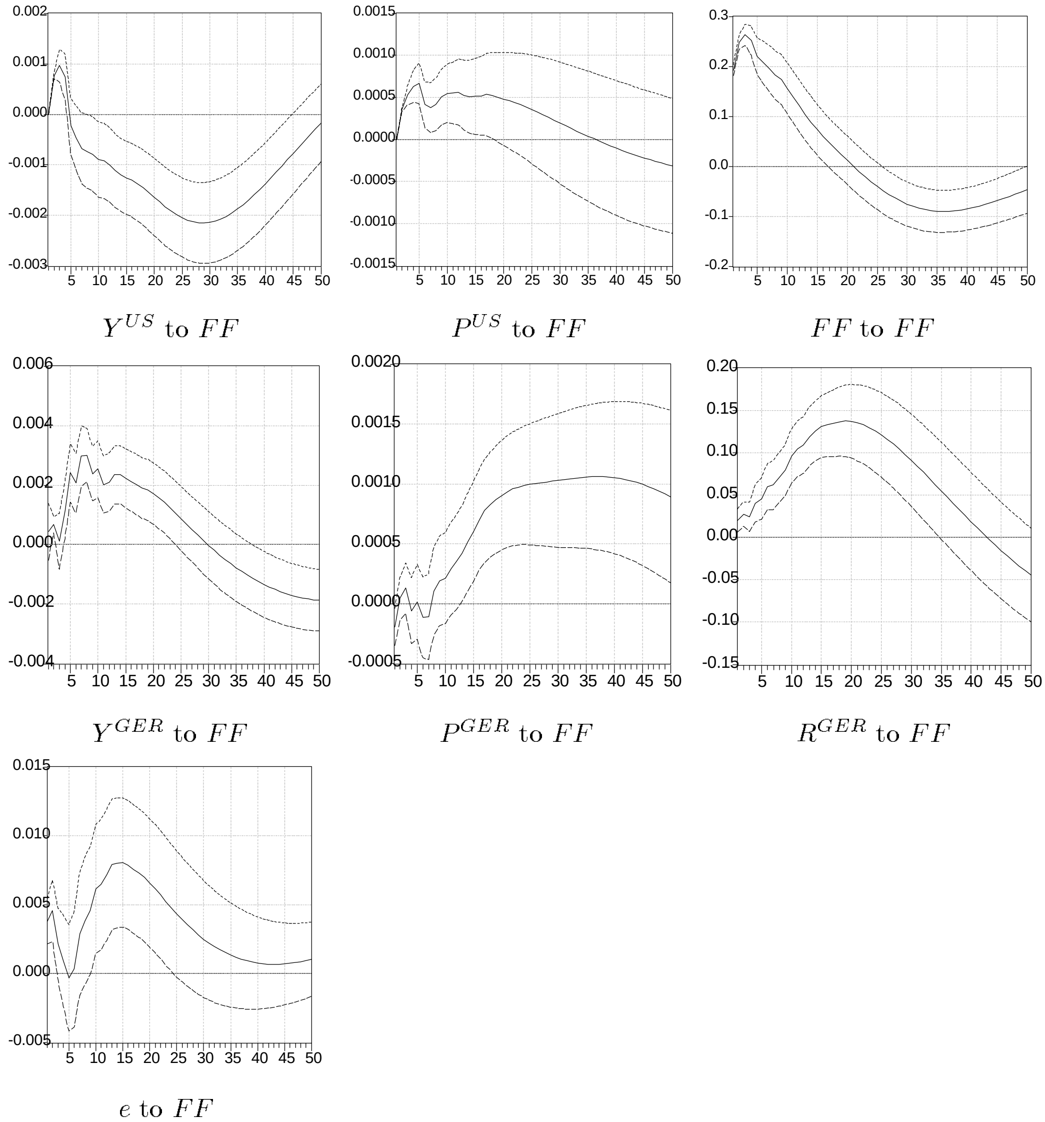
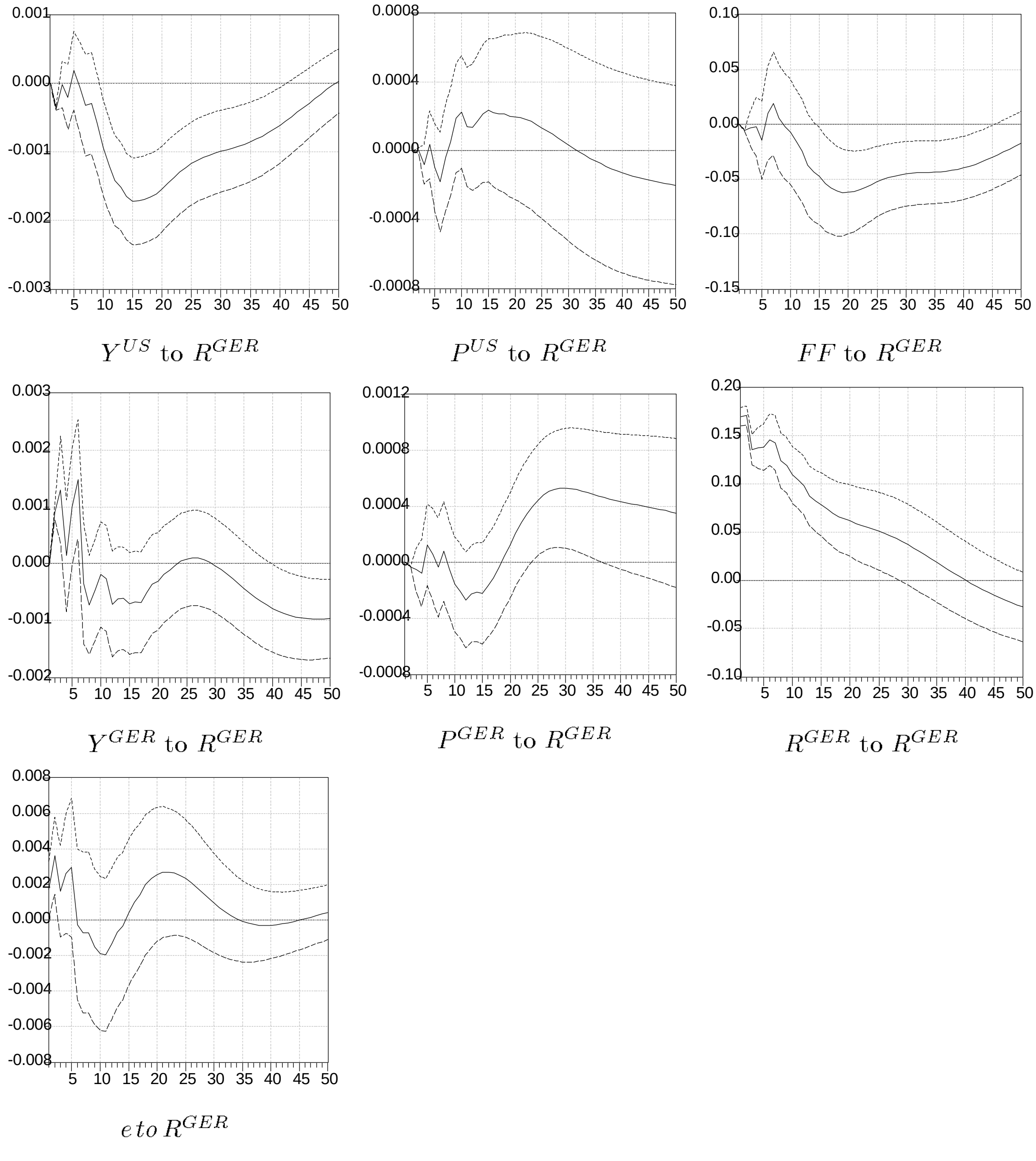


Figure 6.8: impulse responses to German monetary policy shock in the benchmark VAR open-economy model

(dashed lines: 68% confidence interval bands)



We have confirmation of all the facts and puzzles observed in the literature. The analysis of the contemporaneous feedback between variables within the recursive specification provides evidence on the endogeneity of US monetary policy, which reacts significantly to internal conditions, and of the German monetary policy which reacts to both internal conditions and, less significantly to US monetary policy. The exchange rates reacts contemporaneously significantly to US monetary policy (positive interest rate shock in the US induces appreciation of the US dollar vis-a-vis the DM) and to macroeconomic conditions in US and Germany (a positive shock in US industrial production and in German price lead contemporaneously to an appreciation of the US dollar) both it is not contemporaneously significantly affected by German monetary policy.

The analysis of the responses to monetary impulses in the US and Germany confirms all the main findings of the literature namely:

- a significant U-shaped response of US output to US monetary policy
- the existence of a price puzzle both for the US and Germany
- the existence of a forward discount puzzle generated by US monetary policy restriction
- exchange rate puzzle for German monetary policy shock

6.6.2 *Omitted variables*

Our analysis of VAR models of the MTM in close has shown a crucial importance of the Commodity Price Index in the derivation of monetary policy shocks. The arguments made for the inclusion of this variable in close-economy VAR of the MTM are also compelling for open economy VAR. It might be vary well the case that puzzles observed in open-economies are related to mis-specification, via the omission of a commodity price index in the benchmark open-economy VAR. We consider this potential explanation, by concentrating of open economy VAR model linking the US and the German economy.

We then include a commodity price index by keeping the Choleski identification and considering Pcm as a macroeconomic variables influencing both US and German monetary policy. The new impulse responses are reported in figures 6.9 and 6.10.

Figure 6.9: Impulse responses to US monetary policy shock in the benchmark VAR with commodities price index
(dashed lines : 68% interval confidence band)

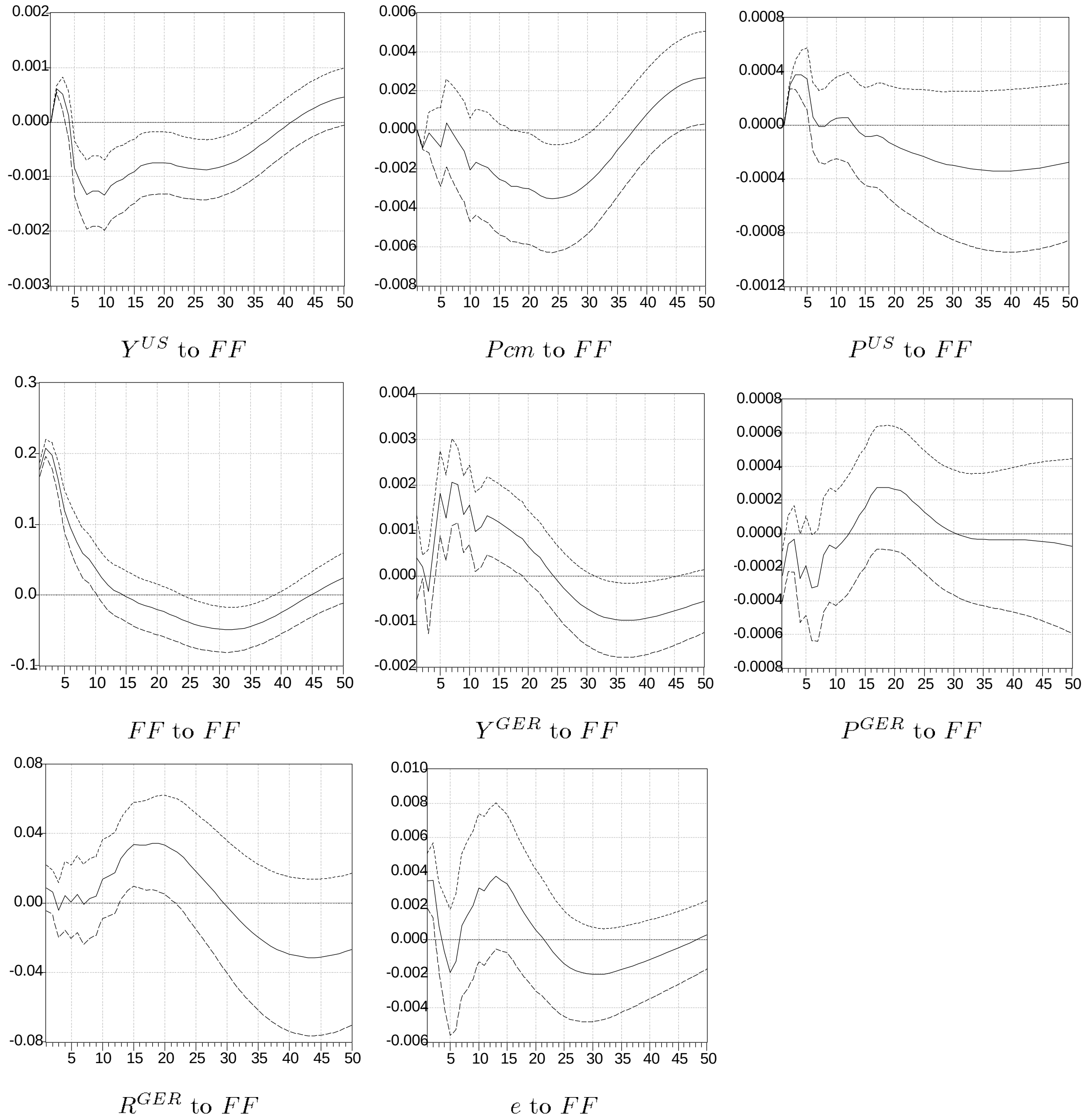
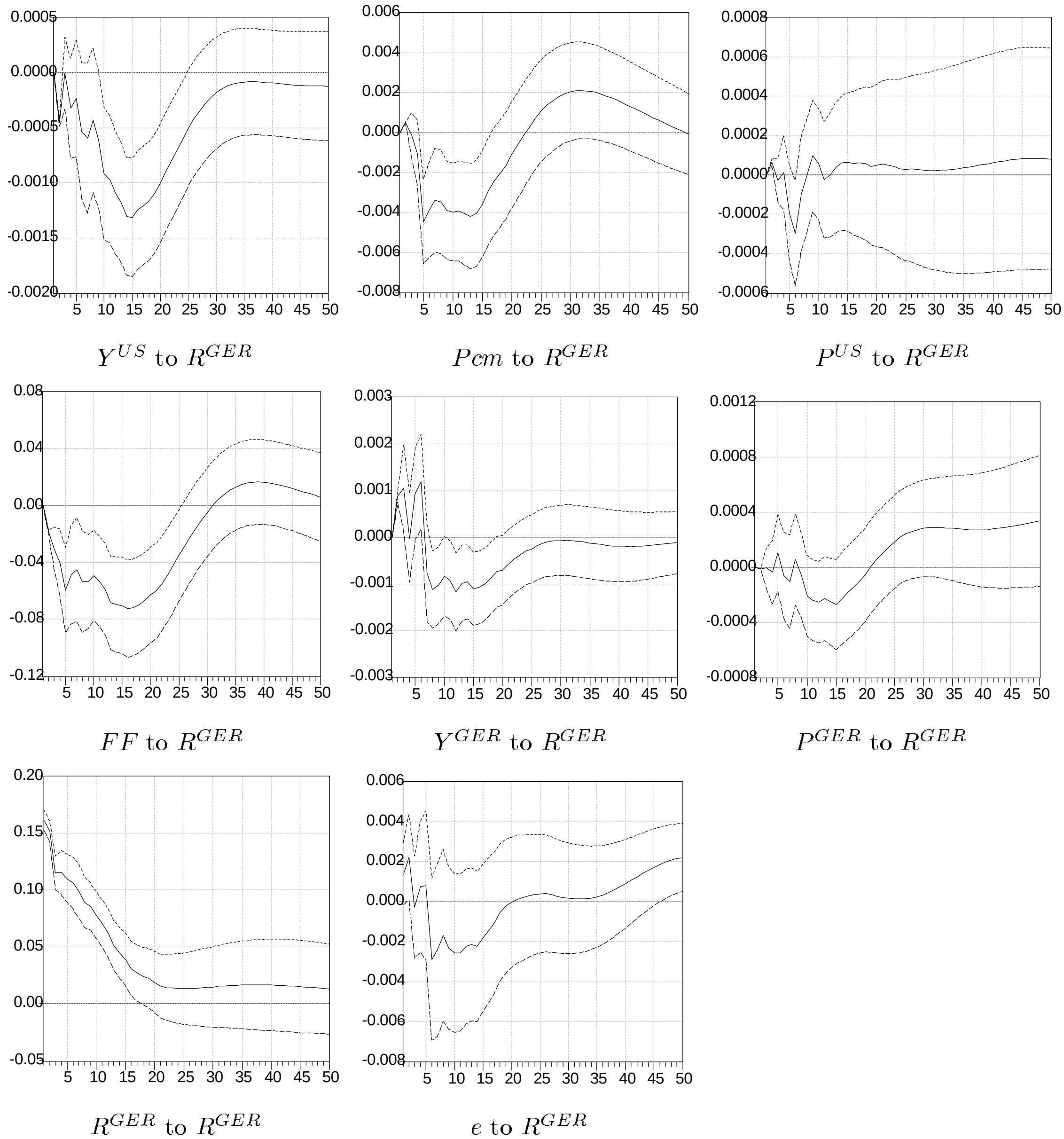


Figure 6.10: Impulse responses to German monetary policy shock in the benchmark VAR with commodities price index

(dashed lines : 68% interval confidence bands)



The results in the two figures show that the inclusion of commodity price solves the price puzzle for both countries, moreover also the forward discount bias puzzle and the exchange rate puzzle tend to disappear. Finally, although we do not observe a symmetric contemporaneous effect of US and German monetary policy on the exchange rate, the impulse response functions of the exchange rate to the two monetary shocks over an horizon of four year show a remarkable degree of symmetry.

Altough the inclusion of commodity prices seems very relevant in fixing many of the empirical problems in open-economy VAR we have open the issue of potential simultaneity between the exchange rate and the policy rate in small open economy and not yet addressed it. We shall consider this issue by looking at the more general problem of assessing the reliability of the measurement of monetary policy with VAR models.

6.7 VAR and non-VAR measures of monetary policy.

Econometric measurement of monetary policy has always been a debated issue. VAR models are linear, constant-parameters autoregressive distributed lag models, bound to include a very limited number of variables with a very parsimonous lag parameterization. The crucial step to derive evidence from the data using VARs, is the possibility of posing identifying restrictions independently from theoretical models. We have illustrated how a consensus has been reached in the case of closed economy and how the same result has not yet been reached for open economies. We have provided an intepretation of this difference in the light of the difficulties in identifying monetary policy shocks in open economies. Recently, VAR based monetary policy shocks have been compared with monetary policy shocks measured by alternative approaches. We think that these developments can be useful not only to evaluate VAR methodologies, but also to help identification when, as in the case of the open economy models, the traditional VAR methods have problems in delivering the necessary number of identifying restrictions.

6.7.1 *Non-VAR measures of monetary policy shocks*

Historically alternative to econometric measurement of monetary policy have been always considered, think for example of qualitative indicators of monetary policy derived adopting the "narrative approach" of Romer and Romer ([44] and [45]). In a recent paper, Leeper [34] shows that even the dummy variable generated by the "narrative approach" (identifying episodes of deliberate monetary contractions) is predictable from past macroeconomic variables, thus reflecting the endogenous response of policy to the economy, and the estimated coefficients cannot provide an unbiased estimate of the response of the macroeconomic variables to a monetary impulse.

Recently the attention of monetary economists has turned to financial markets, which are a potential source of very powerful information and measurement of monetary policy. We shall consider a variety of measures of monetary policy derived from financial market and assess their role in the evaluation of VAR based monetary policy shocks in open and close economies.

A first possible alternative has been originally proposed by Rudebusch [46] and further analyzed by Brunner [9]. Monetary policy shocks are derived from the 30-Day Fed funds future contracts, which have been quoted on the Chicago Board of Trade since October 1988, and are bets on the average overnight fed funds rate for the delivery month, the variable included in benchmark VARs. Shocks are constructed as the difference between the federal funds rate at month t and the 30-day federal funds future at month $t-1$. Such choice is based on the evidence, that the regression of the federal funds rate at t on the 30-day federal funds future at $t-1$ produces an intercept not significantly different from zero, a slope coefficient not significantly different from one, and serially uncorrelated residuals:

$$FF_t = -(0.0436)0.037 + \ (0.007)0.999FFF_{t-1} + \widehat{u}_t$$
$$R^2 = 0.99 \qquad \sigma = 0.145 \qquad DW = 1.86$$

Note that this procedure produces shocks, labelled FFF, which are comparable to the reduced form innovations from the VAR and not to the structural monetary policy shocks, because surprises relative to the information available at the end of month $t-1$ may reflect endogenous policy responses to news about the economy that become available in the course of month t. However if an identification scheme is available, then innovations derived from the future contract can be transformed in the relevant shocks by applying to them the standard VAR identification procedure. A non-trivial problem with this procedure is generated by the fact that Federal Funds future are available on from 1988 onwards, on a more extended sample future on the 1-month euro-dollar are available. Given that the properties of the series generated by 1-month Euro-dollar are very close to the properties of Federal Funds future, the direct measurement based on 1-month Eurodollar could be used on a extended sample.

A second non-VAR measure of policy shocks is based on the work of Skinner and Zettelmeyer [53]. They derive a measure of unanticipated monetary policy shocks by following a two-step methodology: first, using information from central bank reports and newspapers a list of days on which monetary policy announcements occurred is constructed; second, monetary policy shocks are identified with the changes in the three-month interest rate on the days of policy announcements. The validity of such procedure require that (i) short rates (e.g the overnight rate) are affected by policy; (ii) arbitrage is effective between the overnight and the three-month interest rate; (iii) the impact of other news affecting the three-month rate on the day of the policy decision is negligible; (iv) policy actions are not endogenous responses to information that becomes available on the day when the decision is taken. To ensure that conditions (iii) and (iv) are applicable, Skinner and Zettelmeyer go through reports of the policy actions and exclude from their sample those which do no conform to requirement (iii) and (iv). The main problem with the index so obtained is that it can only pin down shocks associated to monetary policy decisions reflected in some action on controlled variables, whereas shocks associated with *no* action (while some action was expected by the markets) are neglected.

An alternative approach which might overcome this problem has been proposed by Bagliano and Favero (1998), applying the methodology set out in Svensson (1994) and Soderlind and Svensson (1997). The methodology is based on the use of instantaneous forward rates as monetary policy indicators. Forward rates are interest rates on investments made at a future date, the settlement date, and expiring at a date further into the future, the maturity date. Instantaneous forward interest rates are the limit as the maturity date and the settlement date approach one another.

To illustrate our derivation of spot rate let us start by the consideration of a *zero-coupon* bond issued at time t with a face value of 1, maturity of m years and price P_{mt}^{ZC}. The simple yield Y_{mt} is related to the price as follows:

$$P_{mt}^{ZC} = \frac{1}{(1+Y_{mt})^m} \tag{6.24}$$

Defining the spot rate r_{mt} as $\log(1+Y_{mt})$, which is the continuously compounded yield, and the discount function D_{mt} as the price at time t of a zero coupon that pays one unit at time $t+m$, we then have :

$$P_{mt}^{ZC} = \exp(-mr_{mt}) = D_{mt} \tag{6.25}$$

Consider now a *coupon* bond that pays a coupon rate of c per cent annually and pays a face value of 1 at maturity. The price of the bond at trade date is given by the following formula:

$$P_{mt} = \sum_{k=1}^{m} cD_{kt} + D_{mt} \tag{6.26}$$

Given the observation of prices of coupon bonds, spot rates on zero coupon equivalent can be derived by fitting a discount function based on the following specification for the spot rates:

$$\begin{aligned} r_{kt} = \beta_0 + \beta_1 \frac{1-\exp\left(-\frac{k}{\tau_1}\right)}{\frac{k}{\tau_1}} & + \beta_2 \left(\frac{1-\exp\left(-\frac{k}{\tau_1}\right)}{\frac{k}{\tau_1}} - \exp\left(-\frac{k}{\tau_1}\right) \right) \\ & + \beta_3 \left(\frac{1-\exp\left(-\frac{k}{\tau_2}\right)}{\frac{k}{\tau_2}} - \exp\left(-\frac{k}{\tau_2}\right) \right) \end{aligned} \tag{6.27}$$

Such specification has been originally introduced by Svensson [58] and it is an extension of the parametrization proposed by Nelson and Siegel [42]. Implied *forward* rates can be calculated from spot rates. A forward rate at time t with trade date $t+t'$ and settlement date $t+T$ can be calculated as the return on an investment strategy based on buying zero-coupon bonds at time t maturing at time $t+T$ and selling at time t zero-coupon bonds maturing at time $t+t'$. The forward rate is related to the spot rate by the following formula:

$$f_{t+T,t+t',t} = \frac{Tr_{T,t} - t'r_{t',t}}{T-t'} \tag{6.28}$$

so the forward rate for a 1-year investment with settlement in 2 years and maturity in 3 years is equal to three times the 3 year spot rate minus twice the two year spot rate. The instantaneous forward rate is the rate on a forward contract with an infinitesimal investment after the settlement date:

$$f_{mt} = \lim_{T\to m} f_{t+T,t+m,t} \tag{6.29}$$

In practice we identify the instantaneous forward rate with an overnight forward rate, a forward rate with maturity one day after the settlement. The relation between instantaneous forward rate and spot rate is then:

$$r_{mt} = \frac{\int_{\tau=t}^{t+m} f_{\tau t} d\tau}{m}$$

or, equivalently

$$f_{mt} = r_{mt} + m\frac{\partial r_{m,t}}{\partial m} \tag{6.30}$$

Given specification (6.27) for the spot rate, the resulting forward function is as follows:

$$f_{kt} = \beta_0 + \beta_1 \exp\left(-\frac{k}{\tau_1}\right) + \beta_2 \frac{k}{\tau_1} \exp\left(-\frac{k}{\tau_1}\right) + \beta_3 \frac{k}{\tau_2} \exp\left(-\frac{k}{\tau_2}\right) \tag{6.31}$$

Therefore as k goes to zero the spot and the forward rate coincide at $\beta_0+\beta_1$ and as k goes to infinity the spot and the forward rate coincide at β_0. The forward rate function features a constant, an exponential term decreasing when β_1 is positive, and two "hump shape" terms. In principle $\beta_0+\beta_1$can be restricted to match the observed overnight rate, but, we do not follow this strategy. A standard practice in the application of this curve-fitting approach is to include the overnight rate in the information set, sometime constraining the fitted overnight rate to match the observed one in estimation. However, a monetary policy shock implies by definition a jump in, at least, the short end of the term structure. Forcing the smooth instantaneous forward rate curve to fit exactly the observed overnight rate would not allow to seize an eventual expected monetary policy action. For this reason, we exclude the overnight rate from the information used for estimation. Then, exploiting the continuity of the functional form, we reconstruct the very short end of the term structure allowing for a gap between the estimated overnight and the observed overnight. Such a gap represents the jump in the very short-end of the term structure associated with expectations of intervention by the central bank.An example can clarify matters. On occasion of the meeting held on the 2nd of December 1993 the Bundesbank reduced the repo rate by 25 basis points. On the close of the markets before the meeting we observed the structure of spot rates relevant to the estimation of our yield curves reported in Table 1

TABLE 1: German Interest rates

date	30/11/93	2/12/93
o/n	6.70	6.35
7-days	6.44	6.31
1-month	6.44	6.31
3-month	6.19	6.06
6-month	5.81	5.75
1-year	5.37	5.25
2-year	5.08	5.03
3-year	5.05	5.02
4-year	5.16	5.15
5-year	5.3	5.29
7-year	5.69	5.68
10-year	6.16	6.17

In figure 11 we report Nelson-Siegel interpolants. More precisely we report the two instantaneous forward curves associated respectively to the spot curve estimated excluding the overnight rate (IFW) and to the spot curve estimated including the overnight rate (IFOY).

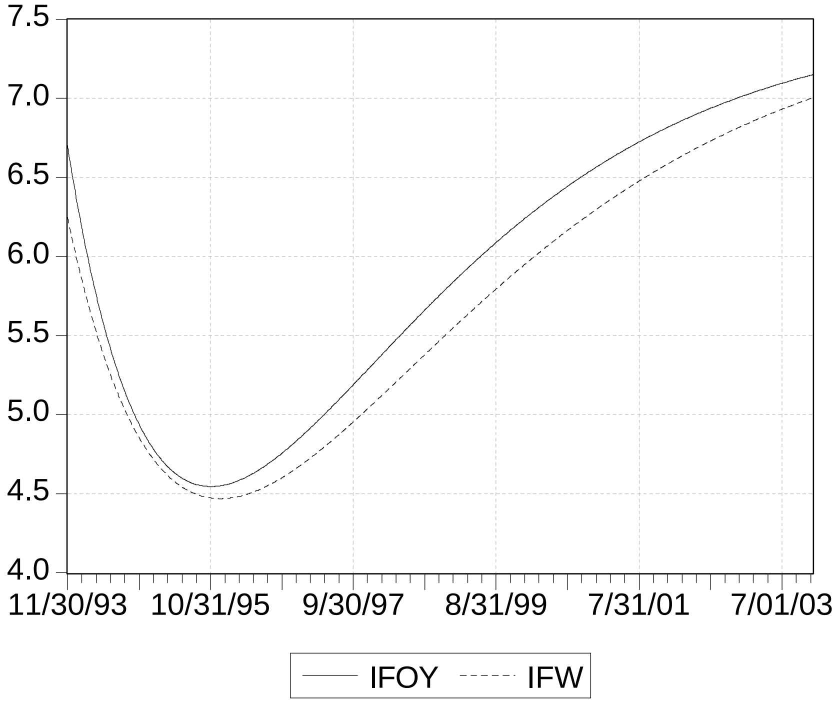


FIG.6.11. Estimated forward-rate curves on the 30/11/93 with and without the overnight rate

fitting the curve on data including the overnight without allowing for a jump in the term structure from the date of the Bundesbank Council meeting afterwards, would have spuriously generated an interest rate shock.

If the pure expectational model is valid and there is no term premium, then instantaneous forward rates at future dates can be interpreted as the expected spot interest rates for those future rates. The observable equivalent of the instantaneous forward rate is the overnight rate.

The following strategy for identification of monetary policy shocks exploits directly the relation between spot rates and instantaneous forward rates

- Exploiting the fact that intervention on policy rates for Germany and US takes place on occasion of regular meetings of the Bundesbank Council and of the FOMC (since 1994), collect data on the term structure of interest rates the day before the monetary policy meetings. Observations on one-day, seven-days rate, 1m euro, 3m euro, 6m euro, 12m euro, 3, 5, 7, and 10-year fixed interest rate swap are available on DATASTREAM and other databases

- estimate a term structure for spot rates and the associated curve of instantaneous ratea

- interpret the instantaneous rate as the overnight rate, and derive from the curve the expected implicit overnight rate for the day after the monetary policy meeting

- derive monetary policy shocks subtracting from the observed overnight rate the day after the policy meeting the overnight rate implicit in the curve the day before the policy meeting

- aggregate the above daily measures (concentrated in a few special days) to construct monthly measures of shocks

There are several difficulties that one should overcome in constructing this measure of monetary policy shocks. Following [3], we illustrate examples of monetary shocks generated by unanticipated *action* or by unanticipated *inaction* by the Bundesbank, and examples of markets' anticipation of Bundesbank behaviour when expectations on monetary policy turned out to be correct and no shocks were observed.We consider the sample 1984-1997.

Consider first July 1988. In this month the Bundesbank Council met twice,on the 14th and on the 28th. During the first Council the Bundesbank didn't take any action, during the second the Council it was decided to raise the Lombard Rate by 50bp. In Figure 6.12 we report the the weekly and the overnight rate, alongwith the monetary policy action(PMA).

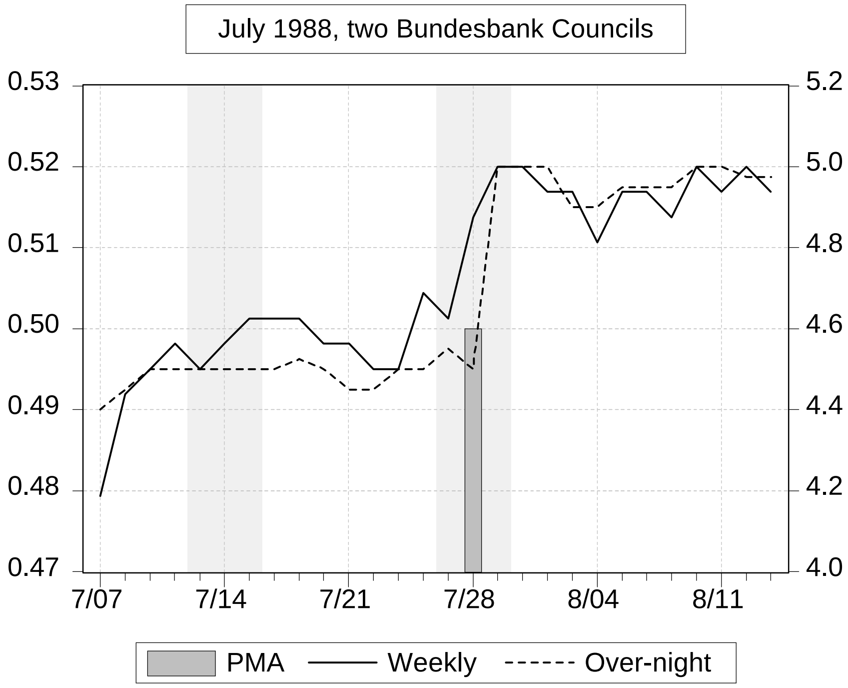


FIG.6.12.Monetary policy interventions and short-term interest rates in Germany. July 1988

We shade areas of three days centered around meetings. We note that no monetary policy action was expected during the first meeting, while some action was expected before the second one. Six days before the meeting the weekly rate contains the first six days of maturity which doesn't include the action and the seventh one which instead does include the action, so the weekly rate should start to "reflect" the monetary policy action six days before the meeting. Of course the weight of the seventh day is one-seventh so the information doesn't appear clearly six days before, but as we approach the date of the council the weight of the action becomes greater and the expectation discloses itself. It can be observed that the weekly rate starts reacting three days before the meeting. It is also possible than the market realizes that the Bundesbank will act only a few days before the Council (say less than six days before), in this case the weekly rate starts reacting later than six days before the Council. The weekly rate should be the best observed interest rate to identify expections on monetary policy actions. In fact Council meetings take place fortnightly and the 1-month rate immediately before any meeting reflects expectations on the outcome of the following two meetings.

The second episode we consider is the tightening of monetary policy occurred after German reunification in January-February 1991. Two meetings were held in this period, the 17th of Janauary and the 2nd of February. As Figure 6.13 clearly shows, the weekly rate increased sharply just before the first Council revealing an expected increase in the interest rates.

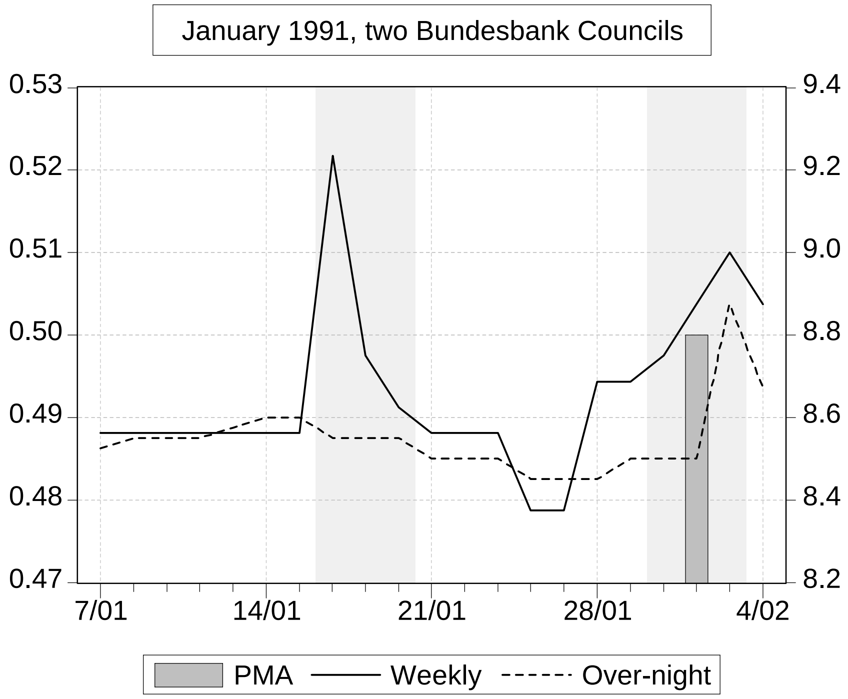


FIG.6.13.Monetary policy interventions and short-term interest rates in Germany. January 1991

The Bundesbank didn't act on that meeting.We immediately observe than the expected tightening happened during the following Council meeting, when the Bundesbank raised the Discount Rate and the Lombard Rate by 50 bp. To summarize, on the fourteenth of January we observed a monetary policy shock arising from an anticipated action that did't occur, meanwhile on the second of February there is no shock as the policy move has been correctly anticipated.

The third episode we single out occurred in December 1991(see Figure 6.14) when the Bundesbank tightened the monetary policy, raising once again the Discount Rate and the Lombard Rate by 50 bp.The dates of the Bundesbank Councils are the 5th and the 19th of December. During the latter meeting the Bundesbank surprised the market, and we observe a shock arising from an unexpected policy action.

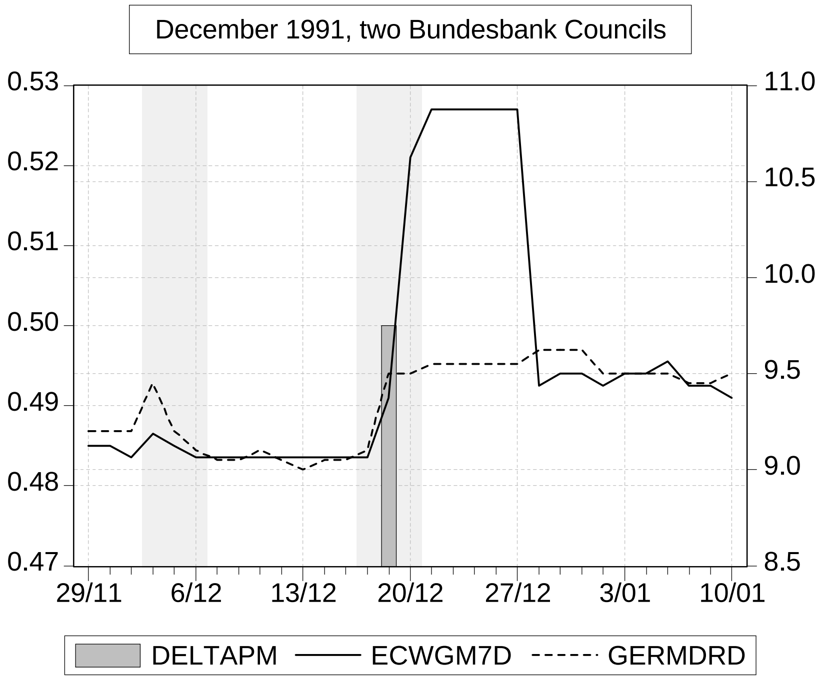


FIG.6.14.Monetary policy interventions and short-term interest rates in Germany. December 1991

The main strength of the methodology based on foward interest rate curves is its flexibility and its capability to capture shocks independently from the specification and parameterization ofa linear autoregressive model. The main limitation of this approach is caused by the volatility of very short-term rates not related to expectations on monetary policy. Figure 6.15-6.16 reports daily observations on the over-night rate and the weekly rate for the estimation sample period used in the VAR.

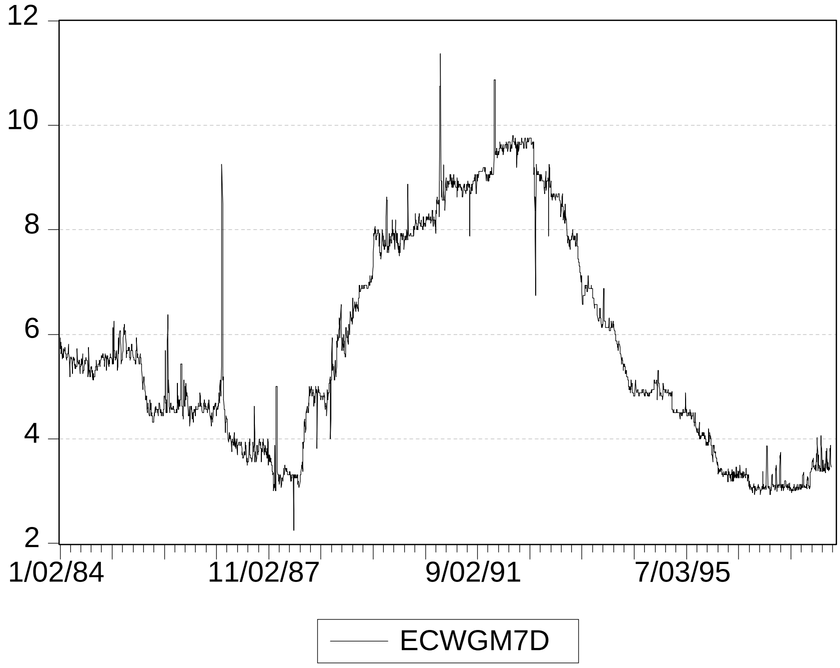


FIG. 6.15.The German 7-days rate

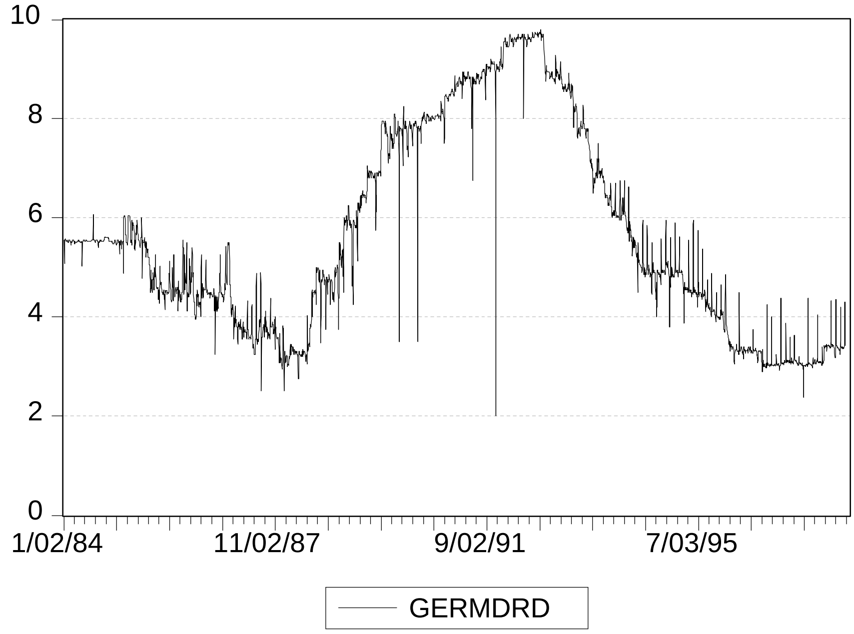


FIG. 6.16 The German overnight rate

We immediately note a number of blips in the series. Those blips could be very damaging to our methodology whenever they happen on occasion of a Bundesbank Council meeting. Most of those blips are generated by banks reserves management which run into a non perfecly liquid markets, such as on the occasion of the last day of the average reserves maintenaince period. We make an effort to render our inference robust to blips. In fact, we have estimated our curves starting from the 7-day rather than the overnight rate, and our methodology of estimation considers the information contained in the whole term structure. However, we have run a further check and avoid to label as policy shocks all unexpected movements in policy rates which have disappeared within a week after the Council Meeting. Such correction led us to single out two outliers in 1988:9 and 1991:12. The 1988:9 outlier, whose determination is described in Figure 13, is the only one of a relevant magnitude.

In Figure 6.17 we report the behaviour of the 7-days and the 1-month rate in the course of September 1988. No policy intervention was decided in September 1988, however just before the meeting of mid September we observe an hike in the 7-days rate. Such hike is not reflected in the term structure for longer maturities (we report 1-month for reference). This hike would have been labelled as a shock by the methodology, however, as it is reversed, within the week after the meeting this episode should be considered as a monetary policy shock.

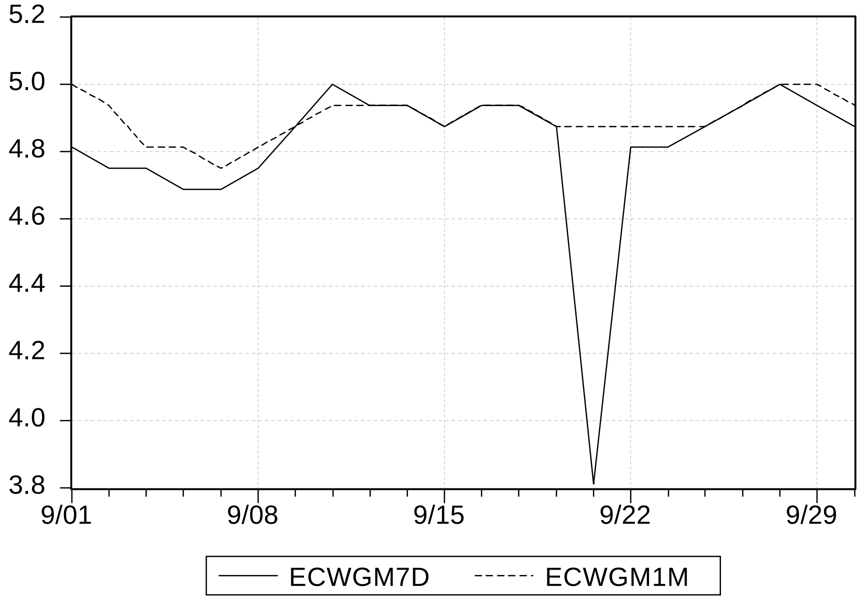


FIG. 6.17. The German 7-days and one-month rate in September 1988

6.8 Empirical results

Non-VAR measures of monetary policy can be directly compared with VAR measures, they can also be used to assess the robustness of the VAR based descriptions of the monetary transmsission mechanism, finally, they can be exploited, within a VAR, to aid identification of other structural shocks. To illustrate these possibilities we consider in turn the close economy(US) case and the open-economy(US-Germany) case.

6.8.1 *Close economy (US)*

To evaluate the role of non-*VAR* based measures of monetary policy shock, we first estimate the close-economy four-variable version of the *VAR* model for the US and compute impulse response functions of all variables to a shock in the Federal funds rate. Our model is specified as follows:

$$\mathbf{A}\begin{pmatrix} Y_t^{US} \\ Pcm_t \\ P_t^{US} \\ FF_t \end{pmatrix} = \mathbf{C}(L)\begin{pmatrix} Y_{t-1}^{US} \\ Pcm_{t-1} \\ P_{t-1}^{US} \\ FF_{t-1} \end{pmatrix} + \mathbf{B}\begin{pmatrix} \nu_t^1 \\ \nu_t^2 \\ \nu_t^3 \\ \nu_t^{FF} \end{pmatrix} \tag{6.32}$$

$$\tag{6.33}$$

where **A** is lower triangular and **B** is diagonal.

The ordering chosen allows for a contemporaneous response of the policy rate to innovations in output, consumer prices and the commodity price level. The orthogonalized residual of the Federal funds rate equation, ν^{FF}, is identified as a monetary policy shock. No structural interpretation is given to the (orthogonalized) residuals from the other equations in the system. We then consider two non-VAR measures of monetary policy: that derived from one-month Eurodollar forward rate ($EUR\$$) and that derived from the estimation of the instantaneous forward rate curve on occasion of FOMC meetings(IFS^{US}). These alternative shocks are plotted with the VAR based shocks in Figures 6.18-6.19.

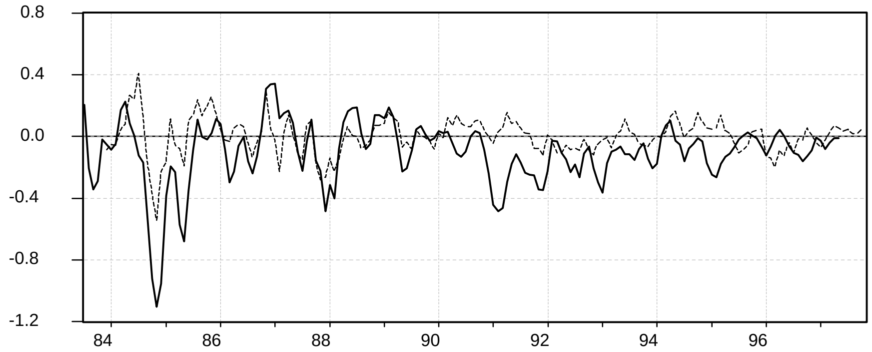


FIG.6.18.Three-month centered moving averages of $EUR\$$ shocks (solid line) and close economy VAR monetary policy shocks (dotted line)

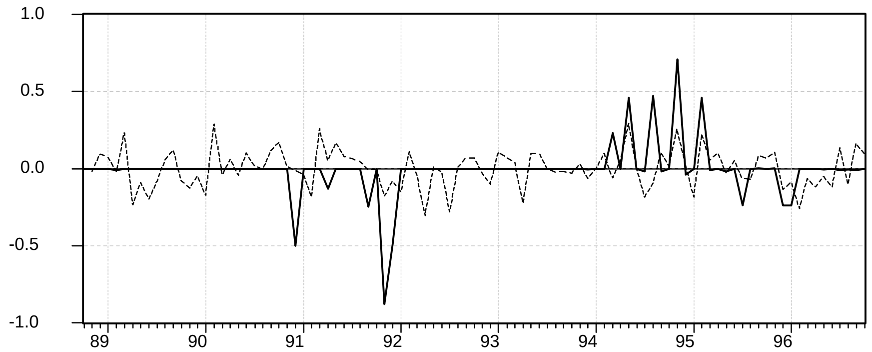


FIG.6.19.IFS^{US} shocks (solid line) and close economy VAR monetary policy shocks (dotted line)

Note that the $EUR\$$ measure is available on a larger sample than the IFS^{US} measure as the practice of modifying monetary policy rates on occasion of given and announced dates started only in the nineties. We report in Table 2 the correlations of VAR and non-VAR measures of monetary policy.

TABLE 2: VAR and non-VAR monetary policy shocks
Sample: 1988(11)-1996(10)

Correlation coefficients (standard errors on the diagonal)

	$EUR\$$	IFS^{US}	ν^{FF}
$EUR\$$	0.185		
IFS^{US}	0.203	0.169	
ν^{FF}	0.352	0.319	0.123

Sample: 1988(11)-1996(10)

	$EUR\$$	ν^{FF}
$EUR\$$	0.277	
ν^{FF}	0.500	0.211

Rudebusch[46] using the Federal Funds future contract obtains very similar results to those obtained in our shorter-sample to conclude that VAR based measured of monetary policy do not make sense. We note that much better results are obtained in the enlarged sample. To provide further evidence we specify a VAR augmented by the non-VAR meausure of monetary policy shocks, considered as an exogenous variable.Following Amisano and Giannini (1996), we represent the estimated system as follows:

$$\mathbf{A}\begin{pmatrix} Y_t^{US} \\ Pcm_t \\ P_t^{US} \\ FF_t \end{pmatrix} = \mathbf{C}(L)\begin{pmatrix} Y_{t-1}^{US} \\ Pcm_{t-1} \\ P_{t-1}^{US} \\ FF_{t-1} \end{pmatrix} + \begin{pmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \end{pmatrix} x_t + \mathbf{B}\begin{pmatrix} \nu_t^1 \\ \nu_t^2 \\ \nu_t^3 \\ \nu_t^{FF} \end{pmatrix} \tag{6.34}$$

$$x = EUR\$, IFS^{US} \tag{6.35}$$

where $\mathbf{A}$ is lower traingular and $\mathbf{B}$ is diagonal. The estimated values of the coefficients g_i are reported in Table 3

TABLE 3:Coefficients on $EUR\$$ and IFS^{US} in the benchmark VAR

Sample 1988(11)-1997(11)

	Y^{US}	Pcm	P^{US}	FF
$EUR\$$	(0.0032)0.0061	(0.0121)0.0055	(1.0633)0.0013	**(0.097)0.468**
IFS^{US}	(0.0031)0.0025	(0.0116)0.0082	(0.0013)0.0009	**(0.099)0.356**

Sample: 1984(1)-1997(11)

	Y^{US}	Pcm	P^{US}	FF
$EUR\$$	(0.0016)0.0026	(0.0006)0.0007	(0.0063)0.0058	**(0.062)0.552**

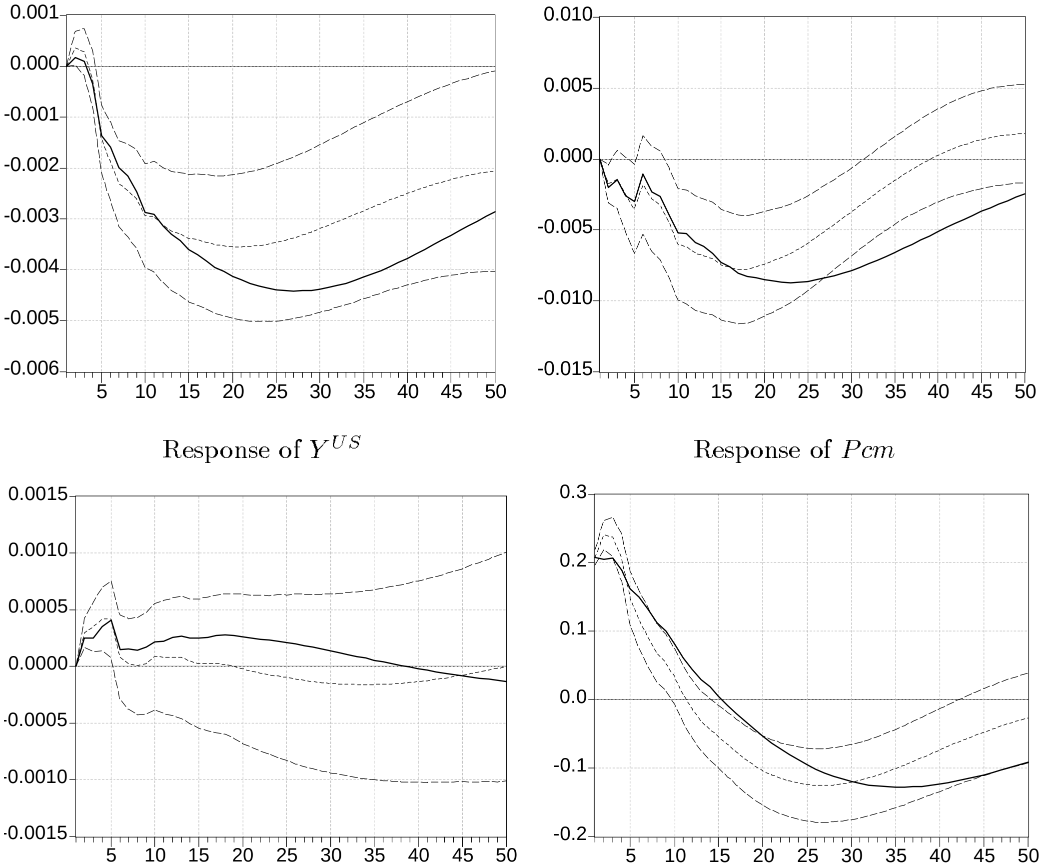
We note that none of the macroeconomic variables responds to the non-VAR monetary policy shocks, while the Federal Fund rates does. As suggested by the correlations between shocks results are stronger on the larger sample. We then concentrate on this sample and compare impulse responses to monetary policy shocks in the traditional benchmark VAR specification with impulses responses to non-VAR monetary policy shocks in our augmented specification.

The results, shown in Figure 6.20, illustrate that a contractionary monetary policy shock produces the expected negative effect on output and a persistent effect on the Federal funds rate.

The inclusion of the commodity price index is successful in solving the price puzzle: the consumer price level does not show a perverse response to restrictive policy.

The pairs of impulse response functions, based on the VAR and the non-VAR shocks, describe a very similar transmission mechanism, supporting the evidence already provided by Brunner (1996) and Bagliano and Favero (1998) with different exogenous measures. Despite a correlation of 0.5 between $EUR\$$ and the measure of policy shock obtained from the benchmark VAR, the dynamic effects of monetary policy show very close features: both measures capture unexpected variations in the policy rate related to monetary policy and the existence of other non-policy disturbances does not change the basic features of the response to a policy shock.

Figure 6.20
Impulse responses to alternative U.S. monetary policy shocks in close economy
Responses to $EUR\$$ shocks (solid line) and to VAR-based structural shocks ν^{FF} (dotted line) with one standard deviation confidence intervals from the benchmark VAR



Response of Y^{US} Response of Pcm

Response of P^{US} Response of FF

6.8.2 *Open economy (US-Germany)*

Let us now consider the open-economy version of the *VAR* system. We begin by a baseline specification which includes the commodity price index:

$$\mathbf{A}\begin{pmatrix} Y_t^{US} \\ Pcm_t \\ P_t^{US} \\ FF_t \\ Y_t^{GER} \\ P_t^{GER} \\ e_t \\ R_t^{GER} \end{pmatrix} = \mathbf{C}(L)\begin{pmatrix} Y_{t-1}^{US} \\ Pcm_{t-1} \\ P_{t-1}^{US} \\ FF_{t-1} \\ Y_{t-1}^{GER} \\ P_{t-1}^{GER} \\ e_{t-1} \\ R_{t-1}^{GER} \end{pmatrix} + \mathbf{B}\begin{pmatrix} \nu_t^1 \\ \nu_t^2 \\ \nu_t^3 \\ \nu_t^{FF} \\ \nu_t^5 \\ \nu_t^6 \\ \nu_t^e \\ \nu_t^{RGER} \end{pmatrix} \tag{6.36}$$

where $\mathbf{A}$ is lower triangular and $\mathbf{B}$ is diagonal.

As we have done for the close economy case, we compare orthogonalized residual of the German call money rate equation (ν^{RGER}) with the non-VAR measure of German monetary policy shocks IFS^{GER}, derived from instateneuos forward rates. Figure 6.21 and Table 4 confirms the results for correlations obtained in the close-economy case.

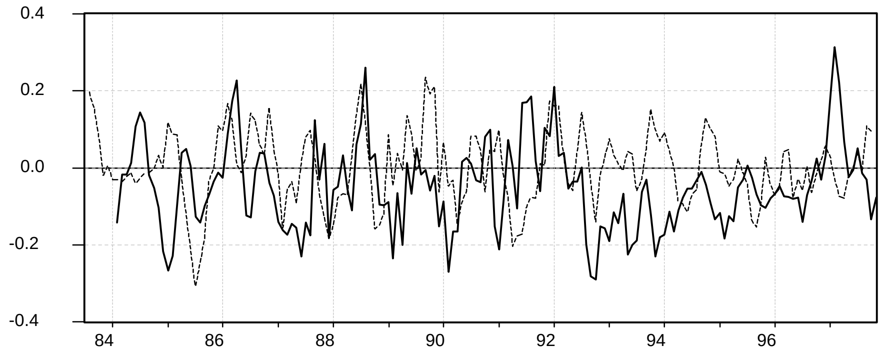


FIG.6.21Three-month centered moving averages of IFS^{GER} shocks (solid line) and open economy *VAR* German monetary policy shocks (dotted line)

TABLE 4: VAR and non-VAR monetary policy shocks
Sample: 1984(1)-1997(11)

	IFS^{GER}	ν^{RGER}
IFS^{GER}	0.194	
ν^{RGER}	0.163	0.169

As in the close-economy case, we augment the previously estimated system by including the exogenous measure of German monetary policy shocks IFS^{GER} described in the previous section.

The open-economy *VAR* is now the following:

$$\mathbf{A}\begin{pmatrix} Y_t^{US} \\ Pcm_t \\ P_t^{US} \\ FF_t \\ Y_t^{GER} \\ P_t^{GER} \\ e_t \\ R_t^{GER} \end{pmatrix} = \mathbf{C}(L)\begin{pmatrix} Y_{t-1}^{US} \\ Pcm_{t-1} \\ P_{t-1}^{US} \\ FF_{t-1} \\ Y_{t-1}^{GER} \\ P_{t-1}^{GER} \\ e_{t-1} \\ R_{t-1}^{GER} \end{pmatrix} + \begin{pmatrix} g_1 \\ g_2 \\ g_3 \\ g_4 \\ g_5 \\ g_6 \\ g_7 \\ g_8 \end{pmatrix} IFS_t^{GER} + \mathbf{B}\begin{pmatrix} \nu_t^1 \\ \nu_t^2 \\ \nu_t^3 \\ \nu_t^{FF} \\ \nu_t^5 \\ \nu_t^6 \\ \nu_t^e \\ \nu_t^{RGER} \end{pmatrix} \quad (6.37)$$

Using our exogenous measure of monetary policy shocks in combination with a Choleski ordering with the German policy rate coming last, we are able to directly address the issue of simultaneity between German monetary policy and the exchange rate. The contemporaneous effect of a monetary policy shock on the exchange rate is given by the coefficient on IFS^{GER} in the exchange rate equation (g_7), while the response of the German interest rate to innovations in the exchange rate is endogenized by the ordering chosen. As shown in Table 5, the simultaneous relations, we do not observe a significant contemporaneous feedback between the German interest rate and the exchange rate in any direction. In our framework, this is a testable proposition rather than an assumed identified restriction. We note that our measure of monetary policy shocks enters significantly in the German policy rate equation and that the contemporaneous response of U.S. output to German monetary policy shocks is small but marginally significant.[13]

TABLE 5:Coefficients on IFS^{GER} and simultaneous responses of e in the benchmark *VAR*

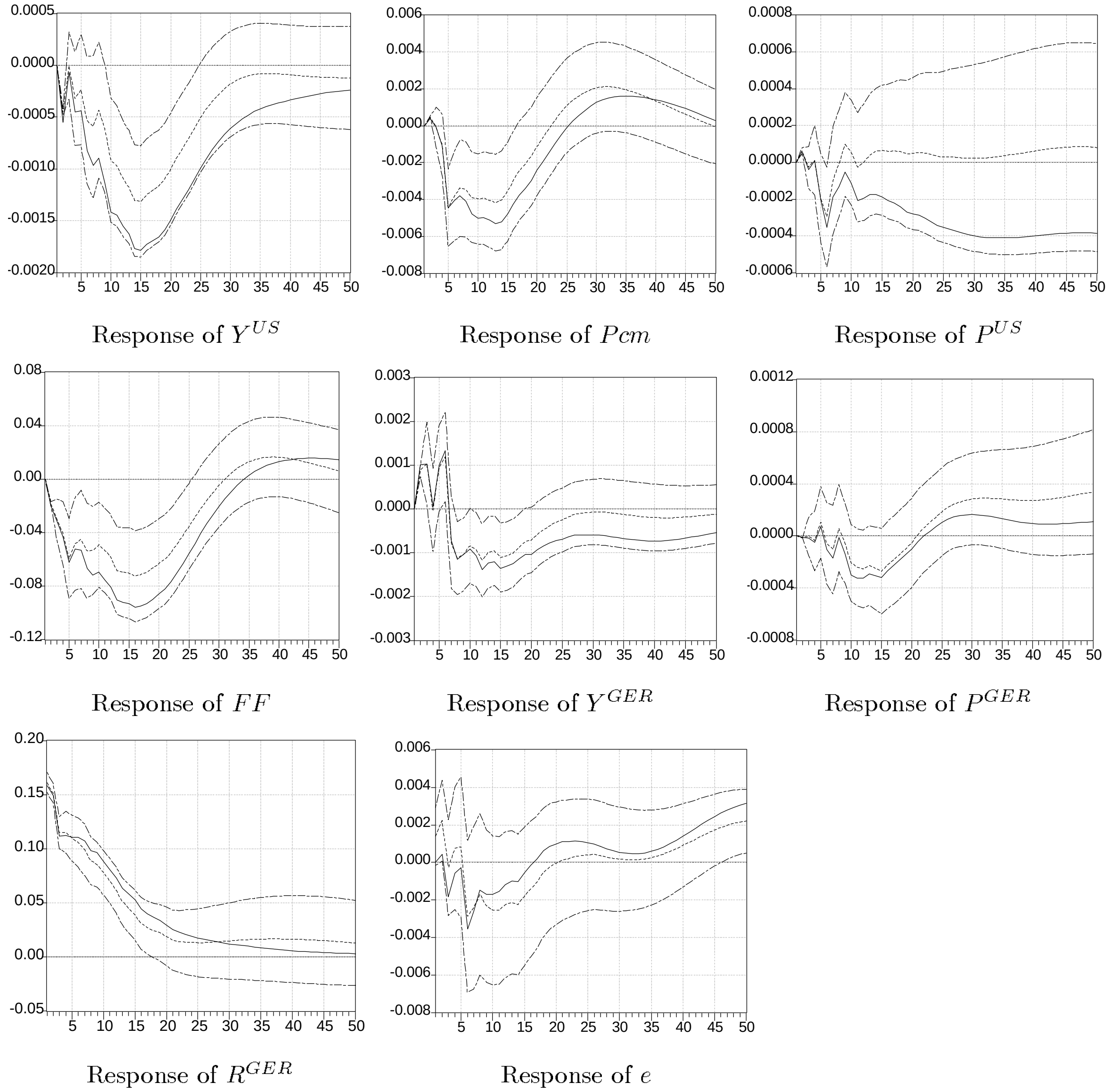
	Y^{US}	Pcm	P^{US}	FF	Y^{GER}	P^{GER}
IFS^{GER}	(0.002)−0.007	(0.008)−0.01	(0.0008)−0.0013	(0.1146)−0.0892	(0.0011)0.00002	(0.007)(
e	**(0.037)1.36**	(0.11)−0.15	(1.09)−0.15	**(0.0083)0.022**	(0.126)0.045	**(0.79)2**

The pair of impulse response functions shown in Figure 6.22, alongwith one-standard deviation bands, confirm qualitatively the results obtained for the close US economy: measuring monetary policy shocks using financial market data does not alter the main features of the monetary transmission mechanism for Germany.

[13] We report impulse responses based on restricting such coefficient to zero; relaxing this restriction does not affect the shape and magnitude of the impulse responses.

Figure 6.22
Impulse responses to alternative German monetary policy shocks in open economy

Responses to IFS^{GER} shocks (solid line) and to VAR-based structural shocks ν^{RGER} (dotted line) with one standard deviation confidence intervals from the benchmark VAR



Response of Y^{US} — Response of Pcm — Response of P^{US}

Response of FF — Response of Y^{GER} — Response of P^{GER}

Response of R^{GER} — Response of e

6.9 Conclusions

The VAR approach to the monetary transmission mechanism is aimed at the derivation of stylized facts to help the selection of the theoretical model to be used for simulating the effects of monetary policy. The identification of parameters in these type of models does not allow to separate deep parameters describing taste and technology from expectational parameters dependent on policy regimes. However, by estimating this type of models on a single policy regime and by describing the responses of variables to structural shocks of interest, it is hoped to derive some stylized facts on the monetary transmission mechanism. Such stylized facts should help the selection of the relevant theoretical model to be used for policy simulation. We have intepreted the choice of concentrating on shocks as a consequence of the impossibility of identifying deep parameters independently from expectational parameters. Unfortunately, structural shocks are not directly observable and the imposition of a set of identifying restrictions is a necessary prerequisite for the analysis. Given the aim of the analysis, it is essential that identifying restrictions are posed independently from specific theories. All the developments of the Choleski ordering that we have discussed in the chapter provide the researcher with tools for achieving this aim. In particular, we have shown how informations from financial markets can be used both to assess robustness of results derived within traditional VAR models of the monetary transmission mechanism and to aid identification in cases when traditional analysis does not deliver a sufficient number of restrictions. Within this framework for analysis it is also natural that the number of identifying restrctions is kept at a minimum. VAR models of the monetary transmission mechanism are very rarely cointegratd VARs. We have seen that multivariate cointegration analysis requires the solution of a long-run identification problem, and that imposing cointegrating restrictions on a VAR in levels increases efficiency in the estimator at the cost of the risk of inconsistency when the incorrect identifying restrictions are imposed. The monetary transmission mechanism is a short-run phenomenon and this explains why researchers prefer to employ unrestricted VAR to evaluate impulse response analysis over a short to medium horizon. Cointegrated VAR are however an almost inevitable choice when the relevant, theory neutral restrictions, are long-run restrictions.

As VAR models are the natural empirical counterparts of dynamic general equilibrium monetary models, their statistical adequacy is not as closely scrutinized as the adequacy of reduced form specification within the LSE approach. In particular, in many of the applications outliers are not removed and non-normality is not an uncommon feature. Parameters stability is also an issue in the debate. As far as identification is concerned, the idea of using restrictions neutral with respect to the theories under scrutiny is nice but not always implementable. In fact, VAR models of the monetary transmission in open economies have not been as successful in establishing stylized facts, probably because of the difficulties in generating a "neutral" identification scheme. Moreover our analysis of the empirical evidence on the monetary transmission mechanism has shown a high level of uncertainty associated with VAR based results. In fact, rather large standard errors are associated to the point estimates of impulse response functions. The more so in the case of VAR in open economies, where practitioners have developed the habit of reporting one-standard deviation bands rather than two standard deviations bands. The main consequence of such uncertainty is that the aim of the exercise, once again model selection, is difficult to achieve in practice.

REFERENCES

Amisano G. and C.Giannini(1996) ”Topics in Structural VAR Econometrics” Springer-Verlag

Bagliano F-C, C.A. Favero(1998) “Measuring Monetary Policy with VAR models: an evaluation”, *European Economic Review*, 42,6, 1069-1112

Bagliano F-C, C.A. Favero(1999) “Information from financial markets and VAR measures of monetary policy”, *European Economic Review*, 43,6, 825-838

Bernanke B.S. and A. Blinder (1992) “The Federal Funds Rate and the channels of monetary transmission”, *American Economic Review,* 82, 901-921

Bernanke B.S. and I. Mihov (1998) “Measuring monetary policy”, *Quarterly Journal of Economics, 113, 3, 869-902*

Bernanke B.S. and I. Mihov (1996) “What Does the Bundesbank Target?”, *NBER* working paper n.5764

Bernanke B.S. and I.Mihov(1998) ”The liquidity effect anf long-run neutrality”, NBER working paper n.6608

Blanchard O.J. and D.T.Quah(1989) ”The dynamic effects of aggregate demand and supply disturbances”, American Economic Review, 79, 655-73

Brunner A. (1996) “Using measures of expectations to identify the effects of a monetary policy shock”, Board of Governors of the Federal Reserve System, International Finance Discussion Paper n.537

Campbell J. (1995) “Some Lessons from the Yield Curve”, *Journal of Economic Perspectives*, 9, 3, 129-152

Christiano L.J. and M. Eichenbaum (1992) “Liquidity Effects and the Monetary Transmission Mechanism”, *American Economic Review*, 82, 2, 346-353

Christiano L.J., M. Eichenbaum and C.L. Evans (1996a) “The Effects of Monetary Policy Shocks: Evidence from the Flow of Funds”, *Review of Economics and Statistics*, 78, 16-34

Christiano L.J., M. Eichenbaum and C.L. Evans (1996b) “Monetary Policy Shocks and their Consequences: Theory and Evidence”, paper presented at ISOM 1996

Christano L.J., M.Eichenbaum and C.Evans (1998) ”Monetary Policy shocks: what have we learned and to what end?” NBER working paper 6400

Clarida R. and J.Gali (1994) ”Sources of real exchange rate fluctuations: how important are nominal shocks ?” Carnegie-Rochester Conference Series on Public Policy 41:1-56

Clarida R. and M.Gertler(1996) ”How the Bundesbank conducts monetary policy”, NBER working paper 5581

Cochrane J.(1994) ”Permanent and Transitory Components of GNP and Stock Prices”, Quarterly Journal Of Economics,241-267

Cushman D.O.and T.Zha(1997) "Identifying monetary policy in a small open economy under flexible exchange rates" Journal of Monetary Economics, 39, 433-448

Doornik J. and D.F. Hendry (1996) *PcFIML: Interactive econometric modelling of dynamic system*, International Thompson Publishing, London

Eichenbaum M. and C.Evans (1995) "Some Empirical Evidence on the effects of monetary policy on real exchange rates" Quarterly Journal of Economics, 110:975-1009

Farmer R.E.A. (1997) "Money in a Real Business Cycle Model", mimeo, Dept. of Economics, UCLA

Faust J. and E.M.Leeper(1997) "When do long run identifying restrctions give reliable results?." Journal of Business and Economic Statistics, 15, 3, 345-353

Faust J. and C.H.Whiteman(1997) "General-to-Specific procedures for fitting a data-admissible, theory inspired, congruent, parsimonious, encompassing, weakly exogenous, identified, strucutral model to the DGP: A translation and critique." Board of Governors of the Federal System, International Finance Discussion Paper 576

Favero C.A., F. Giavazzi and L.Spaventa "High-yields. The spread on German interest rates", The Economic Journal, 107, 956-985

Favero C.A., M. Pifferi and F. Iacone (1996) "Monetary Policy, Forward Rates and Long Rates: Does Germany Differ from the United States?", *CEPR* Discussion Paper n.1456

Goodfriend M. and R.King (1997) "The Newclasical Synthesis and the Role of Monetary Policy", paper presented at the 12th NBER Annual Macroeconomic Conference

Gordon D. and Leeper E.M. (1994) "The Dynamic Impacts of Monetary Policy: an Exercise in Tentative Identification", *Journal of Political Economy*, 102, 1228-1247

Grilli V. and Roubini N.(1996) "Liquidity and Exchange Rates: Puzzling Evidence from the G-7 Countries", mimeo, Yale University

Grilli V. and Roubini V.(1996) "Liquidity Models in Open Economies: Theory and Empirical Evidence" European Economic Review, 4, 847-859

Hamilton J. (1994) Time-Series Analysis, Princeton, Princeton University Press

Hendry D.F. (1996) *Dynamic Econometrics*, Oxford University Press, Oxford

Kim S. and N.Roubini(1997) "Liquidity and Exchange Rates in the G-7 Countries: Evidence from identified VARs", mimeo

Leeper E.M., C.A. Sims and T. Zha (1996) "What does monetary policy do?", available at ftp://ftp.econ.yale.edu/pub/sims/mpolicy

Leeper E.M.(1997) "Narrative and VAR approaches to monetary policy: common identification problems", *Journal of Monetary Economics*,

Lucas R.E. Jr. (1972) "Expectations and the Neutrality of Money", *Journal of Economic Theory*, 4, April, 103-124

Lucas R.E. Jr. (1976) "Econometric Policy Evaluation: A Critique", in K. Brunner and A. Meltzer (eds.) *The Phillips curve and labor markets,* Amsterdam, North-Holland

Mc Callum B.(1994) "A reconsideration of the uncovered interest parity relationship", *Journal of Monetary Economics,* 33:105-132

Mc Callum B.(1999) "Analysis of the monetary transmission mechanism:methodological issues" paper presented at the Deutsche Bundesbank conference "The monetary transmission process: recent developments and lessons for Europe", Frankfurt, March 1999

Mc Callum B. and E.Nelson(1999) "An optimizing IS-LM specification for monetary and business cycle analysis" Journal of Money Credit and Banking, 31

Mellander E., Vredin A. and Warne A.(1993) "Stochastic trends and economic fluctuations in a small open economy" Journal of Applied Econometrics

Mosconi R.(1998) "Malcolm. The theory and practice of cointegration analysis in RATS",Cafoscarina, Venice

Nelson C.R. and A.F. Siegel (1987) "Parsimonious modelling of yield curves", *Journal of Business,* 60, 473-489

Obstfeld M. and K.Rogoff(1996) "Foundations of International Economics", MIT Press, Cambridge

Romer C D. and D.H. Romer (1989) "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz", in O.J. Blanchard and S. Fischer (eds.), *NBER Macroeconomics Annual 1989*, Mit Press, 121-170

Romer C. D. and D.H. Romer (1994) "Monetary Policy Matters", *Journal of Monetary Economics,* 34, 75-88

Rudebusch G.D. (1996) "Do Measures of Monetary Policy in a VAR Make Sense?", Temi di Discussione n.269, Bank of Italy

Schlagenhauf D.E., Wrase J.M.(1995) "Liquidity and Real Activity in a simple open economy model" *Journal of Monetary Economics* 35:431-461

Sims C.A. (1980) "Macroeconomics and Reality", *Econometrica*, 48, 1-48

Sims C.A.(1980) "Comparison of interwar and postwar business cycles: monetarism reconsidered", American Economic Review, May, 250-257

Sims C.A. (1992) "Interpreting the Macroeconomic Time-Series Facts: the Effects of Monetary Policy", *European Economic Review*, 36, 975-1011

Sims C.A., J. H. Stock and M.Watson (1990) "Inference in linear time-series models with some unit roots", *Econometrica*, 58, 113-144

Sims C.A. and Tao Zha (1996) "Does Monetary Policy Generate Recessions ?", mimeo, available at ftp://ftp.econ.yale.edu/pub/sims/mpolicy

Skinner T. and J. Zettelmeyer (1996) "Identification and effects of monetary policy shocks: an alternative approach", mimeo, MIT, september

Smets F.(1996) "Measuring Monetary Policy in the G7 countries: interest rates versus exchange rates" paper presented at CEPR-Banca d'Italia workshop on "Model Specification, Identification and Estimation in Empirical Macroeconomics

Smets F.(1997) "Measuring Monetary Policy Shocks in France, Germany and Italy: the role of the exchange rate" BIS working paper no.42

Soderlind P and Svensson L.E.O.(1997) "New techiques to extract market expectations from financial instruments" Journal of Monetary Economics, 40, 383-42

Strongin S. (1995) "The identification of monetary policy disturbances. Explaining the liquidity puzzle", *Journal of Monetary Economics*, 35, 463-497

Svensson L.E.O. (1994) "Estimating and interpreting forward interest rates: Sweden 1992-1994", *CEPR* Discussion Paper n.1051

Uhlig H.(1997) "What are the effects of monetary policy ? Results from an agnostic identification procedure. http://cwis.kub.nl/~few5/center/staff/uhlig/home.htm

7

INTERTEMPORAL OPTIMISATION AND THE GMM METHOD

7.1 Introduction

The intertemporal optimisation approach to macroeconomic thoery takes the Lucas critique very seriously and is based on the convinction that questions like "How should a central bank respond to shocks in macroeconomic variables?" are to be answered within the framework of quantitative monetary general equilibrium models of the business cycle.

The evaluation of the effects of monetary policy is a question for theoretical models rather than for empirical ad-hoc macroeconometric models. We have seen that VAR based empirical evidence helps the selection of the relevant theoretical model. However, once the model selection problem has been solved, two are the relevant issues: parameterization and simulation. In this chapter we mainly concentrate on parameterization while we shall devote the next chapter to simulation and policy evaluation. The intertemporal optimisation approach takes no interest in the parameters estimated by traditional macroeconometric modelling. In fact, traditional structural econometric modelling delivers parameters which are convolutions of the interesting "deep" parameters describing tastes and technology and of expectational parameters, which are dependent on the specific policy regime.

Interestingly, the intertemporal optimisation-Rational Expectations paradigm generates in a very natural way an alternative econometric approach to estimate deep parameters of interest: the Generalized Methods of Moments (GMM).

This chapter is devoted to the illustration of the empirically relevant aspect of the GMM methodology. We shall consider applications to consumer's behaviour and the estimation of monetary policy rules. We shall start by illustrating the close relationship between the econometric methodology and the intertemporal optimisation achieved by the implementation of the GMM method in the estimation of Euler equations. We shall then consider technically the definition of the estimator, the problems related to the estimation of the covariance matrix and the inference within GMM models.

Having considered the technical aspects of the estimator, we evaluate its success in the new-classical camp and its extremely rare utilization by Keynesian macroeconomist by giving an econometric interpretation to a statement by Mankiw, Rotemberg e Summers [23] who assert that

"... *The major difference between modern neoclassical and traditional Keynesian macroeconomic theories is that the former regard observed levels of employ-*

ment, consumption and output as realizations from dynamic optimizing decisions by both households and firms, while the latter regard them as reflecting constraints on households and firms..."

We shall conclude by showing applications of the GMM approach to the estimation of deep parameters describing (representative) consumer behaviour and to the estimation of deep parameters describing central banks' preferences in monetary policy rules.

7.2 Euler Equations and "closed form solutions":

Consider the standard optimisation problem for the representative consumer:

$$c_{t+i}, A_{t+i} Max E_t i = 0 \sum^{\infty} (1+\delta)^{-i} U(c_{t+i}) \tag{7.1}$$

subject to the following constraints

$$A_{t+i} = (1+r) A_{t+i-1} + y_{t+i} - c_{t+i}$$
$$i \rightarrow \infty \lim E_t A_{t+i} (1+r)^{-i} = 0$$

Where y is disposable labour income, c is consumption of non-durable goods and services, A is wealth (a single financial asset) giving a return of r, U is a utility function featuring both intertemporally and intra-temporal separability. All variables are expressed in real terms. The δ parameter describe the rate of the intertemporal preferences of the representative consumer, who has an infinite horizon and does not face liquidity constraints of any form.Therefore, she can run negative balances of A in any period with the only constraint that the present discounted value of wealth in time t approaches zero as t approaches infinite (transversality condition).Lastly, by E_t we denote expectation formed conditionally upon the information set available at time t.

We solve the intertemporal optimisation problem by finding the maximum of the following Lagrangean function:

$$c_{t+i}, A_{t+i} Max E_t i = 0 \sum^{\infty} (1+\delta)^{-i} G_{t+i} \tag{7.2}$$

$$G_{t+i} = U(c_{t+i}) + \lambda_{t+i} (A_{t+i} - (1+r) A_{t+i-1} - y_{t+i} + c_{t+i})$$

We assume that the real return is non-stochastic and that the utility function is Constant Relative Risk Aversion (CRRA):

$$U\left(c_{t+i}\right)=\frac{c_{t+i}^{1-\gamma}}{1-\gamma} \tag{7.3}$$

where the γ parameter describes the consumer's risk aversion. This specification completes the setup of our problem. We note that there are two parameters which describes tastes and could be defined as deep in Lucas's terminology:γ and δ.

First order conditions for optimality could be stated as follows:

$$E_t\left(c_{t+i}^{-\gamma}\right)-E_t\lambda_{t+i}=0 \qquad \forall i \tag{7.4}$$

$$E_t\lambda_{t+i}-E_t\left(\frac{1+r}{1+\delta}\lambda_{t+i+1}\right)=0 \qquad \forall i$$

By eliminating the Lagrange multipliers and by considering the specific case of i=0, we obtain

$$E_t\left(\frac{1+r}{1+\delta}c_{t+1}^{-\gamma}-c_t^{-\gamma}\right)=0 \tag{7.5}$$

Some consideration on equation (7.5) , known as the Euler equation, are in order. First, this relation clearly confutes the idea that economic theory gives mainly predictions on the long-run behaviour of economic variables: in fact, the Euler equation imposes restriction on the short-run dynamics of economic variables. Second, the only parameters entering equation(7.5) are γ and δ, the "deep" parameters describing consumer's preferences. Third, as equation(7.5) does not represent the "closed form solution" of the intertemporal optimisation problem but just the first order condition for optimality, it cannot interpreted as a consumption function. However, from (7.5) we derive the falsifiable proposition that, under the joint hypothesis of Rational Expectations, the only significant variable in predicting consumption at time t+1 given the inforamation available at time t is consumption at time t. Under our hypotheses the logarithm of consumption behaves as a "random walk" (Hall, [17]). The conditional expectation for time t+1 taken at time t of the expression between brackets in (7.5) is zero, and such expression is orthogonal to any other variable than consumption included in the agent's information set at time t. Labelling f_{t+1} the expression between brackets in (7.5)we have:

$$E_t f_{t+1}=0, \quad E_t f_{t+1}\mathbf{z}_t=0 \tag{7.6}$$

where z is a vector containing any economic variable observable at time t.

Note that the Euler equation does not have any implication for the relation between consumption and other economic variables: the significance of income in explaining contemporaneous fluctuations in consumption is perfectly compatible with our intertemporal optimization model, which only rules out the significance of income at time t in explaining the difference between the marginal utility of savings at time t for time t+1 (if consumer does not consume one unit at time t, she will invest it in the financial asset and have (1+r) unit at time t+1, the discounted value of this quantity being $\frac{1+r}{1+\delta}$)and the marginal utility of consumption at time t. Finally, note that if the rate of intertemporal preference and the interest rate are equal, fluctuations in consumption are determined exclusively by stochastic shocks. To further illustrate the relationship between Euler equation and consumption function and to provide a firmer background to our discussion of econometric methodology we take advantage of the simplicity of our specification to derive analytically a closed form solution to our intertemporal optimisation problem. To simplify matters even further consider "certainty equivalence".

From the first order conditions we derive the following relationship between consumption at time t and consumption in any period following t:

$$c_{t+i} = c_t \left(\frac{1+r}{1+\delta} \right)^{\frac{i}{\gamma}} \tag{7.7}$$

Aggegate now over time the period budget constraint and impose the transversality condition to obtain:

$$i = 0 \sum^{\infty} \frac{c_{t+i}}{(1+r)^i} = i = 0 \sum^{\infty} \frac{y_{t+i}}{(1+r)^i} + (1+r) A_{t-i} \tag{7.8}$$

By using (7.7)in (7.8)to substitute consumption in all future period with an expression in terms of current consumption, we obtain :

$$c_t = (\rho - 1) \left(i = 0 \sum^{\infty} \frac{y_{t+i}}{(1+r)^i} + (1+r) A_{t-i} \right) \tag{7.9}$$

where $\rho = \frac{(1+\delta)^{\frac{1}{\gamma}}}{(1+r)^{\frac{1}{\gamma}-1}}$, and ρ is assumed to be greater than one. When $\delta = r$ expression (7.9)simplifies drastically in the following

$$c_t = r \left(i = 0 \sum^{\infty} \frac{y_{t+i}}{(1+r)^i} + (1+r) A_{t-i} \right) \tag{7.10}$$

Equation (7.10)represents the closed form solution to the intertemporal optimisation problem and it is the structural consumption function for our representative consumer under the hypothesis of certainty equivalence. Note that consumption is function of permanent income, which includes current income, and that the reaction of consumption to the modifications in the real interest rate depends on an income effect and on a substitution effect. The sign of the income effect depends on the consumer's financial position: if the consumer is in debt ($A_{t-1} < 0$) the income effect is negative, while for a consumer in credit ($A_{t-1} > 0$) the income effect is positive. The substitution effect is always negative, as an increase in the interest rate lowers the discounted stream of future income.

The closed form solution is useful to understand the skepticism of newclassical economists towards the use of empirical ad-hoc structural macroeconometric models to simulate the effect of macroeconomic policy. To see the point quickly[14] let us re-write relation (7.9)omitting "perfect foresight":

$$c_t = (\rho - 1)\left(i = 0\sum^{\infty} \frac{E_t y_{t+i}}{(1+r)^i} + (1+r) A_{t-1}\right) + \varepsilon_t \tag{7.11}$$

$$\varepsilon_t = (\rho - 1)\, i = 0\sum^{\infty} \frac{(y_{t+i} - E_t y_{t+i})}{(1+r)^i}$$

in order to interpret in the light of (7.11)traditional ad-hoc macroeconometric models, which do not usually explicitly incorporates expectations, we need to solve for future income in terms of current income. We do so by assuming a simple autoregressive process for income:

$$y_t = a_1 y_{t-1} + u_t \tag{7.12}$$

using repeatedly (7.11)in (7.11)we have:

$$c_t = (\rho - 1)(1+r) A_{t-1} + (\rho - 1)\frac{1+r}{1+r-a_1} y_t + \varepsilon_t \tag{7.13}$$

Parameters in (7.13)are convolutions of the deep parameters contained in ρ and of the expectational parameter a_1, which will change every time the process generating income is subject to modification. Moreover, given the estimation of (7.13) ,the structural parameters describing consumer's tastes, δ and γ, are even not identifiable. Note also that the residual term in (7.13)is, by construnction, autocorrelated. If we can represent autocorrelation in the following, simple, manner

$$\varepsilon_t = \delta \varepsilon_{t-1} + v_t \tag{7.14}$$

then the best representation of the Data Generating Process will then be the following:

[14] This amounts to a little cheating, which simplifies matter greatly without having any substantial effect on our final conclusions.

$$\Delta c_t = (\rho - 1)(1 + r)\Delta A_{t-1} + (\rho - 1)\frac{1 + r}{1 + r - a_1}\Delta y_t - \tag{7.15}$$
$$- (1 - \delta)\left(c_{t-1} - (\rho - 1)(1 + r)A_{t-2} - (\rho - 1)\frac{1 + r}{1 + r - a_1}y_{t-1}\right) + v_t$$

which, obviously, is an Error Correction Mechanism (ECM).

It is clear that when the income generating process is constant a specification like (7.15)will perform extremely well in fitting the data. Note that such specification can be obtained without any reference to the theoretical intertemporal optimisation approach, being derived by LSE type econometric specification search within the class of ECM representations of cointegrating regressions. However, if the Data Generating Process is the one postulated by the intertemporal optimization theory, then the estimated model cannot be used for policy simulation. No empirical question involving simulating the impact on consumption of different policies determining the income process can be meaningfully answered on the basis of the estimation of a model like (7.15) . In fact the estimated parameters are function of the parameters in the income process and they become misleading if the interesting policy to be simulated implies a change in the income generating process. Within the intertemporal optimisation framework the answer to interesting policy question has to be based on the theoretical model rather than on an empirical ad-hoc macroeconometric model. Therefore the impact of different policy on consumption is to be based on the direct simulation of alternative processes for income within the framework of the theoretical model (7.11). Obviously, to implement meaningfully this approach, some estimate of the parameter ρ and hence of the parameters δ and γ, which describes appropriately the preferences of consumers are needed. Now, the Euler equation qualifies immediately as the best relations to be estimated empirically for the identification of the parameters of interest. In fact, it allows identification of the parameters of interest and it does not depend at all on the expectational parameters. Moreover, it allows by its nature the implementation of an estimator: the Generalised Method of Moments. We devote the next section to the econometric analysis of this estimator.

7.3 Estimating Euler equations: The GMM method.

Generlizing the results of the specific problem discussed in the previous section we can represent the first order condition from a generic intertemporal optimisation problem as follows:

$$E_t [f(\mathbf{x}_{t+i}, \boldsymbol{\theta}) \mathbf{z}_t] = 0 \tag{7.16}$$

where $\boldsymbol{\theta}$ is the (px1) vector containing the parameters of interest and $\mathbf{z}$ is the (nx1) vector of variables that theory suggests orthogonal to $f(\mathbf{x}_{t+i}, \boldsymbol{\theta})$. In our example $\boldsymbol{\theta} = (\delta, \gamma)$, $f(\mathbf{x}_{t+i}, \boldsymbol{\theta}) = \left(\frac{1+r}{1+\delta}c_{t+1}^{-\gamma} - c_t^{-\gamma}\right)$and $\mathbf{z}_t$contains any variables observable at time t other than consumption.

It is intuitively clear that a necessary condition for identification of parameters of interest is n≥p, with overidentification in the case of strict inequality and just-identification in the case of equality. If n<p, then the parameters of interest are not identified. Let us concentrate on the over-identification case, of which just-identification is a special case. This is going to be the relevant case in many economic example, as deriving Euler equations from intertemporal optimisation and Rational Expectations usually selects a potentiallly infinite number of valid instruments. Think of our application to the consumer problem: any lagged variable is a valid instrument under the null that the rational-expectations/intertemporal optimisation model is the Data Generating Process.

The estimator is "naturally"derived from (7.16)by substituting population moments with sample moments:

$$\frac{1}{T}t = 1\sum^{T} \left[f(\mathbf{x}_{t+i}, \boldsymbol{\theta})\,\mathbf{z}_t\right] = 0 \tag{7.17}$$

where T is the size of the available sample. Obviously, in case of over-identification (7.17)produces a system of n equations in p unknowns, which does not admit a unique solution. This problem is solved by considering p linear combinations of the n first order conditions and therefore by minimising the "Euclidean distance" between $\frac{1}{T}t = 1\sum^{T} \left[f(\mathbf{x}_{t+i}, \boldsymbol{\theta})\,\mathbf{z}_t\right]$ and the null vector. This implies solving the following minimisation problem:

$$\theta\min\left(t = 1\sum^{T}\left[f(\mathbf{x}_{t+i}, \boldsymbol{\theta})\,\mathbf{z}_t\right]\right)' \mathbf{A}\left(t = 1\sum^{T}\left[f(\mathbf{x}_{t+i}, \boldsymbol{\theta})\,\mathbf{z}_t\right]\right) \tag{7.18}$$

where A is a an appropriate (nxn) weighting matrix. By defining a (Txn) matrix $F(\mathbf{x}_{t+i}, \mathbf{z}_t, \boldsymbol{\theta})$,with typical element $f(\mathbf{x}_{t+i}, \boldsymbol{\theta})\, z_{jt}$, where j=1,...n, t=1,...,T ,the minimisation problem can be re-written as

$$\theta\min i' F(\mathbf{x}_{t+i}, \mathbf{z}_t, \boldsymbol{\theta})\,\mathbf{A}F(\mathbf{x}_{t+i}, \mathbf{z}_t, \boldsymbol{\theta})\, i$$

where i is a (Tx1) identity vector. It can be shown [7] that any symmetric positive definite matrix A will yield a consistent estimate of the vector of parameters of interest. However, Hansen [18] has shown that a necessary (but not sufficient) condition to obtain an symptotically efficient estimate of θ is to set A equal to the inverse of the covariance matrix of the sample moments. The intuition behind this choice is simple: less weight is put on the more imprecise conditions. Therefore if $\Psi = Var\left(t = 1\sum^{T}\left[f(\mathbf{x}_{t+i}, \boldsymbol{\theta})\,\mathbf{z}_t\right]\right)$,GMM estimates are obtained by solving the following minimisation problem:

$$\theta \min \left(t = 1 \sum^{T} \left[f\left(\mathbf{x}_{t+i}, \boldsymbol{\theta} \right) \mathbf{z}_t \right] \right)' \Psi^{-1} \left(t = 1 \sum^{T} \left[f\left(\mathbf{x}_{t+i}, \boldsymbol{\theta} \right) \mathbf{z}_t \right] \right) \tag{7.19}$$

Note that in general, as $\Psi$is function of $\boldsymbol{\theta}$,it would be necessary to proceed in at least two step. In fact, exploiting the fact that any arbitrary weigthing matrix will deliver consistent estimates of $\boldsymbol{\theta}$. This vector of parameters is estimated first, then a $\hat{\Psi}$ is constructed and the minimization (7.19) is then performed. Of course, the two-step procedure is easily extended to an iterative procedure.

Hansen [18] has shown that the minimised criterion function can be also used to test the validity of instruments in case of over-identification, in fact the quantity:

$$J = \left(t = 1 \sum^{T} \left[f\left(\mathbf{x}_{t+i}, \hat{\boldsymbol{\theta}} \right) \mathbf{z}_t \right] \right)' \hat{\Psi}^{-1} \left(t = 1 \sum^{T} \left[f\left(\mathbf{x}_{t+i}, \hat{\boldsymbol{\theta}} \right) \mathbf{z}_t \right] \right) \tag{7.20}$$

is distributed as a χ^2 with n-p degrees of freedom. The quantity (7.20)is known in the literature as the J statistic. GMM is a very general class class of estimators and many of the known estimators can be set up as special cases of GMM. Consider for example the Generalised Instrumental Variables Estimator.

The relevant problem is to estimate the vector of unknown parameters $\boldsymbol{\beta}$ in the linear model :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{7.21}$$

where y is a (Tx1) vector of observations on the dependent variable, $\mathbf{X}$ is a (Txp) matrix of observations on the explanatory variables, $\boldsymbol{\beta}$ is the (px1) vector of parameters of interest, and $\mathbf{u}$ is the (Tx1) vector of observations on the error term with zero mean and variance-covariance matrix equal to $\sigma^2 I$. Assume that $\mathbf{X}$ are not weakly exogenous for the estimation of the parameters of interest, we have then:

$$p \lim \frac{1}{T} \mathbf{X}'\mathbf{u} \neq \mathbf{0} \tag{7.22}$$

However, there exists a $\mathbf{Z}$ matrix containing T observations on n valid instruments, for which we have :

$$p \lim \frac{1}{T} \mathbf{Z}'\mathbf{u} = \mathbf{0} \tag{7.23}$$

Condition (7.23), which defines instruments as valid, gives also a set of orthogonality restrictions to construct a GMM estimate. Let us concentrate on the overidentification case, where n>p. Applying formula (7.19) , the relevant estimate is derived by solving the following problem:

$$\beta Min \left(\mathbf{u}'\mathbf{Z}\Psi^{-1}\mathbf{Z}'\mathbf{u}\right) \tag{7.24}$$

where the appropriate choice for the matrix Ψ is:

$$\Psi = \mathbf{E}\left(\mathbf{Z}'\mathbf{u}\mathbf{u}'\mathbf{Z}\right) = \sigma^2\mathbf{Z}'\mathbf{Z} \tag{7.25}$$

Therefore, the GMM estimate will minimize the following criterion:

$$\beta Min \left(\frac{1}{\sigma^2}\mathbf{u}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{u}\right) \tag{7.26}$$

which admits GIVE as the solution :

$$\hat{\beta}= \left(\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{y} \tag{7.27}$$

Similarly, the J-statistic will take the following form:

$$J = \frac{\hat{\mathbf{u}}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\hat{\mathbf{u}}}{s^2} \tag{7.28}$$

where $\hat{\mathbf{u}}= \mathbf{y} - \mathbf{X}\hat{\beta}$ and $s^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{T}$. (7.28) is distributed as a χ^2 with n-p degrees of freedom and it is the very well known test for the validity of instruments originally proposed by Sargan within the context of the GIVE estimator (see, for example, Sargan (1988)).

7.3.1 *Covariance Matrix Estimation*

So far we have implicitly considered the case in which the empirical moments were serially independent. In general it is worthwhile to relax such assumption, as in many macroeconometric applied cases some dependence in the empirical moments will be generated. Think for example the case of estimation of central bank reaction functions. As we shall see later, a Central Bank's policy rule can be specified by assuming that CB set their instrument, the interest rate, to react to contemporaneous output gap, the difference between implies current and potential output, and to deviation of future expected inflation from the target for inflation. Future inflation is the relevant variable because the existence of lags between monetary action and their effect on the economy makes reacting to contemporaneous target useless. The literature takes the relevant horizon for future inflation to be about one-year. So the following rule could be specified on monthly data:

$$r_t = a_0 + a_1 E_t\left(\pi_{t+12} - \pi_t^*\right) + a_2 E_t\left(y_t - y_t^*\right) + v_t \tag{7.29}$$

where v_t is an exogenous i.i.d. disturbance. To fit (7.29)to the data, the unobservable forecast variables are eliminated by rewriting the rule in terms of realized variables as follows:

$$r_t = a_0 + a_1\left(\pi_{t+12} - \pi_t^*\right) + a_2\left(y_t - y_t^*\right) + \varepsilon_t \tag{7.30}$$

$$\varepsilon_t = a_1\left[E_t\left(\pi_{t+12} - \pi_t^*\right) - \left(\pi_{t+12} - \pi_t^*\right)\right] + a_2\left[E_t\left(y_t - y_t^*\right) - \left(y_t - y_t^*\right)\right] + v_t \tag{7.31}$$

Labelling $\mathbf{z}_t$ the vector of variables within the central bank's information set at the time the interest rate is chosen, we can construct the GMM estimator using the following set of orthogonality condition:

$$E_t\left(\varepsilon_t \mid \mathbf{z}_t\right) = 0 \tag{7.32}$$

however, by construction, the composite disturbance term ε_t features an MA(11) structure and empirical moments cannot be considered as serially independent.

To deal with this case we rewrite Ψ, the covariance matrix of the empirical moments, as follows:

$$\Psi = \lim_{T\to\infty}\left[\frac{1}{T}\sum_{p=1}^{T}\sum_{q=1}^{T} E\left(F_p'\left(\mathbf{x}_{t+i},\mathbf{z}_t,\boldsymbol{\theta}\right)F_q\left(\mathbf{x}_{t+i},\mathbf{z}_t,\boldsymbol{\theta}\right)\right)\right] \tag{7.33}$$

where $F_q\left(\mathbf{x}_{t+i},\mathbf{z}_t,\boldsymbol{\theta}\right)$ is the qth row of the (Txn) matrix $F\left(\mathbf{x}_{t+i},\mathbf{z}_t,\boldsymbol{\theta}\right)$. The first step to find a consistent estimator of Ψ is to define the autocovariances of the empirical moments as follows:

$$\Gamma(j) = \frac{1}{T}\sum_{p=j+1}^{T} E\left(F_p'\left(\mathbf{x}_{t+i},\mathbf{z}_t,\boldsymbol{\theta}\right)F_{p-j}\left(\mathbf{x}_{t+i},\mathbf{z}_t,\boldsymbol{\theta}\right)\right) \; for\; j \geq 0 \tag{7.34}$$

$$\Gamma(j) = \frac{1}{T}\sum_{p=-j+1}^{T} E\left(F_{p+j}'\left(\mathbf{x}_{t+i},\mathbf{z}_t,\boldsymbol{\theta}\right)F_p\left(\mathbf{x}_{t+i},\mathbf{z}_t,\boldsymbol{\theta}\right)\right) \; for\; j \leq 0 \tag{7.35}$$

In terms of the (nxn) matrices $\Gamma(j)$, the right hand side of (7.33)without the limit becomes:

$$\Psi^n = j = -n + 1 \sum^{n-1} \Gamma(j) \tag{7.36}$$

If there were no serial correlations between observations, then only $\Gamma(0)$ would be non-zero and we would have

$$\Psi^n = \Gamma(0) = \frac{1}{T}\ p = 1 \sum^{T} E\left(F_p'(\mathbf{x}_{t+i}, \mathbf{z}_t, \boldsymbol{\theta})\, F_p(\mathbf{x}_{t+i}, \mathbf{z}_t, \boldsymbol{\theta})\right) \tag{7.37}$$

which could be useful to deal with heteroscedasticity in the empirical moments. To see this point empirically let us consider again the case of GIVE with heteroscedastic disturbances.

The relevant problem is to estimate the vector of unknown parameters $\boldsymbol{\beta}$ in the linear model :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{7.38}$$

where y is a (Tx1) vector of observations on the dependent variable, $\mathbf{X}$ is a (Txp) matrix of observations on the explanatory variables, $\boldsymbol{\beta}$ is the (px1) vector of parameters of interest, and $\mathbf{u}$ is the (Tx1) vector of observations on the error term with zero mean and variance-covariance matrix equal to Ω.

Where $\Omega = \begin{bmatrix} \sigma_1^2 & . & 0 \\ . & . & . \\ 0 & 0 & \sigma_T^2 \end{bmatrix}$. As before we assume that $\mathbf{X}$ are not weakly exogenous for the estimation of the parameters of interest, but there exists a $\mathbf{Z}$ matrix containing T observations on n valid instruments. Applying formula (7.19) , the relevant estimate is derived by solving the following problem:

$$\boldsymbol{\beta} Min \left(\mathbf{u}'\mathbf{Z}\Psi^{-1}\mathbf{Z}'\mathbf{u}\right) \tag{7.39}$$

where the appropriate choice for the matrix Ψ is:

$$\Psi = \Gamma(\mathbf{0}) = \mathbf{Z}'\Omega\mathbf{Z} = \frac{\mathbf{1}}{\mathbf{T}}\ \mathbf{p} = \mathbf{1} \sum^{\mathbf{T}} \mathbf{E}\left(u_p^2\right) \mathbf{Z}_p'\mathbf{Z}_p \tag{7.40}$$

which can be consistently estimated by using any consistent estimator of the parameters of interest and by substituting $\mathbf{E}\left(u_p^2\right)$ with the just the square of the corresponding residual.

This estimator is intepretable as an extension of the Heteroscedaticity Consistent estimator proposed by White(1982) within the OLS framework to the GIVE case.

Kdylqj glvfxvvhg wklv dssolfdwlrq ohw xv jr edfn wr wkh jhqhudo fdvh ri vh0 uldo fruuhodwlrq ri xqnqrzq irup1 Dv uhpdunhg e| Gdylgvrq0PdfNlqqrq^:` lw lv whpswlqj wr hvwlpdwh wkh dxwrfrydudqfhv vlpso| e| rplwwlqj wkh h{shfwdwlrqv lq wkdw h{suhvvlrq/ hydoxdwlqj wkh $\hat{s}$ dw vrph frqvlvwhqw suholplqdu| hvwlpdwh ri wkh yhfwru ri sdudphwhuv ri lqwhuhvw dqg wkhq e| vxevwlwxwlqj wkh $\hat{a}$ wkhuhe| re0 wdlqhg lqwr+:=6:, lq rughu wr rewdlq d vxlwdeoh hvwlpdwh ri= Wklv zrxog mxvw jhqudol}hg wkh surfhgxuh lpsohphqwhg lq wkh fdvh @ + 3, =

to eliminate all the non-observable variables from the Euler equation. Therefore, GMM is not empirically feasible anymore. This simple example shows rather clearly why when constraints are introduced in the intertemporal optimisation problem, the GMM method becomes much less popular.

It is fair to say that several solutions have been proposed to the problem generated by the introduction of liquidity constraints, but none of them replicate the neat correspondance between solution to the economic problem and implementation of the econometric methodology obtained in the case of the intertemporal optimisation without market imperfections.

Deaton (1991,1992) observes the impossibility of finding an analitycal solution to (7.48)and proposes to characterise the properties of the numeric solution obtained under the hypothesis of very simple DGP for the income process. However, even for a very simple autoregressive process for income, the computational burden is rather heavy. Pesaran-Smith(1994) propose to approximate the unknown Lagrange multipliers by a general function of observable variables. Within this context it is important to gather institutional information to help the identification of the appropriate functional form and of the appropriate argument for such function. Favero-Pesaran(1994) apply this methodology to the empirical modelling of oil investment using institutional and geological information to identify the appropriate function. Abandoning time-series there is the possibility to revert to panel data to identify liquidity constrained agents Zeldes(1989). Brave attempts to identify the relevance of liquidity constraints using time series data have been proposed by Campbell-Mankiw(1987), and Jappelli-Pagano(1989). Aggregate time series consumption is thought of as the result of the aggregation of consumption by two type of agents: those who are liquidity constrained and those who are not liquidity constrained. To allow aggregation utility is assumed to be quadratic, then the behavoiur of the uncostrained agents is described by the usual Euler equation while constrained agents are assumed to consume all their disposable income in each period. By assuming that a fixed proportion of income accrues to each type of agents the Euler equation for the unconstrained agents and the consumption function for the constrained agents are aggregated into a macroeconomic consumption function, which, interestingly, takes the form of an ECM model. One of the estimated parameter in such consumption function is the proportion of income accruing to the constrained agents. The importance of liquidity constraints in the economy can therefore be empirically evaluated on macro time-series data. It is not clear however why the proportion of income accruing to the liquidity constrained agents is thought of as a parameter rather than a variable.

7.4.2 *Empirical Problems*

The main empirical problem with the GMM approach to estimate structural parameters has been noted by a series of authors (Ghysels-Hall(1990a, 1990b), Garber-King(1983), Oliner S., Rudebusch G. e Sichel D.(1992)) for the US data. In fact it has been observed that in general the parameters estimated on aggregate time-series data by implementing GMM on Euler equations derived by different intertemporal optimisation problem are not stable over time. Such instability is clearly not compatible with their nature of parametrs describing taste and technology suggested by the theoretical models. There are several possible interpretation of instability: it could signal the incorrect specification of the estimated model or it could be generated by the fact that representative agents model are applied on aggregate data without taking proper care of aggregation. This second interpretation has generated a research programme which refrain from estimating the "deep" parameters from aggregate macroeconomic time-series. "Deep" parameters are insteadtaken from microeconometric studies on disaggregated data, using these parameters thoeretical model are then calibrated and simulated, finallly properties of the simulated data are compared with properties of macroeconomic time-series to evaluate the ability of the theoretical model to replicate features of the real data. We shall concentrate on the calibration methodology later on. We conclude now our analysis of the GMM method by looking at empirical applications of this methodology.

7.5 An application to the consumer's problem

The first illustrative example which we consider involves the estimation of the Euler equation (7.5) : $E_t\left(\frac{1+r_{t+1}}{1+\delta}c_{t+1}^{-\gamma}-c_t^{-\gamma}\right)=0$,using the a data set on monthly US data, which is the version of the Hansen-Singleton(1993) data-set made available as a tutorial data-set for Microfit version 4.0(See Pesaran and Pesaran(1997)). The data set is available in Excel format as HS.XLS. It contains monthly data for the sample 1959:3-1978:12 on the following variables:

X1: ratio of consumption in time period t-1 to consumption in time period t

X2: 1+ the one-period real return on stocks

Estimation of the Euler equation is implemented using the appropriate routine in E-Views, using the Bartlett weights and the Newey-West criterion to choose the lag truncation parameter. The following results are obtained:

Method: Generalized Method of Moments

Sample(adjusted): 1959:04 1978:12

Included observations: 237 after adjusting endpoints

Bandwidth: Fixed (4)

Kernel: Bartlett

Convergence achieved after: 3 weight matricies, 4 total coef iterations

C(1)*(X1^C(2))*X2-1

Instrument list: C X1(-1) X2(-1)

Coeff	Estimate	Std. Error	t-Statistic.	Prob
C(1)	0.998082	0.004465	223.5548	0.0000
C(2)	0.891202	1.814987	0.491024	0.6239

S.E. of regression 0.041502 Sum squared resid 0.404766

Durbin-Watson stat 1.828192 J-statistic 0.006453

Note that the parameters are estimated by using the following three orthogonality conditions:

$$E_t\left(\frac{1+r_{t+1}}{1+\delta}\left(\frac{c_t}{c_{t+1}}\right)^{\gamma}-1\right)=0$$

$$E_t\left(\frac{1+r_{t+1}}{1+\delta}\left(\frac{c_t}{c_{t+1}}\right)^{\gamma}-1\right)r_t=0$$

$$E_t\left(\frac{1+r_{t+1}}{1+\delta}\left(\frac{c_t}{c_{t+1}}\right)^{\gamma}-1\right)\left(\frac{c_{t-1}}{c_t}\right)=0$$

therefore we have one over-identifying restrction whose validity can be tested by using the J-statistic. such statistic is easily computed multiplying by the number of observations the J-statistic reported in the E-views output. Such statistic is distributed as a χ^2 with one degree of freedom. Given that the observed value is 1.5294 (237*0.006453), we do not reject the null of validity of instruments. Note that the coefficient of risk aversion is estimated rather poorly, while the discount factor is instead estimated rather precisely. To evalute the relevance of the correction for heteroscedasticity and correlation of unknown form we implement the GMM without such correction. This result is obtained by defining a variable u_t taking a value of zero everywhere and estimating the following model :

$$u_{t+1}=\frac{1+r_{t+1}}{1+\delta}\left(\frac{c_t}{c_{t+1}}\right)^{\gamma}-1+\varepsilon_{t+1} \tag{7.49}$$

The GMM estimates can be obtained by estimating (7.49)by instrumental variables, using the constant, r_t, and $\left(\frac{c_{t-1}}{c_t}\right)$as instruments. The following results are obtained:

Dependent Variable: U
Method: Two-Stage Least Squares
Date: 08/07/98 Time: 10:43
Sample(adjusted): 1959:04 1978:12
Included observations: 237 after adjusting endpoints
Convergence achieved after 8 iterations
U=C(1)*(X1^C(2))*X2-1
Instrument list: X1(-1) X2(-1) C

Coeff	Estimate	Std. Error	t-Statistic.	Prob
C(1)	0.998945	0.004947	201.9470	0.0000
C(2)	0.864734	2.044035	0.423052	0.6726

S.E. of regression 0.041545 Sum squared resid 0.405609
Durbin-Watson stat 1.829335

Results are unaltered. An interesting exercise here is to assess the stability of estimated parameters over time.

7.6 GMM and monetary policy rules

We have already introduced the discussion of the estimation of monetary rules by GMM to illustrate the issue of the possibility of correlation in the sample moments. We now investigate this topic at greater depth, referring to the empirical work by Clarida, Gali and Gertler(1997, 1998). Specification (7.29) ,although useful for some illustrative purposes, is not successful in capturing the observed persistence in the interest rates. Therefore, in the literature the following empirical model is usually specified:

$$r_t^* = \bar{r} + a_1 E_t \left(\pi_{t+12} - \pi^*\right) + a_2 E_t \left(y_t - y_t^*\right) \tag{7.50}$$

$$r_t = (1-\rho)\, r_t^* + \rho r_{t-1} + v_t \tag{7.51}$$

where $r^*$is the target interest rate and a_0 is the equilibrium value for r^*. The partial adjustment mechanism introduced by equation ([?])is justified by the empirical observation of tendency of central banks to smooth interest rates (see Goodfriend(1991)).Moreover a constant target rate of inflation is assumed in the estimated version of the rule. Combining equation (7.50) with (7.51) we derive the following set of orthogonality conditions:

$$E_t \left[r_t - (1-\rho)\, a_0 - a_1 (1-\rho)\, E_t \pi_{t+12} - a_2 (1-\rho)\, E_t \left(y_t - y_t^*\right) - \rho r_{t-1} \mid \mathbf{u}_t\right] = 0 \tag{7.52}$$

Where $\mathbf{u}_t$ includes all the variables in central bank's information set at the time interest rates are chosen. GMM can be used in this framework to estimate the parameters of interest a_0,a_1, a_2 and ρ. The J-test for the validity of over-identifying restrictions can then be used to assess if the simple specification of the monetary policy rule in (7.52) omits important variables which in fact enter the central bank rule. Obvious candidates for the role of omitted variables are monetary aggregates, foreign interest rates, long-term interest rates, exchange-rate fluctuations and possibly stock-markets overvaluation (do central banks care of "irrational exuberance" ?). Moreover the estimation of parameters of interest allows some relevant consideration on monetary policy. In fact, given(7.50) , it is possible to write an equilibrium relation for the real interest rate as follows:

$$rr_t^* = \bar{rr} + (a_1 - 1)\, E_t \left(\pi_{t+12} - \pi^*\right) + a_2 E_t \left(y_t - y_t^*\right) \tag{7.53}$$

Where $\bar{rr}$ is the equilibrium real interest rate, independent from monetary policy. Equation (7.53)illustrates the criticale role of parameter a_1. If $a_1 > 1$ the target real rate is adjusted to stabilize inflation, while with $0 < a_1 < 1$ it instead moves to accomodate inflation: the central bank raises the nominal rate in response to an expected rise in inflation but it does not increase it sufficiently to keep the real rate from declining. Taylor(1998) and Clarida,Gali,Gertler(1998) have shown that $0 < a_1 < 1$ are consistent with the possibility of persistent, self-fulfilling fluctuations in inflation and output. Therefore the value of one for a_1 is an important discriminatory criterion to judge central bank behaviour. Clarida,Gali,Gertler(1998) show that in the pre October 1979 period the FED rule features rules $a_1 < 1$,while the post October 1979 period features $a_1 > 1$. Finally, it is possible to use the fitted values for the parameters $a_0, a_1$to recover an estimate of the central banks' constant target inflation rate π^*. In fact, the empirical model does not allow separate identification of the equilibrium inflation rate and of the equilibrium real interest rate but it does provide a relation between them conditional upon a_0,and a_1. Given that $a_0 = \bar{r} - a_1\pi^*$ and $\bar{rr} = \bar{r} - \pi^*$, we have then

$$\pi^* = \frac{\bar{rr} - a_0}{a_1 - 1} \tag{7.54}$$

which establishes a relation between the target rate of inflation and the equilibrium real interest rate defined by the parameters a_0,and $a_1$in the policy rule. Clarida, Gali, Gertler (1997) set the real interest rate to the average in the sample and use (7.54) to recover the implied value for π^*.

The database CGG contains monthly data for the US and German economy taken from DATASTREAM and from the database on US monthly data used in Sims,Leeper and Zha(1996), which should enable replication of the reaction function estimated by the authors, as well as testing for a number of interesting overidentifying restrictions. The following

variables for the sample 1979:1-1996:12 are available:

GERCMR: German average (of the month) call money rate

GER10Y: redemption yield on German 10-year government bonds

GERCP: German consumer price index

GERINFTAR: Bundesbank announced inflation target (rate of medium term unavoidable inflation)

GERM1: German M1

GERM3: German M3

GERIP: German Industrial Production

PCM: IMF world commodity price index (in US dollars)

SMNBR: US smoothed (by a 36-month moving average) non-borrowed reserves

SMTR: US smoothed (by a 36-month moving average) total reserves

TOTMKUS_DY_01: US stock market dividend-yield

TOTMKUS_PE_01: US stock market price-earning ratio

TOTMKUS_PI_01: US stock market price index

US10Y: redemption yield on US 10-year government bonds

USCP: US consumer price index

USDM: US dollar/ D.Mark exchange rate
USFDTRG: US Federal Funds target
USFF: US average Federal fund rate
USIP:US industrial production
USLABCOSE: US unit labour cost
USM2SA: US M2
USMANHERA: US manufacturing hourly earnings
USOPERATE: US capacity utilisation rate

7.6.1 *The estimation of a baseline policy rule for the FED*

We concentrate first on the US case, trying to replicate the results in CGG(97). A series of empirical problem must be solved in order to perform GMM estimation of the monetary policy rule. The first issue we take is the measurement of the output gap. CGG(97) take deviation of the log of industrial production from a quadratic trend. This is easily obtained by taking the residuals of an OLS regression of the log of industrial production on a constant, a linear trend and a quadratic trend. Such measurement of the cycle would be correct only if the log of industrial production features a deterministic quadratic trend. To check robustness of the definition of the cycle to alternative de-trending methods we compare the original CGG proposal (USGAP1) with the difference between industrial production and an Hodrick-Prescott filter with penalty parameter set to 14400 (USGAP2) and with the demeaned capacity utilization rate(USGAP3). We construct USGAP1, USGAP2, and USGAP3, on the sample 1981:10 1997:12, as we would like to start estimation of the policy rule from 1982:10 (the beginning of the interest rate targeting regime). We the alternative measures of output gaps in Figure 1.

We note that the three different measure do not show evident discrepancies as far as the location of the turning points in the cycle is concerned up to 1990, from 1990 onwards USGAP1 signal a persistent recession, not shared by the other two measures. Obviously such difference does show up in a corresponfing difference in policy rates. Orphanides[26] has considered extensively the problem of measuring the output gap to show that different behaviours by the Fed in the course of the seventies and the eighties can be explained by different measures of the output gaps rather than by different parameters in the reaction function. To keep our results comparable with CGG we keep USGAP1 as the relevant measure of the gap, checking robustness to different detrending choices could be an interesting exercise.

The second empirical problem is the choice of the instruments. Here we follow CGG by taking as instruments the constant, the first six lags, the ninth and the twelvth lag of output gap, the first six lags, the ninth and the twelwth lag of the federal fund rate, the first six lags, the ninth and the twelvth lag of inflation, the first six lags, the ninth and the twelvth lag of the log IMF commodity price index.

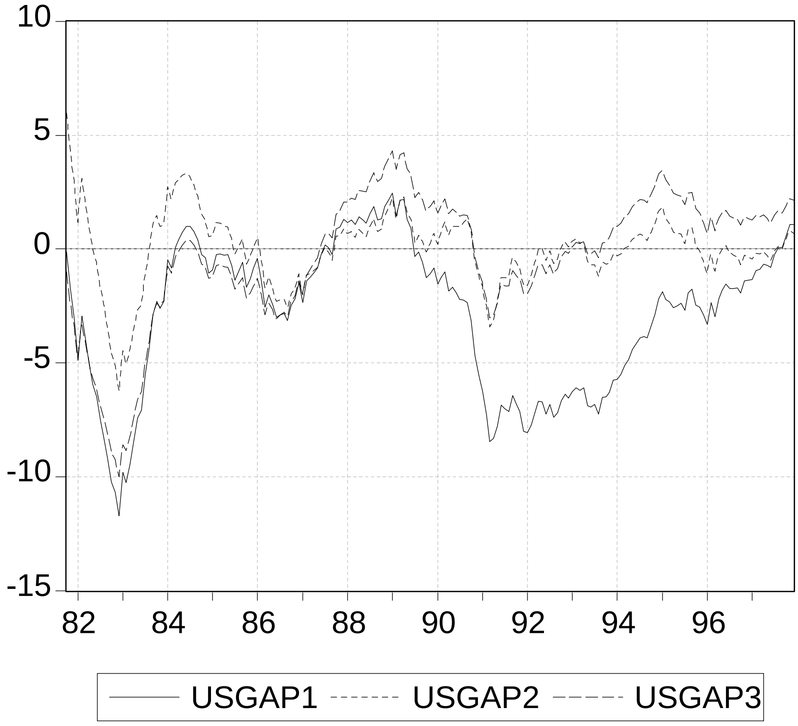


FIG. 7.1. Alternative measures of US output gap

We then implement estimation by GMM, using the correction for heteroscedasticity and autocorrelation of unknown form with a lag truncation parameter of 12 and choosing Bartlett weights to ensure positive definitess of the estimated variance-covariance matrix. The following results are obtained by implementing GMM in E-Views:

Dependent Variable: USFF Method: Generalized Method of Moments
Sample(adjusted): 1982:10 1996:12, 171 observations
No prewhitening
Bandwidth: Fixed (12)
Kernel: Bartlett
Convergence achieved after: 78 weight matricies, 79 total coef iterations
USFF= C(2)*USFF(-1) +(1-C(2))*(C(1)+C(3)*USINFL(+12) +C(4) *USGAP1)
Instrument list: C USGAP1(-1) USGAP1(-2) USGAP1(-3) USGAP1(-4) USGAP1(-5) USGAP1(-6) USGAP1(-9) USGAP1(-12) USINFL(-1)

USINFL(-2) USINFL(-3) USINFL(-4) USINFL(-5) USINFL(-6) USINFL(-9) USINFL(-12) USFF(-1) USFF(-2) USFF(-3) USFF(-4) USFF(-5) USFF(-6) USFF(-9) USFF(-12) DLPCM(-1) DLPCM(-2) DLPCM(-3) DLPCM(-4) DLPCM(-5) DLPCM(-6) DLPCM(-9) DLPCM(-12)

Coeff	Estimate	Std. Error	t-Statistic.	Prob
C(2)	0.92	0.012	73.82	0.0000
C(1)	2.87	0.99	2.90	0.0004
C(3)	1.73	0.25	6.87	0.0000
C(4)	0.66	0.10	6.60	0.0000

R-squared 0.98 Mean dependent var 6.713957

Adjusted R-squared 0.98 S.D. dependent var 2.191514

S.E. of regression 0.28 Sum squared resid 13.74

Durbin-Watson stat 1.06 J-statistic 0.0611

So we have an estimated a_0=2.87,an estimated $a_1 = 1.73$, an estimated $a_2 = 0.66$, while $\rho = 0.92$. Estimates are in line with the one obtained by CGG, with $a_1 > 1$, altough our slightly different. Such difference could be explained by their choice of second order lag in the adjustment, while we restrict to first order dynamics.[16]

The statistic for the validity of instruments, distibuted as a χ^2 with 29 (33 instruments for 4 parameters) takes the value of 10.45 (0.0611*171as the reported statistic in E-Views is divided by the number of observations) and does not reject the null of validity of instruments. If we follow the practice suggested by CGG to derive an estimate for the inflation target by using the estimated parametrs and the average real interest rate over the sample as a proxy for the real equilibrium interest rate, we get a point estimate of 0.5 with a rather wide confidence interval (as the 95 confidence interval for a_0 spans 0.89-4.85). Overall the rule is rather successful in explaining the Fed behaviour as illustrated in Figure 2, where we report observed policy rates and the 95 per cent confidence interval from our estimated equation:

7.6.2 *Does the Fed care for the long-term interest rate ?*

Within the GMM framework it is rather easy to check the importance of omitted variables in the policy rule. In fact if there are important omitted variables from the policy rule, for such variables the orthogonality condition should be violated and the test for the validity of instruments should then reject the null hypothesis. There is a rather wide literature concentranting on the importance of long-term interest rates for the FED explicitly related to their signalling role for "inflation scares". As pointed out by Goodfriend[16], the behaviour of long-term interest rate could be informative on agents expectations for inflation and on the effects of monetary policy on such expectations. Campbell(1995) concentrates on the collapse of bond price in 1994 relating it to movements in the term premium

[16]Checking this empirically could be a useful exercise

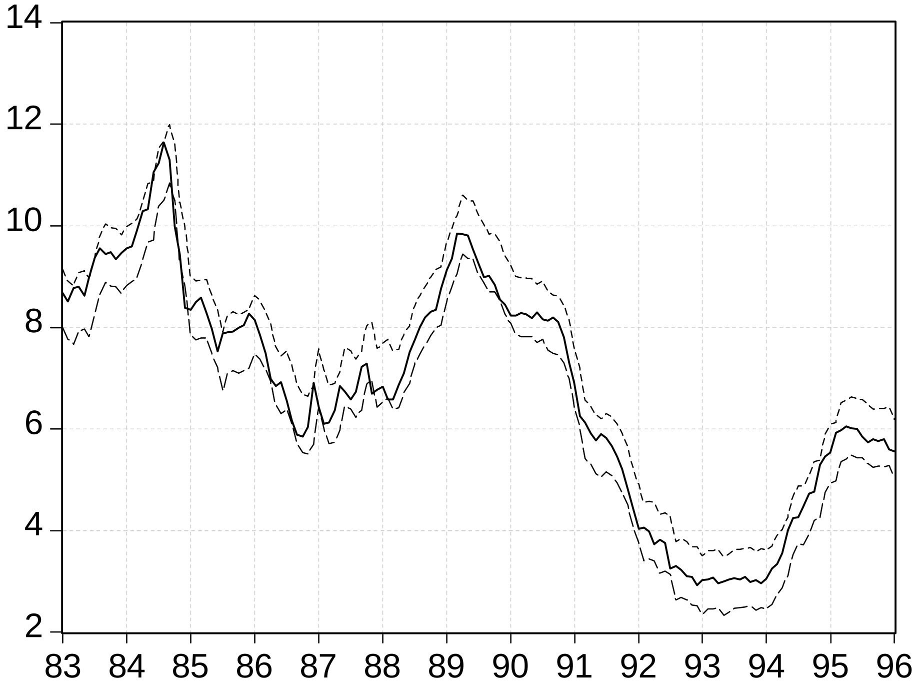


FIG. 7.2. Observed US policy rates and the 95 per cent confidence interval from the estimated policy rule

generated by a rise in expected inflation, not matched by any movement in the same direction in actual inflation. Looking at the 1994 data we see clearly that the Fed reacted lately to the increase in long-term interest rates and it took several tightening steps in the target federal funds rate to convinve markets of the central bank determination in fighting inflation. In fact only after several tightening movements in the policy rate the long-term interest rate started revertig its upwrd trend. All this discussion show that there are good theoretical and policy reason for the Central Bank to monitor long-term interest rates, and the omission of long-term interest rates from the rule seems an obvious candidate for putting our testing procedure at work. We then re-estimate the base-line model by including the level of contemporaneous long-term interest rates in the set of instruments. The following results are obtained:

Dependent Variable: USFF
Method: Generalized Method of Moments
Date: 08/06/98 Time: 15:51
Sample(adjusted): 1982:10 1996:12

Included observations: 171 after adjusting endpoints
No prewhitening
Bandwidth: Fixed (12)
Kernel: Bartlett
Convergence achieved after: 64 weight matricies, 65 total coef iterations
USFF= C(2)*USFF(-1) +(1-C(2))*(C(1)+C(3)*USINFL(+12) +C(4) *USGAP1)
Instrument list: C USGAP1(-1) USGAP1(-2) USGAP1(-3) USGAP1(-4) USGAP1(-5) USGAP1(-6) USGAP1(-9) USGAP1(-12) USINFL(-1) USINFL(-2) USINFL(-3) USINFL(-4) USINFL(-5) USINFL(-6) USINFL(-9) USINFL(-12) USFF(-1) USFF(-2) USFF(-3) USFF(-4) USFF(-5) USFF(-6) USFF(-9) USFF(-12) DLPCM(-1) DLPCM(-2) DLPCM(-3) DLPCM(-4) DLPCM(-5) DLPCM(-6) DLPCM(-9) DLPCM(-12) US10Y US10Y(-1)

Coeff	Estimate	Std. Error	t-Statistic.	Prob
C(2)	0.95	0.007	120.63	0.0000
C(1)	4.23	1.10	3.84	0.0002
C(3)	1.48	0.27	5.37	0.0000
C(4)	0.86	0.11	7.47	0.0000

R-squared 0.98 Mean dependent var 6.713957
Adjusted R-squared 0.98 S.D. dependent var 2.191514
S.E. of regression 0.27 Sum squared resid 12.69
Durbin-Watson stat 1.17 J-statistic 0.067

The point estimates of the parameters are slightly modified but the tests for validity of instruments does not reject the null$(0.067 * 171 = 11.45)$. In the light of this evidence we can conclude that the long-term interest rate affects the Fed behaviour as a leading indicator for future inflation but not as an independent argument of the monetary policy rule.

7.7 Interest rate rules and central banks' preferences

Monetary policy rules like those we have so far considered are empirically successful and useful to show how the GMM methodology is applied. However, they are not in line with our introduction to the GMM methodology in that they are not derived explicitly from an intertemporal optimization problem and therefore no deep parameters describing central banks' preferences are identifiable. In fact it is perhaps surprising that the GMM methodology has been used to estimate reaction functions, while the optimization problem of the central banks provides first order conditions which are instead a more natural object of GMM estimation. Following Svensson[**?**], we consider the simplest possible version of the inflation targeting problem. The central bank faces the following intertemporal optimisation problem:

$$Minimize \quad E_t i = 0 \sum^{\infty} \delta^i L_{t+i} \tag{7.55}$$

where:

$$L = \frac{1}{2}\left[(\pi_t - \pi^*)^2 + \lambda x_t^2\right] \tag{7.56}$$

where E_t denotes expectations conditional upon the information set available at time t , δ is the relevant discount factor, π_t is inflation at time t , π^* is the target level of inflation, x represents deviations of output from its natural level, λ is a parameter which determines the degree of flexibility in inflation targeting. When $\lambda = 0$ the central bank is defined as a strict inflation targeter. As the monetary instrument is the policy rate, i_t , the structure of the economy must be described to obtain an explicit form for the policy rule. We consider the following specification for aggregate supply and demand in a closed economy:

$$x_{t+1} = \beta_x x_t - \beta_r \left(i_t - E_t \pi_{t+1} - \bar{r}\right) + u_{t+1}^d \tag{7.57}$$

$$\pi_{t+1} = \pi_t + \alpha_x x_t + u_{t+1}^s \tag{7.58}$$

As shown in Svensson[21], the first order conditions for optimality may be written as follows:

$$\frac{dL}{di_t} = (E_t \pi_{t+2} - \pi^*) = -\frac{\lambda}{\delta \alpha_x k} E_t x_{t+1} \tag{7.59}$$

$$k = 1 + \frac{\delta \lambda k}{\lambda + \delta \alpha_x^2 k} \tag{7.60}$$

Note that(7.59) deliver the set of orthogonality conditions, which constitute the natural object for GMM estimation. Joint estimation of (7.58) , (7.57) , and (7.59) allows identification and estimation of the parameters describing the structure of the economy and of the parameters describing central banks' preferences. Alternatively (7.58) , (7.57) , and (7.59) , can be used to derive an interest rate rule. In fact, by substituting from (7.58) in (7.57) we obtain:

$$E_t \pi_{t+2} = E_t \pi_{t+1} + \alpha_x [\beta_x x_t - \beta_r (i_t - E_t \pi_{t+1} - \bar{r})] \tag{7.61}$$

and by substituting (7.61)in (7.59) we derive a standard interest rate rule:

$$i_t = \bar{r} + \pi^* + \left(\frac{1 + \alpha_x \beta_r}{\alpha_x \beta_r}\right)(E_t \pi_{t+1} - \pi^*) + \tag{7.62}$$
$$+ \frac{\beta_x}{\beta_r} x_t + \frac{\lambda}{\delta \alpha_x k} \frac{1}{\alpha_x \beta_r} E_t x_{t+1}$$

A number of comments on this rule are in order:

- If the rule is estimated as a single equation, then the fitted parameters are convolutions of the parameters describing central banks preferences (π^*, λ, δ) and of those describing the structure of the economy $\left(\alpha_x, \beta_r, \beta_x, \overline{r}\right)$. Thus the estimated parameters in the interest rate rules are not "deep" in the sense of Lucas (1976).
- As the structure of the economy cannot be identified from the estimation of the rule only, it is impossible to assess if the responses of central banks to output and inflation are consistent with the parameters describing the impact of the policy instrument on these variables. Note, for example, that the estimation of an interest rate rule relating the policy rate to the output gap and to the deviation of expected inflation from target does not help to distinguish a strict inflation targeter($\lambda = 0$, in the terminology of Svensson) , from a flexible inflation targeter $(\lambda > 0)$.
- Econometric identification of the rule requires the timing assumption that the central bank can set policy rates in response to contemporaneous macro variables in the economy, but policy rates do not have a contemporaneous impact on those variables. This assumption is commonly used to identify VAR models of the monetary transmission mechanism.
- In order to make (7.62) consistent with the data, the rule has been interpreted as delivering "target" interest rates, and a sluggish adjustment of actual to target rates has been imposed (Clarida, Gali and Gertler, 1997). Direct estimation of the policy rule does not allow to identify a structure of central bank's preferences which is consistent with interest rate smoothing.
- There is only one empirical implication of the rule which can be confronted with the data independently from the identification of the parameters of interest, namely whether the parameter describing the reaction of policy rates to a gap between expected and target inflation is larger than one. In fact a monetary policy which accommodates changes in inflation, $\frac{\partial i_t}{\partial E_t \pi_{t+1}} \leq 1$,will not in general converge to the target rate π^*. This empirical prediction is the one which has attracted most of the discussion on estimated monetary policy rules (See again Clarida, Gali and Gertler, 1997).

To provide a better mapping from central banks' behavior to their preferences a strategy, closer to the spirit of intertemporal optimisation, seems more appropriate. First, estimate the structure of the economy to identify the parameters of the aggregate supply and demand functions. Second, estimate the Euler equation for the solution of the intertemporal optimisation problem to identify central banks preferences. In this step (and in reference to the simple example analyzed above), given the knowledge of $\alpha_x$and β_r , we can identify directly, from the estimation of the first order conditions (7.59), the λ and π^* associated to each assumed value of the discount rate, δ . Third, test if the monetary policy rule consistent with the structure of the economy and the preferences of the central bank matches the actual behavior of policy rates. This strategy has been followed by Favero and Rovelli[12], whose empirical investigation leads to select a strict inflation targeting with real interest rate smoothing (with estimated relative weigths of about four to one) as the best model to describe the Fed behaviour in the eighties.

REFERENCES

Andrews D.W.K.(1991) "Heteroscedasticity and autocorrelation consistent covariance matrix estimation" Econometrica, 59, 249-254

Campbell J. and Mankiw G. (1987) Permanent Income, Current Income and Consumption NBER Working Paper 2436

Clarida R. and M.Gertler(1996) "How the Bundesbank conducts monetary policy." NBER working paper 5581

Clarida R., J.Gali' and M.Gertler(1997) "Monetary policy rules in practice: some international evidence." CEPR discussion paper series n. 1750

Clarida R., J.Gali' and M.Gertler(1998) "Monetary policy rules and macroeconomic stability: evidence and some theory" CEPR discussion paper series n. 1908

Cumby R.E., J.Huizinga and M.Obstfeld(1983) "Two-step two-stage least squares estimation in models with rational expectations" Journal of Econometrics, 21, 333-355

Davidson R. and J.G.MacKinnon(1993) "Estimation and inference in econometrics", Oxford University Press, Oxford

Deaton A.(1991) "Savings and Liquidity Constraints" Econometrica, 59 (5), 1221-1248

Deaton A. (1992) Understanding Consumption Oxford, University Press

Eviews 3 User's Guide(1997),Quantitative Micro Software

Favero C.A., Pesaran M.H.(1994) "Oil Investment in the North Sea" , Economic Modelling, 11, 308-329

Favero C.A. and R. Rovelli (1999) "Modelling and identifying central banks' preferences" CEPR discussion paper no 2178

Garber, P. e R.K. King (1983) "Deep Structural Escavation? A Critique of the Euler Equation Methods", NBER technical Working Paper no.31

Ghysels E. and Hall A.(1990a) "A Test for Structural Stability of Euler Conditions Parameters Estimated via the Generalized Method of Moments Estimator." International Economic Review, 2, 31, 355-364.

Ghysels E. and A. Hall (1990b) "Are Consumption-Based Intertemporal Capital Asset Pricing Models Structural?" Journal of Econometrics 45,1-2, 121-139

Goodfriend M.(1991) "Interest rate smoothing and the conduct of monetary policy. Carnegie-Rochester Conference on Public Policy, 7-30

Hall R.E.(1978) "Stochastic implications of the Life-Cycle Permanent Income hypothesis: theory and evidence", Journal of Political Economy, 86, 972-987

Hansen L. (1982) "Large Sample Properties of The Generalized Method of Moments Estimators", Econometrica, 50, 1029-1053

Hansen L. and Singleton K. (1982) "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models" Econometrica, 50, 1269-1291

Hansen L.P. and K.J. Singleton(1983) "Stochastic consumption, risk aversion, and the temporal behavoiur of asset returns" Journal of Political Economy, 91, 249-265

Jappelli T. and Pagano M. (1989) "Consumption and Capital Market Imperfections: An International Comparison", American Economic Review, 79, 5, 1089-1105

Leeper E.M., C.A. Sims and T. Zha (1996) "What does monetary policy do?", available at ftp://ftp.econ.yale.edu/pub/sims/mpolicy

Mankiw G., Rotemberg J., and Summers L.(1985) "Intertemporal Substitution in Economics" Quarterly Journal of Economics, 85, 225-251

Newey W.K. and K.D.West(1987) "A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix", Econometrica, 55, 703-708

Oliner S., Rudebusch G. e Sichel D.(1992) "The Lucas Critique Revisited: Assessing the Stability of Empirical Euler Equations", Board of Governors of the Federal Reserve System, D.P. 130

Orphanides A.(1999) "The quest for prosperity without inflation", paper presented at the conference on "Monetary Policy-making under uncertainty", Frankfurt, December 3-4, 1999

Pesaran M.H. and R. Smith (1994) "Theory and Evidence in Economics", Journal of Econometrics

Pesaran M.H. and B.Pesaran(1997) "Working with Microfit 4.0. Interactive econometric analysis", Oxford University Press

Sargan D.(1988) "Lectures on advanced econometric theory", edited by M.Desai, Basil Blackwell, Oxford

Svensson L.E.O. (1997) "Inflation Forecast Targeting: Implementing and Monitoring Inflation Targets", *European Economic Review*, 41, 1111-1146

Svensson L.E.O. (1998) "Monetary Issues for the ESCB". Presented at the Carnegie-Rochester Conference. Available at http://www.iies.su.se/leosven/

Taylor J.B.(1993) "Discretion versus policy rules in practice" Carnegie-Rochester Conference series on Public Policy, 39, 195-214

Taylor J.B.(1998) "The robustness and efficiency of monetary policy rules as guidelines for interest rate setting by the European Central Bank" paper presented at the Conference on Monetary Policy Rules, Stockholm 12-13 June 1998

Zeldes S.P. (1989) "Consumption and Liquidity Constraints: An Empirical Investigation ", Journal of Political Economy, 97, 2, 305-346

8

INTERTEMPORAL OPTIMISATION AND CALIBRATION

8.1 Introduction

to be added ...

8.2 Basic set-up

We follow McCallum and Nelson (1997) and assume that our economy is inhabited by a large number of infinitely-living identical price-taking households. They can be aggregate into a single representative household, whose preferences are summarized by the following intertemporal utility function:

$$U_t = E_t \left[\sum_{s=t}^{\infty} \frac{\beta^{s-t}}{1-\mu} \left[\widetilde{C}_t^{1-\mu} + \theta \left(\frac{\widetilde{M}_t}{P_t} \right)^{1-\mu} \right] \right] \tag{8.1}$$

where $\mu \in (0, \infty)$ is the reciprocal of the intertemporal elasticity of substitution, $\beta \in (0, 1)$ is the intertemporal discount factor, $\widetilde{C}_t \in R^+$ is the real consumption level at date t, $\frac{\widetilde{M}_t}{P_t} \in R^+$ is the stock of real money balances held at the start of period t, and $\theta \in (0, \infty)$ is the relative weight of real money balances in the felicity function. As stated in McCallum and Nelson (1997, p. 15), "... the rationale for the inclusion of is of course that holdings of the economy=s medium of exchange provide transaction services that reduce ... (the) resources needed in 'shopping' for the numerous distinct consumption goods whose aggregate is represented by .$\widetilde{C}_t$".

Following again McCallum and Nelson (1997), we assume that each household produces a single good using the following constant-returns-to-scale Cobb-Douglas production function:

$$\widetilde{Y}_t = a_t \widetilde{K}_t^{1-\alpha} (Z_t \widetilde{n}_t)^{\alpha} \tag{8.2}$$

with $\alpha \in (0, 1)$, where $\widetilde{K}_t \in R^+$ is the stock of capital held by the household at date t, $Z_t = \gamma^t \in R^{++}$ is labor-augmenting exogenous technical progress, $n_t \in [0, 1]$ is the labor input, and $a_t \in R$ is a stochastic measure of Total Factor Productivity (TFP). We assume that the natural logarithm of follows a first-order univariate AR process:

$$\ln (a_{t+1}) = (1-\rho) \ln (a) + \rho \ln (a_t) + \epsilon_t \tag{8.3}$$

where a is the unconditional mean, $\rho \in (0, 1)$ the persistence parameter, and ϵ_t S $N\left(0, \sigma^2\right)$ the iid innovation. By adequately choosing units, we impose $a = 1$.

Each household inelastically supplies one unit of labor to a competitive labor market, from which the same household as a producer purchases the labor input at the real wage rate W_t. Furthermore, a market for a one-period government bond exists. These bonds pay between date $t-1$ and t an interest rate equal to r_t.

The following representative household's budget constraint has to hold each period with probability one:

$$\widetilde{K}_{t+1} + (1+\pi_t)\frac{\widetilde{M}_{t+1}}{P_{t+1}} + \widetilde{B}_{t+1} = \qquad (8.$$

$$(1-\delta)\widetilde{K}_t + \frac{\widetilde{M}_t}{P_t} + (1+r_t)\widetilde{B}_t + a_t\widetilde{K}_t^{1-\alpha}(Z_t\widetilde{n}_t)^{\alpha} - \widetilde{C}_t - (\widetilde{n}_t - 1)W_t -$$

where $\widetilde{B}_t \in R^+$ is the stock of government bonds held at the beginning of period t, P_t is the money price of goods, $\pi_t = (P_{t+1} - P_t)/P_t$ is the inflation rate, V_t is a lump-sum tax levied on the household, and δ is the depreciation rate.

The presence of exogenous technical progress introduces a non-stationary component in the system. This implies that the model will not converge to a steady-state in the long-run. The original non-stationary model can be transformed into a stationary one simply by normalizing all equations with regard to Z_t[17].

Normalizing equation (8.1) gets:

$$U_t = E_t\left[\sum_{s=t}^{\infty}\frac{\widetilde{\beta}^{s-t}}{1-\mu}\left(\widetilde{c}_t^{1-\mu} + \theta(\widetilde{m}_t)^{1-\mu}\right)\right] \qquad (8.5)$$

where $\widetilde{\beta} = \beta\gamma^{1-\mu}$, $\widetilde{c}_t = \widetilde{C}_t/Z_t$, and $\widetilde{m}_t = \widetilde{M}_t/(P_t Z_t)$. To assure the finiteness of our objective function we further impose that $\widetilde{\beta} < 1$.

Similarly, normalizing (8.4) we obtain:

$$\gamma\widetilde{k}_{t+1} + (1+\pi_t)\gamma\widetilde{m}_{t+1} + \gamma\widetilde{b}_{t+1} = \qquad (8.6)$$

$$(1-\delta)\widetilde{k}_t + \widetilde{m}_t + (1+r_t)\widetilde{b}_t + a_t\widetilde{k}_t^{1-\alpha}\widetilde{n}_t^{\alpha} - \widetilde{c}_t - (\widetilde{n}_t - 1)w_t - v_t$$

where small letters identify normalized variables.

8.2.1 *The household's problem*

The representative household solves a stochastic optimal control problem, with consumption and labor as control variables, and capital, money, and bonds as endogenous state variables . Formally, she maximizes (8.5), evaluated at date 0, subject to (8.6) and the initial conditions for all endogenous state variables.

In order to obtain the first order conditions, we form a Lagrangian in expectations:

[17] Since γ is exogenous, the normalization can be easily reversed: the original and the transformed model are isomorphic. Any qualitative conclusion we may reach studying normalized model can be immediately extended to the original one.

$$L = E_t \left\{ \sum_{s=t}^{\infty} \widetilde{\beta}^t \left[\left(\frac{\widetilde{c}_t^{1-\mu} + \theta \widetilde{m}_t^{1-\mu}}{1-\mu} \right) + \widetilde{\lambda}_t \left[(1-\delta) \widetilde{k}_t + \widetilde{m}_t + (1+r_t) \widetilde{b}_t + a_t \widetilde{k}_t^{1-\alpha} \widetilde{n}_t^{\alpha} - \widetilde{c}_t - (\widetilde{n}_t - 1) w_t - v \right. \right. \right. \tag{8.7}$$

where $\widetilde{\lambda}_t$ is a vector of present-value costate variables, and derive it with respect to $\widetilde{c}_t$, $\widetilde{n}_t$, $\widetilde{k}_{t+1}$, $\widetilde{m}_{t+1}$, $\widetilde{b}_{t+1}$ and $\widetilde{\lambda}_t$. The first order conditions are:

$$\widetilde{c}_t^{-\mu} = \widetilde{\lambda}_t \tag{8.8}$$

$$\alpha a_t \widetilde{k}_t^{1-\alpha} \widetilde{n}_t^{\alpha-1} = w_t \tag{8.9}$$

$$\widetilde{\beta} E_t \left[\widetilde{\lambda}_{t+1} (1-\alpha) a_{t+1} \widetilde{k}_{t+1}^{-\alpha} \widetilde{n}_{t+1}^{\alpha} + \widetilde{\lambda}_{t+1} (1-\delta) \right] = \gamma \widetilde{\lambda}_t \tag{8.10}$$

$$\widetilde{\beta} E_t \left(\theta \widetilde{m}_{t+1}^{-\mu} + \widetilde{\lambda}_{t+1} \right) = \left[1 + E_t (\pi_t) \right] \gamma \widetilde{\lambda}_t \tag{8.11}$$

$$\begin{aligned} \gamma \widetilde{k}_{t+1} + (1+\pi_t) \gamma \widetilde{m}_{t+1} + \gamma \widetilde{b}_{t+1} = & \\ & (1-\delta) \widetilde{k}_t + \widetilde{m}_t + (1+r_t) \widetilde{b}_t + a_t \widetilde{k}_t^{1-\alpha} \widetilde{n}_t^{\alpha} - \widetilde{c}_t - (\widetilde{n}_t - 1) w_t - v_t. \end{aligned} \tag{8.12}$$

Conditions (8.8)-(8.12), together with the following trasversality condition:

$$\lim_{t \longrightarrow \infty} E_0 \left[\widetilde{\beta}^t \widetilde{\lambda}_t \left(\widetilde{k}_{t+1} + \widetilde{m}_{t+1} + \widetilde{b}_{t+1} \right) \right] = 0, \tag{8.13}$$

are necessary and sufficient for the household's problem, ie. they completely characterize the sequence of probability measures that solve the household's stochastic optimal control problem. Given that the state variables are always positive, we may rewrite (8.13) as three separated trasversality conditions:

$$\lim_{t \longrightarrow \infty} E_0 \widetilde{\beta}^t \widetilde{\lambda}_t \widetilde{k}_{t+1} = 0, \lim_{t \longrightarrow \infty} E_0 \left[\widetilde{\beta}^t \widetilde{\lambda}_t \widetilde{m}_{t+1} \right] = 0, \lim_{t \longrightarrow \infty} E_0 \left[\widetilde{\beta}^t \widetilde{\lambda}_t \widetilde{b}_{t+1} \right] = 0. \tag{8.14}$$

8.2.2 *The government*

To close the model, we need to specify a government budget constraint that holds each period with probability one. We simply assume that all seigniorage revenues are immediately payed back to the households as lump-sum transfers:

$$-v_t = (1+\pi_t)\,\gamma m_{t+1} - m_t + \gamma b_{t+1} - (1+r_t)\,b_t \tag{8.15}$$

where m_t is the per-household real money supply and b_t is the per-household supply of government bonds.

The dynamics of government bonds has to satisfy also the so-called No-Ponzi-Game (NPG) condition with probability one:

$$\lim_{s\longrightarrow\infty} E_t\left[\prod_{j=t}^{s}(1+r_j)^{-1}\,b_{t+s+1}\right] \geq 0. \tag{8.16}$$

The NPG condition states that the present value of government bonds cannot be strictly negative in the long-run. In other words, it rules out the possibility for the government to repay existing debt contracting always new debt. The (intratemporal) budget constraint (8.15) together with the NPG condition (8.16) forms an intertemporal budget constraint.

For the sake of simplicity, we impose that normalized lump-sum transfers are constant over time, ie. that $v_t = v\ \nabla t$. Furthermore, we assume that the nominal money stock grows at an exogenously given rate η_t:

$$M_t = \prod_{i=0}^{t}\eta_i M_0. \tag{8.17}$$

Finally, we assume that the logarithm of η_t follows a stationary AR process:

$$\ln\left(\eta_{t+1}\right) = (1-\zeta)\ln(\eta) + \zeta\ln(\eta_t) + \epsilon_t \tag{8.18}$$

where η is the unconditional mean, $\zeta \in (0,1\ )$the persistence parameter, and ϵ_t S $N\left(0,\sigma^2\right)$ the iid innovation.

Equation (8.17) can be interpreted as a “degenerated” version of the central banker reaction function, since monetary policy, ie. the growth rate of nominal money balances, does not depend on any endogenous variable. In this framework, then, monetary policy shocks can be modeled as unexpected shocks to the exogenous growth rate of nominal money balances. In other words, they coincide with the iid innovations in (8.18).

8.3 Dynamic equilibrium

To characterize a dynamic equilibrium, we note first of all that, being all households identical, in equilibrium individual and aggregate per-household quantities have to coincide; this implies that all tildes have to be dropped. Furthermore, we impose the following conditions:

$$\widetilde{n}_t = n_t = 1, \quad \widetilde{m}_t = m_t = \frac{M_t}{\gamma^t P_t}, \quad \widetilde{b}_t = b_t \tag{8.19}$$

where in (8.19) M_t is the per-household nominal money supply. Equation (8.19) equates demand and supply for the labor input, the stock of nominal money balances, and the stock of government bonds.

Combining (8.17) and (8.19) with the definition of inflation rate we obtain:

$$1 + \pi_t = \frac{\eta_t}{\gamma} \frac{m_t}{m_{t+1}}. \tag{8.20}$$

There are two other things to note:

1. The variable m_t, a state variable from the household point of view, becomes a forward-looking aggregate decision variable when considered from the aggregate point of view. The reason is the following. The nominal money supply is exogenous, while the demand for real money balances is endogenous. The price level P_t has to equate supply and demand, ie. under rational expectations, has to satisfy the first order condition governing the accumulation of real money balances, equation (8.11). The price level, then, substitutes M_t as an aggregate endogenous variable, with the difference that P_t is not a state variable, but a *forward-looking* variable that can be treated as a costate variable. The variable m_t, then, is the ratio between $M_t\gamma^{-t}$, an exogenous process, and P_t. To stress this point, we may rewrite m_t as $1/p_t$, where $p_t = \gamma^t P_t / M_t$ is a normalized stationary variable.
2. Since both the nominal money supply and the lump-sum transfers are exogenous, bonds have to counterbalance seigniorage in order to keep the government budget balanced. The supply of government bond is then, in some sense, exogenous too. More precisely, it is not under the government control. The demand for government bonds is, however, still endogenous. The real interest rate r_t has to equate supply and demand for bonds, ie. under rational expectations, has to satisfy condition (12). Also , then, becomes an aggregate *forward-looking* decision variable.

We now substitute (16) and (8.20) in the household first order conditions, in order to obtain a system of stochastic difference equations that fully describe a dynamic competitive equilibrium in our economy[18] :

[18] For the sake of simplicity, we dropped equation (??)

$$c_t^{-\mu} = \lambda_t \tag{8.21}$$

$$E_t\left[\lambda_{t+1}\left[(1-\alpha)\, a_{t+1}k_{t+1}^{-\alpha} + 1 - \delta\right]\right] = \frac{\gamma}{\widetilde{\beta}}\lambda_t \tag{8.22}$$

$$E_t\left(\theta p_{t+1}^{\mu} + \lambda_{t+1}\right) = E_t\left(p_{t+1}\right)\frac{\eta_t}{\widetilde{\beta}}\frac{\lambda_t}{p_t} \tag{8.23}$$

$$E_t\left[\lambda_{t+1}\left(1 + r_{t+1}\right)\right] = \frac{\gamma}{\widetilde{\beta}}\lambda_t \tag{8.24}$$

$$\gamma k_{t+1} = (1-\delta)\, k_t + a_t k_t^{1-\alpha} - c_t \tag{8.25}$$

$$\gamma b_{t+1} = \frac{1-\eta_t}{p_t} + (1 + r_t)\, b_t - v_t. \tag{8.26}$$

Furthermore, the trasversality conditions (8.14) can be restated as:

$$\lim_{t\longrightarrow\infty} E_0 \widetilde{\beta}^t \lambda_t k_{t+1} = 0,\ \lim_{t\longrightarrow\infty} E_0\left[\widetilde{\beta}^t \frac{\lambda_t}{p_{t+1}}\right] = 0,\ \lim_{t\longrightarrow\infty} E_0\left[\widetilde{\beta}^t \lambda_t b_{t+1}\right] = 0. \tag{8.27}$$

Note that the last trasversality condition in (8.27) implies (check!) the NPG condition (8.16). Note furthermore that, as in Benassy (1995), the dynamics of consumption and investment is driven by real shocks only (consider equations 8.21, 8.22, and 8.25: they represent a stand-alone Brock-Mirman model). The "real" world is in some sense completely separated from the "monetary" world. The dynamics of government debt, on the other side, is driven by both monetary and real shocks. This "separation" result is of course not robust, depending on our very particular assumptions as far as the structure of preferences and the money creation process are concerned.

In summary, equations (8.22)-(8.26), together with the initial conditions and (8.27), form a system of stochastic difference equations that completely describe the competitive equilibrium allocations for our economy. The solution to such a system is an infinite sequence of conditional probability measures that converge in the long-run to a invariant, or unconditional, distribution; in other words, a sequence $\{P_t\left(c_t, p_t, r_t, \lambda_t, b_t, a_t, \eta_t \ \mathrm{p} k_0, b_0, a_0, \eta_0\right)\}_0^{\infty}$, where each $P_t(\cdot)$ represents a probability measure on R_+^{∞}, converging to $P(c, p, r, \lambda, k, b, 1, \eta)$ as $t \rightarrow \infty$. Given the recursive structure of our system, a solution can be also seen as a set of aggregate decision rules for $c_t, p_t, r_t, \lambda_t, k_{t+1}$ and b_{t+1}, expressed as functions of k_t, b_t, a_t and η_t.

8.4 An IS-LM interpretation

One of the main contributions of McCallum and Nelson (1997) is the prove that the simple addition of an expectational term is sufficient to make the standard IS function match a fully optimizing model, whereas no changes are needed as far as the LM function is concerned.

We can easily replicate their argument in our framework. We can combine (8.23) with (8.21) and (8.22), obtaining:

$$E_t\left(\theta m_{t+1}^{-\mu}+c_{t+1}^{-\mu}\right)=\left[1+E_t\left(\pi_t\right)\right]E_t\left[c_{t+1}^{-\mu}\left(1+r_{t+1}\right)\right] \tag{8.28}$$

Following Sargent (1987, pp. 94-95), we can approximate[19] (8.28) with:

$$E_t\left(\theta m_{t+1}^{-\mu}\right)=E_t\left(c_{t+1}^{-\mu}\right)\left\{\left[1+E_t\left(\pi_t\right)\right]\left[1+E_t\left(r_{t+1}\right)\right]-1\right\} \tag{8.29}$$

Equation (8.29) can be rewritten as:

$$E_t\left(\theta m_{t+1}^{-\mu}\right)=E_t\left(c_{t+1}^{-\mu}\right)E_t\left(i_{t+1}\right) \tag{8.30}$$

where $i_{t+1}=\left(1+\pi_t\right)\left(1+r_{t+1}\right)-1$ is the nominal interest rate between date t and $t+1$. Furthermore, we can combine (8.21) and (8.22) to get:

$$\widetilde{\beta}E_t\left[c_{t+1}^{-\mu}\left(1+r_{t+1}\right)\right]=\gamma c_t^{-\mu} \tag{8.31}$$

Consider now equations (8.30) and (8.31). The first one differs only by a random term from a standard LM function $m_t=LM\left(c_t,i_t\right)$, where the real money balances depend upon a transaction variable and an opportunity cost variable. The second one, instead, can be interpreted as an extended IS function by imposing a further assumption. If, as stated in McCallum and Nelson (1997, pp. 7-10), we are able to approximate fluctuations in income with fluctuations in consumption (at least for business cycle purposes) , then we may substitute to in (8.31), and get an extended IS function of the form $y_t=IS\left[E_t\left(y_{t+1}\right),E_t\left(r_{t+1}\right)\right]$. The previous IS function is non-standard since it incorporates expectational terms for both the income level and the real interest rate. This forward-looking aspect is usually absent in standard $IS-LM$ analysis.

8.5 Calibration

Before focusing on the solution method, we need to choose a value for all deep parameters of the model. To specify a particular value for all exogenous parameters

[19] Given two random variables, x and y, we have that $E(xy)=E(x)E(y)+Cov(x,y)$. Sargent () approximates the conditional covariance term by zero.

is a quite demanding task indeed, since many of these parameters are difficult, or practically impossible, to estimate.

The approach we follow here is known as calibration, and is extensively discussed in Cooley (1997), among others. He states (p. 56) that "calibration is a strategy for finding numerical values for the parameters of artificial economic worlds...[it] uses economic theory extensively as the basis for restricting a general framework and mapping that framework into the measured data."

The main difference between the calibration approach and the standard econometric approach lies in the bidirectional relationship between theory and measurement. First of all, the theory defines the quantities of interest to be measured, and suggests how available measurements have to be reorganized, since they may be not consistent with the theory. For example, the concept of investment in our model is a fairly broad one; since no government or foreign sectors are explicitly modeled, to obtain a measure of investment that matches our theoretical concept we have to reorganize available data, and sum up private fixed investment, private consumption of durable goods, government investment, and net exports. An extensive discussion of this and other topics can be found in Cooley and Prescott (1995), where a standard neoclassical growth model is carefully calibrated to the US economy.

Then, measurements are used to give empirical content to the theory, and in particular to provide empirically based values for unknown parameters; in other words, the parameters are chosen, according to Cooley (1997, p. 58), "so that the behavior of the model economy matches features of the measured data in as many dimensions as there are unknown parameters". In our case, the features we want to match are long-run features of the real and monetary variables, since our main interest are the short-run cyclical properties of the model.

Note that estimation and calibration are not substitutes, but complements. In other words, the calibration approach is mostly useful when a sub-set of the parameters is unobservable or difficult to estimate, while standard econometric methods are still preferable when the parameters are observable and the appropriate data easily available.

Note furthermore that the diffused practice of borrowing parameter values from other studies is admissible only if the data were reorganized in way compatible with our needs, and if the measurements obtained refer to the same theoretical concept. We will follow this practice too, and borrow many parameter values from Cooley and Prescott (1995), Cooley and Hansen (1995), and Gavin and Kydland (1999); we stress, however, that the requirements previously discussed are in our case completely fulfilled.

Now, the complete list of parameters we have to pin down is the following: , the intertemporal elasticity of substitution, $\widetilde{\beta}$, the intertemporal discount factor, θ, the relative weight of real money balances in the felicity function, α, the technology coefficient, δ, the depreciation rate, γ, the long-run growth rate, η, the unconditional mean of money growth, ν, the constant level of normalized transfers, ρ, the persistence parameter for TFP, and ς, the persistence parameter for money growth[20].

A large sub-set of these parameters is easily estimated using available data. In particular, the long-run quarterly growth rate γ can be estimated by fitting a linear trend to the logarithm of quarterly GDP; Cooley and Prescott (1995) obtain $\gamma = 1.004$. The persistence parameter for TFP may be estimated by fitting an AR model on the standard Solow residual. Symmetrically, the parameters governing the stochastic process for money growth can be estimated simply fitting an AR model on the logarithm of the actual money growth rate. Again, Cooley and Prescott (1995) obtain $\rho = 0.95$, while Cooley and Hansen (1995) obtain $\eta = 1.013$ and $\zeta = 0.49$..

Finally, empirical estimates exist also for the elasticity of intertemporal substitution; most authors agree on a figure that lies between 1 and 2, so we chose the standard value of 2 for our experiments.

The remaining parameters, namely $\widetilde{\beta}$, θ, α, ν, and , are left for our calibration exercise. Not that the first two parameters are in principle unobservable. As already anticipated, we choose values for these parameters that make the model reproduce some long-run features of actual US data. First of all, then, we have to find out what the long-run features of the model are.

We immediately recognize that the Cobb-Douglas technology implies a labor share in income constant and equal to . This leads us to choose a value for equal to the long-run labor share in total income. Cooley and Prescott (1995) carefully reconstruct a consistent measure of total income and capital income, obtaining a long-run capital share equal to 0.4. We borrow their result and choose .

We impose then a certainty equivalence assumption, assuring this way that the unconditional mean of the invariant distribution to which the solution tends in the long-run is equal to the steady-state of the deterministic version of our system. The steady-state of this deterministic system can be easily computed dropping all expectations and time indexes from (22)-(27):

$$c^{-\mu} = \lambda \tag{8.32}$$

$$(1-\alpha)\frac{y}{k} + 1 - \delta = \frac{\gamma}{\widetilde{\beta}} \tag{8.33}$$

$$\theta p^{\mu} = \left(\frac{\eta}{\widetilde{\beta}} - 1\right)\lambda \tag{8.34}$$

$$1 + r = \frac{\gamma}{\widetilde{\beta}} \tag{8.35}$$

[20]Note

$$\gamma = 1 - \delta + \frac{i}{k} \tag{8.36}$$

$$\gamma b = \frac{1-\eta}{p} + \frac{\gamma}{\widetilde{\beta}} b - \nu \tag{8.37}$$

Equations (33)-(38) define implicitly the steady-state values for the control, endogenous state and costate variables. Furthermore, we may easily obtain a closed form solution for the steady-state capital-output ratio, the income velocity of money, the consumption share in total income, and the government bond-output ratio. From (34) we get:

$$\frac{k}{y} = \frac{\widetilde{\beta}(1-\alpha)}{\gamma - \widetilde{\beta}(1+\delta)} \tag{8.38}$$

Combining (33) and (35), we obtain:

$$\frac{c}{y}\frac{y}{m} = \left(\frac{\eta - \widetilde{\beta}}{\widetilde{\beta}\theta}\right)^{\frac{1}{\mu}} \tag{8.39}$$

Solving (37) for the investment-capital ratio gets:

$$\frac{i}{k} = \gamma - 1 + \delta \tag{8.40}$$

Finally, from (38):

$$\frac{b}{y} = \left[\frac{\nu}{y} - (1-\eta)\frac{m}{y}\right]\frac{\gamma\widetilde{\beta}}{1-\widetilde{\beta}} \tag{8.41}$$

Combining (39) and (41) we get an expression for the investment share $i/y = (i/k)(k/y)$ and indirectly for the consumption share, $c/y = 1 - i/y$..

Empirical estimates of the long-run capital-output ratio, the government bond-output ratio, the income velocity, and the consumption share are readily available. In particular, Cooley and Prescott (1995) obtain a long-run quarterly capital-output ratio equal to 13.28 and a consumption share equal to 0.75; Gavin and Kydland (1999) report a long-run M1 income velocity equal to 5.3.

Reference for b/y

Manipulating (39)-(42), we can express the parameters $\widetilde{\beta}$, θ, δ, and ν, as a function of these observable long-run properties:

$$\delta = 1 - \gamma + \left(1 - \frac{c}{y}\right)\frac{y}{k} \tag{8.42}$$

$$\widetilde{\beta} = \frac{\gamma}{(1-\alpha)\frac{y}{k} + 1 - \delta} \tag{8.43}$$

$$\theta = \frac{\eta - \widetilde{\beta}}{\widetilde{\beta}}\left(\frac{y}{c}\frac{m}{y}\right)^{\mu} \tag{8.44}$$

$$\frac{\nu}{y} = \frac{1-\widetilde{\beta}}{\gamma\widetilde{\beta}}\frac{b}{y} + (1-\eta)\frac{m}{y} \tag{8.45}$$

The implied values are $\delta = 0.015$, $\widetilde{\beta} = 0.989$, $\theta = 1.22$, and .

As already noted, our certainty equivalence assumption makes the unconditional mean of the invariant distribution to which the system tends in the long-run to match its deterministic steady-state. With this im mind, calibration can be interpreted as a method of moments estimation that focuses only on a limited subset of the parameters, setting only the discrepancy between the first moment of the data and the model to zero.

8.6 The KPR procedure

To obtain the decision rules, we apply the well known King, Plosser and Rebelo (1988, KPR) solution procedure. As anticipated in the previous section, we start by imposing certainty equivalence. As a consequence, the unconditional mean of the invariant distribution to which the solution tends in the long-run corresponds to the deterministic steady-state. Then, we linearize the system around the steady-state and solve it with the Blanchard-Khan algorithm.

8.6.1 *Log-linearization*

To implement operationally this procedure, we start by considering a deterministic version of the first order conditions:

$$c_t^{-\mu} = \lambda_t \tag{8.46}$$

$$\lambda_{t+1}\left[(1-\alpha)\,\alpha_{t+1}k_{t+1}^{-\alpha} + 1 - \delta\right] = \frac{\gamma}{\widetilde{\beta}}\lambda_t \tag{8.47}$$

$$\theta p_{t+1}^{\mu} + \lambda_{t+1} = \frac{\eta}{\widetilde{\beta}} \frac{p_{t+1}}{p_t} \lambda_t \tag{8.48}$$

$$\lambda_{t+1} \left(1 + r_{t+1}\right) = \frac{\gamma}{\widetilde{\beta}} \lambda_t \tag{8.49}$$

$$\gamma k_{t+1} = \left(1 - \delta\right) k_t + \alpha k_t^{1-\alpha} - c_t \tag{8.50}$$

$$\gamma b_{t+1} = \frac{1 - \eta_t}{p_t} + \left(1 + r_t\right) b_t - \nu \tag{8.51}$$

We linearly approximate conditions (47)-(52) with a first-order Taylor approximation around the deterministic steady-state, expressing the approximated conditions in percentage deviations from the steady-state itself.

Consider equation (47), and rewrite it as[21] :

$$e^{-\mu \widetilde{c}_t} = e^{\widetilde{\lambda}_t} \tag{8.52}$$

where $\widetilde{x} = \ln\left(x_t\right)$. The first-order Taylor approximation of (53) is equal to[22] :

$$-\mu e^{-\mu \widetilde{c}} \left(\widetilde{c}_t - \widetilde{c}\right) = e^{\widetilde{\lambda}} \left(\widetilde{\lambda}_t - \widetilde{\lambda}\right) \tag{8.53}$$

Since in steady-state $\exp\left(-\mu \widetilde{c}\right) = \exp\left(\widetilde{\lambda}\right)$, and since $\widetilde{x}_t - \widetilde{x} = \ln\left(x_t / x\right)$, condition (54) can be simplified as[23] :

$$-\mu \widehat{c}_t = \widehat{\lambda}_t \tag{8.54}$$

where $\widehat{x}_t = \ln\left(x_t / x\right)$. Equation (55) is a log-linearized version of condition (47), expressed in percentage deviation from the steady-state, since $\widehat{x}_t$ t $\left(x_t - x\right) / x$.

Consider now condition (48), and rewrite it as:

[21] Along an optimal path both c_t and λ_t are strictly positive.

[22] The first order Taylor expansion of a non-linear function $f\left(x\right)$ around a point x_0 is given by $f\left(x\right) = \Delta f\left(x_0\right)\left(x - x_0\right) + \epsilon\left(x\right)$. This implies that $f\left(x\right)$ t $\Delta f\left(x_0\right)\left(x - x_0\right)$.

[23] There is an easier way to get ():take logs of (), add and subtract $\ln\left(\lambda\right)$ from thwe left-hand side of the result, and consider that $\ln\left(\lambda\right) = -\mu \ln\left(c\right)$. Note thet this approachis feasible only because () is aleady log-linear. Note furthermorethat condition is not an approximation, but simply a transformation.

$$e^{\widetilde{\lambda}_{t+1}}\left[(1-\alpha)\,e^{\widetilde{a}_{t+1}}e^{-\alpha\widetilde{k}_{t+1}}+1-\delta\right]=\frac{\gamma}{\widetilde{\beta}}e^{\widetilde{\lambda}_t} \tag{8.55}$$

The first-order Taylor approximation of (56) around the steady-state is:

$$(1-\alpha)\,e^{\widetilde{\lambda}}e^{-\alpha\widetilde{k}}\left(\widetilde{\lambda}_{t+1}-\widetilde{\lambda}\right)+(1-\alpha)\,e^{\widetilde{\lambda}}e^{-\alpha\widetilde{k}}\widetilde{a}_{t+1}-\alpha_{S_K}e^{\widetilde{\lambda}}e^{-\alpha\widetilde{k}}\left(\widetilde{k}_{t+1}-\widetilde{k}\right)+(1-\delta)\,e^{\widetilde{\lambda}}\left(\widetilde{\lambda}_{t+1}-\widetilde{\lambda}\right)=\varphi e \tag{8.56}$$

Equation (57) can be rewritten as:

$$\left[(1-\alpha)\,\frac{y}{k}+1-\delta\right]\lambda\widehat{\lambda}_{t+1}+(1-\alpha)\,\frac{y}{k}\lambda\widehat{a}_{t+1}-\alpha\,(1-\alpha)\,\frac{y}{k}\lambda\widehat{k}_{t+1}=\frac{\gamma}{\widetilde{\beta}}\lambda\widehat{\lambda}_t \tag{8.57}$$

since $\exp\left(\widetilde{x}\right)=x$ and:

$$(1-\alpha)\,e^{\widetilde{\lambda}}e^{-\alpha\widetilde{k}}=(1-\alpha)\,\lambda k^{-\alpha}=(1-\alpha)\,\lambda\frac{k^{1-\alpha}}{k}=(1-\alpha)\,\frac{y}{k} \tag{8.58}$$

Taking into account that in steady-state:

$$\left[(1-\alpha)\,\frac{y}{k}+1-\delta\right]\lambda=\frac{\gamma}{\widetilde{\beta}}\lambda \tag{8.59}$$

we can divide everything by $(1-\alpha)\,\frac{y}{k}\lambda$ and rewrite (59) as:

$$-\alpha\widehat{k}_{t+1}+\varpi\widehat{\lambda}_{t+1}-\varpi\widehat{\lambda}_t=-\widehat{a}_{t+1} \tag{8.60}$$

where:

$$\varpi=\frac{k}{y}\frac{\gamma}{\widetilde{\beta}\,(1-\alpha)} \tag{8.61}$$

Note that in equation (61), all endogenous state and costate variables are grouped on the left-hand side, while the (unique) exogenous state variable is isolated on the right-hand side. This is done for future notational convenience.

Conditions (49) and (50) can be log-linearly approximated by (check!):

$$\left[\mu\left(1-\frac{\widetilde{\beta}}{\eta}\right)-1\right]\widehat{p}_{t+1}+\widehat{p}_t+\frac{\widetilde{\beta}}{\eta}\widehat{\lambda}_{t+1}-\widehat{\lambda}_t=\widehat{\eta}_t \tag{8.62}$$

$$\frac{r}{1+r}\widehat{r}_{t+1}+\widehat{\lambda}_{t+1}-\widehat{\lambda}_t=0. \tag{8.63}$$

Finally, conditions (51)-(52) can be log-linearly approximated by (again, check!):

$$\gamma\frac{k}{y}\widehat{k}_{t+1}-\left[(1-\delta)\frac{k}{y}+1-\alpha\right]\widehat{k}_t=-\frac{c}{y}\widehat{c}_t+\widehat{a}_t \tag{8.64}$$

$$\gamma\widehat{b}_{t+1}-(1+r)\,\widehat{b}_t+(1-\eta)\,\varphi\widehat{p}_t-r\widehat{r}_t=-\eta\varphi\widehat{\eta}_t \tag{8.65}$$

where $\varphi=(m/y)\,(y/b)$.
Equations (3) and (19) directly imply that:

$$\widehat{a}_{t+1}=\rho\widehat{a}_t+\epsilon_t \tag{8.66}$$

$$\widehat{\eta}_{t+1}=\zeta\widehat{\eta}_t+\epsilon_t. \tag{8.67}$$

8.6.2 *The linearized system*

Defining $\widehat{u}_t=[\widehat{c}_t]$, $\widehat{s}_t=\left[\widehat{k}_t\mid\widehat{b}_t\mid\widehat{p}_t\mid\widehat{r}_t\mid\widehat{\lambda}_t\right]'$, and $\widehat{e}_t=[\widehat{a}_t\mid\widehat{\eta}_t]$, we can rewrite (55) as:

$$M_{\mathrm{u}\,u}\widehat{u}_t=M_{\mathrm{u}\,s}\widehat{s}_t+M_{ue}\widehat{e}_t \tag{8.68}$$

where $M_{\mathrm{u}\,u}=[-\mu]$, $M_{us}=[0\mid0\mid0\mid0\mid1]$,, and$M_{ue}=[0\mid0]$.
Conditions (61) and (63)-(66) can be jointly rewritten as:

$$M_{ss}(L)\,\widehat{s}_{t+1}=M_{su}(L)\,\widehat{u}_{t+1}+M_{se}(L)\,\widehat{e}_{t+1} \tag{8.69}$$

where L is the lag operator, or as:

$$\left(M_{ss}^0+M_{ss}^1\cdot L\right)\widehat{s}_{t+1}=\left(M_{su}^0+M_{su}^1\cdot L\right)\widehat{u}_{t+1}+\left(M_{se}^0+M_{se}^1\cdot L\right)\widehat{e}_{t+1} \tag{8.70}$$

where:

$$M_{ss}^0=\begin{bmatrix}-\alpha & 0 & 0 & 0 & \Delta\\ 0 & 0 & \mu\left(1-\frac{\widetilde{\beta}}{\eta}\right)-1 & 0 & \frac{\widetilde{\beta}}{\eta}\\ 0 & 0 & 0 & \frac{r}{1+r} & 1\\ \gamma\frac{k}{y} & 0 & 0 & 0 & 0\\ 0 & \gamma & 0 & 0 & 0\end{bmatrix} \tag{8.71}$$

$$M_{ss}^{1}=\begin{bmatrix}0 & 0 & 0 & 0 & -\Delta\\ 0 & 0 & 1 & 0 & -1\\ 0 & 0 & 0 & 0 & -1\\ -(1-\delta)\frac{k}{y}+1-\alpha & 0 & 0 & 0 & 0\\ 0 & -(1+r) & (1-\eta)\varphi & -r & 0\end{bmatrix} \tag{8.72}$$

$$M_{su}^{0}=\begin{bmatrix}0\\0\\0\\0\\0\end{bmatrix},M_{su}^{1}=\begin{bmatrix}0\\0\\0\\-\frac{c}{y}\\0\end{bmatrix}.M_{se}^{0}=\begin{bmatrix}-1 & 0\\0 & 0\\0 & 0\\0 & 0\\0 & 0\end{bmatrix},M_{se}^{1}=\begin{bmatrix}0 & 0\\0 & 1\\0 & 0\\1 & 0\\0 & -\eta\varphi\end{bmatrix} \tag{8.73}$$

Assuming is invertible, we can solve (69) for $\widehat{u}_t$:

$$\widehat{u}_t=Q_{us}\widehat{s}_t+Q_{ue}\widehat{e}_t \tag{8.74}$$

where $Q_{us}=M_{\mathrm{u}\,u}^{-1}M_{us}$ and $Q_{ue}=M_{\mathrm{u}\,u}^{-1}M_{ue}$. Evaluating (74) at date $t+1$ and substituting the result into (71) we get:

$$\left(M_{ss}^{0}+M_{ss}^{1}\cdot L\right)\widehat{s}_{t+1}=\left(M_{su}^{0}+M_{su}^{1}\cdot L\right)\left(Q_{us}\widehat{s}_{t+1}+Q_{ue}\widehat{e}_{t+1}\right)+\left(M_{se}^{0}+M_{se}^{1}\cdot L\right)\widehat{e}_{t+1} \tag{8.75}$$

Rearranging terms, we may rewrite (75) as:

$$\left(M_{s}^{0}+M_{s}^{1}\cdot L\right)\widehat{s}_{t+1}=\left(M_{e}^{0}+M_{e}^{1}\cdot L\right)\widehat{e}_{t+1} \tag{8.76}$$

where $M_s^0=M_{ss}^0-M_{su}^0Q_{us}$, $M_s^1=M_{ss}^1-M_{su}^1Q_{us}$, $M_e^0=M_{se}^0+M_{su}^0Q_{ue}$, $M_e^1=M_{se}^1+.M_{su}^1Q_{ue}$.

If M_s^0 is invertible, we can solve for $\widehat{s}_{t+1}$, obtaining:

$$\widehat{s}_{t+1}=W\widehat{s}_t+R\widehat{e}_{t+1}+Q\widehat{e}_t \tag{8.77}$$

where $W=-\left(M_{ss}^{0}\right)^{-1}M_{ss}^{1}$, $R=\left(M_{ss}^{0}\right)^{-1}M_{se}^{0}$, and $Q=.\left(M_{ss}^{0}\right)^{-1}M_{se}^{1}$.

Under our certainty equivalence assumption, randomness can be reintroduced by simply taking the conditional expectation of (77):

$$E_t\left(\widehat{s}_{t+1}\right)=W\widehat{s}_t+RE_t\left(\widehat{e}_{t+1}\right)+Q\widehat{e}_t \tag{8.78}$$

Clearly $E_t\left(\widehat{e}_{t+1}\right)=P\widehat{e}_t$, where:

$$P=\begin{bmatrix}\rho & 0\\ 0 & \zeta\end{bmatrix} \tag{8.79}$$

and (78) becomes:

$$E_t\left(\widehat{s}_{t+1}\right) = W\widehat{s}_t + (RP + Q)\,\widehat{e}_t = W\widehat{s}_t + A\widehat{e}_t \tag{8.80}$$

where $A = RP + Q$.

Equation (80) is a linear system of expectational difference equations, and can be solved by applying the Blanchard-Khan algorithm.

8.6.3 *The Blanchard-Khan algorithm*

If P is the modal matrix of W and μ its canonical form (with the eigenvalues on the diagonal ordered in ascending absolute value), and if P is invertible, we may decompose W as $W = P\mu P^{-1}$. We partition the vector of endogenous state variables as $\widehat{s}_t = \left[\widehat{s}_{1't} \mid \widehat{s}_{2't}\right]'$, where $\widehat{s}_{1't} = \left[\widehat{k}_t \mid \widehat{b}_t\right]'$ contains the backward-looking variables, $\widehat{s}_{2't} = \left[\widehat{p}_t \mid \widehat{r}_t \mid \widehat{\lambda}_t\right]'$ and the forward-looking ones. Let us furthermore partition the matrices W, μ , P^{-1} and A as:

$$W = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}, P^{-1} = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}, \mu = \begin{bmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{bmatrix}, A = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}. \tag{8.81}$$

The dynamics of (80) are governed by the eigenvalues of W. Assuming that the first two eigenvalues are stable (strictly less than one in absolute value) and the last three unstable, the system will be saddle-point stable and there will be only one initial vector of forward-looking variables compatible with the transversality conditions.

Pre-multiplying (80) by P^{-1}, we can transform the original system in a transformed system comprised of two decoupled vectors of difference equations:

$$E_t\left(\widehat{z}_{t+1}\right) = \mu\widehat{z}_t + B\widehat{e}_t \tag{8.82}$$

where $\widehat{z}_t = P^{-1}\widehat{s}_t$ and $B = P^{-1}A$. The transformed system is comprised of two difference equations describing the behaviour of the backward-looking variables, $\widehat{s}_{1t}$, and three difference equations describing the behaviour of the forward-looking variables, $\widehat{s}_{2t}$.

The backward-looking sub-system can be rewritten as:

$$E_t\left(\widehat{z}_{1t+1}\right) = \mu_1\widehat{z}_{1t} + b_1\widehat{e}_t \tag{8.83}$$

where b_1 is implicitly defined by $B = \left[b_1 \mid b_2\right]'$. Since the eigenvalues in μ_1 are less than one in absolute value, equation (83) is stable in the forward direction; furthermore, since $\widehat{s}_{1t}$ is predetermined, the initial conditions completely pin down its solution.

Unfortunately, the forward-looking sub-system:

$$E_t\left(\widehat{z}_{2t+1}\right) = \mu_2 \widehat{z}_{2t} + b_2 \widehat{e}_t \tag{8.84}$$

is stable in the backward direction, since the elements of μ_2 exceed one in absolute value. This means that it is necessary to impose a terminal rather than an initial condition, ie. we have to impose the stochastic TVC.

Rewrite (84) as:

$$\widehat{z}_{2t} = \mu_2^{-1} E_t\left(\widehat{z}_{2t+1}\right) + \mu_2^{-1} b_2 \widehat{e}_t = d_1 E_t\left(\widehat{z}_{2t+1}\right) + d_2 \widehat{e}_t. \tag{8.85}$$

Applying well-known results[24], we can show that the solution to (80) is given by:

$$\widehat{z}_{2t} = -\sum_{k=0}^{\infty} d_1^k d_2 E_t\left(\widehat{e}_{t+k}\right) = -\sum_{k=0}^{\infty} d_1^k d_2 \rho^k \widehat{e}_t = L_{ee} \widehat{e}_t. \tag{8.86}$$

Applying the VEC operator to L_{ee} we easily get:

$$\overrightarrow{L}_{ee} = -\left(1 - \rho' \otimes d_1\right)^{-1} \overrightarrow{d}_2 \tag{8.87}$$

By construction:

$$\widehat{z}_{1t} = q_{11} \widehat{s}_{1t} + q_{12} \widehat{s}_{2t}, \widehat{z}_{2t} = q_{21} \widehat{s}_{1t} + q_{22} \widehat{s}_{2t} \tag{8.88}$$

We can solve the second expression in (88) for $\widehat{s}_{2t}$:

$$\widehat{s}_{2t} = q_{22}^{-1} \widehat{z}_{2t} - q_{22}^{-1} q_{21} \widehat{s}_{1t} \tag{8.89}$$

or:

$$\widehat{s}_{2t} = L_s \widehat{s}_{1t} + L_e \widetilde{e}_t = L_\nu \widetilde{\nu}_t \tag{8.90}$$

where $L_e = q_{22}^{-1} L_{ee}$, $L_s = -q_{22}^{-1} q_{21}$, $L_\nu = [L_s \mid L_e]$, and $\widetilde{\nu}_t = [\widehat{s}_{1t} \mid \widehat{e}_t]'$.

>From (80) we can isolate the first two equations of the original system, given by:

[24] See Sargent (1987).

$$E_t\left(\widehat{s}_{1t+1}\right) = w_{11}\widehat{s}_{1t} + w_{12}\widehat{s}_{2t} + a_1\widehat{e}_t. \qquad (8.91)$$

Since $\widehat{s}_{1t}$ is predetermined in the Blanchard-Kahn sense (expectational error equal to zero), we can rewrite (91) as:

$$\widehat{s}_{1t+1} = w_{11}\widehat{s}_{1t} + w_{12}\widehat{s}_{2t} + a_1\widehat{e}_t. \qquad (8.92)$$

Taking into account (90), we can rearrange (92) as:

$$\widehat{s}_{1t+1} = \left(w_{11} + w_{12}L_s\right)\widehat{s}_{1t} + \left(w_{12}L_e + a_1\right)\widehat{e}_t. \qquad (8.93)$$

Combining (93), (67), and (68) we get:

$$\widetilde{\nu}_{t+1} = M_\nu \widetilde{\nu}_t + i_t \qquad (8.94)$$

where:

$$M_\nu = \begin{bmatrix} \left(w_{11} + w_{12}L_s\right) & \left(w_{12}L_e + a_1\right) \\ 0 & P \end{bmatrix}, i_t = \begin{bmatrix} 0 \\ \left[\epsilon_t \mid \epsilon_t\right]' \end{bmatrix} \qquad (8.95)$$

>From (69) we have that:

$$\widehat{u}_t = Q_{us}\begin{bmatrix} \widehat{s}_{1t} \\ L_\nu \widetilde{\nu}_t \end{bmatrix} + Q_{ue}\widehat{e}_t \qquad (8.96)$$

where $Q_{us} = M_{\mathbf{u}u}^{-1}M_{us}$ and $Q_{ue} = M_{\mathbf{u}u}^{-1}M_{ue}$.
Partitioning the matrix Q_{us} we may rewrite (91) as:

$$\widehat{u}_t = U_\nu \widetilde{\nu}_t \qquad (8.97)$$

where $U_\nu = \left[Q_{us}^1 \mid Q_{ue}\right] + Q_{us}^2 L_\nu$.
Given an exogenous sequence of innovations, the (endogenous and exogenous) state variables evolve according to (94), the control variables according to (97), and the costate variables according to (90). Equation (94) describes a first-order vector autoregression; iterating on (94) and taking into account (90) and (97) we can to recover the sequence of probability distributions that represents the (approximated) solution to our non-linear system of stochastic difference equations.

8.6.4 *Variables of interest*

There are two other variables of interest we would like to recover[25], ie. output and investment, defined as:

$$y_t = a_t k_t^{1-\alpha}, \qquad i_t = a_t k_t^{1-\alpha} - c_t. \tag{8.98}$$

We can log-linearize (98) as:

$$\widehat{y}_t = \widehat{a}_t + (1-\alpha)\,\widehat{k}_t, \qquad \widehat{i}_t = \frac{y}{i}\widehat{a}_t + (1-\alpha)\,\frac{y}{i}\widehat{k}_t - \frac{c}{i}\widehat{c}_t. \tag{8.99}$$

We can write the previous system in a more compact form as:

$$\widehat{f}_t = FV_u \widehat{u}_t + FV_\nu \widehat{\nu}_t + FV_s \widehat{s}_{2t}, \tag{8.100}$$

where $\widehat{f}_t = \left[\widehat{y}_t \mid \widehat{i}_t\right]'$ and:

$$FV_u = \begin{bmatrix} 0 \\ -\frac{c}{i} \end{bmatrix}, FV_\nu = \begin{bmatrix} 1-\alpha & 0 \; 1 \; 0 \\ (1-\alpha)\,\frac{y}{i} & 0 \; \frac{y}{i} \; 0 \end{bmatrix}, FV_s = \begin{bmatrix} 0\,0\,0 \\ 0\,0\,0 \end{bmatrix} \tag{8.101}$$

Substituting (90) and (97) into (100) we can rewrite it more compactly as $\widehat{f}_t = F_\nu \widehat{\nu}_t$, where $F_\nu = FV_u U_\nu + FV_\nu + FV_s L_\nu$. For notational convenience, we define also a new vector $\widehat{h}_t = \left[\widehat{u}_t \mid \widehat{f}_t \mid \widehat{s}_{2t}\right]'$ such that $\widehat{h}_t = H_\nu \widehat{\nu}_t$, where $H_\nu = [U_\nu \mid F_\nu \mid L_\nu]$..

8.6.5 *MATLAB code*

The whole procedure described until now can be easily translated in the MATLAB matrix programming language. First of all, we need a main file, called *main.m*, containing all the building blocks needed to run the KPR procedure itself.

The first part of the file defines the variables' names and their position in the system. Then, it stores the number of variables that belong to each relevant group, ie. the number of controls, states, and so on. Finally, it stores the total number of variables in the system, and the total number of endogenous state and costate variables.

```
% we create a vector containing the names of all the variables in
the model,
```

[25]Note thet, since the real wage is simply output times a constant, their dynamic properties will coincide; there is no value-added in studying the real wage separately, and we simply drop it from our variables of interest.

```
% and store the position of each variable in this vector:
name=['k ';'b ';'a ';'e ';'c ';'y ';'i ';'p ';'r ';'l '];
kp=1; bp=2; ap=3; ep=4; cp=5; yp=6; ip=7; pp=8; rp=9; lp=10;
% we store also the dimension of all relevant vectors (nc: controls,
% ns: states, nl: costates, nn: exo. states, nxf: vars of interest,
% nvar: total number of vars, nt: states+costates):
nc=1; ns=2; nl=3; nn=2; nxf=2; nvar=ns+nl+nc+nn+nxf; nt=nl+ns;
```

The second part of the file stores the complete parameterization used in our exercise, and creates the covariance matrix of the vector of innovations defined as i_t in equation (90).

```
% Our parameterization:
mu=2; % risk aversion
sn=0.6; % alpha
sc=0.75; % c/y share
g=1.004; % growth rate
rky=13.28; % capital-output ratio
rmy=5.3; % money income velocity
rby=0.6; % bonds-output ratio
eta=1.013; % money growth rate
rho=[0.95 0;0 0.49]; % persistence of shocks
covar=[1 0;0 1]; % covariance matrix
% we define the covar. matrix of iota in equation (90):
lmat=zeros(ns+nn);
lmat(ns+1:ns+nn,ns+1:ns+nn)=chol(covar)';
```

The third part solves for the steady-state and obtains the value of all calibrated parameters.

```
% calibration and steady-state:
sk=1-sn; % 1-alpha
si=1-sc; % i/y share
d=1-g+si/rky; % delta
be=g/(sk/rky+1-d); % beta tilde
r=(g?be)/be; % real int. rate
theta=((eta?be)/be)*(rmy/sc)^mu; % pref. param.
rvy=(1?be)/(g*be)*rby+(1-eta)*rmy; % v/y ratio
vpi=(g/be)*(rky/sk); % capital delta
vphi=rmy/rby;
% we store a vector of steady-state values:
yss=rky^(sk/sn); kss=rky*yss; bss=rby*yss;
css=sc*yss; iss=si*yss; lss=css^(-mu);
pss=((eta/be?1)*(lss/theta))^(1/mu);
steady=[kss;bss;1;eta;css;yss;iss;pss;r;lss];
```

Finally, the fourth part stores the matrices that describe the linearized system that will be solved by the KPR procedure.

```
% we define the matrices describing the linearized system:
muu=-mu;
mus=[0,0,0,0,1];
mue=[0,0];
ms0=[-sn,0,0,0,vpi; 0,0,mu*(1-be/eta)-1,0,be/eta; 0,0,0,r/(1+r),1;...
         g*rky,0,0,0,0; 0,g,0,0,0];
ms1=[0,0,0,0,-vpi; 0,0,1,0,-1; 0,0,0,0,-1; -(1-d)*rky-sk,0,0,0,0;...
         0,-g/be,(1-eta)*vphi,-r,0];
mu0=[0;0;0;0;0];
mu1=[0;0;0;-sc;0];
me0=[-1,0;0,0;0,0;0,0;0,0];
me1=[0,0;0,1;0,0;1,0;0,-eta*vphi];
fvu=[0;-sc/si];
fvv=[sk,0,1,0;sk/si,0,1/si,0];
fvs=[0,0,0;0,0,0];
```

The KPR procedure is then implemented by two external functions. The first, *kpr.m*, performs step by step the algorithm described in the previous paragraphs, following as strictly as possible the notation:

```
function [h,mv]=kpr(muu,mus,mue,mu0,mu1,ms0,ms1,me0,me1,fvu,fvv,fvl,rho);
nl=size(fvl,2); % recover the number of costate vars
nt=size(mus,2); % recover the number os state+costate vars
ns=nt-nl; % recover the number of state vars
qus=muu\mus;
que=muu\mue;
qus1=qus(:,1:ns);
qus2=qus(:,ns+1:nt);
msss0=ms0-mu0*qus;
msss1=ms1-mu1*qus;
msse0=me0+mu0*que;
msse1=me1+mu1*que;
w=-msss0\msss1;
a=msss0\(msse0*rho+msse1);
[lv,mv]=bk(w,a,rho,ns);
uv=[[qus1,que]+qus2*lv];
fv=fvu*uv+fvv+fvl*lv;
h=[uv;fv;lv];
end;
```

The second, *bk.m*, performs the core Blanchard-Khan solution algorithm, describe in par. 1.6.3, following again the same notation:

```
function [lv,mv]=bk(w,a,rho,ns)
```

```
nt=size(w,1);
nl=nt-ns;
nn=size(rho,1);
w11=w(1:ns,1:ns);
w12=w(1:ns,ns+1:nt);
[evec,eval]=eig(w);
[mu1,ind]=sort(abs(diag(eval)));
mu=diag(eval);
mu=mu(ind);
p=evec(:,ind);
mu2=diag(mu(ns+1:nt));
ps=p\eye(size(p));
b=ps*a;
d1=mu2\eye(size(mu2));
d2=d1*b(ns+1:nt,:);
lee=-(eye(nl*nn)-kron(rho',d1))\d2(:);
lee=reshape(lee,nl,nn);
ls=-ps(ns+1:nt,ns+1:nt)\ps(ns+1:nt,1:ns);
le=ps(ns+1:nt,ns+1:nt)\lee;
lv=[ls,le];
mv=[w11+w12*ls,w12*le+a(1:ns,:);zeros(nn,ns),rho];
end;
```

Once the *main.m* file has been run, we recall *kpr.m*, which in turn recalls *bk.m*, to obtain the matrices that characterize the approximated solution.

8.7 Impulse response functions

Once the approximated solution for the chosen parameterization is available, we may be interested in studying the effect of an unexpected shock to one of the exogenous state variables, for example a 1% increase in TFP.

Assume that the system at date 0 is in steady-state, ie. $\widehat{\nu}_0 = 0$; iterating on (89 and assuming $\epsilon_j = 0$ for $j = 1, 2, ..., t$, we get that:

$$\widehat{\nu}_t = M_\nu^t \epsilon_0 \tag{8.102}$$

This implies that $d\widehat{\nu}_t / d\epsilon_0'$ represents the effect on of an unexpected shock at date 0, if for $j = 1, 2, ..., t$. The derivative $\partial \widehat{\nu}_{it} / \partial \epsilon_{j0} = [M_\nu^t]_{ij}$, considered as a function of time, is called the impulse response function of the state variable i for shocks to the state variable j, if all other shocks at all other dates are zero. Given (99, it is easy to recover the impulse response functions of all control variables and all variables of interest.

To operationally perform the impulse response analysis, we have first of all to choose the state variable to shock; then, we have to define an initial vector of innovations, for example $\epsilon_0 = [0 \mid 0 \mid 1 \mid 0]'$ if we are assuming a positive shock to TFP. Then, given ϵ_0 and $\widehat{\nu}_0 = 0$, we obtain from (89, $\widehat{u}_1$, $\widehat{f}_1$, and $\widehat{l}_1$ from (98. Assuming $\epsilon_t = 0\ \forall t \geq 1$, we iterate the procedure for finite number T of periods; finally, we plot the simulate series, obtaining the impulse response functions.

The MATLAB code that performs this task is contained in a file called *impulse.m.*

```
clear;
main;
n=60; % Simulation horizon
pr=0; % Flag: if 0, plot on screen, if 1, save on disk
shockto=['A ';'Eta'];
sim=zeros(nvar,n);
[h,mv]=kpr(muu,mus,mue,mu0,mu1,ms0,ms1,me0,me1,fvu,fvv,fvl,rho);
for i=1:2;
% Build a matrix of innovations:
sck=zeros(nn+ns,n+1);
sck(i+2,1)=1;
s=sck(:,1);
% Simulate the system:
for j=1:n;
sim(:,j)=[s;h*s];
s=mv*s+sck(:,j+1);
end;
% Plot the results:
t=1:n;
subplot(3,1,1), hnd=plot(t,sim(yp,:)','k-',t,sim(cp,:)','k-.',t,sim(ip,:)','k:');
legend(hnd,'y','c','i',1);
title(['Shock to ' shockto(i,:)]);
ylabel('% Deviation');
subplot(3,1,2), hnd=plot(t,sim(pp,:)','k:');
legend(hnd,'p',1);
subplot(3,1,3), hnd=plot(t,sim(kp,:)','k?',t,sim(bp,:)','k:');
legend(hnd,'k','b',1);
if pr==0
pause;
else
eval(['print -dbitmap fig' num2str(i)]);
end;
end;
```

The results are plotted in Figures 1-2.

8.8 Stochastic simulations

To estimate the small sample stochastic properties of the model, we can perform a Montecarlo experiment. In other words, we can draw from a random number generator a finite sequence of innovations (the simulation horizon is typically T=100) and iterate to get the simulate series for all exogenous and endogenous variables.

To isolate the dynamic behavior of our model at business cycles frequencies, we filter the simulated series applying the so called Hodrik-Prescott (HP) filter, with a smoothing parameter equal to 1600. Then, we calculate all statistics of interests, for example the relative standard deviation of each variable with regard to output, the autocorrelation coefficient, the correlation coefficient with output, and so on. We repeat this procedure for at least 1000 times, storing each round the results in a matrix. Finally, we summarize the empirical distribution of our statistics of interest calculating their mean, standard deviation, and so on, across the 1000 replications. In Appendix ? we provide a MATLAB program that performs a set of 1000 stochastic simulations on a 100 periods horizon, calculating the simulate distribution of the relative standard deviations, autocorrelations, and correlations with output for all main variables. Results are summarized in Tables 1-3.