A Solution Manual to
The Econometrics of Financial Markets

Petr Adamek

John Y. Campbell

Andrew W. Lo

A. Craig MacKinlay

Luis M. Viceira

Author address:

MIT Sloan School, 50 Memorial Drive, Cambridge, MA 02142–1347

Department of Economics, Harvard University, Littauer Center, Cambridge, MA 02138

MIT Sloan School, 50 Memorial Drive, Cambridge, MA 02142–1347

Wharton School, University of Pennsylvania, 3620 Locust Walk, Philadelphia, PA 19104–6367

Department of Economics, Harvard University, Littauer Center, Cambridge, MA 02138

Contents

List of Figures

List of Tables

Preface

The problems in *The Econometrics of Financial Markets* have been tested in PhD courses at Harvard, MIT, Princeton, and Wharton over a number of years. We are grateful to the students in these courses who served as guinea pigs for early versions of these problems, and to our teaching assistants who helped to prepare versions of the solutions. We also thank Leonid Kogan for assistance with some of the more challenging problems in Chapter 9.

Problems in Chapter 2

Solution 2.1

2.1.1 Recall the martingale property given by (2.1.2) and observe that the mean-squared error of the time-t forecast X_t of price P_{t+1} is

$$\text{(S2.1.1)} \qquad \mathrm{E}[(X_t - P_{t+1})^2 | P_t, \ldots] \;=\; (X_t - P_t)^2 + \mathrm{E}[P_{t+1}^2 - P_t^2 | P_t, \ldots] \;.$$

This expression is minimized by the forecast $X_t \equiv P_t$.

2.1.2 Let $l > 0$. Then

$$\text{(S2.1.2)} \qquad \begin{aligned} \mathrm{E}[(P_t - P_{t-k})(P_{t-k-l} - P_{t-2k-l})] &= \mathrm{E}\big[\, \mathrm{E}[P_t - P_{t-k} | P_{t-k-l}, \ldots] \times \\ &\qquad\qquad (P_{t-k-l} - P_{t-2k-l}) \,\big] \\ &= \mathrm{E}[0(P_{t-k-l} - P_{t-2k-l})] = 0 \;. \end{aligned}$$

Solution 2.2

Denote the martingale property (2.1.2) by M. Then

$$\text{(S2.2.1)} \qquad \mathrm{RW1} \Rightarrow \mathrm{RW2} \Rightarrow \mathrm{M} \Rightarrow \mathrm{RW3},$$

and no other implication holds in general. For example, consider the following counter-examples. Let $\{\xi_n\}_{n=1}^{\infty}$ be a sequence of random variables drawn independently from a uniform distribution over the interval $[-1, 1]$ and $\xi_0 = 0$. Then the process with increments (i) $\epsilon_{2n-1} \equiv \xi_n$ and $\epsilon_{2n} \equiv |\xi_n| - 1/2$ satisfies RW3 but not M; (ii) $\epsilon_n \equiv \xi_n \xi_{n-1}$ satisfies M but not RW2; (iii) $\epsilon_n \equiv n\xi_n$ satisfies RW2 but not RW1.

Solution 2.3

A necessary condition for the log-price process p_t in (2.2.9) to satisfy RW1 is $\alpha + \beta = 1$. Let $c \equiv \alpha + \beta$ and consider the set of all non-RW1 Markov processes (2.2.9), i.e., $c \neq 1$. The restriction CJ $= 1$ is equivalent to $\alpha\beta = c/4$. The constraints $0 \leq \alpha, \beta \leq 1$ are satisfied exactly for $c \in [1, 4/3]$ and therefore the set of all two-state Markov chains represented by the pair (α, β) that cannot support any RW1 process but still yields CJ $= 1$ is simply

$$\text{(S2.3.1)} \qquad \{(1 \pm \sqrt{1 - c^{-1}}, 1 \mp \sqrt{1 - c^{-1}})c/2; 1 < c \leq 4/3\}.$$

Such Markov chains do generate sequences, reversals, etc.

Solution 2.4

For a stationary process, $\mathrm{Var}[Z_t] = \mathrm{Var}[Z_{t-k}]$ and $\mathrm{Cov}[Z_{t-k}, Z_{t-l}] = \mathrm{Cov}[Z_t, Z_{t-l+k}]$. Thus, we have

$$\text{(S2.4.1)} \qquad \mathrm{Var}[Z_t(q)] \;=\; \sum_{k=0}^{q-1} \mathrm{Var}[Z_{t-k}] \;+\; 2\sum_{k=1}^{q-1} (q-k)\mathrm{Cov}[Z_t, Z_{t-k}]$$

which yields (2.4.19). The coefficients of $\mathrm{Cov}[Z_t, Z_{t-k}]$ are simply the number of k-th order autocovariance terms in the variance of the multiperiod return $Z_t(q)$ (recall that this multiperiod return is the sum of q one-period returns). The coefficients decline linearly

Ten individual stocks used for problem 2.5, identified by CRSP permanent number PERMNO, CUSIP identifier, (most recent) ticker symbol and abbreviation of full name.

PERMNO	CUSIP	Ticker	COMPANY NAME
18075	*03203710*	AP	AMPCO-PITTSBURGH CORP.
30840	*21161520*	CUO	CONTINENTAL MATERIALS CORP.
26470	*29265N10*	EGN	ENERGEN CORP.
32096	*36480210*	GAN	GARAN INC.
19174	*37006410*	GH	GENERAL HOST CORP.
12095	*37083810*	GSX	GENERAL SIGNAL CORP.
15747	*45870210*	IK	INTERLAKE CORP.
12490	*45920010*	IBM	INTERNATIONAL BUSINESS MACHS. CORP.
18286	*75510310*	RAY	RAYTECH CORP. DE
15472	*98252610*	WWY	WRIGLEY, WILLIAM JR. CO.

TABLE 2.1. Ten individual stocks for Problem 2.5

Periods P_0 to P_4 for daily and monthly data.

	Daily Periods		*Monthly Periods*	
Period	Calendar Days	Length (Days)	Calendar Days	Length (Months)
A	620703–941230	8179	620731–941230	390
A_1	620703–700923	2045	620731–700831	98
A_2	700924–781027	2045	700930–780929	97
A_3	781030–861128	2044	781031–861031	97
A_4	861201–941230	2045	861128–941230	98

TABLE 2.2. Periods for Daily and Monthly Data

with k until they reach zero for $k = q$ because there are successively fewer and fewer higher-order autocovariances.

From (2.4.19) it is apparent that individual autocorrelation coefficients can be non-zero but their weighted average can be zero. For example, according to (2.4.19), $\mathrm{VR}(3) = 1 + 2(\frac{2}{3}\rho_1 + \frac{1}{3}\rho_2)$, hence a non-random-walk process with $\rho_1 = -\frac{1}{4}$ and $\rho_2 = \frac{1}{2}$ will satisfy $\mathrm{VR}(3) = 1$. Therefore, the variance ratio test will have very lower power against such alternatives, despite the fact that they violate the random walk hypothesis.

Solution 2.5

We consider the daily and monthly returns of the ten individual stocks considered in Chapter 1 (see Table 1.1). We use CRSP daily data consisting of 8,179 days from July 3, 1962 to December 30, 1994 and CRSP monthly data consisting of 390 months from July 31, 1962 to December 30, 1994. For these ten stocks there are 23 missing daily returns and 4 missing monthly returns in our sample. The stocks are identified in Table 2.1, and we shall refer to them by their ticker symbols (value-weighted and equal-weighted indexes will be denoted by VW and EW).

Denote the entire sample period by A and the four consecutive subperiods of approximately equal length by A_1, A_2, A_3, A_4, respectively (note that these periods differ for daily and monthly data). Descriptions of lengths and starting and ending dates of the periods are given in Table 2.2.

Statistics for daily and monthly simple and continuously compounded returns. All the statistics $\hat{\mu}$, $\hat{\sigma}$, and $\hat{\rho}(1)$ are reported in percent.

Security		Simple Returns						Cont. Comp. Returns					
		Daily Sampling			Monthly Sampling			Daily Sampling			Monthly Sampling		
	Period	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}(1)$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}(1)$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}(1)$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\rho}(1)$
VW	A	0.044	0.803	19.4	0.96	4.37	4.8	0.041	0.807	19.3	0.85	4.38	5.8
	A_1	0.035	0.635	25.9	0.76	3.82	6.2	0.033	0.635	25.9	0.68	3.82	7.1
	A_2	0.026	0.814	29.4	0.71	4.65	5.3	0.022	0.813	29.4	0.60	4.60	6.6
	A_3	0.070	0.821	16.0	1.40	4.58	-4.7	0.067	0.822	16.0	1.29	4.55	-4.7
	A_4	0.045	0.913	11.0	0.95	4.36	9.8	0.040	0.822	11.1	0.85	4.49	-7.2
EW	A	0.078	0.685	38.6	1.25	5.67	22.0	0.075	0.687	38.7	1.08	5.67	22.2
	A_1	0.069	0.728	38.8	1.18	5.44	18.9	0.066	0.728	38.8	1.03	5.44	20.3
	A_2	0.060	0.696	49.5	1.37	6.61	19.6	0.057	0.696	49.5	1.16	6.39	21.2
	A_3	0.082	0.644	33.1	1.54	5.51	16.4	0.079	0.646	33.1	1.38	5.55	15.0
	A_4	0.100	0.669	30.4	0.91	5.00	33.6	0.097	0.676	30.9	0.77	5.22	31.7
AP	A	0.053	2.411	-3.7	1.06	10.62	0.3	0.024	2.396	-4.1	0.52	10.31	0.2
	A_1	0.076	2.983	-6.2	1.33	12.14	1.9	0.032	2.952	-6.9	0.63	11.58	2.4
	A_2	0.070	2.313	-6.5	1.54	9.15	-7.1	0.044	2.295	-6.9	1.13	8.82	-6.6
	A_3	0.042	1.701	7.5	0.94	9.50	3.9	0.028	1.694	7.4	0.50	9.34	3.7
	A_4	0.024	2.472	-2.8	0.45	11.37	-0.9	-0.007	2.470	-3.2	-0.18	11.19	-2.1
CUO	A	0.143	5.239	-20.9	1.65	17.76	-7.0	0.009	5.155	-21.8	0.19	16.96	-9.7
	A_1	0.241	6.722	-26.7	2.02	18.92	-1.6	0.022	6.590	-28.5	0.46	17.14	-5.1
	A_2	0.191	6.699	-29.2	1.39	19.60	-11.1	-0.027	6.577	-30.2	-0.25	17.52	-12.4
	A_3	0.140	3.523	9.5	3.11	18.47	-3.5	0.079	3.497	8.4	1.43	18.36	-7.9
	A_4	-0.001	2.692	13.8	0.09	13.22	-14.1	-0.038	2.714	15.2	-0.88	14.49	-12.8
EGN	A	0.054	1.407	-6.8	1.09	5.75	-7.0	0.044	1.405	-6.9	0.94	5.55	-6.8
	A_1	0.022	1.083	-8.3	0.43	3.65	-14.9	0.051	1.080	-8.3	0.36	3.60	-14.7
	A_2	0.047	1.636	-12.4	0.97	6.80	3.2	0.034	1.637	-12.5	0.76	6.19	4.4
	A_3	0.091	1.437	-8.6	1.79	5.45	-12.9	0.081	1.434	-8.8	1.63	5.31	-13.2
	A_4	0.056	1.415	2.8	1.20	6.51	-14.0	0.046	1.411	2.8	0.99	14.49	-12.4
GAN	A	0.079	2.349	4.4	1.65	11.30	2.8	0.051	2.333	4.1	1.03	10.92	4.2
	A_1	0.088	2.886	8.2	1.76	14.12	12.0	0.047	2.852	7.8	0.84	13.34	14.3
	A_2	0.085	2.729	-0.7	1.95	11.71	-5.7	0.047	2.723	-1.2	1.29	11.18	-5.1
	A_3	0.106	1.918	-0.3	1.95	8.64	-2.5	0.088	1.910	-0.2	1.56	8.61	-1.0
	A_4	0.036	1.614	14.8	0.93	9.94	4.9	0.023	1.601	15.1	0.44	9.94	5.5
GH	A	0.070	2.790	-2.2	1.33	11.65	6.3	0.032	2.768	-2.4	0.66	11.53	5.7
	A_1	0.069	3.103	-6.0	1.06	11.91	3.4	0.022	3.074	-6.2	0.35	12.00	3.6
	A_2	0.060	2.936	4.3	1.30	12.68	19.1	0.018	2.890	3.7	0.55	12.05	17.6
	A_3	0.060	2.389	-1.0	3.27	10.94	-12.5	0.126	2.373	-1.2	2.64	10.94	-12.0
	A_4	-0.001	2.677	-6.0	-0.29	10.69	5.3	-0.037	2.682	-5.6	-0.87	10.78	5.4
GSX	A	0.054	1.660	11.6	1.17	8.18	2.7	0.040	1.661	11.7	0.83	8.21	3.7
	A_1	0.063	1.866	7.4	1.45	9.04	1.6	0.046	1.862	7.3	0.89	8.82	2.1
	A_2	0.055	1.710	19.7	1.37	8.96	6.6	0.041	1.710	19.8	1.05	8.85	7.8
	A_3	0.042	1.600	6.5	0.92	6.75	-7.2	0.042	1.599	6.5	0.69	6.67	-5.9
	A_4	0.042	1.436	13.4	1.03	7.74	3.6	0.031	1.443	14.1	0.70	8.30	4.9
IK	A	0.043	2.156	0.4	0.86	9.37	-6.5	0.020	2.145	0.3	0.43	9.22	-5.0
	A_1	0.031	1.395	-0.7	0.69	6.42	-15.1	0.022	1.391	-0.8	0.49	6.31	-14.0
	A_2	0.064	1.475	6.0	0.71	7.19	-2.2	0.040	1.470	5.8	1.12	6.79	-3.9
	A_3	0.102	1.441	8.6	2.12	4.58	-6.4	0.041	1.431	8.5	1.85	7.03	-6.7
	A_4	-0.025	3.518	-1.8	-0.73	14.18	-8.4	0.031	3.498	-2.1	-1.73	14.02	-7.2
IBM	A	0.039	1.423	-0.4	0.81	6.17	6.6	0.029	1.427	-0.4	0.61	6.19	6.9
	A_1	0.068	1.257	6.2	1.39	5.62	6.9	0.060	1.255	6.2	1.22	5.56	7.5
	A_2	0.028	1.355	3.8	0.66	5.97	-1.6	0.019	1.351	3.8	0.48	5.90	-1.4
	A_3	0.058	1.375	-6.7	1.10	5.46	14.4	0.048	1.370	-6.7	0.95	5.37	14.5
	A_4	0.002	1.670	-2.8	0.07	7.35	5.1	-0.012	1.690	-2.8	-0.20	7.57	5.1
RAY	A	0.050	3.388	-0.6	0.83	14.88	-12.0	-0.008	3.362	-1.4	-0.13	13.65	-11.7
	A_1	0.014	1.426	10.3	0.32	6.63	15.2	0.004	1.449	10.0	0.08	6.97	18.0
	A_2	0.062	1.914	12.1	1.53	8.59	-9.9	0.043	1.904	12.1	1.17	8.47	-8.7
	A_3	-0.014	3.051	8.6	-0.43	15.28	-20.7	-0.060	3.027	7.9	-1.57	15.00	-18.1
	A_4	0.137	5.558	-5.6	1.88	22.98	-11.1	-0.014	5.505	-6.6	-0.23	19.83	-12.5
WWY	A	0.072	1.446	5.6	1.51	6.67	2.8	0.061	1.447	5.6	1.29	6.55	2.0
	A_1	0.026	0.864	5.2	0.56	3.61	-3.8	0.022	0.862	5.2	0.50	3.58	-3.2
	A_2	0.036	1.355	12.4	0.90	7.47	17.6	0.027	1.361	12.3	0.62	7.40	15.7
	A_3	0.110	1.510	7.2	2.20	6.21	-10.0	0.099	1.504	7.1	1.99	6.10	-10.3
	A_4	0.116	1.868	0.5	2.49	8.25	-5.4	0.098	1.873	0.7	2.05	8.08	-5.6

TABLE 2.3. Statistics for Daily and Monthly Simple and Continuously Compounded Returns

2.5.1 See the left side of Table 2.3 for the required statistics.

2.5.2 See the right side of Table 2.3. If r denotes net simple return in percent, then $100 \times \log(1 + r/100)$ is the corresponding continuously compounded return in percent.

2.5.3 See Figure 2.1 for the required plots. Returns are truncated to fit in the interval $[-3\%, 3\%]$, i.e., returns smaller than -3% are replaced by a return of -3%, and returns larger than 3% are replaced by a return of 3%.

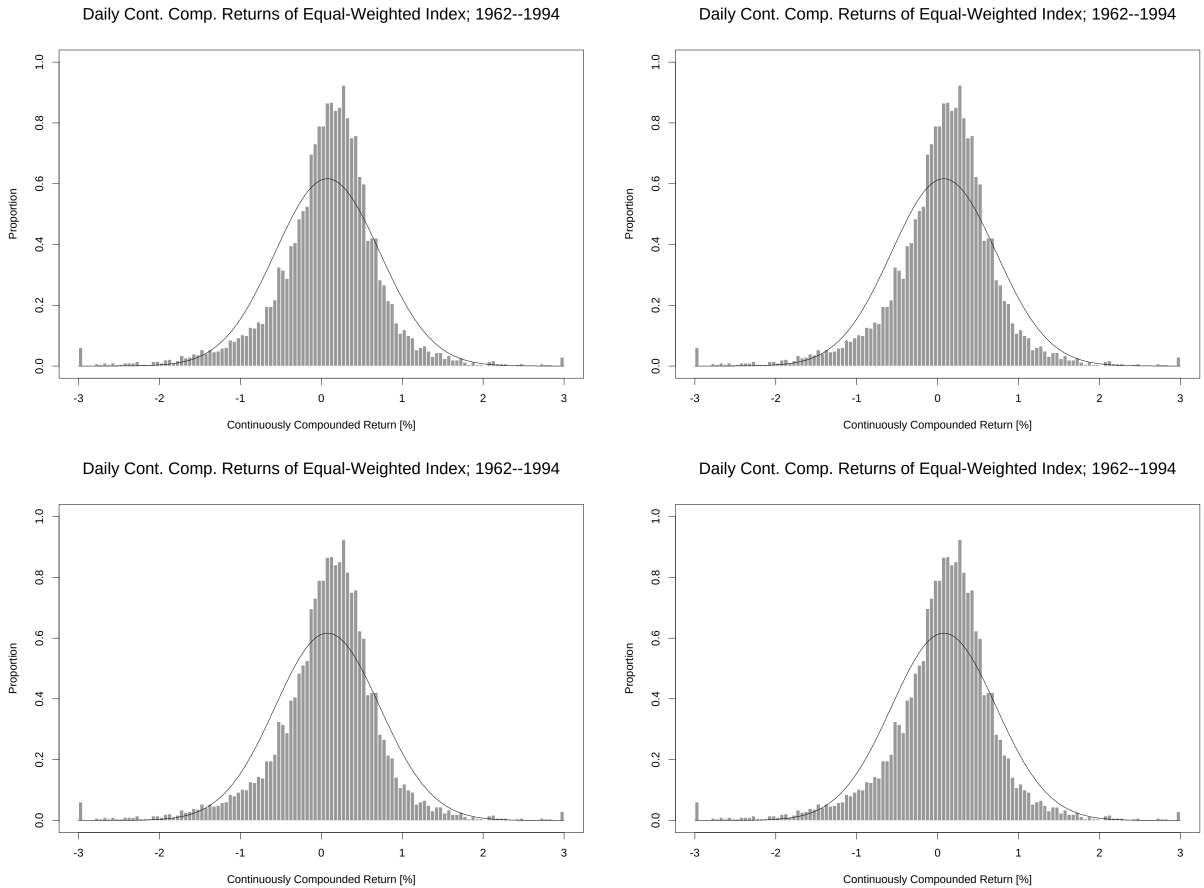

FIGURE 2.1. Histograms of returns on indexes; 1962–1994

2.5.4 Use results from Problems 2.5.1 and 2.5.2 (i.e., Table 2.3), counts from Table 2.2, the assumption that returns are IID, and the asymptotic normality of $\hat{\mu}$ to obtain estimates of the 99% confidence intervals.

2.5.5 Compute the statistics of interest as in Problem 2.5.1. See also Table 1.1 in the text for the statistics for the entire sample period. The variances of your estimates can be estimated via the bootstrap (see Efron and Tibshirani [1993]) under the assumption that returns are temporally IID. Computing the exact variances for estimators of skewness, kurtosis, and the studentized range is possible under certain distributional assumptions for returns, but is quite involved so the bootstrap is the preferred method—it simplifies the estimation greatly: no additional sampling theory is needed). Use the asymptotic normality of your estimators to perform the tests.

Problems in Chapter 3

Solution 3.1

Using (3.1.4), we obtain

$$\text{(S3.1.1)} \qquad \mathrm{E}[r_{it}^o] = \sum_{k=0}^{\infty} \mathrm{E}[X_{it}(k)]\mathrm{E}[r_{i,t-k}] = \sum_{k=0}^{\infty} (1-\pi_i)\pi_i^k \mu_i = \mu_i$$

as in (3.1.9). Observe that

$$\begin{aligned} \mathrm{E}[(r_{it}^o)^2] &= \sum_{k,l=0}^{\infty} \mathrm{E}[X_{it}(k)X_{it}(l)]\mathrm{E}[r_{i,t-k}r_{i,t-l}] \\ &= \sum_{k=0}^{\infty} \mathrm{E}[X_{it}(k)]\mathrm{E}[r_{i,t-k}^2] + \\ \text{(S3.1.2)} \qquad & \quad 2\sum_{k=0}^{\infty}\sum_{l>k} \mathrm{E}[X_{it}(k)X_{it}(l)]\mathrm{E}[r_{i,t-k}]\mathrm{E}[r_{i,t-l}] \\ &= \sigma_i^2 + 2\mu_i^2 \sum_{k=0}^{\infty}\sum_{l>k}(1-\pi_i)\pi_i^k\pi_i^{l-k} \\ &= \sigma_i^2 + 2\mu_i^2/(1-\pi_i) \ , \end{aligned}$$

hence

$$\text{(S3.1.3)} \qquad \mathrm{Var}[r_{it}^o] = \mathrm{E}[(r_{it}^o)^2] - \mu_i^2 = \sigma_i^2 + 2\pi_i\mu_i^2/(1-\pi_i)$$

as in (3.1.10). Next, for $n > 0$ we have

$$\begin{aligned} \mathrm{E}[r_{it}^o r_{i,t-n}^o] &= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \mathrm{E}[X_{it}(k)X_{i,t-n}(l)]\mathrm{E}[r_{i,t-k}r_{i,t-n-l}] \\ \text{(S3.1.4)} \qquad &= \sum_{k=0}^{n-1}\sum_{l=0}^{\infty}(1-\pi_i)\pi_i^k(1-\pi_i)\pi_i^l\mu_i^2 \\ &= \mu_i^2(1-\pi_i^n) \ , \end{aligned}$$

which yields the first part of (3.1.11). For $i \neq j$ and $n \geq 0$ we have

$$\begin{aligned} \mathrm{E}[r_{it}^o r_{j,t-n}^o] &= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \mathrm{E}[X_{it}(k)]\mathrm{E}[X_{j,t-k}(l)] \\ \text{(S3.1.5)} \qquad &= \mu_i\mu_j + \sum_{l=0}^{\infty}(1-\pi_i)\pi_i^{l+n}(1-\pi_j)\pi_j^l\beta_i\beta_j\sigma_f^2 \\ &= \mu_i\mu_j + \pi_i^n\frac{(1-\pi_i)(1-\pi_j)}{1-\pi_i\pi_j}\beta_i\beta_j\sigma_f^2 \end{aligned}$$

and the second part of (3.1.11) follows. Equation (3.1.12) is direct consequence of (3.1.10) and the first part of (3.1.11).

Solution 3.2

Consider the case where the common factor f_t is the (observable) market portfolio. Then the true beta of security i is β_i as in (3.1.1) and the beta computed from observed returns β_i^o is given by

$$\begin{aligned} \beta_i^o &= \operatorname{Cov}[r_{it}^o, f_t]/\operatorname{Var}[f_t] \\ &= \mathrm{E}[f_t \sum_{k=0}^{\infty} X_{it}(k) r_{i,t-k}]/\sigma_f^2 \\ &= (1-\pi_i)\beta_i \ . \end{aligned} \tag{S3.2.1}$$

Thus, the beta will be biased towards 0 if nonsynchronous trading is not properly accounted for.

Solution 3.3

3.3.1 Let P_{it} and Q_{it} denote the unconditional probabilities that $\delta_{it} = 0$ and $\delta_{it} = 1$, respectively and let P_i and Q_i be the corresponding steady-state probabilities. Then (3.5.1) yields

$$P_{it} = \pi_i P_{i,t-1} + (1-\pi_i')Q_{i,t-1} \tag{S3.3.1}$$

and similarly for Z. In steady state, $P_i = P_{it} = P_{i,t-1}$ and $Q_i = Q_{it} = Q_{i,t-1}$, hence

$$\begin{aligned} P_i &= \frac{1-\pi_i'}{2-(\pi_i+\pi_i')} \\ Q_i &= \frac{1-\pi_i}{2-(\pi_i+\pi_i')} \ . \end{aligned} \tag{S3.3.2}$$

Therefore, the unconditional steady-state mean, variance, and first-order autocorrelation of δ_{it} is

$$\mu_{\delta i} = Q_i \tag{S3.3.3}$$

$$\sigma_{\delta i}^2 = \mathrm{E}[\delta_{it}^2] - \mathrm{E}[\delta_{it}]^2 = Q_i(1-Q_i) \tag{S3.3.4}$$

$$\gamma_{\delta ii}(1) = \mathrm{E}[\delta_{it}\delta_{i,t-1}] - \mu_{\delta i}^2 = Q_i(\pi_i' - Q_i) \ . \tag{S3.3.5}$$

3.3.2 To calculate the statistics of observed returns, use (3.1.4). For the mean we have

$$\begin{aligned} \mu_{Ri} &= \sum_{k=0}^{\infty} \mathrm{E}[X_{it}(k)]\mathrm{E}[R_{i,t-k}] \\ &= \mu_i \left(\pi_i' Q_i + \sum_{k=1}^{\infty} P_i \pi_i^{k-1}(1-\pi_i) \right) = \mu_i(\pi_i' Q_i + P_i) \ , \end{aligned} \tag{S3.3.6}$$

for the variance

$$\begin{aligned} \sigma_{Ri}^2 &= \mathrm{E}\left[\left(\sum_{k=0}^{\infty} X_{it}(k) R_{i,t-k} \right)^2 \right] - \mu_{Ri}^2 \\ &= \mu_i^2 \left(\pi_i' Q_i + \sum_{k,l=1}^{\infty} \mathrm{E}[X_{it}\,(\max(k,l))] \right) - \mu_{Ri}^2 \\ &= \mu_i^2 \left(\frac{P_i}{\pi_i(1-\pi_i)} + \pi_i' Q_i - (\pi_i' Q_i + P_i)^2 \right) \ , \end{aligned} \tag{S3.3.7}$$

and for the first-order autocovariance

$$\gamma_{ii}(1) = \mathrm{E}\left[\sum_{k,l=0}^{\infty} X_{it}(k)X_{i,t+1}R_{i,t-k}R_{i,t-k-1}\right] - \mu_{Ri}^2$$

$$= \mu_i^2 \sum_{l=0}^{\infty} \pi_i' \mathrm{E}[X_{i,t-1}(l)] - \mu_{Ri}^2 \tag{S3.3.8}$$

$$= \mu_i^2 \left(\pi_i' - (\pi_i' Q_i + P_i)^2\right) \ .$$

Thus, serial correlation in δ_{it} decreases the mean μ_{Ri} as compared to the case of no nonsynchronous-trading effects, *ceteris paribus*, since $\pi_i' Q_i + P_i < 1$ for $0 < \pi_i, \pi_i' < 1$.

3.3.3 Assume we are given a sequence $\{\delta_{it}\}_{t=0}^{T}$ of no-trade indicators. For convenience, we shall condition on the initial no-trade indicator δ_{i0} (the extension to the general case is straightforward). Denote by $n_{i00}, n_{i01}, n_{i10}$, and n_{i00} the counts of all pairs of consecutive days with no-trade patterns '00','01','10', and '11', respectively. Therefore, $\sum_{j,k=0,1} n_{ijk} = T$. Since δ_{it} follows the Markov process (3.5.1), the log-likelihood function of the sequence $\{\delta_{it}\}_{t=1}^{T}$ is

$$\mathcal{L}\left(\{\delta_{it}\}_{t=1}^{T} | \delta_{i0}\right) = n_{i00}\log\pi_i + n_{i01}\log(1-\pi_i) + n_{i10}\log(1-\pi_i') + n_{11}\log\pi_i' \ . \tag{S3.3.9}$$

The maximum likelihood estimators of π_i, π_i' are

$$\hat{\pi}_i = \frac{n_{i00}}{n_{i00}+n_{i01}} \qquad \hat{\pi}_i' = \frac{n_{i11}}{n_{i10}+n_{i11}} \ . \tag{S3.3.10}$$

and the Fisher information matrix is

$$i(\pi_i, \pi_i') = \mathrm{E}\left[\begin{pmatrix} \frac{n_{i00}}{\pi_i^2} + \frac{n_{i01}}{(1-\pi_i)^2} & 0 \\ 0 & \frac{n_{i10}}{(1-\pi_i')^2} + \frac{n_{i11}}{\pi_i'^2} \end{pmatrix}\right] \tag{S3.3.11}$$

so that our estimates $\hat{\pi}_i, \hat{\pi}_i'$ are asymptotically independent and normal, with asymptotic variances estimated efficiently by

$$\hat{\sigma}_{\hat{\pi}_i}^2 = \left(\hat{\pi}_i(1-\hat{\pi}_i)(n_{i00}+n_{i01})\right)^{-1} \qquad \hat{\sigma}_{\hat{\pi}_i'}^2 = \left(\hat{\pi}_i'(1-\hat{\pi}_i')(n_{i10}+n_{i11})\right)^{-1} \ . \tag{S3.3.12}$$

The results of the empirical analysis are given in Tables 3.1 and 3.2. In Table 3.1, the non-trading counts are reported for six securities using five years of daily data from January 4, 1988 to December 31, 1992. The data was extracted from the CRSP daily master file: out of 1,120 ordinary common shares continuously listed on the NYSE over this time span, 360 did not trade at least on one of the NYSE trading dates, and 56 did not trade on at least 100 days out of 1,517 days in total. Our sample is a randomized selection of six stocks from the latter set. Values for n_{i0} and n_{i1}, defined analogously to n_{i00} etc., are also provided for convenience. Note that n_{i01} and n_{i10} coincide in some cases.

Estimates of π_i and π_i' are given in Table 3.2.

Solution 3.4

Let the Markov process for I_t be given by the transition probability matrix

$$C \equiv \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}, \tag{S3.4.1}$$

with steady-state probabilities of I_t being -1 and 1 given by $P \equiv (1-q)/(2-p-q)$ and $Q \equiv (1-p)/(2-p-q)$, respectively.

Input Data for Problem 3.1.3. Representative sample from infrequently traded (at least 100 no-trade days in the sample interval) ordinary common shares continuously listed on the NYSE from 1988 to 1992. Each stock is identified by its ticker symbol and CUSIP number. Counts of days a stock did not trade n_{i0}, did trade n_{i1}, and patterns of non-trading for all pairs of consecutive days n_{i00}, n_{i01}, n_{i10}, n_{i11} are reported.

Ticker	CUSIP	n_{i0}	n_{i1}	n_{i00}	n_{i01}	n_{i10}	n_{i11}
ZMX	*98991710*	390	1127	129	260	261	867
UNF	*90470810*	244	1273	67	177	177	1096
JII	*47936810*	220	1297	70	150	150	1147
MBC	*59478010*	173	1344	25	148	148	1196
ADU	*02342610*	136	1381	33	103	103	1278
LVI	*50243910*	117	1400	35	82	82	1318

TABLE 3.1. Input Data for Problem 3.1.3

Parameter Estimates for Problem 3.1.3. Maximum likelihood estimates of probabilities π_i, π'_i for representative sample of infrequently traded stock are reported, together with estimates of their standard deviations σ.

Ticker	$\hat{\pi}_i$	$\hat{\sigma}_{\hat{\pi}_i}$	$\hat{\pi}'_i$	$\hat{\sigma}_{\hat{\pi}'_i}$
ZMX	0.327	0.108	0.769	0.071
UNF	0.275	0.143	0.861	0.081
JII	0.318	0.145	0.884	0.087
MBC	0.145	0.216	0.890	0.087
ADU	0.243	0.200	0.925	0.102
LVI	0.299	0.202	0.941	0.114

TABLE 3.2. Resulting Statistics for Problem 3.1.3

Then ΔP_t is a *four*-state Markov process where the quasi-state $\Delta P_t = 0$ is in fact two distinct states according to whether the pair (I_{t-1}, I_t) is $(-1,-1)$ or $(1,1)$. In the steady state, we have the following transition probability matrix:

$$\text{(S3.4.2)} \qquad \begin{pmatrix} 0 & p & 1-p \\ \frac{(1-p)q(1-q)}{p(1-q)+q(1-p)} & \frac{p^2(1-q)+q^2(1-p)}{p(1-q)+q(1-p)} & \frac{p(1-p)(1-q)}{p(1-q)+q(1-p)} \\ 1-q & q & 0 \end{pmatrix}.$$

The moments of ΔP_t are then

$$\text{(S3.4.3)} \qquad \begin{aligned} \mathrm{E}[\Delta P_t] &= 0, \\ \mathrm{Var}[\Delta P_t] &= \frac{2s^2(1-p)(1-q)}{2-p-q}, \\ \mathrm{Cov}[\Delta P_t, \Delta P_{t-k}] &= -\frac{s^2(1-p)(1-q)}{2-p-q} \times \\ &\qquad (1,-1)C^{k-1}(1-p, q-1)', \quad k>0, \\ \mathrm{Corr}[\Delta P_t, \Delta P_{t-k}] &= \frac{-(1,-1)C^{k-1}(1-p,q-1)'}{2(2-p-q)}, \quad k>0\ . \end{aligned}$$

Observe that the first autocorrelation coefficient equals $-1/2$ as in the IID case, but the higher-order autocorrelations are nonzero in general.

Solution 3.5

We will show how discreteness can influence and bias several popular stock price statistics. Consider a stock with a *virtual* price process that follows a continuous geometric Brownian motion, with a net expected annual return and standard deviation of return (*not* continuously compounded) of $\mu = 10\%$ and $\sigma = 20\%$, respectively.

Assume the observer has available daily-sampled prices rounded to the closest eighth of a dollar (or to \$0.125 if the virtual price is less than \$0.125) for a period of ten years. For purposes of this exercise we neglect the complications of non-trading days and assume the year consists of 253 equally-spaced trading days.

We shall focus on the estimator $\hat{\mu}_i$ of the expected annual returns defined as rescaled *arithmetic* average of *daily* returns, and the estimator $\hat{\sigma}$ of the volatility of annual returns defined as a rescaled standard deviation of *daily* returns. The rescaling is as follows: the average of daily returns is multiplied by the number of trading days 253, and the standard deviation of daily returns is multiplied by $\sqrt{253}$. While such estimators might be suitable for slowly-changing and continuous price processes, they are badly biased estimators of the theoretical 10% expected return and 20% standard deviation, respectively.

Expressing the parameters of the underlying geometric Brownian motion process as

$$\mu' = \log \frac{\mu + 1}{\sqrt{1 + \sigma/(\mu + 1)}} \quad \text{(S3.5.1)}$$

$$\sigma' = \log\left(1 + \sigma/(\mu + 1)\right) \quad \text{(S3.5.2)}$$

and running 4,000 replications of the simulation described above, we report the means of the statistics for a hypothetical stock with various initial prices in the Table 3.3.

Estimates of return, standard deviation, and autocorrelation are highly biased for low-priced stocks. Indeed, the hypothetical \$0.25 stock exhibits apparent return of almost 50%. For higher stock prices the discreteness biases subside. Nevertheless, we see that even for high-priced stocks the estimates are still biased due to the way we rescaled daily estimates to yield annual figures (these estimates would be unbiased if we had assumed arithmetic instead of geometric Brownian motion).

Problem 3.5 shows that the effects of price discreteness can be substantial for stock-return statistics and that appropriate care has to be taken to avoid such biases.

Solution 3.6

3.6.1 From the histogram of IBM transaction stock prices on January 4th and 5th, 1988 (Figure 3.1) we observe price clustering around \$120 and \$123. These clusters correspond to trades taking place on different days.

On the other hand, the histogram of price *changes* (Figure 3.2) does not exhibit any apparent clustering, leaving aside the discretization to eighths of dollars (or "ticks"), i.e., the smallest price variation possible from one trade to the next. We see that most of changes fall in the range from -2 to $+2$ ticks.

When we compare the two histograms of price changes conditional on prices falling on an odd or an even eighth (Figure 3.3), we see a different pattern: there are fewer zero-tick price changes that fall on odd eighths than on even eighths, and relatively more one-tick price changes that fall on odd eighths than on even eighths. Overall, even-eighth prices are significantly more frequent then odd-eighth ones. These regularities underscore the potentially important impact that discreteness can have on statistical inference for transactions data.

3.6.2 The histogram of times between trades for IBM stock (Figure 3.4) shows that the majority of trades take place within intervals shorter than one minute. Based on $n = 2{,}746$ time intervals, the estimate of the expected time between trades is $\hat{\mu}_\Delta = 16.86$ and the estimate of the standard deviation of the time between trades is $\hat{\sigma}_\Delta = 19.46$. The 95%

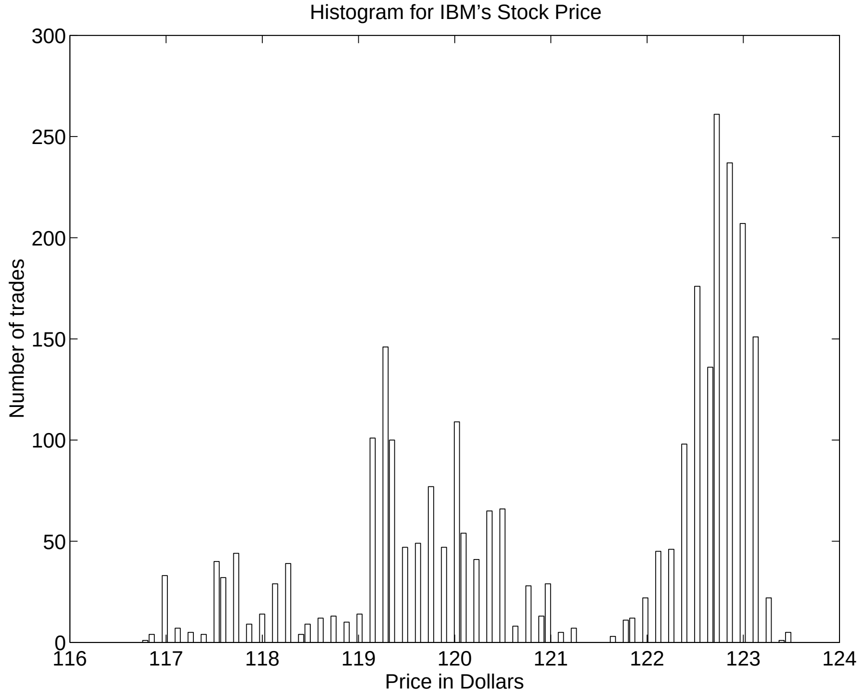

FIGURE 3.1. Histogram for IBM Stock Price

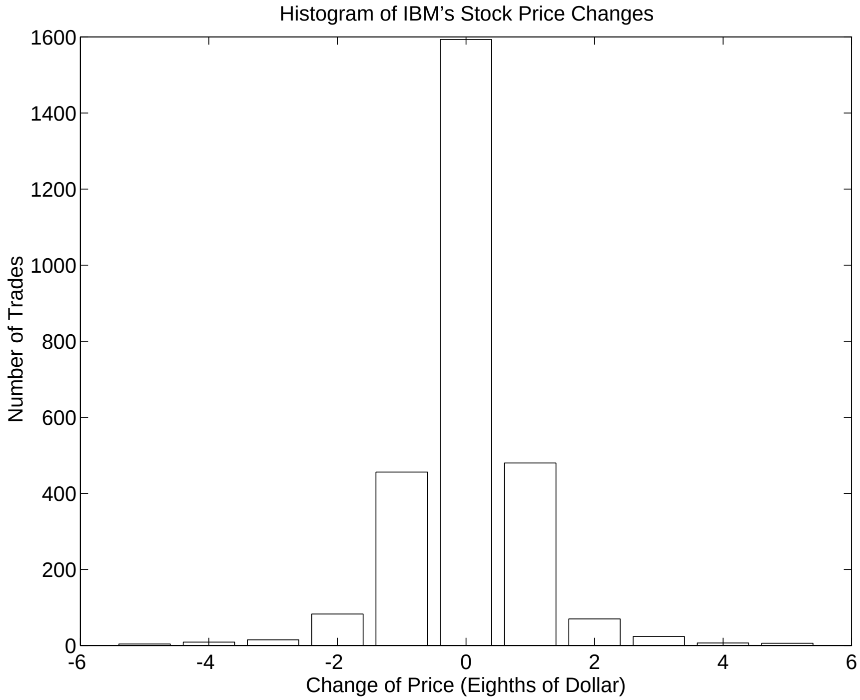

FIGURE 3.2. Histogram of IBM Price Changes

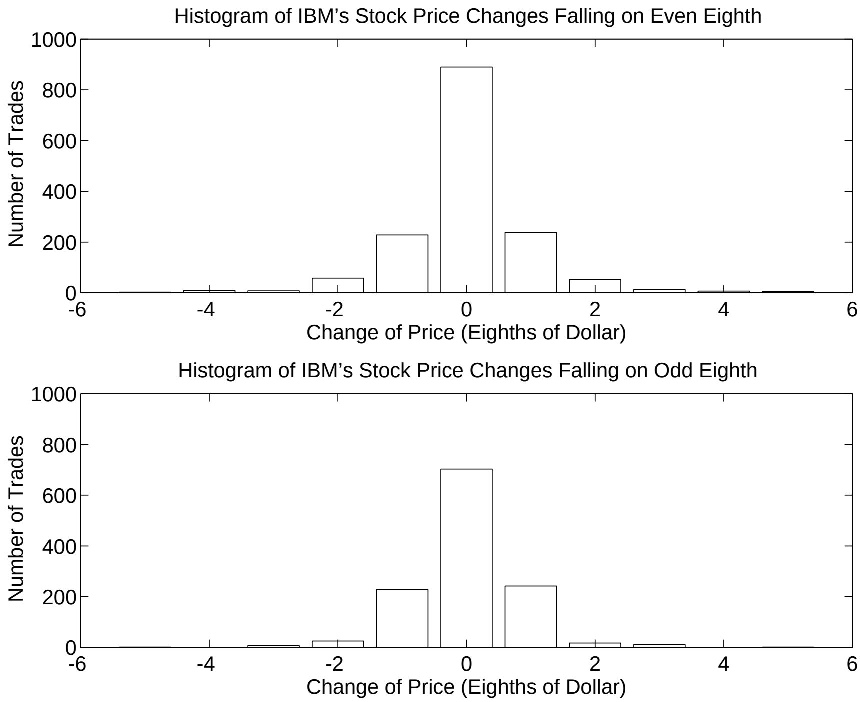

FIGURE 3.3. Histogram of IBM Price Changes Falling on Odd or Even Eighth

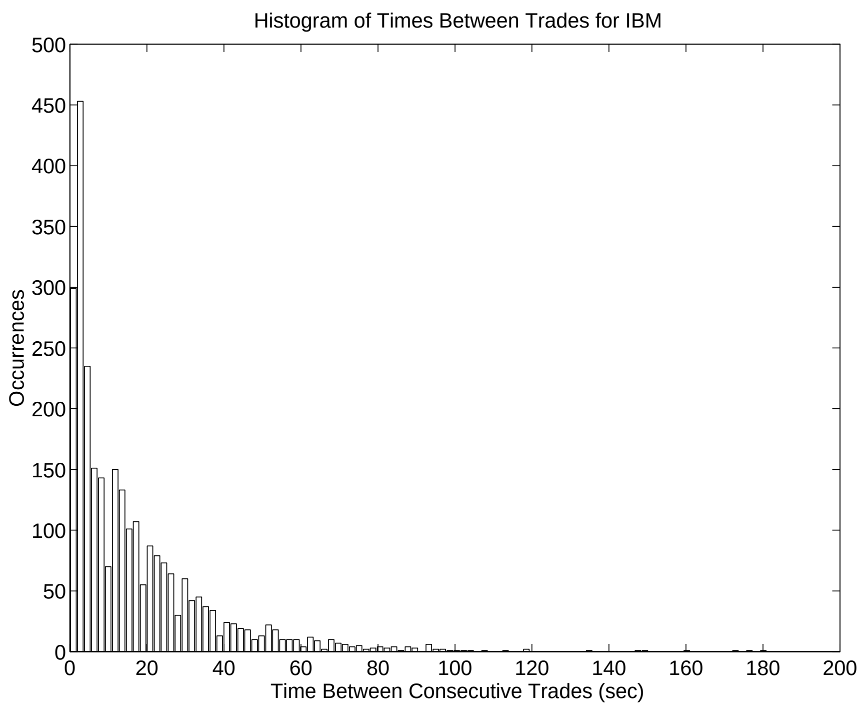

FIGURE 3.4. Histogram of Times Between Trades for IBM

Simulations Results for Problem 3.5. The impact of discretization of prices to a \$1/8 grid on naive estimates of annual mean and standard deviation based on daily returns data is simulated for a hypothetical stock following a continuous time geometric Brownian motion price process with an annual expected return of 10% and an annual standard deviation of 20%. For a low-priced stock discreteness biases are substantial. For a high-priced stock the main source of bias is the misspecification of the process for purposes of estimation, e.g. taking arithmetic means instead of geometric. The statistics are based on 4,000 replications of 10 years of daily price data for each row.

Initial price	Expected Return	Standard Deviation	Autocorrelation
0.25	0.4949 (0.0043)	0.9170 (0.0049)	-0.2609 (0.0010)
0.50	0.2893 (0.0012)	0.6454 (0.0021)	-0.2821 (0.0007)
1.00	0.1847 (0.0006)	0.4537 (0.0012)	-0.2886 (0.0005)
2.00	0.1327 (0.0007)	0.3236 (0.0008)	-0.2734 (0.0005)
5.00	0.1038 (0.0008)	0.2210 (0.0003)	-0.1562 (0.0009)
10.00	0.0977 (0.0009)	0.1918 (0.0001)	-0.0562 (0.0006)
20.00	0.0963 (0.0009)	0.1834 (0.0000)	-0.0167 (0.0003)
50.00	0.0969 (0.0009)	0.1808 (0.0000)	-0.0033 (0.0003)
100.00	0.0934 (0.0009)	0.1805 (0.0000)	-0.0010 (0.0003)

TABLE 3.3. Simulation Results for Problem 3.5

confidence interval for the expected time can be therefore estimated as

$$(\hat{\mu}_\Delta - 1.96\hat{\sigma}_\Delta/n^{1/2}, \hat{\mu}_\Delta + 1.96\hat{\sigma}_\Delta/n^{1/2}) = (16.23, 17.49) \ . \qquad \text{(S3.6.1)}$$

Suppose that trade times follow a Poisson process with parameter λ. That is, assume that the probability P_k of exactly k trades occurring during any one-minute interval is given by

$$P_k = e^{-\lambda}\frac{\lambda^k}{k!} \ . \qquad \text{(S3.6.2)}$$

The sample average time between trades $\hat{\mu}_\Delta$ is a sufficient statistic for λ; in fact, $\hat{\lambda} = 60/\hat{\mu}_\Delta$ is a consistent and efficient estimator of λ. Note that the number 60 is the result of rescaling time from seconds to minutes. For our sample, $\hat{\lambda} = 3.56$. We can map the 95% confidence interval (S3.6.1) of μ_Δ derived above into the following 95% confidence interval for λ

$$\left(\frac{60}{\hat{\mu} + 1.96\hat{\sigma}}, \frac{60}{\hat{\mu} - 1.96\hat{\sigma}}\right) = (3.43, 3.69) \ . \qquad \text{(S3.6.3)}$$

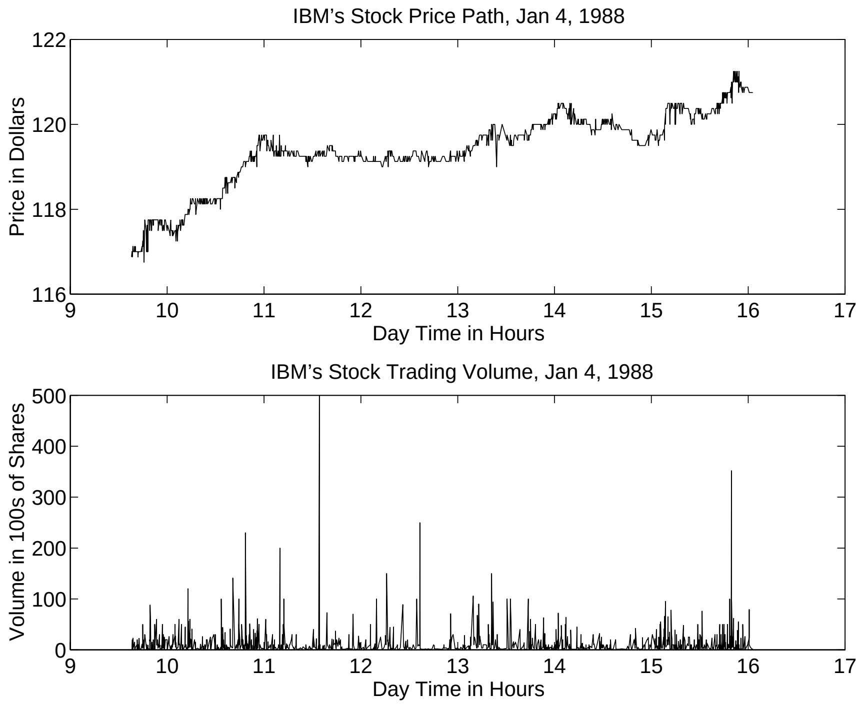

FIGURE 3.5. IBM Price and Volume on Jan 4, 1988

Note that this confidence interval is not centered on $\hat{\lambda}$, but has the advantage of following so directly from the confidence interval for $\hat{\mu}_\Delta$. As $n \to \infty$, both $\hat{\mu}_\Delta$ and $\hat{\lambda}$ are asymptotically normal consistent estimates of μ_Δ and λ.

It also follows from the definition of the Poisson distribution that the probability of no trade during a one-minute interval can be estimated by

$$\hat{P}_0 = e^{-\hat{\lambda}} = 0.0285 \tag{S3.6.4}$$

with a 95% confidence interval of $(0.0248, 0.0324)$.

Dividing the two trading dates into one-minute intervals and counting the number of trades, we get a total of 776 minutes (excluding possible opening and closing lags each day). A trade occurred in 733 of them. Furthermore, 697 minutes in which a trade occurred were immediately preceded by a minute in which a trade occurred as well. Therefore, the estimate of the probability of a trade occurring within a particular minute is 0.0554 with a 95% confidence interval of $(0.0393, 0.0715)$, and the estimate of the probability of a trade occurring within a particular minute conditional on a trade occurring in previous minute is 0.0491 with a 95% confidence interval of $(0.0335, 0.0648)$. Estimates of conditional and unconditional probabilities do not differ statistically significantly, hence we cannot reject the hypothesis of independence on these grounds. On the other hand, there is a statistically significant discrepancy between these sample probabilities and the estimate based on the Poisson assumption. Thus, we can reject the independence of trades in that sense.

3.6.3 Plots of price and volume against time-of-day for both days exhibit certain patterns (Figures 3.5, and 3.6). Price discreteness is visible from its price path; volume exhibits

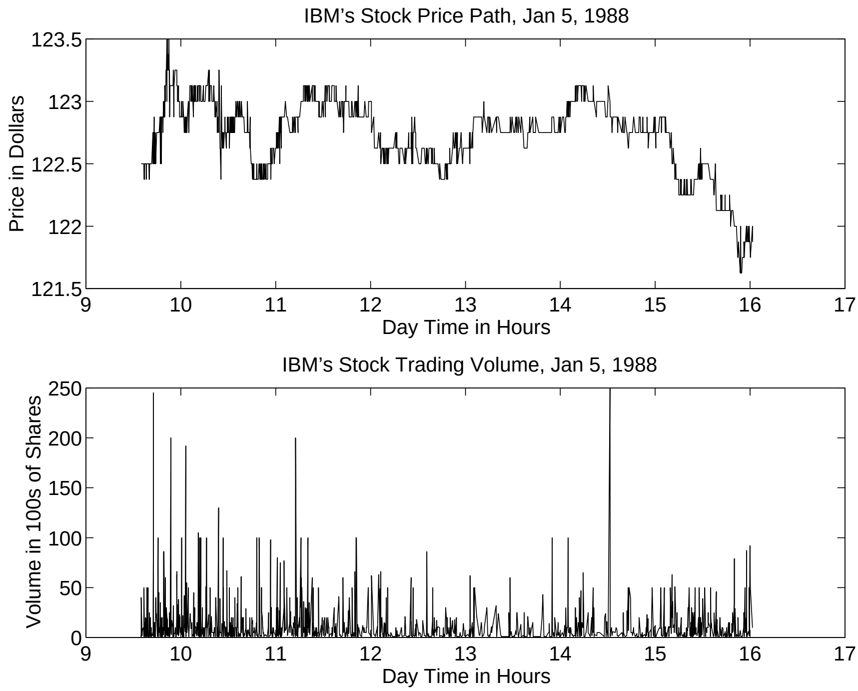

FIGURE 3.6. IBM Price and Volume on Jan 5, 1988

large skewness and kurtosis; there is apparently less volume around lunchtime. Time-of-day phenomena are probably untestable from a sample of two days. There is no *apparent* relationship between price movements and volume visible by naked eye.

Consider the simple qualitative hypothesis that large-volume trades are accompanied by price movements of different magnitude than small-volume trades. Let us partition sample of 2,746 trades into $n_b = 42$ *block* trades (trades that are greater than or equal to 100 round lots) and $n_s = 2{,}704$ smaller trades, and compute the sample means $\hat{\delta}_b$, $\hat{\delta}_s$ and standard errors of *absolute price changes* immediately following the trades, expressed in dollars:

$$\begin{aligned} \hat{\delta}_b &= 0.0446 \quad (0.0102) \\ \hat{\delta}_s &= 0.0675 \quad (0.0019)\ . \end{aligned} \tag{S3.6.5}$$

The difference of these averages is 0.0229 with a standard error of 0.0104, which is significantly different from zero at the 5% level. Therefore, trading volume is indeed linked to subsequent price changes. Note that block trades are followed by smaller price changes than the majority of small volume trades.

3.6.4 Consider the following simple model for estimating the price impact of selling IBM stock. Assume that we cannot distinguish whether a trade was "seller-initiated" or "buyer-initiated" from the data, so that we will relate only the absolute magnitude of trading volume to the absolute magnitude of price change as in the previous part. Moreover, assume that the (absolute) price impact of a trade is *proportional* to volume, *ceteris paribus*, and that errors of measurement are, after division by volume, independent and identically distributed. Under these strong but simple conditions we can estimate efficiently the coefficient of proportionality ρ between volume and its price impact as the sample mean $\hat{\rho}$ of ratios of absolute price changes to volume, according to the Gauss-Markov theorem.

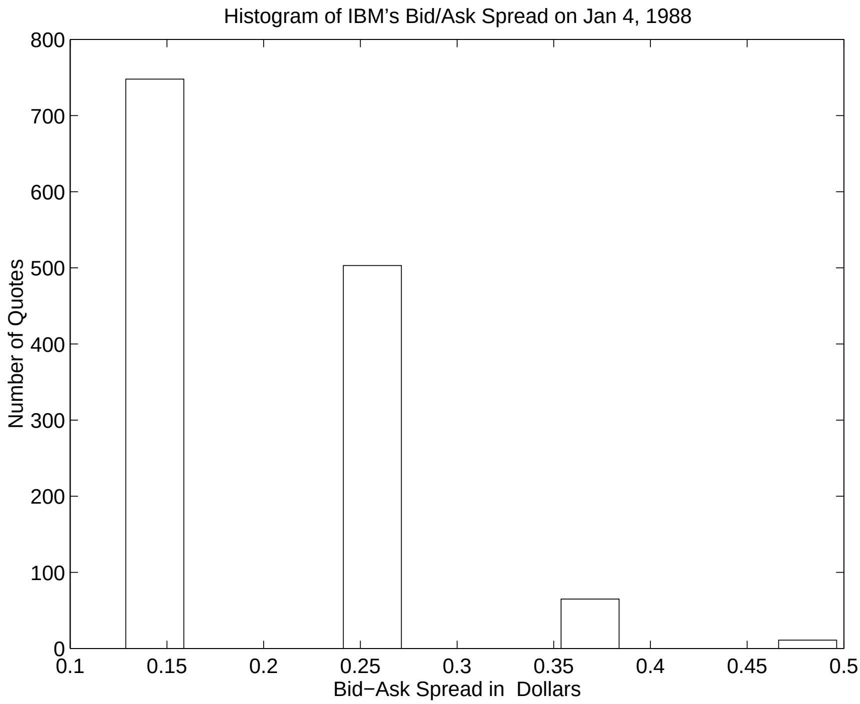

FIGURE 3.7. IBM Bid/Ask Spread Histogram on Jan 5, 1988

In our data $\hat{\rho} = 0.0310$ (in dollars per one round lot) with a standard error of 0.0013. Thus, we can conclude that the seller of one round lot effectively pays three cents per share less than marginal seller. This amount becomes economically interesting in the case of block trades where the loss is of the order of \$3 per \$120-share of stock.

Solution 3.7

3.7.1 The structure of individual bid/ask spread data is particularly simple. During January 4th and 5th, 1988, only four sizes of spreads occurred among 1,327 quotes. There were 748 quotes with a spread of one tick, 502 with two ticks, 65 with three ticks, and 11 with four ticks. A histogram (Figure 3.7) shows that one-tick and two-tick spreads were by far the most common.

Bid-ask spread dynamics are not IID. Table 3.4 displays the empirical distribution of bid/ask spreads both unconditionally and conditionally on the previous quote's spread. It is apparent that the conditional distributions differ significantly from the unconditional distribution.

3.7.2 The question of "causality" between quote revisions and transactions is difficult to answer with the data at hand if we wish to take into account agents' expectations about future events. Thus, for simplicity we shall consider "causality" strictly in the temporal sense: does an increase in the spread come before or after an increase in trading volume, *ceteris paribus*?

First, for simplicity let us measure intensity of transactions activity at any time interval by the number of shares traded in that interval, independently of how the volume is broken up to individual trades and independently of the stock price.

Let us partition the trading day to $n = 1, \ldots, N$, roughly 15-second intervals delimited by a subset of quotes. Let variables s_n, v_n^+ and v_n^- indicate changes in quote spread

Unconditional and Conditional Distributions of Bid/Ask Spreads for IBM stock during January 4th and 5th, 1988. Relative frequencies of bid/ask spreads conditional on preceding quote's spread are expressed in percent. Spreads are denominated in ticks.

Previous	Current Spread			
Spread	1	2	3	4
1	71.4	27.1	1.5	0.0
2	40.0	51.6	7.2	1.2
3	18.5	52.3	23.1	6.2
4	9.1	54.5	27.3	9.1
Any	56.4	37.9	4.9	0.8

TABLE 3.4. Unconditional and Conditional Distributions of Bid/Ask Spreads

and transaction volume related to nth interval. More specifically, let s_n be UP, if quote spread increases between quotes delimiting nth interval, UNCH if spread does not change, and DOWN if spread decreases. Also, v_n^+ is UP if trading volume at interval n is smaller than that in $n+1$, etc. Analogously, v_n^- be UP if trading volume in $n-1$ is smaller than that in n.

Estimates of joint probabilities of s, v^+ and v^- allow some statistical inference about the relation between spreads and transactions. In particular, if quote revisions affect only subsequent transactions but do not influence previous one, the variables s and v^- should be statistically independent. On the other hand, if quotes reflect previous transaction activity, s and v^+ should be statistically independent. The empirical distribution of the 27 triples $[v^-, s, v^+]$, under assumption that their realizations at triples of consecutive 15-second intervals are IID, allow us to test the proposed hypotheses.

However, it is well possible that transaction activity and quote revisions influence temporally *each other*. That being the case, testing existence of unilateral causality may be next to meaningless. Therefore, let us test "causality" in each direction separately, against alternative hypothesis of no relation between quote revisions and transaction activity. In another words, let us test whether variables s and v^+ are dependent, to see whether current quote revisions influence future transactions. Similarly, let us test whether variables s and v^- are dependent, to see whether current quote revision is influenced by past transactions.

Using a standard asymptotic test of independence for a contingency table as described in Rao (1973, pp. 404–412), we have, under the null hypothesis of independence, that:

$$\chi^2_{(r-1)(s-1)} = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{(n_{ij} - n_{i.}n_{.j}/n_{..})^2}{n_{i.}n_{.j}/n_{..}} \tag{S3.7.1}$$

has χ^2 distribution with $(r-1)(s-1)$ degrees of freedom. In our case $r = s = 3$ we have χ^2_4. The contingency tables (Tables 3.5) provide a summary of the data.

It turns out that χ^2_4 statistics for the first table is 13.6, and for the second 17.9 so that we reject the hypothesis of no dependence between s and v^- on 0.9% significance level and that of s and v^+ on 1.2% significance level. Thus we have shown that quote revisions "influence" future transactions, and past transactions "influence" quote revisions.

The next step of the analysis may be to postulate a particular model that involves both effects between transactions and quotes, and perform another round of the statistical analysis.

3.7.3 This part is very similar to 3.7.2, hence we omit the solution.

3.7.4 Let us assume that the investor starts with the bond position and that considers the *quotes* as the relevant price information sense: the investors account only for the

Contingency Tables for causal relationship between transactions activity and quote revisions. The trading day Jan 4, 1988 is divided to quote-to-quote intervals of roughly 15 seconds apart and changes of bid/ask spread together with changes in the trade volume in these intervals are counted. Tables show relationship between past/future changes in transactions activity against the spread change in the intervals.

Past	Current Spread		
Trading	UP	UNCH	DOWN
UP	23	59	10
UNCH	3	12	2
DOWN	21	50	32

Future	Current Spread		
Trading	UP	UNCH	DOWN
UP	20	51	33
UNCH	2	2	2
DOWN	25	68	9

TABLE 3.5. Contingency Tables

rise or decline of the mid-price given as average of the bid/ask spread. Quotes that do not change the mid-price are effectively ignored. Further assume that at the end of the two-day trading period the stock position is liquidated into bonds.

A simulation of such a trading strategy shows: (1) if the investor is allowed to buy and sell at the average, he is left with $101,899 at the end; (2) if the bid/ask prices are used, he is left with $97,769. We see that the bid/ask spread does matter.

It would be difficult to perform any sensible statistical analysis based on the one simulation performed. In particular, it is incorrect to assert that the strategy in (1) that led to nearly a 1% return over one day would dominate a buy-and-hold strategy for a different data set or over a different time span. Nevertheless, the gap between the profits of (1) and (2) are real: frequent trading and large spreads do create significant losses compared to the "frictionless" case.

Problems in Chapter 4

Solution 4.1

Because OLS is consistent, we have $\hat{\delta}_i \to \delta_i$ in probability as $L_1 \to \infty$ from (4.5.3). Thus, for $\hat{\epsilon}_i^*$ from (4.5.7) we have $\hat{\epsilon}_i^* = R_i^* - X_i^* \hat{\delta}_i \to R_i^* - X_i^* \delta_i = \epsilon_i^*$ in probability, as $L_1 \to \infty$. Because abnormal returns ϵ_i^* are independent (across time), the sample abnormal returns $\hat{\epsilon}_i^*$ are asymptotically independent as $L_1 \to \infty$.

Solution 4.2

We assume that the cumulative abnormal return test statistics are calculated using the known standard deviation of the abnormal returns, that the abnormal returns are independent through time and across observations and normally distributed, and that the abnormal returns are measured without parameter sampling error (L_1 is large). Denote $L_2 = 3$ as the length of the event window and N as the number of event observations. Designate group 1 as the observations with low standard deviation and group 2 as the observations with high standard deviation. For the group means and standard deviations we have $\mu_1 = 0.003$, $\mu_2 = 0.003$, $\sigma_1 = 0.03$, and $\sigma_2 = 0.06$ where the subscript indicates the group. $N_1 = 25$ and $N_2 = 25$ are the number of observations in groups 1 and 2 respectively.

To calculate the power against the given alternative, we need to derive the distributions of the test statistics under that alternative. First, we aggregate the abnormal returns over the event window for each observation which gives

(S4.2.1)
$$\mathrm{CAR}_i = \sum_{l=1}^{L_2} \epsilon_{il}^*.$$

Given the assumptions, $\mathrm{E}[\mathrm{CAR}_i] = L_2 \mu_{g(i)}$ and $\mathrm{Var}[\mathrm{CAR}_i] = L_2 \sigma_{g(i)}^2$ where $g(i)$ equals the group of observation i.

Then, we aggregate across observations to form the test statistics (modified to reflect the above assumptions). The aggregation of abnormal returns corresponding to J_1 in (4.4.22) is

(S4.2.2)
$$J_1 = \left[\frac{L_2(N_1\sigma_1^2 + N_2\sigma_2^2)}{N^2}\right]^{-1/2} \left[\frac{1}{N}\sum_{i=1}^{N} \mathrm{CAR}_i\right],$$

The aggregation corresponding to J_2 in (4.4.24) is:

(S4.2.3)
$$J_2 = \sqrt{N}\left[\frac{1}{N}\sum_{i=1}^{N} \frac{\mathrm{CAR}_i}{\sqrt{L_2 \sigma_{g(i)}^2}}\right].$$

Under the specified alternative hypothesis, the distributions of the test statistics J_1 and J_2 are

$$J_1 \sim \mathcal{N}(\mu_1^*, 1) = \mathcal{N}\left(\frac{\sqrt{L_2}(N_1\mu_1 + N_2\mu_2)}{(N_1\sigma_1^2 + N_2\sigma_2^2)^{1/2}}, 1\right), \tag{S4.2.4}$$

$$J_2 \sim \mathcal{N}(\mu_2^*, 1) = \mathcal{N}\left(\frac{\sqrt{L_2}}{\sqrt{N}}\left(N_1\frac{\mu_1}{\sigma_1} + N_2\frac{\mu_2}{\sigma_2}\right), 1\right). \tag{S4.2.5}$$

Substituting in the alternative parameter values, for the means of J_1 and J_2 we have $\mu_1^* = 0.775$ and $\mu_2^* = 0.919$, respectively.

Consider a two-sided test of size α based on J_1 and J_2, respectively, of the null hypothesis $H_0 : [\mu_1 \quad \mu_2] = [0 \quad 0]$ against the alternative hypothesis $H_A : [\mu_1 \quad \mu_2] = [0.003 \quad 0.003]$. Using equation (4.6.1), the powers of the tests, P_1 and P_2 are

$$\begin{aligned} P_1 &= \Pr[J_1 < \Phi^{-1}(\alpha/2)] + \Pr[J_1 > \Phi^{-1}(1-\alpha/2)] \\ &= [\Phi(-\mu_1^* + \Phi^{-1}(\alpha/2))] + [1 - \Phi(\mu_1^* + \Phi^{-1}(1-\alpha/2))], \\ P_2 &= \Pr[J_2 < \Phi^{-1}(\alpha/2)] + \Pr[J_2 > \Phi^{-1}(1-\alpha/2)] \\ &= [\Phi(-\mu_2^* + \Phi^{-1}(\alpha/2))] + [1 - \Phi(\mu_2^* + \Phi^{-1}(1-\alpha/2))]. \end{aligned} \tag{S4.2.6}$$

Evaluation of these expressions for $\alpha = 0.05$ gives $P_1 = 12.1\%$ and $P_2 = 15.1\%$.

Solution 4.3

The solution is the same as for Problem 4.2 except that $\mu_2 = 0.006$ instead of 0.003. Using this value for μ_2, we have $\mu_1^* = 1.162$ and $\mu_2^* = 1.225$ giving $P_1 = 21.3\%$ and $P_2 = 23.3\%$.

Problems in Chapter 5

Solution 5.1

For the regression equation

(S5.1.1) $$R_a = \beta_0 + \beta_1 R_{op} + \beta_2 R_p + \epsilon_p$$

using well-known regression results, we have

(S5.1.2) $$\beta_1 = \mathrm{Cov}[R_a, R_{op}]/\mathrm{Var}[R_{op}] = \beta_{aop},$$

(S5.1.3) $$\beta_2 = \mathrm{Cov}[R_a, R_p]/\mathrm{Var}[R_p] = \beta_{ap},$$

(S5.1.4) $$\beta_0 = \mu_a - (\beta_{aop}\mu_{op} + \beta_{ap}\mu_p),$$

since $\mathrm{Cov}[R_p, R_{op}] = 0$. The result $\beta_2 = \beta_{ap}$ is immediate, thus we need to show that $\beta_1 = 1 - \beta_{ap}$ and $\beta_0 = 0$ to complete the solution.

Let r be the minimum variance portfolio with expected return equal to that of portfolio a, $\mu_a = \mu_r$. From the form of the solution for the minimum variance portfolio weights in (5.2.6), R_r can be expressed as

(S5.1.5) $$R_r = (1-\lambda)R_{op} + \lambda R_p$$

where $\lambda = (\mu_r - \mu_{op})/(\mu_p - \mu_{op})$. Using $\mathrm{Cov}[R_p, R_{op}] = 0$ and $\mu_r = (1-\lambda)\mu_{op} + \lambda\mu_p$ we have

$$\begin{aligned}\beta_{rop} &= \mathrm{Cov}[R_r, R_{op}]/\mathrm{Var}[R_{op}] \\ &= \mathrm{Cov}[(1-\lambda)R_{op} + \lambda R_p, R_{op}]]/\mathrm{Var}[R_{op}] \\ &= (1-\lambda)\end{aligned} \qquad \text{(S5.1.6)}$$

$$\begin{aligned}\beta_{rp} &= \mathrm{Cov}[R_r, R_p]/\mathrm{Var}[R_p] \\ &= \mathrm{Cov}[(1-\lambda)R_{op} + \lambda R_p, R_p]]/\mathrm{Var}[R_p] \\ &= \lambda\end{aligned} \qquad \text{(S5.1.7)}$$

(S5.1.8) $$\mu_r = \beta_{rop}\mu_{op} + \beta_{rp}\mu_p.$$

Portfolio a can be expressed as portfolio r plus an arbitrage (zero-investment) portfolio a^* composed of portfolio a minus portfolio r (long a and short r). The return of a^* is

(S5.1.9) $$R_{a^*} = R_a - R_r.$$

Since $\mu_a = \mu_r$, the expected return of a^* is zero. Because a^* is an arbitrage portfolio with an expected return of zero, for any minimum variance portfolio q, the solution to the optimization problem

(S5.1.10) $$\min_c \mathrm{Var}[R_q + cR_{a^*}]$$

is $c = 0$. Any other solution would contradict q being minimum variance. Noting that $\text{Var}[R_q + cR_{a^*}] = \text{Var}[R_q] + 2c\text{Cov}[R_q, R_{a^*}] + c^2\,\text{Var}[R_{a^*}]$ we have

$$\frac{\partial}{\partial c}\text{Var}[R_q + cR_{a^*}] = 2\text{Cov}[R_q, R_{a^*}] + 2c\,\text{Var}[R_{a^*}]. \tag{S5.1.11}$$

Setting this derivative equal to zero and substituting in the solution $c = 0$ gives

$$\text{Cov}[R_q, R_{a^*}] = 0. \tag{S5.1.12}$$

Thus the return of a^* is uncorrelated with the return of all minimum variance portfolios. Using this result we have

$$\begin{aligned} \text{Cov}[R_a, R_p] &= \text{Cov}[R_r + R_{a^*}, R_p] \\ &= \text{Cov}[R_r, R_p] \end{aligned} \tag{S5.1.13}$$

$$\begin{aligned} \text{Cov}[R_a, R_{op}] &= \text{Cov}[R_r + R_{a^*}, R_{op}]. \\ &= \text{Cov}[R_r, R_{op}] \end{aligned} \tag{S5.1.14}$$

From (S5.1.13) and (S5.1.14) it follows that

$$\beta_{aop} = \beta_{rop} \tag{S5.1.15}$$

$$\beta_{ap} = \beta_{rp}. \tag{S5.1.16}$$

Combining (S5.1.2) with (S5.1.6), (S5.1.7), (S5.1.15) and (S5.1.16) we have $\beta_1 = \beta_{aop} = 1 - \beta_{ap}$. Since $\mu_r = \mu_a$, combining (S5.1.4) with (S5.1.8), (S5.1.15), and (S5.1.16) gives $\beta_0 = 0$ which completes the solution.

Solution 5.2

Begin with the excess return market model from (5.3.1) for N assets. Taking unconditional expectations of both sides and rearranging gives

$$\boldsymbol{\alpha} = \boldsymbol{\mu} - \boldsymbol{\beta}\mu_m. \tag{S5.2.1}$$

Given that the market portfolio is the tangency portfolio, from (5.2.28) we have the $(N \times 1)$ weight vector of the market portfolio

$$\boldsymbol{\omega}_m = \frac{1}{\boldsymbol{\iota}'\Omega^{-1}\boldsymbol{\mu}}\Omega^{-1}\boldsymbol{\mu}. \tag{S5.2.2}$$

Using $\boldsymbol{\omega}_m$ we can calculate the $(N \times 1)$ vector of covariances of the N asset returns with the market portfolio return, the expected excess return of the market, and the variance of the market return,

$$\text{Cov}[\mathbf{Z}, Z_m] = \Omega\boldsymbol{\omega}_m = \frac{1}{\boldsymbol{\iota}'\Omega^{-1}\boldsymbol{\mu}}\boldsymbol{\mu} \tag{S5.2.3}$$

$$\mu_m = \boldsymbol{\omega}_m'\boldsymbol{\mu} = \frac{\boldsymbol{\mu}'\Omega^{-1}\boldsymbol{\mu}}{\boldsymbol{\iota}'\Omega^{-1}\boldsymbol{\mu}} \tag{S5.2.4}$$

$$\text{Var}[Z_m] = \boldsymbol{\omega}_m'\Omega\boldsymbol{\omega}_m = \frac{\boldsymbol{\mu}'\Omega^{-1}\boldsymbol{\mu}}{(\boldsymbol{\iota}'\Omega^{-1}\boldsymbol{\mu})^2}. \tag{S5.2.5}$$

Combining (S5.2.3) and (S5.2.5) we have

$$\boldsymbol{\beta}_m = \frac{\text{Cov}[\mathbf{Z}, Z_m]}{\text{Var}[Z_m]} = \frac{\boldsymbol{\iota}'\Omega^{-1}\boldsymbol{\mu}}{\boldsymbol{\mu}'\Omega^{-1}\boldsymbol{\mu}}\boldsymbol{\mu}. \tag{S5.2.6}$$

and combining (S5.2.6) and (S5.2.4) we have

$$\boldsymbol{\beta}_m\mu_m = \boldsymbol{\mu}. \tag{S5.2.7}$$

From (S5.2.1) and (S5.2.7) the result $\boldsymbol{\alpha} = 0$ is immediate.

Solution 5.3

The solution draws on the statistical analysis of Section 5.3. The calculations for three selected stocks are left to the reader.

Solution 5.4

Let $\mathbf{Z}_t^*$ be a $(N+1 \times 1)$ vector of excess asset returns with mean $\boldsymbol{\mu}^*$ and covariance matrix Ω^*. Designate asset $N+1$ as the market portfolio m. Assume that Ω^* is full rank. (If the market portfolio is a combination of the N included assets, this assumption can be met by eliminating one asset.)

From (5.2.28) the tangency portfolio q of these $N+1$ assets has weight vector

(S5.4.1) $$\boldsymbol{\omega}_q = \frac{1}{\boldsymbol{\iota}'\Omega^{*-1}\boldsymbol{\mu}^*}\Omega^{*-1}\boldsymbol{\mu}^*.$$

Using straight forward algebra we have

(S5.4.2) $$\frac{\mu_q^2}{\sigma_q^2} = \frac{(\boldsymbol{\omega}_q'\boldsymbol{\mu}^*)^2}{\boldsymbol{\omega}_q'\Omega^*\boldsymbol{\omega}_q} = \boldsymbol{\mu}^{*\prime}\Omega^{*-1}\boldsymbol{\mu}^*$$

The covariance matrix Ω^* can be partitioned in the first N assets and the market portfolio,

(S5.4.3) $$\Omega^* \equiv \begin{bmatrix} \Omega & \boldsymbol{\beta}\sigma_m^2 \\ \boldsymbol{\beta}'\sigma_m^2 & \sigma_m^2 \end{bmatrix}$$

(S5.4.4)

(S5.4.5) $$\equiv \begin{bmatrix} \boldsymbol{\beta}\boldsymbol{\beta}'\sigma_m^2 + \Sigma & \boldsymbol{\beta}\sigma_m^2 \\ \boldsymbol{\beta}'\sigma_m^2 & \sigma_m^2 \end{bmatrix}$$

where $\Omega = \boldsymbol{\beta}\boldsymbol{\beta}'\sigma_m^2 + \Sigma$ is substituted.

Using the formula for a partitioned inverse (see Morrison (1990) page 69) we have

(S5.4.6) $$\Omega^{*-1} \equiv \begin{bmatrix} \Sigma^{-1} & \Sigma^{-1}\boldsymbol{\beta} \\ \boldsymbol{\beta}'\Sigma^{-1} & \frac{1}{\sigma_m^2} + \boldsymbol{\beta}'\Sigma^{-1}\boldsymbol{\beta} \end{bmatrix}$$

(S5.4.7)

Using $\boldsymbol{\mu}^{*\prime} = [\boldsymbol{\mu}' \;\, \mu_m]$ and (S5.4.6) we have

(S5.4.8) $$\boldsymbol{\mu}^{*\prime}\Omega^{*-1}\boldsymbol{\mu}^* = \frac{\mu_m^2}{\sigma_m^2} + (\boldsymbol{\mu} - \boldsymbol{\beta}\mu_m)'\Sigma^{-1}(\boldsymbol{\mu} - \boldsymbol{\beta}\mu_m).$$

Substituting $\boldsymbol{\alpha} = \boldsymbol{\mu} - \boldsymbol{\beta}\mu_m$ gives

(S5.4.9) $$\boldsymbol{\mu}^{*\prime}\Omega^{*-1}\boldsymbol{\mu}^* = \frac{\mu_m^2}{\sigma_m^2} + \boldsymbol{\alpha}'\Sigma^{-1}\boldsymbol{\alpha}.$$

From (S5.4.2) and (S5.4.9) we have

(S5.4.10) $$\boldsymbol{\alpha}'\Sigma^{-1}\boldsymbol{\alpha} = \frac{\mu_q^2}{\sigma_q^2} - \frac{\mu_m^2}{\sigma_m^2}$$

which is the result in (5.5.3).

Problems in Chapter 6

Solution 6.1

Let the number of portfolios in the set be K and let $\mathbf{R}_{Kt}$ be the $(K \times 1)$ vector of time period t returns for the portfolios. Since the entire minimum variance boundary can be generated from the K portfolios, for any value of the constant $\mu^{\dagger}$, there exists a combination of the portfolios with expected return $\mu^{\dagger}$ which is minimum variance with respect to the K portfolios plus the N assets. Choose $\mu^{\dagger}$ to be any value but the global minimum variance portfolio expected return (see equation (5.2.11)) and denote this portfolio op. Corresponding to op is a minimum variance portfolio p whose return is uncorrelated with the return of op (see Section 5.2). Since p and op are minimum variance portfolios their returns are linear combinations of the elements of $\mathbf{R}_{Kt}$,

$$R_{pt} = \mathbf{R}'_{Kt}\boldsymbol{\omega}_p^K \tag{S6.1.1}$$

$$R_{opt} = \mathbf{R}'_{Kt}\boldsymbol{\omega}_{op}^K, \tag{S6.1.2}$$

where $\boldsymbol{\omega}_p^K$ and $\boldsymbol{\omega}_{op}^K$ are $(K \times 1)$ vectors of portfolio weights. Because p and op are uncorrelated minimum variance portfolios, we have

$$\boldsymbol{\mu} = \boldsymbol{\iota}\mu_{op} + \boldsymbol{\beta}_p(\mu_p - \mu_{op}) \tag{S6.1.3}$$

where

$$\begin{aligned}
\boldsymbol{\beta}_p &= \frac{\text{Cov}[\mathbf{R}_t, R_{pt}]}{\sigma_p^2} \\
&= \frac{1}{\sigma_p^2}\text{Cov}[\mathbf{R}_t, \mathbf{R}'_{Kt}\boldsymbol{\omega}_p^K] \\
&= \frac{1}{\sigma_p^2}\text{Cov}[\mathbf{R}_t, \mathbf{R}'_{Kt}]\boldsymbol{\omega}_p^K. \qquad \text{(S6.1.4)}
\end{aligned}$$

(See Section 5.2.) Substituting (S6.1.4) into (S6.1.3) gives

$$\boldsymbol{\mu} = \boldsymbol{\iota}\mu_{op} + \text{Cov}[\mathbf{R}_t, \mathbf{R}'_{Kt}]\boldsymbol{\omega}_p^K \frac{(\mu_p - \mu_{op})}{\sigma_p^2}. \tag{S6.1.5}$$

Analogous to (S6.1.5) for the K portfolios we have

$$\boldsymbol{\mu}_K = \boldsymbol{\iota}\mu_{op} + \text{Cov}[\mathbf{R}_{Kt}, \mathbf{R}'_{Kt}]\boldsymbol{\omega}_p^K \frac{(\mu_p - \mu_{op})}{\sigma_p^2}. \tag{S6.1.6}$$

Rearranging (S6.1.6) gives

$$\boldsymbol{\omega}_p^K \frac{(\mu_p - \mu_{op})}{\sigma_p^2} = \text{Cov}[\mathbf{R}_{Kt}, \mathbf{R}'_{Kt}]^{-1}(\boldsymbol{\mu}_K - \boldsymbol{\iota}\mu_{op}). \tag{S6.1.7}$$

Substituting (S6.1.7) into (S6.1.5) gives

$$\boldsymbol{\mu} = \boldsymbol{\iota}\mu_{op} + \text{Cov}[\mathbf{R}_t, \mathbf{R}'_{Kt}]\text{Cov}[\mathbf{R}_{Kt}, \mathbf{R}'_{Kt}]^{-1}(\boldsymbol{\mu}_K - \boldsymbol{\iota}\mu_{op}). \tag{S6.1.8}$$

Now consider the multivariate regression of N assets on K factor portfolios,

$$\mathbf{R}_t = \mathbf{a} + \mathbf{B}\mathbf{R}_{Kt} + \boldsymbol{\epsilon}_t \tag{S6.1.9}$$

where $\mathbf{a}$ is the $(N \times 1)$ intercept vector, $\mathbf{B}$ is the $(N \times K)$ matrix of factor regression coefficients, and $\boldsymbol{\epsilon}_t$ is the time period t residual vector. From regression theory we have

$$\mathbf{B} = \text{Cov}[\mathbf{R}_t, \mathbf{R}'_{Kt}]\text{Cov}[\mathbf{R}_{Kt}, \mathbf{R}'_{Kt}]^{-1} \tag{S6.1.10}$$

$$\mathbf{a} = \boldsymbol{\mu} - \mathbf{B}\boldsymbol{\mu}_K. \tag{S6.1.11}$$

From (S6.1.8) and (S6.1.10), we have

$$\begin{aligned} \boldsymbol{\mu} &= \boldsymbol{\iota}\mu_{op} + \mathbf{B}(\boldsymbol{\mu}_K - \boldsymbol{\iota}\mu_{op}) \\ &= (\boldsymbol{\iota} - \mathbf{B}\boldsymbol{\iota})\mu_{op} + \mathbf{B}\boldsymbol{\mu}_K \end{aligned} \tag{S6.1.12}$$

Since (S6.1.12) holds for different values of μ_{op} it must be the case that $(\boldsymbol{\iota} - \mathbf{B}\boldsymbol{\iota}) = 0$, that is the factor regression coefficients for each asset, including asset a, sum to one. If $(\boldsymbol{\iota} - \mathbf{B}\boldsymbol{\iota}) = 0$, then (S6.1.12) reduces to $\boldsymbol{\mu} = \mathbf{B}\boldsymbol{\mu}_K$ and thus from (S6.1.11) we have $\mathbf{a} = 0$, that is the regression intercept will be zero for all assets including asset a.

Solution 6.2

Let $\boldsymbol{\mu}^*$ and Ω^* be the mean excess return vector and the covariance matrix respectively for the N assets and portfolio p,

$$\boldsymbol{\mu}^{*\prime} \equiv \begin{bmatrix} \boldsymbol{\mu} \\ \\ \mu_p \end{bmatrix} \tag{S6.2.1}$$

$$\Omega^* \equiv \begin{bmatrix} \Omega & \boldsymbol{\beta}'\sigma_p^2 \\ \\ \boldsymbol{\beta}\sigma_p^2 & \sigma_p^2 \end{bmatrix} \tag{S6.2.2}$$

(S6.2.3)

$$\equiv \begin{bmatrix} \boldsymbol{\beta}\boldsymbol{\beta}'\sigma_p^2 + \Sigma & \boldsymbol{\beta}'\sigma_p^2 \\ \\ \boldsymbol{\beta}\sigma_p^2 & \sigma_p^2 \end{bmatrix} \tag{S6.2.4}$$

where $\Omega = \boldsymbol{\beta}\boldsymbol{\beta}'\sigma_p^2 + \Sigma$ and $\Sigma = \boldsymbol{\delta}\boldsymbol{\delta}'\sigma_h^2 + \mathbf{I}\sigma_\epsilon^2$.

Given the $N+1$ assets, the maximum squared Sharpe ratio is $\boldsymbol{\mu}^{*\prime}\Omega^{*-1}\boldsymbol{\mu}^*$ which is the squared Sharpe ratio of the tangency portfolio. As demonstrated in problem **5.4**, given $\boldsymbol{\mu} = \mathbf{a} + \boldsymbol{\beta}\mu_p$ and $\Omega = \boldsymbol{\beta}\boldsymbol{\beta}'\sigma_p^2 + \Sigma$ this ratio can be expressed as

$$s_I^2 = \boldsymbol{\mu}^{*\prime}\Omega^{*-1}\boldsymbol{\mu}^* = \frac{\mu_p^2}{\sigma_p^2} + \mathbf{a}'\Sigma^{-1}\mathbf{a} \tag{S6.2.5}$$

where s_I^2 is the maximum squared Sharpe ratio for economy I, $I = A, B$. Analytically inverting $\Sigma = \boldsymbol{\delta}\boldsymbol{\delta}'\sigma_h^2 + \mathbf{I}\sigma_\epsilon^2$ and simplifying, s_I^2 can be expressed as

$$s_I^2 = s_p^2 + \frac{1}{\sigma_\epsilon^2}\left[\mathbf{a}'\mathbf{a} + \frac{\sigma_h^2(\mathbf{a}'\boldsymbol{\delta})^2}{(\sigma_\epsilon^2 + \sigma_h^2\boldsymbol{\delta}'\boldsymbol{\delta})}\right]. \tag{S6.2.6}$$

where s_p^2 is the squared Sharpe ratio of portfolio p.

Solution 6.3

Using (S6.2.6) and the cross-sectional distributional properties of the elements of $\mathbf{a}$ and $\boldsymbol{\delta}$, an approximation for the maximum squared Sharpe measure for each economy can be derived. For both economies, $\frac{1}{N}\mathbf{a}'\mathbf{a}$ converges to σ_a^2, and $\frac{1}{N}\boldsymbol{\delta}'\boldsymbol{\delta}$ converges to σ_a^2. For economy A, $\frac{1}{N^2}(\mathbf{a}'\boldsymbol{\delta})^2$ converges to $\sigma_{\mathbf{a}}^4$, and for economy B, $\frac{1}{N}(\mathbf{a}'\boldsymbol{\delta})^2$ converges to σ_a^4. Substituting these limits into (S6.2.6) gives approximations of the maximum squared Sharpe measures squared for each economy. Substitution into (S6.2.6) gives

$$s_A^2 = s_p^2 + \frac{N\sigma_a^2}{\sigma_\epsilon^2 + N\sigma_h^2\sigma_a^2} \tag{S6.3.1}$$

(S6.3.2) $$s_B^2 = s_p^2 + N\frac{\sigma_a^2}{\sigma_\epsilon^2}[1 - \frac{\sigma_h^2\sigma_a^2}{\sigma_\epsilon^2 + N\sigma_h^2\sigma_a^2}].$$

Thus we have the squared Sharpe ratios for economies A and B, respectively. The squared Sharpe ratios for large N follow from (S6.3.1) and (S6.3.2). For economy A we have

(S6.3.3) $$s_A^2 = s_p^2 + \frac{1}{\sigma_h^2},$$

and for economy B we have

(S6.3.4) $$s_B^2 = s_p^2 + N[\frac{\sigma_a^2}{\sigma_\epsilon^2}].$$

The maximum squared Sharpe measure is bounded as N increases for economy A and unbounded for economy B. Examples of economies A and B are discussed in section 6.6.3.

Problems in Chapter 7

Solution 7.1

Each period, the corporation repurchases shares worth λX while the total stock is worth $V \equiv X/(1-(1+R)^{-1}) = (1+R)X/R$. Therefore, the number of shares outstanding follows the "law of motion"

$$N_{t+1} = \left(1 - \lambda \frac{R}{1+R}\right) N_t. \tag{S7.1.1}$$

7.1.1 Price per share is $P_t \equiv V/N_t$, and dividend per share is $D_t \equiv (1-\lambda)X/N_{t+1}$. (Note that dividends are paid after repurchases, on the remaining shares only). Hence the growth rate of dividends per share, G, satisfies

$$1 + G = \frac{N_t}{N_{t+1}} = \frac{1+R}{1+R(1-\lambda)}. \tag{S7.1.2}$$

Dividends per share grow, even though total dividends do not, because the number of shares is shrinking over time.

7.1.2 The dividend-price ratio is

$$DP \equiv D_t/P_t = \frac{(1-\lambda)X}{V}\left(\frac{N_t}{N_{t+1}}\right). \tag{S7.1.3}$$

Manipulating this equation yields

$$DP = \frac{R-G}{1+R}, \tag{S7.1.4}$$

which is consistent with equation (7.1.9) after accounting for the fact that prices here are cum-dividend, whereas the discussion in the text applies to ex-dividend prices.

7.1.3 This follows immediately from the results of the previous subsection, because if price is the present value of discounted future dividends, including the dividend paid today, then

$$P_t = \sum_{t=0}^{\infty} \frac{D_{t+i}}{(1+R)^i} = \sum_{t=0}^{\infty} D_t \frac{(1+G)^i}{(1+R)^i} = D_t\left(\frac{1+R}{R-G}\right), \tag{S7.1.5}$$

which is what we showed in the previous subsection.

Intuitively, shares must have the same value to shareholders who sell shares to the repurchasing firm and to shareholders who do not. A shareholder who sells a fraction λ of his shares to the firm each period receives a constant fraction of total dividends and repurchase payments, that is, a constant fraction of the firm's cash flow. A shareholder who sells no shares to the firm receives a growing fraction of total dividends, because the total number of shares is shrinking over time. The value of the shares is the same in either case.

Solution 7.2

7.2.1 Denoting the discount rate $R = e^r$, we can write

$$\text{(S7.2.1)} \qquad F_t = \mathrm{E}_t[\sum_{\tau=t+1}^{\infty} \frac{D_\tau}{R^{\tau-t}}] = \sum_{n=1}^{\infty} \mathrm{E}_t[e^{d_t + \sum_{k=1}^{n}(\mu+\xi_{t+k}-r)}]$$
$$= D_t \sum_{n=1}^{\infty} e^{n(\mu+\sigma^2/2-r)}.$$

The condition $\mu + \sigma^2/2 < r$ is necessary for the sum to converge. It follows that the ratio of fundamental value to dividend is

$$\text{(S7.2.2)} \qquad \frac{F_t}{D_t} = \frac{e^{\mu+\sigma^2/2-r}}{1 - e^{\mu+\sigma^2/2-r}}.$$

7.2.2 Since

$$\text{(S7.2.3)} \qquad F_t + cD_t^\lambda = e^{-r}\mathrm{E}[F_{t+1} + cD_{t+1}^\lambda + D_{t+1}]$$

and

$$\text{(S7.2.4)} \qquad F_t = e^{-r}\mathrm{E}_t[F_{t+1} + D_{t+1}],$$

we have

$$\text{(S7.2.5)} \qquad D_t^\lambda = e^{-r}\mathrm{E}_t[D_{t+1}^\lambda].$$

Since

$$\text{(S7.2.6)} \qquad \mathrm{E}_t[D_{t+1}^\lambda] = e^{\mathrm{E}_t[\lambda d_{t+1}] + \mathrm{Var}_t[\lambda d_{t+1}]/2} = e^{\lambda\mu + \lambda d_t + \lambda^2\sigma^2/2},$$

we get a quadratic equation for the parameter λ,

$$\text{(S7.2.7)} \qquad \lambda^2\sigma^2/2 + \lambda\mu - r = 0.$$

For such a parameter λ, the price process $P_t = F_t + cD_t^\lambda$ indeed gives the same expected rate of return as the process $P_t = F_t$.

7.2.3 The Froot-Obstfeld bubble requires a very specific dividend process. However, the bubble is strongly correlated with the dividend, capturing the effect of dividend "overreaction". The bubble never bursts for a strictly positive dividend stream.

Solution 7.3

7.3.1 Using approximate formula (7.1.30), we have for $k > 1$

$$\mathrm{Cov}[r_t, r_{t+k}]$$
$$\text{(S7.3.1)} \qquad = \mathrm{E}\left[\left(x_{t-1} + \eta_{d,t} - \frac{\rho\xi_t}{1-\rho\phi}\right)\left(x_{t+k-1} + \eta_{d,t+k} - \frac{\rho\xi_{t+k}}{1-\rho\phi}\right)\right]$$
$$= \mathrm{E}\left[x_{t-1}x_{t+k-1} - \frac{\rho\xi_t}{1-\rho\phi}x_{t+k-1}\right].$$

Because $x_t = \sum_{n=0}^{\infty} \phi^n \xi_{t-n}$, we have

$$\text{(S7.3.2)} \qquad \mathrm{E}[x_{t-1}x_{t+k-1}] = \frac{\phi^k\sigma_\xi^2}{1-\phi^2}$$

and

$$\text{(S7.3.3)} \qquad \mathrm{E}[\xi_t x_{t+k-1}] = \phi^{k-1}\sigma_\xi^2.$$

Thus the return autocovariance is

$$\text{(S7.3.4)} \qquad \mathrm{Cov}[r_t, r_{t+k}] = \phi^{k-1}\left(\frac{\phi}{1-\phi^2} - \frac{\rho}{1-\rho\phi}\right)\sigma_\xi^2.$$

This is negative when $\phi < \rho$. The autocorrelation of stock returns is determined by the balance of two opposing effects. Expected stock returns are positively autocorrelated, and this creates positive autocorrelation in realized stock returns. However innovations in expected future stock returns are negatively correlated with current unexpected stock returns, and this creates negative autocorrelation in realized stock returns. The latter effect dominates when $\phi < \rho$.

7.3.2 Assume now that

$$\text{(S7.3.5)} \qquad \text{Cov}[\eta_{d,t}, \xi_t] = \sigma_{\eta,\xi} > 0.$$

We have

$$\text{(S7.3.6)} \qquad \text{Cov}[r_t, r_{t+k}] = \phi^{k-1}\left(\frac{\sigma_{\eta,\xi}}{\sigma_\xi^2} + \frac{\phi}{1-\phi^2} - \frac{\rho}{1-\rho\phi}\right)\sigma_\xi^2.$$

If $\sigma_{\eta,\xi}$ is large enough, the first term can dominate the others, giving positive return autocovariances.

Solution 7.4

7.4.1 Equation (7.1.19) implies that for $p_t = v_t$,

$$\text{(S7.4.1)} \qquad r_{t+1} \approx k + \rho p_{t+1} + (1-\rho)d_{t+1} - p_t = k + \rho\mu + \rho\epsilon_{t+1}.$$

We see that r_{t+1} is just a constant plus a white noise component — the log stock return r_{t+1} is therefore unforecastable.

7.4.2 Let us rewrite the formula for v_t and substitute in the dividend rule. We get

$$\text{(S7.4.2)} \qquad v_t - d_t = \frac{1-\lambda}{\rho}(v_{t-1} - d_{t-1}) + \mu - \frac{c}{\rho} - \frac{\eta_t}{\rho} + \epsilon_t,$$

so the log dividend-price ratio $d_t - v_t$ follows an AR(1) process with persistence coefficient $(1-\lambda)/\rho$.

7.4.3 The log dividend-price ratio is

$$\text{(S7.4.3)} \qquad d_t - p_t = d_t - (v_t - \gamma(d_t - v_t)) = (1+\gamma)(d_t - v_t).$$

Because $d_t - v_t$ is an AR(1) process and the log dividend-price ratio $d_t - p_t$ is a (positive) multiple of $d_t - v_t$, it is also an AR(1) process.

The approximate log stock return can be rewritten, using the formulas for d_t and p_t, as

$$\begin{aligned} r_{t+1} &\approx k + \rho(p_{t+1} - d_{t+1}) + (d_{t+1} - p_t) \\ &= k + \rho(1+\gamma)(v_{t+1} - d_{t+1}) + c + (1-\lambda+\gamma)(d_t - v_t) + \eta_{t+1} \\ \text{(S7.4.4)} \qquad &= \text{constant } + \frac{\lambda\gamma}{1+\gamma}(d_t - p_t) - \gamma\eta_{t+1} + \rho(1+\gamma)\epsilon_{t+1}. \end{aligned}$$

Setting $x_t \equiv (\lambda\gamma/(1+\gamma))(d_t - p_t)$, we get a model of the form (7.1.27) and (7.1.28), with x_t being the optimal forecasting variable for r_{t+1} up to a constant.

7.4.4 From the above we have

$$\text{(S7.4.5)} \qquad r_{t+1} - \text{E}_t[r_{t+1}] = -\gamma\eta_{t+1} + \rho(1+\gamma)\epsilon_{t+1},$$

and

$$\text{(S7.4.6)} \qquad x_{t+1} - \text{E}_t[x_{t+1}] = \lambda\gamma\left(\frac{\eta_{t+1}}{\rho} - \epsilon_{t+1}\right),$$

so that the covariance of the innovations is always negative.

Solution 7.5

Let us denote the expectation conditional on the full information set at time t as $\mathrm{E}_t[\cdot]$ and the expectation conditional on information J_t as $\mathrm{E}_{J_t}[\cdot]$. Thus, we have $\mathrm{E}[\mathrm{E}_t[\cdot]] = \mathrm{E}[\cdot]$, $\mathrm{E}_{J_t}[\mathrm{E}_t[\cdot]] = \mathrm{E}_{J_t}[\cdot]$, and so forth, by the law of iterated expectations. In particular, $\mathrm{E}[p_t] = \mathrm{E}[\mathrm{E}_t[p_t^*]] = \mathrm{E}[p_t^*]$. Note that the following "prices" are expectations listed in order of decreasing conditioning information: p_t^*, $p_t = \mathrm{E}_t[p_t^*]$, $\hat{p}_t = \mathrm{E}_{J_t}[p_t^*]$, and $\mathrm{E}[p_t^*]$.

7.5.1 Calculate

$$\begin{aligned} \mathrm{Var}[p_t] &= \mathrm{E}[((\mathrm{E}_t[p_t^*] - \mathrm{E}_{J_t}[p_t^*]) + (\mathrm{E}_{J_t}[p_t^*] - \mathrm{E}[p_t^*]))^2] \\ &= \mathrm{E}[\mathrm{E}_t[p_t^*] - \mathrm{E}_{J_t}[p_t^*])^2] + \mathrm{E}[(\mathrm{E}_{J_t}[p_t^*] - \mathrm{E}[p_t^*])^2] \\ &\geq \mathrm{E}[(\mathrm{E}_{J_t}[p_t^*] - \mathrm{E}[\mathrm{E}_{J_t}[p_t^*]])^2] \\ &= \mathrm{Var}[\hat{p}_t]. \end{aligned} \tag{S7.5.1}$$

where the cross term at the second step was eliminated using the fact that

$$\mathrm{E}[\mathrm{E}_{J_t}[(\mathrm{E}_t[p_t^*] - \mathrm{E}_{J_t}[p_t^*])(\mathrm{E}_{J_t}[p_t^*] - \mathrm{E}[p_t^*])]] = 0 \tag{S7.5.2}$$

as $\mathrm{E}_{J_t}[p_t^*] - \mathrm{E}_{J_t}[p_t^*]$ conditional on J_t is a constant and $\mathrm{E}_{J_t}[\mathrm{E}_t[p_t^*] - \mathrm{E}_{J_t}[p_t^*]] = 0$. Calculations in other parts of the problem are similar.

Intuitively, a price forecast based on less information is less volatile.

7.5.2 Calculate

$$\begin{aligned} \mathrm{Var}[p_t^* - \hat{p}_t] &= \mathrm{E}[((p_t^* - p_t) + (p_t - \hat{p}_t))^2] \\ &= \mathrm{E}[(p_t^* - p_t)^2] + \mathrm{E}[(p_t - \hat{p}_t)^2] \\ &= \mathrm{Var}[p_t^* - p_t]. \end{aligned} \tag{S7.5.3}$$

It follows that

$$\mathrm{Var}[p_t^* - \hat{p}_t] \geq \mathrm{Var}[p_t - \hat{p}_t] \tag{S7.5.4}$$

and

$$\mathrm{Var}[p_t^* - \hat{p}_t] \geq \mathrm{Var}[p_t - \hat{p}_t] \tag{S7.5.5}$$

as was to be shown.

Intuitively, a forecast based on inferior information has a larger error variance. Also, the error variance for a forecast of the actual realization of an uncertain price is larger than the error variance for a forecast of a superior-information forecast.

Stock prices, referred to in Problem 7.5.1, are usually considered nonstationary so that their conditional variances do not converge to a finite unconditional variance. On the other hand, forecast errors, referred to in Problem 7.5.2 may plausibly be assumed stationary. Therefore, the framework of Problem 7.5.2 seems more suitable for econometric analysis.

7.5.3 Note that $\hat{p}_{t+1}$ is defined to be $\mathrm{E}_{J_t}[p_{t+1}^*]$. Using the approximation

$$\mathrm{E}[r_{t+1}] = \mathrm{E}[\mathrm{E}_{J_t}[r_{t+1}]] = \mathrm{E}[\hat{r}_{t+1}] \tag{S7.5.6}$$

that follows from (7.1.19) we get

$$\begin{aligned} \mathrm{Var}[r_{t+1}] &= \mathrm{E}[((r_{t+1} - \hat{r}_{t+1}) + (\hat{r}_{t+1} - \mathrm{E}[\hat{r}_{t+1}]))^2] \\ &= \mathrm{E}[(r_{r+1} - \hat{r}_{t+1})^2] + \mathrm{E}[(\hat{r}_{t+1} - \mathrm{E}[\hat{r}_{t+1}])^2] \\ &\geq \mathrm{Var}[\hat{r}_{t+1}]. \end{aligned} \tag{S7.5.7}$$

Intuitively, the variance of a return forecast is less volatile than the return itself. Just as in Problem 7.5.2, this result is more useful than that in Problem 7.5.1 because the stochastic processes for returns do not seem to have the unit roots characteristic of price processes.

Problems in Chapter 8

Solution 8.1

8.1.1 Recall that $\beta_{\overline{M}} = \Omega^{-1}(\iota - \overline{M}\mathrm{E}[\iota + R_t])$, $M_t^* = \overline{M} + (R_t - \mathrm{E}[R_t])'\beta_{\overline{M}}$. Therefore

$$\text{(S8.1.1)} \quad \mathrm{E}[M_t^*(\overline{M})^2] = \overline{M}^2 + \beta_{\overline{M}}'\Omega\beta_{\overline{M}} = \overline{M}^2 + (\iota - \overline{M}\mathrm{E}[\iota + R_t])'\Omega^{-1}(\iota - \overline{M}\mathrm{E}[\iota + R_t]).$$

Note in particular that $\mathrm{E}[M_t^*(\overline{M})^2] \geq \overline{M}^2$.

First, we will show that in the market augmented by a risk-free asset with return $1 + R_{Ft} = 1/\overline{M}$, there exists a benchmark portfolio with return

$$\text{(S8.1.2)} \qquad 1 + R_{bt} = \frac{M_t^*(\overline{M})}{\mathrm{E}[M_t^*(\overline{M})^2]}.$$

Consider a portfolio with dollar weights $\beta_{\overline{M}}$ on the risky assets and $\overline{M}^2 - \mathrm{E}[\iota + R_t]'\beta_{\overline{M}}$ on the risk free asset. Such a portfolio has payoff $M_t^*(\overline{M})$ and value

$$\text{(S8.1.3)} \qquad \begin{aligned} & \iota'\beta_{\overline{M}} + \overline{M}^2 - \mathrm{E}[\iota + R_t]'\beta_{\overline{M}} \\ = \; & \overline{M}^2 + (\iota - \overline{M}\mathrm{E}[\iota + R_t])'\Omega^{-1}(\iota - \overline{M}\mathrm{E}[\iota + R_t]) \\ = \; & \mathrm{E}[M_t^*(\overline{M})^2]. \end{aligned}$$

Thus the portfolio return is exactly $1 + R_{bt}$, and the proof of existence is complete.

Next, consider any portfolio R_{pt} such that $\mathrm{E}[R_{pt}] = \mathrm{E}[R_{bt}]$. The properties of $M_t^*(\overline{M})$ and R_{bt} imply that

$$\text{(S8.1.4)} \qquad \begin{aligned} \mathrm{Var}[R_{pt}] &= \mathrm{E}[((R_{pt} - R_{bt}) + (R_{bt} - \mathrm{E}[R_{bt}]))^2] \\ &= \mathrm{E}[(R_{pt} - R_{bt})^2] + \mathrm{E}[(R_{bt} - \mathrm{E}[R_{bt}])^2] \\ &= \mathrm{Var}[R_{pt} - R_{bt}] + \mathrm{Var}[R_{bt}]. \end{aligned}$$

The only nontrivial step was to eliminate the cross term

$$\text{(S8.1.5)} \qquad \begin{aligned} & \mathrm{E}[(R_{pt} - R_{bt})(R_{bt} - \mathrm{E}[R_{bt}])] \\ = \; & \mathrm{E}[R_{pt} - R_{bt}]\frac{M_t^*(\overline{M})}{\mathrm{E}[M_t^*(\overline{M})^2]} - \mathrm{E}[R_{pt} - R_{bt}]\mathrm{E}[1 + R_{bt}] \\ = \; & 0. \end{aligned}$$

Thus $\mathrm{Var}[R_{pt}] \geq \mathrm{Var}[R_{bt}]$ whenever $\mathrm{E}[R_{pt}] = \mathrm{E}[R_{bt}]$, so the benchmark portfolio is on the mean-variance frontier.

8.1.2 First note that $\mathrm{E}[(M_t(\overline{M}) - M_t^*(\overline{M}))(1 + R_{pt})] = 0$ for any portfolio return R_{pt}. It follows that

$$\text{(S8.1.6)} \qquad \mathrm{Cov}[M_t(\overline{M}), R_{pt}] = \mathrm{Cov}[M_t^*(\overline{M}), R_{pt}] \leq \mathrm{Var}[M_t^*(\overline{M})]^{1/2}\,\mathrm{Var}[R_{pt}]^{1/2},$$

where equality is attained iff $M_t^*(\overline{M})$ is perfectly correlated with R_{pt}. In particular, we have $\mathrm{Cov}[M_t(\overline{M}), R_{bt}] = \mathrm{Var}[M_t^*(\overline{M})]^{1/2}\,\mathrm{Var}[R_{bt}]^{1/2}$ and therefore

$$\text{(S8.1.7)} \qquad \mathrm{Corr}[M_t(\overline{M}), R_{pt}] \leq \frac{\mathrm{Var}[M_t^*(\overline{M})]^{1/2}}{\mathrm{Var}[M_t(\overline{M})]^{1/2}} = \mathrm{Corr}[M_t(\overline{M}), R_{bt}],$$

so that R_{bt} is a maximum correlation portfolio among all R_{pt}'s with respect to any stochastic discount factor $M_t(\overline{M})$.

8.1.3 From $\mathrm{E}[M_t^*(\overline{M})(\iota + R_{it})] = 1$ we get

$$\text{(S8.1.8)} \qquad \mathrm{Cov}[M_t^*(\overline{M}), R_{it}] = 1 - \mathrm{E}[M_t^*(\overline{M})]\mathrm{E}[1 + R_{it}]$$

and similarly for R_{bt}. Therefore

$$\text{(S8.1.9)} \qquad \frac{\mathrm{Cov}[M_t^*(\overline{M}), R_{it}]}{\mathrm{Cov}[M_t^*(\overline{M}), R_{bt}]} = \frac{1 - \mathrm{E}[M_t^*(\overline{M})]\mathrm{E}[1 + R_{it}]}{1 - \mathrm{E}[M_t^*(\overline{M})]\mathrm{E}[1 + R_{bt}]}.$$

Note that $\mathrm{E}[M_t^*(\overline{M})] = \overline{M}$ and $M_t^*(\overline{M}) = c(1 + R_{bt})$, where $c = \mathrm{E}[M_t^*(\overline{M})^2]^{-1} > 0$ is a constant. Thus, the above expression simplifies to

$$\text{(S8.1.10)} \qquad \frac{\mathrm{Cov}[R_{bt}, R_{it}]}{\mathrm{Cov}[R_{bt}, R_{bt}]} = \frac{1/\overline{M} - \mathrm{E}[1 + R_{it}]}{1/\overline{M} - \mathrm{E}[1 + R_{bt}]}$$

which yields (8.1.17).

8.1.4 Indeed,

$$\text{(S8.1.11)} \qquad \mathrm{E}[1 + R_{bt}] = \mathrm{E}\left[\frac{M_t^*(\overline{M})}{\mathrm{E}[M_t^*(\overline{M})^2]}\right] = \frac{\overline{M}}{\mathrm{E}[M_t^*(\overline{M})^2]},$$

and

$$\begin{aligned} \sigma(R_{bt}) &= \left(\frac{\mathrm{E}[M_t^*(\overline{M})(1 + R_{bt})]}{\mathrm{E}[M_t^*(\overline{M})^2]} - \frac{\overline{M}^2}{(\mathrm{E}[M_t^*(\overline{M})^2])}\right)^{1/2} \\ \text{(S8.1.12)} \qquad &= \frac{\left(\mathrm{E}[M_t^*(\overline{M})^2] - \overline{M}^2\right)^{1/2}}{\mathrm{E}[M_t^*(\overline{M})^2]} \end{aligned}$$

so that

$$\text{(S8.1.13)} \qquad \frac{\sigma(R_{bt})}{\mathrm{E}[1 + R_{bt}]} = \left(\frac{\mathrm{E}[M_t^*(\overline{M})^2]}{\overline{M}^2} - 1\right)^{1/2}.$$

Similarly,

$$\begin{aligned} \text{(S8.1.14)} \qquad \frac{1/\overline{M} - \mathrm{E}[1 + R_{bt}]}{\sigma(R_{bt})} &= \frac{\frac{\mathrm{E}[M_t^*(\overline{M})^2]}{\overline{M}} - \overline{M}}{\left(\mathrm{E}[M_t^*(\overline{M})^2] - \overline{M}^2\right)^{1/2}} \\ &= \left(\frac{\mathrm{E}[M_t^*(\overline{M})^2]}{\overline{M}^2} - 1\right)^{1/2}. \end{aligned}$$

8.1.5 Note that

$$\text{(S8.1.15)} \qquad \mathrm{E}[M_t(\overline{M})] = \overline{M} = \mathrm{E}[M_t^*(\overline{M})]$$

and

$$\begin{aligned} \mathrm{E}[M_t(\overline{M})^2] &= \mathrm{E}[\big(M_t^*(\overline{M}) + \big(M_t(\overline{M}) - M_t^*(\overline{M})\big)\big)^2] \\ \text{(S8.1.16)} \qquad &= \mathrm{E}[M_t^*(\overline{M})^2] + \mathrm{E}[\big(M_t(\overline{M}) - M_t^*(\overline{M})\big)^2] \\ &\geq \mathrm{E}[M_t^*(\overline{M})^2]. \end{aligned}$$

Therefore $\sigma\left(M_t(\overline{M})\right) \geq \sigma\left(M_t^*(\overline{M})\right)$ and finally

$$\text{(S8.1.17)} \qquad \frac{\sigma(R_{bt})}{\mathrm{E}[1 + R_{bt}]} = \frac{\sigma\left(M_t^*(\overline{M})\right)}{\mathrm{E}[M_t^*(\overline{M})]} \leq \frac{\sigma\left(M_t(\overline{M})\right)}{\mathrm{E}[M_t(\overline{M})]}.$$

Solution 8.2

8.2.1 Assume a representative-agent utility function as in (8.2.2),

$$u(\{C\}) = \sum_{j=0}^{\infty} \delta^j \frac{C_{t+j}^{1-\gamma}}{1-\gamma} = \sum_{j=0}^{\infty} \delta^j v(C), \tag{S8.2.1}$$

and consider the maximization problem

$$\max_{\{C\},\{w\}} \mathrm{E}_t[u(\{C\})] \tag{S8.2.2}$$

subject to $W_{t+1} + C_{t+1} = w_t'(\iota + R_{t+1})W_t$ and $W_t \geq 0$.

Consider a single intertemporal sub-problem involving incremental investment of an amount x in a specific asset i from period t to $t+1$, at the cost of time-t consumption, the proceeds of the investment to be consumed at $t+1$:

$$\max_x \mathrm{E}_t[v(C_t - x) + \delta v(C_{t+1} + x(1 + R_{i,t+1}))]. \tag{S8.2.3}$$

At the optimum of the previous problem, $x = 0$ has to be optimal here, so that

$$\frac{\partial}{\partial x}\mathrm{E}_t[v(C_t - x) + \delta v(C_{t+1} + x(1 + R_{i,t+1}))]\bigg|_{x=0} = 0, \tag{S8.2.4}$$

which implies

$$\mathrm{E}_t[C_t^{-\gamma} + \delta(1 + R_{i,t+1})C_{t+1}^{-\gamma}] = 0, \tag{S8.2.5}$$

from which (8.2.3) follows. Note that (8.2.3) is not only necessary but also sufficient for the optimum once it holds for all i's and t's.

Assuming that asset returns and consumption are jointly log-normal, the quantity

$$(1 + R_{i,t+1})\delta(C_{t+1}/C_t)^{-\gamma} \tag{S8.2.6}$$

is also log-normal and therefore by taking logs of (8.2.3)

(S8.2.7)

$$\mathrm{E}_t[\log\left((1 + R_{i,t+1})\delta(C_{t+1}/C_t)^{-\gamma}\right)] + \frac{1}{2}\mathrm{Var}_t[\log\left((1 + R_{i,t+1})\delta(C_{t+1}/C_t)^{-\gamma}\right)] = 0,$$

so that

$$\begin{aligned}\mathrm{E}_t[r_{i,t+1}] + \log\delta - \gamma\mathrm{E}_t[\Delta c_{t+1}] + \\ \frac{1}{2}\left(\mathrm{Var}_t[r_{i,t+1}] + \gamma^2\,\mathrm{Var}_t[\Delta c_{t+1}] - 2\gamma\mathrm{Cov}_t[r_{i,t+1}\Delta c_{t+1}]\right) = 0\end{aligned} \tag{S8.2.8}$$

which gives (8.2.5).

Assuming that conditional variances and covariances

$$\begin{aligned} V_{ii} &= \mathrm{Var}_t[r_{i,t+1}], \\ V_{cc} &= \mathrm{Var}_t[\Delta c_{t+1}], \\ V_{ic} &= \mathrm{Cov}_t[r_{i,t+1}, \Delta c_{t+1}] \end{aligned} \tag{S8.2.9}$$

are all constants, we can write (8.2.5) as

$$\mathrm{E}_t[r_{i,t+1}] = \gamma\mathrm{E}_t[\Delta c_{t+1}] + \left(-\log\delta - \frac{1}{2}\left(V_{ii} + \gamma^2 V_{cc} - 2\gamma V_{ic}\right)\right), \tag{S8.2.10}$$

which is a linear function of $\mathrm{E}_t[\Delta c_{t+1}]$ with slope coefficient γ—the coefficient of risk aversion for the power utility function. This solves part (i). Subtracting (8.2.6), the riskfree asset equation, we get as in (8.2.7)

$$\mathrm{E}_t[r_{i,t+1} - r_{f,t+1}] + \frac{1}{2}V_{ii} = \gamma V_{ic}, \tag{S8.2.11}$$

so that the "premium" of the asset is proportional to the conditional covariance of the log asset return with consumption growth, with coefficient of proportionality γ. This solves part (ii).

8.2.2 Part (i). Let aggregate equity e pay a log dividend equal to log aggregate consumption, so that $\Delta d_{e,t} = \Delta c_t$. From the previous part we know that $\mathrm{E}_t[r_{e,t+1+j}] = \gamma \mathrm{E}_t[\Delta c_{t+1+j}]$, up to a constant. Then (7.1.25) implies that

(S8.2.12)

$$r_{e,t+1+j} - \mathrm{E}_t r_{e,t+1+j} = \Delta c_{t+1} - \mathrm{E}_t \Delta c_{t+1} + (1-\gamma) \sum_{j=1}^{\infty} \rho^j \left[\mathrm{E}_{t+1} \Delta c_{t+1+j} - \mathrm{E}_t \Delta c_{t+1+j} \right].$$

By simple algebraic manipulation of the process for Δc_{t+1}, we obtain the following expression for Δc_{t+1+j}:

$$\Delta c_{t+1+j} = \mu \left(\sum_{i=0}^{j} \phi^i \right) + \phi^{j+1} \Delta c_t + \sum_{i=0}^{j} \phi^{j-i} u_{t+i+1}, \tag{S8.2.13}$$

so by the Law of Iterated Expectations we have that

$$\mathrm{E}_t \Delta c_{t+1+j} = \mu \left(\sum_{i=0}^{j} \phi^i \right) + \phi^{j+1} \Delta c_t \tag{S8.2.14}$$

and

$$\mathrm{E}_{t+1} \Delta c_{t+1+j} = \mu \left(\sum_{i=0}^{j} \phi^i \right) + \phi^{j+1} \Delta c_t + \phi^j u_{t+1}, \tag{S8.2.15}$$

which in turn implies

$$\mathrm{E}_{t+1} \Delta c_{t+1+j} - \mathrm{E}_t \Delta c_{t+1+j} = \phi^j u_{t+1}. \tag{S8.2.16}$$

Substituting in this expression, and noting that $\Delta c_{t+1} - \mathrm{E}_t \Delta c_{t+1} = u_{t+1}$, we obtain

$$\begin{aligned} r_{e,t+1+j} - \mathrm{E}_t r_{e,t+1+j} &= u_{t+1} + (1-\gamma) \sum_{j=1}^{\infty} \rho^j \phi^j u_{t+1} \\ &= u_{t+1} + (1-\gamma) \frac{\rho\phi}{1-\rho\phi} u_{t+1} \\ &= \left(\frac{1-\gamma\rho\phi}{1-\rho\phi} \right) u_{t+1}. \end{aligned} \tag{S8.2.17}$$

Part (ii). For a real consol paying a fixed real dividend we have that $\Delta d_{i,t+1+j} = 0$, so the unexpected return is influenced only by changes in expected future interest rates. Similar reasoning as in part (i) gives the unexpected real consol bond return as

$$r_{b,t+1+j} - \mathrm{E}_t r_{b,t+1+j} = \left(\frac{-\gamma\rho\phi}{1-\rho\phi} \right) u_{t+1}. \tag{S8.2.18}$$

8.2.3 Part (i). From equation (8.2.7), the equity premium is given by γV_{ce}, where

$$\begin{aligned} V_{ce} &= \mathrm{Cov}_t \left(\Delta c_{t+1} - \mathrm{E}_t \Delta c_{t+1}, r_{e,t+1} - \mathrm{E}_t r_{e,t+1} \right) \\ &= \mathrm{Cov} \left(u_{t+1}, \frac{1-\gamma\rho\phi}{1-\rho\phi} u_{t+1} \right) \\ &= \frac{1-\gamma\rho\phi}{1-\rho\phi} \sigma_u^2, \end{aligned} \tag{S8.2.19}$$

and we may write $\mathrm{Cov}(\cdot,\cdot)$ instead of $\mathrm{Cov}_t(\cdot,\cdot)$ because the process for Δc_{t+1} is homoskedastic.

Similarly, the consol bond premium is γV_{cb}, where

$$V_{cb} = \frac{-\gamma\rho\phi}{1-\rho\phi} \sigma_u^2. \tag{S8.2.20}$$

Part (ii). The bond premium has the opposite sign to ϕ because a positive ϕ implies that a positive endowment shock increases future consumption more than current consumption, so real interest rates rise and bond prices fall when consumption rises. Real bonds thus provide a hedge against endowment risk and they have a negative premium. The bond premium is proportional to the square of γ because a larger γ both increases the variability of real interest rates and bond returns, and increases the premium required by investors for bearing a unit of risk.

Part (iii). The premium of equity over the consol is

$$\gamma\left(V_{ce}-V_{be}\right)=\frac{\gamma\sigma_u^2}{(1-\rho\phi)}, \tag{S8.2.21}$$

so the equity premium is just the bond premium plus a premium related to dividend uncertainty, which is always positive and proportional to γ.

Part (iv). The lesson for the equity premium literature is that models with high degrees of risk aversion tend to imply a high bond premium as well as a high equity premium. This is a counterfactual implication.

Solution 8.3

8.3.1 The second-period endowment is m with probability $\frac{1}{2}$ and $(1-a)m$ with probability $\frac{1}{2}$. This can be written as

$$w=\begin{cases} m, & \frac{1}{2} \\ (1-a)m, & \frac{1}{2} \end{cases} \tag{S8.3.1}$$

where w is both the individual and aggregate second period endowment.

Consider buying ϵ of the asset. The asset price is paid in the second period, so the expected utility cost is

$$\begin{aligned} \frac{1}{2}U'\left(m\right)p\epsilon+\frac{1}{2}U'\left(\left(1-a\right)m\right)p\epsilon &= \frac{1}{2}p\epsilon\left[\frac{1}{m}+\frac{1}{\left(1-a\right)m}\right] \\ &= \frac{1}{2}p\epsilon\left[\frac{2-a}{\left(1-a\right)m}\right]. \end{aligned} \tag{S8.3.2}$$

The expected utility gain is

$$\frac{1}{2}U'\left(m\right)m\epsilon+\frac{1}{2}U'\left(\left(1-a\right)m\right)\left(1-a\right)m\epsilon=\frac{1}{2}\epsilon\left[1+1\right] \tag{S8.3.3}$$

$$=\epsilon. \tag{S8.3.4}$$

In equilibrium, the expected utility cost must equal the expected utility benefit, so

$$\frac{1}{2}p\epsilon\left[\frac{2-a}{\left(1-a\right)m}\right]=\epsilon, \tag{S8.3.5}$$

which implies

$$p=\left[\frac{2\left(1-a\right)}{2-a}\right]m. \tag{S8.3.6}$$

The expected gross return on the claim is the ratio between its expected payoff and its price:

$$1+R^{(a)}=\frac{\frac{1}{2}m+\frac{1}{2}\left(1-a\right)m}{p}=1+\frac{a^2}{4\left(1-a\right)}, \tag{S8.3.7}$$

which rises with a (a measure of aggregate risk) as we would expect.

8.3.2 Now we have that the individual second period endowment is:

$$w^I=\begin{cases} m, & \frac{1}{2} \\ m, & \left(1-b\right)\frac{1}{2} \\ \left(1-\frac{a}{b}\right)m, & b\frac{1}{2} \end{cases} \tag{S8.3.8}$$

while the aggregate endowment is still as before:

$$w^A = \begin{cases} m, & \frac{1}{2} \\ (1-a)m, & \frac{1}{2} \end{cases} \tag{S8.3.9}$$

Note that we must have $b > a$ so the individual endowment is always non-negative and log utility is defined. If $b = 1$ then we are back to the previous case.

Since all agents have the same utility function and face the same probability of being in each group, they all have the same expected endowment and are identical ex-ante. However, ex-post their endowments will differ,so there will be ex-post heterogeneity.

As before, consider buying ϵ of the asset. The expected utility cost is

$$\begin{aligned} &\frac{1}{2}U'(m)\,p\epsilon + \frac{1}{2}(1-b)\,U'(m)\,p\epsilon + \frac{1}{2}bU'\left(\left(1-\frac{a}{b}\right)m\right)p\epsilon \\ &= \frac{1}{2}\frac{p\epsilon}{m}\left[1+(1-b)+\frac{b}{1-\frac{a}{b}}\right] = \frac{1}{2}\frac{p\epsilon}{m}\left[2+\frac{a}{1-\frac{a}{b}}\right]. \end{aligned} \tag{S8.3.10}$$

The expected utility gain is

$$\begin{aligned} &\frac{1}{2}U'(m)\,m\epsilon + \frac{1}{2}(1-b)\,U'(m)\,(1-a)\,m\epsilon + \frac{1}{2}bU'\left(\left(1-\frac{a}{b}\right)m\right)(1-a)\,m\epsilon \\ &= \frac{1}{2}\epsilon\left[1+(1-b)(1-a)+\frac{b(1-a)}{1-\frac{a}{b}}\right]. \end{aligned} \tag{S8.3.11}$$

In equilibrium, expected cost equals expected gain so

$$\frac{1}{2}\frac{p\epsilon}{m}\left[2+\frac{a}{1-\frac{a}{b}}\right] = \frac{1}{2}\epsilon\left[1+(1-b)(1-a)+\frac{b(1-a)}{1-\frac{a}{b}}\right]. \tag{S8.3.12}$$

Thus

$$p = \left[\frac{2(b-a)+a^2(1-b)}{2(b-a)+ba}\right]m, \tag{S8.3.13}$$

which gives the previous result when $b = 1$, and $p = (1-a)\,m$ when $b = a$.

The expected gross return on the claim is :

$$1+R^{(b)} = \frac{\frac{1}{2}m+\frac{1}{2}(1-a)\,m}{p} \tag{S8.3.14}$$

$$= \frac{(2-a)\left[2(b-a)+ba\right]}{4(b-a)+2a^2(1-b)}, \tag{S8.3.15}$$

so when $b = 1$ we obtain the same result as before,

$$1+R^{(b)} = \frac{(2-a)^2}{4(1-a)} = 1+R^{(a)}, \tag{S8.3.16}$$

and when $b = a$,

$$1+R^{(b)} = \frac{(2-a)}{2(1-a)} = \frac{2}{(2-a)}(1+R^{(a)}). \tag{S8.3.17}$$

Since $0 < a \leq 1$, $R^{(b)} \geq R^{(a)}$. Therefore, heterogeneity in the form of individual uninsurable risk increases the expected return on the asset.

8.3.3 The literature on representative agent models tends to find that average stock returns are higher than can be explained with plausible degrees of risk aversion. Uninsurable individual risk might be one explanation.

This problem is based on N. Gregory Mankiw's "The Equity Premium and the Concentration of Aggregate Shocks", Journal of Financial Economics, September 1986. Mankiw shows that one gets similar results for any utility function with $U''' > 0$. Quadratic utility has $U''' = 0$ ("certainty equivalence") and uninsurable individual risk has no effect.

The result also depends on the fact that there is more dispersion of individual endowments in bad times than in good times.

Problems in Chapter 9

Solution 9.1

Without loss of generality, let us consider the random variable $p_n(T)$ (the derivation for $p_n(t)$ is analogous). Denote the moment-generating functions of the increments ϵ_k in (9.1.1) and $p_n(T)$ in (9.1.2) as $M_\epsilon(\tau)$ and $M_p(\tau)$ respectively, where

$$M_\epsilon(\tau) = \mathrm{E}[e^{\tau\epsilon_k}] \tag{S9.1.1}$$

$$= pe^{\tau\Delta} + qe^{-\tau\Delta} \tag{S9.1.2}$$

$$M_p(\tau) = \mathrm{E}[e^{\tau p_n(T)}] = \mathrm{E}[e^{\tau\sum_{k=1}^n \epsilon_k}] \tag{S9.1.3}$$

$$= \mathrm{E}[\prod_{k=1}^n e^{\tau\epsilon_k}] = \prod_{k=1}^n \mathrm{E}[e^{\tau\epsilon_k}] \tag{S9.1.4}$$

$$= \left[\pi e^{\tau\Delta} + (1-\pi)e^{-\tau\Delta} \right]^n \tag{S9.1.5}$$

Recall from (9.1.7) that $\pi = \frac{1}{2}(1+\frac{\mu\sqrt{h}}{\sigma})$ and $\Delta = \sigma\sqrt{h}$. This implies

$$M_p(\tau) = \left[\frac{1}{2}(1+\frac{\mu\sqrt{h}}{\sigma})e^{\tau\sigma\sqrt{h}} + \frac{1}{2}(1-\frac{\mu\sqrt{h}}{\sigma})e^{-\tau\sigma\sqrt{h}} \right]^n \tag{S9.1.6}$$

$$= \left[\cosh(\tau\sigma\sqrt{h}) + \frac{\mu\sqrt{h}}{\sigma}\sinh(\tau\sigma\sqrt{h}) \right]^n \tag{S9.1.7}$$

where $\cosh(x)$ and $\sinh(x)$ are the hyperbolic sine and cosine functions

$$\cosh(x) = \frac{e^x + e^{-x}}{2} = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots$$

$$\sinh(x) = \frac{e^x - e^{-x}}{2} = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots$$

Simplifying and letting $n \to \infty$ yields

$$M_p(\tau) = \left[1 + (\frac{\tau^2\sigma^2}{2} + \mu\tau)\frac{T}{n} + o(\frac{1}{n}) \right]^n \tag{S9.1.8}$$

$$\to e^{(\mu\tau + \frac{\tau^2\sigma^2}{2})T} \tag{S9.1.9}$$

which is the moment-generating function for a normal random variable with mean μT and variance $\sigma^2 T$.

Solution 9.2

Denote by $\boldsymbol{\theta} \equiv [\ \mu \ \ \sigma^2\ ]'$ and observe that

$$\mathcal{I}(\boldsymbol{\theta}) = \lim_{n\to\infty} -\mathrm{E}\left[\frac{1}{n}\frac{\partial^2\mathcal{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \right] = \left[\begin{array}{cc} \frac{T}{n\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{array} \right] . \tag{S9.2.1}$$

Therefore, the inverse is simply

(S9.2.2) $$\mathcal{I}^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{n\sigma^2}{T} & 0 \\ 0 & 2\sigma^4 \end{bmatrix} .$$

Solution 9.3

Consider n observations in the interval $[0, T]$ equally spaced at intervals $h \equiv T/n$, and let $p(0) = 0$ to simplify the algebra. Let $p_k \equiv p(kh)$. Using (9.3.48) we find that

(S9.3.1) $$p_k - e^{-\gamma h} p_{k-1} = \mu h(k - e^{-\gamma h}(k-1)) + \sigma \int_{kh-h}^{kh} e^{-\gamma(kh-s)} dB(s) .$$

Now let $\rho = e^{-\gamma h}$. Then we can rewrite (S9.3.1) as

(S9.3.2) $$p_k - \rho p_{k-1} - \mu h(k - \rho(k-1)) = \sigma_\epsilon \epsilon_k ,$$

where $\epsilon_k \sim \mathcal{N}(0, 1)$ and

(S9.3.3) $$\sigma_\epsilon^2 = \text{Var}\left[\sigma \int_{kh-h}^{kh} e^{-\gamma(kh-s)} dB(s)\right] = \frac{\sigma^2}{\gamma}(1 - e^{-2\gamma h}) .$$

We now derive the maximum likelihood estimators $\hat{\mu}$, $\hat{\rho}$ and $\hat{\sigma}_\epsilon^2$ from which we can obtain $\hat{\gamma}$ and $\hat{\sigma}^2$ as:

(S9.3.4) $$\hat{\gamma} = -\frac{1}{h}\log(\hat{\rho}), \quad \hat{\sigma}^2 = \frac{2\hat{\gamma}\hat{\sigma}_\epsilon^2}{(1 - e^{-2\hat{\gamma} h})}$$

by the Principle of Invariance (see Zehna [1966]). The log-likelihood function is given by

(S9.3.5) $$\mathcal{L}(\mu, \rho, \sigma_\epsilon^2) = -\frac{n}{2}\log(2\pi\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2}\sum_{k=1}^{n}[p_k - \rho p_{k-1} - \mu h(k - \rho(k-1))]^2$$

and the necessary first-order conditions for the maximum of the log-likelihood function are

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \mu} &= \frac{1}{\sigma_\epsilon^2}\sum_{k=1}^{n}[p_k - \rho p_{k-1} - \mu h(k - \rho(k-1))]h(k - \rho(k-1)) = 0, \\
\frac{\partial \mathcal{L}}{\partial \rho} &= \frac{1}{\sigma_\epsilon^2}\sum_{k=1}^{n}[p_k - \rho p_{k-1} - \mu h(k - \rho(k-1))](p_{k-1} - \mu h(k-1)) = 0, \\
\frac{\partial \mathcal{L}}{\partial \sigma_\epsilon^2} &= -\frac{n}{2}\frac{1}{\sigma_\epsilon^2} + \frac{1}{2\sigma_\epsilon^4}\sum_{k=1}^{n}[p_k - \rho p_{k-1} - \mu h(k - \rho(k-1))]^2 = 0.
\end{aligned}$$

These conditions can be written as a system of equations in $(\hat{\mu}, \hat{\rho}, \hat{\sigma}_\epsilon^2)$:

(S9.3.6) $$\hat{\mu} = \frac{\sum_{k=1}^{n}(p_k - \hat{\rho} p_{k-1})(k - \hat{\rho}(k-1))}{\sum_{k=1}^{n}[k - \hat{\rho}(k-1)]^2},$$

(S9.3.7) $$\hat{\rho} = \frac{\sum_{k=1}^{n}(p_k - \hat{\mu} hk)(p_{k-1} - \hat{\mu} h(k-1))}{\sum_{k=1}^{n}[p_{k-1} - \hat{\mu} h(k-1)]^2},$$

(S9.3.8) $$\hat{\sigma}_\epsilon^2 = \frac{1}{n}\sum_{k=1}^{n}[p_k - \hat{\rho} p_{k-1} - \hat{\mu} h(k - \hat{\rho}(k-1))]^2.$$

From here we will assume that the trend μ is known exactly. Then we can calculate

$$\begin{aligned}\frac{\partial^2\mathcal{L}}{\partial\rho^2} &= -\frac{1}{\sigma_\epsilon^2}\sum_{k=1}^n[p_{k-1}-\mu h(k-1)]^2,\\ \frac{\partial^2\mathcal{L}}{\partial\rho\partial\sigma_\epsilon^2} &= -\frac{1}{\sigma_\epsilon^2}\frac{\partial\mathcal{L}}{\partial\rho},\\ \frac{\partial^2\mathcal{L}}{\partial(\sigma_\epsilon^2)^2} &= \frac{n}{2}\frac{1}{\sigma_\epsilon^4}-\frac{1}{\sigma_\epsilon^6}\sum_{k=1}^n[p_k-\rho p_{k-1}-\mu h(k-\rho(k-1))]^2.\end{aligned}$$

But observe that

$$\begin{aligned}\lim_{n\to\infty}-\mathrm{E}\left[\frac{1}{n}\left.\frac{\partial^2\mathcal{L}}{\partial\rho^2}\right|_{(\hat\rho,\hat\sigma_\epsilon^2)}\right] &= \frac{\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n(p_{k-1}-\mu h(k-1))^2}{\lim_{n\to\infty}\hat\sigma_\epsilon^2}\\ &= \frac{1}{1-e^{-2\gamma h}},\\ \lim_{n\to\infty}-\mathrm{E}\left[\frac{1}{n}\left.\frac{\partial^2\mathcal{L}}{\partial\rho\partial\sigma_\epsilon^2}\right|_{(\hat\rho,\hat\sigma_\epsilon^2)}\right] &= 0,\\ \lim_{n\to\infty}-\mathrm{E}\left[\frac{1}{n}\left.\frac{\partial^2\mathcal{L}}{\partial(\sigma_\epsilon^2)^2}\right|_{(\hat\rho,\hat\sigma_\epsilon^2)}\right] &= \lim_{n\to\infty}-\mathrm{E}\left[\frac{1}{n}(-\frac{n}{2}\frac{1}{\hat\sigma_\epsilon^4})\right]=\frac{1}{2\sigma_\epsilon^4}\ .\end{aligned}$$

Using (9.3.7), we conclude that

$$\sqrt{n}(\hat\rho-\rho) \overset{a}{\sim} \mathcal{N}(0,1-e^{-2\gamma h}), \tag{S9.3.9}$$

$$\sqrt{n}(\hat\sigma_\epsilon^2-\sigma_\epsilon^2) \overset{a}{\sim} \mathcal{N}(0,2\sigma_\epsilon^4). \tag{S9.3.10}$$

The asymptotic distribution of γ and σ^2 can now be obtained using (S9.3.4) and the delta method described in the Appendix A.4:

$$\sqrt{n}(\hat\gamma-\gamma) \overset{a}{\sim} \mathcal{N}\left(0,\frac{1}{h^2}(e^{2\gamma h}-1)\right), \tag{S9.3.11}$$

$$\sqrt{n}(\hat\sigma^2-\sigma^2) \overset{a}{\sim} \mathcal{N}\left(0,2\sigma^4\left(1+\frac{e^{4\gamma h}}{2\gamma^2h^2}[1-e^{-2\gamma h}(1+2\gamma h)]^2\right)\right). \tag{S9.3.12}$$

To derive the continuous-record asymptotics of $\hat\gamma$ and $\hat\sigma_\epsilon^2$, we let $n\to\infty$ while T is held fixed, hence $h=T/n\to 0$. Since

$$\hat\gamma=-\frac{1}{h}\log(1-(1-\hat\rho))=\frac{1}{h}(1-\hat\rho)+o(h)=\frac{(1-\hat\rho)}{\frac{1}{n}}+o(\frac{1}{n}), \tag{S9.3.13}$$

we conclude that

$$\hat\gamma \overset{a}{\approx} -\frac{\sum_{k=1}^n(p_k-p_{k-1}-\mu\frac{1}{n})(p_{k-1}-\mu\frac{k-1}{n})}{\frac{1}{n}\sum_{k=1}^n[p_{k-1}-\mu\frac{k-1}{n}]^2}\ . \tag{S9.3.14}$$

The denominator converges to

$$\int_0^T(p(s)-\mu s)^2\,ds, \tag{S9.3.15}$$

while the numerator converges to

$$\int_0^T(p(s)-\mu s)\,dp(s)-\mu\int_0^T(p(s)-\mu s)\,ds. \tag{S9.3.16}$$

We conclude that

$$\hat\gamma \overset{a}{\approx} -\frac{\int_0^T(p(s)-\mu s)\,dp(s)-\mu\int_0^T(p(s)-\mu s)\,ds}{\int_0^T(p(s)-\mu s)^2\,ds}. \tag{S9.3.17}$$

which simplifies to

$$\hat{\gamma} \overset{a}{\approx} -\frac{\int_0^T q(s)\,dq(s)}{\int_0^T q(s)^2\,ds}. \tag{S9.3.18}$$

where $q(t) \equiv p(t) - \mu t$. Finally, it can be shown that

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n}\sum_{k=1}^{n}[q_k - \hat{\rho} q_{k-1}]^2 \overset{a}{\approx} \sigma^2. \tag{S9.3.19}$$

Solution 9.4

9.4.1 The maximum likelihood estimates (9.3.27), (9.3.28) are evaluated using daily returns from January 2, 1991 to December 29, 1995 and assuming $h = 1/253$, i.e., 253 trading periods in a year. The riskfree interest rate is set to $r = 5\%$. The estimates are:

$$\hat{\sigma}^2 = 0.074 \quad , \quad \hat{\mu} \equiv r - \frac{\hat{\sigma}^2}{2} = 0.0131 \ .$$

9.4.2 Two variants of the Monte Carlo method are used:

1. The *crude* method of Section 9.4.1.
2. The *antithetic variates* method of Section 9.4.4.

The initial stock price (the closing price on December 29, 1995) is \$91.375. 100,000 replications are used in both cases ($m = 100,000$). The number of discrete intervals is $n = 253$.

The *crude* Monte Carlo method produces an estimate according to (9.4.6) of:

$$\hat{H}(0) = \$17.70 \ .$$

The standard deviation of the estimate $\hat{H}(0)$ is estimated according to (9.4.10) as:

$$\hat{\sigma}_y(253) = \$19.60 \ .$$

Therefore, according to (9.4.8), a 95% confidence interval is

$$\$17.58 \le H(0) \le \$17.82 \ .$$

The minimum number of replications necessary to yield a price estimate within \$0.05 of the true price is estimated according to (9.4.9):

$$m \ge 5.905 \times 10^5 \ .$$

The *antithetic variates* method produces an estimate according to (9.4.13):

$$\hat{H}(0) = \$17.63 \ .$$

Standard deviation of the estimate $\hat{H}(0)$ is estimated according to (9.4.15):

$$\hat{\sigma}_y(253) = \$8.56 \ .$$

As a result, according to (9.4.8), a 95% confidence interval is

$$\$17.58 \le H(0) \le \$17.69 \ .$$

The minimum number of replications necessary to yield a price estimate within \$0.05 of the true price is estimated according to (9.4.9) as:

$$m \ge 1.126 \times 10^5 \ .$$

9.4.3 The closed-form solution for the option price is given by the Goldman-Sosin-Gatto formula (9.4.11) and is evaluated using the estimate of σ^2 obtained in Problem 9.4.1:

$$H(0) = \$18.91 \ .$$

The difference between the theoretical price $H(0)$ and our estimate $\hat{H}(0)$ arises from the difference between the maximum of discretely-sampled and continuously-sampled prices. Specifically, the theoretical price $H(0)$ of the option is evaluated under the assumption that the option allows one to sell the stock at the maximum price observed over the course of the entire year. The estimate $\hat{H}(0)$ was obtained under the assumption that only daily closing prices are used to evaluate the maximum. Obviously, the first definition always leads to a higher option price than the second.

In the context of this particular problem the second definition of the option (the one used in Monte Carlo simulations) is more relevant, since it is based on the definition of the actual option. The Goldman-Sosin-Gatto formula is a continuous-time approximation to this option. Therefore, the Monte Carlo estimator of the option price should be used to decide whether to accept or reject CLM's proposal.

Problems in Chapter 10

Solution 10.1

10.1.1 Prices of the zero-coupon bonds are $P_A = e^{-8\times 0.091} \approx 0.4829$ and $P_B = e^{-9\times 0.080} \approx 0.4868$ per dollar of their face values. Since nominal interest rates cannot be negative, the finding that $P_A < P_B$ implies an arbitrage opportunity and is inconsistent with any expectations theory.
10.1.2 Prices of zero-coupon bonds are now $P'_A = e^{-7\times 0.091} \approx 0.5289$ and $P'_B = e^{-8\times 0.08} \approx 0.5273$ per dollar of their face values. As $P'_A \geq P'_B$ in this case, the prices do not imply an arbitrage opportunity and may be consistent with the pure expectations hypothesis.
10.1.3 Let us assume the coupon payments are annual and are made at the end of the year. Consider first the case analogous to Problem 10.1.1. Prices do not now imply an arbitrage opportunity. As an example, assume that *all* one- to eight-year zero-coupon bonds have price P_8 per one dollar of their face value, and that the nine-year zero-coupon bond has price P_9. Under these assumptions we can express the prices as

$$P_8 = \frac{P_A}{1 + 8\times 0.08} \approx 0.2944, \quad \text{(S10.1.1)}$$

$$P_9 = \frac{P_B - 8\times 0.08 P_8}{1 + 0.08} \approx 0.2752. \quad \text{(S10.1.2)}$$

We see that, under this non-stochastic term structure given by P_8 and P_9, all interest rates are nonnegative and $P_8 \geq P_9$, so that no arbitrage opportunity exists.

Now, consider the case analogous to Problem 10.1.2. Assume that *all* one- to seven-year zero-coupon bonds have price P'_7 per one dollar of their face value and that eight-year zero-coupon bond has price P'_8. Under these assumptions we can express the prices as

$$P'_7 = \frac{P'_A}{1 + 7\times 0.08} \approx 0.3390, \quad \text{(S10.1.3)}$$

$$P'_8 = \frac{P'_B - 7\times 0.08 P'_7}{1 + 0.08} \approx 0.3125. \quad \text{(S10.1.4)}$$

$P'_7 \geq P'_8$, so again there is no arbitrage opportunity.

Note however that the assumptions required to rationalize these bond prices are rather extreme, since they require zero nominal interest rates between one and eight years. The loglinear approximate model for coupon bonds presented in (10.1.20) gives a different answer. This model effectively imposes "smoothness" on the term structure. Equation (10.1.20) allows us to compute the implicit n-period-ahead 1-period log forward rate given the coupon-bond duration D_{cnt} in (10.1.10), which in turn requires the coupon-bond price P_{cnt} in (10.1.9).

For the data in Problem 10.1.1 we have

$$\begin{aligned} P_{c8t} &= .9171;\ D_{c9t} = 6.1186 \text{ years} \\ P_{c9t} &= .9797;\ D_{c9t} = 6.7212 \text{ years}, \end{aligned} \quad \text{(S10.1.5)}$$

so (10.1.20) gives

$$f_{8t} \approx -3.1684\% < 0$$

Notice that the bonds are not selling at par, so it is not correct to use the simpler formula for D_{cnt} that obtains in this case. This result again implies under the log pure expectations hypothesis a negative one-period log yield 8 periods ahead.

Similarly, for the data in Problem 10.1.2 we have

$$\text{(S10.1.6)} \qquad \begin{aligned} P_{c7t} &= .9245;\ D_{c7t} = 5.5615 \text{ years} \\ P_{c8t} &= .9813;\ D_{c8t} = 6.1876 \text{ years,} \end{aligned}$$

so (10.1.20) gives

$$f_{7t} \approx -1.7711\% < 0$$

which implies under the log pure expectations hypothesis a negative one-period log yield 7 periods ahead. Using this approach, we find that coupon bonds violate the log PEH more than zero-coupon bonds. The reason is that the duration of coupon bonds does not increase linearly with their maturity, but increases at a decreasing rate. That is, $D_{c,n+1,t} - D_{c,n,t} < 1$. This in turn makes it easier to get negative forward rates for given yields.

Solution 10.2

10.2.1 Assume the postulated process and simplify notation, introducing $a_t \equiv y_{t1} - y_{1,t-1}$ and $b_t \equiv y_{2t} - y_{1t}$. The equations of the model can then be written as

$$\text{(S10.2.1)} \qquad \begin{array}{ll} \text{I} & a_t = \lambda b_t + \epsilon_t, \\ \text{II} & b_t = \dfrac{1}{2}\mathrm{E}_t[a_{t+1}] + x_t, \\ \text{III} & x_t = \phi x_{t-1} + \eta_t, \\ \text{IV} & a_t = \gamma x_t + \epsilon_t. \end{array}$$

¿From the first and fourth equations we get $b_t = \gamma\lambda^{-1}x_t$; from the third and fourth equations we get $\mathrm{E}_t[a_{t+1}] = \gamma\phi x_t$; the second equation then gives an expression for the coefficient γ in terms of the other parameters of the model,

$$\text{(S10.2.2)} \qquad \gamma = \left(\frac{2\lambda}{2 - \lambda\phi}\right).$$

It is straightforward to verify that with this value for γ, the y_{1t} process satisfies all the equations of the model, provided that $\lambda\phi < 2$.

10.2.2 Using notation from Problem 10.2.1, the regression has the form

$$\text{(S10.2.3)} \qquad a_{t+1}/2 = \alpha + \beta b_t + u_{t+1}.$$

As $\mathrm{E}_t[a_{t+1}] = \gamma\phi x_t$ and $b_t = \gamma\lambda^{-1}x_t$, we see that the population parameters are $\alpha = 0$ and $\beta = \lambda\phi/2$. Clearly $\beta < 1$ since we have required $\lambda\phi < 2$.

10.2.3 Assume the process of the given form and simplify notation, introducing $a_t \equiv y_{1t} - y_{1,t-1}$ and $b_t \equiv y_{nt} - y_{1t}$. Note that

$$\text{(S10.2.4)} \qquad y_{n,t+1} - y_{nt} = b_{t+1} + a_{t+1} - b_t.$$

The equations of the model and of the postulated process are then

$$\text{(S10.2.5)} \qquad \begin{array}{ll} \text{I} & a_t = \lambda b_t + \epsilon_t, \\ \text{II} & b_t = (n-1)\mathrm{E}_t[b_{t+1} + a_{t+1} - b_t] + x_t, \\ \text{II} & x_t = \phi x_{t-1} + \eta_t, \\ \text{IV} & a_t = \gamma x_t + \epsilon_t. \end{array}$$

¿From the first and fourth equations we get $b_t = \gamma\lambda^{-1}x_t$; from the third we get $\mathrm{E}_t[b_{t+1}] = \gamma\lambda^{-1}\phi x_t$; from the third and fourth we get $\mathrm{E}_t[a_{t+1}] = \gamma\phi x_t$; and the second equation then gives the condition for the parameter γ:

$$\gamma = \frac{\lambda}{n - (n-1)\phi(1+\lambda)}. \tag{S10.2.6}$$

It is straightforward to verify that with this value of γ, the y_{1t} process satisfies all the equations of the model, provided that $(1+\lambda)\phi(n-1) < n$.

In our notation, the regression takes the form

$$b_{t+1} + a_{t+1} - b_t = \alpha + \beta \frac{b_t}{n-1} + u_{t+1}. \tag{S10.2.7}$$

As $\mathrm{E}[b_{t+1}+a_{t+1}-b_t] = (\gamma\lambda^{-1}\phi+\gamma\phi-\gamma\lambda^{-1})x_t$ and $b_t = \gamma\lambda^{-1}x_t$, we see that the population parameters are $\alpha = 0$ and $\beta = (1+\lambda)\phi(n-1) - (n-1)$. The parameter restrictions we have imposed allow β to be either positive or negative.

10.2.4 The model does explain why short-rate regressions of the type explored in Problem 10.2.2 give coefficients positive but less than one, while long-rate regressions of the type explored in Problem 10.2.3 often give negative coefficients. The underlying mechanism is a time-varying term premium, interacting with the desire of the monetary authority to smooth interest rates.

A limitation of this model is that it assumes a nonstationary interest rate process, which has unsatisfactory long-run properties. For example, with probability one the interest rate eventually becomes negative. Bennett McCallum, "Monetary Policy and the Term Structure of Interest Rates", NBER Working Paper No. 4938, 1994, works out a stationary version of this model; the algebra is more complicated but the properties of the model are similar.

Problems in Chapter 11

Solution 11.1

11.1.1 We assume throughout the problem that bond prices are determined by the homoskedastic lognormal model implied by equations (11.1.5) and (11.1.3),

$$-m_{t+1} = x_t + \beta\xi_{t+1} \qquad \text{(S11.1.1)}$$

$$x_{t+1} = (1-\phi)\,\mu + \phi x_t + \xi_{t+1}, \qquad \text{(S11.1.2)}$$

with $\xi_t \sim N(0,\sigma^2)$, but to fit the current term structure of interest rates we assume instead that the state variable follows the process given in equation (11.3.4):

$$x_{t+i} = x_{t+i-1} + g_{t+i} + \xi_{t+i}. \qquad \text{(S11.1.3)}$$

A useful way to relate the deterministic drift terms g_{t+i} and the parameters of the true pricing model when fitting the term structure of interest rates is to compute the forward rates implied by the assumed model (S11.1.1) and (S11.1.3), and compare them with those implied by the true model (S11.1.1) and (S11.1.2). To compute the forward rates implied by the assumed model we need first to compute the log bond prices, since $f_{n,t} = p_{n,t} - p_{n+1,t}$. Using equality (11.0.2) and the lognormal property of the stochastic discount factor, we have that

$$\begin{aligned} p_{n,t} &= \log \mathrm{E}_t\left[\prod_{i=1}^{n} M_{t+i}\right] \\ &= \mathrm{E}_t\left[\sum_{i=1}^{n} m_{t+i}\right] + \frac{1}{2}\,\mathrm{Var}_t\left(\sum_{i=1}^{n} m_{t+i}\right). \end{aligned}$$

But from the assumed model for the state variable (S11.1.3) we have

$$x_{t+i} = x_t + \sum_{j=1}^{i} g_{t+j} + \sum_{j=1}^{i} \xi_{t+j}, \qquad \text{(S11.1.4)}$$

so

$$\begin{aligned} \sum_{i=1}^{n} m_{t+i} &= -\sum_{i=1}^{n} x_{t+i-1} - \beta \sum_{i=1}^{n} \xi_{t+i} \\ &= -n\,x_t - \sum_{i=1}^{n} (n-i)\,g_{t+i} - \sum_{i=1}^{n} (\beta+n-i)^2\,\xi_{t+i} \end{aligned}$$

and

$$p_{n,t} = -n\,x_t - \sum_{i=1}^{n} (n-i)\,g_{t+i} + \frac{1}{2}\sum_{i=1}^{n} (\beta+n-i)\,\sigma^2. \qquad \text{(S11.1.5)}$$

We can now use (S11.1.5) to compute forward rates implied by the assumed model:

$$\begin{aligned} f_{n,t} &= p_{n,t} - p_{n+1,t} \\ &= x_t + \sum_{i=1}^{n} g_{t+i} - \frac{1}{2}\,(\beta+n)^2\,\sigma^2. \qquad \text{(S11.1.6)} \end{aligned}$$

Comparing (S11.1.6) with equation (11.1.14), that gives us the forward rates implied by the true model, we find immediately that the drift terms g_{t+i} are related to the parameters of the true model by the following expression:

$$\sum_{i=1}^{n} g_{t+i} = -\left(1-\phi^n\right)\left(x_t-\mu\right) - \frac{1}{2}\left[\left(\beta+\frac{1-\phi^n}{1-\phi}\right)^2 - (\beta+n)^2\right]\sigma^2. \tag{S11.1.7}$$

11.1.2 Since $r_{1,t+1} = -\mathrm{E}_t[m_{t+1}] - \mathrm{Var}_t\left(m_{t+1}\right)/2$, the short term interest rates at $(t+1)$ implied by the assumed model and the true model are the same:

$$r_{1,t+1} = x_t - \frac{1}{2}\beta^2\sigma^2. \tag{S11.1.8}$$

The dynamics of the state variable in the true model, given by (S11.1.2), and (S11.1.8) imply that future short rates equal:

$$\begin{aligned} r_{1,t+n+1} &= \mu\left(1-\phi^n\right) + \phi^n x_t + \sum_{i=1}^{n}\phi^{n-i}\xi_{t+i} - \frac{1}{2}\beta^2\sigma^2 \\ &= r_{1,t+1} - \left(1-\phi^n\right)\left(x_t-\mu\right) + \sum_{i=1}^{n}\phi^{n-i}\xi_{t+i}, \end{aligned}$$

so the expected future log short rates in the true model are

$$\mathrm{E}_t\left[r_{1,t+n+1}\right] = r_{1,t+1} - \left(1-\phi^n\right)\left(x_t-\mu\right). \tag{S11.1.9}$$

The dynamics of the state variable in the assumed model, given by (S11.1.3), imply:

$$\begin{aligned} r_{1,t+n+1} &= x_t + \sum_{i=1}^{n} g_{t+i} + \sum_{i=1}^{n}\xi_{t+i} - \frac{1}{2}\beta^2\sigma^2 \\ &= r_{1,t+1} + \sum_{i=1}^{n} g_{t+i} + \sum_{i=1}^{n}\xi_{t+i}, \end{aligned}$$

so expected future log short rates under the assumed model are

$$\mathrm{E}_t\left[r_{1,t+n+1}\right] = r_{1,t+1} + \sum_{i=1}^{n} g_{t+i}. \tag{S11.1.10}$$

Therefore, if we choose the drift terms so

$$\sum_{i=1}^{n} g_{t+i} = -\left(1-\phi^n\right)\left(x_t-\mu\right), \tag{S11.1.11}$$

the assumed model will be able to reproduce the expected short rates. However, by comparing (S11.1.7) and (S11.1.11) we can see that it is not possible to choose drift terms so they match simultaneously both current forward rates and expected future log short rates, since

$$\left(\beta+\frac{1-\phi^n}{1-\phi}\right)^2 \neq (\beta+n)^2$$

unless $\phi \to 1$, i.e., unless the state variable in the true model follows a random walk. It is also interesting to note that the set of deterministic drifts that matches expected future log short rates—see equation (S11.1.11)—converges to $-(x_t-\mu)$ as $n \to \infty$, while the set of deterministic drifts that matches forward rates—see equation (S11.1.7)— tends to $-\infty$ as $n \to \infty$. Therefore, if we choose the drift terms so they reproduce the forward rate structure of the true model, this will result in expected future log short rates declining without bound as we increase the horizon, while the true model implies that the expected future log short rates converge to a finite constant.

11.1.3 ¿From equation (11.1.8) and (S11.1.2), the time t conditional variance of log bond prices at time $t+1$ implied by the true bond pricing model is

$$\begin{aligned} \mathrm{Var}_t\left(p_{n,t+1}\right) &= B_{n-1}^2 \mathrm{E}_t\left[x_{t+1} - \mathrm{E}_t x_{t+1}\right]^2 \\ &= \left(\frac{1-\phi^n}{1-\phi}\right)^2 \sigma^2, \end{aligned} \tag{S11.1.12}$$

while from (S11.1.5) and (S11.1.3), the time t conditional variance of log bond prices at time $t+1$ implied by the assumed bond pricing model is

$$\mathrm{Var}_t\left(p_{n,t+1}\right) = n^2\sigma^2. \tag{S11.1.13}$$

Hence (S11.1.13) cannot be equal to (S11.1.12) unless $\phi \to 1$, i.e. unless the state variable follows a random walk in the true model. Moreover, for $n > 1$, the conditional variance of log bond prices implied by the assumed model is larger than the conditional variance implied by the true model and, while the true model implies that the conditional variance of log bond prices is bounded at $\sigma^2/(1-\phi)^2$ as $n \to \infty$, the assumed model implies an unbounded conditional variance.

11.1.4 Section 11.3.3 shows that the price of a European call option written on a zero-coupon that matures $n+\tau$ periods from now, with n periods to expiration and strike price X, is given under the true model by

$$C_{nt}\left(X\right) = P_{n+\tau,t}\,\Phi\left(d_1\right) + X\,P_{n,t}\,\Phi\left(d_2\right),$$

where $P_{n,t} = \exp\{p_{n,t}\} = \exp\{A_n + B_n x_t\}$ is the price of the bond, $\Phi(\circ)$ denotes the cumulative distribution function of a standard normal random variable,

$$\begin{aligned} d_1 &= \frac{p_{n+\tau,t} - x - p_{n,t} + \mathrm{Var}_t\left(p_{\tau,t+n}\right)/2}{\sqrt{\mathrm{Var}_t\left(p_{\tau,t+n}\right)}}, \\ d_2 &= d_1 - \sqrt{\mathrm{Var}_t\left(p_{\tau,t+n}\right)}, \end{aligned}$$

$x = \log(X)$ and

$$\begin{aligned} \mathrm{Var}_t\left(p_{\tau,t+n}\right) &= B_\tau^2\,\mathrm{Var}_t\left(x_{t+n}\right) \\ &= \left(\frac{1-\phi^\tau}{1-\phi}\right)^2 \left(\frac{1-\phi^{2n}}{1-\phi^2}\right)\sigma^2. \end{aligned} \tag{S11.1.14}$$

In our assumed model we use the same formula to value the option, except that we need to compute $\mathrm{Var}_t\left(p_{\tau,t+n}\right)$ under our assumed process for the state variable (S11.1.3). From (S11.1.5), we have

$$\begin{aligned} \mathrm{Var}_t\left(p_{\tau,t+n}\right) &= n^2\,\mathrm{Var}_t\left(x_{t+n}\right) \\ &= \tau^2 n\sigma^2, \end{aligned} \tag{S11.1.15}$$

where the second line follows from (S11.1.4).

Obviously, (S11.1.15) differs from (S11.1.14), unless $\phi \to 1$, so in general the assumed model will misprice options. For $\tau > 1$ and/or $n > 1$, it will overstate the volatility of the future log bond price, hence overvaluing the option. This overvaluation increases with the expiration date of the option and/or the maturity of the underlying bond. This is true no matter what combination of the drift parameters we choose. Backus, Foresi and Zin (1996) use this result to caution against the popular practice among practitioners of augmenting standard arbitrage-free bond pricing models with time-dependent parameters to fit exactly the yield curve. This augmentation may seriously misprice state-contingent claims, even though it is able to exactly reproduce the prices of some derivative securities.

Solution 11.2

11.2.1 The homoskedastic single-factor term-structure model of Section 11.1.1 holds:

$$-m_{t+1} = x_t + \beta\xi_{t+1} \tag{S11.2.1}$$

$$x_{t+1} = (1-\phi)\mu + \phi x_t + \xi_{t+1}. \tag{S11.2.2}$$

Thus, the price function for an n-period bond is

$$-p_{nt} = A_n + B_n x_t \tag{S11.2.3}$$

with

$$B_n = 1 + \phi B_{n-1} = \frac{1-\phi^n}{1-\phi}, \tag{S11.2.4}$$

$$A_n - A_{n-1} = (1-\phi)\mu B_{n-1} - (\beta + B_{n-1})^2\sigma^2/2, \tag{S11.2.5}$$

and $A_0 = B_0 = 0$.

Equation (11.3.15) in CLM gives the price at time t of an n-period forward contract on a zero coupon-bond which matures at time $t+n+\tau$ as $G_{\tau nt} = P_{\tau+n,t}/P_{nt}$. Taking logs, $g_{\tau nt} = p_{\tau+n,t} - p_{nt}$. Substituting out $p_{\tau+n,t}$ and p_{nt} using (S11.2.3)-(S11.2.5) yields:

$$-g_{\tau nt} = (A_{n+\tau} - A_n) + (B_{n+\tau} - B_n)x_t. \tag{S11.2.6}$$

Thus, the pricing function for an n-period forward contract on a zero coupon-bond which matures at time $t+n+\tau$ is given by:

$$-g_{\tau nt} = A^g_{\tau n} + B^g_{\tau n}x_t, \tag{S11.2.7}$$

with

$$\begin{aligned} A^g_{\tau n} &= A_{n+\tau} - A_n \\ B^g_{\tau n} &= B_{n+\tau} - B_n, \end{aligned}$$

where (S11.2.4) and (S11.2.5) can be used to write $A_{n+\tau}$, $B_{n+\tau}$ as functions of A_n, B_n. Clearly, the log forward price $g_{\tau nt}$ is affine in the state variable x_t.

In order to show that the log futures price $h_{\tau nt}$ is also affine in the state variable we can use equation (11.3.10) in CLM:

$$H_{\tau nt} = \mathrm{E}_t\left[M_{t+1}H_{\tau,n-1,t+1}/P_{1t}\right] \tag{S11.2.8}$$

Taking logs and assuming joint lognormality:

$$h_{\tau nt} = \mathrm{E}_t\left[m_{t+1} + h_{\tau,n-1,t+1} - p_{1t}\right] + \frac{1}{2}\mathrm{Var}_t\left[m_{t+1} + h_{\tau,n-1,t+1} - p_{1t}\right]. \tag{S11.2.9}$$

Let us first determine $h_{\tau 1t}$. Since $h_{\tau,0,t+1} = p_{\tau,t+1}$we have that:

$$h_{\tau 1t} = \mathrm{E}_t\left[m_{t+1} + p_{\tau,t+1} - p_{1t}\right] + \frac{1}{2}\mathrm{Var}_t\left[m_{t+1} + p_{\tau,t+1} - p_{1t}\right] \tag{S11.2.10}$$

Substituting out m_{t+1} using (S11.2.1) and $p_{\tau,t+1}$, p_{1t} using (S11.2.3) and (S11.2.2) yields:

$$\begin{aligned} h_{\tau 1t} \;=\; & \mathrm{E}_t\Big[-x_t - \beta\xi_{t+1} - A_\tau - B_\tau(1-\phi)\mu - \\ & B_\tau\phi x_t - B_\tau\xi_{t+1} + x_t - \beta^2\sigma^2/2\Big] \;+ \\ & \frac{1}{2}\mathrm{Var}_t\Big[-x_t - \beta\xi_{t+1} - A_\tau - B_\tau(1-\phi)\mu - \\ & B_\tau\phi x_t - B_\tau\xi_{t+1} + x_t - \beta^2\sigma^2/2\Big]\,. \end{aligned} \tag{S11.2.11}$$

Since $\mathrm{E}_t\xi_{t+1} = 0$ and $\mathrm{Var}_t\,\xi_{t+1} = \sigma^2$ it follows that:

$$-h_{\tau 1t} = A^h_{\tau 1} + B^h_{\tau 1}x_t \tag{S11.2.12}$$

with

$$A^h_{\tau 1} = A_\tau + (1-\phi)\mu B_\tau - \sigma^2/2\left[-\beta^2 + (\beta + B_\tau)^2\right]$$

$$B^h_{\tau 1} = \phi B_\tau.$$

Let us now solve for $h_{\tau nt}$. We guess that $-h_{\tau nt} = A^h_{\tau n} + B^h_{\tau n} x_t$. We proceed to verify our guess. At the same time we derive formulas for the coefficients $A^h_{\tau n}$, $B^h_{\tau n}$ as functions of the term structure coefficients A_n, B_n. Proceeding as above:

$$h_{\tau nt} = \mathrm{E}_t\left[m_{t+1} + h_{\tau,n-1,t+1} - p_{1t}\right] + \frac{1}{2}\mathrm{Var}_t\left[m_{t+1} + h_{\tau,n-1,t+1} - p_{1t}\right] \quad \text{(S11.2.13)}$$

and using our guess to substitute out for $h_{\tau,n-1,t+1}$:

$$\begin{aligned} h_{\tau nt} \quad = \quad & \mathrm{E}_t\Big[-x_t - \beta\xi_{t+1} - A^h_{\tau,n-1} - B^h_{\tau,n-1}(1-\phi)\mu - \\ & \qquad B^h_{\tau,n-1}\phi x_t - B^h_{\tau,n-1}\xi_{t+1} + x_t - \beta^2\sigma^2/2\Big] \; + \\ & \frac{1}{2}\mathrm{Var}_t\Big[-x_t - \beta\xi_{t+1} - A^h_{\tau,n-1} - B^h_{\tau,n-1}(1-\phi)\mu - \\ & \qquad B^h_{\tau,n-1}\phi x_t - B^h_{\tau,n-1}\xi_{t+1} + x_t - \beta^2\sigma^2/2\Big]. \end{aligned} \quad \text{(S11.2.14)}$$

We obtain:

(S11.2.15)

$$-h_{\tau nt} = A^h_{\tau,n-1} + B^h_{\tau,n-1}(1-\phi)\mu - \sigma^2/2\left[-\beta^2 + (\beta + B^h_{\tau,n-1})^2\right] + \phi B^h_{\tau,n-1} x_t.$$

Thus

$$A^h_{\tau n} = A^h_{\tau,n-1} + B^h_{\tau,n-1}(1-\phi)\mu - \sigma^2/2\left[-\beta^2 + (\beta + B^h_{\tau,n-1})^2\right],$$

$$B^h_{\tau n} = \phi B^h_{\tau,n-1}.$$

Solving recursively and using $B^h_{\tau 1} = \phi B_\tau$ yields

$$A^h_{\tau n} - A^h_{\tau,n-1} = B^h_{\tau,n-1}(1-\phi)\mu - \sigma^2/2\left[-\beta^2 + (\beta + \phi^{n-1}B_\tau)^2\right], \quad \text{(S11.2.16)}$$

$$B^h_{\tau n} = \phi^n B_\tau.$$

This completes Part 11.2.1.

11.2.2 The log ratio of forward to futures prices is given by

$$g_{\tau nt} - h_{\tau nt} = (A_n - A_{n+\tau}) + (B_n - B_{n+\tau})x_t + A^h_{\tau n} + B^h_{\tau n} x_t. \quad \text{(S11.2.17)}$$

In order to show that this is constant we need to show that $B_n - B_{n+\tau} + B^h_{\tau n} = 0$.

Straightforward algebra gives us:

$$\begin{aligned} B_n - B_{n+\tau} + B^h_{\tau n} \quad &= \quad B_n - B_{n+\tau} + \phi^n B_\tau = \\ &= \quad \frac{1 - \phi^n - 1 + \phi^{n+\tau} + \phi^n - \phi^{n+\tau}}{1-\phi} = 0. \end{aligned} \quad \text{(S11.2.18)}$$

Showing that the ratio of forward to future prices is greater than one is equivalent to showing that the log ratio is greater than zero. In order to do so we write

$$\begin{aligned} g_{\tau nt} - h_{\tau nt} \quad &= \quad A_n - A_{n+\tau} + A^h_{\tau n} = \\ &= \quad A_{n-1} + (1-\phi)\mu B_{n-1} - (\beta + B_{n-1})^2\sigma^2/2 \\ & \quad\;\; - A_{n+\tau-1} - (1-\phi)\mu B_{n+\tau-1} + (\beta + B_{n+\tau-1})^2\sigma^2/2 \\ & \quad\;\; + A^h_{t,n-1} + (1-\phi)\mu B^h_{\tau,n-1} - \sigma^2/2\left[-\beta^2 + (\beta + B^h_{\tau,n-1})^2\right]. \end{aligned}$$

It can easily be checked that the terms in $(1-\phi)\mu$ add up to zero. Given the recursive nature of the problem and remembering that $g_{\tau 1t} - h_{\tau 1t} = 0$ we have that

$$g_{\tau nt} - h_{\tau nt} = -\sigma^2/2 \sum_{j=0}^{n-1} \left[(\beta + B_j)^2 - (\beta + B_{j+\tau})^2 - \beta^2 + (\beta + \phi^j B_\tau)^2 \right]. \tag{S11.2.19}$$

Using the fact that $\phi^j B_\tau = B_{j+\tau} - B_j$, after some algebra, we obtain

$$g_{\tau nt} - h_{\tau nt} = \frac{\sigma^2(1-\phi^\tau)(1-\phi^{n-1})(1-\phi^n)\phi}{(1-\phi)^3(1+\phi)}. \tag{S11.2.20}$$

Thus we have that

$$\begin{aligned} g_{\tau nt} - h_{\tau nt} &> 0 \quad \text{when} \quad 0 < \phi < 1 \\ g_{\tau nt} - h_{\tau nt} &< 0 \quad \text{when} \quad -1 < \phi < 0. \end{aligned} \tag{S11.2.21}$$

The difference between a futures contract and a forward contract is that the first is marked to market each period during the life of the contract, so that the purchaser of a futures contract receives the futures price increase or pays the futures price decrease each period. When interest rates are random, these mark-to-market payments may be correlated with interest rates. When $0 < \phi < 1$, so that $\mathrm{Cov}(h_{\tau nt}, y_{1t}) < 0$, the purchaser of a futures contract tends to receive the futures price increase at times when interest rates are low, and tends to pay the futures price decrease at times when interest rates are high, making the futures contract worth less than the forward contract. On the other hand, when $-1 < \phi < 0$, the futures contract will be worth more than the forward contract since its purchaser tends to receive price increases when interest rates are high (so that the money can be invested at a high rate of return).

11.2.3 The parameter values for this part, $\phi = 0.98$ and $\sigma^2 = 0.00051^2$, can be found in Section 11.2.2, page 453 (and not in Section 11.1.2).

Problems in Chapter 12

Solution 12.1

12.1.1 There are several criteria with which random number generators can be judged:

- Stochastic *quality* of apparent randomness, as reflected in the probabilistic properties of generated sample and assessed by batteries of statistical tests of independence, goodness-of-fit to specific probability distributions, etc.
- Computational *efficiency*, in terms of cost of implementation, resource requirements, volume of output per second, volume of output in absolute terms, all without deterioration of stochastic quality.
- *Portability* of the algorithm.
- *Reproducibility* of random series (based on the initial "seed" of the random number generator).

The ultimate introduction to the science and art of pseudorandom number generation is Chapter 3 of D. E. Knuth's classics *The Art of Computer Programming* 1969, 1981, where the most influential and comprehensive study of the subject is to be found.

One example of the many recent treatises on the state of the art is Fishman (1996), which emphasizes pseudorandom number generators in Chapter 7. High-quality pseudorandom number generators also emerge in cryptography. Cryptographically secure generators, related to *stream ciphers* and *one-way hash functions* achieve extraordinary stochastic quality, generally at the expense of increasing computation costs. See for example Schneier (1996, Chapters 16–18).

There exist batteries of statistical tests intended to measure stochastic quality of pseudorandom number generators; these include tests such as chi-square, Kolmogorov-Smirnov, frequency, serial, gap, permutation, run, moments, serial correlation, and especially spectral tests (see Knuth [1969] for details); or, for example, an omnibus test assessing joint independence and one- to three-dimensional uniformity, assembled by Fishman (1996, Section 7.12).

12.1.2 Generally, very well researched and tested MLCG generators constitute an accepted pragmatic compromise among the criteria imposed on pseudorandom number generators discussed in Problem 12.1.1. The proper choice of parameters of MLCG generators is essential, and theoretical guidelines are readily available in Knuth (1969) and elsewhere. The quality of the tent- and logistic-map generators is inferior for most purposes, as most standard statistical tests of randomness will show.

The extra modification of using parameters like 1.99999999 instead of 2 etc. patches the most obvious flaw of the tent- and logistic-map generators: with real numbers represented in binary form using finite-length mantissas, repetitive multiplication by 2 deteriorates quality of the sequence rapidly, i.e., the sequence degenerates in time that is proportional to the mantissa length. In most practical cases, though, the use of well-researched pseudorandom number generators with solid theoretical guarantees of quality, such as MLCG, is indicated.

If the quality of even properly chosen MLCG is not sufficient for an application at hand, one may consider using some other classes of well-tested generators with balanced

Table Estimates of kernel-regression betas of IBM relative to S&P 500 based on monthly return data from 1965:1 to 1994:12. Each estimate is local to a particular level of S&P 500 monthly return.

SP500 [%]	$\hat{\beta}_{\text{IBM,SP500}}$
−15	1.366
−10	1.395
−5	0.689
0	0.666
5	0.806
10	0.531
15	1.994

TABLE 12.1. IBM Betas Relative to S&P 500

quality-cost tradeoffs, for example, Marsaglia's lagged-Fibonacci generators (see Marsaglia and Zaman [1991]).

Solution 12.2

Equations (12.4.1) and (12.4.3) describe one unit. Our case involves ten such units with $J \equiv 5$. The output layer is given by equation (12.4.4) with $K \equiv 10$. For simplicity, choose $h(\cdot)$ to be the identity, in accord with the discussion on pages 514–542. Thus, the nonlinear model has 60 parameters to fit. Using a nonlinear optimization technique of choice, find the parameter values that attain the minimum (beware of local minima!) in-sample root-mean-squared-error (RMSE) of the one-step-ahead estimate, with identical weights given to each datapoint of S&P 500 returns from 1926:1 to 1985:12. Then, apply the fitted perceptron parameters on data in period from 1986:1 to 1994:12.

The RMSE will be substantially larger in the out-of sample period than in the in-sample period. The out-of-sample RMSE 60-parameter perceptron will probably not be drastically smaller than out-of-sample RMSE of a linear model with less immodest number of parameters (say, ten (10); consider an OLS regression with five lagged returns and their squares as explanatory variables), but the in-sample RMSE of the former will be noticeably smaller than RMSE of the latter. This phenomenon can be related to concept of "overfitting" which occurs when lack of structural, qualitative information of the data generating stochastic process is countered by increase in number of *ad hoc* degrees of freedom in the model: this procedure results in excellent in-sample fit while out-of-sample performance stays mediocre.

Solution 12.3

First implement the kernel regression estimator $\hat{m}_h(x)$ according to formula (12.3.9) with a Gaussian kernel $K_h(x)$ as in (12.3.10). Second, determine optimal bandwidth by minimizing the cross-validation function $\text{CV}(h)$ as in (12.3.13), based on estimator $\hat{m}_h(x)$ and given historical S&P 500 and IBM monthly returns.

Numerically, the appropriate bandwidth for period from 1965:1 to 1994:12 is $h = 1.49\%$ (the scale is in monthly returns of S&P 500). The resulting regression is plotted (Figure 12.1).

The analog of the conventional beta estimate here is the quantity $\partial\hat{m}_h(x)/\partial x$, evaluated at particular level of S&P 500 return x. See Section (12.3.3) for a detailed discussion of average derivative estimators.

Let us replace derivative by its discrete analog with a step length difference of 1% of the S&P 500 monthly return. The resulting estimates of β's for different levels of S&P 500 returns is shown in Table 12.1.

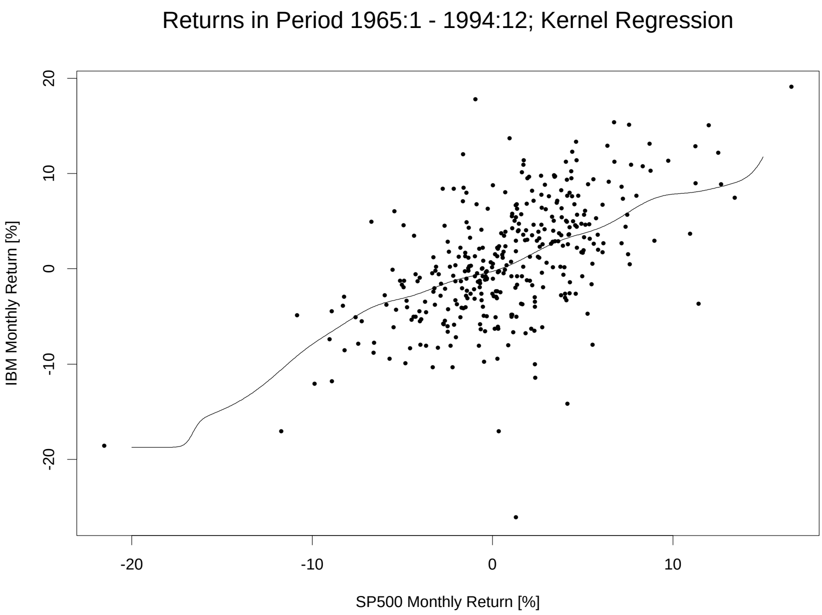

FIGURE 12.1. Kernel Regression of IBM Returns on S&P 500 Returns

We see that the local estimates of beta vary considerably, most likely due to the relatively small number of datapoints in the estimation, possible variation in beta over time, or genuine nonlinearity of the relation between IBM and S&P 500 monthly returns.

Some advantages of kernel regression relative to ordinary least squares are: cross-validation allows for nonparametric, adaptive and asymptotically consistent estimation of the true relation between IBM and S&P 500 returns even when this relation is not linear; the kernel estimator $\hat{m}(x)$ conveys more information about the relationship than a single parameter (β) and allows easy visualization of the relation.