Lecture Notes for Empirical Finance 2003 (second year PhD course in Stockholm)

Paul Söderlind[1]

January 2003 (some typos corrected later)

[1]University of St. Gallen and CEPR. *Address:* s/bf-HSG, Rosenbergstrasse 52, CH-9000 St. Gallen, Switzerland. *E-mail:* Paul.Soderlind@unisg.ch. Document name: EFAll.TeX.

Contents

1 GMM Estimation of Mean-Variance Frontier

Reference: Hamilton (1994); Greene (2000); Campbell, Lo, and MacKinlay (1997); Cochrane (2001)

These notes are about how we can use GMM and the delta method to estimate a mean-variance frontier and also calculate a confidence band for the frontier. The purpose is to provide a quick repetition of GMM and the delta method—and to teach you how to implement it in MatLab (or some other programming language). The approach in these notes is therefore very practical and contains essentially no new ingredients for those who have taken the PhD courses in econometrics and finance. The combination of ingredients, as well as the implementation, might be new, however.

1.1 GMM Estimation of Means and Covariance Matrices

Let R_t be a vector of net returns of N assets. We want to estimate the mean vector and the covariance matrix. The moment conditions for the mean vector are

$$\mathrm{E}\, R_t - \mu = \mathbf{0}_{N\times 1}, \tag{1.1}$$

and the moment conditions for the second moment matrix are

$$\mathrm{E}\, R_t R_t' - \Gamma = \mathbf{0}_{N\times N}. \tag{1.2}$$

Example 1 *With $N = 2$ we have*

$$\mathrm{E}\begin{bmatrix} R_{1t} \\ R_{2t} \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ and } \mathrm{E}\begin{bmatrix} R_{1t}^2 & R_{1t}R_{2t} \\ R_{2t}R_{1t} & R_{2t}^2 \end{bmatrix} - \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

The second moment matrix is symmetric and contains therefore only $N(N+1)/2$ unique elements. Although there is some book keeping involved in getting rid of the non-unique elements (and then getting back to the full second moment matrix again), it is often worth the trouble—in particular if we later want to apply the delta method (take derivatives with respect to the parameters). The standard way of doing that is through the vech operator.

Remark 2 *(The vech operator) vech(A) where A is $m \times m$ gives an $m(m+1)/2 \times 1$ vector with the elements on and below the principal diagonal A stacked on top of each other (column wise). For instance, vech* $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{22} \end{bmatrix}$.

Remark 3 *(Duplication matrix) The duplication matrix D_m is defined such that for any symmetric $m \times m$ matrix A we have* $\mathrm{vec}(A) = D_m vech(A)$. *The duplication matrix is therefore useful for "inverting" the vech operator (the step from* $\mathrm{vec}(A)$ *to A is trivial). For instance, to continue the example of the vech operator*

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{21} \\ a_{22} \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{21} \\ a_{22} \end{bmatrix} \text{ or } D_2 vech(A) = \mathrm{vec}(A).$$

We therefore pick out the unique elements in (1.2) by

$$\mathrm{E}\,\mathrm{vech}(R_t R_t') - \mathrm{vech}(\Gamma) = \mathbf{0}_{N(N+1)/2\times 1}. \tag{1.3}$$

Example 4 *Continuing the previous example, we get*

$$\mathrm{E}\begin{bmatrix} R_{1t}^2 \\ R_{2t}R_{1t} \\ R_{2t}^2 \end{bmatrix} - \begin{bmatrix} \Gamma_{11} \\ \Gamma_{21} \\ \Gamma_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Stack (1.1) and (1.3) and substitute the sample mean for the population expectation to get the GMM estimator

$$\frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} R_t \\ \mathrm{vech}(R_t R_t') \end{bmatrix} - \begin{bmatrix} \hat{\mu} \\ \mathrm{vech}(\hat{\Gamma}) \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{N\times 1} \\ \mathbf{0}_{N(N+1)/2\times 1} \end{bmatrix}, \text{ or} \tag{1.4}$$

$$\frac{1}{T}\sum_{t=1}^{T} x_t - \hat{\beta} = \mathbf{0}. \tag{1.5}$$

Remark 5 *(MatLab coding) Let R be a $T \times N$ matrix with R_t' in row t. Let Z be a $T \times N(N+1)/2$ matrix with* $\mathrm{vech}(R_t R_t')'$ *in row t. Then,* `mean([R,Z])'` *calculates the*

column vector $\hat{\beta}$. To be concrete, if $T = 3$ and $N = 2$, then the R and Z matrices are

$$R = \begin{bmatrix} R_{1t} & R_{2t} \\ R_{1t+1} & R_{2t+1} \\ R_{1t+2} & R_{2t+2} \end{bmatrix} \text{ and } Z = \begin{bmatrix} R_{1t}^2 & R_{2t}R_{1t} & R_{2t}^2 \\ R_{1t+1}^2 & R_{2t+1}R_{1t+1} & R_{2t+1}^2 \\ R_{1t+2}^2 & R_{2t+1}R_{1t+2} & R_{2t+2}^2 \end{bmatrix}.$$

Remark 6 *(More MatLab coding) It is easy to get the returns data into the R matrix. Once that is done, the simplest (but not the fastest) way to construct the Z matrix is by a loop. First, get a vech function (pre-installed in Octave, but not in MatLab). Then, use the following code:*

```
Z = zeros(T,N*(N+1)/2);
for i=1:T;
  Z(i,:) = vech(R(i,:)'*R(i,:))';
end;
```

Remark 7 *(More MatLab coding, continued) To understand what the code does, note that if R(2,:) is row 2 of R, then*

$$\begin{aligned} R(2,:)' * R(2,:) &= \begin{bmatrix} R_{1t+1} & R_{2t+1} \end{bmatrix}' \begin{bmatrix} R_{1t+1} & R_{2t+1} \end{bmatrix} \\ &= \begin{bmatrix} R_{1t+1} \\ R_{2t+1} \end{bmatrix} \begin{bmatrix} R_{1t+1} & R_{2t+1} \end{bmatrix} \\ &= \begin{bmatrix} R_{1t+1}^2 & R_{1t+1}R_{2t+1} \\ R_{2t+1}R_{1t+1} & R_{2t+1}^2 \end{bmatrix}. \end{aligned}$$

1.2 Asymptotic Distribution of the GMM Estimators

1.2.1 GMM in General

In general, the sample moment conditions in GMM are written

$$\bar{m}(\beta) = \frac{1}{T}\sum_{t=1}^{T} m_t(\beta) = \mathbf{0}, \tag{1.6}$$

where $\bar{m}(\beta)$ is short hand notation for the sample average and where the value of the moment conditions clearly depend on the parameter vector. We let β_0 denote the true value of the parameter vector. The GMM estimator is

$$\hat{\beta} = \arg\min \bar{m}(\beta)' W \bar{m}(\beta), \tag{1.7}$$

where W is some symmetric positive definite weighting matrix.

GMM estimators are typically asymptotically normally distributed, with a covariance matrix that depends on the covariance matrix of the moment conditions (evaluated at the true parameter values) and the possibly non-linear transformation of the moment conditions that defines the estimator. Let S_0 be the covariance matrix of $\sqrt{T}\bar{m}(\beta_0)$ (evaluated at the true parameter values)

$$S_0 = \lim_{T\to\infty} \text{Cov}\left[\sqrt{T}\bar{m}(\beta_0)\right] = \sum_{s=-\infty}^{\infty} \text{Cov}\left[m_t(\beta_0), m_{t-s}(\beta_0)\right], \tag{1.8}$$

and let G_0 be the probability limit of the gradient of the sample moment conditions with respect to the parameters (also evaluated at the true parameters)

$$G_0 = \text{plim}\, \frac{\partial \bar{m}(\beta_0)}{\partial \beta'}. \tag{1.9}$$

We the have that

$$\begin{aligned} \sqrt{T}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V) \text{ if } W = S_0^{-1}, \text{ where} \\ V = \left(G_0' S_0^{-1} G_0\right)^{-1}. \end{aligned} \tag{1.10}$$

The choice of weighting matrix is irrelevant if the model is exactly identified (as many moment conditions as parameters), so (1.10) can be applied to this case (even if we did not specify any weighting matrix at all).

The Newey-West estimator is commonly used to estimate the covariance matrix S_0. It essentially evaluates m_t at the point estimates instead of at β_0 and then calculates

$$\widehat{\text{Cov}}\left(\sqrt{T}\bar{m}\right) = \sum_{s=-n}^{n} \left(1 - \frac{|s|}{n+1}\right) \widehat{\text{Cov}}\,(m_t, m_{t-s}) \tag{1.11}$$

$$= \widehat{\text{Cov}}\,(m_t, m_t) + \sum_{s=1}^{n} \left(1 - \frac{s}{n+1}\right) \left(\widehat{\text{Cov}}\,(m_t, m_{t-s}) + \widehat{\text{Cov}}\,(m_t, m_{t-s})'\right) \tag{1.12}$$

where n is a finite "bandwidth" parameter and where m_t is short hand notation for $m_t(\hat{\beta})$.

Example 8 *With $n = 1$ in (1.11) the Newey-West estimator becomes*

$$\widehat{\text{Cov}}\left(\sqrt{T}\bar{m}\right) = \widehat{\text{Cov}}\left(m_t, m_t\right) + \frac{1}{2}\left(\widehat{\text{Cov}}\left(m_t, m_{t-1}\right) + \widehat{\text{Cov}}\left(m_t, m_{t-1}\right)'\right).$$

Remark 9 *(MatLab coding) Suppose we have an $T \times K$ matrix M with m_t' in row t. We want to calculate $\widehat{\text{Cov}}\left(m_t, m_{t-s}\right)$. The first step is to make all the columns in M have zero means and the second is to calculate $\Sigma_{t=s+1}^{T} m_t m_{t-s}'/T$ as in*

```
Mb = M - repmat(mean(M),T,1);       %has zero means
Cov_s = Mb(s+1:T,:)'*Mb(1:T-s,:)/T;
```

In practice, the gradient G_0 is also approximated by using the point estimates and the available sample of data.

1.2.2 Application to the Means and Second Moment Matrix

To use the results in the previous section to estimate the covariance matrix of $\hat{\beta}$ in (1.5), we note a few things. First, The gradient G_0 in (1.9) is an identity matrix, so the covariance matrix in (1.10) simplifies to $V = S_0$. Second, we can estimate S_0 by letting $m_t = x_t - \hat{\beta}$ (see (1.5)) and then use (1.12), perhaps with $n = 1$ as in the example (there is little autocorrelation in returns data).

1.3 Mean-Variance Frontier

Equipped with (estimated) mean vector μ and the unique elements of the second moments matrix, we can calculate the mean variance frontier of the (risky) assets.

1.3.1 Calculate the Mean-Variance Frontier

With $\hat{\beta}$, it is straightforward to calculate the mean-variance frontier. It should be thought of (and coded) as a function that returns the lowest possible standard deviation for a given required return and parameters β (μ and vech(Γ)). We denote this function $g(\beta, \mu_p)$, where μ_p is vector of required returns and note that it should return a vector with as many elements as in the vector μ_p.

1.3.2 Applying the Delta Method on the Mean-Variance Frontier

Recall the definition of the delta method. For notational convenience, we suppress the μ_p argument.

Remark 10 *(Delta method) Consider an estimator $\hat{\beta}_{k\times 1}$ which satisfies*

$$\sqrt{T}\left(\hat{\beta} - \beta_0\right) \xrightarrow{d} N\left(0, V_{k\times k}\right),$$

and suppose we want the asymptotic distribution of a transformation of β

$$\gamma_{q\times 1} = g\left(\beta\right),$$

where $g\left(.\right)$ is has continuous first derivatives. The result is

$$\sqrt{T}\left[g\left(\hat{\beta}\right) - g\left(\beta_0\right)\right] \xrightarrow{d} N\left(0, \Psi_{q\times q}\right), \text{ where}$$

$$\Psi = \frac{\partial g\left(\beta_0\right)}{\partial \beta'} V \frac{\partial g\left(\beta_0\right)'}{\partial \beta}, \text{ where } \frac{\partial g\left(\beta\right)}{\partial \beta'} = \begin{bmatrix} \frac{\partial g_1(\beta)}{\partial \beta_1} & \cdots & \frac{\partial g_1(\beta)}{\partial \beta_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_q(\beta)}{\partial \beta_1} & \cdots & \frac{\partial g_q(\beta)}{\partial \beta_k} \end{bmatrix}_{q\times k}$$

To apply this to the mean variance frontier, note the following. The estimator $\hat{\beta}$ contains the estimated means, $\hat{\mu}$, and the unique elements in the second moment matrix, vech($\hat{\Gamma}$), and the V is the covariance matrix of $\sqrt{T}\hat{\beta}$ discussed in Section 1.2.2. The function $g(\beta, \mu_p)$ defines the minimum standard deviation, at the vector of required returns μ_p, as a function of the parameters β.

These derivatives can typically be very messy to calculate analytically, but numerical approximations often work fine. A simple code can be structured as follows: First, calculate the $q \times 1$ vector $g(\hat{\beta}, \mu_p)$ at the point estimates by one call of the function g (note that the function should return a vector with as many elements as in μ_p). Second, set up an empty $q \times k$ matrix to fill with the partial derivatives. Third, change the first element in $\hat{\beta}$, denoted $\hat{\beta}_1$, to $1.05\hat{\beta}_1$, and call the new parameter vector $\tilde{\beta}$. Calculate the $q \times 1$ vector $g(\tilde{\beta}, \mu_p)$, and then $\Delta = [g(\tilde{\beta}, \mu_p) - g(\hat{\beta}, \mu_p)]/(0.05\hat{\beta}_1)$ and fill the first column in the empty matrix with Δ. Repeat for the other parameters. Here is a simple code:

```
gbhat = g(bhat,mup);
Dg_db = zeros(q,k);
for j = 1:k;        %loop over columns (parameters)
```

```
  bj          = bhat;
  bj(j)       = 1.05*bhat(j);
  Dg_db(:,j) = (g(bj,mup)- gbhat)/(0.05*bhat(j));
end;
```

1.3.3 Recover the Covariance Matrix

One of the steps that need to be solved when we define the $g(\beta, \mu_p)$ function is how we can reconstruct the covariance matrix of R from the means and the unique elements in the second moment matrix. This calculation involves two straightforward steps. First, we get the vec of the estimated second moment matrix, vec($\hat{\Gamma}$), from its unique elements in vech($\hat{\Gamma}$) by premultiplying the latter with the duplication matrix of order N, D_N. Then, reshuffle the vec($\hat{\Gamma}$) to get $\hat{\Gamma}$ (in MatLab you would use the `reshape` command). Second, recall the definition of the covariance matrix of the vector R with mean vector μ

$$\begin{aligned}\Sigma = \text{Cov}(R) &= \text{E}\,(R-\mu)\,(R-\mu)' \\ &= \text{E}\,RR' - \mu\mu', \end{aligned} \tag{1.13}$$

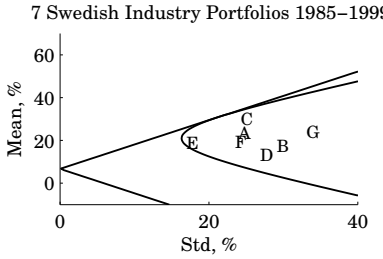
which gives the covariance matrix.

Figures 1.1 and 1.2 show some empirical results. The uncertainty is lowest for the minimum variance portfolio (in a normal distribution, the uncertainty about an estimated variance is increasing in the true variance, $\text{Var}(\sqrt{T}\hat{\sigma}^2) = 2\sigma^4$).

Remark 11 *(Var($\sqrt{T}\hat{\sigma}^2$) as an application of GMM). To estimate the variance of a zero mean variable we specify the moment condition $m_t = x_t^2 - \sigma^2$. If data is iid (all autocorrelations drop put), we have $S_0 = \text{Var}(m_t) = \text{E}(x_t^2 - \sigma^2)^2$. By expanding we get $\text{E}(x_t^4 + \sigma^2 - 2x_t^2\sigma^2) = \text{E}\,x_t^4 - \sigma^4$. From the properties of normally distributed variables we know that $\text{E}\,x_t^4 = 3\sigma^4$, which means $S_0 = 2\sigma^4$. Note that the Jacobian is -1, so the GMM formula says $\sqrt{T}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{d} N(0, 2\sigma^4)$.*

Bibliography

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay, 1997, *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey.



Portf	Std	Mean	SR
A	23.98	23.45	0.98
B	29.12	17.48	0.60
C	24.29	30.41	1.25
D	26.90	13.59	0.50
E	17.02	19.23	1.13
F	23.52	19.53	0.83
G	33.11	24.24	0.73

Figure 1.1: Mean-Variance frontier of Swedish industry portfolios (weekly returns, $100\sqrt{52}$Variance is plotted against $100 * 52*$mean)

Cochrane, J. H., 2001, *Asset Pricing*, Princeton University Press, Princeton, New Jersey.

Greene, W. H., 2000, *Econometric Analysis*, Prentice-Hall, Upper Saddle River, New Jersey, 4th edn.

Hamilton, J. D., 1994, *Time Series Analysis*, Princeton University Press, Princeton.

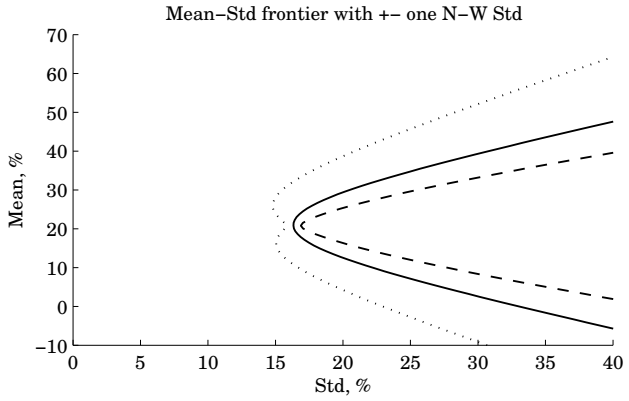


Figure 1.2: Mean-Variance frontier fo Swedish industry portfolios — with confidence band

2 Predicting Asset Returns

Reference: Cuthbertson (1996) 5 and 6; Cochrane (2001) 20.1; Campbell, Lo, and MacKinlay (1997) 2 and 7

2.1 A Little Financial Theory and Predictability

The traditional interpretation of autocorrelation in asset returns is that there are some "irrational traders." For instance, feedback trading would create positive short term autocorrelation in returns.

A more modern interpretation is that there are movements in expected (excess) returns because of change in riskiness. To see how this fits into the SDF framework, let R_{t+1} be the gross return on any asset. The canonical asset pricing equation then says

$$\mathrm{E}_t\, m_{t+1} R_{t+1} = 1, \tag{2.1}$$

where m_{t+1} is the stochastic discount factor.

By using the fact that $\mathrm{Cov}(x, y) = \mathrm{E}(x - \mathrm{E}\,x)(y - \mathrm{E}\,y) = \mathrm{E}\,xy - \mathrm{E}\,x\,\mathrm{E}\,y$ (which applies also to conditional moments), we can rewrite (2.1) as

$$\mathrm{E}_t\, m_{t+1}\, \mathrm{E}_t\, R_{t+1} + \mathrm{Cov}_t(m_{t+1}, R_{t+1}) = 1. \tag{2.2}$$

Let R_{ft+1} be the return on an asset with zero conditional correlation with the SDF, that is, a riskfree asset. It satisfies

$$\mathrm{E}_t\, m_{t+1}\, \mathrm{E}_t\, R_{ft+1} = 1. \tag{2.3}$$

Subtracting (2.3) from (2.2), rearranging, and using the implication from (2.3) that $\mathrm{E}_t\, R_{ft+1} = 1/\,\mathrm{E}_t\, m_{t+1}$ gives

$$\mathrm{E}_t(R_{t+1} - R_{ft+1}) = -\,\mathrm{Cov}_t(m_{t+1}, R_{t+1})\,\mathrm{E}_t\, R_{ft+1}. \tag{2.4}$$

This says that the expected excess return will vary unless the product of the conditional

covariance and the expected riskfree rate is constant. Note that a changing expected excess return is the same as that the future excess return is forecastable. Also note that if the right hand side of (2.4) is constant, then $\mathrm{E}_t\, R_{t+1}$will move one-for-one with $\mathrm{E}_t\, R_{ft+1}$, so any predictability in the riskfree rate carries over to the risky return.

We still have an "annoying" $\mathrm{E}_t\, R_{ft+1}$ term in (2.4). Consider nominal returns and suppose we are in a situation where inflation is expected to be much higher in the future (change of monetary policy regime). Even if the covariance in (2.4) is unchanged (as it would be if this is a deterministic increase in inflation which has no real effects), the expected excess return changes (as long as the covariance is non-zero) since $\mathrm{E}_t\, R_{ft+1}$ increases. Clearly, this effect disappears if we study real returns instead or if we define the excess return as R_{t+1}/R_{ft+1} instead, as we implicitly do when we analyze continuously compounded returns.

For instance, if the SDF and the return are (conditionally) lognormally distributed, then we can arrive at a simple expression for the excess continuously compounded return. Let $r_t = \ln R_t$. If $\ln m_{t+1} + r_{t+1}$ is normally distributed, then we can write the pricing equation as

$$\begin{aligned} 1 &= \mathrm{E}_t \exp(\ln m_{t+1} + r_{t+1}) \\ &= \exp[\mathrm{E}_t \ln m_{t+1} + \mathrm{E}_t\, r_{t+1} + \mathrm{Var}_t(\ln m_{t+1} + r_{t+1})/2]. \end{aligned} \tag{2.5}$$

Take logs to get

$$0 = \mathrm{E}_t \ln m_{t+1} + \mathrm{E}_t\, r_{t+1} + \mathrm{Var}_t(\ln m_{t+1})/2 + \mathrm{Var}_t(r_{t+1})/2 + \mathrm{Cov}_t(\ln m_{t+1}, r_{t+1}) \tag{2.6}$$

The same applies to the riskfree asset with return r_{ft+1}, except that the covariance with the SDF is zero. The expected excess (continuously compounded) return is then

$$\mathrm{E}_t\, r_{t+1} - \mathrm{E}_t\, r_{ft+1} = -\mathrm{Var}_t(r_{t+1})/2 + \mathrm{Var}_t(r_{ft+1})/2 - \mathrm{Cov}_t(\ln m_{t+1}, r_{t+1}). \tag{2.7}$$

This expression relates the excess return to the second moments, in particular, the covariance of the log SDF and the log return.

As an example, consider a simple consumption-based model to start thinking about these issues. Suppose we want to maximize the expected discounted sum of utility

$\mathrm{E}_t\, \Sigma_{s=0}^{\infty} \beta^s u(c_{t+s})$. Let Q_t be the consumer price index in t. Then, we have

$$m_{t+1} = \begin{cases} \beta \frac{u'(c_{t+1})}{u'(c_t)} \frac{Q_t}{Q_{t+1}} \text{ if returns are nominal} \\ \beta \frac{u'(c_{t+1})}{u'(c_t)} \text{ if returns are real.} \end{cases} \tag{2.8}$$

This shows that to get a time-varying risk premium in (2.7), the conditional covariance of the log marginal utility and the return must change (disregarding the Jensen's inequality terms).

2.1.1 Application: Mean-Variance Portfolio Choice with Predictable Returns

If there are non-trivial market imperfections, then predictability can be used to generate economic profits. If there are no important market imperfections, then predictability of excess returns should be thought of as predictable movements in risk premia. We will typically focus on this second interpretation.

As a simple example, consider a small mean-variance investor whose preferences differ from the average investor: he is not affected by risk that creates the time variation in expected returns. Suppose he can invest in two assets: a risky asset with return R_{t+1} and a risk free asset with return R_f and zero variance. The return on the portfolio is $R_{p,t+1} = \alpha R_{t+1} + (1-\alpha) R_f$. The utility is quadratic in terms of wealth: $\mathrm{E}_t\, R_{p,t+1} - \mathrm{Var}_t(R_{p,t+1})k/2$. Substituting gives that the maximization problem is

$$\max_\alpha \alpha\, \mathrm{E}_t\, R_{t+1} + (1-\alpha) R_f - \frac{k}{2}\alpha^2\, \mathrm{Var}_t(R_{t+1}).$$

The first order condition is

$$\begin{aligned} 0 &= \mathrm{E}_t\, R_{t+1} - R_f - k\alpha\, \mathrm{Var}_t(R_{t+1}) \text{ or} \\ \alpha &= \frac{1}{k} \frac{\mathrm{E}_t\, R_{t+1} - R_f}{\mathrm{Var}_t(R_{t+1})}. \end{aligned}$$

The weight on the risky asset is clearly increasing in the excess return and decreasing in the variance. If we compare two investors of this type and the same k, but with different investment horizons, then the portfolio weight α is the same if the ratio of the mean and variance is unchanged by the horizon. This is the case if returns are iid.

To demonstrate the last point note that the two period excess return is approximately equal to the sum of two one period returns $R_{t+1} - R_f + R_{t+2} - R_f$. With iid returns the

mean of this two period excess return is $2\,\mathrm{E}(R_{t+1} - R_f)$ and the variance is $2\,\mathrm{Var}(R_{t+1})$ since the covariance of the returns is zero. We therefore get

$$\alpha \text{ for 1-period horizon} = \frac{1}{k}\frac{\mathrm{E}_t\, R_{t+1} - R_f}{\mathrm{Var}_t(R_{t+1})},$$
$$\alpha \text{ for 2-period horizon} = \frac{1}{k}\frac{2\,\mathrm{E}_t(R_{t+1} - R_f)}{2\,\mathrm{Var}_t(R_{t+1})},$$

which are the same.

With correlated returns we still have the same two period mean, but the variance is now $\mathrm{Var}(R_{t+1} + R_{t+2}) = 2\,\mathrm{Var}(R_{t+1}) + 2\,\mathrm{Cov}(R_{t+1}, R_{t+2})$. This gives the portfolio weight on the risky asset

$$\alpha \text{ for 2 period horizon} = \frac{1}{k}\frac{2\,\mathrm{E}_t(R_{t+1} - R_f)}{2\,\mathrm{Var}_t(R_{t+1}) + 2\,\mathrm{Cov}_t(R_{t+1}, R_{t+2})}.$$

With mean reversion in prices the covariance is negative, so the weight on the risky asset is larger for the two period horizon than for the one period horizon.

2.2 Empirical U.S. Evidence on Stock Return Predictability

Reference: Bodie, Kane, and Marcus (1999) 12; Cuthbertson (1996) 6.1; and Campbell, Lo, and MacKinlay (1997) 2 and 7; Cochrane (2001) 20.1

The two most common methods for investigating the predictability of stock returns are to calculate autocorrelations and to construct simple dynamic portfolios and see if they outperform passive portfolios. The dynamic portfolio could, for instance, be a simple filter rule that calls for rebalancing once a month by buying (selling) assets which have increased (decreased) more than $x\%$ the last month. If this portfolio outperforms a passive portfolio, then this is evidence of some positive autocorrelation ("momentum") on a one-month horizon. The following points summarize some evidence which seems to hold for both returns and returns in excess of a riskfree rate (an interest rate).

1. The empirical evidence suggests some, but weak, positive autocorrelation in *short horizon* returns (one day up to a month) — probably too little to be able to trade on. The autocorrelation is stronger for small than for large firms (perhaps no autocorrelation at all for weekly or longer returns in large firms). This implies that

equally weighted stock indices have larger autocorrelation than value-weighted indices. (See Campbell, Lo, and MacKinlay (1997) Table 2.4.)

2. Stock indices have more positive autocorrelation than (most of) the individual stocks: there must be fairly strong cross-autocorrelations across individual stocks. (See Campbell, Lo, and MacKinlay (1997) Tables 2.7 and 2.8.)

3. There seems to be negative autocorrelation for *multi-year* stock returns, for instance in 5-year US returns for 1926-1985. It is unclear what drives this result, however. It could well be an artifact of just a few extreme episodes (Great Depression). Moreover, the estimates are very uncertain as there are very few (non-overlapping) multi-year returns even in a long sample—the results could be just a fluke.

4. The aggregate stock market returns, that is, the returns on a value-weighted stock index, seem to be forecastable on the medium horizon by various *information variables*. In particular, future stock returns seems to be predictable by the current dividend-price ratio and earnings-price ratios (positively, one to several years), or by the interest rate changes (negatively, up to a year). For instance, the coefficient of determination (usually denoted R^2, but should not to be confused with the return used above) for predicted two-year returns on the US stock market on current dividend-price ratio is around 0.3 for the 1952-1994 sample. (See Campbell, Lo, and MacKinlay (1997) Tables 7.1-2.) This evidence suggests that expected returns may very well be time varying and correlated with the business cycle.

5. Even if short-run returns, R_{t+1}, are fairly hard to forecast, it is often fairly easy to forecast volatility as measured by $|R_{t+1}|$ or R_{t+1}^2 (for instance, using ARCH or GARCH models). For an example, see Bodie, Kane, and Marcus (1999) Figure 13.7. This could possibly be used for dynamic trading strategies on options which directly price volatility. For instance, buying both a call and a put option (a "straddle" or a "strangle"), is a bet on a large price movement (in any direction).

6. It is sometimes found that stock prices behave differently in periods with high volatility than in more normal periods. Granger (1992) reports that the forecasting performance is sometimes improved by using different forecasting models for these two regimes. A simple and straightforward way to estimate a model for peri-

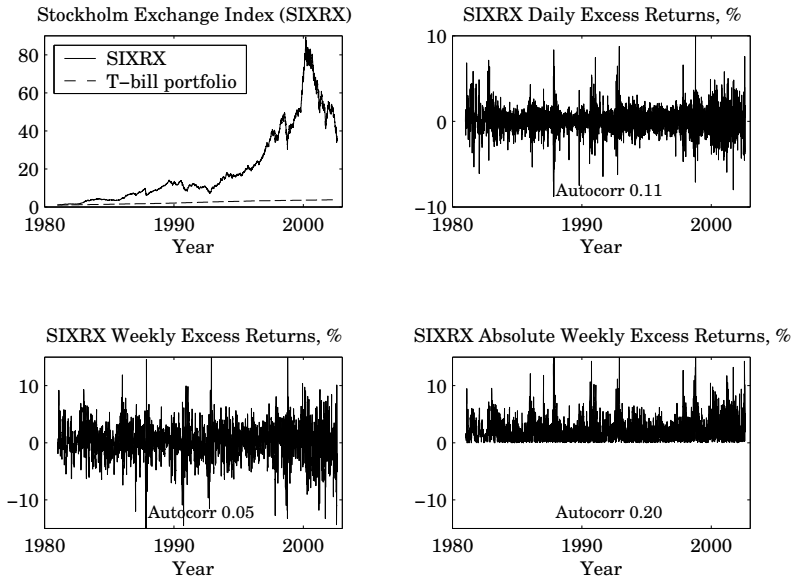


Figure 2.1: Value weighted stock index, Stockholm

ods of normal volatility is to simply throw out data for volatile periods (and other exceptional events).

7. It is important to assess forecasting models in terms of their out-of-sample forecast performance. Too many models seems to fit data in-sample, but most of them fail in out-of-sample tests. Forecasting models are of no use if they cannot forecast.

8. There are also a number of strange patterns ("anomalies") like the small-firms-in-January effect (high returns on these in the first part of January) and the book-to-market (high returns on firms with high book/market value of the firm's equity).

For some simple Swedish evidence, see Figure 2.1.

2.3 Prices, Returns, and Predictability

This section sets up a few simple time series models of asset prices in order to illustrate what they imply for returns autocorrelations.

Let the gross return of an asset be $R_{t+1} = P_{t+1}/P_t$, where we interpret the prices as including any dividends (so P_{t+1} equals the "raw" price plus any dividends in $t+1$). Take logs to get the continuously compounded return $r_{t+1} = p_{t+1} - p_t$.

Suppose that p_t is a *random walk without drift*, $p_t = p_{t-1} + \varepsilon_t$. Then, the first difference, that is, the return r_t, is white noise, ε_t, and therefore unpredictable.

Now, assume instead that p_t is a *random walk with drift*, $p_t = \mu + p_{t-1} + \varepsilon_t$. Then, the return is a constant plus white noise, $\mu + \varepsilon_t$, so movements of the returns around the mean are unpredictable.

We now let p_t be the sum of a random walk and a temporary component (with perfectly correlated innovations)

$$p_t = u_t + \theta\varepsilon_t, \text{ where } u_t = u_{t-1} + \varepsilon_t \tag{2.9}$$

$$= u_{t-1} + (1+\theta)\varepsilon_t \tag{2.10}$$

Note that it is the same white noise process, ε_t, that drives both the random walk and the temporary component. We can clearly write p_{t+1} as

$$p_{t+1} = u_{t+1} + \theta\varepsilon_{t+1} \tag{2.11}$$

$$= u_t + \varepsilon_{t+1} + \theta\varepsilon_{t+1}$$

$$= u_{t-1} + \varepsilon_t + (1+\theta)\varepsilon_{t+1} \tag{2.12}$$

The responses of p_t and p_{t+1} to ε_t are therefore $(1+\theta)\varepsilon_t$ and ε_t, respectively—see *Figure 2.2*.

From (2.10) and (2.12), the return is

$$r_{t+1} = p_{t+1} - p_t = (1+\theta)\varepsilon_{t+1} - \theta\varepsilon_t, \text{ so} \tag{2.13}$$

$$r_t = (1+\theta)\varepsilon_t - \theta\varepsilon_{t-1}. \tag{2.14}$$

The responses of r_t and r_{t+1} to ε_t are therefore $(1+\theta)\varepsilon_t$ and $-\theta\varepsilon_t$, respectively—see

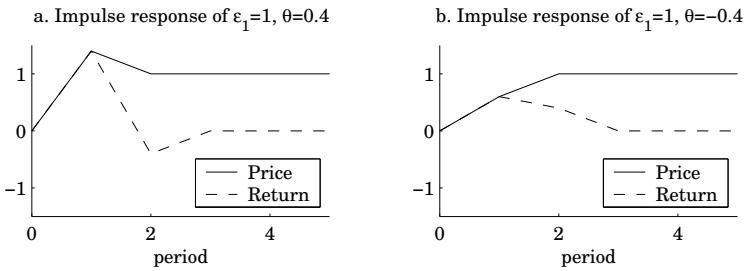


Figure 2.2: Impulse reponses when price is random walk plus temporary component, (2.9)

Figure 2.2. Since ε_t is white noise, the autocovariance of returns is

$$\text{Cov}(r_{t+1}, r_t) = -\theta(1+\theta)\,\text{Var}(\varepsilon_t), \tag{2.15}$$

but higher-order autocovariance, for instance, $\text{Cov}(r_{t+2}, r_t)$, are all zero.

If the price is a random walk, $\theta = 0$, then the autocovariance is zero. If the initial effect is positive and larger than the long run effect (partial mean reversion in the price level), $\theta > 0$, then the returns are negatively autocorrelated. The intuition is that news in t drive up p_t more than p_{t+1} (some mean reversion in the price level), so returns are above normal in the first period but negative in the second period (see Figure 2.2.a). One possible interpretation of this would be a positive piece of news that drive up today's price, but also makes the asset less risky and therefore gives lower (required) expected future returns. In contrast, if the initial price effect is positive but lower than the long run effect, $-1 < \theta < 0$, then the autocorrelation of returns is positive (see Figure 2.2.b).

2.4 Autocorrelations

Reference: Campbell, Lo, and MacKinlay (1997) 2

2.4.1 Autocorrelation Coefficients and the Box-Pierce Test

The autocovariances of the y_t process can be estimated as

$$\hat{\gamma}_s = \frac{1}{T}\sum_{t=1+s}^{T}(y_t - \bar{y})(y_{t-s} - \bar{y})', \tag{2.16}$$

$$\text{with } \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t. \tag{2.17}$$

(We typically divide by T in even if we have only $T - s$ full observations to estimate γ_s from.) Autocorrelations are then estimated as

$$\hat{\rho}_s = \hat{\gamma}_s/\hat{\gamma}_0.$$

The sampling properties of $\hat{\rho}_s$ are complicated, but there are several useful large sample results for Gaussian processes (these results typically carry over to processes which are similar to the Gaussian—a homoskedastic process with finite 6th moment is typically enough, see Priestley (1981) 5.3 or Brockwell and Davis (1991) 7.2-7.3). When the true autocorrelations are all zero (not ρ_0, of course), then for any i and j different from zero

$$\sqrt{T}\begin{bmatrix}\hat{\rho}_i\\ \hat{\rho}_j\end{bmatrix} \to^d N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\right). \tag{2.18}$$

If instead, the process is an MA(q), so $\rho_1, ..., \rho_q$ may be non-zero, but $\rho_k = 0$ for $k > q$, then it can be shown that

$$\sqrt{T}\hat{\rho}_k \to^d N\left(0, \textstyle\sum_{s=-q}^{q}\rho_s^2\right) \text{ for } k > q. \tag{2.19}$$

These results can be used to construct tests for both single autocorrelations (t-test or χ^2 test) and several autocorrelations at once (χ^2 test).

Example 12 *(t-test) We want to test the hypothesis that $\rho_1 = 0$. Since the $N(0,1)$ distribution has 5% of the probability mass below -1.65 and another 5% above 1.65, we can reject the null hypothesis at the 10% level if $\sqrt{T}|\hat{\rho}_1| > 1.65$. With $T = 100$, we therefore need $|\hat{\rho}_1| > 1.65/\sqrt{100} = 0.165$ for rejection, and with $T = 1000$ we need $|\hat{\rho}_1| > 1.65/\sqrt{1000} \approx 0.053$.*

The *Box-Pierce test* follows directly from the result in (2.18), since it shows that $\sqrt{T}\hat{\rho}_i$

and $\sqrt{T}\hat{\rho}_j$ are iid N(0,1) variables. Therefore, the sum of the square of them is distributed as an χ^2 variable. The test statistic typically used is

$$Q_L = T\sum_{s=1}^{L}\hat{\rho}_s^2 \to^d \chi_L^2. \tag{2.20}$$

Example 13 *(Box-Pierce) Let $\hat{\rho}_1 = 0.165$, and $T = 100$, so $Q_1 = 100 \times 0.165^2 = 2.72$. The 10% critical value of the χ_1^2 distribution is 2.71, so the null hypothesis of no autocorrelation is rejected.*

The choice of lag order in (2.20), L, should be guided by theoretical considerations, but it may also be wise to try different values. There is clearly a trade off: too few lags may miss a significant high-order autocorrelation, but too many lags can destroy the power of the test (as the test statistic is not affected much by increasing L, but the critical values increase).

The main problem with these tests is that the assumptions behind the results in (2.18) may not be reasonable. For instance, data may be heteroskedastic. One way of handling this is to make use of the GMM framework.

2.4.2 Box-Pierce as an Application of GMM*

This section discusses how GMM can be used to test if a series is autocorrelated. The analysis focuses on first-order autocorrelation, but it is straightforward to extend it to higher-order autocorrelation.

Consider a scalar random variable x_t with a zero mean (it is easy to extend the analysis to allow for a non-zero mean). Consider the moment conditions

$$m_t(\beta) = \begin{bmatrix} x_t^2 - \sigma^2 \\ x_t x_{t-1} - \rho\sigma^2 \end{bmatrix}, \text{ so } \bar{m}(\beta) = \frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} x_t^2 - \sigma^2 \\ x_t x_{t-1} - \rho\sigma^2 \end{bmatrix}, \text{ with } \beta = \begin{bmatrix} \sigma^2 \\ \rho \end{bmatrix}. \tag{2.21}$$

σ^2 is the variance and ρ the first-order autocorrelation so $\rho\sigma^2$ is the first-order autocovariance. We want to test if $\rho = 0$. We could proceed along two different routes: estimate ρ and test if it is different from zero or set ρ to zero and then test overidentifying restrictions. We analyze how the first of these two approaches works when the null hypothesis of $\rho = 0$ is true.

We estimate both σ^2 and ρ by using the moment conditions (2.21) and then test if $\rho = 0$. To do that we need to calculate the asymptotic variance of $\hat{\rho}$ (there is little hope of being able to calculate the small sample variance, so we have to settle for the asymptotic variance as an approximation).

We have an exactly identified system so the weight matrix does not matter—we can then proceed as if we had used the optimal weighting matrix (all those results apply).

To find the asymptotic covariance matrix of the parameters estimators, we need the probability limit of the Jacobian of the moments and the covariance matrix of the moments—evaluated at the true parameter values. Let $\bar{m}_i(\beta_0)$ denote the ith element of the $\bar{m}(\beta)$ vector—evaluated at the true parameter values. The probability of the Jacobian is

$$D_0 = \operatorname{plim}\begin{bmatrix} \partial\bar{m}_1(\beta_0)/\partial\sigma^2 & \partial\bar{m}_1(\beta_0)/\partial\rho \\ \partial\bar{m}_2(\beta_0)/\partial\sigma^2 & \partial\bar{m}_2(\beta_0)/\partial\rho \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ -\rho & -\sigma^2 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -\sigma^2 \end{bmatrix}, \tag{2.22}$$

since $\rho = 0$ (the true value). Note that we differentiate with respect to σ^2, not σ, since we treat σ^2 as a parameter.

The covariance matrix is more complicated. The definition is

$$S_0 = \mathrm{E}\left[\frac{\sqrt{T}}{T}\sum_{t=1}^{T} m_t(\beta_0)\right]\left[\frac{\sqrt{T}}{T}\sum_{t=1}^{T} m_t(\beta_0)\right]'.$$

Assume that there is no autocorrelation in $m_t(\beta_0)$ (which means, among other things, that volatility, x_t^2, is not autocorrelated). We can then simplify as

$$S_0 = \mathrm{E}\, m_t(\beta_0) m_t(\beta_0)'.$$

This assumption is stronger than assuming that $\rho = 0$, but we make it here in order to illustrate the asymptotic distribution. To get anywhere, we assume that x_t is iid $N(0, \sigma^2)$. In this case (and with $\rho = 0$ imposed) we get

$$\begin{aligned} S_0 &= \mathrm{E}\begin{bmatrix} x_t^2 - \sigma^2 \\ x_t x_{t-1} \end{bmatrix}\begin{bmatrix} x_t^2 - \sigma^2 \\ x_t x_{t-1} \end{bmatrix}' = \mathrm{E}\begin{bmatrix} (x_t^2 - \sigma^2)^2 & (x_t^2 - \sigma^2)x_t x_{t-1} \\ (x_t^2 - \sigma^2)x_t x_{t-1} & (x_t x_{t-1})^2 \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{E}\, x_t^4 - 2\sigma^2\mathrm{E}\, x_t^2 + \sigma^4 & 0 \\ 0 & \mathrm{E}\, x_t^2 x_{t-1}^2 \end{bmatrix} = \begin{bmatrix} 2\sigma^4 & 0 \\ 0 & \sigma^4 \end{bmatrix}. \end{aligned} \tag{2.23}$$

To make the simplification in the second line we use the facts that $\mathrm{E}\, x_t^4 = 3\sigma^4$ if $x_t \sim$

$N(0,\sigma^2)$, and that the normality and the iid properties of x_t together imply $\mathrm{E}\, x_t^2 x_{t-1}^2 = \mathrm{E}\, x_t^2\, \mathrm{E}\, x_{t-1}^2$ and $\mathrm{E}\, x_t^3 x_{t-1} = \mathrm{E}\, \sigma^2 x_t x_{t-1} = 0$.

By combining (2.22) and (2.23) we get that

$$\begin{aligned}\mathrm{ACov}\left(\sqrt{T}\begin{bmatrix}\hat{\sigma}^2\\ \hat{\rho}\end{bmatrix}\right) &= \left(D_0' S_0^{-1} D_0\right)^{-1}\\ &= \left(\begin{bmatrix}-1 & 0\\ 0 & -\sigma^2\end{bmatrix}'\begin{bmatrix}2\sigma^4 & 0\\ 0 & \sigma^4\end{bmatrix}^{-1}\begin{bmatrix}-1 & 0\\ 0 & -\sigma^2\end{bmatrix}\right)^{-1}\\ &= \begin{bmatrix}2\sigma^4 & 0\\ 0 & 1\end{bmatrix}. \end{aligned} \tag{2.24}$$

This shows the standard expression for the uncertainty of the variance and that the $\sqrt{T}\hat{\rho}$. Since GMM estimators typically have an asymptotic distribution we have $\sqrt{T}\hat{\rho} \to^d N(0,1)$, so we can test the null hypothesis of no first-order autocorrelation by the test statistic

$$T\hat{\rho}^2 \sim \chi_1^2. \tag{2.25}$$

This is the same as the *Box-Pierce test for first-order autocorrelation*.

This analysis shows that we are able to arrive at simple expressions for the sampling uncertainty of the variance and the autocorrelation—provided we are willing to make strong assumptions about the data generating process. In particular, we assumed that data was iid $N(0,\sigma^2)$. One of the strong points of GMM is that we could perform similar tests without making strong assumptions—provided we use a correct estimator of the asymptotic covariance matrix S_0 (for instance, Newey-West).

2.4.3 Autoregressions

An alternative way of testing autocorrelations is to estimate an AR model

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + ... + a_p y_{t-p} + \varepsilon_t, \tag{2.26}$$

and then test if all the coefficients are zero with a χ^2 test. This approach is somewhat less general than the Box-Pierce test, but most stationary time series processes can be well approximated by an AR of relatively low order. To account for heteroskedasticity and other problems, it can make sense to estimate the covariance matrix of the parameters by

an estimator like Newey-West.

2.4.4 Autoregressions versus Autocorrelations*

If is straightforward to see the relation between autocorrelations and the AR model when the AR model is the true process. This relation is given by the *Yule-Walker equations*.

For an AR(1), the autoregression coefficient is simply the first autocorrelation coefficient. For an AR(2), $y_t = a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t$, we have

$$\begin{aligned}\begin{bmatrix}\mathrm{Cov}(y_t, y_t)\\ \mathrm{Cov}(y_{t-1}, y_t)\\ \mathrm{Cov}(y_{t-2}, y_t)\end{bmatrix} &= \begin{bmatrix}\mathrm{Cov}(y_t, a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t)\\ \mathrm{Cov}(y_{t-1}, a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t)\\ \mathrm{Cov}(y_{t-2}, a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t)\end{bmatrix}\\ &= \begin{bmatrix}a_1\,\mathrm{Cov}(y_t, y_{t-1}) + a_2\,\mathrm{Cov}(y_t, y_{t-2}) + \mathrm{Cov}(y_t, \varepsilon_t)\\ a_1\,\mathrm{Cov}(y_{t-1}, y_{t-1}) + a_2\,\mathrm{Cov}(y_{t-1}, y_{t-2})\\ a_1\,\mathrm{Cov}(y_{t-2}, y_{t-1}) + a_2\,\mathrm{Cov}(y_{t-2}, y_{t-2})\end{bmatrix}, \text{ or}\\ \begin{bmatrix}\gamma_0\\ \gamma_1\\ \gamma_2\end{bmatrix} &= \begin{bmatrix}a_1\gamma_1 + a_2\gamma_2 + \mathrm{Var}(\varepsilon_t)\\ a_1\gamma_0 + a_2\gamma_1\\ a_1\gamma_1 + a_2\gamma_0\end{bmatrix}. \end{aligned} \tag{2.27}$$

To transform to autocorrelation, divide through by γ_0. The last two equations are then

$$\begin{bmatrix}\rho_1\\ \rho_2\end{bmatrix} = \begin{bmatrix}a_1 + a_2\rho_1\\ a_1\rho_1 + a_2\end{bmatrix} \text{ or } \begin{bmatrix}\rho_1\\ \rho_2\end{bmatrix} = \begin{bmatrix}a_1/(1-a_2)\\ a_1^2/(1-a_2) + a_2\end{bmatrix}. \tag{2.28}$$

If we know the parameters of the AR(2) model (a_1, a_2, and $\mathrm{Var}(\varepsilon_t)$), then we can solve for the autocorrelations. Alternatively, if we know the autocorrelations, then we can solve for the autoregression coefficients. This demonstrates that testing that all the autocorrelations are zero is essentially the same as testing if all the autoregressive coefficients are zero. Note, however, that the transformation is non-linear, which may make a difference in small samples.

2.4.5 Variance Ratios

The 2-period variance ratio is the ratio of $\mathrm{Var}(y_t + y_{t-1})$ to $2\,\mathrm{Var}(y_t)$. To simplify notation, let y_t have a zero mean (or be demeaned), so $\mathrm{Cov}(y_t, y_{t-s}) = \mathrm{E}\, y_t y_{t-s}$, with short hand

notation γ_s. We can then write the variance ratio as

$$VR_2 = \frac{\mathrm{E}(y_t + y_{t-1})^2}{2\,\mathrm{E}\,y_t^2} \tag{2.29}$$

$$= \frac{\mathrm{E}\,y_t^2 + \mathrm{E}\,y_{t-1}^2 + \mathrm{E}\,y_t y_{t-1} + \mathrm{E}\,y_{t-1} y_t}{2\,\mathrm{E}\,y_t^2}$$

$$= \frac{\gamma_0 + \gamma_0 + \gamma_1 + \gamma_{-1}}{2\gamma_0}. \tag{2.30}$$

Let $\rho_s = \gamma_s/\gamma_0$ be the sth autocorrelation. We can then write (2.30) as

$$= \frac{1}{2}(\rho_{-1} + 2 + \rho_1) \tag{2.31}$$

$$= 1 + \rho_1, \tag{2.32}$$

where we exploit the fact that the autocorrelation (and autocovariance) function is symmetric around zero, so $\rho_{-s} = \rho_s$ (and $\gamma_{-s} = \gamma_s$). Note, however, that this does not hold for cross autocorrelations ($\mathrm{Cov}(x_t, y_{t-s}) \neq \mathrm{Cov}(x_{t-s}, y_t)$).

It is clear from (2.32) that if y_t is not serially correlated, then the variance ratio is unity; a value above one indicates positive serial correlation and a value below one indicates negative serial correlation.

We can also consider longer variance rations, where we sum q observations in the numerator and then divide by $q\,\mathrm{Var}(y_t)$. To see the pattern of what happens, consider $q = 3$ and assume that y_t has a zero mean so we can use second moments for covariances. In that case, the numerator $\mathrm{Var}(y_t + y_{t-1} + y_{t-2}) = \mathrm{E}(y_t + y_{t-1} + y_{t-2})(y_t + y_{t-1} + y_{t-2})$ is

$$\underbrace{\left(\mathrm{E}\,y_t^2 + \mathrm{E}\,y_t y_{t-1} + \mathrm{E}\,y_t y_{t-2}\right)}_{\gamma_0+\gamma_1+\gamma_2} + \underbrace{\left(\mathrm{E}\,y_{t-1} y_t + \mathrm{E}\,y_{t-1}^2 + \mathrm{E}\,y_{t-1} y_{t-2}\right)}_{\gamma_{-1}+\gamma_0+\gamma_1} + \underbrace{\left(\mathrm{E}\,y_{t-2} y_t + \mathrm{E}\,y_{t-2} y_{t-1} + \mathrm{E}\,y_{t-2}^2\right)}_{\gamma_{-2}+\gamma_{-1}+\gamma_0}$$

$$= \gamma_{-2} + 2\gamma_{-1} + 3\gamma_0 + 2\gamma_1 + \gamma_2. \tag{2.33}$$

It can be noted that the (numerator of the) variance ratios are fairly similar to the uncertainty of a sample average.

The variance ratio is therefore

$$VR_3 = \frac{1}{3\gamma_0}(\gamma_{-2} + 2\gamma_{-1} + 3\gamma_0 + 2\gamma_1 + \gamma_2)$$

$$= \frac{1}{3}(\rho_{-2} + 2\rho_{-1} + 3 + 2\rho_1 + \rho_2). \tag{2.34}$$

Comparing VR_2 in (2.31) and VR_3 in (2.34) suggests that VR_q is a similar weighted average of autocorrelations, where the weights are tent-shaped (1 at lag 0, $1/q$ at lag $q-1$, and zero at lag q and beyond). In fact, it can be shown that we have

$$VR_q = \frac{\mathrm{Var}\left(\sum_{s=0}^{q-1} y_{t-s}\right)}{q\,\mathrm{Var}(y_t)} \tag{2.35}$$

$$= \sum_{s=-(q-1)}^{q-1} \left(1 - \frac{|s|}{q}\right)\rho_s \text{ or} \tag{2.36}$$

$$= 1 + 2\sum_{s=1}^{q-1}\left(1 - \frac{s}{q}\right)\rho_s. \tag{2.37}$$

Note that we could equally well let the summation in (2.36) run from $-q$ to q since the weight $1 - |s|/q$, is zero for that lag. The same holds for (2.37).

It is immediate that no autocorrelation means that $VR_q = 1$ for all q. If all autocorrelations are non-positive, $\rho_s \leq 0$, then $VR_q \leq 1$, and vice versa.

Example 14 *(Variance ratio of an AR(1)) When $y_t = ay_{t-1} + \varepsilon_t$ where ε_t is iid white noise, then*

$$VR_2 = \frac{1}{2}a + 1 + \frac{1}{2}a = 1 + a \text{ and}$$
$$VR_3 = \frac{1}{3}a^2 + \frac{2}{3}a + 1 + \frac{2}{3}a + \frac{1}{3}a^2 = 1 + \frac{4}{3}a + \frac{2}{3}a^2$$
$$VR_4 = \frac{1}{4}a^3 + \frac{2}{4}a^2 + \frac{3}{4}a + 1 + \frac{3}{4}a + \frac{2}{4}a^2 + \frac{1}{4}a^3 = 1 + \frac{3}{2}a + a^2 + \frac{1}{2}a^3.$$

See Figure 2.3 *for a numerical example.*

The estimation of VR_q is done by replacing the population variances in (2.35) with the sample variances, or the autocorrelations in (2.37) by the sample autocorrelations.

The sampling distribution of $\widehat{VR}_q$ under the null hypothesis that there is no autocorrelation follows directly from the sampling distribution of the autocorrelation coefficient.

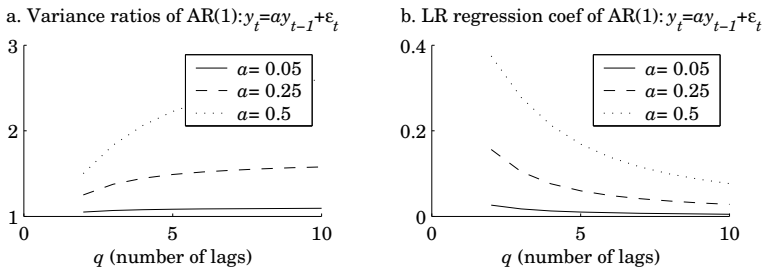


Figure 2.3: Variance ratio and long run autocorrelation of AR(1)

For instance, if the assumptions behind (2.18) are satisfied, then we have that

$$\sqrt{T}\left(\widehat{VR}_q - 1\right) = 2\sum_{s=1}^{q-1}\left(1 - \frac{s}{q}\right)\sqrt{T}\hat{\rho}_s \tag{2.38}$$

is a linear combination of (asymptotically) uncorrelated $N(0,1)$ variables (the $\sqrt{T}\hat{\rho}_s$). It follows that

$$\sqrt{T}\left(\widehat{VR}_q - 1\right) \to^d N\left[0, \sum_{s=1}^{q-1} 4\left(1 - \frac{s}{q}\right)^2\right]. \tag{2.39}$$

For instance, we have

$$\sqrt{T}\left(\widehat{VR}_2 - 1\right) \to^d N(0,1) \text{ and } \sqrt{T}\left(\widehat{VR}_3 - 1\right) \to^d N(0, 20/9). \tag{2.40}$$

These distributional results depend on the assumptions behind the results in (2.18). One way of handling deviations from those assumptions is to estimate the autocorrelations and their covariance matrix with GMM.

The results in CLM Table 2.5 and 2.6 (weekly CRSP stock index returns, early 1960s to mid 1990s) show variance ratios above one and increasing with the number of lags, q. The results for individual stocks in CLM Table 2.7 show variance ratios close to, or even below, unity. Cochrane 20 Table 20.5–6 report weak evidence for more mean reversion in multi-year returns (annual NYSE stock index,1926 to mid 1990s).

2.4.6 Long-Run Autoregressions

Consider an AR(1) of two-period sums (returns)

$$y_{t+1} + y_{t+2} = a + b_2(y_{t-1} + y_t) + \varepsilon_{t+2}. \tag{2.41}$$

The least squares population regression coefficient is

$$b_2 = \frac{\text{Cov}(y_{t+1} + y_{t+2}, y_{t-1} + y_t)}{\text{Var}(y_{t-1} + y_t)}. \tag{2.42}$$

The numerator can be written

$$\begin{aligned}\text{Cov}(y_{t+1} + y_{t+2}, y_{t-1} + y_t) &= \underbrace{\text{Cov}(y_{t+1}, y_{t-1} + y_t)}_{\gamma_2+\gamma_1} + \underbrace{\text{Cov}(y_{t+2}, y_{t-1} + y_t)}_{\gamma_3+\gamma_2} \\ &= \gamma_1 + 2\gamma_2 + \gamma_3.\end{aligned} \tag{2.43}$$

We can therefore write (2.42) as

$$\begin{aligned}b_2 &= \frac{\gamma_1 + 2\gamma_2 + \gamma_3}{\text{Var}(y_{t-1} + y_t)} \\ &= \frac{2\,\text{Var}(y_{t+1})}{\text{Var}(y_{t-1} + y_t)} \frac{\gamma_1 + 2\gamma_2 + \gamma_3}{2\,\text{Var}(y_{t+1})} \\ &= \frac{1}{VR_2}\frac{\rho_1 + 2\rho_2 + \rho_3}{2}.\end{aligned} \tag{2.44}$$

Let us now try an AR(1) for three-period sums instead

$$y_{t+1} + y_{t+2} + y_{t+3} = a + b_3(y_{t-2} + y_{t-1} + y_t) + \varepsilon_{t+3}. \tag{2.45}$$

The numerator in the regression coefficient is then

$$\text{Cov}(y_{t+1} + y_{t+2} + y_{t+3}, y_{t-2} + y_{t-1} + y_t) = (\gamma_3 + \gamma_2 + \gamma_1) + (\gamma_4 + \gamma_3 + \gamma_2) + (\gamma_5 + \gamma_4 + \gamma 3) \tag{2.46}$$

$$= \gamma_1 + 2\gamma_2 + 3\gamma_3 + 2\gamma_4 + \gamma_5.$$

The regression coefficient is therefore

$$b_3 = \frac{\gamma_1 + 2\gamma_2 + 3\gamma_3 + 2\gamma_4 + \gamma_5}{\text{Var}\left(y_{t+1} + y_{t+2} + y_{t+3}\right)}$$
$$= \frac{1}{VR_3}\frac{\rho_1 + 2\rho_2 + 3\rho_3 + 2\rho_4 + \rho_5}{3}. \tag{2.47}$$

The general pattern that emerges from these expressions is that the slope coefficient in the model

$$\sum_{s=1}^{q} y_{t+s} = a + b_q \sum_{s=1}^{q} y_{t+s-q} + \varepsilon_{t+q} \tag{2.48}$$

is

$$b_q = \frac{1}{VR_q}\sum_{s=-(q-1)}^{q-1}\left(1 - \frac{|s|}{q}\right)\rho_{q+s}. \tag{2.49}$$

Note that the autocorrelations are displaced by the amount q. As for the variance ratio, the summation could run from $-q$ to q instead, since the weight, $1 - |s|/q$, is zero for that lag.

Equation (2.49) shows that the variance ratio and the AR(1) coefficient of long-run returns are closely related. A bit of manipulation (and using the fact that $\rho_{-s} = \rho_s$) shows that

$$1 + b_q = \frac{VR_{2q}}{VR_q}. \tag{2.50}$$

If the variance ratio increases with the horizon, then this means that the long-run returns are positively autocorrelated.

Example 15 *(Long-run autoregression of an AR(1)) When $y_t = ay_{t-1} + \varepsilon_t$ where ε_t is iid white noise, then the variance ratios are as in Example (14), and we know that $\rho_{q+s} = a^{q+s}$. From (2.44) we then have*

$$b_2 = \frac{1}{VR_2}\frac{a + 2a^2 + a^3}{2}$$
$$= \frac{1}{1+a}\frac{a + 2a^2 + a^3}{2}.$$

See Figure 2.3 *for a numerical example. For future reference, note that we can simplify to get $b_2 = (1+a)\,a/2$.*

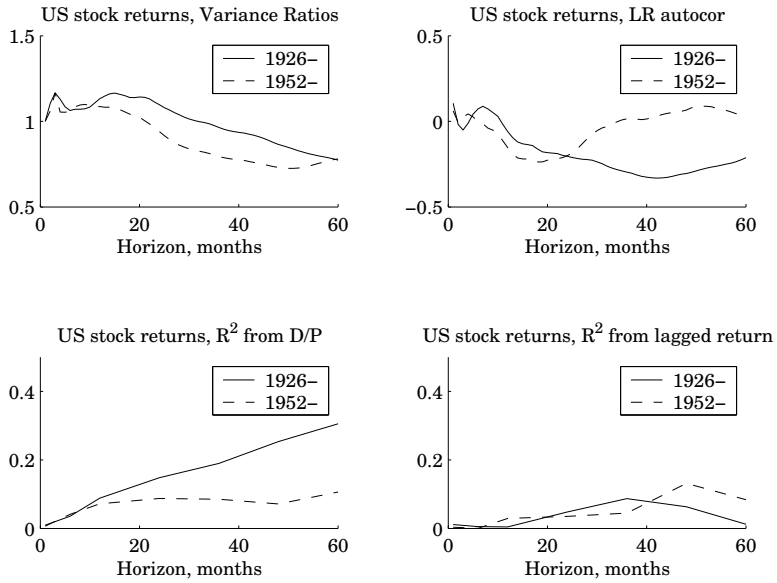


Figure 2.4: US stock returns in excess of T-bill rate, 1926:07-2001:12

Example 16 *(Trying (2.50) on an AR(1)) From Example (14) we have that*

$$\frac{VR_4}{VR_2} - 1 = \frac{1 + \frac{3}{2}a + a^2 + \frac{1}{2}a^3}{1+a} - 1$$
$$= \frac{1}{2}(1+a)\,a,$$

which is b_2 in Example 15.

2.5 Other Predictors

There are many other, perhaps more economically plausible, possible predictors of future stock returns. For instance, both the dividend-price ratio and nominal interest rates have

been used to predict long-run returns, and lagged short-run returns on other assets have been used to predict short-run returns.

2.5.1 Lead-Lags

Reference: Campbell, Lo, and MacKinlay (1997) 2

Stock indices have more positive autocorrelation than (most) the individual stocks: there should therefore be fairly strong cross-autocorrelations across individual stocks. (See Campbell, Lo, and MacKinlay (1997) Tables 2.7 and 2.8.) Indeed, this is also what is found in US data where weekly returns of large size stocks forecast weekly returns of small size stocks.

2.5.2 Prices and Dividends: Accounting Identity

Reference: Campbell, Lo, and MacKinlay (1997) 7 and Cochrane (2001) 20.1.

The gross return, R_{t+1}, is defined as

$$R_{t+1} = \frac{D_{t+1} + P_{t+1}}{P_t}, \text{ so } P_t = \frac{D_{t+1} + P_{t+1}}{R_{t+1}}. \tag{2.51}$$

Substituting for P_{t+1} (and then P_{t+2}, ...) gives

$$P_t = \frac{D_{t+1}}{R_{t+1}} + \frac{D_{t+2}}{R_{t+1}R_{t+2}} + \frac{D_{t+3}}{R_{t+1}R_{t+2}R_{t+3}} + \dots \tag{2.52}$$

$$= \sum_{j=1}^{\infty} \frac{D_{t+j}}{\prod_{k=1}^{j} R_{t+k}}, \tag{2.53}$$

provided the discounted value of P_{t+j} goes to zero as $j \to \infty$. This is simply an accounting identity. It is clear that a high price in t must lead to low future returns and/or high future dividends—which (by rational expectations) also carry over to expectations of future returns and dividends.

It is sometimes more convenient to analyze the price-dividend ratio. Dividing (2.52) and (2.53) by D_t gives

$$\frac{P_t}{D_t} = \frac{1}{R_{t+1}}\frac{D_{t+1}}{D_t} + \frac{1}{R_{t+1}R_{t+2}}\frac{D_{t+2}}{D_{t+1}}\frac{D_{t+1}}{D_t} + \frac{1}{R_{t+1}R_{t+2}R_{t+3}}\frac{D_{t+3}}{D_{t+2}}\frac{D_{t+2}}{D_{t+1}}\frac{D_{t+1}}{D_t} + \dots \tag{2.54}$$

$$= \sum_{j=1}^{\infty}\prod_{k=1}^{j}\frac{D_{t+k}/D_{t+k-1}}{R_{t+k}}. \tag{2.55}$$

As with (2.53) it is just an accounting identity. It must therefore also hold in expectations. Since expectations are good (the best?) predictors of future values, we have the implication that the asset price should predict a discounted sum of future dividends, (2.53), and that the price-dividend ratio should predict a discounted sum of future changes in dividends.

2.5.3 Prices and Dividends: Approximating the Accounting Identity

We now log-linearize the accounting identity (2.55) in order to tie it more closely to the (typically linear) econometrics methods for detecting forecastability.

Rewrite (2.51) as

$$R_{t+1} = \frac{D_{t+1} + P_{t+1}}{P_t} = \frac{P_{t+1}}{P_t}\left(1 + \frac{D_{t+1}}{P_{t+1}}\right) \text{ or in logs}$$

$$r_{t+1} = p_{t+1} - p_t + \ln\left[1 + \exp(d_{t+1} - p_{t+1})\right]. \tag{2.56}$$

Make a first order Taylor approximation of the last term around a steady state value of $d_{t+1} - p_{t+1}$, denoted $\overline{d-p}$,

$$\ln\left[1 + \exp(d_{t+1} - p_{t+1})\right] \approx \ln\left[1 + \exp(\overline{d-p})\right] + \frac{\exp(\overline{d-p})}{1 + \exp(\overline{d-p})}\left[d_{t+1} - p_{t+1} - (\overline{d-p})\right]$$

$$\approx \text{ constant} + (1-\rho)\left(d_{t+1} - p_{t+1}\right), \tag{2.57}$$

where $\rho = 1/[1 + \exp(\overline{d-p})] = 1/[1 + \overline{D/P})]$. If the average dividend-price ratio is 4%, then $\rho = 1/1.04 \approx 0.96$.

We use (2.57) in (2.56) and forget about the constant. The result is

$$\begin{aligned} r_{t+1} &\approx p_{t+1} - p_t + (1-\rho)(d_{t+1} - p_{t+1}) \\ &= \rho p_{t+1} - p_t + (1-\rho) d_{t+1}, \end{aligned} \tag{2.58}$$

where $0 < \rho < 1$.

Add and subtract d_t from the right hand side and rearrange

$$r_{t+1} \approx \rho(p_{t+1} - d_{t+1}) - (p_t - d_t) + (d_{t+1} - d_t), \text{ or} \tag{2.59}$$

$$p_t - d_t \approx \rho(p_{t+1} - d_{t+1}) + (d_{t+1} - d_t) - r_{t+1} \tag{2.60}$$

This is a (forward looking, unstable) difference equation, which we can solve recursively forward. Provided $\lim_{s\to\infty} \rho^s (p_{t+s} - d_{t+s}) = 0$, the solution is

$$p_t - d_t \approx \sum_{s=0}^{\infty} \rho^s [(d_{t+1+s} - d_{t+s}) - r_{t+1+s}]. \tag{2.61}$$

(Trying to solve for the log price level instead of the log price-dividend ratio is problematic since the condition $\lim_{s\to\infty} \rho^s p_{t+s} = 0$ may not be satisfied.) As before, a high price-dividend ratio must imply future dividend growth and/or low future returns.

In the exact solution (2.54), dividends and returns which are closer to the present show up more times than dividends and returns far in the future. In the approximation (2.61), this is captured by giving a higher weight (higher ρ^s).

2.5.4 Dividend-Price Ratio as a Predictor

One of the most successful attempts to forecast long-run return is by using the dividend-price ratio

$$\sum_{s=1}^{q} r_{t+s} = \alpha + \beta_q (d_t - p_t) + \varepsilon_{t+q}. \tag{2.62}$$

For instance, CLM Table 7.1, report R^2 values from this regression which are close to zero for monthly returns, but they increase to 0.4 for 4-year returns (US, value weighted index, mid 1920s to mid 1990s).

By comparing with (2.61), we see that the dividend-ratio in (2.62) is only asked to predict a finite (unweighted) sum of future returns—dividend growth is disregarded. We

should therefore expect (2.62) to work particularly well if the horizon is long (high q) and if dividends are stable over time.

From (2.61) we get that

$$\text{Var}(p_t - d_t) \approx \text{Cov}\left(p_t - d_t, \sum_{s=0}^{\infty} \rho^s (d_{t+1+s} - d_{t+s})\right) - \text{Cov}\left(p_t - d_t, \sum_{s=0}^{\infty} \rho^s r_{t+1+s}\right), \tag{2.63}$$

which shows that the variance of the price-dividend ratio can be decomposed into the covariance of price-dividend ratio with future dividend change minus the covariance of price-dividend ratio with future returns. This expression highlights that if $p_t - d_t$ is not constant, then it must forecast dividend growth and/or returns.

The evidence in Cochrane suggests that $p_t - d_t$ does not forecast future dividend growth, so that forecastability of future returns explains the variability in the dividend-price ratio. This fits very well into the findings of the R^2 of (2.62). To see that, recall the following fact.

Remark 17 *(R^2 from a least squares regression) Let the least squares estimate of β in $y_t = x_t'\beta_0 + u_t$ be $\hat{\beta}$. The fitted values $\hat{y}_t = x_t'\hat{\beta}$. If the regression equation includes a constant, then $R^2 = \widehat{Corr}(y_t, \hat{y}_t)^2$. In a simple regression where $y_t = a + bx_t + u_t$, where x_t is a scalar, $R^2 = \widehat{Corr}(y_t, x_t)^2$.*

2.5.5 Predictability but No Autocorrelation

The evidence for US stock returns is that long-run returns may perhaps be predicted by using dividend-price ratio or interest rates, but that the long-run autocorrelations are weak (long run US stock returns appear to be "weak-form efficient" but not "semi-strong efficient"). Both CLM 7.1.4 and Cochrane 20.1 use small models for discussing this case. The key in these discussions is to make changes in dividends unforecastable, but let the return be forecastable by some state variable ($\text{E}_t\, d_{t+1+s} - \text{E}_t\, d_{t+s} = 0$ and $\text{E}_t\, r_{t+1} = r + x_t$), but in such a way that there is little autocorrelation in returns. By taking expectations of (2.61) we see that price-dividend will then reflect expected future returns and therefore be useful for forecasting.

2.6 Trading Strategies

Another way to measure predictability and to illustrate its economic importance is to calculate the return of a dynamic trading strategy, and then measure the "performance" of this strategy in relation to some benchmark portfolios. The trading strategy should, of course, be based on the variable that is supposed to forecast returns.

A common way (since Jensen, updated in Huberman and Kandel (1987)) is to study the performance of a portfolio by running the following regression

$$R_{1t} - R_{ft} = \alpha + \beta(R_{mt} - R_{ft}) + \varepsilon_t,\ \mathrm{E}\,\varepsilon_t = \mathbf{0} \text{ and } \mathrm{Cov}(R_{mt} - R_{ft}, \varepsilon_t) = \mathbf{0}, \quad (2.64)$$

where $R_{1t} - R_{ft}$ is the excess return on the portfolio being studied and $R_{mt} - R_{ft}$ the excess returns of a vector of benchmark portfolios (for instance, only the market portfolio if we want to rely on CAPM; returns times conditional information if we want to allow for time-variation in expected benchmark returns). Neutral performance (mean variance intersection) requires $\alpha = 0$, which can be tested with a t or F test.

2.7 Maximally Predictable Portfolio

As a way to calculate an upper bound on predictability, Lo and MacKinlay (1997) construct maximally predictable portfolios. The weights on the different assets in this portfolio can also help us to understand more about how the predictability works.

Let Z_t be an $n \times 1$ vector of demeaned returns

$$Z_t = R_t - \mathrm{E}\,R_t, \quad (2.65)$$

and suppose that we (somehow) have constructed rational forecasts $\mathrm{E}_{t-1}\,Z_t$ such that

$$Z_t = \mathrm{E}_{t-1}\,Z_t + \varepsilon_t, \text{ where } \mathrm{E}_{t-1}\,\varepsilon_t = 0,\ \mathrm{Var}_{t-1}(\varepsilon_t\varepsilon_t') = \Sigma. \quad (2.66)$$

Consider a portfolio $\gamma' Z_t$. The R^2 from predicting the return on this portfolio is (as usual) the fraction of the variability of $\gamma' Z_t$ that is explained by $\gamma'\,\mathrm{E}_{t-1}\,Z_t$

$$\begin{aligned} R^2(\gamma) &= 1 - \mathrm{Var}(\gamma'\varepsilon_t)/\,\mathrm{Var}(\gamma' Z_t) \\ &= [\mathrm{Var}(\gamma' Z_t) - \mathrm{Var}(\gamma'\varepsilon_t)]/\,\mathrm{Var}(\gamma' Z_t) \\ &= \mathrm{Var}(\gamma'\,\mathrm{E}_{t-1}\,Z_t)/\,\mathrm{Var}(\gamma' Z_t) \\ &= \gamma'\,\mathrm{Cov}(\mathrm{E}_{t-1}\,Z_t)\gamma/\gamma'\,\mathrm{Cov}(Z_t)\gamma. \end{aligned} \quad (2.67)$$

The covariance in the denominator can be calculated directly from data, but the covariance matrix in the numerator clearly depends on the forecasting model we use (to create $\mathrm{E}_{t-1}\,Z_t$).

The portfolio (γ vector) that gives the highest R^2 is the eigenvector (normalized to sum to unity) associated with the largest eigenvalue (also the value of R^2) of $\mathrm{Cov}(Z_t)^{-1}\,\mathrm{Cov}(\mathrm{E}_{t-1}\,Z_t)$.

Example 18 *(One forecasting variable) Suppose there is only one predictor, x_{t-1},*

$$Z_t = \beta x_{t-1} + \varepsilon_t,$$

where β is $n \times 1$. This means that $\mathrm{E}_{t-1}\,Z_t = \beta x_{t-1}$, so $\mathrm{Cov}(\mathrm{E}_{t-1}\,Z_t) = \mathrm{Var}(x_{t-1})\beta\beta'$ and that $\mathrm{Cov}(Z_t) = \mathrm{Var}(x_{t-1})\beta\beta' + \Sigma$. We can therefore write (2.67) as

$$R^2(\gamma) = \frac{\gamma'\,\mathrm{Var}(x_{t-1})\beta\beta'\gamma}{\gamma'\,\mathrm{Var}(x_{t-1})\beta\beta'\gamma + \gamma'\Sigma\gamma}.$$

The first order conditions for maximum then gives (this is very similar to the calculations of the minimum variance portfolio in mean-variance analysis)

$$\gamma = \Sigma^{-1}\beta/\mathbf{1}'\Sigma^{-1}\beta,$$

where $\mathbf{1}$ is an $n \times 1$ vector of ones. In particular, if Σ (and therefore Σ^{-1}) is diagonal, then the portfolio weight of asset i is β_i divided by the variance of the forecast error of asset i: assets which are hard to predict get smaller weights. We also see that if the sign of β_i is different from the sign of $\mathbf{1}'\Sigma^{-1}\beta$, then it gets a negative weight. For instance, if $\mathbf{1}'\Sigma^{-1}\beta > 0$, so that most assets move in the same direction as x_{t-1}, then asset i gets a negative weight if it moves in the opposite direction ($\beta_i < 0$).

A Testing Rational Expectations*

Reference: Cuthbertson (1996) 5.4

A.0.1 Rational Expectations and Survey Data

If we have survey data, then we can perform a direct test of rational expectations. Consider the scalar variable x_t and suppose we have survey data on the expected value in the next period, $\mathrm{E}_t^p D_{t+1}$. If expectations are rational, then the forecast error $D_{t+1} - \mathrm{E}_t^p D_{t+1}$ should be unforecastable. In particular, if z_t is known in t, then

$$\mathrm{E}[(D_{t+1} - \mathrm{E}_t^p D_{t+1})z_t] = 0. \tag{A.1}$$

This implies that a regression of the forecast errors on any information known in t must give only zero coefficients. Of course, z_t could be a constant so (A.1) says that the mean forecast error should be zero. In practice, this means running the regression

$$D_{t+1} - \mathrm{E}_t^p D_{t+1} = \alpha' Z_t + \varepsilon_{t+1}, \tag{A.2}$$

where $z_t \in Z_t$ and testing if all coefficients (α) are zero.

There are, of course, a few problems with this approach. First, samples can be small (that is always the curse of applied statistics), so that the sample estimate of (A.1) is different from zero by mere chance—in particularly if there was a major surprise during the sample.

Second, the survey respondents may not be the agents that we are most interested in. For instance, suppose the traders actually believe in the efficient market hypothesis at that the current stock price equals traders' expectations of future stock prices. However, a survey of stock price expectations, even if sent to firms on the financial market, is more likely to be answered by financial forecasters/analysts—who may have other expectations. Third, even if the survey respondents are the persons we are interested in (or are likely to share the beliefs with those that we care about), they may choose to miss-report for strategic reasons.

A.0.2 Orthogonality Restrictions and Cross-Equation Restrictions

Suppose we are (somehow) equipped with the correct model, which gives an expression for the expected value

$$\mathrm{E}_t^p D_{t+1} = \gamma' x_t. \tag{A.3}$$

For instance, D_t could be an asset return, so this would be a model of expected asset returns. In the crudest case, it could be that the expected asset return equals the risk free interest rate plus a constant risk premium.

We can test the rational expectations hypothesis by using this expression in (A.1)-(A.2), that is, by an *orthogonality test*. If we know the values of γ, then we can simply calculate a time series of $\mathrm{E}_t^p D_{t+1}$ and then use it in (A.2). If we do not know γ (but are still sure about which variables that belong to the model), then we can combine (A.3) and (A.2) to get the regression equation

$$D_{t+1} = \gamma' x_t + \alpha' Z_t + \varepsilon_{t+1}, \tag{A.4}$$

and test if all elements in α are zero.

This approach can also suffer from the small sample problem, but the really serious problem is probably the specification of the forecasting model. Suppose (A.3) is wrong, and that one of the variables in Z_t belongs to the asset pricing model: (A.2) and (A.4) are then likely to reject the null hypothesis of rational expectations.

A.0.3 Cross-Equation Restrictions

Consider an asset price that is a discounted value of a stream of future "dividends." In the simplest case, the discounting factor (expected return) is a constant, β, and the asset "dies" after the next dividend. In this case, the asset price is

$$P_t = \beta \, \mathrm{E}_t^p D_{t+1}. \tag{A.5}$$

Suppose we have data on a time sequence of such assets (with constant discount factor). Suppose also that the true forecasting equation for D_{t+1} is

$$D_{t+1} = \gamma_1 x_{1t} + \gamma_2 x_{2t} + \upsilon_{t+1}, \text{ so } \mathrm{E}\, D_{t+1} = \gamma_1 x_{1t} + \gamma_2 x_{2t}, \tag{A.6}$$

where we let x_t be a 2×1 vector to simplify the exposition. The only difference between (A.3) and (A.6) is that the latter is the true data generating process, whereas the latter is an equation for how the agents on the financial market form their expectations. If these are identical, that is, if expectations are rational, then we can use (A.6) in (A.5) to get

$$P_t = \beta\gamma_1 x_{1t} + \beta\gamma_2 x_{2t} + u_t \tag{A.7}$$

$$= \delta_1 x_{1t} + \delta_2 x_{2t} + u_t. \tag{A.8}$$

We could estimate the four coefficients in (A.8) and (A.6): γ_1, γ_2, δ_1, and δ_2. Note, however, that there are only three "deep" parameters: β, γ_1, and γ_2, so there are cross-equation ((A.8) and (A.6)) restrictions that can be imposed and tested. The simplest way to do that is to use GMM and use the theoretical equations (A.6) and -(A.7) to specify the following moment conditions

$$m_t = \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} (D_{t+1} - \gamma_1 x_{1t} - \gamma_2 x_{2t})\, x_{1t} \\ (D_{t+1} - \gamma_1 x_{1t} - \gamma_2 x_{2t})\, x_{2t} \\ (P_t - \beta\gamma_1 x_{1t} - \beta\gamma_2 x_{2t})\, x_{1t} \\ (P_t - \beta\gamma_1 x_{1t} - \beta\gamma_2 x_{2t})\, x_{2t} \end{bmatrix}. \tag{A.9}$$

There are four moment conditions, but only three parameters. We can therefore use GMM to estimate the parameters and then test if the single overidentifying restriction is satisfied.

Bibliography

Bodie, Z., A. Kane, and A. J. Marcus, 1999, *Investments*, Irwin/McGraw-Hill, Boston, 4th edn.

Brockwell, P. J., and R. A. Davis, 1991, *Time Series: Theory and Methods*, Springer Verlag, New York, second edn.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay, 1997, *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey.

Cochrane, J. H., 2001, *Asset Pricing*, Princeton University Press, Princeton, New Jersey.

Cuthbertson, K., 1996, *Quantitative Financial Economics*, Wiley, Chichester, England.

Granger, C. W. J., 1992, "Forecasting Stock Market Prices: Lessons for Forecasters," *International Journal of Forecasting*, 8, 3–13.

Huberman, G., and S. Kandel, 1987, "Mean-Variance Spanning," *Journal of Finance*, 42, 873–888.

Lo, A. W., and A. C. MacKinlay, 1997, "Maximizing Predictability in the Stock and Bond Markets," *Macroeconomic Dynamics*, 1, 102–134.

Priestley, M. B., 1981, *Spectral Analysis and Time Series*, Academic Press.

3 Linear Factor Models

3.1 Testing CAPM (Single Excess Return Factor)

Reference: Cochrane (2001) 12.1; Campbell, Lo, and MacKinlay (1997) 5

3.1.1 One Asset at a Time—Traditional ML/LS Approach (iid Returns)

Let $R_{it}^e = R_{it} - R_{ft}$ be the excess return on asset i in excess over the riskfree asset, and let $f_t = R_{mt} - R_{ft}$ be the excess return on the market portfolio. CAPM with a riskfree return says that $\alpha_i = 0$ in

$$R_{it}^e = \alpha + \beta f_t + \varepsilon_{it}, \text{ where } \mathrm{E}\,\varepsilon_{it} = 0 \text{ and } \mathrm{Cov}(f_t, \varepsilon_{it}) = 0. \tag{3.1}$$

See Cochrane (2001) 9.1 for several derivations of CAPM.

This has often been investigated by estimating with LS, and then testing (with a t-test) if the intercept is zero. If the disturbance is iid normally distributed, then this approach is the ML approach.

This test of CAPM can be given two interpretations. If we assume that R_{mt} is the correct benchmark, then it is a test of whether asset R_{it} is correctly priced. Alternatively, if we assume that R_{it} is correctly priced, then it is a test of the mean-variance efficiency of R_{mt} (compare the Roll critique).

Remark 19 *Consider the regression equation $y_t = x_t'b_0 + u_t$. With iid errors that are independent of all regressors (also across observations), the LS estimator, $\hat{b}_{Ls}$, is asymptotically distributed as*

$$\sqrt{T}(\hat{b}_{Ls} - b_0) \xrightarrow{d} N(\mathbf{0}, \sigma^2 \Sigma_{xx}^{-1}), \text{ where } \sigma^2 = \mathrm{E}\,u_t^2 \text{ and } \Sigma_{xx} = \mathrm{E}\,\frac{1}{T}\Sigma_{t=1}^{T} x_t x_t'.$$

When the regressors are just a constant (equal to one) and one variable regressor, f_t, so

$x_t = [1,\ f_t]'$, then we have

$$\Sigma_{xx} = \mathrm{E}\,\frac{1}{T}\Sigma_{t=1}^{T} x_t x_t' = \mathrm{E}\,\frac{1}{T}\Sigma_{t=1}^{T} \begin{bmatrix} 1 & f_t \\ f_t & f_t^2 \end{bmatrix} = \begin{bmatrix} 1 & \mathrm{E}\,f_t \\ \mathrm{E}\,f_t & \mathrm{E}\,f_t^2 \end{bmatrix}, \text{ so}$$

$$\sigma^2 \Sigma_{xx}^{-1} = \frac{\sigma^2}{\mathrm{E}\,f_t^2 - (\mathrm{E}\,f_t)^2} \begin{bmatrix} \mathrm{E}\,f_t^2 & -\mathrm{E}\,f_t \\ -\mathrm{E}\,f_t & 1 \end{bmatrix} = \frac{\sigma^2}{\mathrm{Var}(f_t)} \begin{bmatrix} \mathrm{Var}(f_t) + (\mathrm{E}\,f_t)^2 & -\mathrm{E}\,f_t \\ -\mathrm{E}\,f_t & 1 \end{bmatrix}.$$

(In the last line we use $\mathrm{Var}(f_t) = \mathrm{E}\,f_t^2 - (\mathrm{E}\,f_t)^2$.)

This remark fits well with the CAPM regression (3.1). The variance of the intercept estimator is therefore

$$\mathrm{Var}[\sqrt{T}(\hat{\alpha} - \alpha_0)] = \left\{1 + \frac{(\mathrm{E}\,f_t)^2}{\mathrm{Var}\,(f_t)}\right\} \mathrm{Var}(\varepsilon_{it}) \tag{3.2}$$

$$= [1 + (SR_f)^2]\,\mathrm{Var}(\varepsilon_{it}), \tag{3.3}$$

where SR_f is the Sharpe ratio of the market portfolio (recall: f_t is the excess return on market portfolio). We see that the uncertainty about the intercept is high when the disturbance is volatile and when the sample is short, but also when the Sharpe ratio of the market is high. Note that a large market Sharpe ratio means that the market asks for a high compensation for taking on risk. A bit uncertainty about how risky asset i is then translates in a large uncertainty about what the risk-adjusted return should be.

Remark 20 *If $\sqrt{T}(\hat{\alpha} - \alpha_0) \to^d N(0, s^2)$, then $\hat{\alpha}$ has approximately the distribution $N(\alpha_0, s^2/T)$ in a large sample. (Writing $(\hat{\alpha} - \alpha_0) \to^d N(\alpha_0, s^2/T)$ is not very meaningful since this distribution is degenerate in the limit as the variance is zero.)*

This remark can be used to form a t-test since

$$\frac{\hat{\alpha}}{\mathrm{Std}(\hat{\alpha})} = \frac{\hat{\alpha}}{\sqrt{[1 + (SR_f)^2]\,\mathrm{Var}(\varepsilon_{it})/T}} \xrightarrow{d} N(0, 1) \text{ under } H_0\text{: } \alpha_0 = 0. \tag{3.4}$$

Note that this is the distribution under the null hypothesis that the true value of the intercept is zero, that is, that CAPM is correct (in this respect, at least).

Fact 21 *(Quadratic forms of normally distributed random variables) If the $n \times 1$ vector $X \sim N(0, \Sigma)$, then $Y = X'\Sigma^{-1}X \sim \chi_n^2$. Therefore, if the n scalar random variables*

X_i, $i = 1, ..., n$, are uncorrelated and have the distributions $N(0, \sigma_i^2)$, $i = 1, ..., n$, then $Y = \Sigma_{i=1}^n X_i^2/\sigma_i^2 \sim \chi_n^2$.

Instead of a t-test, we can use the equivalent chi-square test

$$\frac{\hat{\alpha}^2}{\text{Var}(\hat{\alpha})} = \frac{\hat{\alpha}^2}{[1 + (SR_f)^2]\,\text{Var}(\varepsilon_{it})/T} \xrightarrow{d} \chi_1^2 \text{ under } H_0\text{: } \alpha_0 = 0. \tag{3.5}$$

The chi-square test is equivalent to the t-test when we are testing only one restriction, but it has the advantage that it also allows us to test several restrictions at the same time. Both the t-test and the chi–square tests are Wald tests (estimate unrestricted model and then test the restrictions). We discuss how to set up LM and LR tests below.

The test statistic (3.5) can be used to demonstrate the alternative interpretation of the test: that is tests if R_{mt} is mean-variance efficient. To do that, note that from (3.1) we have

$$\text{Cov}\begin{bmatrix} R_{it}^e \\ f_t \end{bmatrix} = \begin{bmatrix} \beta^2\,\text{Var}(f_t) + \text{Var}(\varepsilon_{it}) & \beta\,\text{Var}(f_t) \\ \beta\,\text{Var}(f_t) & \text{Var}(f_t) \end{bmatrix}, \text{ and } \text{E}\begin{bmatrix} R_{it}^e \\ f_t \end{bmatrix} = \begin{bmatrix} \alpha_i + \beta\,\text{E}\,f_t \\ \text{E}\,f_t \end{bmatrix}. \tag{3.6}$$

Suppose we use this information to construct a mean-variance frontier for R_{it} and R_{mt}, and we find the tangency portfolio, with excess return R_{qt}^e. By using (3.6) in the standard formulae, we get that the squared Sharpe ratio for the tangency portfolio is (try it)

$$\frac{(\text{E}\,R_{qt}^e)^2}{\text{Var}(R_{qt}^e)} = \frac{\alpha_i^2}{\text{Var}(\varepsilon_{it})} + \frac{(\text{E}\,f_t)^2}{\text{Var}(f_t)}. \tag{3.7}$$

If we now replace the population parameters with the sample estimates, we can use this expression to see that the test statistic in (3.5) is

$$\frac{\hat{\alpha}^2}{\text{Var}(\hat{\alpha})} = \frac{(SR_q)^2 - (SR_f)^2}{[1 + (SR_f)^2]/T}. \tag{3.8}$$

This shows that if the market portfolio has the same (squared) Sharpe ratio as the tangency portfolio of the mean-variance frontier of R_{it} and R_{mt} (so the market portfolio is mean-variance efficient also when we take R_{it} into account) then the test statistic, $\hat{\alpha}^2/\,\text{Var}(\hat{\alpha})$, is zero. This result is due to Gibbons, Ross, and Shanken (1989). For a discussion and critique, see MacKinlay (1995).

It is also possible to construct small sample test (that does not rely on any asymptotic results), which may be a better approximation of the correct distribution in real-life samples—provided the strong assumptions are (almost) satisfied. The most straightforward modification is to transform (3.5) into an $F_{1,T-1}$-test. This is the same as using a t-test in (3.4) since it is only one restriction that is tested (recall that if $Z \sim t_m$, then $Z^2 \sim F(1, m)$).

A common finding is that these tests tend to reject a true null hypothesis too often when the critical values from the asymptotic distribution are used: the actual small sample *size of the test* is thus larger than the asymptotic (or "nominal") size (see Campbell, Lo, and MacKinlay (1997) Table 5.1). To study the power of the test (the frequency of rejections of a false null hypothesis) we have to specify an alternative data generating process (for instance, how much extra return in excess of that motivated by CAPM) and the size of the test (the critical value to use). Once that is done, it is typically found that these tests require a substantial deviation from CAPM and/or a long sample to get good power.

3.1.2 Several Assets and Non-Spherical Errors (GMM)—Wald Test

To test n assets at the same time when the errors are non-iid we make use of the GMM framework. The case where $n = 1$ will be a special case. A special case of that special case is when returns are also iid. The results in this section will then coincide with those in Section 3.1.1—but only then.

Let $f_t = R_{mt} - R_{ft}$ and stack the expressions (3.1) for $i = 1, \ldots, n$ as

$$\begin{bmatrix} R_{1t}^e \\ \vdots \\ R_{nt}^e \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} f_t + \begin{bmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{nt} \end{bmatrix}, \ \text{E}\,\varepsilon_{it} = 0 \text{ and } \text{Cov}(f_t, \varepsilon_{it}) = 0, \tag{3.9}$$

or more compactly

$$R_t^e = \alpha + \beta f_t + \varepsilon_t, \ \text{E}\,\varepsilon_t = \mathbf{0}_{n\times 1} \text{ and } \text{Cov}(f_t, \varepsilon_t) = \mathbf{0}_{1\times n}, \tag{3.10}$$

where α and β are $n \times 1$ vectors.

If the disturbances are iid normally distributed, then LS equation by equation is the maximum likelihood estimator of the system (this is a SURE system with identical regressors; see, for instance, Greene (2000) 15.4)—and the tests would be a straightforward

extension of the tests in Section 3.1.1. Instead, we focus on the more complicated case with non-iid errors (normally distributed or not).

The $2n$ GMM moment conditions are that, at the true values of α and β,

$$\mathrm{E}\, m_t(\alpha,\beta) = \mathrm{E}\begin{bmatrix} \varepsilon_t \\ f_t\varepsilon_t \end{bmatrix} = \mathrm{E}\begin{bmatrix} R_t^e - \alpha - \beta f_t \\ f_t(R_t^e - \alpha - \beta f_t) \end{bmatrix} = \mathbf{0}_{2n\times 1}. \tag{3.11}$$

There are as many parameters as moment conditions, so the GMM estimator picks values of α and β such that the sample analogues of (3.11) are satisfied exactly

$$\bar{m}(\alpha,\beta) = \frac{1}{T}\sum_{t=1}^{T} m_t(\alpha,\beta) = \frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} R_t^e - \hat{\alpha} - \hat{\beta} f_t \\ f_t(R_t^e - \hat{\alpha} - \hat{\beta} f_t) \end{bmatrix} = \mathbf{0}_{2n\times 1}, \tag{3.12}$$

which gives the LS estimator. No attempt is done to do ML estimation in order to gain efficiency.

Remark 22 *Let the parameter vector in the moment condition have the true value b_0. Define*

$$S_0 = ACov\left[\sqrt{T}\bar{m}\,(b_0)\right] \text{ and } D_0 = \operatorname{plim}\frac{\partial \bar{m}(b_0)}{\partial b'}.$$

We know that (in most cases when the estimator solves $\min \bar{m}\,(b)'\, S_0^{-1}\bar{m}\,(b)$ or when the model is exactly identified), the distribution of the GMM estimator $\hat{b}$ is

$$\sqrt{T}(\hat{b} - b_0) \xrightarrow{d} N\left[\mathbf{0}_{k\times 1}, \left(D_0' S_0^{-1} D_0\right)^{-1}\right].$$

If D_0 is invertible, then $\left(D_0' S_0^{-1} D_0\right)^{-1} = D_0^{-1} S_0 (D_0')^{-1} = D_0^{-1} S_0 (D_0^{-1})'$ (recall $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$ and $(A')^{-1} = (A^{-1})'$, if conformable).

With point estimates and their sampling distribution it is straightforward to set up a Wald test for the hypothesis that all elements in α are zero

$$\hat{\alpha}'\,\mathrm{Var}(\hat{\alpha})^{-1}\hat{\alpha} \xrightarrow{d} \chi_n^2. \tag{3.13}$$

Details on the Wald Test*

To be concrete, consider the case with two assets (1 and 2) so the parameter vector is $b = [\alpha_1, \alpha_2, \beta_1, \beta_2]'$. Write out (3.12) as

$$\begin{bmatrix} \bar{m}_1 \\ \bar{m}_2 \\ \bar{m}_3 \\ \bar{m}_4 \end{bmatrix} = \begin{bmatrix} \frac{1}{T}\Sigma_{t=1}^T (R_{1t}^e - \hat{\alpha}_1 - \hat{\beta}_1 f_t) \\ \frac{1}{T}\Sigma_{t=1}^T (R_{2t}^e - \hat{\alpha}_2 - \hat{\beta}_2 f_t) \\ \frac{1}{T}\Sigma_{t=1}^T f_t(R_{1t}^e - \hat{\alpha}_1 - \hat{\beta}_1 f_t) \\ \frac{1}{T}\Sigma_{t=1}^T f_t(R_{2t}^e - \hat{\alpha}_2 - \hat{\beta}_2 f_t) \end{bmatrix} = \mathbf{0}_{4\times 1}. \tag{3.14}$$

The Jacobian is

$$\begin{aligned} \frac{\partial \bar{m}(\alpha,\beta)}{\partial[\alpha_1,\alpha_2,\beta_1,\beta_2]'} &= \begin{bmatrix} \partial\bar{m}_1/\partial\alpha_1 & \partial\bar{m}_1/\partial\alpha_2 & \partial\bar{m}_1/\partial\beta_1 & \partial\bar{m}_1/\partial\beta_2 \\ \partial\bar{m}_2/\partial\alpha_1 & \partial\bar{m}_2/\partial\alpha_2 & \partial\bar{m}_2/\partial\beta_1 & \partial\bar{m}_2/\partial\beta_2 \\ \partial\bar{m}_3/\partial\alpha_1 & \partial\bar{m}_3/\partial\alpha_2 & \partial\bar{m}_3/\partial\beta_1 & \partial\bar{m}_3/\partial\beta_2 \\ \partial\bar{m}_4/\partial\alpha_1 & \partial\bar{m}_4/\partial\alpha_2 & \partial\bar{m}_4/\partial\beta_1 & \partial\bar{m}_4/\partial\beta_2 \end{bmatrix} \\ &= -\begin{bmatrix} 1 & 0 & \frac{1}{T}\Sigma_{t=1}^T f_t & 0 \\ 0 & 1 & 0 & \frac{1}{T}\Sigma_{t=1}^T f_t \\ \frac{1}{T}\Sigma_{t=1}^T f_t & 0 & \frac{1}{T}\Sigma_{t=1}^T f_t^2 & 0 \\ 0 & \frac{1}{T}\Sigma_{t=1}^T f_t & 0 & \frac{1}{T}\Sigma_{t=1}^T f_t^2 \end{bmatrix}. \end{aligned} \tag{3.15}$$

Note that, in this case with a linear model, the Jacobian does not involve the parameters that we want to estimate. This means that we do not have to worry about evaluating the Jacobian at the true parameter values. The probability limit of (3.15) is simply

$$D_0 = -\begin{bmatrix} 1 & 0 & \mathrm{E}\, f_t & 0 \\ 0 & 1 & 0 & \mathrm{E}\, f_t \\ \mathrm{E}\, f_t & 0 & \mathrm{E}\, f_t^2 & 0 \\ 0 & \mathrm{E}\, f_t & 0 & \mathrm{E}\, f_t^2 \end{bmatrix} \tag{3.16}$$

$$= -\mathrm{E}\begin{bmatrix} 1 & f_t \\ f_t & f_t^2 \end{bmatrix} \otimes I_2 = -\mathrm{E}\left(\begin{bmatrix} 1 \\ f_t \end{bmatrix}\begin{bmatrix} 1 \\ f_t \end{bmatrix}'\right) \otimes I_2, \tag{3.17}$$

where $\otimes$ is the Kronecker product. For n assets, change I_2 to I_n.

Remark 23 *(Kronecker product) If A and B are matrices, then*

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}.$$

Note that D_0 is an invertible matrix, so we can use the result in Remark 22 to write the covariance matrix of the $2n \times 1$ vector of parameters (n parameters in α and another n in β) as

$$\text{ACov}\left(\sqrt{T}\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\end{bmatrix}\right) = D_0^{-1}S_0(D_0^{-1})'. \tag{3.18}$$

The asymptotic covariance matrix of $\sqrt{T}$ times the sample moment conditions, evaluated at the true parameter values, that is at the true disturbances, is defined as

$$S_0 = \text{ACov}\left(\frac{\sqrt{T}}{T}\sum_{t=1}^{T} m_t\right) = \text{ACov}\left(\sqrt{T}\begin{bmatrix} \frac{1}{T}\Sigma_{t=1}^{T}\varepsilon_{1t} \\ \frac{1}{T}\Sigma_{t=1}^{T}\varepsilon_{2t} \\ \frac{1}{T}\Sigma_{t=1}^{T}f_t\varepsilon_{1t} \\ \frac{1}{T}\Sigma_{t=1}^{T}f_t\varepsilon_{2t}\end{bmatrix}\right) = \sum_{s=-\infty}^{\infty} R(s), \text{ where} \tag{3.19}$$

$$R(s) = \text{E}\, m_t m_{t-s}' = \text{E}\begin{bmatrix}\varepsilon_{1t}\\ \varepsilon_{2t}\\ f_t\varepsilon_{1t}\\ f_t\varepsilon_{2t}\end{bmatrix}\begin{bmatrix}\varepsilon_{1t-s}\\ \varepsilon_{2t-s}\\ f_{t-s}\varepsilon_{1t-s}\\ f_{t-s}\varepsilon_{2t-s}\end{bmatrix}'. \tag{3.20}$$

With n assets, we can write (3.20) in terms of the $n \times 1$ vector ε_t as

$$R(s) = \text{E}\, m_t m_{t-s}' = \text{E}\begin{bmatrix}\varepsilon_t\\ f_t\varepsilon_t\end{bmatrix}\begin{bmatrix}\varepsilon_{t-s}\\ f_{t-s}\varepsilon_{t-s}\end{bmatrix}' = \text{E}\begin{bmatrix}\varepsilon_t\varepsilon_{t-s}' & \varepsilon_t\varepsilon_{t-s}'f_{t-s}\\ f_t\varepsilon_t\varepsilon_{t-s}' & f_t\varepsilon_t\varepsilon_{t-s}'f_{t-s}\end{bmatrix}. \tag{3.21}$$

The Newey-West estimator is often a good estimator of S_0, but the estimation can sometimes be simplified, and the performance of the test improved, by imposing (correct, of course) restrictions on the $R(s)$ matrices.

We get the expression for the covariance matrix of the GMM estimators by using (3.3) and (3.19)-(3.21) in (3.18).

Remark 24 *(Special case 1: f_t is independent of ε_{t-s}, errors are iid, and $n = 1$) With*

these assumptions $R(s) = \mathbf{0}_{2n\times 2n}$ if $s \neq 0$, and $S_0 = \begin{bmatrix} 1 & \text{E}\, f_t \\ \text{E}\, f_t & \text{E}\, f_t^2 \end{bmatrix}\text{Var}(\varepsilon_{it})$. Combining with (3.17) gives

$$ACov\left(\sqrt{T}\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\end{bmatrix}\right) = \begin{bmatrix} 1 & \text{E}\, f_t \\ \text{E}\, f_t & \text{E}\, f_t^2 \end{bmatrix}^{-1}\text{Var}(\varepsilon_{it}),$$

which is the same expression as $\sigma^2\Sigma_{xx}^{-1}$ in Remark 19.

Remark 25 *(Special case 2: as in Special case 1, but $n \geq 1$) With these assumptions $R(s) = \mathbf{0}_{2n\times 2n}$ if $s \neq 0$, and $S_0 = \begin{bmatrix} 1 & \text{E}\, f_t \\ \text{E}\, f_t & \text{E}\, f_t^2 \end{bmatrix} \otimes \text{E}\,\varepsilon_t\varepsilon_t'$. Combining with (3.17) (recall $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ and $(A \otimes B)(C \otimes D) = AC \otimes BD$, if conformable) gives*

$$ACov\left(\sqrt{T}\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\end{bmatrix}\right) = \begin{bmatrix} 1 & \text{E}\, f_t \\ \text{E}\, f_t & \text{E}\, f_t^2 \end{bmatrix}^{-1} \otimes \left(\text{E}\,\varepsilon_t\varepsilon_t'\right).$$

Size and Power of Tests

The size (using asymptotic critical values) and power in small samples is often found to be disappointing. Typically, these tests tend to reject a true null hypothesis too often (see Campbell, Lo, and MacKinlay (1997) Table 5.1) and the power to reject a false null hypothesis is often fairly low. These features are especially pronounced when the sample is small and the number of assets, n, is low. One useful rule of thumb is that a *saturation ratio* (the number of observations per parameter) below 10 (or so) is likely to give poor performance of the test. In the test here we have nT observations, $2n$ parameters in α and β, and $n(n+1)/2$ unique parameters in S_0, so the saturation ratio is $T/(2 + (n+1)/2)$. For instance, with $T = 60$ and $n = 10$ or at $T = 100$ and $N = 20$, we have a saturation ratio of 8, which is very low (compare Table 5.1 in CLM).

One possible way of dealing with the wrong size of the test is to use critical values from simulations of the small sample distributions (Monte Carlo simulations or bootstrap simulations).

Choice of Portfolios and some Empirical Evidence

This type of test is typically done on portfolios of assets, rather than on the individual assets themselves. There are several econometric and economic reasons for this. The econometric techniques we apply need the returns to be (reasonably) stationary in the sense that they have approximately the same means and covariance (with other returns) throughout the sample (individual assets, especially stocks, can change character as the company moves into another business). It might be more plausible that size or industry portfolios are stationary in this sense. Individual portfolios are typically very volatile, which makes it hard to obtain precise estimate and to be able to reject anything.

It sometimes makes economic sense to sort the assets according to a characteristic (size or perhaps book/market)—and then test if the model is true for these portfolios. Rejection of the CAPM for such portfolios may have an interest in itself.

For some empirical evidence, see Campbell, Lo, and MacKinlay (1997) Table 5.3 and Cochrane (2001) 20.2 (in particular, Figure 20.8).

3.1.3 Several Assets and Non-Spherical Errors (GMM)—LM Test

We could also construct an "LM test" instead by imposing $\alpha = \mathbf{0}$ in the moment conditions (3.11) and (3.12). The sample moment conditions are then

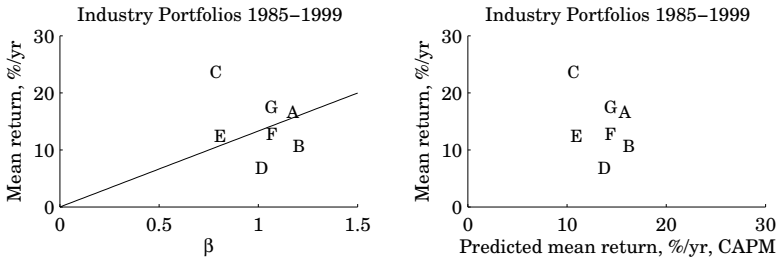
$$\bar{m}(\alpha, \beta) = \frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} R_t^e - \hat{\beta} f_t \\ f_t(R_t^e - \hat{\beta} f_t) \end{bmatrix} = \mathbf{0}_{2n\times 1}. \tag{3.22}$$

Since there are $q = 2n$ moment conditions, but only n parameters (the β vector), this model is overidentified. We therefore need a weighting matrix in the GMM loss function.

Once we have estimated the model, we can test the n overidentifying restrictions that all $q = 2n$ moment conditions are satisfied at the estimated n parameters $\hat{\beta}$. If not, the restrictions (null hypothesis) that $\alpha = \mathbf{0}$ is rejected. If we used the efficient weighting matrix, then the test statistic is of a particularly simple form.

Remark 26 *When the GMM estimator solves the quadratic loss function $\bar{m}(\beta)' S_0^{-1} \bar{m}(\beta)$ (that is, when the optimal weighting matrix has been used), then the J test statistic is*

$$T\bar{m}(\hat{\beta})' S_0^{-1} \bar{m}(\hat{\beta}) \xrightarrow{d} \chi^2_{q-k}, \tag{3.23}$$



Asset	WaldStat	pval
all	14.90	0.04
1	0.43	0.51
2	1.23	0.27
3	7.25	0.01
4	1.54	0.21
5	0.95	0.33
6	0.11	0.74
7	0.32	0.57

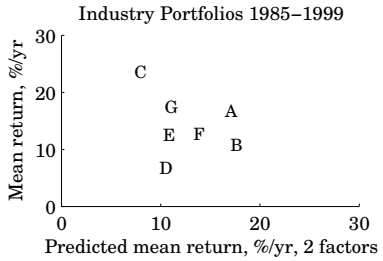
Figure 3.1: CAPM, Swedish industry portfolios, weekly data 1985-1999

where q is the number of moment conditions (elements in $\bar{m}(\beta)$) and k is the number of parameters.

Note that the J test ($TJ(\hat{\beta}^{restricted}) \sim \chi^2_{q-k}$, if $W = S_0^{-1}$) and D test ($T[J(\hat{\beta}^{restricted}) - J(\hat{\beta}^{less\ restricted})] \sim \chi^2_s$, if $W = S_0^{-1}$) coincide in this case, since the loss function value of the unrestricted model (3.12) is zero as there are as many parameters as moment conditions.

3.2 Testing Multi-Factor Models (Factors are Excess Returns)

Reference: Cochrane (2001) 12.1; Campbell, Lo, and MacKinlay (1997) 6.2.1



Asset	WaldStat	pval
all	18.78	0.01
1	0.01	0.91
2	2.06	0.15
3	10.91	0.00
4	0.43	0.51
5	1.02	0.31
6	0.03	0.87
7	1.16	0.28

Figure 3.2: Two-factor model, Swedish industry portfolios, weekly data 1985-1999

3.2.1 General Approach

When the K factors, f_t, are excess returns, the null hypothesis typically says that $\alpha_i = 0$ in

$$R^e_{it} = \alpha_i + \beta_i' f_t + \varepsilon_{it}, \text{ where } \mathrm{E}\,\varepsilon_{it} = 0 \text{ and } \mathrm{Cov}(f_t, \varepsilon_{it}) = \mathbf{0}_{K\times 1}. \tag{3.24}$$

and β_i is now an $K \times 1$ vector. The CAPM regression (3.1) is a special case when the market excess return is the only factor. In other models like ICAPM (see Cochrane (2001) 9.2), we typically have several factors. A standard atheoretical way to extract K factors from N portfolios is to pick the first K principal components. We stack the returns for n assets to get

$$\begin{bmatrix} R^e_{1t} \\ \vdots \\ R^e_{nt} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} \beta_{11} & \dots & \beta_{1K} \\ \vdots & \ddots & \vdots \\ \beta_{n1} & \dots & \beta_{nK} \end{bmatrix} \begin{bmatrix} f_{1t} \\ \vdots \\ f_{Kt} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{nt} \end{bmatrix}, \text{ or}$$

$$R^e_t = \alpha + \beta f_t + \varepsilon_t, \text{ where } \mathrm{E}\,\varepsilon_t = \mathbf{0}_{n\times 1} \text{ and } \mathrm{Cov}(f_t, \varepsilon_t) = \mathbf{0}_{K\times n}, \tag{3.25}$$

where α is $n \times 1$ and β is $n \times K$.

When we test one asset at a time (as we do in (3.24)), then we can still apply a t-test or chi-square test as in (3.4) and (3.8), except that $\mathrm{Var}(\hat{\alpha}_i)$ is a more complicated expression than before. More generally, we test the null hypothesis with the Wald test (3.13)—using the appropriate $\mathrm{Var}(\hat{\alpha}_i)$ or $\hat{S}_0$ matrix (how they should be estimated depends on the features of the data).

For instance, if the errors are iid and independent of the factors (across periods too), then the standard LS expression for the covariance matrix of the parameters apply. If these conditions are not satisfied, then we need to adjust the covariance estimator with a Newey-West approach or something similar. This is most easily accomplished by treating the estimation problem as a GMM problem.

Fact 27 *(Principal component analysis) The first (sample) principal component of the $N \times 1$ vector R_t is $w_1' R_t$ where w_1 is the eigenvector associated with the largest eigenvalue of $\Sigma = \mathrm{Cov}(R_t)$. This value of w_1 solves the problem $\max_w w'\Sigma w$ subject to the normalization $w'w = 1$. The jth principal component solves the same problem, but under the additional restriction that $w_i' w_j = 0$ for all $i < j$. The solution is the eigenvector associated with the jth largest eigenvalue. This means that the first K principal components are those (normalized) linear combinations that account for as much of the variability as possible—and that the principal components are uncorrelated.*

Details*

With only one asset and two factors (f_{1t} and f_{2t}) the moment conditions for GMM are

$$\mathrm{E}\,m_t(\alpha, \beta) = \mathrm{E}\begin{bmatrix} \varepsilon_{it} \\ f_{1t}\varepsilon_{it} \\ f_{2t}\varepsilon_{it} \end{bmatrix} = \mathbf{0}_{3\times 1}, \tag{3.26}$$

and with two assets and two factors we have

$$\mathrm{E}\,m_t(\alpha, \beta) = \mathrm{E}\begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ f_{1t}\varepsilon_{1t} \\ f_{1t}\varepsilon_{2t} \\ f_{2t}\varepsilon_{1t} \\ f_{2t}\varepsilon_{2t} \end{bmatrix} = \mathrm{E}\begin{bmatrix} R^e_{1t} - \alpha_1 - \beta_{11} f_{1t} - \beta_{12} f_{2t} \\ R^e_{2t} - \alpha_2 - \beta_{21} f_{1t} - \beta_{22} f_{2t} \\ f_{1t}(R^e_{1t} - \alpha_1 - \beta_{11} f_{1t} - \beta_{12} f_{2t}) \\ f_{1t}(R^e_{2t} - \alpha_2 - \beta_{21} f_{1t} - \beta_{22} f_{2t}) \\ f_{2t}(R^e_{1t} - \alpha_1 - \beta_{11} f_{1t} - \beta_{12} f_{2t}) \\ f_{2t}(R^e_{2t} - \alpha_2 - \beta_{21} f_{1t} - \beta_{22} f_{2t}) \end{bmatrix} = \mathbf{0}_{6\times 1}. \tag{3.27}$$

Note that β_{ij} shows how the ith asset depends on the jth factor.

If we let $\varepsilon_t = [\varepsilon_{1t}, \varepsilon_{2t}]'$ and $f_t = [f_{1t}, f_{2t}]'$, then these moment conditions can be

written

$$\mathrm{E}\, m_t(\alpha,\beta) = \mathrm{E}\left(\begin{bmatrix} 1 \\ f_t \end{bmatrix} \otimes \varepsilon_t\right) = \mathrm{E}\left(\begin{bmatrix} 1 \\ f_t \end{bmatrix} \otimes (R_t^e - \alpha - \beta f_t)\right) = \mathbf{0}_{6\times 1}. \quad (3.28)$$

With n assets and K factors, change $\mathbf{0}_{6\times 1}$ to $\mathbf{0}_{n(1+K)\times 1}$. Note that this expression looks similar to (3.11)—the only difference is that f_t may now be a vector (and we therefore need to use the Kronecker product). It is then intuitively clear that the expressions for the asymptotic covariance matrix of $\hat{\alpha}$ and $\hat{\beta}$ will look very similar too. See Appendix for details on how to code this GMM problem.

To show that, consider the Jacobian of the moment conditions in (3.27)

$$\frac{\partial \bar{m}(\alpha,\beta)}{\partial[\alpha,\beta]'} = \begin{bmatrix} \partial\bar{m}_1/\partial\alpha_1 & \partial\bar{m}_1/\partial\alpha_2 & \partial\bar{m}_1/\partial\beta_{11} & \partial\bar{m}_1/\partial\beta_{21} & \partial\bar{m}_1/\partial\beta_{12} & \partial\bar{m}_1/\partial\beta_{22} \\ \vdots & \ddots & & & & \vdots \\ & & & & & \\ & & & & & \\ & & & & & \\ \partial\bar{m}_6/\partial\alpha_1 & \cdots & & & & \partial\bar{m}_6/\partial\beta_{22} \end{bmatrix}. \quad (3.29)$$

Note that, the first two columns are the derivatives with respect to the intercepts, the next two columns are the derivatives with respect to the coefficients of the first factor, f_{1t}, and the last two columns are the derivatives with respect to the second factor, f_{2t}. Using (3.27) and then evaluating the probability limit gives

$$\begin{aligned} D_0 &= -\mathrm{E}\begin{bmatrix} 1 & 0 & f_{1t} & 0 & f_{2t} & 0 \\ 0 & 1 & 0 & f_{1t} & 0 & f_{2t} \\ f_{1t} & 0 & f_{1t}^2 & 0 & f_{1t}f_{2t} & 0 \\ 0 & f_{1t} & 0 & f_{1t}^2 & 0 & f_{1t}f_{2t} \\ f_{2t} & 0 & f_{2t}f_{1t} & 0 & f_{2t}^2 & 0 \\ 0 & f_{2t} & 0 & f_{2t}f_{1t} & 0 & f_{2t}^2 \end{bmatrix} \\ &= -\mathrm{E}\begin{bmatrix} 1 & f_{1t} & f_{2t} \\ f_{1t} & f_{1t}^2 & f_{1t}f_{2t} \\ f_{2t} & f_{2t}f_{1t} & f_{2t}^2 \end{bmatrix} \otimes I_2 = -\mathrm{E}\left(\begin{bmatrix} 1 \\ f_t \end{bmatrix}\begin{bmatrix} 1 \\ f_t \end{bmatrix}'\right) \otimes I_2. \end{aligned} \quad (3.30)$$

With n assets, change I_2 to I_n. This is clearly the same as (3.17), even if f_t may contain

several elements here.

The asymptotic covariance matrix of ($\sqrt{T}$ times) the moment conditions is

$$S_0 = \mathrm{ACov}\left(\frac{\sqrt{T}}{T}\sum_{t=1}^{T} m_t\right) = \mathrm{ACov}\left(\frac{\sqrt{T}}{T}\sum_{t=1}^{T}\begin{bmatrix} 1 \\ f_t \end{bmatrix} \otimes \varepsilon_t\right) = \sum_{s=-\infty}^{\infty} R(s), \text{ where} \quad (3.31)$$

$$R(s) = \mathrm{E}(m_t m_{t-s}') = \mathrm{E}\left[\left(\begin{bmatrix} 1 \\ f_t \end{bmatrix} \otimes \varepsilon_t\right)\left(\begin{bmatrix} 1 \\ f_{t-s} \end{bmatrix} \otimes \varepsilon_{t-s}\right)'\right]. \quad (3.32)$$

which is essentially the same as (3.19)-(3.21).

Remark 28 *(Easy coding of GMM Problem (3.28)) This exactly identified linear system can be estimated by LS, equation by equation. Once that is done, plug in the fitted residuals in the moment conditions, that is, (3.28) without the expectations operator*

$$m_t(\alpha,\beta) = \begin{bmatrix} 1 \\ f_t \end{bmatrix} \otimes \varepsilon_t.$$

We can then estimate the covariance matrix (3.31)-(3.32) by, for instance, the Newey-West method. A harder, but more general approach (works also in overidentified models) is discussed in the Appendix.

To set up a J ("LM") test (3.23), we change the moment conditions (3.28) by imposing $\alpha = \mathbf{0}$. We then have $n(1+K)$ moment conditions, but only nK parameters, so there are n overidentifying restrictions to test (essentially that we get the mean returns right even with zero intercepts).

3.3 Testing Multi-Factor Models (General Factors)

Reference: Cochrane (2001) 12.2; Campbell, Lo, and MacKinlay (1997) 6.2.3 and 6.3

3.3.1 General Approach

Linear factor models imply that all expected excess returns are linear functions of the same vector of factor risk premia

$$\mathrm{E}\, R_{it}^e = \beta_i'\lambda, \text{ where } \lambda \text{ is } K \times 1, \text{ for } i = 1, \ldots n, \text{ or} \tag{3.33}$$

$$\mathrm{E}\begin{bmatrix} R_{1t}^e \\ \vdots \\ R_{nt}^e \end{bmatrix} = \begin{bmatrix} \beta_{11} & \ldots & \beta_{1K} \\ \vdots & \ddots & \vdots \\ \beta_{n1} & \ldots & \beta_{nK} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_K \end{bmatrix}, \text{ or}$$

$$\mathrm{E}\, R_t^e = \beta\lambda, \tag{3.34}$$

where β is $n \times K$.

The old way of testing this is to do a two-step estimation: first, estimate the β_i vectors in a time series model like (3.25) (equation by equation); second, use $\hat{\beta}_i$ as regressors in a regression equation of the type (3.33) with a residual added: $\Sigma_{t=1}^T R_{it}^e/T = \hat{\beta}_i'\lambda + u_i$. It is then tested if $u_i = 0$ for all assets $i = 1, \ldots, n$. This approach is often called a *cross-sectional* regression while the previous tests are called *time series* regression.

The main problem of the cross-sectional approach is that we have to account for the fact that the regressors in the second step, $\hat{\beta}_i$, are just estimates and therefore contain estimation errors. This errors-in-variables problem is likely to have two effects: *(i)* bias the estimates of λ downwards and the mean of the fitted residuals upwards; and *(ii)* invalidate the standard expression the test of the second step regression.

A way to handle these problems is to combine the moment conditions for the regression function (3.28) (to estimate β) with (3.34) (to estimate λ) to get a joint system

$$\mathrm{E}\, m_t(\alpha, \beta, \lambda) = \mathrm{E}\begin{bmatrix} \begin{bmatrix} 1 \\ f_t \end{bmatrix} \otimes (R_t^e - \alpha - \beta f_t) \\ R_t^e - \beta\lambda \end{bmatrix} = \mathbf{0}_{n(1+K+1)\times 1}. \tag{3.35}$$

We can then test the overidentifying restrictions of the model. There are $n(1+K+1)$ moment condition (for each asset we have one moment condition for the constant, K moment conditions for the K factors, and one moment condition corresponding to the restriction on the linear factor model). There are only $n(1+K)+K$ parameters (n in α, nK in β and K in λ). We therefore have $n-K$ overidentifying restrictions which can be tested with a chi-square test. For instance, if we use the efficient weighting matrix,

then the test (3.23) is used. Note that this is a non-linear estimation problem, since the parameters in β multiply the parameters in λ.

Details*

For instance, with two assets and one factor we have the moment conditions

$$\mathrm{E}\, m_t(\alpha_1, \alpha_2, \beta_1, \beta_2, \lambda) = \mathrm{E}\begin{bmatrix} (R_{1t}^e - \alpha_1 - \beta_1 f_t) \\ (R_{2t}^e - \alpha_2 - \beta_2 f_t) \\ f_t(R_{1t}^e - \alpha_1 - \beta_1 f_t) \\ f_t(R_{2t}^e - \alpha_2 - \beta_2 f_t) \\ R_{1t}^e - \beta_1\lambda \\ R_{2t}^e - \beta_2\lambda \end{bmatrix} = \mathbf{0}_{6\times 1}. \tag{3.36}$$

There are then 6 moment conditions and 5 parameters, so there is one overidentifying restriction to test. Note that with one factor, then we need at least two assets for this testing approach to work ($n - K = 2 - 1$). In general, we need at least one more asset than factors.

From the GMM estimation using (3.35) we get estimates of the factor risk premia and also the variance-covariance of them. This allows us to characterize the risk factors and to test if they priced (each of them or perhaps all jointly) by using a Wald test.

Alternatively, we can combine the moment equations so they become equal to the number of parameters, for instance by specifying an $(n(1+K)+K) \times [n(1+K+1)]$ matrix A and combine as $A\,\mathrm{E}\, m_t = \mathbf{0}_{(n(1+K)+K)\times 1}$. One possibility is to let the upper left block of A be an identity matrix and just combine the last n moment conditions, $R_t^e - \beta\lambda$, to just K moment conditions. For instance, combining as $\beta'(R_t^e - \beta\lambda)$ gives the traditional cross-section approach—but where we take the uncertainty about the generated betas into account. This does not generate any overidentifying restrictions, but it still allows us to test hypothesis about λ. For instance, the moment conditions in (3.36) could be combined

as

$$\mathrm{E}\, m_t(\alpha_1, \alpha_2, \beta_1, \beta_2, \lambda) = \mathrm{E}\begin{bmatrix} (R^e_{1t} - \alpha_1 - \beta_1 f_t) \\ (R^e_{2t} - \alpha_2 - \beta_2 f_t) \\ f_t(R^e_{1t} - \alpha_1 - \beta_1 f_t) \\ f_t(R^e_{2t} - \alpha_2 - \beta_2 f_t) \\ \beta_1(R^e_{1t} - \beta_1\lambda) + \beta_2(R^e_{2t} - \beta_2\lambda) \end{bmatrix} = \mathbf{0}_{5\times 1}, \quad (3.37)$$

which has as many parameters as moment conditions.

3.3.2 Alternative Formulation of Moment Conditions using $\alpha = \beta(\lambda - \mathrm{E}\, f_t)$

The test of the general multi-factor models is sometimes written on a slightly different form (see, for instance, Campbell, Lo, and MacKinlay (1997) 6.2.3, but adjust for the fact that they look at returns rather than excess returns). To illustrate this, note that the regression equations (3.25) imply that

$$\mathrm{E}\, R^e_t = \alpha + \beta\, \mathrm{E}\, f_t. \quad (3.38)$$

Equate (3.38) and (3.33) to get

$$\alpha = \beta(\lambda - \mathrm{E}\, f_t), \quad (3.39)$$

which is another way of summarizing the restrictions that the linear factor model gives. We can then rewrite the moment conditions (3.35) as (substitute for α and skip the last set of moments)

$$\mathrm{E}\, m_t(\beta, \lambda) = \mathrm{E}\left[\begin{bmatrix} 1 \\ f_t \end{bmatrix} \otimes (R^e_t - \beta(\lambda - \mathrm{E}\, f_t) - \beta f_t)\right] = \mathbf{0}_{n(1+K)\times 1}. \quad (3.40)$$

Note that there are $n(1 + K)$ moment conditions and $nK + K$ parameters (nK in β and K in λ), so there are $n - K$ overidentifying restrictions (as before).

Details*

To see how these are used, consider the case when there are two asset and one factor. Use the restrictions (3.39) in the moment conditions for that case (compare with (3.14)) to get

$$\mathrm{E}\, m_t(\beta_1, \beta_2, \lambda) = \mathrm{E}\begin{bmatrix} R^e_{1t} - \beta_1(\lambda - \mathrm{E}\, f_t) - \beta_1 f_t \\ R^e_{2t} - \beta_2(\lambda - \mathrm{E}\, f_t) - \beta_2 f_t \\ f_t[R^e_{1t} - \beta_1(\lambda - \mathrm{E}\, f_t) - \beta_1 f_t] \\ f_t[R^e_{2t} - \beta_2(\lambda - \mathrm{E}\, f_t) - \beta_2 f_t] \end{bmatrix} = \mathbf{0}_{4\times 1}. \quad (3.41)$$

This gives 4 moment conditions, but only three parameters, so there is one overidentifying restriction to test—just as with (3.36).

3.3.3 What if the Factor is a Portfolio?

It would (perhaps) be natural if the tests discussed in this section coincided with those in Section 3.2 when the factors are excess returns. That is *almost* so. The difference is that we here estimate the $K \times 1$ vector λ (factor risk premia) as a vector of free parameters, while the tests in Section 3.2 *impose* $\lambda = \mathrm{E}\, f_t$. If we were to put this restriction on (3.40), then we are back to the LM test of the multifactor model where (3.40) specifies $n(1 + K)$ moment conditions, but includes only nK parameters (in β)—we gain one degree of freedom for every element in λ that we avoid to estimate. If we do not impose the restriction $\lambda = \mathrm{E}\, f_t$, then the tests are not identical and can be expected to be a bit different (in small samples, in particular).

3.3.4 Empirical Evidence

See Campbell, Lo, and MacKinlay (1997) 6.5 (Table 6.1 in particular) and Cochrane (2001) 20.2.

One of the more interesting studies is Fama and French (1993) (see also Fama and French (1996)). They construct 25 stock portfolios according to two characteristics of the firm: the size and the book-value-to-market-value ratio (BE/ME). In June each year, they sort the stocks according to size and BE/ME. They then form a 5×5 matrix of portfolios, where portfolio ij belongs to the ith size quantile *and* the jth BE/ME quantile.

They run a traditional CAPM regression on each of the 25 portfolios (monthly data 1963–1991)—and then study if the expected excess returns are related to the betas as they

should according to CAPM (recall that CAPM implies $\mathrm{E}\, R^e_{it} = \beta_i \lambda$ where λ is the risk premium (excess return) on the market portfolio). However, there is little relation between $\mathrm{E}\, R^e_{it}$ and β_i (see Cochrane (2001) 20.2, Figure 20.9). This lack of relation (a cloud in the $\beta_i \times \mathrm{E}\, R^e_{it}$ space) is due to the combination of two features of the data. First, *within a BE/ME quantile*, there is a positive relation (across size quantiles) between $\mathrm{E}\, R^e_{it}$ and β_i—as predicted by CAPM (see Cochrane (2001) 20.2, Figure 20.10). Second, *within a size quantile* there is a negative relation (across BE/ME quantiles) between $\mathrm{E}\, R^e_{it}$ and β_i—in stark contrast to CAPM (see Cochrane (2001) 20.2, Figure 20.11).

Fama and French (1993) also try a multi-factor model. They find that a three-factor model fits the 25 stock portfolios fairly well (two more factors are needed to also fit the seven bond portfolios that they use). The three factors are: the market return, the return on a portfolio of small stocks minus the return on a portfolio of big stocks (SMB), and the return on a portfolio with high BE/ME minus the return on portfolio with low BE/ME (HML). This three-factor model is rejected at traditional significance levels (see Campbell, Lo, and MacKinlay (1997) Table 6.1 or Fama and French (1993) Table 9c), but it can still capture a fair amount of the variation of expected returns (see Cochrane (2001) 20.2, Figures 20.12–13).

Chen, Roll, and Ross (1986) use a number of macro variables as factors—along with traditional market indices. They find that industrial production and inflation surprises are priced factors, while the market index might not be.

3.4 Fama-MacBeth

Reference: Cochrane (2001) 12.3; Campbell, Lo, and MacKinlay (1997) 5.8; Fama and MacBeth (1973)

The Fama and MacBeth (1973) approach is a bit different from the regression approaches discussed so far—although is seems most related to what we discussed in Section 3.3. The method has three steps, described below.

- First, estimate the betas β_i ($i = 1, \ldots, n$) from (3.24) (this is a time-series regression). This is often done on the whole sample—assuming the betas are constant. Sometimes, the betas are estimated separately for different sub samples (so we could let $\hat{\beta}_i$ carry a time subscript in the equations below).

- Second, run a cross sectional regression for every t. That is, for period t, estimate λ_t from the cross section (across the assets $i = 1, \ldots, n$) regression

$$R^e_{it} = \lambda_t' \hat{\beta}_i + \varepsilon_{it}, \tag{3.42}$$

  where $\hat{\beta}_i$ are the regressors. (Note the difference to the traditional cross-sectional approach discussed in Section 3.3, where the second stage regression regressed $\mathrm{E}\, R^e_{it}$ on $\hat{\beta}_i$, while the Fama-French approach runs one regression for every time period.)

- Third, estimate the averages

$$\hat{\varepsilon}_i = \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_{it} \text{ for } i = 1, \ldots, n, \text{ (across time, for every asset)} \tag{3.43}$$

$$\hat{\varepsilon}_t = \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_{it} \text{ for } t = 1, \ldots, T \text{ (across assets, for every period)} \tag{3.44}$$

The second step, using $\hat{\beta}_i$ as regressors, creates an errors-in-variables problem since $\hat{\beta}_i$ are estimated, that is, measured with an error. The effect of this is typically to bias the estimator of λ_t towards zero (and any intercept, or mean of the residual, is biased upward). One way to minimize this problem, used by Fama and MacBeth (1973), is to let the assets be portfolios of assets, for which we can expect that some of the individual noise in the first-step regressions to average out—and thereby make the measurement error in $\hat{\beta}$ smaller. If CAPM is true, then the return of an asset is a linear function of the market return and an error which should be uncorrelated with the errors of other assets—otherwise some factor is missing. If the portfolio consists of 20 assets with equal error variance in a CAPM regression, then we should expect the portfolio to have an error variance which is 1/20th as large.

We clearly want portfolios which have different betas, or else the second step regression (3.42) does not work. Fama and MacBeth (1973) choose to construct portfolios according to some initial estimate of asset specific betas. Another way to deal with the errors-in-variables problem is adjust the tests. Jagannathan and Wang (1996) and Jagannathan and Wang (1998) discuss the asymptotic distribution of this estimator.

We test the model by testing if all $\varepsilon_i = 0$ or if the average ε_i is zero. For the latter,

Fama and MacBeth (1973) suggest that the test should use the following estimator of the variance

$$\text{Var}\left(\frac{\sqrt{T}}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t\right)=\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_t-\bar{\hat{\varepsilon}}_t\right)^2, \tag{3.45}$$

where $\bar{\hat{\varepsilon}}_t$ is the sample average of ε_t, $\bar{\hat{\varepsilon}}_t = \Sigma_{t=1}^{T}\varepsilon_t/T$. Note that this is the standard formula for the variance of $\sqrt{T}$ times a sample average for an iid variable (which equals the variance of the series itself). It is, of course, unclear if that assumption is valid.

If, instead, we want to test if $\varepsilon_i = 0$ for all assets ($i = 1, \ldots, n$), we need the whole covariance matrix for the vector of ε_i. They are estimated under the same iid assumption as above

$$\text{Cov}\left(\frac{\sqrt{T}}{T}\sum_{t=1}^{T}\begin{bmatrix}\varepsilon_1\\ \vdots\\ \varepsilon_n\end{bmatrix}\right)=\frac{1}{T}\sum_{t=1}^{T}\left(\begin{bmatrix}\varepsilon_{1t}\\ \vdots\\ \varepsilon_{nt}\end{bmatrix}-\begin{bmatrix}\overline{\varepsilon_{1t}}\\ \vdots\\ \overline{\varepsilon_{nt}}\end{bmatrix}\right)\left(\begin{bmatrix}\varepsilon_{1t}\\ \vdots\\ \varepsilon_{nt}\end{bmatrix}-\begin{bmatrix}\overline{\varepsilon_{1t}}\\ \vdots\\ \overline{\varepsilon_{nt}}\end{bmatrix}\right)', \tag{3.46}$$

where $\overline{\varepsilon_{it}} = \Sigma_{t=1}^{T}\varepsilon_{it}/T$. Once we have this covariance matrix, we can test the hypothesis that $\varepsilon_i = 0$ for all assets ($i = 1, \ldots, n$) by the usual chi-square test.

Fama and MacBeth (1973) found, among other things, that the squared beta is not significant in the second step regression, nor is a measure of non-systematic risk.

A Coding of the GMM Problem in Section 3.2

This section describes how the GMM problem with the moment conditions (3.28) can be programmed.

A.1 Exactly Identified System

The moments are

$$m_t(\alpha,\beta)=\begin{bmatrix}1\\ f_t\end{bmatrix}\otimes\left(R_t^e-\begin{bmatrix}\alpha & \beta\end{bmatrix}\begin{bmatrix}1\\ f_t\end{bmatrix}\right).$$

We want to rewrite this on a form where the parameters are collected in a vector, so that we easily can apply matrix algebra to find the parameter estimates. To do that, notice the

following rules for Kronecker products: *(i)* $\text{vec}(ABC) = (C'\otimes A)\text{vec}(B)$; *(ii)* $(A\otimes B)' = A'\otimes B'$; and *(iii)* if a is $m\times 1$ and b is $n\times 1$, then $a\otimes b = (a\otimes I_n)b$. The first rule allows to write

$$I_n\begin{bmatrix}\alpha & \beta\end{bmatrix}\begin{bmatrix}1\\ f_t\end{bmatrix}=\underbrace{\left(\begin{bmatrix}1\\ f_t\end{bmatrix}'\otimes I_n\right)}_{x_t'}\underbrace{\text{vec}(\begin{bmatrix}\alpha & \beta\end{bmatrix})}_{b}.$$

The third rule allows us two rewrite the moments as

$$m_t(\alpha,\beta)=\underbrace{\left(\begin{bmatrix}1\\ f_t\end{bmatrix}\otimes I_n\right)}_{x_t}\left(R_t^e-\begin{bmatrix}\alpha & \beta\end{bmatrix}\begin{bmatrix}1\\ f_t\end{bmatrix}\right),$$

where the second rule helped us identify the leading term as the transpose of x_t' in the previous equation. We therefore have

$$m_t(\alpha,\beta)=\underbrace{\left(\begin{bmatrix}1\\ f_t\end{bmatrix}\otimes I_n\right)}_{x_t}\left(R_t^e-\underbrace{\left(\begin{bmatrix}1\\ f_t\end{bmatrix}'\otimes I_n\right)}_{x_t'}b\right).$$

To compare with the LS case, let $y_t = R_t^e$. We can then write the moment conditions as

$$\begin{aligned}m_t(\alpha,\beta) &= x_t\left(y_t - x_t'b\right)\\ &= x_t y_t - x_t x_t' b.\end{aligned}$$

In the exactly identified case, we can solve for the b vector from $\Sigma_{t=1}^{T}m_t/T = \mathbf{0}$. To do that, define

$$\Sigma_{xy}=\frac{1}{T}\sum_{t=1}^{T}x_t y_t \text{ and } \Sigma_{xx}=\frac{1}{T}\sum_{t=1}^{T}x_t x_t', \text{ so}$$
$$\Sigma_{t=1}^{T}m_t/T = \Sigma_{xy}-\Sigma_{xx}b = \mathbf{0} \text{ or } \hat{b}=\Sigma_{xx}^{-1}\Sigma_{xy}.$$

This looks very similar to the standard LS, but there is a hitch: x_t is an $n(1+K)\times n$ matrix, not a vector. It is still possible to compute the vector Σ_{xy} (dimension $(1+K)n\times 1$) and the matrix Σ_{xx} (dimension $(1+K)n\times(1+K)n$) efficiently, since they have very

simple structures. In particular, note that

$$\Sigma_{xx} = \frac{1}{T}\sum_{t=1}^{T} x_t x_t' = \frac{1}{T}\sum_{t=1}^{T}\left(\begin{bmatrix} 1 \\ f_t \end{bmatrix}\begin{bmatrix} 1 \\ f_t \end{bmatrix}'\right) \otimes I_n.$$

(This follows from the rule $(A \otimes B)(C \otimes D) = AC \otimes BD$, if conformable.) If we put $\begin{bmatrix} 1 \\ f_t \end{bmatrix}'$ in row t of the $T \times (1+K)$ matrix X, then

$$X'X = \sum_{t=1}^{T}\left(\begin{bmatrix} 1 \\ f_t \end{bmatrix}\begin{bmatrix} 1 \\ f_t \end{bmatrix}'\right),$$

we can compute Σ_{xx} as $(X'X/T) \otimes I_n$.

Also, note that

$$\Sigma_{xy} = \frac{1}{T}\sum_{t=1}^{T} x_t y_t = \frac{1}{T}\sum_{t=1}^{T} \text{vec}\left(R_t^e \begin{bmatrix} 1 \\ f_t \end{bmatrix}'\right).$$

(This follows from same rule as above and that $b \otimes a = \text{vec}(ab')$ if vectors.) If we put $R_t^{e\prime}$ in row t of the $T \times n$ matrix R, then

$$\text{vec}(R'X/T) = \frac{1}{T}\sum_{t=1}^{T} \text{vec}\left(R_t^e \begin{bmatrix} 1 \\ f_t \end{bmatrix}'\right) = \Sigma_{xy}.$$

A.2 Overidentified system

When the system is overidentified, then the moment conditions are of the form $m_t(\alpha, \beta) = z_t\left(y_t - x_t'b\right)$, where there are more elements in z_t than in x_t. The GMM approach is then to combine the moment conditions with a weighting matrix, W. To see how that works in this context, let

$$\Sigma_{zy} = \frac{1}{T}\sum_{t=1}^{T} z_t y_t \text{ and } \Sigma_{zx} = \frac{1}{T}\sum_{t=1}^{T} z_t x_t', \text{ so } \frac{1}{T}\sum_{t=1}^{T} m_t = \Sigma_{zy} - \Sigma_{zx}b.$$

The GMM loss function is then

$$\bar{m}'W\bar{m} = (\Sigma_{zy} - \Sigma_{zx}b)'W(\Sigma_{zy} - \Sigma_{zx}b),$$

and the first order conditions with respect to b are

$$-\Sigma_{zx}'W(\Sigma_{zy} - \Sigma_{zx}\hat{b}) = \mathbf{0},$$

so the GMM estimator is

$$\hat{b} = (\Sigma_{zx}'W\Sigma_{zx})^{-1}\Sigma_{zx}'W\Sigma_{zy}.$$

As an example of an overidentified system, consider the case where we restrict the intercepts to be zero, $\alpha = 0$. The moment conditions can then be written

$$m_t(\alpha, \beta) = \underbrace{\left(\begin{bmatrix} 1 \\ f_t \end{bmatrix} \otimes I_n\right)}_{z_t}\left(R_t^e - \underbrace{\left(f_t' \otimes I_n\right)}_{x_t'}b\right).$$

In this case, we have

$$\Sigma_{zx} = \frac{1}{T}\sum_{t=1}^{T} z_t x_t' = \frac{1}{T}\sum_{t=1}^{T}\left(\begin{bmatrix} 1 \\ f_t \end{bmatrix} f_t'\right) \otimes I_n.$$

It is straightforward to calculate Σ_{zx} efficiently by putting $\begin{bmatrix} 1 \\ f_t \end{bmatrix}'$ in row t of the $T \times (1+K)$ matrix Z and f_t' in row t of the $T \times K$ matrix X and the calculating

$$Z'X = \sum_{t=1}^{T}\left(\begin{bmatrix} 1 \\ f_t \end{bmatrix} f_t'\right), \text{ so } \Sigma_{zx} = (Z'X/T) \otimes I_n.$$

Finally, Σ_{zy} can be calculated as before, that is,

$$\text{vec}(R'X/T) = \Sigma_{xy}.$$

Bibliography

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay, 1997, *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey.

Chen, N.-F., R. Roll, and S. A. Ross, 1986, "Economic Forces and the Stock Market," *Journal of Business*, 59, 383–403.

Cochrane, J. H., 2001, *Asset Pricing*, Princeton University Press, Princeton, New Jersey.

Fama, E., and J. MacBeth, 1973, "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 71, 607–636.

Fama, E. F., and K. R. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.

Fama, E. F., and K. R. French, 1996, "Multifactor Explanations of Asset Pricing Anomalies," *Journal of Finance*, 51, 55–84.

Gibbons, M., S. Ross, and J. Shanken, 1989, "A Test of the Efficiency of a Given Portfolio," *Econometrica*, 57, 1121–1152.

Greene, W. H., 2000, *Econometric Analysis*, Prentice-Hall, Upper Saddle River, New Jersey, 4th edn.

Jagannathan, R., and Z. Wang, 1996, "The Conditional CAPM and the Cross-Section of Expectd Returns," *Journal of Finance*, 51, 3–53.

Jagannathan, R., and Z. Wang, 1998, "A Note on the Asymptotic Covariance in Fama-MacBeth Regression," *Journal of Finance*, 53, 799–801.

MacKinlay, C., 1995, "Multifactor Models Do Not Explain Deviations from the CAPM," *Journal of Financial Economics*, 38, 3–28.

4 Linear Factor Models: SDF vs Beta Methods

4.1 Linear SDF $\Leftrightarrow$ Linear Factor Model

Reference: Cochrane (2001) 6.3 and 13.5

4.1.1 General Case

Consider a linear model for the SDF

$$m_t = a + b'(f_t - \mathrm{E}\, f_t), \tag{4.1}$$

where f_t is a $K \times 1$ vector. This is equivalent to a linear factor model for the assets where the expected excess return of asset i can be written

$$\mathrm{E}\, R^e_{it} = \lambda' \beta_i, \tag{4.2}$$

where β_i is a $K \times 1$ vector with regression coefficients of R^e_{it} on f_t and λ is a $K \times 1$ vector of factor risk premia.

To see this, expand the asset pricing equation for excess return i, and substitute for m_t by using (4.1)

$$\begin{aligned} 0 &= \mathrm{E}(m_t R^e_{it}) \\ &= \mathrm{E}\, m_t\, \mathrm{E}\, R^e_{it} + \mathrm{Cov}(m_t,\, R^e_{it}) \\ &= a\, \mathrm{E}\, R^e_{it} + b'\, \mathrm{Cov}(f_t,\, R^e_{it}). \end{aligned} \tag{4.3}$$

Rewrite the last term by sticking in $\mathrm{Var}(f_t)\, \mathrm{Var}(f_t)^{-1}$ before the covariance matrix

$$\begin{aligned} 0 &= a\, \mathrm{E}\, R^e_{it} + b'\, \mathrm{Var}(f_t)\, \mathrm{Var}(f_t)^{-1}\, \mathrm{Cov}(f_t,\, R^e_{it}) \text{ or} \\ \mathrm{E}\, R^e_{it} &= \underbrace{-\frac{1}{a} b'\, \mathrm{Var}(f_t)}_{\lambda'} \underbrace{\mathrm{Var}(f_t)^{-1}\, \mathrm{Cov}(f_t,\, R^e_{it})}_{\beta_i}. \end{aligned} \tag{4.4}$$

If we know b in (4.1), then (4.4) shows how to construct λ so (4.2) holds. Conversely, if

we know λ in (4.2), then (4.4) shows how to construct b so (4.1) holds. (To identify a in (4.1) we need to use at least one raw (not excess) return, for instance, the riskfree return.)

To be precise, from (4.4) we have

$$\lambda = -\frac{1}{a}\operatorname{Var}(f_t)b, \tag{4.5}$$

since the variance matrix is symmetric. Both λ and b are $K \times 1$ vectors, so knowledge of one (and a) allows us to calculate the other.

A Discussion of λ and b

Consider (4.5) the simple case when there are two factors

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = -\frac{1}{a}\begin{bmatrix} \operatorname{Var}(f_{1t}, f_{1t}) & \operatorname{Cov}(f_{1t}, f_{2t}) \\ \operatorname{Cov}(f_{1t}, f_{2t}) & \operatorname{Var}(f_{2t}, f_{2t}) \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}. \tag{4.6}$$

If the factors are uncorrelated, then λ_j is just a scaling of b_j so they contain the same information. This is, of course, also true in the one factor case.

If the factors are correlated (the more likely case), then the relation is more complicated. In particular, we can have $\lambda_j \neq 0$, but $b_j = 0$, which means that the jth factor does not affect the SDF (and therefore not the pricing of assets) but it still has a factor risk premium. As an example, suppose f_t are excess return, so $\lambda = \mathrm{E}\, f_t$. It is quite possible that $\mathrm{E}\, f_{jt} \neq 0$, but that the SDF is not a function of it so f_{jit} has no importance for the pricing of assets. This may seem odd—but it should serve as a warning: we cannot understand a linear factor model by studying the factor risk premia only: what matters is the beta times the factor risk premia as in (4.2).

4.1.2 CAPM

Suppose the SDF is a linear function of the market excess return $m_t = a + bR^e_{mt}$. This makes sense, for instance, when the utility function is quadratic in wealth (or in consumption, provided that consumption is proportional to wealth) so that marginal utility is linear in the market return.

To derive the traditional CAPM formulation from that SDF, use $f_t = R^e_{mt}$ in (4.4) to

get

$$\mathrm{E}\, R^e_{it} = \underbrace{-\frac{1}{a}b'\operatorname{Var}(R^e_{mt})}_{\lambda}\underbrace{\operatorname{Var}(R^e_{mt})^{-1}\operatorname{Cov}(R^e_{mt}, R^e_{it})}_{\beta_i}, \tag{4.7}$$

which does not look like the CAPM relation at all. To rewrite it, note that (4.7) must apply to R^e_{mt} too

$$\mathrm{E}\, R^e_{mt} = -\frac{1}{a}b'\operatorname{Var}(R^e_{mt})\underbrace{\operatorname{Var}(R^e_{mt})^{-1}\operatorname{Cov}(R^e_{mt}, R^e_{mt})}_{1} \tag{4.8}$$

$$= -\frac{1}{a}b'\operatorname{Var}(R^e_{mt}) \tag{4.9}$$

$$= \lambda. \tag{4.10}$$

The key is, of course, that regressing R^e_{mt} on itself must give a coefficient equal to unity, that is, the beta of the market portfolio must be unity. We note that the factor risk premium must equal the excess return of the market, $\lambda = \mathrm{E}\, R^e_{mt}$. We can use this result to substitute for the λ term in (4.7)

$$\mathrm{E}\, R^e_{it} = \mathrm{E}\, R^e_{mt}\beta_i, \tag{4.11}$$

which is the traditional CAPM formulation.

4.1.3 Multi-Factor Model with Excess Returns as Factors

We now try the case where $m_t = a + b'f_t$, where f_t is a vector of excess returns. As in the CAPM case, (4.4) must apply to these excess returns too, which allows to to identify the vector of factor risk premia as $\lambda = \mathrm{E}\, f_t$.

This follows from the fact that the multiple regression of the jth element in f_t, f_{jt}, on the vector f_t gives a vector of regression coefficients with all zeros except that the jth element is one. Therefore, for this factor, (4.4) just picks out element j of λ.

Example 29 *Suppose we run the regression*

$$f_{1t} = a + \begin{bmatrix} f_{1t} \\ f_{2t} \end{bmatrix}'\beta + \varepsilon_{1t}.$$

The probability limit of $\hat{\beta}_{LS}$ is then (the probability limit allows us to cast this in terms of

the population moment—otherwise use the sample analogues)

$$\hat{\beta}_{LS} = \begin{bmatrix} \text{Var}(f_{1t}) & \text{Cov}(f_{1t},\, f_{2t}) \\ \text{Cov}(f_{1t},\, f_{2t}) & \text{Var}(f_{2t}) \end{bmatrix}^{-1} \begin{bmatrix} \text{Var}(f_{1t}) \\ \text{Cov}(f_{1t},\, f_{2t}) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

4.2 Estimating Explicit SDF Models

Reference: Cochrane (2001) 13; Dahlquist and Söderlind (1999)

4.2.1 The Mechanics of Estimating Linear SDF Models

To make the notation simpler, let the model of the SDF be

$$m_t = h_t'\gamma, \tag{4.12}$$

where the $(1 + K) \times 1$ vector h_t contains a constant and the other factors. If we let x_t be the payoffs and p_{t-1} the prices of the n assets, then the pricing equations are

$$\text{E}\, x_t m_t = \text{E}\, p_{t-1}, \tag{4.13}$$

which we will take to be the moment conditions. Combining gives the sample moment conditions

$$\frac{1}{T}\sum_{t=1}^{T} g_t(\gamma) = \bar{g}(\gamma) = \mathbf{0}_{n\times 1}, \text{ where} \tag{4.14}$$

$$g_t = x_t m_t - p_{t-1} = x_t h_t'\gamma - p_{t-1}. \tag{4.15}$$

To simplify the notation a bit let

$$S_{xh} = \frac{1}{T}\sum_{t=t}^{T} x_t h_t' \text{ and } S_p = \frac{1}{T}\sum_{t=t}^{T} p_{t-1}. \tag{4.16}$$

We can then write the moment conditions as

$$\bar{g}(\gamma) = S_{xh}\gamma - S_p. \tag{4.17}$$

We need to specify a symmetric and positive definite weighting matrix, W, in the GMM loss function if there are more moment conditions than parameters (if $n \geq 1 + K$).

The loss function is then

$$J = \bar{g}(\gamma)' W \bar{g}(\gamma) = \left(S_{xh}\gamma - S_p\right)' W \left(S_{xh}\gamma - S_p\right). \tag{4.18}$$

The first order conditions are

$$\begin{aligned} \mathbf{0}_{(1+K)\times 1} = \frac{\partial J}{\partial \gamma} &= \left(\frac{\partial \bar{g}(\hat{\gamma})}{\partial \gamma'}\right)' W \bar{g}(\hat{\gamma}) \\ &= S_{xh}' W \left(S_{xh}\hat{\gamma} - S_p\right), \text{ so} \\ \hat{\gamma} &= \left(S_{xh}' W S_{xh}\right)^{-1} S_{xh}' W S_p. \end{aligned} \tag{4.19}$$

This shows that it is easy to estimate the linear SDF model since it only involves iterating on the weighting matrix (if that is what you want to do) and to solve a system of linear equations in every iteration. (It is not obvious that you actually want to perform the explicit inversion in (4.19)—there are typically faster and numerically more robust methods of solving systems of linear equations).

Note that (4.19) is really a kind of GLS cross-sectional regression (LS if $W = I$) where the average prices $S_p = \Sigma_{t=1}^{T} p_{t-1}/T$ are regressed on the second moment of the payoff and factors, $S_{xh} = \Sigma_{t=1}^{T} x_t h_t'/T$. This makes sense since the pricing equations (4.19) says that it is the second moments that explain prices. Since the optimization problem is to minimize squared average price errors (possibly weighted if $W \neq I$), the result is a LS or GLS of average prices on their determinants.

4.2.2 The Problem with Excess Returns*

The approach above cannot be used directly if all the assets are excess returns. In that case $S_p = \mathbf{0}$, so $\hat{\gamma} = \mathbf{0}$, which must be wrong. The problem is that there is nothing to tie down the mean of the SDF. To handle this case, we need to impose a mean of the SDF or add moment conditions that tie down the mean, for instance, some "raw" returns instead of just the excess returns (a "riskfree" return might be a good choice).

If we choose to impose a mean of the SDF, then we return to the formulation in (4.1), where we choose a value for a. The sample moment conditions can then be written

$$\frac{1}{T}\sum_{t=1}^{T} g_t(\gamma) = \bar{g}(\gamma) = \mathbf{0}_{n\times 1}, \text{ where } g_t = x_t m_t - p_{t-1} = x_t a + x_t(f_t - \text{E}\, f_t)' b - p_{t-1}. \tag{4.20}$$

To simplify, let

$$S_x = \frac{1}{T}\sum_{t=t}^{T} x_t,\ S_{xf} = \frac{1}{T}\sum_{t=t}^{T} x_t(f_t - \mathrm{E}\, f_t)' \text{ and } S_p = \frac{1}{T}\sum_{t=t}^{T} p_{t-1}. \tag{4.21}$$

We can then write the moment conditions as

$$\bar{g}(b) = S_x a + S_{xf} b - S_p. \tag{4.22}$$

The GMM loss function is

$$J = \bar{g}(\gamma)' W \bar{g}(\gamma) = \left(S_x a + S_{xf} b - S_p\right)' W \left(S_x a + S_{xf} b - S_p\right), \tag{4.23}$$

with first order conditions

$$\begin{aligned} \mathbf{0}_{K\times 1} &= S_{xf}' W \left(S_x a + S_{xf}\hat{b} - S_p\right), \text{ so} \\ \hat{b} &= \left(S_{xf}' W S_{xf}\right)^{-1} S_{xf}' W \left(S_p - S_x a\right). \end{aligned} \tag{4.24}$$

This approach works even if the payoffs are excess returns, that is, when prices are zero so $S_p = \mathbf{0}$, provided we choose $a \neq 0$ (a zero value would not make sense anyway). We can also note that in this case, S_x equals the average excess returns, so the GMM estimation essentially amounts to regressing average excess returns on the second moments of the average excess returns with the factors.

Note that although $\hat{b}$ depends on the choice of a, the normalized fitted "pricing errors" of excess returns do not. To see that, set $p_{t-1} = S_p = \mathbf{0}$, and use (4.24) in (4.20) to get the fitted errors

$$\hat{g}_t = x_t a - x_t(f_t - \mathrm{E}\, f_t)' \left(S_{xf}' W S_{xf}\right)^{-1} S_{xf}' W S_x a. \tag{4.25}$$

This shows that all pricing errors (all periods, all asses) are scaled by a—so the choice of a does not affect how the pricing errors of different assets are traded off in the the optimization problem (note that all elements in $\hat{b}$ are scaled proportionally with a).

This last result also helps see that there is nothing to tie down the mean of the SDF when the payoffs are excess returns: if the moment conditions $\mathrm{E}\, g_t = 0$ are satisfied at one choice of a, then they are also satisfied at any other choice. In practice, this means

that the first order conditions of the loss function with respect to a

$$0 = S_x' W \left(S_x \hat{a} + S_{xf}\hat{b} - S_p\right) \tag{4.26}$$

is satisfied at any $\hat{a}$ if $S_p = \mathbf{0}$. To see that, substitute for $\hat{b}$ using (4.24) and note that we then get an equation where $\hat{a}$ is irrelevant.

4.3 SDF Models versus Linear Factor Models Again

Reference: Ferson (1995); Jagannathan and Wang (2002) (theoretical results); Cochrane (2001) 15 (empirical comparison).

The test of the linear factor model and the test of the linear SDF model are (generally) not the same: they test the same implications of the models, but in slightly different ways. The moment conditions look a bit different—and combined with non-parametric methods for estimating the covariance matrix of the sample moment conditions, the two methods can give different results (in small samples, at least).

There is one case where we know that the tests of the linear factor model and the SDF model are identical: when the factors are excess returns and the SDF is constructed to price these factors as well (see, Ferson (1995), Bekaert and Urias (1996), and Söderlind (1999)). Asymptotically, they are always the same, as showed by Jagannathan and Wang (2002).

To give an intuition for this result, consider the case when we have one asset and one factor (for instance the market excess return). The moment conditions for the linear factor model $R_{it}^e = \beta_i f_t + \varepsilon_{it}$ where $\mathrm{E}(\varepsilon_{it}) = \mathrm{E}(f_t \varepsilon_{it}) = 0$ are then

$$\mathrm{E}\begin{bmatrix} \varepsilon_{it} \\ \varepsilon_{it} f_t \end{bmatrix} = \mathrm{E}\begin{bmatrix} R_{it}^e - \beta_i f_t \\ (R_{it}^e - \beta_i f_t) f_t \end{bmatrix} = \mathbf{0}_{2\times 1}, \tag{4.27}$$

so we have two moment conditions and one parameter.

Suppose the SDF model is $m_t = 1 + b f_t$, where f_t is an excess return. The moment conditions for the asset pricing equations $\mathrm{E}\left(R_{it}^e m_t\right) = \mathrm{E}(f_t m_t) = 0$ are then

$$\mathrm{E}\begin{bmatrix} f_t m_t \\ R_{it}^e m_t \end{bmatrix} = \mathrm{E}\begin{bmatrix} f_t + b f_t^2 \\ R_{it}^e + b R_{it}^e f_t \end{bmatrix} = \mathbf{0}_{2\times 1}, \tag{4.28}$$

so we once again have two moment conditions and one parameter.

To show that testing (4.27) is the same as testing (4.28), suppose we put all the weight on the first moment condition and then test if the second moment condition is satisfied (we will soon the opposite). From the first line in (4.27) we get $\beta_i = \mathrm{E}\, R^e_{it} / \mathrm{E}\, f_t$, so the second line becomes

$$\mathrm{E}\, R^e_{it} f_t - \frac{\mathrm{E}\, R^e_{it}}{\mathrm{E}\, f_t} \mathrm{E}\, f_t^2 = u_{LF2}, \tag{4.29}$$

where we indicate the moment condition by u_{LF2}. Similarly, From the first line in (4.28) we get $b = -\mathrm{E}\, f_t / \mathrm{E}\, f_t^2$, so the second line becomes

$$\mathrm{E}\, R^e_{it} - \frac{\mathrm{E}\, f_t}{\mathrm{E}\, f_t^2} \mathrm{E}\, R^e_{it} f_t = u_{SDF2}. \tag{4.30}$$

Multiply by $-\mathrm{E}\, f_t^2 / \mathrm{E}\, f_t$

$$-\frac{\mathrm{E}\, f_t^2}{\mathrm{E}\, f_t} \mathrm{E}\, R^e_{it} + \mathrm{E}\, R^e_{it} f_t = -\frac{\mathrm{E}\, f_t^2}{\mathrm{E}\, f_t} u_{SDF2}, \tag{4.31}$$

which is the same as (4.29). This shows that deviation from the linear factor model, u_{LF2}, is proportional to the deviation from the SDF model—testing if the deviation is zero amounts to the same thing. It is easy to show that the same holds if we instead make the second moment condition hold exactly.

4.4 Conditional SDF Models

Reference: Cochrane (2001) 8; Ferson and Schadt (1996); Dahlquist and Söderlind (1999)

4.4.1 How to Incorporate Conditioning Information

There are several ways to incorporate conditioning information in the linear factor model.

Case 1: Factors Are Complicated Functions

The simplest, and perhaps most common way, of introducing conditional information is to simply state that the factors are not just the usual market indices or macro economic series: the factors are non-linear functions of them (this is sometimes called "scaled factors" to indicate that we scale the original factors with instruments). For instance, if R^e_{mt} is the

return on the market portfolio and z_{t-1} is something else which is thought to be important for asset pricing (use theory), then the factors could be

$$f_{1t} = R^e_{mt} \text{ and } f_{2t} = z_{t-1} R^e_{mt}. \tag{4.32}$$

An alternative interpretation of this is that we have only one factor, but that the coefficient of the factor is time varying, for instance, so the SDF is

$$\begin{aligned} m_t &= \gamma_1 R^e_{mt} + \gamma_2 z_{t-1} R^e_{mt} \\ &= a R^e_{mt} + b_t R^e_{mt}, \text{ where } b_t = \gamma_2 z_{t-1}. \end{aligned} \tag{4.33}$$

Case 2: Take the Asset Pricing Equation Seriously

The asset pricing relation really says that

$$p_{t-1} = \mathrm{E}_{t-1}\, x_t m_t, \tag{4.34}$$

where x_t is the payoff in t of an asset that costs p_{t-1} in $t-1$. This is hard to implement since moment estimators rely on using sample averages in place of the population expectation—and we certainly do not have repeated observations of period $t-1$.

However, (4.34) implies that, for any "instrument" z_{t-1} which is in the information set in t, we have

$$\mathrm{E}\, z_{t-1} p_{t-1} = \mathrm{E}\, z_{t-1} x_t m_t. \tag{4.35}$$

In particular, it certainly holds for $z_{t-1} = 1$ for all t, so (4.34) implies

$$\mathrm{E}\, p_{t-1} = \mathrm{E}\, x_t m_t, \tag{4.36}$$

which is a way of deriving the unconditional model. However, if we want to test more implications of the asset pricing model, then we could use (4.35) as moment conditions. This also has some intuitive appeal since we can think of $z_{t-1} x_t$ as the payoffs of managed portfolios, that is, portfolios whose size depends on the instruments z_{t-1}.

Many types of conditioning information has been used, for instance, lagged market returns and the slope of the yield curve. Some care must be taken in choosing the instruments, however. For instance, consider the case when the payoffs are returns, so (4.35)

can be written

$$\mathrm{E}\, z_{t-1} = \mathrm{E}\, z_{t-1} R_t m_t. \tag{4.37}$$

If z_{t-1} is orthogonal to $R_t m_t$, then this simplifies to

$$\mathrm{E}\, z_{t-1} = \mathrm{E}\, z_{t-1}\, \mathrm{E}\, R_t m_t \text{ or } 1 = \mathrm{E}\, R_t m_t, \tag{4.38}$$

which is the unconditional pricing equation once again. In theory, it cannot hurt to add several such moment conditions, but it creates a multicollinearity problem which can play havoc with the GMM estimators. The point is that instruments that uncorrelated with the returns or the pricing kernel are likely to be uninformative—and perhaps even worse.

4.4.2 Estimating Conditional SDF Models

The conditional SDF model can be estimated in the same way as the unconditional model (4.14)—if we just reinterpret the variables. Suppose we want to estimate the model (4.35) where the SDF is as in (4.12). Combing gives the moment conditions

$$\frac{1}{T}\sum_{t=1}^{T} g_t(\gamma) = \bar{g}(\gamma) = \mathbf{0}, \text{ where} \tag{4.39}$$

$$g_t = z_{t-1} x_t h_t' \gamma - z_{t-1} p_{t-1}. \tag{4.40}$$

This is just like the unconditional case (4.14)–(4.17), except that we have $z_{t-1}x_t$ and $z_{t-1}p_{t-1}$ instead of x_t and p_{t-1}—both prices and payoffs are scaled by the instrument z_{t-1}.

4.4.3 An Example of a Conditional SDF Model

Dahlquist and Söderlind (1999) evaluates mutual fund performance with conditional SDFs. In one case, the SDF is a linear function of five the returns on industry portfolios and a portfolio of short-term bonds (the variables in h_t), and the instruments (all lagged) are a detrended interest rate, the excess return on a bond portfolio, and the general stock market return (the variables in z_{t-1}). The test assets are a large set of Swedish mutual funds—and the main issue is whether the the SDF can price (conditionally) these mutual funds. If not, this could be interpreted as if the mutual funds have abnormal returns compared to passive portfolios formed by combining the six portfolios in the SDF. The results indicate

that most funds have positive, but insignificant, abnormal returns.

In an extension, the SDF is also allowed to be a function of the same six portfolios returns as before *and* products of these returns with some (other) instruments, x_{t-1}. The interpretation of this is that the mutual fund returns are then evaluated against dynamic portfolios, where the composition of these dynamic portfolios is changed linearly by the new instruments x_{t-1}. It can be shown that any mutual fund that would use any such strategy is assigned a neutral performance, and that abnormal performance requires something better than that.

Bibliography

Bekaert, G., and M. S. Urias, 1996, "Diversification, Integration and Emerging Market Closed-End Funds," *Journal of Finance*, 51, 835–869.

Cochrane, J. H., 2001, *Asset Pricing*, Princeton University Press, Princeton, New Jersey.

Dahlquist, M., and P. Söderlind, 1999, "Evaluating Portfolio Performance with Stochastic Discount Factors," *Journal of Business*, 72, 347–383.

Ferson, W. E., 1995, "Theory and Empirical Testing of Asset Pricing Models," in Robert A. Jarrow, Vojislav Maksimovic, and William T. Ziemba (ed.), *Handbooks in Operations Research and Management Science* . pp. 145–200, North-Holland, Amsterdam.

Ferson, W. E., and R. Schadt, 1996, "Measuring Fund Strategy and Performance in Changing Economic Conditions," *Journal of Finance*, 51, 425–461.

Jagannathan, R., and Z. Wang, 2002, "Empirical Evaluation of Asset Pricing Models: A Comparison of the SDF and Beta Methods," *Journal of Finance*, 57, 2337–2367.

Söderlind, P., 1999, "An Interpretation of SDF Based Performance Measures," *European Finance Review*, 3, 233–237.

5 Weighting Matrix in GMM

Reference: Cochrane (2001) 11.5-11.6.

5.1 Arguments for a Prespecified Weighting Matrix

- Robustness ($W = I \approx OLS$, $W = S^{-1} \approx GLS$). Problem with $W = S^{-1}$ if S is wrong. If you use the iterative approach to get the (approximately) efficient weighting matrix: always check point estimates in every iteration (large changes indicate potential problems).
- Near singularity of S: $W = S^{-1}$ perhaps not robust (see above). Even if S is correct, the loss function can be a bit counter intuitive (see below).
- When you know what you want to minimize, for instance, some economically interesting moment conditions.
- When you want to use one set of moment conditions to estimate the parameters, and then test if another set of moment conditions are satisfied at these parameter estimates.
- Want to compare (loss function value of) non-nested models. For instance, compare the loss function values from using two different candidate SDFs to price some assets. Note: this is an informal comparison.

Some of these points are discussed below.

5.2 Near Singularity of the Weighting Matrix

Let $\bar{g}_i$ be the sample average of the ith moment condition. Consider GMM estimation with two moment conditions and an optimal weighting matrix $W = S^{-1}$, where the

covariance matrix of the two moment conditions is

$$S = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \tag{5.1}$$

so both moment conditions have the same variance and the covariance (and here correlation) ρ.

The loss function is

$$J = \begin{bmatrix} \bar{g}_1 \\ \bar{g}_2 \end{bmatrix}' \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}^{-1} \begin{bmatrix} \bar{g}_1 \\ \bar{g}_2 \end{bmatrix} = \frac{1}{1-\rho^2}\left(\bar{g}_1^2 + \bar{g}_2^2 - 2\rho\bar{g}_1\bar{g}_2\right) \tag{5.2}$$

It is clear that this loss function punishes squared average pricing errors for each asset ($\bar{g}_1^2$ and $\bar{g}_2^2$), but it also puts a lot of weight on getting the two average pricing errors having the same signs (then $\bar{g}_1\bar{g}_2 > 0$) and the loss values is lower. As ρ approaches unity, this aspect is almost as important as getting the squared average pricing errors down. In fact, we can write

$$J = \frac{1}{1-\rho^2}\left[\bar{g}_1\left(\bar{g}_1 - \rho\bar{g}_2\right) + \bar{g}_2\left(\bar{g}_2 - \rho\bar{g}_1\right)\right], \tag{5.3}$$

so for ρ close to one, much of the effort in the minimization goes to make the difference between $\bar{g}_1 - \bar{g}_2$ as small as possible. Is that bad? Perhaps not if ρ is estimated with precision, but it can be a problem if ρ is overestimated. The parameters will then move towards making asset 1 and asset 2 equally bad.

5.3 Very Different Variances

We change the covariance matrix in (5.1) by making the two moment conditions uncorrelated and with variances σ_1^2 and σ_2^2, respectively. With the efficient weighting matrix, the loss function is

$$J = \begin{bmatrix} \bar{g}_1 \\ \bar{g}_2 \end{bmatrix}' \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}^{-1} \begin{bmatrix} \bar{g}_1 \\ \bar{g}_2 \end{bmatrix} = \frac{g_1^2}{\sigma_1^2} + \frac{g_2^2}{\sigma_2^2}. \tag{5.4}$$

This is the classical GLS case: the parameter estimates will tend to reflect the moment condition with the lowest variance. It is not self-evident that this is what you want. For instance, suppose the moment conditions are the pricing errors for two different financial

assets

$$\begin{bmatrix} \bar{g}_1 \\ \bar{g}_2 \end{bmatrix} = \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} R_{1t+1}m_{t+1} - 1 \\ R_{2t+1}m_{t+1} - 1 \end{bmatrix},$$

where m_t is a model SDF which depends on some variables and parameters (which you want to estimate), R_{it+1} is the ex post return in $t+1$ of asset i. Less volatile assets (small $\text{Var}(R_{it+1})$) are likely to have smaller variance of $R_{it+1}m_{t+1} - 1$ so the loss function will put a large weight on them. In contrast, assets with volatile ex post returns will not influence the estimates very much. This may not be what you want.

There is one clear case where we *want* to take the relative variability into account: when the moment conditions have different scales because of the way data is scaled. For instance, one return may be like 0.02 while another is like 2 (%). However, the perhaps best way to deal with this case is to scale the data before estimation.

5.4 $(\mathrm{E}\, x_t x_t')^{-1}$ as Weighting Matrix

Reference: Hansen and Jagannathan (1997) and Cochrane (2001) 11.5

Remark 30 *(Projection) Suppose we want to forecast the scalar y using the $k \times 1$ vector x and that we restrict the forecasting rule to be linear $\hat{y} = x'\beta$. This rule is a linear projection if β satisfies the orthogonality conditions* $\mathrm{E}[x(y - x'\beta)] = \mathbf{0}_{k\times 1}$*, that is, if* $\beta = (\mathrm{E}\, xx')^{-1}\mathrm{E}\, xy$*. A linear projection minimizes* $\mathrm{E}[y - k(y)]^2$ *within the class of linear $k(x)$ functions. Note that the orthogonality conditions means that*

$$y = x'\beta + \varepsilon, \text{ where } \mathrm{E}(x\varepsilon) = \mathbf{0}_{k\times 1}.$$

It may make sense to use $(\mathrm{E}\, x_t x_t')^{-1}$ as a weighting in GMM when the moment conditions define pricing errors from some model of the SDF. The reason is that the loss function is then (the square of) the smallest distance between the SDF used and the space of admissible SDFs (those that price the assets).

A problem, however, is that the second moment matrix might be close to singular (often more so than the covariance matrix of returns: $\mathrm{E}\, x_t x_t' = \mathrm{Cov}(x_t) + \mathrm{E}\, x\, \mathrm{E}\, x'$, where $\mathrm{E}\, x\, \mathrm{E}\, x'$ is an outer product and therefore has a rank of 1).

5.4.1 Projections of y_t and m_t onto the Space of Payoffs

This section establishes some preliminary results which we need later. Let x_t be an $n \times 1$ vector of payoffs used, with corresponding price in the previous period p_{t-1}. The payoff space, X_t, consists of all linear combinations of x_t. Let y_t be the SDF used in the model, which should be understood as a short hand notation for some function of data and parameter (which are estimated). As an example, with a consumption-based model with constant relative risk aversion we have $y_t = (C_t/C_{t-1})^{-\gamma}$, where the risk aversion γ is estimated.

Let x_t^* be the unique admissible SDF that is in the space of payoffs (the existence of this unique SDF follows from LOP)

$$x_t^* = x_t'\alpha \text{ and } p_{t-1} = \mathrm{E}_{t-1}\, x_t x_t^*. \tag{5.5}$$

Take unconditional expectations of both sides of the price equation to get

$$\begin{aligned} \mathrm{E}\, p_{t-1} &= \mathrm{E}\, x_t x_t^* && (5.6)\\ &= \mathrm{E}\, x_t x_t'\alpha, \text{ so} && (5.7)\\ \alpha &= (\mathrm{E}\, x_t x_t')^{-1}\mathrm{E}\, p_{t-1}. && (5.8) \end{aligned}$$

This defines an SDF that prices the assets (and which is in the space of payoffs). We can also note that any SDF (in the space of payoffs or not) that prices the assets, the admissible SDFs, can be written

$$m_t = x_t^* + \varepsilon_t = x_t'\alpha + \varepsilon_t, \tag{5.9}$$

where $\mathrm{E}\, \varepsilon_t x_t = 0$ (which implies $\mathrm{E}\, x_t m_t = \mathrm{E}\, x_t x_t^* = \mathrm{E}\, p_{t-1}$). Note that x_t^* is the projection of m_t onto the space of payoffs.

Consider the candidate SDF that we want to test, y_t. It can always be decomposed as

$$y_t = \text{ Proj}(y_t|X_t) + u_t = x_t'\beta + u_t \tag{5.10}$$

where $\mathrm{E}\, u_t x_t = 0$. To be precise, the projection is

$$\begin{aligned} \text{Proj}(y_t|X_t) &= x_t'\beta \text{ where} && (5.11)\\ \beta \text{ solves } \mathrm{E}[x_t(y_t - x_t'\beta)] &= \mathbf{0}_{n\times 1} \text{ so } \beta = (\mathrm{E}\, x_t x_t')^{-1}\mathrm{E}\, x_t y_t. && (5.12) \end{aligned}$$

5.4.2 GMM Loss Function Value = Distance between y_t and m_t

This section shows that the GMM loss function value with $(\mathrm{E}\, x_t x_t')^{-1}$ as the weighting matrix equals the distance between the candidate SDF and the closest point in the space of admissible SDFs. The squared distance between y_t and any admissible SDF, m_t, is defined as

$$\delta^2 = \mathrm{E}(y_t - m_t)^2, \tag{5.13}$$

which should be read as $\mathrm{E}[(y_t - m_t)^2]$. By using the previous definitions we can write (5.13) as

$$\delta^2 = \mathrm{E}[\mathrm{Proj}(y_t|X_t) + u_t - (x_t^* + \varepsilon_t)]^2 = \mathrm{E}[\mathrm{Proj}(y_t|X_t) - x_t^*]^2 + \mathrm{E}(u_t - \varepsilon_t)^2, \tag{5.14}$$

which follows from that u_t and ε_t are uncorrelated with x_t and therefore with both $\mathrm{Proj}(y_t|X_t)$ and x_t^* (which are linear combinations of x_t).

The distance in (5.13) or (5.14) is different for different choices of the admissible SDF, m_t. The choice of m_t that minimizes the distance is $m_t^0 = x_t^* + u_t$, that is, with $\varepsilon_t = u_t$ in (5.9). This is permissible since u_t satisfies $\mathrm{E}\, u_t x_t = 0$, so this m_t^0 indeed prices the assets. We have thus shown that

$$\begin{aligned}
\min_{m_t \text{ st } \mathrm{E} m_t x_t = \mathrm{E}\, p_{t-1}} \mathrm{E}(y_t - m_t)^2 &= \mathrm{E}[\mathrm{Proj}(y_t|X_t) - x_t^*]^2 \\
&= \mathrm{E}\left[x_t'(\mathrm{E}\, x_t x_t')^{-1}\, \mathrm{E}\, x_t y_t - x_t'(\mathrm{E}\, x_t x_t')^{-1}\, \mathrm{E}\, p_{t-1}\right]^2 \\
&= \mathrm{E}\left[x_t'(\mathrm{E}\, x_t x_t')^{-1}\, (\mathrm{E}\, x_t y_t - \mathrm{E}\, p_{t-1})\right]^2 .
\end{aligned} \tag{5.15}$$

To simplify, note the following remark.

Remark 31 *Let z and w be $n \times 1$ vectors. Then, $z'w$ is a scalar, so $(z'w)^2 = (z'w)'z'w = w'zz'w$.*

The expression with brackets in (5.15) fits this remark: $x_t'(\mathrm{E}\, x_t x_t')^{-1}$ corresponds to z'

and $(\mathrm{E}\, x_t y_t - \mathrm{E}\, p_{t-1})$ to w. Since $(\mathrm{E}\, x_t x_t')^{-1}$ is symmetric, we can then rewrite (5.15) as

$$\begin{aligned}
\min_{m_t \text{ st } \mathrm{E} m_t x_t = \mathrm{E}\, p_{t-1}} \mathrm{E}(y_t - m_t)^2 &= \mathrm{E}\left[(\mathrm{E}\, x_t y_t - \mathrm{E}\, p_{t-1})'\, (\mathrm{E}\, x_t x_t')^{-1} x_t x_t' (\mathrm{E}\, x_t x_t')^{-1}\, (\mathrm{E}\, x_t y_t - \mathrm{E}\, p_{t-1})\right] \\
&= (\mathrm{E}\, x_t y_t - \mathrm{E}\, p_{t-1})'\, (\mathrm{E}\, x_t x_t')^{-1}\, \mathrm{E}\, x_t x_t' (\mathrm{E}\, x_t x_t')^{-1}\, (\mathrm{E}\, x_t y_t - \mathrm{E}\, p_{t-1}) \\
&= (\mathrm{E}\, x_t y_t - \mathrm{E}\, p_{t-1})'\, (\mathrm{E}\, x_t x_t')^{-1}\, (\mathrm{E}\, x_t y_t - \mathrm{E}\, p_{t-1}) .
\end{aligned} \tag{5.16}$$

This shows that the (squared) distance between the candidate SDF, y_t, and the closest admissible SDF is the loss function of the GMM that uses $(\mathrm{E}\, x_t x_t')^{-1}$ as weighting matrix.

By choosing to compare y_t with this particular admissible SDF we give y_t its best chance of looking good. The idea is that all admissible SDFs are valid benchmarks, so we compare our candidate SDF, y_t, to the closest point in space of admissible stochastic discount factors (m_t^0).

5.4.3 GMM Loss Function Value = Max Relative Pricing Error of y_t

This section shows that the GMM loss function value with $(\mathrm{E}\, x_t x_t')^{-1}$ as the weighting matrix also equals the maximum relative pricing error (induced by the candidate SDF, y_t) among all portfolios of x_t.

Remark 32 *The Cauchy-Schwartz inequality is that for the random variables x and y we have $(\mathrm{E}\, xy)^2 \leq \mathrm{E}\, x^2\, \mathrm{E}\, y^2$.*

The error in the price of some asset i due to using the candidate SDF y_t is

$$\mathrm{E}\, x_{it} y_t - \mathrm{E}\, p_{it-1} = \mathrm{E}(y_t x_{it}) - \mathrm{E}(m_t x_{it}) = \mathrm{E}[(y_t - m_t)x_{it}]. \tag{5.17}$$

Note that the left hand side is one of the elements in (5.16).

From the Cauchy-Schwartz inequality we know that

$$(\mathrm{E}\, x_{it} y_t - \mathrm{E}\, p_{it-1})^2 = \{\mathrm{E}[(y_t - m_t)x_{it}]\}^2 \leq \mathrm{E}(y_t - m_t)^2\, \mathrm{E}\, x_{it}^2. \tag{5.18}$$

Divide by $\mathrm{E}\, x_{it}^2$ to get

$$\frac{(\mathrm{E}\, x_{it} y_t - \mathrm{E}\, p_{it-1})^2}{\mathrm{E}\, x_{it}^2} = \frac{\{\mathrm{E}[(y_t - m_t)x_{it}]\}^2}{\mathrm{E}\, x_{it}^2} \leq \mathrm{E}(y_t - m_t)^2. \tag{5.19}$$

As before, the right hand side is the squared distance between y_t and any admissible SDF, m_t. This expression says that the relative (to the payoff) squared pricing error induced by the candidate SDF, y_t, must be weakly smaller than the squared distance between y_t and any admissible SDF, m_t.

In particular, (5.19) must hold for the admissible SDF that minimizes the distance in (5.14), that is, for $m_t^0 = x_t^* + u_t$. We can now show (see below) that there is actually a portfolio of the original n assets (that is, a linear combination of x_t) with payoff x_{it} such that the weak inequality in (5.19) holds with equality if we use $m_t = m_t^0$. Since we have already shown that $\mathrm{E}(y_t - m_t^0)^2$ equals the value of the GMM loss function with $(\mathrm{E}\, x_t x_t')^{-1}$ as the weighting matrix, we have then shown that the GMM loss function value also equals the highest relative pricing error (due to the candidate SDF, y_t) among all portfolios of x_t.

To show this, note (see previous section) that we have

$$y_t - m_t^0 = \mathrm{Proj}(y_t|X_t) + u_t - (x_t^* + u_t) = \mathrm{Proj}(y_t|X_t) - x_t^* = x_{it}. \tag{5.20}$$

Now, pick the portfolio with payoff

$$x_{it} = \mathrm{Proj}(y_t|X_t) - x_t^*, \tag{5.21}$$

where $\mathrm{Proj}(y_t|X_t)$ is defined in (5.11) and x_t^* in (5.5). It is clearly possible to form this portfolio since both $\mathrm{Proj}(y_t|X_t)$ and x_t^* are linear combinations (portfolios) of the payoffs, x_t. Use (5.21) and (5.20) in (5.19) to see that this choice of x_{it} makes (5.19) with $m_t = m_t^0$ hold with equality.

We have thus shown that $\mathrm{E}(y_t - m_t^0)^2$ is the maximum of the squared pricing error relative to the second moment of the payoff that the candidate SDF induces. All other portfolios have a ratio that is lower than $\mathrm{E}(y_t - m_t^0)^2$.

Bibliography

Cochrane, J. H., 2001, *Asset Pricing*, Princeton University Press, Princeton, New Jersey.

Hansen, L. P., and R. Jagannathan, 1997, "Assessing Specification Errors in Stochastic Discount Factor Models," *Journal of Finance*, 52, 557–590.

6 Consumption-Based Asset Pricing

6.1 Introduction

This chapter studies the two main financial building blocks of simulation models: the consumption-based asset pricing model and the definition of asset classes. The aim is to discuss what different modelling choices imply for asset prices. For instance, what is the effect of different utility functions, investment technologies, monetary policies, and leverage on risk premia and yield curves?

The emphasis is on surveying existing models and discussing the main mechanisms behind the results. I therefore choose to work with stylized facts and simple analytical pricing expressions. There are no simulations or advanced econometrics in this chapter.[1] The following two examples should give the flavour. First, I use simple pricing expressions and scatter plots to show that the consumption based asset pricing model cannot explain the cross-sectional variation of Sharpe ratios. Second, I discuss how the slope of the real yield curve is driven by the autocorrelation of consumption by studying explicit log-linear pricing formulas of just two assets: a one-period bond and a one-period forward contract.

The plan of the chapter is as follows. Section 6.2 deals with the consumption-based asset pricing model. It studies if the model is compatible with historical consumption and asset data. From a modelling perspective, the implicit question is: if my model could create a realistic consumption process and has defined realistic assets (for instance, levered equity), would it then predict reasonable asset returns? Section 6.3 deals with how assets are defined in simulation models and what that implies for pricing. I discuss yield curves (real and nominal), claims on consumption (one-period and multi-period), options, and levered equity. Section 6.4 summarises the main findings. Technical details are found in a number of appendices at the end.

[1] See Rouwenhorst (1995) for a survey of simulation evidence; Ferson (1995) and Reffett and Schorfheide (2000) for surveys of careful econometric modelling.

6.2 Problems with the Consumption-Based Asset Pricing Model

This part of the chapter takes a hard look at the consumption-based asset pricing model since it is one of the building blocks in general equilibrium models. The approach is to derive simple analytical pricing expressions and to study stylized facts—with the aim of conveying the intuition for the results.

The first sections below look at earlier findings on the equity premium puzzle and the riskfree rate puzzle and study if they are stable across different samples (see the surveys of Bossaert (2002), Campbell (2001), Cochrane (2001), and Smith and Wickens (2002)). The following sections present new evidence on the ability of the consumption-based model to account for the predictability of returns and the cross-sectional variation (across portfolios formed on industries, firm size, dividend-price ratio, book-to-market ratio, and country). The last section discusses if alternative models of the stochastic discount factor (including Epstein-Zin utility, habit persistence, and idiosyncratic risk) can help solving the problems of the consumption-based model.

6.2.1 The Asset Pricing Equation

This section sets the stage for studying whether the consumption-based asset pricing model can fit the data. The basic pricing equation is specified and some stylized facts are presented.

The basic asset pricing equation says

$$\mathrm{E}_{t-1}(R_t M_t) = 1, \tag{6.1}$$

where R_t is the gross return of holding an asset from period $t-1$ to t, M_t is a stochastic discount factor (SDF). E_{t-1} denotes the expectations conditional on the information in period $t-1$, that is, when the investment decision is made. This equation holds for any asset that are freely traded without transaction costs (or taxes), even if markets are incomplete.[2]

[2]The existence of such an SDF is guaranteed if the law of one price holds (portfolios with the same payoffs have the same prices). The SDF is unique if, in addition, markets are complete (all risk can be insured). See, Harrison and Kreps (1979), Ross (1978), and Hansen and Richard (1987); for textbook treatments see LeRoy and Werner (2001) and Cochrane (2001).

It is convenient to rewrite (6.1) as[3]

$$\mathrm{E}_{t-1}(R_t) = 1/\mathrm{E}_{t-1}(M_t) - \mathrm{Cov}_{t-1}(R_t, M_t)/\mathrm{E}_{t-1}(M_t), \tag{6.2}$$

which expresses the conditional expectations of the return in terms of the conditional expectation of the SDF and the conditional covariance of the return and the SDF.[4]

Two applications of this equation will be particularly useful in the subsequent analysis. First, a (gross) riskfree rate, R_{ft}, has no conditional covariance with the stochastic discount factor (but it may still be a random variable), so (6.2) implies that the expected riskfree rate is the inverse of the expected SDF

$$\mathrm{E}_{t-1}(R_{ft}) = 1/\mathrm{E}_{t-1}(M_t). \tag{6.3}$$

Second, consider the gross returns on two assets (one could be a riskfree asset) and let R^e be the excess return, that is, the difference between the two returns. Apply (6.2) on both returns and take the difference to get

$$\mathrm{E}_{t-1}(R^e_t) = -\mathrm{Cov}_{t-1}(R^e_t, M_t)/\mathrm{E}_{t-1}(M_t), \tag{6.4}$$

which shows that a positive risk premium requires a negative correlation with the SDF.

In a consumption-based model, (6.1) is the Euler equation for optimal saving in $t-1$ where M_t is the ratio of marginal utilities in t and $t-1$, $M_t = \beta u'(C_t)/u'(C_{t-1})$.[5] I will focus on the case where the marginal utility of consumption is a function of consumption only, which is by far the most common formulation. This allows for other terms in the utility function, for instance, leisure and real money balances, but they have to be additively separable from the consumption term. With constant relative risk aversion (CRRA) γ, the stochastic discount factor is

$$M_t = \beta(C_t/C_{t-1})^{-\gamma}, \text{ so} \tag{6.5}$$

$$\ln(M_t) = \ln(\beta) - \gamma\Delta c_t, \text{ where } \Delta c_t = \ln(C_t/C_{t-1}). \tag{6.6}$$

The second line is only there to introduce the convenient notation Δc_t for the consumption

[3]$\mathrm{Cov}(x, y) = \mathrm{E}(xy) - \mathrm{E}(x)\,\mathrm{E}(y)$.

[4]A conditional variance, $\mathrm{Var}_{t-1}(x_t)$, is the variance of the forecast error $x_t - \mathrm{E}_{t-1}(x_t)$; a conditional covariance, $\mathrm{Cov}_{t-1}(x_t, y_t)$, is the covariance of the forecast errors $x_t - \mathrm{E}_{t-1}(x_t)$ and $y_t - \mathrm{E}_{t-1}(y_t)$.

[5]See Breeden (1979) and Lucas (1978) for early analyses of the consumption-based model.

growth rate.

The next few sections study if the pricing model consisting of (6.1) and (6.5) can fit historical data. To be clear about what this entails, note the following. First, general equilibrium considerations will not play any role in the analysis: the production side will not be even mentioned. Instead, the focus is on one of the building blocks of an otherwise unspecified model. Second, complete markets are not assumed. The key assumption is rather that the basic asset pricing equation (6.1) holds for the assets I analyse (broad US/international stock/bond portfolios). This means that the representative US investor can trade in these assets without transaction costs and taxes (clearly an approximation). Third, the properties of historical (ex post) data are assumed to be good approximations of what investors expected. In practice, this assumes both rational expectations and that the sample is large enough for the estimators (of various moments) to be precise.

My approach is to derive simple expressions which highlights the basic mechanisms and makes the analysis more easily adaptable to other data sets (sample periods/countries/assets). These simple expressions are obtained by making assumptions about the distribution of consumption and returns.

As an example of the pros and cons of this approach, suppose we want to analyse how different values of the risk aversion γ affects the unconditional mean of the SDF, $\mathrm{E}(M_t)$. If we make no assumptions about the distribution of consumption, then we must essentially take a data set on consumption growth, construct the variable on the right hand side of (6.5) for a value of γ and estimate its mean. We then repeat this for other values of γ and then describe (for instance, by plotting) how the mean varies with γ. This approach is cumbersome and the findings are hard to generalize. In contrast, by assuming that consumption growth is normally distributed, then we can express $\mathrm{E}(M_t)$ as a simple function of γ and the mean and standard deviation of consumption growth.[6] This second approach makes it much easier to interpret and extend the findings, although it comes at the price of only getting approximate answers (since data is probably not exactly normally distributed).

Table 6.1 shows the key statistics for quarterly US real returns and consumption growth for the sample periods 1947–2001 and 1970–2001. The means and standard deviations are expressed in percent, and have been annualized by multiplying the mean by four and the standard deviation by two (as would be appropriate for an annual sum of a

[6]The result is $\mathrm{E}(M_t) = \beta \exp[-\gamma\,\mathrm{E}(\Delta c_t) + \gamma^2\sigma(\Delta c_t)^2/2]$.

	Mean	Std	Sharpe ratio	Auto-corr	Correlation with Δc_t	Correlation with Δc_{t+1}
1947–2001						
Consumption growth, Δc_t	1.94*	1.10*		0.20*		0.20*
Stock market excess return	7.93*	15.98*	0.50*	0.05	0.14*	0.22*
Long gov bond excess return	1.15	9.38*	0.12	−0.04	−0.10	0.08
1-month T-bill	1.00*	1.57*		0.58*	0.13	0.27*
1970–2001						
Consumption growth, Δc_t	1.89*	0.91*		0.41*		0.41*
Stock market excess return	6.23*	17.87*	0.35*	0.01	0.16*	0.32*
Long gov bond excess return	3.06	11.43*	0.27	−0.08	−0.10	0.21*
1-month T-bill	1.56*	1.44*		0.69*	0.10	0.23*

Table 6.1: **US quarterly real returns and consumption growth (annualized, %).** This table shows summary statistics of US real returns and consumption growth rate for 1947Q2–2001Q4 and 1970Q1–2001Q4. A star (*) denotes significance at the 5% level, based on GMM/delta method inference, using a Newey-West estimator with one lag (see Appendix B). Consumption is real per capita consumption of services and non-durable goods. The consumption growth rate in quarter t is $\Delta c_t = \ln(C_t/C_{t-1})$, where C_t is the consumption level in quarter t. Returns are real returns. Excess returns are real returns in excess of the real return on a one-month Treasury Bill. To annualize, quarterly means are multiplied by 4 and quarterly standard deviations by 2, as appropriate if the variable is a random walk. See Appendix A for details on data sources and transformations.

quarterly iid variable). Two sample periods are used since it is of interest to see how stable the results are, but also because some of the international data used later in the section is only available from 1970.

The table uses a star (*) to indicate which numbers that are significantly different from zero at the 5% significance level. The tests are based on estimating the moments by GMM, using a Newey and West (1987) estimator for the sampling uncertainty of the moments, and then applying the delta-method (see Appendix B). This means that the tests are only valid asymptotically—but are consistent even if data is not iid (as assumed by standard tests).

We see, among other things, that consumption has a standard deviation of only 1% (annualized), the stock market has had an average excess return (over a T-bill) of 6–8%

(annualized), and that returns are only weakly correlated with consumption growth. These figures will be important in the following sections. Two correlations with consumption growth are shown, since it is unclear if returns should be related to what is recorded as consumption this quarter or the next. The reason is that consumption is measured as a flow during the quarter, while returns are measured at the end of the quarter.

6.2.2 The Equity Premium Puzzle

This section studies if the consumption-based asset pricing model can explain the historical risk premium on the US stock market.

To discuss the historical average excess returns, it is convenient to work with the unconditional pricing equation $\mathrm{E}(R_t M_t) = 1$, which follows from applying the law of iterated expectations on (6.1). We can then derive expressions like (6.2)–(6.4) for unconditional moments instead.

To highlight the basic problem with the consumption-based model and to simplify the exposition, I assume that the excess return, R_t^e, and the log SDF, $\ln(M_t)$, have a bivariate normal distribution.[7] Stein's lemma[8] then gives

$$\mathrm{Cov}(R_t^e, M_t) = \mathrm{Cov}[R_t^e, \ln(M_t)] \times \mathrm{E}(M_t). \tag{6.7}$$

We use this in the unconditional version of (6.4) to get the excess return and Sharpe ratio as

$$\mathrm{E}(R_t^e) = -\mathrm{Cov}[R_t^e, \ln(M_t)], \text{ and} \tag{6.8}$$

$$\mathrm{E}(R_t^e)/\sigma(R_t^e) = -\rho[R_t^e, \ln(M_t)] \times \sigma[\ln(M_t)], \tag{6.9}$$

where $\rho[R_t^e, \ln(M_t)]$ is the correlation of the excess return and the log SDF, and $\sigma(R_t^e)$ and $\sigma[\ln(M_t)]$ are the standard deviations of the excess return and log SDF respectively.

The log SDF in (6.6) is linear in consumption growth so the assumption of normality boils down to assuming that R_t^e and Δc_t have a bivariate normal distribution. We can then

[7] This does not imply (but is implied by) normally distributed gross returns, which would eventually violate limited liability. Even so, that would only happen once every 3000 years or so with the means and volatilities in Table 6.1. I just wonder if Psusennes I (Pharao 1044BC to 994BC) allowed limited liability. In any case, his gold mask (Egyptian Museum, Cairo) is second only to Tutanchamon's.

[8] Stein's lemma says that if x and y have a bivariate normal distribution and $h(y)$ is a differentiable function such that $\mathrm{E}[|h'(y)|] < \infty$, then $\mathrm{Cov}[x, h(y)] = \mathrm{Cov}(x, y)\,\mathrm{E}[h'(y)]$. This holds also for conditional distributions. In (6.7), $x = R_t^e$, $y = \ln(M_t)$, and $h(y) = \exp(y)$ so $h'(y) = M_t$.

write (6.8)–(6.9) as

$$\mathrm{E}(R_t^e) = \mathrm{Cov}(R_t^e, \Delta c_t)\gamma, \text{ and} \tag{6.10}$$

$$\mathrm{E}(R_t^e)/\sigma(R_t^e) = \rho(R_t^e, \Delta c_t) \times \sigma(\Delta c_t)\gamma, \tag{6.11}$$

This expression is convenient since it allows us to split the Sharpe ratio into three separate factors: the correlation of returns and consumption growth, the volatility of consumption growth, and the risk aversion. In particular, the first two factors can be estimated directly from data without knowing any preference parameters and the volatility is a constant in a cross-section (something we will use later).

Equation (6.11) shows that to get a positive expected excess return, the asset must be "risky" in the sense that it is positively correlated with consumption growth (negatively correlated with the SDF). In that case, the asset tends to give a high return when marginal utility is low, so investors demand a risk premium. This effect is strong if consumption is very volatile and/or investors are strongly risk averse.

Table 6.1 shows that the stock market has a Sharpe ratio around 0.35–0.5 (annualized) and that the consumption growth rate has a standard deviation of around $1\% = 0.01$ (annualized). The lowest value of γ which makes equation (6.11) hold is then $\gamma = 35$, and this would only work if the stock market were a worthless hedge against consumption fluctuations so the correlation is one.[9]

The basic problem with the consumption-based asset pricing model is that investors enjoy a fairly stable consumption series (either because income is smooth or because it is easy/inexpensive to smooth consumption by changing savings), so only an extreme risk aversion can motivate why investors require such a high equity premium. This is the *equity premium puzzle* stressed by Mehra and Prescott (1985) (although they approach the issue from another angle—more about that later on).

Still higher values of γ are needed if the correlation in (6.11) is lower than one—as it certainly is. Table 6.1 shows that the correlation is around 0.1–0.3, where the higher value is for the correlation with Δc_{t+1}. If we want to be nice to the consumption-based model, then we can use the higher value. With a Sharpe ratio of 0.35 (the lowest value,

[9] This is related to the approach in Hansen and Jagannathan (1991). Rewrite (6.4) as $\mathrm{E}(R_t^e)/\sigma(R_t^e) = -\rho(R_t^e, M_t)\sigma(M_t)/\mathrm{E}(M_t)$, which shows that the following "volatility bound" must hold: $\sigma(M_t)/\mathrm{E}(M_t) \geq |\mathrm{E}(R_t^e)|/\sigma(R_t^e)$. Their approach is essentially to search in a set of returns for the portfolio with the largest Sharpe ratio and then study if a given model of the SDF satisfies the volatility bound. In the case considered here there is only one asset.

once again to be nice) and a standard deviation of consumption of 0.01, (6.11) becomes $0.35 = 0.3 \times 0.01\gamma$, which requires $\gamma = 117$.

The low correlation therefore adds a *correlation puzzle* (see, for instance, Cochrane (2001)) to the equity premium puzzle: the consumption-based model makes the aggregate US stock market look pretty safe since almost all risk is non-systematic; it would indeed take a high risk aversion coefficient to motivate the historical risk premium.

If the consumption-based model cannot (with reasonable parameter values) explain the risk premium on the aggregate US stock market—what about other "aggregate" assets like US government bonds? Table 6.1 shows that the Sharpe ratio for bond returns is 0.12 for the long sample but as high as 0.27 for the shorter (and later) sample—but the sampling variability is so large so neither of these figures is significantly (on the 5% level) different from zero. In any case, if we use the point estimates we need a risk aversion coefficient (γ) of around 10 or 25 to make equation (6.11) hold—which looks much more promising than for equity. However, the correlation of the bond return and consumption growth is very low, so the equity premium puzzle is also a bond premium puzzle.

The Equity Premium Puzzle over Time

The results for the two samples in Table 6.1 are somewhat different, which immediately raises the question of how sensitive the conclusions are to the sample period. *Figure 6.1* provides some answers by plotting the key statistics for different samples.

Figure 6.1.a shows recursive estimates of the Sharpe ratio of the aggregate US stock market and the Sharpe ratio predicted by the CRRA model when the risk aversion is 100, that is, $\rho(R^e, \Delta c) \times \sigma(\Delta c) \times 100$ from (6.11). The correlations are for Δc_{t+1}. The recursive estimation means that the results for 1957Q2 use data for 1947Q2–1957Q2, the results for 1957Q3 add one data point, etc. The results for 2001Q4 are therefore the same as those in the top panel of Table 6.1. Figure 6.1.b shows the same statistics, but estimated on a moving data window of 10 years. For instance, the results for 1980Q2 are for the sample 1971Q3–1980Q2. Finally, Figure 6.1.c uses a moving data window of 5 years.

Together these figures give the impression that there are fairly long swings in the data. This fundamental uncertainty should serve as a warning against focusing on the fine details of the data. It could also be used as an argument for using longer data series—provided we are willing to assume that the economy has not undergone important regime changes.

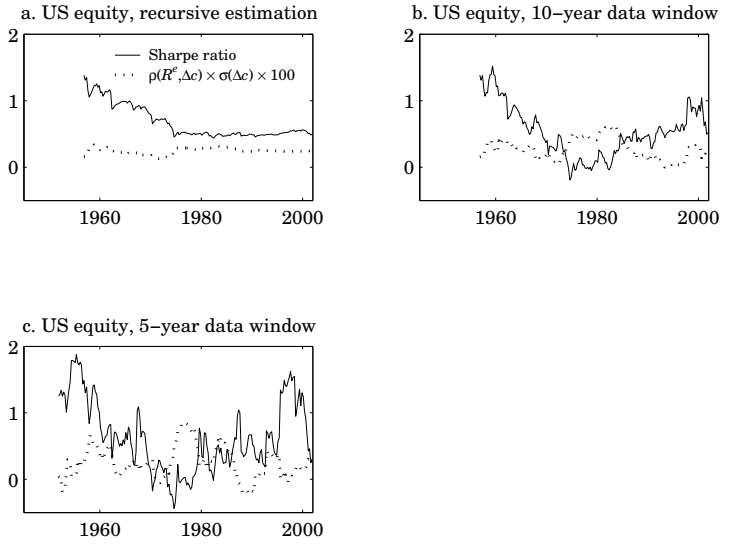


Figure 6.1: **The equity premium puzzle for different samples.** Subfigure a shows recursive estimations (longer and longer sample) of the Sharpe ratio of the aggregate US stock market and of the correlation with Δc_{t+1} times the standard deviation of Δc_{t+1}. Subfigures b and c show the same statistics calculated on a moving data window of 10 years and 5 years respectively. See Appendix A for details on data sources and transformations.

There are other interesting results in these figures. In particular, the actual Sharpe ratio and the prediction of the consumption-based model move in different directions—which is something we will return to later in discussing time-variation of risk premia (Section 6.2.5).

The Mehra-Prescott Approach

The Mehra and Prescott (1985) approach to analyzing the consumption-based asset pricing model is somewhat less direct than what we did above. However, it has had a large impact on the research community, so it is worth summarising.

The starting point is a Lucas (1978) economy where (aggregate) consumption equals

the exogenous (aggregate) endowment. To see how this works, write the gross return as $R_t = (P_t + D_t)/P_{t-1}$, where P_t is the asset price and D_t the dividend (per asset) in period t. We can then write (6.1) as $P_{t-1} = \mathrm{E}_{t-1}[M_t(P_t + D_t)]$. Iterating forward, applying the law of iterated expectations, ruling out bubbles, and using the CRRA SDF in (6.5) give

$$P_{t-1} = \mathrm{E}_{t-1}\left[\beta(C_t/C_{t-1})^{-\gamma} D_t + \beta^2 (C_{t+1}/C_{t-1})^{-\gamma} D_{t+1} + \ldots\right]. \tag{6.12}$$

The price of a real one-period bond is found by setting $D_t = 1$ and all future dividends to zero. The price of a claim to the stream of future non-storable endowments—which is interpreted as a proxy of a broad stock market index—is found by setting D_t equal to C_t. With a time series process for consumption, and values of β and γ, it is straightforward to calculate asset returns.

Mehra and Prescott (1985) construct a time series process by postulating that consumption growth follows a two-state Markov process: C_t/C_{t-1} is either low or high and the probability of which depends on whether C_{t-1}/C_{t-2} was low or high. This makes it very simple to calculate the expectations in (6.12): there is only one state variable and it can only take two values. In spite of this, the process allows both autocorrelation and time-varying volatility. The model parameters are calibrated to fit the mean, standard deviation, and autocorrelation in consumption growth. With this machinery, Mehra and Prescott (1985) find that a very high value of γ is needed to explain the historical equity premium—basically because consumption is fairly smooth.

This approach is interesting, but it is a fairly indirect method for studying the equity premium, and it adds a few very strong assumptions, in particular, about the consumption process and that the stockmarket corresponds to a claim on future consumption. In any case, several extensions have been made, for instance to investigate if more extreme consumption growth processes (fat tails or "crash states") can rescue the model—but so far with limited success (see, for instance, Salyer (1998) and Bidarkota and McCulloch (2003)).

6.2.3 The Riskfree Rate Puzzle

The CRRA utility function has the special feature that the intertemporal elasticity of substitution is the inverse of the risk aversion, that is, $1/\gamma$. The choice of the risk aversion parameter, for instance, to fit the equity premium, will therefore have direct effects on the

riskfree rate.

A key feature of any consumption-based asset pricing model, or any consumption/saving model for that matter, is that the riskfree rate governs the time slope of the consumption profile. From the asset pricing equation for a riskfree asset (6.3) we have $\mathrm{E}_{t-1}(R_{ft})\,\mathrm{E}_{t-1}(M_t) = 1$, which in the CRRA model becomes

$$\mathrm{E}_{t-1}(R_{ft})\,\mathrm{E}_{t-1}[\beta(C_t/C_{t-1})^{-\gamma}] = 1. \tag{6.13}$$

Note that we must use the conditional asset pricing equation —at least as long as we believe that the riskfree asset is a random variable. A riskfree asset is defined by having a zero conditional covariance with the SDF, which means that it is regarded as riskfree at the time of investment ($t-1$). In practice, this means a real interest rate (perhaps approximated by the real return on a T-bill since the innovations in inflation are small), which may well have a nonzero unconditional covariance with the SDF.[10] Indeed, in Table 6.1 the real return on a T-bill is as correlated with consumption growth as the aggregate US stockmarket.

To make progress, assume that the log SDF and the log riskfree rate $r_{ft} = \ln(R_{ft})$ have a bivariate normal distribution—which is somewhat different from our earlier assumption, but should (once again) be seen as just a an approximation. We can then write (6.13) as[11]

$$\mathrm{E}_{t-1}(r_{ft}) + \sigma^2_{t-1}(r_{ft})/2 = -\ln(\beta) + \gamma\,\mathrm{E}_{t-1}(\Delta c_t) - \gamma^2\sigma^2_{t-1}(\Delta c_t)/2. \tag{6.14}$$

To relate this equation to historical data, we take unconditional expectations to get

$$\mathrm{E}(r_{ft}) + \mathrm{E}\,\sigma^2_{t-1}(r_{ft})/2 = -\ln(\beta) + \gamma\,\mathrm{E}(\Delta c_t) - \gamma^2\,\mathrm{E}\,\sigma^2_{t-1}(\Delta c_t)/2. \tag{6.15}$$

Before we try to compare (6.15) with data, several things should be noted. First, the log gross rate is very close to a traditional net rate ($\ln(1+z) \approx z$ for small z), so it makes sense to compare with the data in Table 6.1. Second, we can safely disregard the variance terms since they are very small, at least as long as we are considering reasonable values of γ. Although the average conditional variances are not directly observable, we know that

[10] As a very simple example, let $x_t = z_{t-1} + \varepsilon_t$ and $y_t = z_{t-1} + u_t$ where ε_t are u_t uncorrelated with each other and with z_{t-1}. If z_{t-1} is observable in $t-1$, then $\mathrm{Cov}_{t-1}(x_t, y_t) = 0$, but $\mathrm{Cov}(x_t, y_t) = \sigma^2(z_{t-1})$.

[11] If $x \sim N(\mu, \sigma^2)$ and $y = \exp(x)$ then $\mathrm{E}(y^k) = \exp(k\mu + k^2\sigma^2/2)$.

they must be smaller than the unconditional variances[12], which are very small in Table 6.1. In fact, the variances are around 0.0001 whereas the means are around 0.01–0.02.

According to (6.15) there are two ways to reconcile a positive consumption growth rate with a low real interest rate (2%/year and 1%–1.5% respectively in Table 6.1): investors may prefer to consume later rather than sooner ($\beta > 1$) or they are willing to substitute intertemporally without too much compensation ($1/\gamma$ is high, that is, γ is low). However, fitting the equity premium requires a high value of γ, so investors must be implausibly patient if (6.15) is to hold. For instance, with $\gamma = 25$ (which is a very conservative guess of what we need to fit the equity premium) equation (6.15) says $0.01 = -\ln(\beta) + 25 \times 0.02$ (ignoring the variance terms) for the long sample, which requires $\beta \approx 1.6$. The shorter sample gives a very similar number. This is the *riskfree rate puzzle* stressed by Weil (1989). The basic intuition for this result is that it is hard to reconcile a steep slope of the consumption profile and a low compensation for postponing consumption if people are insensitive to intertemporal prices—unless they are extremely patient (actually, unless they prefer to consume later rather than sooner).

Another implication of a high risk aversion is that the real interest rate should be very volatile, which it is not. According to Table 6.1 the standard deviation of the real interest rate is 1.5 times the standard deviation of consumption growth. From (6.14) the volatility of the (expected) riskfree rate should be

$$\sigma[\mathrm{E}_{t-1}(r_{ft})] = \gamma\sigma[\mathrm{E}_{t-1}(\Delta c_t)], \tag{6.16}$$

if the variance terms are constant. The standard deviation of expected real interest rate is γ times the standard deviation of expected consumption growth. We cannot observe the conditional expectations directly, and therefore not estimate their volatility. However, a simple example is enough to demonstrate that high values of γ are likely to imply counterfactually high volatility of the real interest rate.

Suppose that consumption growth is an AR(1) process. In that case, $\sigma[\mathrm{E}_{t-1}(\Delta c_t)]$ equals $\rho(\Delta c_t, \Delta c_{t+1})\sigma(\Delta c_t)$, which from Table 6.1 is 0.22% for the long sample and 0.37% for the short sample.[13] With $\gamma = 25$, (6.16) then implies that the standard deviation of the expected real interest rate should be around 5%–9%. Table 6.1 shows that the

[12]Let $\mathrm{E}(y|x)$ and $\mathrm{Var}(y|x)$ be the expectation and variance of y conditional on x. The unconditional variance is then $\mathrm{Var}(y) = \mathrm{Var}[\mathrm{E}(y|x)] + \mathrm{E}[\mathrm{Var}(y|x)]$.

[13]If $x_t = \alpha x_{t-1} + \varepsilon_t$, where ε_t is iid, then $\mathrm{E}_{t-1}(x_t) = \alpha x_{t-1}$, so $\sigma(\mathrm{E}_{t-1}\, x_t) = \alpha\sigma(x_{t-1})$.

unconditional standard deviation, which is an upper bound (see footnote 12), is around 1.5%. This shows that an intertemporal elasticity of substitution of 1/25 is not compatible with the relatively stable real return on T-bills.

6.2.4 The Cross-Section of Returns

The previous section demonstrated that the consumption-based model has a hard time explaining the risk premium on a broad equity portfolio—essentially because consumption growth is too smooth to make stocks look particularly risky. However, the model *does* predict a positive equity premium, even if it is not large enough. This suggests that the model may be able to explain the relative risk premia across assets, even if the scale is wrong. In that case, the model would still be useful for some issues. This section takes a closer look at that possibility by focusing on the relation between the Sharpe ratio and the correlation with consumption growth in a cross-section of asset returns.

The key equation is (6.11), which I repeat here for ease of reading

$$\mathrm{E}(R_t^e)/\sigma(R_t^e) = \rho(R_t^e, \Delta c_t) \times \sigma(\Delta c_t)\gamma.$$

The volatility of the log SDF, $\sigma(\Delta c_t)\gamma$, is the same for all assets, so the cross-sectional variation in Sharpe ratios is due to cross-sectional variation in the correlation of the portfolio excess return and consumption growth, $\rho(R_t^e, \Delta c_t)$. We can therefore study the cross-sectional implication of (6.11) by estimating the linear regression equation

$$\mathrm{E}(R_t^e)/\sigma(R_t^e) = a_0 + a_1\rho(R_t^e, \Delta c_t) + u, \tag{6.17}$$

where the observations are different portfolios. According to (6.11) the slope coefficient a_1 should be positive (since $\sigma(\Delta c_t)\gamma$ is) and the coefficient of determination R^2 should be high (unity, except for sampling variability).

In a sense, this regression is similar to the traditional cross-sectional regressions of returns on factors with unknown factor risk premia (see, for instance, Cochrane (2001) chap 12 or Campbell, Lo, and MacKinlay (1997) chap 6).

Figure 6.2 studies different cross-sections of assets by plotting the Sharpe ratio against the correlation with consumption growth. To save some space, only results for Δc_{t+1} are shown. To be honest, the corresponding results for Δc_t look even less flattering. Figure 6.2.a shows evidence for 10 US industry portfolios (telecom, utilities, etc) for the sample

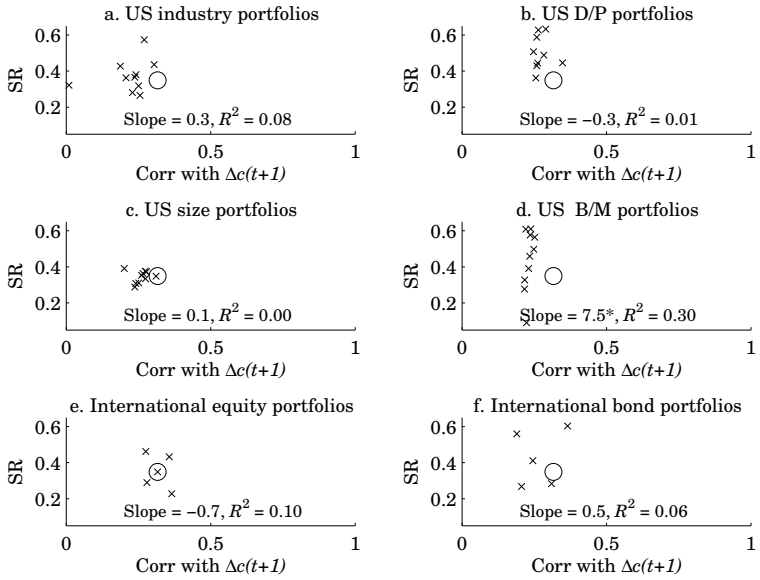


Figure 6.2: **Sharpe ratios and consumption correlation, 1970–2001.** Subfigure a shows Sharpe ratios and correlations with Δc_{t+1} for 10 US industry portfolios for 1970Q1–2001Q4. The large circle is the aggregate US stock market. The numbers at the bottom are slope coefficients and R^2 from a regression of the Sharpe ratio on a constant and the correlation. Signficance at the 5% level is marked by a star *. Subfigures b–d are similar to subfigure a but for US dividend/price, size, and book-to-market portfolios respecively. Subfigures e–f are for international equity and bond portfolios respectively. See Appendix A for details on data sources and transformations.

1970Q1–2001Q4; see Appendix A for details on data. There seems to be a weak positive relation between the Sharpe ratio and the correlation with consumption. This is confirmed by the numbers at the bottom of the figure: the slope coefficient in the estimated regression equation (6.17) is 0.3, but it is not significantly different from zero at the 5% level (would have been marked by a star * as in Figure 6.2.d) and the R^2 is only 0.08. The result is no better if the odd portfolio with zero correlation (oil, gas, and coal) is excluded.

Another way to see essentially the same result is to compare the industry portfolios

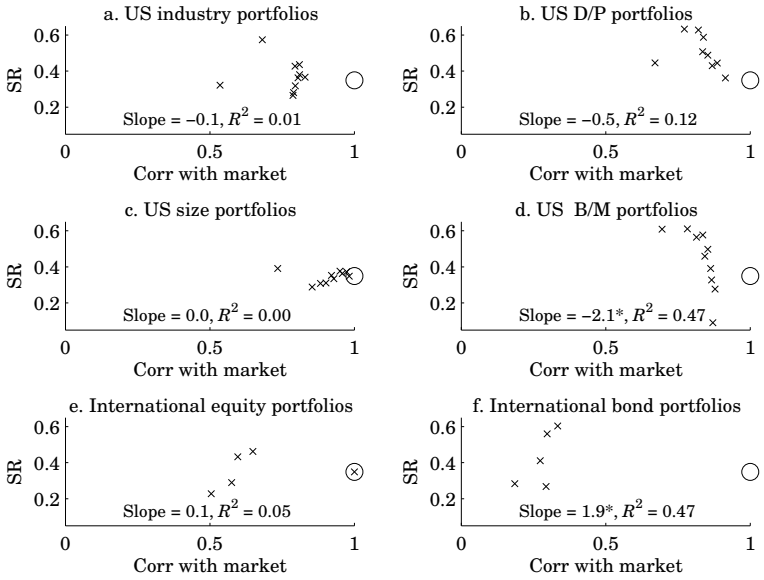


Figure 6.3: **Sharpe ratios and market correlation, 1970–2001.** These figures show Sharpe ratios plotted against the correlation with the aggregate US market excess return. See Figure 6.2 for details.

with the aggregate US market (marked by a large circle in the figure). In this figure, the market happens to have the highest correlation with consumption growth and should according to (6.11) therefore have the highest Sharpe ratio. In fact, six of the ten portfolios have higher Sharpe ratios, of which three are considerably higher.

There is also very little support for the model on US dividend/price and size portfolios (deciles of D/P and size respectively): the R^2s are virtually zero in Figures 6.2.b–c. The model seems to work somewhat better on US book-to-market portfolios (Figure 6.2.d), but fails again on the international equity and bond portfolios (Figures 6.2.e–f).

It is clear that the consumption-based model has problems with explaining the cross-section of Sharpe ratios. The main problem is that there are considerable differences in Sharpe ratios, but not in correlations with consumption growth. To put it bluntly:

systematic risk seems to be something else than covariance with consumption growth. However, the model is not a complete failure: 4 of 6 figures at least get a positive slope. To get a perspective on the magnitude of the problems, I will compare with the traditional Capital Asset Pricing Model (CAPM, Lintner (1965), Mossin (1968), and Sharpe (1964)).

The CAPM is most easily derived by postulating that the SDF is an affine function of the market excess return, R^e_{mt},

$$M_t = a + bR^e_{mt}. \tag{6.18}$$

(With a constant riskfree rate, this is the same as assuming an affine function of the market return.). Apply the unconditional version of (6.4) on the market excess return and some other excess return, R^e_t, and combine the results to get the standard CAPM expression

$$\mathrm{E}(R^e_t) = \frac{\mathrm{Cov}(R^e_t, R^e_{mt})}{\sigma^2(R^e_{mt})}\mathrm{E}(R^e_{mt}). \tag{6.19}$$

This says that the excess return on an asset is a regression coefficient (R^e_t regressed on R^e_{mt}) times the market excess return.

For a direct comparison with the consumption-based model in (6.11), it is more convenient to write (6.19) as

$$\mathrm{E}(R^e_t)/\sigma(R^e_t) = \rho(R^e_t, R^e_{mt}) \times \mathrm{E}(R^e_{mt})/\sigma(R^e_{mt}), \tag{6.20}$$

which says that the Sharpe ratio is a positive linear function (with a slope equal to the Sharpe ratio of the market, that is, around 0.5) of the correlation of the asset with the market.

Figure 6.3 illustrates (6.20). The figure has the same structure as Figure 6.2, except that the correlation is now with the excess return on the aggregate US market. These correlations are typically much higher than the correlations with consumption growth. However, it is hard to argue that the correlations with the stock market are much better at explaining the cross-section of Sharpe ratios: the slope coefficient is now positive in only 2 of 6 cases. In particular, CAPM gives odd results (negative coefficients) for the dividend-price and the book-to-market portfolios (see Fama and French (1993) for a detailed analysis of these portfolios).

The conclusion is that the consumption-based model is not good at explaining the cross-section of Sharpe ratios, but it is no worse than CAPM—if it is any comfort.

6.2.5 Time-Variation in Risk Premia

In contrast to the traditional interpretation of "efficient markets," it has been found that excess returns might be somewhat predictable—at least in the long run (a couple of years). In particular, Fama and French (1988a) and Fama and French (1988b) have argued that future long-run returns can be predicted by the current dividend-price ratio and/or current returns.

Table 6.2 illustrates this by showing the R^2 from the regressions

$$R^e_{t+k}(k) = a_0 + a_1 x_t + u_{t+k}, \text{ where } x_t = D_t/P_t \text{ or } R^e_t(k), \tag{6.21}$$

where $R^e_t(k)$ is the annualized k-quarter excess return of the aggregate US stock market and D_t/P_t is the dividend-price ratio (see Appendix A for details on data).

Horizon in quarters, (k)	1	4	8	12	16	20
1947–2001						
D_t/P_t	0.02*	0.10*	0.13*	0.16*	0.20*	0.27*
$R^e_t(k)$	0.00	0.03	0.03*	0.04	0.13*	0.07*
1970–2001						
D_t/P_t	0.01	0.04	0.01	0.00	0.00	0.00
$R^e_t(k)$	0.00	0.06*	0.05	0.00	0.03	0.03

Table 6.2: **Predicting long-run excess returns on the aggregate US stockmarket.** This table shows results from regressing future k-quarter returns, $R^e_{t+k}(k)$, on the current dividend-price ratio (D_t/P_t) or the current k-quarter return, $R^e_t(k)$. The samples are 1947Q2–2001Q4 and 1970Q1–2001Q4. A star (*) denotes a significance at the 5% level, based on GMM/delta method inference and a Newey-West estimator with $k-1$ lags. See Appendix A for details on data sources and transformations.

It seems as if the dividend-price ratio has some explanatory power for future returns—at least for long horizons in the long sample (first panel, 1947–2001). For instance, the R^2 is around 0.20–0.25 for the 4–5 year horizons. However, the effect vanishes in the late sample (second panel, 1970–2001). The lagged return is a fairly weak predictor in both samples.

This evidence suggests that excess returns may perhaps have a predictable component,

that is, that (ex ante) risk premia are changing over time. To see how that fits with the consumption-based model, assume that consumption growth is conditionally normally distributed. We then get a conditional Sharpe ratio as (compare with (6.11))

$$\mathrm{E}_{t-1}(R_t^e)/\sigma_{t-1}(R_t^e) = \rho_{t-1}(R_t^e, \Delta c_t) \times \sigma_{t-1}(\Delta c_t)\gamma. \tag{6.22}$$

By multiplying both sides by $\sigma_{t-1}(R_t^e)$ we see that the conditional expected excess return should equal the conditional covariance times the risk aversion.

It is clear from the earlier Figures 6.1.a–c that the consumption-based model probably cannot generate plausible movements in risk premia. In those figures the conditional moments in (6.22) are approximated by estimates on different data windows (that is, different subsamples). Although this is a crude approximation, the results are revealing: the actual Sharpe ratio and the prediction of the consumption-based model move in different directions on all frequencies. In Figure 6.1.a (longer and longer sample) the actual Sharpe ratio is decreasing over time, whereas the prediction is mostly increasing. In Figure 6.1.b the actual Sharpe ratio (on 10-year subsamples) reaches a minimum during the late 1970s and early 1980s—at the same time as the prediction is at maximum. The 5-year subsamples in Figure 6.1.c seems to pick up business cycle movements, and the two series clearly move in different directions.

6.2.6 Refinements of the Consumption-Based Model

There are many suggestions for how the problems with the consumption-based can be solved. One major strand proposes changes to the utility function; another to how we measure the volatility of returns and consumption.

If we focus on the utility function, we need a high risk aversion (to get a high equity premium), a high elasticity of intertemporal elasticity of substitution (to get a low and stable real interest rate), and maybe also time-variation of the risk aversion (to fit the time-variation in expected returns). This could be produced by using a functional form that explicitly separates risk aversion from the intertemporal elasticity of substitution as in Epstein and Zin (1989a) or by certain types of habit persistence models as in Campbell and Cochrane (1999).

If we instead focus on the measurement of volatility, we need a high consumption volatility (to get a high equity premium) and time-variation in the volatility of returns

and/or consumption (to get time-variation in expected returns). The first could be achieved by taking into account the effects of uninsurable idiosyncratic shocks as in Mankiw (1986); the second by a model of time-variation of second moments (typically a GARCH or regime switching model) as in, for instance, Bansal and Lundblad (2002).

There are of course many unresolved issues. First, some of the proposed "fixes" have serious side effects. For instance, some of the habit persistence models create implausible implications for consumption smoothing and optimal fiscal policy (see, Lettau and Uhlig (2000), Lettau and Uhlig (2002), and Ljunqvist and Uhlig (2000)).

Second, it is unclear if these modifications of the SDF can improve the consumption-based model's ability to account for the cross-section of Sharpe ratios discussed in Section 6.2.4, since the cross-sectional variation often depends on the correlation with consumption growth only—just as in the CRRA model. The next few sections demonstrate this for some well-known models of time nonseparable utility (Epstein and Zin (1989a)), habit persistence (Campbell and Cochrane (1999)), and idiosyncratic shocks (Constantinides and Duffie (1996)).

Epstein-Zin Utility

The basic idea of the recursive utility function in Epstein and Zin (1989b) is to form a certainty equivalent of future utility as $Z_t = [\mathrm{E}_t(U_{t+1}^{1-\gamma})]^{1/(1-\gamma)}$ where γ is the risk aversion—and then use a CES aggregator function to govern the intertemporal trade-off between current consumption and the certainty equivalent: $U_t = [(1-\delta)C_t^{1-1/\psi} + \delta Z_t^{1-1/\psi}]^{1/(1-1/\psi)}$ where ψ is the elasticity of intertemporal substitution.

Epstein and Zin (1989b) show that if all wealth is marketable, so the budget restriction can be written $W_{t+1} = R_{mt+1}(W_t - C_t)$ where R_{mt+1} is the market return (in a broad sense), then the Euler equation for an asset with return R_t is

$$1 = \mathrm{E}_{t-1}[\beta^\theta (C_t/C_{t-1})^{-\theta/\psi} R_{mt}^{\theta-1} R_t], \text{ where } \theta = (1-\gamma)/(1-1/\psi). \tag{6.23}$$

This gives rise to rather complicated pricing expressions, but it is straightforward to show that the basic mechanisms are as in the CRRA model, except that the elasticity of intertemporal substitution is no longer forced to be the inverse of the risk aversion.

To illustrate that (6.23) has (approximately) the same implications for risky assets as the CRRA model, assume that consumption-wealth ratio is constant (for instance, because

the market return is iid). This effectively closes down all interaction between the risk and intertemporal substitution (see Svensson (1989) and Campbell (1993)). It follows from the budget constraint that consumption growth and the market return are proportional so (6.23) can be written

$$\mathrm{E}_{t-1}[(C_t/C_{t-1})^{-\gamma} R_t] = \text{constant}, \tag{6.24}$$

which is exactly the same as the CRRA model. The Epstein-Zin model therefore inherits the problem with explaining the cross-section of Sharpe ratios. This is also approximately true when the consumption-wealth ratio is not too variable. Of course, if there are large predictable movements in the market return, then it is no longer true.

To see that the intertemporal substitution is governed by the parameter ψ, close down all systematic risk by assuming that the market return is a riskfree rate (set $R_{mt} = R_t = R_{ft}$ in (6.23)) and that log consumption and the log real interest rate, r_{ft}, have a bivariate normal distribution. We then get

$$\mathrm{E}_{t-1}(r_{ft}) + \theta\sigma^2_{t-1}(r_{ft})/2 = -\ln(\beta) + \frac{1}{\psi}\mathrm{E}_{t-1}(\Delta c_t) - \frac{\theta}{\psi^2}\sigma^2_{t-1}(\Delta c_t)/2. \tag{6.25}$$

This is the same relation between the real interest rate and consumption growth as in the CRRA case (see (6.14))—except that the elasticity of intertemporal substitution is no longer forced to be the inverse of the risk aversion.

This analysis shows that the Epstein-Zin model could help us to solve the riskfree rate puzzle (since we have a separate parameter for the elasticity of intertemporal substitution), but it can probably not do much with the equity premium (unless we are willing to accept a very high risk aversion) or the cross-sectional variation in Sharpe ratios.

Habit Persistence

The utility function in the habit persistence model of Campbell and Cochrane (1999) has the same functional form as the CRRA model, but the argument is the difference between consumption and a habit level, $C_t - X_t$, instead of just consumption. Variation in consumption and habit can both contribute to variation in marginal utility, which affects asset prices.

Campbell and Cochrane (1999) parameterize the habit in terms of the "surplus ratio" $S_t = (C_t - X_t)/C_t$, which measures how much aggregate consumption exceeds the habit. Since this ratio is external to the investor, marginal utility becomes $(C_t - X_t)^{-\gamma} =$

$(C_t S_t)^{-\gamma}$. We can therefore write the log SDF as

$$\ln(M_t) = \ln(\beta) - \gamma(\Delta s_t + \Delta c_t), \tag{6.26}$$

where s_t is the log surplus ratio. The process for s_t is assumed to be a non-linear AR(1)

$$s_t = \phi s_{t-1} + \lambda(s_{t-1})\Delta c_t + \text{constant}, \tag{6.27}$$

where $\lambda(s_{t-1})$ is a decreasing function of s_{t-1} which controls the sensitivity to (aggregate) consumption growth. The function $\lambda(s_{t-1})$ is always non-negative, so the log surplus ratio reacts positively to consumption growth (the main effect of letting λ depend on s_{t-1} is to make the risk premium move with the business cycle). Because of (6.27), a move in Δc_t will typically be accompanied by a move of Δs_t in the same direction, which increases the volatility of the SDF (6.26). In other words, a move in consumption drives the habit X_t in the opposite direction so $C_t - X_t$ varies a great deal.

Consumption growth is assumed to be conditionally normally distributed, so the conditional Sharpe ratio is (compare with (6.9))

$$\mathrm{E}_{t-1}(R^e_t)/\sigma_{t-1}(R^e_t) = -\rho_{t-1}[R^e_t, \ln(M_t)] \times \sigma_{t-1}[\ln(M_t)]. \tag{6.28}$$

From (6.26) and (6.27) we note that the innovation in the log SDF is $-\gamma[\lambda(s_{t-1}) + 1]$ times the innovation in Δc_t. Since $\lambda(s_{t-1})$ is positive and known in $t-1$ this is very similar to the CRRA model, except that the conditional volatility of the log SDF (the last term in (6.28)) is scaled up. This could help explain a higher risk premium. In a way, it is very similar to just increasing the risk aversion γ (although Campbell and Cochrane (1999) show that their model do not suffer from the some of the drawbacks of the CRRA model with a high γ).

However, (6.28) has the same implication for the cross-sectional variation of Sharpe ratios as the CRRA model. The reason is that the conditional correlation in (6.28) is the same as in the CRRA model since the innovation in the log SDF is proportional to the innovation in consumption growth (and the factor of proportionality cancels in the correlation).

Idiosyncratic Risk

A quite different attempt to solve the problems with the consumption-based asset pricing model is to argue that aggregating individual investors into a (fictitious) representative investor is a non-trivial task. Investors who face idiosyncratic risk that they cannot insure against (incomplete markets) may be reluctant to hold risky assets. Since macro data averages out idiosyncratic components, the volatility of aggregate consumption would then underestimate true consumption volatility of a representative investor.

However, it is surprisingly hard to make this story stick. The first problem is that only shocks that cannot be hedged qualify as idiosyncratic shocks: hedging means averaging out. One way of hedging consumption (if not income) is to have a savings account (see Telmer (1993)). Moreover, income can sometimes be partially hedged by trading in existing assets if they are correlated with income. In practice, this means that idiosyncratic shocks are probably a quite smaller than the income shocks estimated from micro data.

The second problem is that idiosyncratic shocks may still not matter unless their distribution is strongly tied to state of the economy. To develop an example, note that the basic asset pricing equation (6.1) must hold for every investor (after all, it is his first order condition), so $\mathrm{E}_{t-1}(R_t M_{jt}) = 1$ where M_{jt} is the SDF of investor j. The conditional version of the asset pricing equation with normally distributed consumption growth (6.10) for investor j is then

$$\mathrm{E}_{t-1}(R_t^e) = \mathrm{Cov}_{t-1}(R_t^e, \Delta c_{jt})\gamma, \tag{6.29}$$

where Δc_{jt} is consumption growth of investor j.

We let the innovation in the consumption investor j have two components: an aggregate innovation (common to all investors) and an idiosyncratic innovation. The idiosyncratic shock cannot be correlated with existing assets (discussed above), so it contributes nothing to the conditional covariance in (6.29): the risk premium will only depend on the covariance of the return with the aggregate innovation—as in the CRRA model. In this case the idiosyncratic shocks do not matter for the risk premium (although they will typically decrease the real interest rate as investors increase the precautionary saving).

This result is quite general. For instance, it holds for the Campbell and Cochrane (1999) model discussed before, and in any other model with a lognormal SDF where the idiosyncratic component affects the log SDF linearly (this follows directly from (6.8)).[14]

[14] Lettau (2002) shows that this property holds for any distribution of the consumption innovation if utility is CRRA and the the distribution of the idiosyncratic component is not affected by the aggregate shock.

To make idiosyncratic shocks matter for risk premia, they will either have to be something less than completely idiosyncratic (for instance, affecting a group of investors who collectively have market power; see Den Haan (2001)) or their distribution must depend on the aggregate shock (see Mankiw (1986) and Constantinides and Duffie (1996)).

In the latter class of models, a higher risk premium requires that the volatility of the idiosyncratic shocks is stochastic and that the innovations to this volatility and of aggregate consumption are negatively correlated: bad times mean more risk. To illustrate this, consider the CRRA model where the log SDF for investor j is

$$\ln(M_{jt}) = \ln(\beta) - \gamma \Delta c_{jt}. \tag{6.30}$$

Let the innovation in consumption growth for investor j be

$$\Delta c_{jt} - \mathrm{E}_{t-1}(\Delta c_{jt}) = \varepsilon_t + u_{jt}, \tag{6.31}$$

where ε_t is the aggregate component and u_{jt} the individual component which has a variance which depends on the realization of ε_t.

With (6.30)–(6.31) the asset pricing equation for investor j says that

$$\begin{aligned} 0 &= \mathrm{E}_{t-1}(R_t^e M_{jt}) && (6.32)\\ &= \mathrm{E}_{R,\varepsilon,u}\left\{R_t^e \exp\left[\ln(\beta) - \gamma\,\mathrm{E}_{t-1}(\Delta c_t) - \gamma\varepsilon_t - \gamma u_{jt}\right]\right\}. && (6.33) \end{aligned}$$

The operator $\mathrm{E}_{R,\varepsilon,u}$ is used instead of E_{t-1} to indicate that the expectations involves integrating over the random variables R_t^e, ε_t, and u_{jt}. Suppose that the distribution of u_{jt} conditional on the aggregate innovation, ε_t, is normal with zero mean and variance $v_{t-1} + \lambda(\varepsilon_t)$ where v_{t-1} is known in $t-1$ and $\lambda(\varepsilon_t)$ depends on the aggregate shock. Integrating over u_{jt} then gives

$$0 = \mathrm{E}_{R,\varepsilon}\left\{R_t^e \exp\left[\ln(\beta) - \gamma\,\mathrm{E}_{t-1}(\Delta c_t) + \gamma^2 v_{t-1} - \gamma\varepsilon_t + \gamma^2\lambda(\varepsilon_t)/2\right]\right\}, \tag{6.34}$$

where $\mathrm{E}_{R,\varepsilon}$ indicates that we now need to integrate over R_t^e and ε_t. (This first integration works since the asset return is independent of the realization of the idiosyncratic shock, even if not of its variance.)

I follow the

approximation in Lettau (2002) of assuming that ε_t and $\lambda(\varepsilon_t)$ have a bivariate normal distribution. This is only an approximation since it does not rule out negative variances, but it has the advantage of giving a very straightforward expression. The term in square brackets in (6.34) can then be interpreted as a normally distributed log SDF of the investor, which allows us to use the decomposition of the Sharpe ratio in (6.28) which I repeat here for easy of reading

$$\mathrm{E}_{t-1}(R_t^e)/\sigma_{t-1}(R_t^e) = -\rho_{t-1}[R_t^e, \ln(M_t)] \times \sigma_{t-1}[\ln(M_t)].$$

In this model, the conditional volatility and correlation are

$$\sigma_{t-1}[\ln M_t)] = \gamma\sigma_{t-1}[\varepsilon_t - \gamma\lambda(\varepsilon_t)/2] \tag{6.35}$$

$$-\rho_{t-1}[R_t^e, \ln(M_t)] = \rho_{t-1}[R_t^e, \varepsilon_t - \gamma\lambda(\varepsilon_t)/2]. \tag{6.36}$$

Note that setting $\lambda(\varepsilon_t) = 0$ gives the same expressions as without idiosyncratic shocks. In fact, if $\lambda(\varepsilon_t)$ is a constant, that is, if the volatility of the idiosyncratic shocks do not depend on the aggregate shock, then we also get the same expressions as without idiosyncratic shocks.

If $\lambda(\varepsilon_t)$ is decreasing in ε_t so idiosyncratic risk is larger in bad times, then the volatility in (6.35) is larger than without idiosyncratic shocks. This could help explaining the equity premium puzzle—although there are doubts about whether the idiosyncratic risk is sufficiently volatile to be quantitatively important (see Lettau (2002) and Cogley (1998)).

However, this mechanism is unlikely to help the model to explain the cross-sectional variation in Sharpe ratios—which depends only on the correlation (6.36): as a first-order approximation this correlation is the same as when $\lambda(\varepsilon_t)$ is constant. For instance, if $\lambda(\varepsilon_t) = a + b\varepsilon_t$ with $b < 0$, then this is exactly so. This gives the same implications for the cross-sectional dispersion of Sharpe ratios as the model without idiosyncratic shocks—at least unless the idiosyncratic dispersion is a very non-linear function of the aggregate state.

An additional aspect of models with idiosyncratic risk is that they allow us to study the effects of transaction costs in a serious way: heterogenous agents with idiosyncratic costs trade, representative agents do not. The first effect of transaction costs is to transform the first order condition of an investor to a set of inequalities $1/(1+\tau) \leq \mathrm{E}_{t-1}(R_t M_{jt}) \leq 1+\tau$ where τ is the (proportional) cost of a round-trip (the upper limit is for buying the

asset in $t-1$ and selling in t; the lower limit is for the opposite). This has been used to modify the Hansen and Jagannathan (1991) bounds (see He and Modest (1995) and Luttmer (1996)) with mixed results. One of the key assumptions in that analysis seems to be the assumption length of the period: a 0.5% transaction cost is quite substantive compared to a 1% monthly (expected) return, but small compared to a 12% annual return. We therefore need to find the equilibrium (including the endogenous trading frequency) in order to analyse the importance of transaction costs. So far, the findings indicate that the effect on prices is relatively small (unless borrowing/lending is very costly) but that the effect on turnover is large (see Heaton and Lucas (1996) and Vayanos (1998)).

6.3 Assets in Simulation Models

This part of the chapter discusses how different types of assets are incorporated into simulation models. The emphasis is on discussing the key properties of the assets when they are priced by a consumption-based model.

I first analyse a number of different asset classes: short- and long-lived claims on aggregate consumption, options and levered claims, as well as short and long real/nominal interest rates. To economise on space, I leave out some important asset classes like foreign exchange and real estate (see, for instance, Smith and Wickens (2002), Kollman (2001), Zimmermann (1999) and Hendershott (1998)). At the end, I discuss some other important features of how assets are modelled: time-variation in risk premia, the effect of non-Walrasian labour contracts, and the effect of adding an asset to an existing model.

The analysis has the same basic approach as in the first part of the chapter. First, the model consists of the basic asset pricing equation (6.1) and the consumption-based stochastic discount factor (6.5). In practice, this means that we are studying equilibrium conditions on the joint behaviour of consumption growth and returns on assets that can be freely traded. This does not assume that markets are complete. However, adding an asset to a model with incomplete markets may change the consumption process, so the results we derive are only valid for assets that could be traded when the equilibrium was determined (see Section 6.3.8 for a more detailed discussion). Second, analytical pricing relations are derived by making assumptions about the distribution of returns and consumption and sometimes also by approximating complicated expressions.

Although this approach is useful for providing intuition for how different types of

assets will work in a simulation model, it is clearly only a first step. In particular, this chapter never formulates and solves a full model. To do that, the other contributions to this book, the references found below, and the text by Altug and Labadie (1994) should all be very useful.

6.3.1 Preliminaries

Many simulation models are close to log-linear and are typically fed with normally distributed shocks. This means that most variables (including returns) are approximately lognormally distributed—sometimes exactly so. I therefore derive new expressions for risk premia and Sharpe ratios (the expressions in previous sections assumed normally, not lognormally, distributed returns). These expressions will be used repeatedly in the following sections.

The asset pricing equation with the CRRA SDF is $1 = \mathrm{E}_{t-1}\{\beta \exp[-\gamma \Delta c_t + r_t]\}$, where $r_t = \ln(R_t)$ is the log return. If these variables have a bivariate normal distribution, it follows (see footnote 11) that the risk premium and Sharpe ratios are

$$\mathrm{E}_{t-1}(\tilde{r}_t^e) = \mathrm{Cov}_{t-1}(\gamma \Delta c_t, r_t^e), \text{ and} \tag{6.37}$$

$$\mathrm{E}_{t-1}(\tilde{r}_t^e)/\sigma_{t-1}(r_t^e) = \rho_{t-1}(\Delta c_t, r_t^e) \times \sigma_{t-1}(\Delta c_t)\gamma. \tag{6.38}$$

The tilde is here to indicate that the risk premium is defined as $\mathrm{E}_{t-1}(\tilde{r}_t^e) = \mathrm{E}_{t-1}(r_t^e) + \sigma_{t-1}^2(r_t)/2$, that is, the expected excess log return over the real interest rate, adjusted for Jensen's inequality. This definition of an excess return is perhaps slightly non-standard, but it captures all the aspects of interest here.

In the following sections I will make use of the fact that only innovations matter for the risk premia, since it is only innovations that are risky. For the log excess return, this means that only the innovation in the risky asset payoff matters.[15]

6.3.2 Claims on Aggregate Consumption

This section analyses the pricing of claims on aggregate consumption. Many models have considered such assets—sometimes as an approximation of equity (see, for instance, Lucas (1978) and the references in Section 6.2.2). The main purpose of the discussion is

[15]The log excess return, r_t^e is the log payoff minus the lagged log assset price (measured in real terms) minus the log real interest rate. Only the first component has an innovation.

to highlight that the risk premium on long-lived consumption claims depends crucially on the autocorrelation of consumption growth.

We first discuss a one-period claim and then move on to multiperiod claims. The real price in $t-1$ of a claim on aggregate consumption in t satisfies the following asset pricing equation for CRRA utility (see (6.12))

$$P_{ct-1} = \mathrm{E}_{t-1}[\beta(C_t/C_{t-1})^{-\gamma} C_t] \text{ or} \tag{6.39}$$

$$\ln(P_{ct-1}) = \ln(C_{t-1}) + \ln(\beta) + (1-\gamma)\,\mathrm{E}_{t-1}(\Delta c_t) + (1-\gamma)^2 \sigma_{t-1}^2(\Delta c_t)/2, \tag{6.40}$$

where the second line exploits lognormality. The payoff in t is obviously C_t, so the return is C_t/P_{ct-1}. Instead of evaluating this by brute force (it is not very hard, just messy), we use (6.37). Note that the innovation in the excess log return of the consumption claim, r_{ct}^e, equals the innovation in consumption growth so (6.37) gives

$$\mathrm{E}_{t-1}(\tilde{r}_{ct}^e) = \sigma_{t-1}^2(\Delta c_t)\gamma. \tag{6.41}$$

This asset has the highest possible Sharpe ratio ($\sigma_{t-1}(\Delta c_t)\gamma$) since the return is perfectly negatively correlated with the SDF—this is yet another version of the Hansen and Jagannathan (1991) bound. This means that if the model cannot generate a high Sharpe ratio for the one-period consumption claim, then it cannot do so for any other asset either.

Most simulation models have rather focused on long-lived claims on a consumption stream, that is, a portfolio of different claims. What is the expected return on holding such a portfolio between two periods (the holding period return)?

It is easy to get the basic intuition by considering a claim that gives aggregate consumption in t and $t+1$. When traded in $t-1$ this is a portfolio of two claims: a claim on C_t (discussed above) and a claim on C_{t+1}. I will call the the latter a "C_{t+1} strip." The time lines of these assets are shown in Figure 6.4. The holding period return (between $t-1$ and t) on this portfolio is clearly the sum of the holding period returns on its two components.

We already know the (holding period) return on the one-period claim from (6.41), so I now turn to the C_{t+1} strip. This strip becomes a one-period claim in t and its price will then be like in (6.40) but with time subscripts advanced one period

$$\ln(P_{ct}) = \ln(C_t) + \ln(\beta) + (1-\gamma)\,\mathrm{E}_t(\Delta c_{t+1}) + (1-\gamma)^2 \sigma_t^2(\Delta c_{t+1})/2. \tag{6.42}$$

Claim on C_t:	buy claim	get C_t	
Claim on C_{t+1}:	buy claim	keep or sell claim	if kept, get C_{t+1}

Figure 6.4: **Time line of $t-1$ claims on C_t and C_{t+1}.**

This is the only term in the excess log holding period return (between $t-1$ and t) with an innovation. Applying (6.37) therefore gives the expected holding period return (in excess of a real interest rate) of the strip

$$\mathrm{E}_{t-1}(\tilde{r}^e_{ht}) = \gamma\sigma^2_{t-1}(\Delta c_t) + \gamma(1-\gamma)\,\mathrm{Cov}_{t-1}[\Delta c_t, \mathrm{E}_t(\Delta c_{t+1})], \tag{6.43}$$

provided Jensen's inequality term is constant (or at least uncorrelated with the innovation in consumption growth).

The second term in (6.43) is zero if consumption growth is not autocorrelated[16] or if the utility function is logarithmic $\gamma = 1$. In this case, the strip has the same risk premium as the one-period claim in (6.41). This should come as no surprise since lack of autocorrelation makes time irrelevant and log utility makes the discount rate effect perfectly balance the "dividend" effect of any news.[17]

With positive autocorrelation in consumption (as Table 6.1 suggests), the strip has a lower expected return than the one-period claim—provided the risk aversion is larger than unity. The intuition is as follows. The price in (6.42) is high when expected consumption growth, $\mathrm{E}_t\,\Delta c_{t+1}$, is low: although the expected payoff is low, it will be worth a lot since it comes at a time of scarcity (with log utility these effects cancel out). Now, if $\mathrm{E}_t\,\Delta c_{t+1}$ is low, then period t consumption growth, Δc_t, is typically also low (they are positively correlated). This means that the strip generates a high holding period return in t (a capital

[16] $\mathrm{Cov}_{t-1}[\Delta c_t, \mathrm{E}_t(\Delta c_{t+1})] = \mathrm{Cov}_{t-1}(\Delta c_t, \Delta c_{t+1})$ since $\Delta c_{t+1} - \mathrm{E}_t(\Delta c_{t+1})$ is not correlated with anything known in t.

[17] Abel (1999) assumes iid consumption growth, but reintroduces autocorrelation in the SDF through habit persistence.

gain) when consumption is scarce (high marginal utility): it is a fairly good hedge and will therefore not require a high risk premium. It is even conceivable that it could have a negative risk premium.

This extends to longer-lived claims on streams of consumption. Such assets will have a fairly low (high) risk premium if consumption is positively (negatively) autocorrelated for most leads/lags (see Campbell (1986)).

6.3.3 Options and Levered Claims

This section discusses options and option-like assets. Fairly few simulation models have studied traditional options directly, but some models have dealt with option-type assets like levered equity. However, simulation models could be a useful framework for studying and solving option pricing problems. I will illustrate this by considering a very simple case: a European call option that expires next period.

On the expiry date, the owner of a European call option has the right (not the obligation) to buy a prespecified asset for a prespecified (strike) price, X. If the asset payoff happens to be Y_t, then the payoff of the option is zero if $Y_t < X$ (the option is not exercised) and $Y_t - X$ otherwise.

With the CRRA SDF, the price of the call option must be

$$\text{Call option price}_{t-1} = \mathrm{E}_{t-1}[\beta(C_t/C_{t-1})^{-\gamma}\max(0, Y_t - X)]. \tag{6.44}$$

If consumption growth and the log asset payoff, $\ln(Y_t)$, have a bivariate normal distribution, then the solution to this equation is the Black and Scholes (1973)/Merton (1973) formula (see Huang and Litzenberger (1988) or Söderlind and Svensson (1997) for straightforward and not at all tedious calculations). This result is quite natural when we compare with the dynamic hedging approach typically used for deriving the Black-Scholes formula, since both approaches fundamentally rely on lognormality of the asset payoff.[18] This framework has been used by, for instance, Söderlind (2003) to study the implications of monetary policy on bond options in a simple analytical real business cycle model.

To get intuition for the results on option and option-like assets, it is convenient to

[18] The result is easily derived by using the following facts: if $x \sim N(\mu, \sigma^2)$ and $y = \exp(x)$ then $\mathrm{E}(y^k) = \exp(k\mu + k^2\sigma^2/2)$ and $\mathrm{E}(y^k|y > a) = \mathrm{E}(y^k)\Phi(r\sigma - a_0)/\Phi(-a_0)$, where $a_0 = [\ln(a) - \mu]/\sigma$; see Johnson, Kotz, and Balakrishnan (1994).

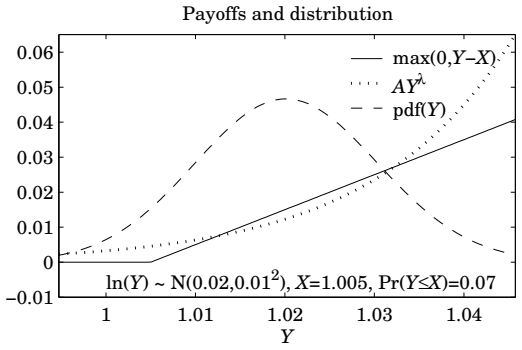


Figure 6.5: **Example of option payoff and power approximation.** This figure shows the payoff of a call option with strike price X and a power approximation. The value of the underlying asset is the random variable y with probability density function pdf(y).

approximate the nonlinear payoff $\max(0, Y_t - X)$ with a power function

$$Z_t = Ay_t^{\lambda}, \text{ with } \lambda = 1/[1 - X/\,\mathrm{E}_{t-1}(Y_t)] \text{ for } X < \mathrm{E}_{t-1}(Y_t). \tag{6.45}$$

Abel (1999) shows that this approximation gives a coefficient of variation (standard deviation divided by the mean) of the payoff which is very close to the option payoff—at least for reasonably low values of X (so the probability of exercise is high). This approximate payoff is particularly convenient when Y_t is lognormally distributed since the lognormality carries over to Ay_t^{λ}: if $\ln(Y_t)$ is normally distributed, so is $\ln(A) + \lambda \ln(Y_t)$.

Figure 6.5 illustrates how (6.45) works for a payoff with the same distribution as US annual consumption growth (mean of 2% and a standard deviation of 1%), where the exercise price, X, is such that there is a 93% chance of exercise. The distribution of Y_t is also shown (it is lognormal, but looks almost like a normal distribution since the mean is so far from zero). The constant A in the power function payoff is chosen so that both payoffs have the same means of around 1.5%. It is straightforward to calculate (using the facts in footnote 18) that both the option payoff and the approximation have standard deviations close to 1%.We now use this approximation to study the implications of firm debt.

Many simulation models assume that the equity value equals the full value of a firm. However, most firms are financed by both equity and debt. Debt-to-assets (leverage) ratios vary a good deal, but values around 0.5 seems to be fairly typical for historical US data. Since equity is not a claim on the full value of the firm, only the residual claim, it has option-like features which is likely to make it more risky.

To illustrate the main effect of leverage, I compare the one-period claim on consumption discussed in Section 6.3.2 with a levered claim whose payoff is $\max(0, C_t - X)$ where X is the debt. Leverage reduces the payoff by subtracting a fixed amount (the debt), except that limited liability means that the payoff cannot be negative. A zero payoff (a return of -100%) is awfully bad for investors, so the levered claim will be considered very risky.

From (6.45) the levered claim has approximately the payoff AC_t^{λ}. This means that the innovation in the log payoff (which is what matters for the risk premium) is λ times the innovation in unlevered claim. It then follows directly from (6.41) that the risk premium of the levered claim, $\mathrm{E}_{t-1}(\tilde{r}^e_{zt})$, is λ times the risk premium on the unlevered claim

$$\mathrm{E}_{t-1}(\tilde{r}^e_{zt}) = \lambda \sigma^2_{t-1}(\Delta c_t)\gamma. \tag{6.46}$$

The levered claim has a higher risk premium since it is more volatile ($\lambda > 1$) , but the Sharpe ratio is the same.

If we interpret $X/\,\mathrm{E}_{t-1}(Y_t)$ as the leverage ratio which is (on average) around 0.5, then Abel's suggestion in (6.45) gives $\lambda = 2$. With this number the risk premium on levered claims should be twice as high as the risk premium on unlevered claims.

It is more difficult to model levered long-lived claims, since the claim can only go "bankrupt" once, so numerical methods are called for. A few simulation models have incorporated leverage and arrived at somewhat different conclusions. For instance, Rouwenhorst (1995) finds a very small effect of leverage, while Jermann (1998) finds that it adds around one percentage point to the equity premium. The main reason for the difference seems to be that the firm value is very stable in Rouwenhorst's model.

6.3.4 The Real Yield Curve

This section studies the yield curve of real interest rates. The price of a k-period real bond (with a unit payoff k periods ahead) must satisfy the asset pricing equation (see (6.12))

$$B_{kt-1} = \mathrm{E}_{t-1}[\beta^k (C_{t+k-1}/C_{t-1})^{-\gamma}]. \tag{6.47}$$

1-period rate:	pay B_{1t-1}	get 1	
2-period rate:	pay B_{2t-1}		get 1
Forward rate:	write contract	pay B_{2t-1}/B_{1t-1}	get 1

Figure 6.6: **Time line of how to get different interest rates.**

I will consider two interest rates that can be calculated from this expression. First, the one-period (log) interest rate $r_{1t-1} = \ln(1/B_{1t-1})$, is the return on holding a real bond from $t-1$ to t. Note that (according to convention) it carries the date of the investment (when the interest rate is agreed on), not the date of payoff as for most other assets. By exploiting the lognormality we get

$$r_{1t-1} = -\ln(\beta) + \gamma \, \mathrm{E}_{t-1}(\Delta c_t) - \gamma^2 \sigma^2_{t-1}(\Delta c_t)/2. \tag{6.48}$$

Real interest rates are time-varying only if expected consumption growth is (disregarding time-variation in Jensen's inequality term).

Second, we can create a portfolio of one-period and two-period real (inflation-indexed) bonds to guarantee the investor a known (as of $t-1$) interest rate on an investment between t and $t+1$, that is, a real forward rate.

The forward-spot parity for an asset without intermediate dividends says that the contracted forward price (to be paid next period) is

$$\text{Forward price} = \frac{\text{Spot price}}{\text{Price of bond maturing next period}}.$$

The intuition is that the forward contract is like buying the asset today, but on credit.

The "spot way" of making sure that we get one unit (of the basket of goods) in $t+1$ is to buy a two-period bond in $t-1$ at the price B_{2t-1}. The price of a bond maturing next period is B_{1t-1}, so the forward price is B_{2t-1}/B_{1t-1}. This is illustrated in Figure 6.6.

We then define the forward rate as the rate of return on this investment: since the bond

pays off one unit in $t+1$ and the investment was B_{2t}/B_{1t} in t, the gross forward rate is B_{1t-1}/B_{2t-1}. From (6.47) we calculate the log forward rate, $f_{t-1} = \ln(B_{1t-1}/B_{2t-1})$, as

$$f_{t-1} = -\ln(\beta) + \gamma \, \mathrm{E}_{t-1}(\Delta c_{t+1}) - \gamma^2 \sigma^2_{t-1}(\Delta c_{t+1})/2 - \gamma^2 \, \mathrm{Cov}_{t-1}(\Delta c_t, \Delta c_{t+1}). \tag{6.49}$$

Both the short interest rate and the forward rate can vary over time. In a one-factor model of the yield curve (see, for instance, Vasicek (1977), Cox, Ingersoll, and Ross (1985), and Backus, Foresi, and Zin (1998)) they move in parallel. In (6.48)–(6.49) parallel movements require that $\mathrm{E}_{t-1}(\Delta c_t)$ and $\mathrm{E}_{t-1}(\Delta c_{t+1})$ are affine functions of a single variable, *the* factor. For instance, in the special case where expected consumption is an AR(1) with positive autocorrelation, then the real interest rates would actually follow a discrete version of the Vasicek (1977) model. That holds approximately in many simple real business cycle models, since their dynamics are well captured by a single dynamic state variable (see Cogley and Nason (1993)). This may sound awfully restrictive, but from the perspective of modelling the real yield curve, it may be acceptable. It still a matter of debate of whether the one-factor model can be rejected or not.

We now turn to studying the *average slope of the yield curve*, that is, the average over time. The risk premium is the forward rate (6.49) minus the expected (as of $t-1$) future short rate (6.48) (but with time subscripts forwarded one period)

$$f_{t-1} - \mathrm{E}_{t-1}(r_{1t}) = -\gamma^2 \, \mathrm{Cov}_{t-1}(\Delta c_t, \Delta c_{t+1}) + \gamma^2 [\sigma^2_t(\Delta c_{t+1}) - \sigma^2_{t-1}(\Delta c_{t+1})]/2, \tag{6.50}$$

provided future conditional volatilities are known in advance.

The forward rate will be lower than the expected future short rate if consumption growth is positively autocorrelated (as Table 6.1 suggests), disregarding the Jensen's inequality terms. The intuition is the same as in Section 6.3.2: the forward contract will generate a capital gain in t if $\mathrm{E}_t(\Delta c_{t+1})$ is low (the one-period interest rate is then low, so the price of a bond is high) and this will typically happen when consumption is scarce in t (positive autocorrelation).

This extends to longer interest rates, so the real yield curve will slope downwards (upwards) if consumption growth is positively (negatively) autocorrelated for most leads/lags. In other words, with positive autocorrelation in consumption growth, the long real (or inflation-indexed) bond is a good hedge and therefore "safer" than the strategy of rolling over short interest rates (see Campbell (1986), Backus, Gregory, and Zin (1989), Backus,

Foresi, and Zin (1998), and Campbell and Viceira (2002)).

6.3.5 Money Supply and the Nominal Interest Rates

This section discusses the effect of monetary policy on nominal interest rates. Simulation models with nominal variables typically introduce money through a cash-in-advance constraint, by sticking real money balances into the utility function, or by a shopping-time technology (see, for instance, Labadie (1989) and Giovannini and Labadie (1991) for asset pricing analysis with cash-in-advance constraints). In many cases, the mere introduction of money has no effect on the real equilibrium, and in most other cases the effect is only temporary and negligible (see, for instance, Cooley and Hansen (1989)).

It is only when nominal rigidities are present that the real variables, including real assets, are affected—and this also gives monetary policy an important role for asset prices. For instance, a policy that aims at stabilizing output is likely to achieve some short-run stabilization of consumption, which may affect risk premia in general. On the other hand, exogenous monetary policy shifts add another source of uncertainty (as argued by Friedman and many others) which may work in the opposite direction. In any case, nominal bonds are likely to be more directly affected by monetary policy than other assets, so this section focuses on the nominal yield curve.

The real price of a nominal bond is as in (6.47) except that the real payoff is not unity but one divided by the nominal price index in $t+k-1$, Q_{t+k-1} Multiplying by today's price level, Q_{t-1}, converts the real price to a nominal bond price

$$B^{\$}_{kt-1} = \mathrm{E}_{t-1}[\beta^k (C_{t+k-1}/C_{t-1})^{-\gamma} Q_{t-1}/Q_{t+k-1}]. \tag{6.51}$$

I assume that log consumption and the inflation rate, $\pi_t = \ln(Q_t/Q_{t-1})$, has a bivariate normal distribution. The (log) one-period nominal interest rate, $i_{1t-1} = \ln(1/B^{\$}_{1t-1})$, is then

$$i_{1t-1} = -\ln(\beta) + \mathrm{E}_{t-1}(\gamma \Delta c_t + \pi_t) - \sigma^2_{t-1}(-\gamma \Delta c_t - \pi_t)/2 \tag{6.52}$$

Using the real interest rate (6.48) gives

$$i_{1t-1} - \mathrm{E}_{t-1}(\pi_t) - r_{1t-1} = -\mathrm{Cov}_{t-1}(\gamma \Delta c_t, \pi_t) - \sigma^2_{t-1}(\pi_t)/2, \tag{6.53}$$

which is a modern version of the Fisher equation: the nominal interest rate equals the sum of expected inflation, the real interest rate and an inflation risk premium (and a Jensen's

inequality term).

The Fisher equation (6.53) shows that the *movements in the nominal interest rate* come from at least two sources: expected inflation may change and the real interest rate too. Of course, in many traditional macro models the real interest rate is very stable so movements in the nominal interest rate reflect similar movements in inflation expectations (the "Fisher effect").

The modelling of monetary policy is certainly crucial for the movements nominal interest rate. In models with optimal monetary policy it is often found that it is necessary to put interest rate volatility in the loss function of the policy maker (see, for instance, Söderlind (2001), and Söderlind (1999)); otherwise the interest rate becomes much too volatile compared with data.[19]

In models with exogenous (non-optimal) monetary policy it can easily happen that the implied nominal interest rate has strange dynamics. As an example, consider what could happen in the model used in Cooley and Hansen (1989) and Cooley and Hansen (1995). The utility function is logarithmic ($\gamma = 1$), so the real interest rate equals the expected consumption growth plus a constant (see (6.48)). The cash-in-advance constraint makes the nominal value of consumption equal the exogenous money supply. Together, these facts mean that $\mathrm{E}_{t-1}(\pi_t) + r_{1t}$ on the left hand side of (6.53) equals the expected money supply growth—and this carries over to the nominal interest rate. The money supply growth is modelled as an exogenous AR(1) process with positive autocorrelation, so this behaviour is inherited by the nominal interest rate (see footnote 13). It might seems innocent to change the autocorrelation coefficient to zero (so money supply becomes a random walk), but this would actually imply a constant nominal interest rate, which is slightly odd.

We now turn to discussing *the risk premium on the nominal bond* in (6.53). It is positive if the covariance of consumption growth and inflation is negative (which seems to be the case in US data). The reason is that the real return on the nominal bond (nominal return minus inflation) then tend to be low when consumption is scarce (see Cochrane and Piazzesi (2002a) for a recent empirical study of bond risk premia).

This risk premium is likely to depend on many modelling choices. With predetermined inflation (which is often used in New Keynesian models), there is no risk premium

[19] Several articles have also analysed the effects on real aggregates and the term structure of monetary policy shocks; see, for instance, Rudebusch (1998) and Cochrane and Piazzesi (2002b).

at all. Monetary policy is also likely to affect the risk premium. For instance, in new Keynesian models with some kind of Phillips curve (sticky, but forward looking, price setting, see, for instance, Clarida, Galí, and Gertler (1999)) output (consumption) and inflation move in the same direction if demand shocks (random movements in the time preference rate) dominate, but in opposite directions if supply shocks (random movements in productivity or factor input prices) dominate. If the monetary policy objective is to minimize some linear combination of variances of output and inflation without any regard to interest rate volatility, then it is well known that monetary policy should neutralize all demand shocks but only some of the supply shocks. This will tend to make risk premium positive.

6.3.6 Time-Variation in Risk Premia

This section briefly discusses how a simulation model can generate time-variation in risk premia. Recall from (6.22) that the Sharpe ratio in the CRRA model is

$$\begin{aligned} \mathrm{E}_{t-1}(R_t^e) &= \mathrm{Cov}_{t-1}(R_t^e, \Delta c_t) && (6.54)\\ &= \rho_{t-1}(R_t^e, \Delta c_t) \times \sigma_{t-1}(\Delta c_t) \times \sigma_{t-1}(R_t^e)\gamma. && (6.55) \end{aligned}$$

To get movements in risk premia in the consumption-based model, the conditional covariance of the excess return and consumption growth must be changing over time. Few simulation models generate this because of a combination of two features: the shocks are typically homoskedastic and the dynamics is close to log-linear (so consumption growth and log returns are linear transformations of the shocks). The result is that all expected asset returns will (almost) move in parallel with the expected riskfree rate.

This can be changed by introducing heteroskedastic shocks. However, from (6.55) we see that time-variation in the volatility of returns only (with constant correlations) will make expected return move, but not Sharpe ratios. Similarly, time-variation in consumption volatility leads to proportional increases of the expected return and Sharpe ratios for all assets.

6.3.7 The Value of a Firm: Labour Contracts and Installation Costs

This section discusses how non-Walrasian labour contracts and installation costs (for new capital) affects the equity value. We put aside the issue of how the total value of a firm is

split between equity holders and creditors in order to focus on the value of their combined stake. The traditional way of modelling a firm is to assume that it hires labour and finances investment through retained earnings (no debt). Any extra cash flow is distributed as dividends

$$D_t = Y_t - w_t L_t - I_t, \tag{6.56}$$

where Y_t is output, w_t the real wage rate, L_t labour input, and I_t investment costs. In reality most firms seem to smooth dividends, so D_t should probably not be directly compared with dividends, but with cash flows or earnings. Typically, the number of shares is kept constant and normalized to unity.

Most simulation models have a Cobb-Douglas production function $Y = Z_t K_t^{\alpha} L_t^{1-\alpha}$, where Z_t is the technology level, K_t capital stock, and L_t is labour input. The first order condition for hiring labour is that the real wage rate equals the marginal product of labour, $w_t = (1-\alpha)Y/L_t$. This does not require a competitive labour market, only that the firm can hire freely at the going wage rate ("right to manage"). With this assumption, we can write (6.56) as

$$D_t = \alpha Y_t - I_t. \tag{6.57}$$

As side remark, we note that a high capital share, α, makes dividends fairly similar to consumption (the aggregate resource constraint is $Y_t = I_t + C_t$), so the consumption claim might be a good proxy for equity in that (admittedly unrealistic) case. In other cases, the dynamics of dividends and consumption can be quite different but many models still seem to generate too stable dividends (see, for instance, Rouwenhorst (1995)). The basic reason is that neither aggregate output nor aggregate investments are nearly as volatile as cash flows and earnings on listed firms.

One possibility of making the cash flow process more realistic is to allow for labour hoarding (or other types of labour contracts that stabilizes the wage bill) and variation in the capacity utilization rate. Labour hoarding would make the wage bill relatively constant, rather than proportional to output as above. As a consequence, the cash flow would fluctuate more in response to movements in output. For instance, Danthine and Donaldson (2002) study the implications different types of labour contracts in a setting where the priority status of wage claims creates an operational leverage (see also Section 6.3.3 for a discussion of leverage)—and find that this can create a sizeable risk premium.

It has also been suggested that investments should be subject to installation costs.

Equations (6.56) and (6.57) are still true (I_t is the investment cost, not the gross addition of capital), but the behaviour of output, investment, and consumption will change. In particular, installation costs will make consumption more volatile since it makes consumption smoothing more costly (see, for instance, Jermann (1998) and Lettau (2000)). For instance, when it is prohibitively expensive to change the investment from some steady state level, then there is no consumption smoothing. Also, cash flows in (6.57) will react stronger to output than before, since investments will no longer move in the same direction as output. Both these effects tend to increase the equity premium, although it should be admitted that increased consumption volatility is something of a cheat since that is not a free parameter. (It could also be noted that installation costs leads to movements in Tobin's q, which seems to be a salient feature of data.)

6.3.8 Adding Assets to a Model

This section discusses if we really need fully-fledged equilibrium models to analyse the pricing of assets. The asset price in (6.12) is $P_t = \mathrm{E}_t[\Sigma_{s=1}^{\infty}\beta^s(C_{t+s}/C_t)^{-\gamma}D_{t+s}]$, where D_{t+s} is the "dividend" in $t+s$. On a mechanical level, we probably need some kind of model to calculate the the joint distribution of C_{t+s} and D_{t+s}. For instance, to price is a one-period real bond ($D_{t+1} = 1$ and $D_{t+s} = 0$ for $s > 1$) we need the model to calculate $\mathrm{E}_t[\beta(C_{t+1}/C_t)^{-\gamma}]$.

However, if the introduction of the asset does not change the process of consumption, then there is no need for explicitly including the asset (in the budget restriction and investment opportunity set of investors) when solving the model. This is the case when we are using a model with complete markets (the stochastic discount factor is unique in this case; see the references in footnote 2). In such a model, we could always analyse the pricing of any asset, for instance, some really exotic derivative, without solving the model anew.

We are in a completely different situation when we introduce non-redundant assets, that is, when we add a non-trivial asset to a model with incomplete markets. In that case we have solve the model with all the assets explicitly incorporated since their existence is likely to affect the equilibrium consumption process. Of course, specifying an exogenous process for nonstorable endowments circumvents the problem, which may explain why that approach is so popular.

Most (simulation based) studies of imperfect markets have concentrated on idiosyncratic income risk that cannot be insured and on borrowing constraints/transaction costs.

The key findings and references are given in Section 6.2.6. Other studies have developed solution methods and analysed a different set of incomplete markets. For instance, Judd and Guu (2001) show that introducing an option on the stock market can lower the equity premium considerably (see also Den Haan (1996) and Marcet and Singleton (1999) for solution methods and some examples).

6.4 Summary

The first part of this chapter studies the consumption-based asset pricing model with constant relative risk aversion (CRRA) by comparing the model predictions to the moments (means, variances, and correlations) of US consumption growth and returns on broad asset classes.

It is shown that the smooth consumption growth rate observed in the US makes it hard to explain the equity premium, unless the risk aversion is very high (this is the equity premium puzzle discussed in Mehra and Prescott (1985)). However, a very high risk aversion seems somewhat implausible and it creates other problems. In particular, a large risk aversion coefficient makes it hard to reconcile average consumption growth with the average return on riskfree assets (this is the riskfree rate puzzle discussed in Weil (1989)) and the low volatility of real interest rates (or approximations of it).

One reaction to these problems is to say that the consumption-based model may fail to explain the magnitude of the risk premium, but that it at least can be used to understand qualitative differences in average return. For instance, the model correctly predicts higher returns on equity than on bonds. One of main issues in the first part of the chapter is how far that argument can be extended. Data on Sharpe ratios for different portfolios sorted on industry, dividend-price ratio, size, book-to-market, and geographical location show that the consumption-based model is unable to explain the cross-sectional variation of Sharpe ratios. The basic problems seems to be that there is a lot of dispersion in Sharpe ratios, but most portfolio returns are only weakly correlated with aggregate consumption growth. It is also shown that the consumption-based model fails badly in explaining time-variation of equity premia. Although the covariance of consumption and equity returns changes over time, it does so in completely the wrong way compared with excess returns.

All these problems have spurred interest in developing the consumption-based model by changing the utility function and/or introducing heterogenous agents. Several of these

suggestions (Epstein-Zin preferences, habit formation, and idiosyncratic risk) are analysed at the end of the first part of the chapter. It is found that they are quite successful in solving some of the problems, but that several remain unsolved. In particular, none of the suggestions make much headway with the cross-sectional variation of Sharpe ration.

The second part of the chapter is a theoretical analysis of the equilibrium relation between different asset returns and the consumption process. Simple pricing expressions (based on assumptions about lognormal returns and consumption growth) highlights the basic mechanism of (otherwise) fairly complicated pricing problems. Among the results, we find the following.

First, a positive autocorrelation of consumption growth is likely to give low risk premia on long-lived consumption claims and a negatively sloped real yield curve. Second, the properties of the Fisher equation and the risk premium on nominal bonds depend crucially on monetary policy. For instance, shocks to aggregate demand typically move consumption and inflation in the same direction, so the standard recipe for monetary policy is cushion all such shocks. However, this is likely to increase the inflation risk premium since it makes inflation and consumption more negatively correlated (so the real return on the nominal bond tend to be low when consumption is scarce). Third, options and levered claims will typically have higher risk premia than the underlying asset/unlevered claim because their returns are more volatile.

Simulation models have made fine contributions to the analysis of asset pricing. In particular, they have taught us a great deal about how the properties of the consumption-based asset model with constant relative risk aversion (CRRA). However, it is clear that the CRRA model has serious problems on many accounts and much of the current research is trying models with different utility functions and/or different measures of consumption. We still know too little about the properties of these new models, for instance, what side effects they have. This is likely to be the next area where simulation models make valuable contributions to the asset pricing literature.

A Data Appendix

The nominal *US stock returns* are from the website of French (2001). These monthly returns are converted to quarterly returns by multiplying the monthly gross returns, for instance, the gross returns for January, February, and March are multiplied to generate

a quarterly gross return. The portfolios are formed from NYSE, AMEX, and NASDAQ firms. The aggregate stock market return is a value weighted return of all available returns. The (equally-weighted) portfolios based on industry, size, dividend/price, or book-to-market are from the same data set, but with the firms sorted on the respective characteristic. The ten industry portfolios are for consumer nondurable, consumer durables, oil, chemicals, manufacturing, telecom, utilities, wholesale and retail, financial services, and other. The ten size portfolios are for deciles of firm market values; the D/P portfolios are for deciles of dividend/price; and the B/M portfolios are for deciles of book value/market values.

The dividend-price ratio for S&P500 is from Shiller (2000).

The nominal return on *long US government bonds* is from Ibbotson Associates.

The *international stock and long government bond returns* are from Ibbotson Associates, but come originally from Morgan Stanley and IMF respectively. The international data is for France, Germany, Japan, and United Kingdom, and has been converted into US dollar returns.

Real returns are calculated by dividing the nominal gross return by the gross inflation rate over the same period. Inflation is calculated from the seasonally adjusted CPI for all urban consumers (available at http://www.stls.frb.org/fred/).

Quarterly growth of *real consumption per capita* of nondurables and services is calculated from the seasonally adjusted number in NIPA Table 8.7 (available at http://www.bea.doc.gov/bea/dn1.htm). The growth rate is calculated as a weighted average of the growth rate of nondurables and the growth rate of services (chained 1996 dollars), where the (time-varying) weight is the relative (current dollar) size of nondurables in relation to services.

The annualized k-quarter excess return used in the discussion of stock returns predictability are calculated as $R_t^e(k) = (R_t R_{t-1} ... R_{t-k+1})^{4/k} - (R_{ft} R_{ft-1} ... R_{ft-k+1})^{4/k}$, where R_t is the quarterly gross return of the aggregate US stock market (see above) and R_{ft} is the quarterly gross return on a short T-bill.

B Econometrics Appendix

The tests in Table 6.1 are based on t-tests from a GMM estimation and the delta method.

Testing of the mean, standard deviation, and Sharpe ratio is done in the following

way. First, moment conditions for the mean (μ) and second moment (χ) of the random variable x_t are defined as

$$\sum_{t=1}^{T} m_t(\beta)/T = \mathbf{0}_{2\times 1} \text{ where } m_t = \begin{bmatrix} x_t - \mu \\ x_t^2 - \chi \end{bmatrix} \text{ and } \beta = \begin{bmatrix} \mu \\ \chi \end{bmatrix}.$$

GMM estimation is trivial and the estimator is asymptotically normally distributed around the true value β_0 as

$$\sqrt{T}(\hat{\beta} - \beta_0) \xrightarrow{d} N(\mathbf{0}_{2\times 1}, S_0),$$

where S_0 is the variance-covariance matrix of $\Sigma_{t=1}^{T} m_t(\beta_0)/\sqrt{T}$, which can be consistently estimated by using the point estimate of β instead of β_0 and then applying the Newey and West (1987) estimator. Second, the standard deviation and Sharpe ratio are calculated as functions of β

$$\begin{bmatrix} \sigma(x) \\ \mathrm{E}(x)/\sigma(x) \end{bmatrix} = g(\beta) = \begin{bmatrix} (\chi - \mu^2)^{1/2} \\ \frac{\mu}{(\chi-\mu^2)^{1/2}} \end{bmatrix}, \text{ so } \frac{\partial g(\chi, \mu)}{\partial \beta'} = \begin{bmatrix} \frac{-\mu}{(\chi-\mu^2)^{1/2}} & \frac{1}{2(\chi-\mu^2)^{1/2}} \\ \frac{\chi}{(\chi-\mu^2)^{3/2}} & \frac{-\mu}{2(\chi-\mu^2)^{3/2}} \end{bmatrix}.$$

From the delta method we then have

$$\sqrt{T}[g(\hat{\beta}) - g(\beta_0)] \xrightarrow{d} N\left(0, \frac{\partial g(\beta_0)}{\partial \beta'} S_0 \frac{\partial g(\beta_0)'}{\partial \beta}\right),$$

which is used to construct asymptotic t-tests.

A similar approach is used for *testing the correlation* of x_t and y_t, $\rho(x_t, y_t)$. For expositional simplicity, assume that both variables have zero means. The variances and the covariance can then be estimated by the moment conditions

$$\sum_{t=1}^{T} m_t(\beta)/T = \mathbf{0}_{3\times 1} \text{ where } m_t = \begin{bmatrix} x_t^2 - \sigma_{xx} \\ y_t^2 - \sigma_{yy} \\ x_t y_t - \sigma_{xy} \end{bmatrix} \text{ and } \beta = \begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{xy} \end{bmatrix}.$$

The covariance matrix of these estimators is estimated as before. The correlation is a simple function of these parameters

$$\rho(x, y) = g(\beta) = \frac{\sigma_{xy}}{\sigma_{xx}^{1/2}\sigma_{yy}^{1/2}}, \text{ so } \frac{\partial g(\chi, \mu)}{\partial \beta'} = \begin{bmatrix} -\frac{1}{2}\frac{\sigma_{xy}}{\sigma_{xx}^{3/2}\sigma_{yy}^{1/2}} & -\frac{1}{2}\frac{\sigma_{xy}}{\sigma_{xx}^{1/2}\sigma_{yy}^{3/2}} & \frac{1}{\sigma_{xx}^{1/2}\sigma_{yy}^{1/2}} \end{bmatrix}.$$

The delta method formula (see above) is the applied to this case as well.

Bibliography

Abel, A. B., 1999, "Risk Premia and Term Premia in General Equilibrium," *Journal of Monetary Economics*, 43, 3–33.

Altug, S., and P. Labadie, 1994, *Dynamic Choice and Asset Markets*, Academic Press.

Backus, D., S. Foresi, and S. Zin, 1998, "Arbitrage opportunities in arbitrage-free models of bond pricing," *Journal of Business and Economic Statistics*, 16, 13–26.

Backus, D., A. Gregory, and S. Zin, 1989, "Risk Premiums in the Term Structure; Evidence from Artificial Economies," *Journal of Monetary Economics*, 24, 371–399.

Bansal, R., and C. Lundblad, 2002, "Market Efficiency, Fundamental Values, and the Size of the Risk Premium in Global Equity Markets," *Journal of Econometrics*, 109, 195–237.

Bidarkota, P. V., and J. H. McCulloch, 2003, "Consumption Asset Pricing with Stable Shocks - Exploring a Solution and Its Implications for Mean Equity Returns," *Journal of Economic Dynamics and Control*, 27, 399–421.

Black, F., and M. Scholes, 1973, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637–659.

Bossaert, P., 2002, *The Paradox of Asset Pricing*, Princeton University Press.

Breeden, D. T., 1979, "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 7, 265–96.

Campbell, J. Y., 1986, "Bond and Stock Returns in a Simple Exchange Model," *Quarterly Journal of Economics*, 101, 785–804.

Campbell, J. Y., 1993, "Intertemporal Asset Pricing without Consumption Data," *American Economic Review*, 83, 487–512.

Campbell, J. Y., 2001, "Consumption-Based Asset Pricing," in George Constantinides, Milton Harris, and Rene Stultz (ed.), *Handbook of the Economics of Finance*, North-Holland, Amsterdam.

Campbell, J. Y., and J. H. Cochrane, 1999, "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107, 205–251.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay, 1997, *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey.

Campbell, J. Y., and L. M. Viceira, 2002, *Strategic Asset Allocation: Portfolio Choice of Long-Term Investors*, Oxford University Press.

Clarida, R., J. Galí, and M. Gertler, 1999, "The Science of Monetary Policy: A New Keynesian Perspective," *Journal of Economic Literature*, 37, 1661–1707.

Cochrane, J. H., 2001, *Asset Pricing*, Princeton University Press, Princeton, New Jersey.

Cochrane, J. H., and M. Piazzesi, 2002a, "Bond Risk Premia," mimeo, University of Chicago.

Cochrane, J. H., and M. Piazzesi, 2002b, "The Fed and Interest Rates—a High-Frequency Identification," *AER Papers and Proceedings*, 92, 90–95.

Cogley, T., 1998, "Idiosyncratic Risk and Equity Premium: Evidence from the Consumer Expenditure Survey," mimeo, Federal Reserve Bankl of San Francisco.

Cogley, T., and J. M. Nason, 1993, "Impulse Dynamics and Propagation Mechanisms in Real Busines Cycle Model," *Economics Letters*, 43, 77–81.

Constantinides, G. M., and D. Duffie, 1996, "Asset Pricing with Heterogeneous Consumers," *The Journal of Political Economy*, 104, 219–240.

Cooley, T. F., and G. D. Hansen, 1989, "The Inflation Tax in a Real Business Cycle Model," *The American Economic Review*, 79, 733–748.

Cooley, T. F., and G. D. Hansen, 1995, "Money and the Business Cycle," in Thomas F. Cooley (ed.), *Frontiers of Business Cycle Research*, Princeton University Press, Princeton, New Jersey.

Cox, J. C., J. E. Ingersoll, and S. A. Ross, 1985, "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385–407.

Danthine, J.-P., and J. B. Donaldson, 2002, "Labour Relations and Asset Returns," *Review of Economic Studies*, 69, 41–64.

Den Haan, W. J., 1996, "Heterogeneity, Aggregate Uncertainty, and the Short-Term Interest Rate," *Journal of Business and Economic Statistics*, 14, 399–411.

Den Haan, W. J., 2001, "The Importance of the Number of Different Agents in a Heterogenous Asset-Pricing Model," *Journal of Economic Dynamics and Control*, 25, 721–746.

Epstein, L. G., and S. E. Zin, 1989a, "Substitution, Risk Aversion, and the Temporal Behavior of Asset Returns: A Theoretical Framework," *Econometrica*, 57, 937–969.

Epstein, L. G., and S. E. Zin, 1989b, "Substitution, Risk Aversion, and the Temporal Behavior of Asset Returns: An Empirical Analysis," *Journal of Political Economy*, 99, 263–286.

Fama, E. F., and K. R. French, 1988a, "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22, 3–25.

Fama, E. F., and K. R. French, 1988b, "Permanent and Temporary Components of Stock Prices," *Journal of Political Economy*, 96, 246–273.

Fama, E. F., and K. R. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.

Ferson, W. E., 1995, "Theory and Empirical Testing of Asset Pricing Models," in Robert A. Jarrow, Vojislav Maksimovic, and William T. Ziemba (ed.), *Handbooks in Operations Research and Management Science* . pp. 145–200, North-Holland.

French, K. R., 2001, "U.S. Research Returns Data," http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/.

Giovannini, A., and P. Labadie, 1991, "Asset Prices and Interest Rates in Cash-in-Advance Models," *Journal of Political Economy*, 99, 1215–1251.

Hansen, L. P., and R. Jagannathan, 1991, "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy*, 99, 225–262.

Hansen, L. P., and S. F. Richard, 1987, "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models," *Econometrica*, 55, 587–613.

Harrison, M. J., and D. M. Kreps, 1979, "Martingales and Arbitrage in Multiperiod Securities Markets," *Journal of Economic Theory*, 20, 381–408.

He, H., and D. M. Modest, 1995, "Market Frictions and Consumption-Based Asset Pricing," *Journal of Political Economy*, 103, 94–117.

Heaton, J. C., and D. J. Lucas, 1996, "Evaluating the Effects of Incomplete Markets on Risk-Sharing and Asset Pricing," *Journal of Political Economy*, 104, 443–487.

Hendershott, P. H., 1998, "Equilibrium Models in Real Estate Research: A Survey," *Journal of Real Estate Literature*, 6, 13–25.

Huang, C.-F., and R. H. Litzenberger, 1988, *Foundations for Financial Economics*, Elsevier Science Publishing, New York.

Jermann, U. J., 1998, "Asset Pricing in Production Economies," *Journal of Monetary Economics*, 41, 257–275.

Johnson, N. L., S. Kotz, and N. Balakrishnan, 1994, *Continuous Univariate Distributions*, Wiley, New York, 2nd edn.

Judd, K. L., and S.-M. Guu, 2001, "Asymptotic Methods for Asset Market Equilibrium Analysis," *Economic Theory*, 18, 127–157.

Kollman, R., 2001, "The Exchange Rate in a Dynamic-Optimizing Business Cycle Model with Nominal Rigidities: A Quantitative Investigation," *Journal of International Economics*, 55, 243–262.

Labadie, P., 1989, "Stochastic Inflation and the Equity Premium," *Journal of Monetary Economics*, 24, 277–298.

LeRoy, S. F., and J. Werner, 2001, *Principles of Financial Economics*, Cambridge University Press.

Lettau, M., 2000, "Inspecting the Mechanism: Closed-form Solutions for Asset Prices in Real Busines Cycle Models," mimeo, Federal Reserve Bank of New York.

Lettau, M., 2002, "Idiosyncratic Risk and Volatility Bounds, or Can Models with Idiosyncratic Risk Solve the Equity Premium Puzzle," Review of Economics and Statistics, forthcoming.

Lettau, M., and H. Uhlig, 2000, "Can Habit Formation Be Reconciled with Business Cycle Facts," *Review of Economic Dynamics*, 3, 79–99.

Lettau, M., and H. Uhlig, 2002, "Sharpe Ratios and Preferences: An Analytical Approach," *Macroeconomic Dynamics*, 6, 242–265.

Lintner, J., 1965, "The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics*, 47, 13–37.

Ljunqvist, L., and H. Uhlig, 2000, "Tax Policy and Aggregate Demand Management Under Catching Up with the Joneses," *American Economic Review*, 90, 356–366.

Lucas, R. E., 1978, "Asset Prices in an Exchange Economy," *Econometrica*, 46, 1426–14.

Luttmer, E. G. J., 1996, "Asset pricing in Economies with Frictions," *Econometrica*, 64, 1439–1467.

Mankiw, G. N., 1986, "The Equity Premium and the Concentration of Aggregate Shocks," *Journal of Financial Economics*, 17, 211–219.

Marcet, A., and K. J. Singleton, 1999, "Equilibrium Asset Prices and Savings of Heterogeneous Agents in the Presence of Incomplete Markets and Portfolio Constraints," *Macroeconomic Dynamics*, 3, 243–277.

Mehra, R., and E. Prescott, 1985, "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145–161.

Merton, R. C., 1973, "Rational Theory of Option Pricing," *Bell Journal of Economics and Management Science*, 4, 141–183.

Mossin, J., 1968, "Equilibrium in a Capital Asset Market," *Econometrica*, 35, 768–783.

Newey, W. K., and K. D. West, 1987, "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Reffett, K., and F. Schorfheide, 2000, "Evaluating Asset Pricing Implications of DSGE Models," mimeo, Arizona State University.

Ross, S. A., 1978, "A Simple Approach to the Valuation of Risky Streams," *Journal of Business*, 51, 453–475.

Rouwenhorst, K., 1995, "Asset Pricing Implications of Equilibrium Business Cycle Models," in Thomas F. Cooley (ed.), *Frontiers of Business Cycle Research*, Princeton University Press, Princeton, New Jersey.

Rudebusch, G. D., 1998, "Do measures of monetary policy in a VAR make sense?," *International Economic Review*, 39, 907–931.

Salyer, K. D., 1998, "Crash States and the Equity Premium: Solving One Puzzle Raises Another," *Journal of Economic Dynamics and Control*, 22, 955–965.

Sharpe, W. F., 1964, "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19, 425–442.

Shiller, R. J., 2000, *Irrational Exuberance*, Princeton University Press, data available at http://www.econ.yale.edu/ shiller/data.htm.

Smith, P. N., and M. R. Wickens, 2002, "Asset Pricing with Observable Stochastic Discount Factors," Discussion Paper No. 2002/03, University of York.

Söderlind, P., 1999, "Solution and Estimation of RE Macromodels with Optimal Policy," *European Economic Review*, 43, 813–823.

Söderlind, P., 2001, "Monetary Policy and the Fisher Effect," *Journal of Policy Modeling*, 23, 491–495.

Söderlind, P., 2003, "Monetary Policy and Bond Option Pricing in an Analytical RBC Model," WP 447, Stockholm School of Economics, Forthcoming in *Journal of Economics and Business*.

Söderlind, P., and L. E. O. Svensson, 1997, "New Techniques to Extract Market Expectations from Financial Instruments," *Journal of Monetary Economics*, 40, 383–420.

Svensson, L. E. O., 1989, "Portfolio Choice with Non-expected Utility in Continuous Time," *Economics Letters*, 40, 313–317.

Telmer, C. I., 1993, "Asset Pricing Puzzles and Incomplete Markets," *Journal of Finance*, 48, 1803–1832.

Vasicek, O. A., 1977, "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177–188.

Vayanos, D., 1998, "Transaction Costs and Asset Prices: A Dynamic Equilibrium Model," *Review of Financial Studies*, 11, 1–58.

Weil, P., 1989, "The Equity Premium Puzzle and the Risk-Free Rate Puzzle," *Journal of Monetary Economics*, 24, 401–421.

Zimmermann, C., 1999, "International Business Cycles and Exchange Rates," *Review of International Economics*, 74, 682–698.

7 ARCH and GARCH

Reference: Campbell, Lo, and MacKinlay (1997) 12.2; Hamilton (1994) 21; Greene (2000) 18.5; Hentschel (1995); Franses and van Dijk (2000)

7.1 Test of ARCH Effects

Suppose we have estimated a model, and obtained the fitted residuals, $\hat{u}_t$ for $t = 1, \ldots, T$. Engle's test of ARCH (autoregressive conditional heteroskedasticity) is perhaps the most straightforward. It amounts to running the regression

$$\hat{u}_t^2 = a_0 + \sum_{s=1}^{q} a_i \hat{u}_{t-s}^2 + v_t. \tag{7.1}$$

Under the null hypothesis of no ARCH effects, all slope coefficients are zero and the R^2 of the regression is zero. These overidentifying restrictions can be tested by noting that, under the null hypothesis, $TR^2 \sim \chi_q^2$.

It is straightforward to phrase Engle's test in terms of GMM moment conditions. We simply use a first set of moment conditions to estimate the parameters of the regression model, and then test if the following additional (ARCH related) moment conditions are satisfied at those parameters

$$\mathrm{E}\begin{bmatrix} u_{t-1}^2 \\ \vdots \\ u_{t-s}^2 \end{bmatrix} (u_t^2 - a_0) = \mathbf{0}_{s\times 1}. \tag{7.2}$$

An alternative test (see Harvey (1989) 259–260), is to apply a Box-Ljung test on $\hat{u}_t^2$, to see if the squared fitted residuals are autocorrelated. We just have to adjust the degrees of freedom in the asymptotic chi-square distribution by subtracting the number of parameters estimated in the regression equation.

These tests for ARCH effects will typically capture GARCH (see below) effects as well.

7.2 ARCH Models

A very simple ARCH(1) model is given by

$$y_t = x_t'b + u_t, \ \ \mathrm{E}(u_t|x_t) = 0 \tag{7.3}$$

where x_t is a $k \times 1$ vector. The residual can be written

$$u_t = v_t\sigma_t \text{ with } v_t \sim \text{iid with } \mathrm{E}_{t-1}\, v_t = 0 \text{ and } \mathrm{E}_{t-1}\, v_t^2 = 1, \tag{7.4}$$

and the conditional variance is generated by

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2, \text{ with } \alpha_0 > 0 \text{ and } 0 \leq \alpha_1 < 1. \tag{7.5}$$

(Note that my convention for the time subscript of the conditional variance is different from that used in CLM. They use σ_{t-1}^2 to indicate the variance in t.)

The non-negativity restrictions are needed in order to guarantee $\sigma_t^2 > 0$. The upper bound $\alpha_1 < 1$ is needed in order to make the conditional variance stationary. To see this, use $\sigma_t^2 = \mathrm{E}_{t-1}\, u_t^2$ to write (7.5) as

$$\mathrm{E}_{t-1}\, u_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2, \text{ which implies} \tag{7.6}$$

$$\mathrm{E}\, u_t^2 = \alpha_0 + \alpha_1\, \mathrm{E}\, u_{t-1}^2 \text{ or } \mathrm{E}\, u_t^2 = \alpha_0/(1-\alpha_1) \text{ (if } \alpha_1 < 1). \tag{7.7}$$

A value of $\alpha_1 < 1$ is clearly needed to make the difference equation stable and hence the unconditional variance finite.

It is clear that the unconditional distribution of u_t is non-normal. Even if we assume that v_t is iid $N(0,1)$, we get that the conditional distribution of u_t is $N(0, \sigma_t^2)$, so the unconditional distribution of u_t is a mixture of normal distributions with different variances. It can be shown that the result is a distribution which have fatter tails than a normal distribution with the same variance (excess kurtosis).

Equation (7.6) shows how we can use information in $t-1$ to forecast u_t^2. By leading the time subscript one period we have

$$\begin{aligned} \mathrm{E}_t\, u_{t+1}^2 &= \alpha_0 + \alpha_1 u_t^2, \text{ so} \\ \mathrm{E}_{t-1}\, u_{t+1}^2 &= \alpha_0 + \alpha_1\, \mathrm{E}_{t-1}\, u_t^2 \\ &= \alpha_0(1+\alpha_1) + \alpha_1^2 u_{t-1}^2, \end{aligned} \tag{7.8}$$

where we used (7.6) to substitute for $E_{t-1} u_t^2$. Similarly, for u_{t+2}^2 we get from (7.8)

$$\begin{aligned} E_{t-1} u_{t+2}^2 &= \alpha_0 + \alpha_1 E_{t-1} u_{t+1}^2 \\ &= \alpha_0(1 + \alpha_1 + \alpha_1^2) + \alpha_1^3 u_{t-1}^2. \end{aligned} \tag{7.9}$$

The pattern for using information in $t-1$ to forecast future volatility is now clear: it works just like an AR(1). In fact, rearranging (7.8) and (7.9) shows that the forecasting equation can be written

$$E_{t-1} u_{t+s}^2 - \frac{\alpha_0}{1-\alpha_1} = \alpha_1^{s+1}\left(u_{t-1}^2 - \frac{\alpha_0}{1-\alpha_1}\right), \tag{7.10}$$

which describes the evolution of the $E_{t-1} u_{t+s}^2$ around its unconditional mean as an AR(1). The parameter α_1 is therefore crucial for the persistence of shocks to volatility.

Note that (7.3) is valid linear regression model (since the residual and regressors are uncorrelated), so LS is consistent (and the most efficient linear estimator of b). Since the heteroskedasticity is not related to the regressors, the standard LS expression for the covariance of the LS estimator of b should also continue to hold. However, there are more efficient estimators than LS. Moreover, even if we are not too interested in efficiency, we might be interested in the heteroskedasticity for its own sake. This suggests that we may want to estimate the full model (7.3)–(7.5) by ML or GMM.

The most common way to estimate the model is to assume that $v_t \sim$iid $N(0,1)$ and to set up the likelihood function. The log likelihood is easily found, since the model is conditionally Gaussian. It is

$$\ln \mathcal{L} = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln \sigma_t^2 - \frac{1}{2}\sum_{t=1}^{T}\frac{u_t^2}{\sigma_t^2} \text{ if } v_t \text{ is iid } N(0,1). \tag{7.11}$$

By plugging in (7.3) for u_t and (7.5) for σ_t^2, the likelihood function is written in terms of the data and model parameters. The likelihood function is then maximized with respect to the parameters. Note that we need a starting value of $\sigma_1^2 = \alpha_0 + \alpha_1 u_0^2$. The most convenient (and common) way is maximize the likelihood function conditional on a y_0 and x_0. That is, we actually have a sample running from $t = 0$ to $t = T$, but observation $t = 0$ is only used to construct a starting value of σ_1^2. The optimization should preferably impose the constraints in (7.5). The MLE is consistent.

Remark 33 *(Coding the ARCH(1) ML estimation.) A straightforward way of coding the estimation problem (7.3)–(7.5) and (7.11) is as follows. First, guess values of the parameters b (a vector), and α_0, and α_1. The guess of b can be taken by LS estimation of (7.3), and the guess of α_0 and α_1 from LS estimation of $\hat{u}_t^2 = \alpha_0 + \alpha_1 \hat{u}_{t-1}^2 + \varepsilon_t$ where $\hat{u}_t$ are the fitted residuals from the LS estimation of (7.3). Second, loop over the sample (first $t = 1$, then $t = 2$, etc.) and estimate calculate $\hat{u}_t$ from (7.3) and σ_t^2 from (7.5). Plug in these numbers in (7.11) to find the likelihood value. Third, make better guesses of the parameters and do the first two steps again. Repeat until the likelihood value converges (at a maximum).*

Remark 34 *(Imposing parameter constraints on ARCH(1).) To impose the restrictions in (7.5) when the previous remark is implemented, iterate over values of $(b, \tilde{\alpha}_0, \tilde{\alpha}_1)$ and let $a_0 = \tilde{\alpha}_0^2$ and $a_1 = \exp(\tilde{a}_1)/[1 + \exp(\tilde{a}_1)]$. This approach is particularly useful when a numerical optimization routine is used.*

If is often found that the fitted normalized residuals, $\hat{u}_t/\sigma_t$, still have too fat tails compared with $N(0,1)$. Estimation using other likelihood functions, for instance, for a t-distribution, can then be used. Or the estimation can be interpreted as a quasi-ML (is typically consistent, but requires different calculation of the covariance matrix of the parameters).

Another possibility is to estimate the model by GMM using, for instance, the following the moment conditions (where σ_t^2 is given by (7.5))

$$E\begin{bmatrix} x_t u_t \\ u_t^2 - \sigma_t^2 \\ u_{t-1}^2(u_t^2 - \sigma_t^2) \end{bmatrix} = \mathbf{0}_{(k+2)\times 1}, \text{ where } u_t = y_t - x_t' b. \tag{7.12}$$

It is straightforward to add more lags to (7.5). For instance, an ARCH(p) would be

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \ldots + \alpha_p u_{t-p}^2. \tag{7.13}$$

We then have to add more moment conditions to (7.12), but the form of the likelihood function is the same except that we now need p starting values and that the upper boundary constraint should now be $\Sigma_{j=1}^{p}\alpha_j \leq 1$.

7.3 GARCH Models

Instead of specifying an ARCH model with many lags, it is typically more convenient to specify a low-order GARCH (Generalized ARCH) model. The GARCH(1,1) is simple and surprisingly general

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \text{with } \alpha_0 > 0;\ \alpha_1,\ \beta_1 \geq 0;\ \text{and } \alpha_1 + \beta_1 < 1. \tag{7.14}$$

The non-negativity restrictions are needed in order to guarantee $\sigma_t^2 > 0$. The upper bound $\alpha_1 + \beta_1 < 1$ is needed in order to make the σ_t^2 stationary. To see this, use $\sigma_t^2 = \mathrm{E}_{t-1}\, u_t^2$ to write (7.14) as

$$\begin{aligned} \mathrm{E}_{t-1}\, u_t^2 &= \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1\, \mathrm{E}_{t-2}\, u_{t-1}^2, \text{ so} \\ \mathrm{E}\, u_t^2 &= \alpha_0 + \alpha_1\, \mathrm{E}\, u_{t-1}^2 + \beta_1\, \mathrm{E}\, u_{t-1}^2 \text{ or } \mathrm{E}\, u_t^2 = \alpha_0/(1-\alpha_1-\beta_1). \end{aligned} \tag{7.15}$$

The bound $\alpha_1 + \beta_1 < 1$ is needed to make the difference equation stable and therefore the unconditional variance finite.

The GARCH(1,1) corresponds to an ARCH(∞) with geometrically declining weight, which is easily seen by solving (7.14) recursively by substituting for σ_{t-s}^2

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 \left(\alpha_0 + \alpha_1 u_{t-2}^2 + \beta_1 \sigma_{t-2}^2\right) \\ &= \alpha_0 (1+\beta_1) + \alpha_1 u_{t-1}^2 + \beta_1 \alpha_1 u_{t-2}^2 + \beta_1^2 \sigma_{t-2}^2 \\ &= \vdots \\ &= \frac{\alpha_0}{1-\beta_1} + \alpha_1 \sum_{j=0}^{\infty} \beta_1^j u_{t-1-j}^2. \end{aligned} \tag{7.16}$$

This suggests that a GARCH(1,1) might be a reasonable approximation of a high-order ARCH.

Repeated substitutions show that the forecasting equation for future volatility is

$$\mathrm{E}_{t-1}\, u_{t+s}^2 - \frac{\alpha_0}{1-\alpha_1-\beta_1} = (\alpha_1+\beta_1)^{s+1} \left(u_{t-1}^2 - \frac{\alpha_0}{1-\alpha_1-\beta_1}\right), \tag{7.17}$$

which is of the same form as for the ARCH model (7.10), but where the sum of α_1 and β_1 being the AR(1) parameter.

To estimate the model consisting of (7.3) and (7.14) we can still use the likelihood

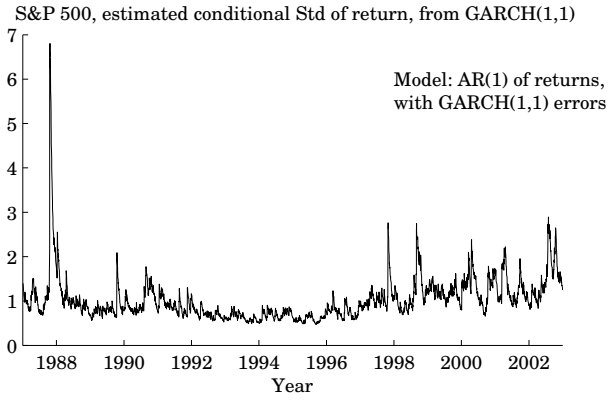


Figure 7.1: Conditional standard deviation, estimated by GARCH(1,1) model.

function (7.11) and do a MLE. We typically create the starting value of u_0^2 as in the ARCH model (use y_0 and x_0 to create u_0), but this time we also need a starting value of σ_0^2. It is often recommended that we use $\sigma_0^2 = \mathrm{Var}(\hat{u}_t)$, where $\hat{u}_t$ are the residuals from a LS estimation of (7.3). It is also possible to assume another distribution than $N(0,1)$.

Remark 35 *(Imposing parameter constraints on GARCH(1,1).) To impose the restrictions in (7.14), iterate over values of* $(b, \tilde{\alpha}_0, \tilde{\alpha}_1, \tilde{\beta}_1)$ *and let* $a_0 = \tilde{\alpha}_0^2$, $a_1 = \exp(\tilde{a}_1)/[1 + \exp(\tilde{a}_1) + \exp(\tilde{\beta}_1)]$, *and* $\beta_1 = \exp(\tilde{\beta}_1)/[1 + \exp(\tilde{a}_1) + \exp(\tilde{\beta}_1)]$.

To estimate the GARCH(1,1) with GMM, we can, for instance, use the following moment (where σ_t^2 is given by (7.2))

$$\mathrm{E}\begin{bmatrix} x_t u_t \\ u_t^2 - \sigma_t^2 \\ u_{t-1}^2 (u_t^2 - \sigma_t^2) \\ u_{t-2}^2 (u_t^2 - \sigma_t^2) \end{bmatrix} = \mathbf{0}_{(k+3)\times 1}, \text{ where } u_t = y_t - x_t' b. \tag{7.18}$$

7.4 Non-Linear Extensions

A very large number of extensions have been suggested. I summarize a few of them, which can be estimated by using the likelihood function (7.11) to do a MLE.

An asymmetric GARCH (Glosten, Jagannathan, and Runkle (1993)) can be constructed as

$$\sigma_t^2 = \alpha_0+\alpha_1 u_{t-1}^2+\beta_1\sigma_{t-1}^2+\gamma\delta(u_{t-1}>0)u_{t-1}^2, \text{ where } \delta(q) = \begin{cases} 1 \text{ if } q \text{ is true} \\ 0 \text{ else.} \end{cases} \quad (7.19)$$

This means that the effect of the shock u_{t-1}^2 is α_1 if the shock was negative and $\alpha_1+\gamma$ if if the shock was positive.

The EGARCH (exponential GARCH, Nelson (1991)) sets

$$\ln\sigma_t^2 = a_0 + \beta_1 \ln\sigma_{t-1}^2 + \gamma\frac{u_{t-1}}{\sigma_{t-1}} + \alpha\frac{|u_{t-1}|}{\sigma_{t-1}}. \quad (7.20)$$

Apart from being written in terms of the log (which is a smart trick to make $\sigma_t^2 > 0$ hold without any restrictions on the parameters), this is an asymmetric model as long as $\gamma \neq 0$. The last term makes makes volatility a function of $|u_{t-1}|$, so both negative and positive values of u_{t-1} affects the volatility in the same way. The linear term in u_{t-1} modifies this. In particular, if $\gamma < 0$, then the volatility increases more as a response to a negative u_{t-1} ("bad news") than to a positive u_{t-1}.

Hentschel (1995) estimates several models of this type, as well as a very general formulation on daily stock index data for 1926 to 1990 (some 17,000 observations). Most standard models are rejected in favour of a model where σ_t depends on σ_{t-1} and $|u_{t-1}-b|^{3/2}$.

7.5 (G)ARCH-M

It can make sense to let the conditional volatility enter the mean equation—for instance, as a proxy for risk which may influence the expected return. We can modify the "mean equation" (7.3) to include the conditional variance σ_t^2 (taken from any of the models for heteroskedasticity) as a regressor

$$y_t = x_t'b + \varphi\sigma_t^2 + u_t, \;\; \mathrm{E}(u_t|x_t,\sigma_t) = 0. \quad (7.21)$$

Note that σ_t^2 is predetermined, since it is a function of information in $t-1$. This model can be estimated by using the likelihood function (7.11) to do MLE.

Remark 36 *(Coding of (G)ARCH-M) We can use the same approach as in Remark 33, except that we use (7.21) instead of (7.3) to calculate the residuals (and that we obviously also need a guess of φ).*

7.6 Multivariate (G)ARCH

Let the model (7.3) be a multivariate model where y_t and u_t are $n\times 1$ vectors. We define the conditional (on the information set in $t-1$) covariance matrix of u_t as

$$\Sigma_t = \mathrm{E}_{t-1}\, u_t u_t'. \quad (7.22)$$

It may seems as if a multivariate version of the GARCH(1,1) model would be simple, but it is not because it contains far too many parameters. Although we only need to care about the unique elements of Σ_t, that is, vech(Σ_t), this still gives very many parameters

$$\mathrm{vech}(\Sigma_t) = A_0 + A_1\mathrm{vech}(u_{t-1}u_{t-1}') + B_1\mathrm{vech}(\Sigma_{t-1}). \quad (7.23)$$

For instance, with $n=2$ we have

$$\begin{bmatrix} \sigma_{11t} \\ \sigma_{21t} \\ \sigma_{22t} \end{bmatrix} = A_0 + A_1 \begin{bmatrix} u_{1t-1}^2 \\ u_{1t-1}u_{2t-1} \\ u_{2t-1}^2 \end{bmatrix} + B_1 \begin{bmatrix} \sigma_{11t-1} \\ \sigma_{21t-1} \\ \sigma_{21t-1} \end{bmatrix}, \quad (7.24)$$

A_0 is 3×1, A_1 is 3×3, and B_1 is 3×3. This gives 21 parameters, which is already hard to manage. We have to limit the number of parameters. We also have to find a way to impose restrictions so Σ_t is positive definite (compare the restrictions of positive coefficients in (7.14)).

One model that achieves both these aims is the *diagonal model*, which assumes that A_1 and B_1 are diagonal. This means that every element of vech(Σ_t) follows a univariate process. With $n=2$, this gives $3+3+3=9$ parameters (in A_0, A_1, and A_2, respectively). To make sure that Σ_t is positive definite we have to impose further restrictions, but they are relatively simple (rewrite the system on matrix form, $\Sigma_t = a_0+a_1u_{t-1}u_{t-1}'+b_1\Sigma_{t-1}$,

and require that a_0, a_1, and b_1 are all positive definite matrices). The obvious drawback of this model is that there is no spillover of volatility from one variable to another.

The *BEKK model* makes Σ_t positive definite by specifying a quadratic form

$$\Sigma_t = C'C + A'u_{t-1}u'_{t-1}A + B'\Sigma_{t-1}B, \tag{7.25}$$

where C is lower triangular and A and B are $n \times n$ matrices. With $n = 2$, this gives $3 + 4 + 4 = 11$ parameters (in C, A, and B, respectively).

The *constant correlation model* makes every variance follow a univariate GARCH(1,1) model and covariances are constant correlations times the two standard deviations. With $n = 2$, this gives 7 parameters (2×3 GARCH parameters and one correlation), which is convenient. The price is, of course, the assumption of no movements in the correlations. To get a positive definite Σ_t, each individual GARCH model must generate a positive variance, and that all the estimated (constant) correlations are between -1 and 1.

Bibliography

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay, 1997, *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey.

Franses, P. H., and D. van Dijk, 2000, *Non-linear Time Series Models in Empirical Finance*, Cambridge University Press.

Glosten, L. R., R. Jagannathan, and D. Runkle, 1993, "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779–1801.

Greene, W. H., 2000, *Econometric Analysis*, Prentice-Hall, Upper Saddle River, New Jersey, 4th edn.

Hamilton, J. D., 1994, *Time Series Analysis*, Princeton University Press, Princeton.

Harvey, A. C., 1989, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Hentschel, L., 1995, "All in the Family: Nesting Symmetric and Asymmetric GARCH Models," *Journal of Financial Economics*, 39, 71–104.

Nelson, D. B., 1991, "Conditional Heteroskedasticity in Asset Returns," *Econometrica*, 59, 347–370.

8 Financial Applications of ARCH and GARCH Models

These notes summarize some financial applications of ARCH/GARCH models. The idea is to show models where the ARCH/GARCH is integrated with financial models.

8.1 "Fundamental Values and Asset Returns in Global Equity Markets," by Bansal and Lundblad

Reference: Bansal and Lundblad (2002) (BL)

This paper studies how stock indices for five major markets are related to news about future cash flows (dividends and/or earnings). It uses monthly data 1973–1998 for France, Germany, Japan, UK, US, and a world market index.

BL argue that their present value model (stock price equals the present value of future cash flows) can account for observed volatility of equity returns and the cross-correlation across markets. This is an interesting result since most earlier present value models have generated too small movements in returns—and also too small correlations across markets. The crucial features of the model are a predictable long-run component in cash flows and time-varying systematic risk.

8.1.1 A Model of the SDF

If the log SDF, m_{t+1}, and the log gross return, r_{it+1}, have a (conditional) bivariate normal distribution, then the asset pricing equation says

$$\begin{aligned} 1 &= \mathrm{E}_t \exp(m_{t+1} + r_{it+1}) = \exp[\mathrm{E}_t\, m_{t+1} + \mathrm{E}_t\, r_{it+1} + \sigma_t^2(m_{t+1} + r_{it+1})/2], \text{ so} \\ \mathrm{E}_t\, r_{it+1} &= -\mathrm{E}_t\, m_{t+1} - \mathrm{Cov}_t(m_{t+1}, r_{it+1}) - \sigma_t^2(m_{t+1})/2 - \sigma_t^2(r_{it+1})/2. \end{aligned} \tag{8.1}$$

For a log riskfree rate (known in t) this simplifies to

$$r_{ft+1} = -\mathrm{E}_t\, m_{t+1} - \sigma_t^2(m_{t+1})/2. \tag{8.2}$$

The expected log "excess return," $\mathrm{E}_t\, r^e_{it+1} = \mathrm{E}_t\, r_{it+1} - r_{ft+1}$, is therefore

$$\mathrm{E}_t\, r^e_{it+1} = -\mathrm{Cov}_t(m_{t+1}, r^e_{it+1}) - \sigma_t^2(r^e_{it+1})/2. \tag{8.3}$$

BL eventually makes the assumption that a conditional (international) CAPM holds for log returns. This means that the log SDF is linear in the riskfree rate, r_{ft+1}, and the surprise in the log excess return on the world market equity index, r^e_{mt+1},

$$m_{t+1} = -r_{ft+1} - \lambda(r^e_{mt+1} - \mathrm{E}_t\, r^e_{mt+1}). \tag{8.4}$$

Using (8.4) in (8.3) gives

$$\begin{aligned} \mathrm{E}_t\, r^e_{it+1} &= \lambda\, \mathrm{Cov}_t(r^e_{mt+1}, r^e_{it+1}) - \sigma_t^2(r^e_{it+1})/2 \\ &= \lambda\beta_i \sigma_t^2(r^e_{mt+1}) - \sigma_t^2(r^e_{it+1})/2, \end{aligned} \tag{8.5}$$

where $\beta_i = \mathrm{Cov}_t(r^e_{mt+1}, r^e_{it+1})/\sigma_t^2(r^e_{mt+1})$ is assumed to be constant. The last term is a Jensen's inequality term.

We can rewrite (8.5) a bit more by noting that it implies that we have a linear factor model (with $\alpha_i = 0$, see lecture notes on linear factor models)

$$r^e_{it+1} = \alpha_i + \beta_i r^e_{mt+1} + \varepsilon_{it+1}, \;\; \mathrm{E}_t\, \varepsilon_{it+1} = 0, \sigma_t^2(\varepsilon_{it+1}) = \sigma^2(\varepsilon_{it+1}), \text{ and } \mathrm{Cov}_t(r^e_{mt+1}, \varepsilon_{it+1}) = 0, \tag{8.6}$$

where it is assumed that the idiosyncratic noise is homoskedastic. This expression can be used to understand the Jensen's inequality term in (8.5)—although that is probably not very important. In any case, (8.6) implies that $\sigma_t^2(r^e_{it+1}) = \beta_i^2 \sigma_t^2(r^e_{mt+1}) + \sigma^2(\varepsilon_{it+1})$, so (8.5) can be written

$$\begin{aligned} \mathrm{E}_t\, r^e_{it+1} &= \lambda\beta_i \sigma_t^2(r^e_{mt+1}) - [\beta_i^2 \sigma_t^2(r^e_{mt+1}) + \sigma^2(\varepsilon_{it+1})]/2 \\ &= \left(\lambda\beta_i - \beta_i^2/2\right) \sigma_t^2(r^e_{mt+1}) - \sigma^2(\varepsilon_{it+1})/2. \end{aligned} \tag{8.7}$$

8.1.2 Approximation of the Return

We define a gross return as

$$R_{it+1} = \frac{D_{it+1} + P_{it+1}}{P_{it}} = \frac{P_{it+1}}{P_{it}}\left(1 + \frac{D_{it+1}}{P_{it+1}}\right) \text{ or in logs}$$
$$r_{it+1} = p_{it+1} - p_{it} + \ln\left[1 + \exp(d_{it+1} - p_{it+1})\right]. \tag{8.8}$$

The last (nonlinear) term is then approximated by constant plus $(1-\rho_i)\,(d_{it+1} - p_{it+1})$, with $\rho = 1/(1+\overline{D/P})$ where $\overline{D/P}$ is the average dividend-price ratio. Add and subtract $d_{it+1} - d_{it}$ to get (constant suppressed, see lecture notes on predictability)

$$r_{it+1} \approx \rho_i \underbrace{(p_{it+1} - d_{it+1})}_{z_{it+1}} - \underbrace{(p_{it} - d_{it})}_{z_{it}} + \underbrace{(d_{it+1} - d_{it})}_{g_{it+1}}. \tag{8.9}$$

Take expectations as of t. Use (8.7) to substitute for $\mathrm{E}_t\, r_{it+1} = r_{ft+1} + \mathrm{E}_t\, r^e_{it+1}$ (disregard the constant $\sigma^2(\varepsilon_{it+1})$) to get

$$r_{ft+1} + (\lambda\beta_i - \beta_i^2/2)\sigma_t^2(r^e_{mt+1}) \approx \rho_i\, \mathrm{E}_t\, z_{it+1} - z_{it} + \mathrm{E}_t\, g_{it+1}. \tag{8.10}$$

Rearrange a bit to get

$$z_{it} \approx \rho_i\, \mathrm{E}_t\, z_{it+1} + \mathrm{E}_t\, \varkappa_{it+1}, \text{ where } \mathrm{E}_t\, \varkappa_{it+1} = \mathrm{E}_t\, g_{it+1} - r_{ft+1} - (\lambda\beta_i - \beta_i^2/2)\sigma_t^2(r^e_{mt+1}). \tag{8.11}$$

Substitute for $\mathrm{E}_t\, z_{it+1}$, and then for $\mathrm{E}_t\, z_{it+2}$ and so forth to get

$$p_{it} - d_{it} = z_{it} \approx \sum_{s=1}^{\infty} \rho_i^{s-1}\, \mathrm{E}_t\, \varkappa_{it+s}. \tag{8.12}$$

If we have information about $\mathrm{E}_t\, g_{it+s}$, $\mathrm{E}_t\, r_{ft+s}$, and $\mathrm{E}_t\, \sigma^2_{t-1+s}(r_{mt+1})$ for $s = 1, 2, ...$, then it is straightforward to solve for z_{it} in terms of the parameters of the time series process (assuming no bubbles). In BL this turns out to be a fairly simple solution since all these three processes are modelled as univariate first-order time-series processes. It is therefore a solution in terms of the current values of the three variables.

This solution can be thought of the "fundamental" price-dividend ratio. The main theme of the paper is see how well this fundamental $p - d$ ratio can explain the actual $p - d$ ratio.

8.1.3 A Benchmark Case

As a benchmark for comparison, consider the case when $\mathrm{E}_t\, \varkappa_{it+s}$ in (8.12) equals a constant. This would happen when the growth rate of cash flow is unpredictable, the riskfree rate is constant, and the market risk premium is too (which here requires that the conditional variance of the market return is constant). In this case, the price-dividend ratio is constant, so the log return equals the cash flow growth plus a constant—see (8.9).

This benchmark case would not be very successful in matching the observed volatility and correlation (across markets) of returns: cash flow growth seems to be a lot less volatile than returns and also a lot less correlated across markets.

What if we allowed for predictability of cash flow growth, but still kept the assumptions of constant real interest rate and market risk premium? Large movements in predictable cash flow growth could then generate large movements in returns, but hardly the correlation across markets. However, large movements in the market risk premium would contribute to both. It is clear that both mechanisms are needed to get a correlation between zero and one. It can also be noted from (8.11) that the returns will be more correlated during volatile periods—since this drives up the market risk premium which is a common component in all returns.

8.1.4 Cash Flow Dynamics

Remark 37 *An ARMA(1,1) model is*

$$y_t = a y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}, \textit{ where } \varepsilon_t \textit{ is white noise.}$$

The model can be written on MA form as

$$y_t = \varepsilon_t + \sum_{s=1}^{\infty} a^{s-1}(a+\theta)\varepsilon_{t-s}.$$

The autocorrelations are

$$\rho_1 = \frac{(1+a\theta)(a+\theta)}{1+\theta^2+2a\theta}, \textit{ and } \rho_s = a\rho_{s-1} \textit{ for } s = 2, 3, \ldots$$

and the conditional expectations are

$$\mathrm{E}_t\, y_{t+s} = a^{s-1}(a y_t + \theta\varepsilon_t)\ s = 1, 2, \ldots$$

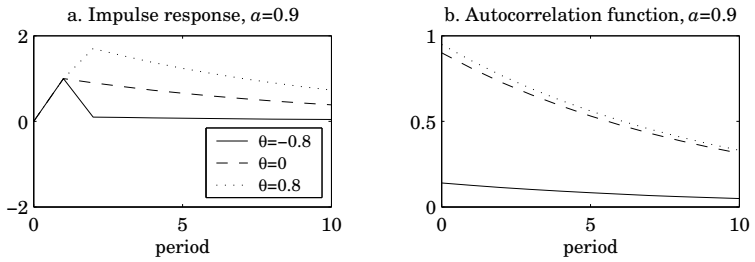


Figure 8.1: Impulse response and autcorrelation functions of ARMA(1,1)

The growth rate of cash flow, g_{it}, is modelled as an ARMA(1,1). The estimation results show that the AR parameter is around 0.97 and that the MA parameter is around -0.93. This means that the growth rate is almost an iid process with very low autocorrelation—but only almost. Since the MA parameter is not negative enough to make the sum of the AR and MA parameters zero, a positive shock to the growth rate will have a long-lived effect (even if small). See *Figure 8.1.a*.

8.1.5 Time Series Process of the SDF and Riskfree Rate

The log excess market return, r^e_{mt+1}, could be modelled as a GARCH(1,1)-M. The "M" part is consistent with the model of the SDF since applying (8.7) on the market return (and noting that $\beta_m = 1$) gives

$$\mathrm{E}_t\, r^e_{mt+1} = (\lambda - 1/2)\sigma^2_t(r^e_{mt+1}). \tag{8.13}$$

The choice of a GARCH(1,1) is based on econometric criteria and the fact that is creates a simple way to calculate the expected future volatility (to be used in (8.12)). We typically define the "market price of risk" as the Sharpe ratio on the market. From (8.13), we see that this is $(\lambda - 1/2)\sigma_t(r_{mt+1})$, so $(\lambda - 1/2)$ can be interpreted as the risk aversion

(including the Jensen's inequality term).

In practice, BL model the excess return (rather than the excess log return)

$$R^e_{mt+1} = \lambda\sigma^2_{mt} + \varepsilon_{mt+1},\ \mathrm{E}_t\,\varepsilon_{mt+1} = 0 \text{ and } \mathrm{Var}_t(\varepsilon_{mt+1}) = \sigma^2_{mt}, \tag{8.14}$$

$$\sigma^2_{mt} = \zeta + \gamma\varepsilon^2_{mt} + \delta\sigma^2_{mt-1}, \tag{8.15}$$

where σ^2_{mt} should be a good approximation of $\sigma^2_t(r_{mt+1})$.

BL also model r_{ft+1} as an AR(1).

8.1.6 Estimation

The model is estimated by GMM. The moment conditions are the following.

1. LS moment conditions for estimating α_i and β_i in (8.6).
2. The score vector (ML first order conditions) for estimating $(\lambda, \zeta, \gamma, \delta)$ in the GARCH(1,1)-M in (8.14)–(8.15).
3. LS moment conditions for estimating the AR(1) parameters of the log riskfree rate.
4. Moment conditions for estimating the ARMA parameters of dividend growth.
5. Moment conditions for z_{it} minus the predicted value from the solution of (8.12). This predicted value turns out to be a linear function of σ^2_{mt}, r_{ft}, g_{it} plus the innovation in g_{it} (this is the sum necessary for making forecasts). The coefficient of this linear combination are non-linear functions of the other parameters of the model. This is the "overidentified" part of the model, which is interesting to test. That is, can the fundamental (fitted) price-dividend ratio really explain the actual ratio?

8.1.7 Results

1. The hypothesis $\alpha_i = 0$ for all i (the five country indices) cannot be rejected.
2. Most of the parameters are precisely estimated, except λ (the risk aversion).
3. Market volatility is very persistent.
4. Cash flow has a small, but very persistent effect of news.

5. The overidenifying restrictions are rejected , but the model still seems able to account for quite a bit of the data: the volatility and correlation (across countries) of the fundamental price-dividend ratios are quite similar to those in the data. Note that the cross correlations are driven by the common movements in R_{ft+1} and σ_{mt}^2.

6. Some questions:

   (a) inconsistent handling of inflation adjustment?

   (b) how much of correlations across countries is driven by σ_{mt}^2 and how much by R_{ft+1}?

   (c) is there really a long-lived component in g_{it} or is this an artifact of how the the dividend series has been "deseasonalized"?

   (d) Is σ_{mt}^2 priced?

8.2 "A Closed-Form GARCH Option Valuation Model" by Heston and Nandi

Reference: Heston and Nandi (2000) (HN); see also the work by Duan and coauthors (references in HN)

This paper derives an option price formula for an asset that follows a GARCH process. This is applied to S&P 500 index options, and it is found that the model works well compared to a Black-Scholes formula.

8.2.1 Option Price Formula

Over the period from t to $t+\Delta$ the change of log asset price minus a riskfree rate (including dividends/accumulated interest), that is, the continuously compounded excess return, follows a kind of GARCH(1,1)-M process

$$\ln S_t - \ln S_{t-\Delta} - r = \lambda h_t + \sqrt{h_t} z_t, \text{ where } z_t \text{ is iid } N(0,1) \tag{8.16}$$

$$h_t = \omega + \beta_1 h_{t-\Delta} + \alpha_1 (z_{t-\Delta} - \gamma_1 \sqrt{h_{t-\Delta}})^2. \tag{8.17}$$

The conditional variance would be a standard GARCH(1,1) process if $\gamma_1 = 0$. The additional term makes the response of h_t to an innovation symmetric around $\gamma_i \sqrt{h_{t-i\Delta}}$ instead of around zero. (HN also treat the case when the process is of higher order.)

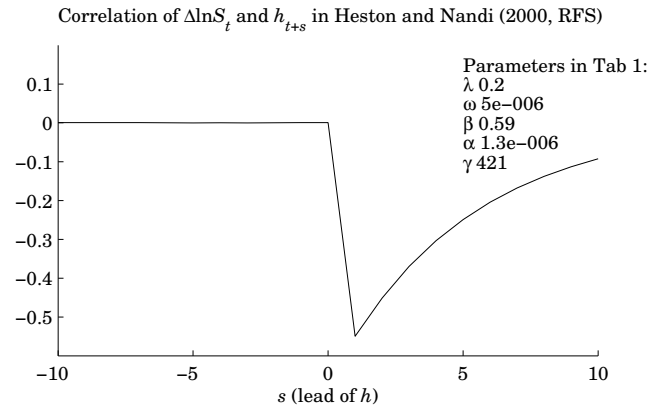


Figure 8.2: Simulated correlations of $\Delta \ln S_t$ and h_{t+s}

If $\gamma_1 > 0$ then the return, $\ln S_t - \ln S_{t-\Delta}$, is negatively correlated with subsequent volatility $h_{t+\Delta}$—as often observed in data. To see this, note that the effect on the return of z_t is linear, but that a negative z_t drives up the conditional variance $h_{t+\Delta} = \omega + \beta_1 h_t + \alpha_1 (z_t - \gamma_1 \sqrt{h_t})^2$ more than a positive z_t (if $\gamma_1 > 0$). The effect on the correlations is illustrated in *Figure 8.2*.

The process (8.16)–(8.17) does of course mean that the conditional (as of $t-\Delta$) distribution of the log asset price $\ln S_t$ is normally distributed. This is not enough to price options on this asset, since we cannot use a dynamic hedging approach to establish a no-arbitrage price since there are (by the very nature of the discrete model) jumps in the price of the underlying asset. Recall that the price on a call option with strike price K is

$$C_{t-\Delta} = \mathrm{E}_{t-\Delta} \{ M_t \max [S_t - K, 0] \}. \tag{8.18}$$

If $\ln S_t$ and $\ln M_t$ has bivariate normal, then the call option price can be written

$$C_{t-\Delta} = \mathrm{E}^*_{t-\Delta} \{ \max [S_t - K, 0] \}, \tag{8.19}$$

where $\mathrm{E}^*_{t-\Delta}$ is the expectations operator for the risk neutral distribution—which is such that $\ln S_t$ is normally distributed (with the same variance as in the objective distribution,

but a different mean). This means that the Black-Scholes formula holds. See, for instance, Huang and Litzenberger (1988).

Remark 38 *(Illustration of risk neutral distribution.) A risk neutral distribution is such that the expected payoff equals today's asset value. As an illustration, suppose the log stochastic discount factor ($m_t = \ln M_t$) and the log asset payoff ($x_t = \ln X_t$) have a bivariate normal distribution (conditional on the information set in $t-\Delta$). The asset price ($P_{t-\Delta}$) is then*

$$\begin{aligned} P_{t-\Delta} &= \mathrm{E}_{t-\Delta}\, M_t X_t = \mathrm{E}_{t-\Delta} \exp(m_t + x_t) \\ &= \exp\left[\mathrm{E}_{t-\Delta}\, m_t + \mathrm{E}_{t-\Delta}\, x_t + \mathrm{Var}_{t-\Delta}(m_t)/2 + \mathrm{Var}_{t-\Delta}(x_t)/2 + \mathrm{Cov}_{t-\Delta}(m_t, x_t)\right] \\ &= \exp\left[\mathrm{E}_{t-\Delta}\, m_t + \mathrm{Var}_{t-\Delta}(m_t)/2\right] \exp\left[\mathrm{E}_{t-\Delta}\, x_t + \mathrm{Var}_{t-\Delta}(x_t)/2 + \mathrm{Cov}_{t-\Delta}(m_t, x_t)\right]. \end{aligned}$$

For a real bond with a known payoff $X_t = 1$, the last term is unity. Since the price of the real bond is $\exp\left(-r_{ft-\Delta}\Delta\right)$, we can rewrite the equation as

$$\begin{aligned} P_{t-\Delta} &= \exp(-r_{ft-\Delta}\Delta) \exp\left[\mathrm{E}_{t-\Delta}\, x_t + \mathrm{Cov}_{t-\Delta}(m_t, x_t) + \mathrm{Var}_{t-\Delta}(x_t)/2\right] \\ &= \exp(-r_{ft-\Delta}\Delta)\, \mathrm{E}^*_{t-\Delta}\, X_t, \end{aligned}$$

where the distribution that defines the $\mathrm{E}^*_{t-\Delta}$ operator is such that $\ln X_t = x_t \sim N[\mathrm{E}_{t-\Delta}\, x_t + \mathrm{Cov}_{t-\Delta}(m_t, x_t), \mathrm{Var}_{t-\Delta}(x_t)]$. This shows that the risk neutral distribution has the same variance, but a different mean. If the covariance is negative so the asset has a positive risk premium (the price is low, see above), then the risk neutral distribution has a lower mean than the objective distribution.

This is effectively what HN assume, since they assume that the risk neutral distribution is normal

Assumption: the price in $t-\Delta$ of call option expiring in t follows BS.

HN show that the risk neutral process must then be as in (8.16)–(8.17), but with γ_1 replaced by $\gamma_1^* = \gamma_1 + \lambda + 1/2$ and λ replaced by $-1/2$ (not in γ_1^*, of course). This means that they use the assumption about the conditional (as of $t-\Delta$) distribution of S_t to build up a conditional (as of $t-\Delta$) distribution of S_{t+s} for any s. It is clear that if we know that the risk neutral process is of the form (8.16)–(8.17), but with other parameters, then

we can calculate the risk neutral distribution for any horizon (by clever tricks or simply by using (8.16)–(8.17) to do Monte Carlo simulations).

Once we have a risk neutral process it is (in principle, at least) straightforward to derive any option price (for any time to expiry). For a European call option with strike price K and expiry at date T, the result is

$$\begin{aligned} C_t(S_t, r, K, T) &= e^{-r(T-t)}\, \mathrm{E}^*_t \max\left[S_T - K, 0\right] && (8.20) \\ &= S_t P_1 - e^{-r(T-t)} K P_2, && (8.21) \end{aligned}$$

where P_1 and P_2 are two risk neutral probabilities (implied by the risk neutral version of (8.16)–(8.17), see above). It can be shown that P_2 is the risk neutral probability that $S_T > K$, and that P_1 is the delta, $\partial C_t(S_t, r, K, T)/\partial S_t$. In practice, HN calculate these probabilities by first finding the risk neutral characteristic function of S_T, $f(\phi) = \mathrm{E}^*_t \exp(i\phi \ln S_T)$, where $i^2 = -1$, and then inverting to get the probabilities.

Remark 39 *(Characteristic function and the pdf.) The characteristic function of a random variable x is*

$$\begin{aligned} f(\phi) &= \mathrm{E} \exp(i\phi x) \\ &= \textstyle\int_x \exp(i\phi x) pdf(x) dx, \end{aligned}$$

where $pdf(x)$ is the pdf. This is a Fourier transform of the pdf (if x is a continuous random variable). The pdf can therefore be recovered by the inverse Fourier transform as

$$pdf(x) = \frac{1}{2\pi} \textstyle\int_{-\infty}^{\infty} \exp(-i\phi x) f(\phi) d\phi.$$

8.2.2 Application to S&P 500 Index Option

Returns on the index are calculated by using official index plus dividends. The riskfree rate is taken to be a synthetic T-bill rate created by interpolating different bills to match the maturity of the option. Weekly data for 1992–1994 are used (created by using lots of intraday quotes for all Wednesdays).

HN estimate the "GARCH(1,1)-M" process (8.16)–(8.17) with ML on daily data on the S&P500 index returns. It is found that the β_i parameter is large, α_i is small, and that $\gamma_1 > 0$ (as expected). The latter seems to be important for the estimated h_t series (see Figures 1 and 2).

Instead of using the "GARCH(1,1)-M" process estimated from the S&P500 index returns, all the model parameters are subsequently estimated from option prices. Recall that the probabilities P_1 and P_2 in (8.21) depend (nonlinearly) on the parameters of the risk neutral version of (8.16)-(8.17). The model parameters can therefore be estimated by minimizing the sum (across option price observation) squared pricing errors.

In one of several different estimations, HN estimate the model on option data for the first half 1992 and then evaluate the model by comparing implied and actual option prices for the second half of 1992. These implied option prices use the model parameters estimated on data for the first half of the year and an estimate of h_t calculated using these parameters and the latest S&P 500 index returns. The performance of this model is compared with a Black-Scholes model (among other models,) where the implied volatility in week $t-1$ is used to price options in price t. This exercise is repeated for 1993 and 1994.

It is found that the GARCH model outperforms (in terms of MSE) the B-S model. In particular, it seems as if the GARCH model gives much smaller errors for deep out-of-the-money options (see Figures 2 and 3). HN argue that this is due to two aspects of the model: the time-profile of volatility (somewhat persistent, but mean-reverting) and the negative correlation of returns and volatility.

Bibliography

Bansal, R., and C. Lundblad, 2002, "Market Efficiency, Fundamental Values, and the Size of the Risk Premium in Global Equity Markets," *Journal of Econometrics*, 109, 195–237.

Heston, S. L., and S. Nandi, 2000, "A Closed-Form GARCH Option Valuation Model," *Review of Financial Studies*, 13, 585–625.

Huang, C.-F., and R. H. Litzenberger, 1988, *Foundations for Financial Economics*, Elsevier Science Publishing, New York.

9 Models of Short Interest Rates

9.1 SDF and Yield Curve Models

A yield curve model is essentially a model for the conditional mean of the SDF. This section summarizes some basic results.

9.1.1 Preliminaries

Reference: Cochrane (2001) 19.3; Campbell, Lo, and MacKinlay (1997) 11.1

The price in t of a bill paying 1 unit of account (for sure) in $t+n$ is

$$P_{nt} = \mathrm{E}_t\, M_{t+1}M_{t+2}\ldots M_{t+n}1 = \mathrm{E}_t \exp(m_{t+1}+m_{t+2}+\ldots+m_{t+n})1, \tag{9.1}$$

where m_{t+1} is the log SDF. (To see this note that $P_{2t} = \mathrm{E}_t\, M_{t+1}P_{1t+1}$ and that $P_{1t+1} = \mathrm{E}_{t+1}\, M_{t+2}1$. Combining gives $P_{2t} = \mathrm{E}_t\, M_{t+1}\,\mathrm{E}_{t+1}\, M_{t+2} = \mathrm{E}_t\, M_{t+1}M_{t+2}$ by iterated expectations.)

A continuously compounded n-period interest rate, y_{nt} satisfies

$$\exp(-ny_{nt}) = P_{nt}. \tag{9.2}$$

The SDF is a real SDF if the interest rates are real, otherwise it is a nominal SDF.

9.1.2 When the log SDF is Normally Distributed

Assume that $m_{t+1}+m_{t+2}+\ldots+m_{t+n}$ in (9.1) is conditionally normally distributed

$$(m_{t+1}+m_{t+2}+\ldots+m_{t+n})|\Omega_t \sim N[\mu_{nt}, s_{nt}^2]. \tag{9.3}$$

We can then evaluate (9.1)–(9.2) as

$$\exp(-ny_{nt}) = P_{nt} = \mathrm{E}_t \exp(m_{t+1}+m_{t+2}+\ldots+m_{t+n}) = \exp(\mu_{nt}+s_{nt}^2/2), \text{ or}$$

$$y_{nt} = -\frac{\mu_{nt}}{n} - \frac{s_{nt}^2}{2n}. \tag{9.4}$$

Example 40 *(When m_t is iid.) When m_t is iid $N(\mu, \sigma^2)$, then $\mu_{nt} = n\mu$ and $s_{nt}^2 = n\sigma^2$ for all n and t. From (9.4) we then get $y_{nt} = -\mu - \sigma^2/2$ so all interest rates are equal (across maturity and time). This would happen, for instance, in a consumption-based model with CRRA utility and iid consumption growth.*

9.1.3 When the log SDF is an AR(1)

Suppose the log SDF is an AR(1)

$$m_{t+1} = \rho m_t + \varepsilon_{t+1}, \text{ where } \varepsilon_{t+1} \text{ is iid } N(0, \sigma^2). \tag{9.5}$$

I have suppressed the intercept in order to minimize the notational clutter. This means that the log SDF has a zero mean, so the SDF level has a mean slightly above one. This process gives

$$\mathrm{E}_t\, m_{t+1} = \rho m_t \text{ and } \mathrm{Var}_t(m_{t+1}) = \sigma^2. \tag{9.6}$$

The expectation and variance in (9.6) correspond to μ_{1t} and s_{1t}^2 in (9.4) (note $n = 1$), so we can write the one-period interest rate as

$$y_{1t} = -\rho m_t - \frac{\sigma^2}{2}. \tag{9.7}$$

This shows that y_{1t} is an AR(1), since m_t is. Specifically, use (9.7) in (9.5) to get

$$y_{1t+1} = \rho y_{1t} - \frac{\sigma^2}{2}(1-\rho) - \rho\varepsilon_{t+1}, \text{ where } \varepsilon_{t+1} \text{ is iid } N(0, \sigma^2). \tag{9.8}$$

This is a discrete time version of the model in Vasicek (1977).

To derive the two-period interest rate, we note from (9.5) that

$$\begin{aligned} m_{t+1} + m_{t+2} &= m_{t+1} + (\rho m_{t+1} + \varepsilon_{t+2}) = (1+\rho)m_{t+1} + \varepsilon_{t+2}, \text{ so} \\ \mathrm{E}_t(m_{t+1} + m_{t+2}) &= (1+\rho)\,\mathrm{E}_t\, m_{t+1} = (1+\rho)\rho m_t, \text{ and} \qquad (9.9) \\ \mathrm{Var}_t(m_{t+1} + m_{t+2}) &= (1+\rho)^2\,\mathrm{Var}_t(m_{t+1}) + \sigma^2 = (1+\rho)^2\sigma^2 + \sigma^2. \qquad (9.10) \end{aligned}$$

The expectation and variance in (9.9)–(9.10) correspond to μ_{2t} and s_{2t}^2 in (9.4) (note $n = 2$). We can then write the two-period interest rate as

$$\begin{aligned} y_{2t} &= -\frac{(1+\rho)\rho m_t}{2} - \frac{(1+\rho)^2\sigma^2 + \sigma^2}{4}, \text{ or (by using (9.7))} \\ &= \frac{(1+\rho)y_{1t}}{2} - \frac{(1+\rho)^2\sigma^2 + \sigma^2 + (1+\rho)\sigma^2}{4}. \end{aligned} \tag{9.11}$$

This shows that both one-period and two-period rates are affine functions of m_t—the only state variable in this model. This holds also for longer interest rates. We can therefore write all interest rates as an affine function of the short rate (or any other rate, for that matter). This affine function need not be monotonic, but all changes in the yield curve give parallel shifts.

Three of the most important testable implications of this model are that *(i)* the short rate is a homoskedastic AR(1); *(ii)* the shifts in the yield curve are parallel; and *(iii)* there are cross sectional restrictions on how interest rates respond to changes in the short interest rate (for instance, y_{2t} and y_{3t} have coefficients which are different functions of ρ)—and how the slope of the yield curve looks on average (see the second term in (9.11)).

9.1.4 When the log SDF is an AR(1) with Square-Root Heteroskedasticity

We now change the AR(1) in (9.5) to allow for conditional heteroskedasticity

$$m_{t+1} - \delta = \rho(m_t - \delta) + \varepsilon_{t+1}, \text{ where } \varepsilon_{t+1} \sim N(0, -m_t\sigma^2), \tag{9.12}$$

where we cheat a bit by assuming that $m_t \leq 0$. (Recall that $1 = \mathrm{E}_t\, M_{t+1}R_{ft+1}$, so the gross riskfree rate is $1/\mathrm{E}_t\, M_{t+1}$ which is typically larger than unity. The assumption made in (9.12) amounts to assuming that $1/M_{t+1} \geq 1$.) It could actually happen that a big ε_t makes m_t positive, but that is very unlikely. This is a "square-root process" (see Cox, Ingersoll, and Ross (1985)) where the conditional standard deviation is proportional to $\sqrt{-m_t}$. It implies

$$\mathrm{E}_t\, m_{t+1} = \rho m_t + (1-\rho)\,\delta \text{ and } \mathrm{Var}_t(m_{t+1}) = -m_t\sigma^2. \tag{9.13}$$

We use this in (9.4) for $n = 1$ to get

$$\begin{aligned} y_{1t} &= -\rho m_t - (1-\rho)\,\delta + \frac{m_t \sigma^2}{2} \\ &= -m_t(\rho - \sigma^2/2) - (1-\rho)\,\delta. \end{aligned} \tag{9.14}$$

The short interest rate is once again an affine function of m_t. It can be shown that all longer rates also are affine functions of m_t too (the assumption of conditional, as of t, normality of $m_{t+1} + m_{t+2} + \ldots + m_{t+n}$ does not work, however). Therefore, all changes in the yield curve are still parallel shifts.

Would it be possible to generalize the heteroskedasticity in (9.12) and preserve the affine structure? No, the only cases that generate an affine structure are $\varepsilon_{t+1} \sim N(0, (-m_t)^{2\gamma}\,\sigma^2)$ for $\gamma = 0$ and $\gamma = 1/2$. In all other cases, (9.14) is nonlinear in m_t.

To find the implications for the conditional variance, we use (9.14) in (9.12) to get

$$y_{1t+1} = \rho y_{1t} - (\rho - \sigma^2/2)\varepsilon_{t+1}, \text{ where } \varepsilon_{t+1} \sim N\left[0,\, y_{1t}\frac{\sigma^2}{\rho - \sigma^2/2}\right], \tag{9.15}$$

which shows that the short interest rate is a heteroskedastic AR(1), where the variance is proportional to the level of the interest rate.

To get away from the feature that the conditional variance is linear in the interest rate level, we need at least a two-factor model, but it could still be an affine model. Another advantage of a multi-factor model is that it can generate more interesting movements in the yield curve than the parallel shifts in affine one-factor models.

For instance, the model of Longstaff and Schwarz (1992) can be generated by a log SDF that is a sum of two AR(1) processes with square-root heteroskedasticity. In that case, the rank of the covariance matrix of the interest rates is equal to two, which is an important testable implication. More generally, the rank of the conditional covariance matrix is equal to the number of factors, K. This means that we can express the spot interest rate (for any maturity) as an affine function of K other interest rates, which should be relatively straightforward to test (think of setting up a system of n interest rates and testing if they can all be explained by two of them—similar to a linear factor model for stocks).

9.2 "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate" by Chan et al (1992)

Reference: Chan, Karolyi, Longstaff, and Sanders (1992) (CKLS); Dahlquist (1996)

The models that CKLS study have the following dynamics (under the objective probability measure) of the one-period interest rate, y_{1t}

$$y_{1t+1} - y_{1t} = \alpha + \beta y_{1t} + \varepsilon_{t+1}, \text{ where } \mathrm{E}_t\,\varepsilon_{t+1} = 0 \text{ and } \mathrm{E}_t\,\varepsilon_{t+1}^2 = \sigma^2 y_{1t}^{2\gamma}. \tag{9.16}$$

This formulation nests several well-known models. It is an approximation of the diffusion process

$$dr(t) = [\beta_0 + \beta_1 r(t)]dt + \sigma r(t)^{\gamma} dW(t),$$

where $W(t)$ is a Wiener process. (For an introduction to the issue of being more careful with estimating a continuous time model on discrete data, see Campbell, Lo, and MacKinlay (1997) 9.3 and Harvey (1989) 9. In some cases, like the homoskedastic AR(1), there is no approximation error because of the discrete sampling. In other cases, there is an error.)

CKLS estimate the model with GMM, using the following moment conditions

$$g_t(\alpha, \beta, \gamma, \sigma^2) = \begin{bmatrix} \varepsilon_{t+1} \\ \varepsilon_{t+1}^2 - \sigma^2 y_{1t}^{2\gamma} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ y_{1t} \end{bmatrix} = \begin{bmatrix} \varepsilon_{t+1} \\ \varepsilon_{t+1} y_{1t} \\ \varepsilon_{t+1}^2 - \sigma^2 y_{1t}^{2\gamma} \\ (\varepsilon_{t+1}^2 - \sigma^2 y_{1t}^{2\gamma}) y_{1t} \end{bmatrix}, \tag{9.17}$$

so there are four moment conditions and four parameters. The choice of the instruments (1 and y_{1t}) is clearly somewhat arbitrary since any variables in the information set in t would do.

CKLS estimate this model in various forms (imposing different restrictions on the parameters) on monthly data on one-month T-bill rates for 1964–1989. They find that both $\hat{\alpha}$ and $\hat{\beta}$ are close to zero (in the unrestricted model $\hat{\beta} < 0$ and almost significantly different from zero—indicating mean-reversion). They also find that $\hat{\gamma} > 1$ and significantly so. This is problematic for the affine one-factor models, since they require $\gamma = 0$ or $\gamma = 0.5$. A word of caution: the estimated parameter values suggest that the interest rate is non-stationary, so the properties of GMM are not really known. In particular, the estimator is

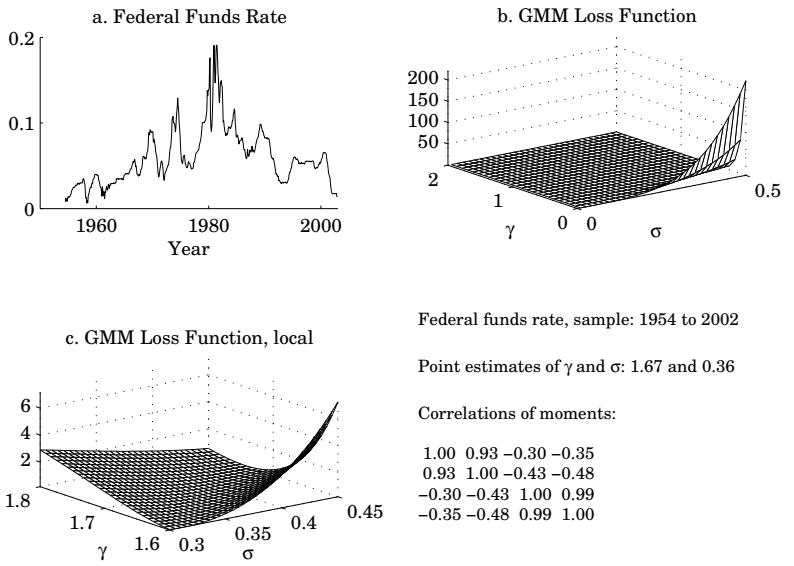


Figure 9.1: Federal funds rate, monthly data, $\alpha = \beta = 0$ imposed

probably not asymptotically normally distributed—and the model could easily generate extreme interest rates.

Some results obtained from re-estimating the model on a longer data set are found in Figure 9.1. In this figure, $\alpha = \beta = 0$ is imposed, but the results are very similar if this is relaxed. One of the first thing to note is that the loss function is very flat in the $\gamma \times \sigma$ space—the parameters are not pinned down very precisely by the model/data. Another way to see this is to note that the moments in (9.17) are very strongly correlated: moment 1 and 2 have a very strong correlation, and this is even worse for moments 3 and 4. The latter two moment conditions are what identifies σ^2 from γ, so it is a serious problem for the estimation. The reason for these strong correlations is probably that the interest rate series is very persistent so, for instance, ε_{t+1} and $\varepsilon_{t+1} y_{1t}$ look very similar (as y_{1t} tends

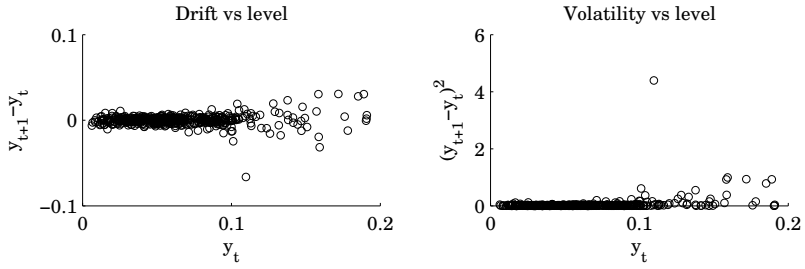


Figure 9.2: Federal funds rate, monthly data

to be fairly constant due to the persistence).

Figure 9.2, which shows cross plots of the interest rate level and the change and volatility in the interest rate, suggests that some of the results might be driven by outliers. There is, for instance, a big volatility outlier in May 1980 and most of the data points with high interest rate and high volatility are probably from the Volcker deflation in the early 1980s. It is unclear if that particular episode can be modelled as belonging to the same regime as the rest of the sample (in particular since the Fed let the interest rate fluctuate a lot more than before). Maybe this episode needs special treatment.

Bibliography

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay, 1997, *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey.

Chan, K. C., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders, 1992, "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," *Journal of Finance*, 47, 1209–1227.

Cochrane, J. H., 2001, *Asset Pricing*, Princeton University Press, Princeton, New Jersey.

Cox, J. C., J. E. Ingersoll, and S. A. Ross, 1985, "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385–407.

Dahlquist, M., 1996, "On Alternative Interest Rate Processes," *Journal of Banking and Finance*, 20, 1093–1119.

Harvey, A. C., 1989, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Longstaff, F. A., and E. S. Schwarz, 1992, "Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model," *Journal of Finance*, 47, 1259–1282.

Vasicek, O. A., 1977, "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177–188.

10 Kernel Density Estimation and Regression

10.1 Non-Parametric Regression

Reference: Campbell, Lo, and MacKinlay (1997) 12.3; Härdle (1990); Pagan and Ullah (1999)

10.1.1 Simple Kernel Regression

Non-parametric regressions are used when we are unwilling to impose a parametric form on the regression equation—and we have a lot of data.

Let the scalars y_t and x_t be related as

$$y_t = m(x_t) + \varepsilon_t,\ \varepsilon_t \text{ is iid and } \mathrm{E}\,\varepsilon_t = \mathrm{Cov}\left[m(x_t), \varepsilon_t\right] = 0, \tag{10.1}$$

where $m()$ is a, possibly non-linear, function.

Suppose the sample had 3 observations (say, $t = 3$, 27, and 99) with exactly the same value of x_t, say 2.1. A natural way of estimating $m(x)$ at $x = 2.1$ would then be to average over these 3 observations as we can expect average of the error terms to be close to zero (iid and zero mean).

Unfortunately, we seldom have repeated observations of this type. Instead, we may try to approximate the value of $m(x)$ (x is a single value, 1.9, say) by averaging over observations where x_t is close to x. The general form of this type of estimator is

$$\hat{m}(x) = \frac{\sum_{t=1}^{T} w(x_t - x) y_t}{\sum_{t=1}^{T} w(x_t - x)}, \tag{10.2}$$

where $w(x_t - x)/\Sigma_{t=1}^{T} w(x_t - x)$ is the weight given to observation t. Note that denominator makes the weights sum to unity.

As an example of a $w(.)$ function, it could give equal weight to the k values of x_t which are closest to x and zero weight to all other observations (this is the "k-nearest neighbor" estimator, see Härdle (1990) 3.2). As another example, the weight function could be defined so that it trades off the expected squared errors, $\mathrm{E}[y_t - \hat{m}(x)]^2$, and the

expected squared acceleration, $\mathrm{E}[d^2\hat{m}(x)/dx^2]^2$. This defines a cubic spline (and is often used in macroeconomics, where $x_t = t$ and is then called the Hodrick-Prescott filter).

A *Kernel regression* uses a pdf as the weight function $w(.)$. The pdf of $N(0, h^2)$ is commonly used, where the choice of h allows us to easily vary the relative weights of different observations. This weighting function is positive so all observations get a positive weight, but the weights are highest for observations close to x and then tapers of in a bell-shaped way. A low value of h means that the weights taper off fast—the weight function is then a normal pdf with a low variance. With this particular kernel, we the following estimator of $m(x)$ at a point x

$$\hat{m}(x) = \frac{\sum_{t=1}^{T} K_h(x_t - x)y_t}{\sum_{t=1}^{T} K_h(x_t - x)}, \text{ where } K_h(x_t - x) = \frac{\exp[-(x_t - x)^2/(2h^2)]}{h\sqrt{2\pi}}. \quad (10.3)$$

Note that $K_h(x_t - x)$ corresponds to the weighting function $w(x_t - x)$ in (10.2).

In practice we have to estimate $\hat{m}(x)$ at a finite number of points x. This could, for instance, be 100 evenly spread points in the interval between the min and max values observed in the sample. Special corrections might be needed if there are a lot of observations stacked close the boundary of the support of x (see Härdle (1990) 4.4).

Kernel regressions are typically consistent, provided longer samples are accompanied by smaller values of h, so the weighting function becomes more and more local as the sample size increases. It can be shown (see Härdle (1990) 3.1) that the mean squared error of the estimator is approximately (for general kernel functions)

$$\begin{aligned}\mathrm{E}[\hat{m}(x) - m(x)]^2 &= \text{variance} + \text{bias}^2 \\ &= \frac{\mathrm{Var}(x)}{hT} \times \text{“peakedness of kernel”} + h^4 \times \text{“curvature of } m(x)\text{”}.\end{aligned} \quad (10.4)$$

A smaller h increases the variance (we effectively use fewer data points to estimate $m(x)$) but decreases the bias of the estimator (it becomes more local to x). If h decreases less than proportionally with the sample size (so hT in the denominator of the first term increases with T), then the variance goes to zero and the estimator is consistent (since the bias in the second term decreases as h does).

It can also be noted that the bias increases with the curvature of the $m(x)$ function (that is, the average $|d^2m(x)/dx^2|$). This makes sense, since rapid changes of the slope of $m(x)$

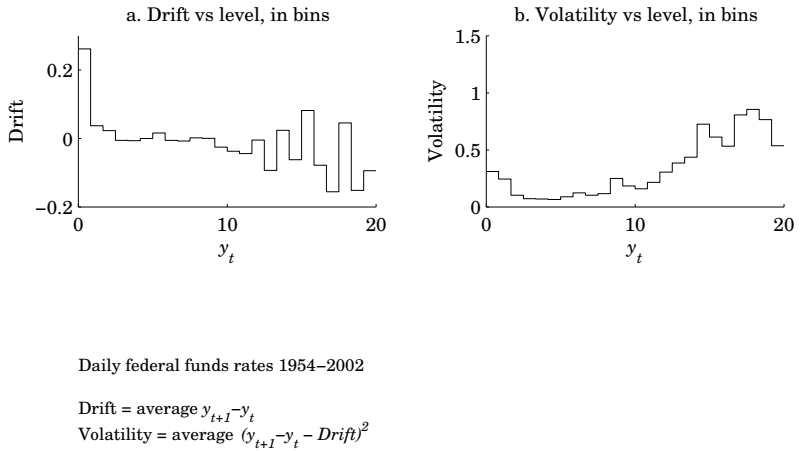


Figure 10.1: Federal funds rate, daily data 1954-2002

makes it hard to get it right by averaging of nearby x values. The variance is a function of the kernel only and not of the $m(x)$ function. In particular, the more concentrated the kernel is (larger “peakedness”) around x (for a given h), the less information is used in forming the average around x, and the uncertainty is therefore larger. See Section 10.2.3 for a further discussion of the choice of h.

See *Figures 10.1–10.2* for an example. Note that the volatility is defined as the square of the drift minus expected drift (from the same estimation method).

10.1.2 Multivariate Kernel Regression

Suppose that y_t depends on two variables (x_t and z_t)

$$y_t = m(x_t, z_t) + \varepsilon_t, \ \varepsilon_t \text{ is iid and } \mathrm{E}\,\varepsilon_t = 0. \quad (10.5)$$

This makes the estimation problem much harder since there are typically few observations in every bivariate bin (rectangle) of x and z. For instance, with as little as a 20 intervals of each of x and z, we get 400 bins, so we need a large sample to have a reasonable number

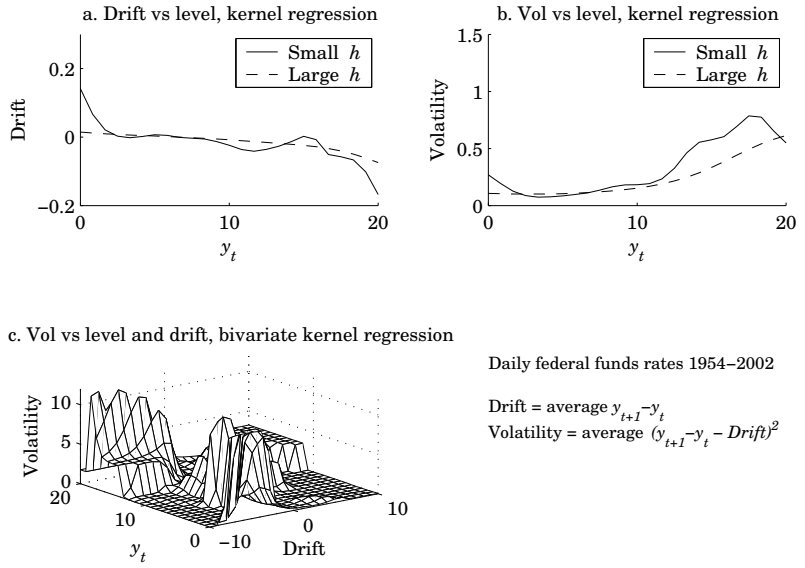


Figure 10.2: Federal funds rate, daily data 1954-2002

of observations in every bin.

In any case, the most common way to implement the kernel regressor is to let

$$\hat{m}(x,z) = \frac{\sum_{t=1}^{T} K_{h_x}(x_t - x)K_{h_z}(z_t - z)y_t}{\sum_{t=1}^{T} K_{h_x}(x_t - x)K_{h_z}(z_t - z)}, \tag{10.6}$$

where $K_{h_x}(x_t - x)$ and $K_{h_z}(z_t - z)$ are two kernels like in (10.3) and where we may allow h_x and h_y to be different (and depend on the variance of x_t and y_t). In this case. the weight of the observation (x_t, z_t) is proportional to $K_{h_x}(x_t - x)K_{h_z}(z_t - z)$, which is high if both x_t and y_t are close to x and y respectively.

See *Figure 10.2* for an example.

10.2 Estimating and Testing Distributions

Reference: Harvey (1989) 260, Davidson and MacKinnon (1993) 267, Silverman (1986); Mittelhammer (1996), DeGroot (1986)

10.2.1 Parametric Tests of Normal Distribution

The skewness, kurtosis and Bera-Jarque test for normality are useful diagnostic tools. For an iid normally distributed variable, $x_t \sim$iid $N\left(\mu, \sigma^2\right)$, they are

$$\begin{array}{lcll} & & \text{Test statistic} & \text{Distribution} \\ \text{skewness} & = & \frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_t-\mu}{\sigma}\right)^3 & N\left(0, \frac{6}{T}\right) \\ \text{kurtosis} & = & \frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_t-\mu}{\sigma}\right)^4 & N\left(3, \frac{24}{T}\right) \\ \text{Bera-Jarque} & = & \frac{T}{6}\text{skewness}^2 + \frac{T}{24}(\text{kurtosis}-3)^2 & \chi^2_2. \end{array} \tag{10.7}$$

This is implemented by using the estimated mean and standard deviation. The distributions stated on the right hand side of are under the null hypothesis that $x_t \sim$iid $N\left(\mu, \sigma^2\right)$.

The intuition for the χ^2_2 distribution of the Bera-Jarque test is that both the skewness and kurtosis are, if properly scaled, $N(0,1)$ variables. It can also be shown that they, under the null hypothesis, are uncorrelated. The Bera-Jarque test statistic is therefore a sum of the square of two uncorrelated $N(0,1)$ variables, which has a χ^2_2 distribution.

The Bera-Jarque test can also be implemented as a test of overidentifying restrictions in GMM. Formulate the moment conditions

$$g(\mu,\sigma^2) = \frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} x_t - \mu \\ (x_t-\mu)^2 - \sigma^2 \\ (x_t-\mu)^3 \\ (x_t-\mu)^4 - 3\sigma^4 \end{bmatrix}, \tag{10.8}$$

which should all be zero if x_t is $N(\mu, \sigma^2)$. We can estimate the two parameters, μ and σ^2, by using the first two moment conditions only, and then test if all four moment conditions are satisfied. It can be shown that this is the same as the Bera-Jarque test if x_t is indeed iid $N(\mu, \sigma^2)$.

10.2.2 Non-Parametric Tests of General Distributions

The *Kolmogorov-Smirnov* test is designed to test if an empirical distribution function, EDF(x) conforms with a theoretical cdf, $F(x)$. The empirical distribution function is defined as the fraction of observations which are less or equal to x, that is,

$$\text{EDF}(x) = \frac{1}{T}\sum_{t=1}^{T}\delta(x_t \leq x), \text{ where } \delta(q) = \begin{cases} 1 \text{ if } q \text{ is true} \\ 0 \text{ else.} \end{cases} \tag{10.9}$$

(The EDF(x_t) and $F(x_t)$ are sometimes plotted against the sorted (in ascending order) sample $\{x_t\}_{t=1}^{T}$.)

Define the absolute value of the maximum distance

$$D_T = \max_{x_t} |\text{EDF}(x_t) - F(x_t)|. \tag{10.10}$$

We reject the null hypothesis that EDF(x) $= F(x)$ if $\sqrt{T}D_t > c$, where c is a critical value which can be calculated from

$$\lim_{T\to\infty} \Pr\left(\sqrt{T}D_T \leq c\right) = 1 - 2\sum_{i=1}^{\infty}(-1)^{i-1}e^{-2i^2c^2}. \tag{10.11}$$

It can be approximated by replacing ∞ with a large number (for instance, 100). For instance, $c = 1.35$ provides a 5% critical value. There is a corresponding test for comparing two empirical cdfs.

Pearson's χ^2 *test* does the same thing as the K-S test but for a discrete distribution. Suppose you have K categories with N_i values in category i. The theoretical distribution predicts that the fraction p_i should be in category i, with $\sum_{i=1}^{K} p_i = 1$. Then

$$\sum_{i=1}^{K}\frac{(N_i - Tp_i)^2}{Tp_i} \sim \chi^2_{K-1}. \tag{10.12}$$

There is a corresponding test for comparing two empirical distributions.

10.2.3 Kernel Density Estimation

Reference: Silverman (1986)

A histogram is just a count of the relative number of observations that fall in (pre-

specified) non-overlapping intervals. If we also divide by the width of the interval, then the area under the histogram is unity, so the scaled histogram can be interpreted as a density function. For instance, if the intervals ("bins") are a wide, then the scaled histogram can be defined as

$$g(x|x \text{ is in bin}_i) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{a}\delta(x_t \text{ is in bin}_i). \tag{10.13}$$

Note that the area under $g(x)$ indeed integrates to unity.

This is clearly a kind of weighted (with binary weights) average of all observations. As with the kernel regression, we can gain efficiency by using a more sophisticated weighting scheme. In particular, using a pdf as a weighting function is often both convenient and efficient. The $N(0, h^2)$ is often used. The kernel density estimator of the pdf at some point x is then

$$\hat{f}(x) = \frac{1}{Th\sqrt{2\pi}}\sum_{t=1}^{T}\exp[-(x - x_t)^2/(2h^2)]. \tag{10.14}$$

The value $h = \text{Std}(x_t)1.06T^{-1/5}$ is sometimes recommended, since it can be shown to be the optimal choice (in MSE sense) if data is normally distributed and the $N(0, h^2)$ kernel is used. A more ambitious approach is to choose h by a leave-one-out cross-validation technique (see Silverman (1986)).

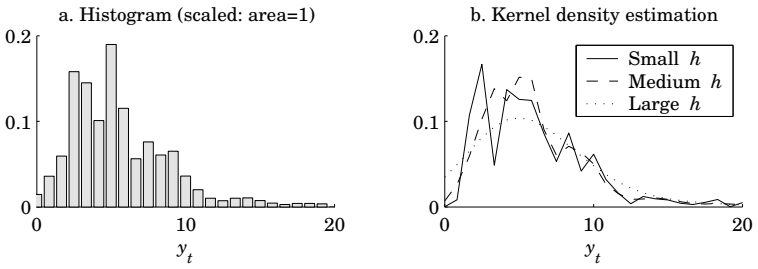
The results on bias and variance in (10.4) are approximately true also for the kernel density estimation if we interpret $m(x)$ as the pdf of x.

The easiest way to handle a bounded support of x is to transform the variable into one with an unbounded support, estimate the pdf for this variable, and then use the "change of variable" technique to transforms to the pdf of the original variable.

We can also estimate multivariate pdfs. Let x_t be a $d \times 1$ matrix and $\hat{\Omega}$ be the estimated covariance matrix of x_t. We can then estimate the pdf at a point x by using a multivariate Gaussian kernel as

$$\hat{f}(x) = \frac{1}{Th^d(2\pi)^{d/2}|\hat{\Omega}|^{1/2}}\sum_{t=1}^{T}\exp[-(x - x_t)'\hat{\Omega}^{-1}(x - x_t)/(2h^2)]. \tag{10.15}$$

The value $h = 0.96T^{-1/(d+4)}$ is sometimes recommended.



Daily federal funds rates 1954–2002

K–S (against N(μ,σ²)): sqrt(T)D = 14.9
Skewness: 1.258
kurtosis: 5.22
Bera–Jarque: 8301

Figure 10.3: Federal funds rate, daily data 1954-2002

10.3 "Testing Continuous-Time Models of the Spot Interest Rate," by Ait-Sahalia (1996)

Reference: Ait-Sahalia (1996)

This paper tests several models of the short interest rate by using a non-parametric technique.

1. The first step of the analysis is to estimate the unconditional distribution of the short interest rate by a kernel density estimator like (10.14). The estimated pdf at the value r is denoted $\hat{\pi}_0(r)$.

2. The second step is to estimate the parameters in a short rate model (for instance, Vasicek's model) by making the unconditional distribution implied by the model parameters (denoted $\pi(\theta, r)$ where θ is a vector of the model parameters and r a value of the short rate) as close as possible to the non-parametric estimate obtained in step 1. This is done by choosing the model parameters as

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{T} \sum_{t=1}^{T} [\pi(\theta, r_t) - \hat{\pi}_0(r)]^2. \tag{10.16}$$

3. The model is tested by using a scaled version of the minimized value of the right hand side of (10.16) as a test statistic (it has an asymptotic normal distribution).

4. It is found that most standard models are rejected (daily data on 7-day Eurodollar deposit rate, June 1973 to February 1995, 5,500 observations), mostly because actual mean reversion is much more non-linear in the interest rate level than suggested by most models. The actual drift and volatility are much like those in Figure 10.2.

10.4 "Nonparametric Estimation of Stat-Price Densities Implicit in Financial Asset Prices," by Ait-Sahalia and Lo (1998)

Reference: Ait-Sahalia and Lo (1998)

This paper estimates non-parametric option price functions and calculates the implicit risk-neutral distribution as the second partial derivative of this function with respect to the strike price.

1. First, the call option price, H_{it}, is estimated as a multivariate kernel regression

$$H_{it} = m(S_t, X, \tau, r_{\tau t}, \delta_{\tau t}) + \varepsilon_{it}, \tag{10.17}$$

where S_t is the price of the underlying asset, X is the strike price, τ is time to expiry, $r_{\tau t}$ is the interest rate between t and $t + \tau$, and $\delta_{\tau t}$ is the dividend yield (if any) between t and $t + \tau$. It is very hard to estimate a five-dimensional kernel regression, so various ways of reducing the dimensionality are tried. For instance, by making $m()$ a function of the forward price, $S_t[\tau \exp(r_{\tau t} - \delta_{\tau t})]$, instead of S_t, $r_{\tau t}$, and $\delta_{\tau t}$ separably.

2. Second, the implicit riskneutral pdf of the future asset price is calculated as $\partial^2 m(S_t, X, \tau, r_{\tau t}, \delta_{\tau t})/\partial X^2$, properly scaled so it integrates to unity.

3. This approach is used on daily data for Jan 1993 to Dec 1993 on S&P 500 index options (14,000 observations). They find interesting patterns of the implied moments (mean, volatility, skewness, and kurtosis) as the time to expiry changes. In particular, the non-parametric estimates suggest that distributions for longer horizons have increasingly larger skewness and kurtosis: whereas the distributions for short horizons are not too different from normal distributions, this is not true for longer horizons. (See their Fig 7.)

4. They also argue that there is little evidence of instability in the implicit pdf over their sample.

Bibliography

Ait-Sahalia, Y., 1996, "Testing Continuous-Time Models of the Spot Interest Rate," *Review of Financial Studies*, 9, 385–426.

Ait-Sahalia, Y., and A. W. Lo, 1998, "Nonparametric Estimation of Stat-Price Densities Implicit in Financial Asset Prices," *Journal of Finance*, 53, 499–547.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay, 1997, *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey.

Davidson, R., and J. G. MacKinnon, 1993, *Estimation and Inference in Econometrics*, Oxford University Press, Oxford.

DeGroot, M. H., 1986, *Probability and Statistics*, Addison-Wesley, Reading, Massachusetts.

Härdle, W., 1990, *Applied Nonparametric Regression*, Cambridge University Press, Cambridge.

Harvey, A. C., 1989, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Mittelhammer, R. C., 1996, *Mathematical Statistics for Economics and Business*, Springer-Verlag, New York.

Pagan, A., and A. Ullah, 1999, *Nonparametric Econometrics*, Cambridge University Press.

Silverman, B. W., 1986, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

21 Finite-difference Solution of Option Prices

21.1 Black-Scholes

We are interested in the price of a European call option with strike price K. The price of the underlying asset follows a geometric Brownian motion

$$dx(t) = \alpha x(t)\,dt + \sigma x(t)\,dW(t), \tag{21.1}$$

where $W(t)$ is a Wiener process. Suppose the risk free rate is a constant, r. The notation (for instance, that the underlying asset price is denoted by x) follows the convention in most texts on finite-difference methods.

Equilibrium requires that the option price function, $C(x,t)$ satisfies the following partial differential equation (PDE)

$$0 = -rC(x,t) + \frac{\partial C(x,t)}{\partial t} + rx(t)\frac{\partial C(x,t)}{\partial x} + \frac{1}{2}x(t)^2\sigma^2\frac{\partial^2 C(x,t)}{\partial x^2}, \tag{21.2}$$

with the boundary conditions (at expiration date T)

$$C(x,T) = \max[x(T) - K, 0]. \tag{21.3}$$

We know that the solution is

$$C(x,t) = x(t)\Phi(d_1) - e^{-r(T-t)}K\Phi\left(d_1 - \sigma\sqrt{T-t}\right), \text{ where} \tag{21.4}$$
$$d_1 = \frac{\ln[x(t)/K] + \left(r + \sigma^2/2\right)(T-t)}{\sigma\sqrt{T-t}}.$$

It will also turn out to be useful to note that

$$C(0,t) = 0 \text{ and } \lim_{x\to\infty} C(x,t) = \lim_{x\to\infty} x - e^{-r(T-t)}K. \tag{21.5}$$

21.2 Finite-difference Methods

References: Duffie (1992), appendix 10.H; Wilmott, Howison, and Dewynne (1995) 8; Press, Teukolsky, Vetterling, and Flannery (1992) 19.2; Hull (2000) 16.8

The partial differential equation we will solve for $F(x,t)$ is

$$0 = -r(x,t)F(x,t) + \frac{\partial F(x,t)}{\partial t} + \mu(x,t)\frac{\partial F(x,t)}{\partial x} + \frac{1}{2}\sigma(x,t)^2\frac{\partial^2 F(x,t)}{\partial x^2} + h(x,t). \tag{21.6}$$

This is more general than the Black-Scholes model, and will allow us to price more complicated derivatives.

Example 41 *(The Black-Scholes model.) To translate from the Black-Scholes model in (21.2) to the pde in (21.6), note that $F(x,t) = C(x,t)$, $r(x,t) = r$, $\mu(x,t) = rx(t)$, and $\sigma(x,t)^2 = x(t)^2\sigma^2$. Also, there are no dividends so $h(x,t)$ should be zero (except possibly at the x grid boundaries—more about that later).*

The idea of any finite difference method is to replace the derivatives with finite differences defined on a grid. We will calculate the function $F(x,t)$ at the discrete set of points (x_i, t_j) with $i = 1, ..., N$ and $j = 1, ..., M$ (typically from today to the expiration date). The distance between x_i and x_{i-1} equals the constant Δx and similarly for $t_j - t_{j-1} = \Delta t$. These are the so called mesh sizes.

Example 42 *(The Black-Scholes model.) The x grid for the Black-Scholes model runs from $x_1 = 0$ (or something very close to zero) to $x_N =$ a large number; the t grid runs from $t_1 = t$ (trade date) to $t_M = T$ (expiration date).*

The solution is obtained by solving one period at a time, starting with t_M and working backward in time. In the last period, the expiration date, we typically know the solution. For the Black-Scholes model, it is given by (21.3). The rest of this section is about how to solve resursively for $t_{M-1}, t_{M-2}, ..., t_1$.

We will consider three different methods of approximating the derivatives in (21.6). The last one (used in the Crank-Nicolson method) is probably the best. In all three methods, the time derivative is approximated by a forward difference

$$\frac{\partial F(x_i,t_j)}{\partial t} \approx \frac{F(x_i,t_{j+1}) - F(x_i,t_j)}{\Delta t}. \tag{21.7}$$

Clearly, this approximation does not work at the upper boundary of the t grid, that is, at $t_M = T$, but this does not matter since the solution for T is given by the boundary conditions type. For instance, for a call option the boundart condition at T is (21.3).

21.2.1 Method 1: The Explicit Finite-difference Method

In the explicit method, $\partial F(x,t)/\partial x$ is approximated by a centered difference

$$\begin{aligned}\frac{\partial F(x_i,t_j)}{\partial x} &\approx \frac{F(x_{i+1},t_{j+1})-F(x_i,t_{j+1})}{2\Delta x}+\frac{F(x_i,t_{j+1})-F(x_{i-1},t_{j+1})}{2\Delta x}\\ &= \frac{F(x_{i+1},t_{j+1})-F(x_{i-1},t_{j+1})}{2\Delta x}\end{aligned} \tag{21.8}$$

Note that the time argument is t_{j+1} everywhere. It would be more natural to use t_j instead, as on the left hand side—but this is the key simplification of the explicit method. More about that later. It is clear that this finite difference does not work at the boundaries of the x grid. For $i = 1$ (lower boundary of x grid), we will never have $F(x_{i-1},t_{j+1})$, and for $i = N$ (upper boundary of x grid) we will never have $F(x_{i+1},t_{j+1})$. We therefore have to supply boundary conditions also in the x dimension. For instance, for the Black-Scholes problem, these are given by (21.5).

The second derivative, $\partial^2 F(x,t)/\partial x^2$ is approximated as a first difference of first differences

$$\begin{aligned}\frac{\partial^2 F(x_i,t_j)}{\partial x^2} &\approx \frac{\left[F(x_{i+1},t_{j+1})-F(x_i,t_{j+1})\right]}{(\Delta x)^2}-\frac{\left[F(x_i,t_{j+1})-F(x_{i-1},t_{j+1})\right]}{(\Delta x)^2}\\ &= \frac{F(x_{i+1},t_{j+1})-2F(x_i,t_{j+1})+F(x_{i-1},t_{j+1})}{(\Delta x)^2}.\end{aligned} \tag{21.9}$$

As before, it is evaluated at t_{j+1}, not at t_j.

Use (21.7), (21.8), and (21.9) in the PDE (21.6) to get (at the point (x_i,t_j))

$$\begin{aligned}0 = &-r(x_i,t_j)F(x_i,t_j)+\frac{F(x_i,t_{j+1})-F(x_i,t_j)}{\Delta t}+\mu(x_i,t_j)\frac{F(x_{i+1},t_{j+1})-F(x_{i-1},t_{j+1})}{2\Delta x}\\ &+\frac{1}{2}\sigma(x_i,t_j)^2\frac{F(x_{i+1},t_{j+1})-2F(x_i,t_{j+1})+F(x_{i-1},t_{j+1})}{(\Delta x)^2}+h(x_i,t_j).\end{aligned} \tag{21.10}$$

This equation allows us to solve explicitly for $F(x_i,t_j)$ in terms of things we already know: the functions r, μ, σ, and h as well as the solution obtained in the previous time

"round", $F(x_{i-1},t_{j+1})$, $F(x_i,t_{j+1})$, and $F(x_{i+1},t_{j+1})$. Note that this means that we start by solving for period $t_M = T$, then solve for period t_{M-1}, and so forth.

Reshuffling (21.10) gives

$$\begin{aligned}&\left[r(x_i,t_j)+\frac{1}{\Delta t}\right]F(x_i,t_j)=\left[\frac{\sigma(x_i,t_j)^2}{2(\Delta x)^2}-\frac{\mu(x_i,t_j)}{2\Delta x}\right]F(x_{i-1},t_{j+1})+\\ &\left[\frac{1}{\Delta t}-\frac{\sigma(x_i,t_j)^2}{(\Delta x)^2}\right]F(x_i,t_{j+1})+\left[\frac{\sigma(x_i,t_j)^2}{2(\Delta x)^2}+\frac{\mu(x_i,t_j)}{2\Delta x}\right]F(x_{i+1},t_{j+1})+h(x_i,t_j).\end{aligned} \tag{21.11}$$

Dividing both sides by $r(x_i,t_j)+1/\Delta t$ shows that we can solve for $F(x_i,t_j)$ in terms of things that we already know.

This can be expressed in matrix form, by stacking $F(x_i,t_j)$, $i = 1,2,...,N$, into the column vector F_j^* (the j index is for the period, t_j) and similarly stack $h(x_i,t_j)$, $i = 1,2,...,N$, into the column vector h_j^*. Then, stacking the equations in (21.11), divided by $r(x_i,t_j)+1/\Delta t$, gives

$$F_j^* = \hat{B}_j F_{j+1}^* + \hat{C}_j h_j^*, \tag{21.12}$$

where $\hat{B}_j$ and $\hat{C}_j$ are two matrices which are straightforward transformations of the functions of r, μ, and σ—see (21.11). Clearly, the first and last rows ($i = 1$ and $i = N$) in $\hat{B}_j$ and $\hat{C}_j$ needs special treatment in order to impose the x boundary conditions. Once these matrices are specified, we can start with a known, F_M^* and then calculate F_{M-1}^*, F_{M-2}^*, ..., F_1^* by repeatedly applying (21.12).

Example 43 *(The Black-Scholes model, call option.) To implement the Black-Scholes model, follow Example 41, and set $h(x_i,t_j) = 0$ for all interior points in the x grid ($1 < i < N$). The grid boundary conditions in (21.5) can be implemented as follows. At the lower bound ($i = 1$), set all elements in the first line of $\hat{B}_j$ to zero, except the first one which should be unity; set the first line in $\hat{C}_j$ to zeros ($h(x_1,t_j)$ can then be anything). At the upper bound ($i = N$), set all elements in the last line of both $\hat{B}_j$ and $\hat{C}_j$ to zero,except the last elements, which should be unity; set $h(x_N,t_j) = x_N - e^{-r(T-t)}K$.*

The problem with this straightforward method is that it is often unstable—in the sense that rounding errors (all computers use finite precision) can accumulate very fast and

completely distort the solution. This is typically the case when the x grid is fine but the t grid is coarse.

21.2.2 Method 2: The Implicit Finite-difference Method

The implicit method approximates the derivatives with respect to x at t_j instead of t_{j+1}. Equations (21.8) and (21.9) are therefore replaced by

$$\frac{\partial F\left(x_i,t_j\right)}{\partial x}\approx\frac{F\left(x_{i+1},t_j\right)-F\left(x_{i-1},t_j\right)}{2\Delta x} \tag{21.13}$$

$$\frac{\partial^2 F\left(x_i,t_j\right)}{\partial x^2}\approx\frac{F\left(x_{i+1},t_j\right)-2F\left(x_i,t_j\right)+F\left(x_{i-1},t_j\right)}{(\Delta x)^2}. \tag{21.14}$$

This small change leads to a much more stable algorithm than the explicit method. However, it is also a bit more complicated, since we now have to solve a system of simultaneous equations at each time step. To see that, use (21.7), (21.13), and (21.14) in the PDE (21.6) to get

$$\begin{aligned}0=&-r\left(x_i,t_j\right)F\left(x_i,t_j\right)+\frac{F\left(x_i,t_{j+1}\right)-F\left(x_i,t_j\right)}{\Delta t}+\mu\left(x_i,t_j\right)\frac{F\left(x_{i+1},t_j\right)-F\left(x_{i-1},t_j\right)}{2\Delta x}\\&+\frac{1}{2}\sigma\left(x_i,t_j\right)^2\frac{F\left(x_{i+1},t_j\right)-2F\left(x_i,t_j\right)+F\left(x_{i-1},t_j\right)}{(\Delta x)^2}+h\left(x_i,t_j\right).\end{aligned} \tag{21.15}$$

This is more complicated than before, since the things we already know (of r, μ, σ, h, and $F\left(x_i,t_{j+1}\right)$) only determine a relation between three elements in the F_j^* vector: $F\left(x_{i-1},t_j\right)$, $F\left(x_i,t_j\right)$, and $F\left(x_{i+1},t_j\right)$. This is in contrast to the explicit method (21.10), where the things we already know explicitly determines $F\left(x_i,t_j\right)$ alone. To see this, reshuffle (21.15) as

$$\begin{aligned}&\left[\frac{\sigma\left(x_i,t_j\right)^2}{2(\Delta x)^2}-\frac{\mu\left(x_i,t_j\right)}{2\Delta x}\right]F\left(x_{i-1},t_j\right)+\left[-r\left(x_i,t_j\right)-\frac{1}{\Delta t}-\frac{\sigma\left(x_i,t_j\right)^2}{(\Delta x)^2}\right]F\left(x_i,t_j\right)+\\&\left[\frac{\sigma\left(x_i,t_j\right)^2}{2(\Delta x)^2}+\frac{\mu\left(x_i,t_j\right)}{2\Delta x}\right]F\left(x_{i+1},t_j\right)=-\frac{F\left(x_i,t_{j+1}\right)}{\Delta t}-h\left(x_i,t_j\right).\end{aligned} \tag{21.16}$$

We therefore have to change (21.12) to

$$\tilde{A}_j F_j^*=\tilde{B}_j F_{j+1}^*+\tilde{C}_j h_j^*, \tag{21.17}$$

where, once again, $\tilde{B}_j$ and $\tilde{C}_j$ are very simple, and where $\tilde{A}_j$ summarizes how the elements in F_j^* are jointly determined—see (21.16). This shows that we can once again calculate F_j^* from h_j^* and F_{j+1}^*, but that we have to solve a system of simultaneous linear equations at each time step (since $\tilde{A}_j$ is not a diagonal matrix).

The solution of the simultaneous equations is fairly simple, however. With few elements in the x grid (small N) you can use any standard method for solving linear systems of equations, including your favorite matrix inversion technique. However, with a many elements it is useful to exploit the special structure of the $\tilde{A}_j$ matrix. Note from (21.15) that the ith line in $\tilde{A}_j$ has non-zero elements only in columns $i-1$, i, and $i+1$. The matrix is tridiagonal (non-zero elements only on the diagonal, super-diagonal, and sub-diagonal). Using a simple computer code, this feature allows us to solve the system of equations very fast, even if there are many elements in x grid.

The implicit method is very stable, but can be a bit slow.

21.2.3 Method 3: The Crank-Nicolson Method

We can speed up the implicit method, but keep the stability, by approximating the derivatives as an average of the explicit and implicit methods. That is, $\partial F\left(x_i,t_j\right)/\partial x$ is taken to be the average of (21.8) and (21.13), and $\partial^2 F\left(x_i,t_j\right)/\partial x^2$ is taken to be the average of (21.9) and (21.14). Of course, this will give us the same structure of the problem as in the implicit method (21.17), but with other coefficients.

Denoting the right hand side d_{ij} (as does Duffie), we can write the expression as

$$a_{ij}F\left(x_{i-1},t_j\right)+b_{ij}F\left(x_i,t_j\right)+c_{ij}F\left(x_{i+1},t_j\right)=d_{ij},\ \text{for } 1<i<N \tag{21.18}$$

where the coefficients a_{ij}, b_{ij}, and c_{ij} are similar (but not identical) to the expression in brackets in (21.16); d_{ij} is a bit messier, but can easily be derived by combining (21.11) and (21.16). The precise expressions are given in Duffie (1992), appendix 10.H. This equation holds for all interior points in the x grid, but not for the first or the last points ($i=1$ and $i=N$), which we will return to shortly.

It is easier to see the pattern in (21.18) by writing it out for three consecutive values of i. Suppose $N=5$, then the interior points in the x grid are $i=2$, 3, and 4. For these,

(21.18) can be written

$$\begin{bmatrix} a_{2j}F\left(x_1,t_j\right)+b_{2j}F\left(x_2,t_j\right)+c_{2j}F\left(x_3,t_j\right) \\ a_{3j}F\left(x_2,t_j\right)+b_{3j}F\left(x_3,t_j\right)+c_{3j}F\left(x_4,t_j\right) \\ a_{4j}F\left(x_3,t_j\right)+b_{4j}F\left(x_4,t_j\right)+c_{4j}F\left(x_5,t_j\right) \end{bmatrix} = \begin{bmatrix} d_{2j} \\ d_{3j} \\ d_{4j} \end{bmatrix}, \text{ or}$$

$$\begin{bmatrix} a_{2j} & b_{2j} & c_{2j} & 0 & 0 \\ 0 & a_{3j} & b_{3j} & c_{3j} & 0 \\ 0 & 0 & a_{4j} & b_{4j} & c_{4j} \end{bmatrix} \begin{bmatrix} F\left(x_1,t_j\right) \\ F\left(x_2,t_j\right) \\ F\left(x_3,t_j\right) \\ F\left(x_4,t_j\right) \\ F\left(x_5,t_j\right) \end{bmatrix} = \begin{bmatrix} d_{2j} \\ d_{3j} \\ d_{4j} \end{bmatrix}. \tag{21.19}$$

We now deal with the (lower and upper) boundaries of the x grid. For the first point in the x grid, x_1, we set $a_{1j}=0$ (there is no value of $F\left(x_0,t_j\right)$) so the equation is like

$$\begin{bmatrix} b_{1j} & c_{1j} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} F\left(x_1,t_j\right) \\ F\left(x_2,t_j\right) \\ F\left(x_3,t_j\right) \\ F\left(x_4,t_j\right) \\ F\left(x_5,t_j\right) \end{bmatrix} = d_{1j}. \tag{21.20}$$

To impose the boundary condition at x_1 we can pick values of b_{1j}, c_{1j}, and d_{1j}.

Example 44 *In the Black-Scholes model, the $x_1=0$ grid boundary conditions is that a call option has a zero price, which we can impose by $b_{1j}=0$, $c_{1j}=1$, and $d_{1j}=0$.*

For the upper x grid boundary we set $c_{Nj}=0$ (there is no value of $F\left(x_{N+1},t_j\right)$), so the equation is like

$$\begin{bmatrix} 0 & 0 & 0 & a_{Nj} & b_{Nj} \end{bmatrix} \begin{bmatrix} F\left(x_1,t_j\right) \\ F\left(x_2,t_j\right) \\ F\left(x_3,t_j\right) \\ F\left(x_4,t_j\right) \\ F\left(x_5,t_j\right) \end{bmatrix} = d_{5j}. \tag{21.21}$$

The coefficients a_{Nj}, b_{Nj}, and d_{Nj} are picked to satisfy the boundary conditions.

Example 45 *(The Black-Scholes model, call option.) In the Black-Scholes model, the $x_N\approx\infty$ (in practice a high finite number) boundary conditions is that a call has the price $x_N-e^{-r(T-t)}K$, which we can impose by $a_{Nj}=0$, $b_{Nj}=1$, and $d_{Nj}=x_N-e^{-r(T-t)}K$.*

We can then combine (21.19)–(21.21) to get the complete system (for period t_j)

$$\begin{bmatrix} b_{1j} & c_{1j} & 0 & 0 & 0 \\ a_{2j} & b_{2j} & c_{2j} & 0 & 0 \\ 0 & a_{3j} & b_{3j} & c_{3j} & 0 \\ 0 & 0 & a_{4j} & b_{4j} & c_{4j} \\ 0 & 0 & 0 & a_{Nj} & b_{Nj} \end{bmatrix} \begin{bmatrix} F\left(x_1,t_j\right) \\ F\left(x_2,t_j\right) \\ F\left(x_3,t_j\right) \\ F\left(x_4,t_j\right) \\ F\left(x_5,t_j\right) \end{bmatrix} = \begin{bmatrix} d_{1j} \\ d_{2j} \\ d_{3j} \\ d_{4j} \\ d_{5j} \end{bmatrix}, \tag{21.22}$$

which is a tridiagonal system of equations, which is easy to solve. We solve this for the $F\left(x_i,t_j\right)$ vector, which is then used for solving for the $F\left(x_i,t_{j-1}\right)$ vector and so forth until we get the $F\left(x_i,t_1\right)$ vector. Of course, each of these rounds of solving require setting up new coefficients a_{ij}, b_{ij}, c_{ij}, and d_{ij}.

21.3 Early Exercise

We know that it is never optimal to exercise an American call option early if there are no dividends, but it can be optimal to exercise an American put option early, in which case the payoff is $K-x\left(t\right)$

A simple way of calculating the American put price is to amend the method described above by taking F_j^* to be not the solution of (21.17), but the maximum of the solution of (21.17) and $K-x\left(t_j\right)$. This value is then used in the next round, that is, for t_{j-1}.

In the Black-Scholes model we know that the European put option has the price

$$P\left(x,t\right)=e^{-r(T-t)}K\Phi\left(-d_1+\sigma\sqrt{T-t}\right)-x\left(t\right)\Phi\left(-d_1\right), \tag{21.23}$$

where d_1 is given in (21.4). The boundary conditions for the put option price are

$$P\left(x,T\right)=\max\left[K-x\left(T\right),0\right],\ P\left(0,t\right)=e^{-r(T-t)}K, \text{ and } \lim_{x\to\infty}P\left(x,t\right)=0. \tag{21.24}$$

Example 46 *(The Black-Scholes model, put option.) In the Black-Scholes model, the x grid boundary conditions means $(a_{1j},b_{1j},c_{1j},d_{1j})=(0,1,0,e^{-r(T-t)}K)$ and $(a_{Nj},b_{Nj},c_{Nj},d_{Nj})=$*

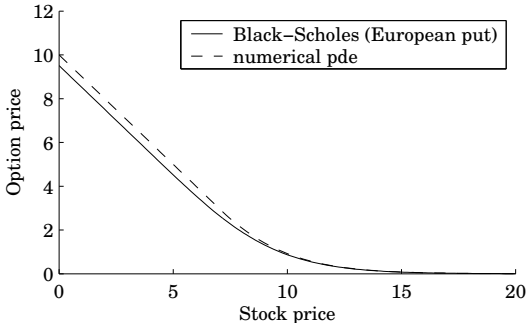


Figure 21.1: Prices on American put options

$(0, 1, 0, 0)$. *Figure 21.1 shows an example with* $r = 0.1$, $\sigma = 0.4$, *and* $K = 10$ *(taken from Wilmott, Howison, and Dewynne (1995), p 176).*

Bibliography

Duffie, D., 1992, *Dynamic Asset Pricing Theory*, Princeton University Press, Princeton.

Hull, J. C., 2000, *Options, Futures, and Other Derivatives*, Prentice-Hall, Upper Saddle River, NJ.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992, *Numerical Recipes in Fortran*, Cambrdige University Press, Cambrdige, 2nd edn.

Wilmott, P., S. Howison, and J. Dewynne, 1995, *The Mathematics of Financial Derivatives*, Cambrdige University Press, Cambridge.

Reading List

Main references: Cochrane (2001) (C) and Campbell, Lo, and MacKinlay (1997) (CLM).

For an introduction to many of the issues, see Cuthbertson (1996), and for a thorough econometrics text on asset pricing, see Gourieroux and Jasiak (2001).

Papers marked with (*) are required reading. More references are given in my lecture notes.

Background

1. *Campbell (2000)

GMM

1. *C 10-11
2. *Lecture notes

Keywords: moment conditions, loss function, asymptotic distribution, J and D tests, Newey-West covariance estimator, choice of weighting matrix

Predictability of Asset Returns

1. *C 20.1
2. *CLM 2 and 7
3. *Lecture notes
4. Cuthbertson (1996) 5 and 6
5. Keywords: autocorrelations, Box-Pierce test, variance ratio, long-run autoregressions, lead-lags, dividend-price ratio as predictor

Linear Factor Models

1. *CLM 5-6
2. *C 12-16, and 20.2-3
3. C 6 and 8 (theory on linear factor models versus SDF models)
4. Gibbons, Ross, and Shanken (1989) (interpretation of CAPM test statistic)
5. Fama and French (1993) and Fama and French (1996) (size, book-to-market)
6. Ferson (1995) (overview)
7. Jagannathan and Wang (2002) (SDF vs linear factor model)

Keywords: CAPM, Jensen's alpha, ML/LS approach, non-iid returns, GMM, size and power, multifactor model, Fama-MacBeth, discount factor model, mean-variance intersection

Consumption-Based Asset Pricing

1. *Lecture notes
2. *CLM 8
3. *C 21
4. Campbell (2001)

Keywords: equity premium puzzle, risk-free rate puzzle, habit persistance, recursive utility, idiosyncratic shocks

Models of Changing Volatility

1. *CLM 12.2
2. *Bansal and Lundblad (2002)

3. *Heston and Nandi (2000)

Keywords: ARCH, GARCH, link between first and second moments, CAPM-GARCH, option-GARCH

Interest Rate Models

1. *CLM 11
2. *Chan, Karolyi, Longstaff, and Sanders (1992)

Keywords: SDF model of yield curve, Vasicek, CIR, non-linear GMM estimation

Testing Distributions and Kernel Regressions

1. *CLM 12.3
2. *Ait-Sahalia (1996) (non-parametric estimation of short rate models)
3. *Ait-Sahalia and Lo (1998) (non-parametric estimation of option price function)
4. Lo, Mamaysky, and Wang (2000) (a modern look at technical analysis)

Keywords: kernel regression, kernel density estimation, Kolmogorov-Smirnov, option pricing

Bibliography

Ait-Sahalia, Y., 1996, "Testing Continuous-Time Models of the Spot Interest Rate," *Review of Financial Studies*, 9, 385–426.

Ait-Sahalia, Y., and A. W. Lo, 1998, "Nonparametric Estimation of Stat-Price Densities Implicit in Financial Asset Prices," *Journal of Finance*, 53, 499–547.

Bansal, R., and C. Lundblad, 2002, "Market Efficiency, Fundamental Values, and the Size of the Risk Premium in Global Equity Markets," *Journal of Econometrics*, 109, 195–237.

Campbell, J. Y., 2000, "Asset Pricing at the Millenium," *Journal of Finance*, 55, 1515–1567.

Campbell, J. Y., 2001, "Consumption-Based Asset Pricing," in George Constantinides, Milton Harris, and Rene Stultz (ed.), *Handbook of the Economics of Finance*, North-Holland, Amsterdam.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay, 1997, *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey.

Chan, K. C., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders, 1992, "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," *Journal of Finance*, 47, 1209–1227.

Cochrane, J. H., 2001, *Asset Pricing*, Princeton University Press, Princeton, New Jersey.

Cuthbertson, K., 1996, *Quantitative Financial Economics*, Wiley, Chichester, England.

Fama, E. F., and K. R. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.

Fama, E. F., and K. R. French, 1996, "Multifactor Explanations of Asset Pricing Anomalies," *Journal of Finance*, 51, 55–84.

Ferson, W. E., 1995, "Theory and Empirical Testing of Asset Pricing Models," in Robert A. Jarrow, Vojislav Maksimovic, and William T. Ziemba (ed.), *Handbooks in Operations Research and Management Science* . pp. 145–200, North-Holland, Amsterdam.

Gibbons, M., S. Ross, and J. Shanken, 1989, "A Test of the Efficiency of a Given Portfolio," *Econometrica*, 57, 1121–1152.

Gourieroux, C., and J. Jasiak, 2001, *Financial Econometrics: Problems, Models, and Methods*, Princeton University Press.

Heston, S. L., and S. Nandi, 2000, "A Closed-Form GARCH Option Valuation Model," *Review of Financial Studies*, 13, 585–625.

Jagannathan, R., and Z. Wang, 2002, "Empirical Evaluation of Asset Pricing Models: A Comparison of the SDF and Beta Methods," *Journal of Finance*, 57, 2337–2367.

Lo, A. W., H. Mamaysky, and J. Wang, 2000, "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation," *Journal of Finance*, 55, 1705–1765.