Lecture Notes in Empirical Finance (PhD): Linear Factor Models

Paul Söderlind[1]

8 June 2006

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

Contents

Chapter 3

Linear Factor Models

3.1 CAPM Tests: Overview

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

Let $R^e_{it} = R_{it} - R_{ft}$ 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

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

The economic importance of a non-zero intercept (α) is that the tangency portfolio changes if the test asset is added to the investment opportunity set. See Figure 3.1 for an illustration.

The basic test of CAPM is to estimate (3.1) on a single asset and then test if the intercept is zero. This can easily be extended to several assets, where we test if all the intercepts are zero.

Notice that the 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 (this is the approach in mutual fund evaluations). 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).

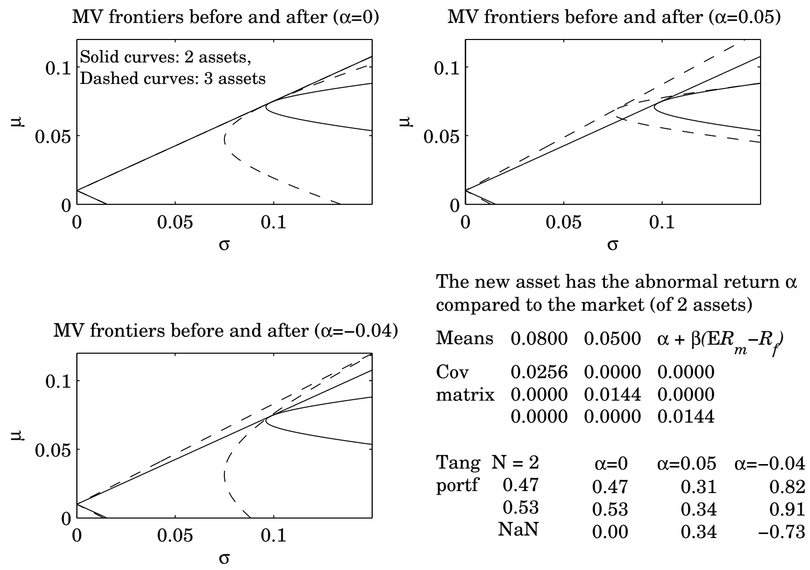


Figure 3.1: MV frontiers with 2 and 3 assets

3.2 Testing CAPM: Traditional LS Approach

3.2.1 CAPM with One Asset: Traditional LS Approach

If the residuals in the CAPM regression are iid, then the traditional LS approach is just fine: estimate (3.1) and form a t-test of the null hypothesis that the intercept is zero. If the disturbance is iid normally distributed, then this approach is the ML approach.

To understand the properties of the LS approch, we use the results in the following remark.

Remark 1 *(Covariance matrix of LS estimator) 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*

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

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

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

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

(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

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

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

where SR_f^2 is the squared 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 gives a large uncertainty about what the risk-adjusted return should be.

The t-test of the hypothesis that $\alpha_0 = 0$ is then

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

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).

Remark 2 *(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

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

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).

It is quite straightforward to use the properties of minimum-variance frontiers (see Gibbons, Ross, and Shanken (1989), and MacKinlay (1995)) to show that the test statistic in (3.5) can be written

$$\frac{\hat{\alpha}_i^2}{\mathrm{Var}(\hat{\alpha}_i)} = \frac{(\widehat{SR}_c)^2 - (\widehat{SR}_f)^2}{[1 + (\widehat{SR}_f)^2]/T}, \tag{3.6}$$

where SR_f is the Sharpe ratio of the market portfolio and SR_c is the Sharpe ratio of the tangency portfolio when investment in both the market return and asset i is possible. (Recall that the tangency portfolio is the portfolio with the highest possible Sharpe ratio.) 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, $\hat{\alpha}_i^2 / \mathrm{Var}(\hat{\alpha}_i)$, is zero—and CAPM is not rejected.

Proof. (of (3.6)) From the CAPM regression (3.1) we have

$$\mathrm{Cov} \begin{bmatrix} R_{it}^e \\ R_{mt}^e \end{bmatrix} = \begin{bmatrix} \beta_i^2 \sigma_m^2 + \mathrm{Var}(\varepsilon_{it}) & \beta_i \sigma_m^2 \\ \beta_i \sigma_m^2 & \sigma_m^2 \end{bmatrix}, \text{ and } \begin{bmatrix} \mu_i^e \\ \mu_m^e \end{bmatrix} = \begin{bmatrix} \alpha_i + \beta_i \mu_m^e \\ \mu_m^e \end{bmatrix}.$$

Suppose we use this information to construct a mean-variance frontier for both R_{it} and R_{mt}, and we find the tangency portfolio, with excess return R_{ct}^e. It is straightforward to show that the square of the Sharpe ratio of the tangency portfolio is $\mu^{e\prime} \Sigma^{-1} \mu^e$, where μ^e is the vector of expected excess returns and Σ is the covariance matrix. By using the covariance matrix and mean vector above, we get that the squared Sharpe ratio for the tangency portfolio, $\mu^{e\prime}\Sigma^{-1}\mu^{e}$, (using both R_{it} and R_{mt}) is

$$\left(\frac{\mu_c^e}{\sigma_c}\right)^2 = \frac{\alpha_i^2}{\text{Var}(\varepsilon_{it})} + \left(\frac{\mu_m^e}{\sigma_m}\right)^2,$$

which we can write as

$$(SR_c)^2 = \frac{\alpha_i^2}{\text{Var}(\varepsilon_{it})} + (SR_m)^2.$$

Use the notation $f_t = R_{mt} - R_{ft}$ and combine this with (3.3) and to get (3.6). ∎

It is also possible to construct small sample test (that do 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 $Z \sim t_n$, then $Z^2 \sim F(1,n)$).

An alternative testing approach is to use an LR or LM approach: restrict the intercept in the CAPM regression to be zero and estimate the model with ML (assuming that the errors are normally distributed). For instanc, for an LR test, the likelihood value (when $\alpha = 0$) is then compared to the likelihood value without restrictions.

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.

3.2.2 CAPM with Several Assets: Traditional LS Approach

Suppose we have n test assets. Stack (3.1) expressions (3.1) for $i = 1,\ldots,n$ as Let $f_t = R_{mt} - R_{ft}$ and stack the expressions (3.1) for $i = 1,\ldots,n$ as

$$\begin{bmatrix} R_{1t}^e \\ \vdots \\ R_{nt}^e \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} f_t + \begin{bmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{nt} \end{bmatrix}, \text{ E}\,\varepsilon_{it} = 0 \text{ and } \text{Cov}(f_t, \varepsilon_{it}) = 0. \quad (3.7)$$

This is a system of seemingly unrelated regressions (SUR)—with the same regressor (see, for instance, Greene (2003) 14). In this case, the efficient estimator (GLS) is LS on each equation separately. Moreover, the covariance matrix of the coefficients is particularly simple.

To see what the covariances of the coefficients are, write each of the the regression equations in (3.7) on a traditional form

$$R_{it}^e = x_t'\theta_i + \varepsilon_{1t}, \text{ where } x_t = \begin{bmatrix} 1 \\ f_t \end{bmatrix}. \quad (3.8)$$

If we define

$$\Sigma_{xx} = \text{plim} \sum\nolimits_{t=1}^{T} x_t x_t'/T, \text{ and } \sigma_{ij} = \text{plim} \sum\nolimits_{t=1}^{T} \varepsilon_{it}\varepsilon_{jt}/T, \quad (3.9)$$

then the asymptotic covariance matrix of the vectors $\hat{\theta}_i$ and $\hat{\theta}_j$ (assets i and j) is $\sigma_{ij}\Sigma_{xx}^{-1}/T$, that is,

$$\text{ACov}(\sqrt{T}\hat{\theta}) = \begin{bmatrix} \sigma_{11} & \ldots & \sigma_{1n} \\ \vdots & & \vdots \\ \sigma_{n1} & \ldots & \hat{\sigma}_{nn} \end{bmatrix} \otimes \Sigma_{xx}^{-1}. \quad (3.10)$$

In a large sample, the estimator is normally distributed. Therefore, under the null hypothesis (that ther intecepts are zero)) we have the following result. From Remark 1 we know that the upper left element of Σ_{xx}^{-1} equals $1 + SR^2$, so

$$\sqrt{T}\hat{\alpha} \to^d N\left[\mathbf{0}_{n\times 1}, \Sigma(1+SR^2)\right], \text{ where } \Sigma = \begin{bmatrix} \sigma_{11} & \ldots & \sigma_{1n} \\ \vdots & & \vdots \\ \sigma_{n1} & \ldots & \hat{\sigma}_{nn} \end{bmatrix}. \quad (3.11)$$

In practice we use the sample moments for the covariance matrix.

To test the null hypothesis that all intercepts are zero, we use the test statistic

$$T\hat{\alpha}'(1+SR^2)\Sigma^{-1}\hat{\alpha} \sim \chi^2_n. \tag{3.12}$$

As for the case of a single asset, it is straightforward to do an LR or LM test instead. Assuming the errors are normally distributed, a restricted model ($\alpha = \mathbf{0}$) is estimated by ML (LS actually), and the properties of the likelihood function are used for testing.

3.3 Testing CAPM: GMM

3.3.1 CAPM with Several Assets: GMM and a Wald Test

To test n assets at the same time when the errors are non-iid we make use of the GMM framework. A special case is when the residuals are iid. The results in this section will then coincide with those in Section 3.2.

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

$$\begin{bmatrix} R^e_{1t} \\ \vdots \\ R^e_{nt} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} f_t + \begin{bmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{nt} \end{bmatrix}, \text{ E}\,\varepsilon_{it} = 0 \text{ and } \text{Cov}(f_t, \varepsilon_{it}) = 0, \tag{3.13}$$

or more compactly

$$R^e_t = \alpha + \beta f_t + \varepsilon_t, \text{ E}\,\varepsilon_t = \mathbf{0}_{n\times 1} \text{ and } \text{Cov}(f_t, \varepsilon_t) = \mathbf{0}_{1\times n}, \tag{3.14}$$

where α and β are $n \times 1$ vectors. Clearly, setting $n = 1$ gives the case of a single test asset.

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

$$\text{E}\, g_t(\alpha, \beta) = \mathbf{0}_{2n\times 1}, \text{ where} \tag{3.15}$$

$$g_t(\alpha, \beta) = \begin{bmatrix} \varepsilon_t \\ f_t\varepsilon_t \end{bmatrix} = \begin{bmatrix} R^e_t - \alpha - \beta f_t \\ f_t\left(R^e_t - \alpha - \beta f_t\right) \end{bmatrix}. \tag{3.16}$$

There are as many parameters as moment conditions, so the GMM estimator picks values of α and β such that the sample analogues of (3.15) are satisfied exactly

$$\bar{g}(\hat{\alpha}, \hat{\beta}) = \frac{1}{T}\sum_{t=1}^{T} g_t(\hat{\alpha}, \hat{\beta}) = \frac{1}{T}\sum_{t=1}^{T} \begin{bmatrix} R^e_t - \hat{\alpha} - \hat{\beta} f_t \\ f_t(R^e_t - \hat{\alpha} - \hat{\beta} f_t) \end{bmatrix} = \mathbf{0}_{2n\times 1}, \tag{3.17}$$

which gives the LS estimator. For the inference, we allow for the possibility of non-iid errors. (With iid errors we get the same results as in Section 3.2, at least asymptotically.)

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

$$\hat{\alpha}' \text{ Var}(\hat{\alpha})^{-1}\hat{\alpha} \overset{d}{\to} \chi^2_n. \tag{3.18}$$

Remark 3 *(Easy coding of the GMM Problem (3.17)) Estimate by LS, equation by equation. Then, plug in the fitted residuals in (3.16) to generate time series of the moments (will be important for the tests).*

Remark 4 *(Distribution of GMM) Let the parameter vector in the moment condition have the true value b_0. Define*

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

When the estimator solves $\min \bar{g}\,(b)'\,S_0^{-1}\,\bar{g}\,(b)$ *or when the model is exactly identified, the distribution of the GMM estimator is*

$$\sqrt{T}(\hat{b} - b_0) \overset{d}{\to} N\left(\mathbf{0}_{k\times 1}, V\right), \textit{ where } V = \left(D_0' S_0^{-1} D_0\right)^{-1} = D_0^{-1} S_0 (D_0^{-1})'.$$

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 (3.15) as

$$\begin{bmatrix} \bar{g}_1(\alpha, \beta) \\ \bar{g}_2(\alpha, \beta) \\ \bar{g}_3(\alpha, \beta) \\ \bar{g}_4(\alpha, \beta) \end{bmatrix} = \frac{1}{T}\sum\nolimits_{t=1}^{T} \begin{bmatrix} R^e_{1t} - \alpha_1 - \beta_1 f_t \\ R^e_{2t} - \alpha_2 - \beta_2 f_t \\ f_t(R^e_{1t} - \alpha_1 - \beta_1 f_t) \\ f_t(R^e_{2t} - \alpha_2 - \beta_2 f_t) \end{bmatrix} = \mathbf{0}_{4\times 1}, \tag{3.19}$$

where $\bar{g}_1(\alpha, \beta)$ denotes the sample average of the first moment condition.

The Jacobian is

$$\frac{\partial \bar{g}(\alpha,\beta)}{\partial[\alpha_1,\alpha_2,\beta_1,\beta_2]'} = \begin{bmatrix} \partial\bar{g}_1/\partial\alpha_1 & \partial\bar{g}_1/\partial\alpha_2 & \partial\bar{g}_1/\partial\beta_1 & \partial\bar{g}_1/\partial\beta_2 \\ \partial\bar{g}_2/\partial\alpha_1 & \partial\bar{g}_2/\partial\alpha_2 & \partial\bar{g}_2/\partial\beta_1 & \partial\bar{g}_2/\partial\beta_2 \\ \partial\bar{g}_3/\partial\alpha_1 & \partial\bar{g}_3/\partial\alpha_2 & \partial\bar{g}_3/\partial\beta_1 & \partial\bar{g}_3/\partial\beta_2 \\ \partial\bar{g}_4/\partial\alpha_1 & \partial\bar{g}_4/\partial\alpha_2 & \partial\bar{g}_4/\partial\beta_1 & \partial\bar{g}_4/\partial\beta_2 \end{bmatrix}$$

$$= -\frac{1}{T}\sum\nolimits_{t=1}^{T} \begin{bmatrix} 1 & 0 & f_t & 0 \\ 0 & 1 & 0 & f_t \\ f_t & 0 & f_t^2 & 0 \\ 0 & f_t & 0 & f_t^2 \end{bmatrix}. \tag{3.20}$$

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 (3.20) is simply the expected value, which can written as

$$D_0 = -\operatorname{E}\begin{bmatrix} 1 & f_t \\ f_t & f_t^2 \end{bmatrix} \otimes I_2 = -\operatorname{E}\left(\begin{bmatrix} 1 \\ f_t \end{bmatrix}\begin{bmatrix} 1 \\ f_t \end{bmatrix}'\right) \otimes I_2, \tag{3.21}$$

where $\otimes$ is the Kronecker product. For n assets, change I_2 to I_n. (The last expression applies also to the case of several factors.)

Remark 5 *(Kronecker product) If A and B are matrices, then*

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

From Remark 4, we can write the covariance matrix of the $2n \times 1$ vector of parameters (n parameters in α and another n in β) as

$$\operatorname{ACov}\left(\sqrt{T}\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}\right) = D_0^{-1} S_0 (D_0^{-1})' \tag{3.22}$$

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

$$S_0 = \operatorname{ACov}\left(\frac{\sqrt{T}}{T}\sum_{t=1}^{T} g_t(\alpha,\beta)\right) = \sum_{s=-\infty}^{\infty} R(s), \text{ where} \tag{3.23}$$

$$R(s) = \operatorname{E} g_t(\alpha,\beta) g_{t-s}(\alpha,\beta)'. \tag{3.24}$$

With n assets, we can write (3.24) in terms of the $n \times 1$ vector ε_t as

$$\begin{aligned} R(s) &= \operatorname{E} g_t(\alpha,\beta) g_{t-s}(\alpha,\beta)' \\ &= \operatorname{E}\begin{bmatrix} \varepsilon_t \\ f_t\varepsilon_t \end{bmatrix}\begin{bmatrix} \varepsilon_{t-s} \\ f_{t-s}\varepsilon_{t-s} \end{bmatrix}' = \operatorname{E}\left[\left(\begin{bmatrix} 1 \\ f_t \end{bmatrix} \otimes \varepsilon_t\right)\left(\begin{bmatrix} 1 \\ f_{t-s} \end{bmatrix} \otimes \varepsilon_{t-s}\right)'\right]. \end{aligned} \tag{3.25}$$

(The last expression applies also to the case of several factors.)

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

Example 6 *(Special case 1: f_t is independent of ε_{t-s}, errors are iid, and $n = 1$) With these assumptions $R(s) = \mathbf{0}_{2\times 2}$ if $s \neq 0$, and $S_0 = \begin{bmatrix} 1 & \operatorname{E} f_t \\ \operatorname{E} f_t & \operatorname{E} f_t^2 \end{bmatrix} \operatorname{Var}(\varepsilon_{it})$. Combining with (3.21) gives*

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

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

Example 7 *(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 $S_0 = \begin{bmatrix} 1 & \operatorname{E} f_t \\ \operatorname{E} f_t & \operatorname{E} f_t^2 \end{bmatrix} \otimes \operatorname{E}\varepsilon_t\varepsilon_t'$. Combining with (3.21) gives*

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

This follows from the facts that $(A\otimes B)^{-1} = A^{-1}\otimes B^{-1}$ and $(A\otimes B)(C\otimes D) = AC\otimes BD$ (if conformable).

3.3.2 CAPM and Several Assets: GMM and an LM Test

We could also construct an "LM test" instead by imposing $\alpha = \mathbf{0}$ in the moment conditions (3.15) and (3.17). The moment conditions are then

$$\mathrm{E}\, g(\beta) = \mathrm{E}\begin{bmatrix} R_t^e - \beta f_t \\ f_t(R_t^e - \beta f_t) \end{bmatrix} = \mathbf{0}_{2n\times 1}. \tag{3.26}$$

Since there are $q = 2n$ moment conditions, but only n parameters (the β vector), this model is overidentified.

We could either use a weighting matrix in the GMM loss function or combine the moment conditions so the model becomes exactly identified.

With a weighting matrix, the estimator solves

$$\min_b \bar{g}(b)' W \bar{g}(b), \tag{3.27}$$

where $\bar{g}(b)$ is the sample average of the moments (evaluated at some parameter vector b), and W is a positive definite (and symmetric) weighting matrix. 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 restriction (null hypothesis) that $\alpha = \mathbf{0}_{n\times 1}$ is rejected.

To combine the moment conditions so the model becomes exactly identified, premultiply by a matrix A to get

$$A_{n\times 2n}\, \mathrm{E}\, g(\beta) = \mathbf{0}_{n\times 1}. \tag{3.28}$$

The model is then tested by testing if all $2n$ moment conditions in (3.26) are satisfied at this vector of estimates of the betas. This is the GMM analogue to a classical LM test.

For instance, to effectively use only the last n moment conditions in the estimation, we specify

$$A\, \mathrm{E}\, g(\beta) = \begin{bmatrix} 0_{n\times n} & I_n \end{bmatrix} \mathrm{E}\begin{bmatrix} R_t^e - \beta f_t \\ f_t(R_t^e - \beta f_t) \end{bmatrix} = \mathbf{0}_{n\times 1}. \tag{3.29}$$

This clearly gives the classical LS estimator without an intercept

$$\hat{\beta} = \frac{\sum_{t=1}^T f_t R_t^e / T}{\sum_{t=1}^T f_t^2 / T}. \tag{3.30}$$

Example 8 *(Combining moment conditions, CAPM on two assets) With two assets we can combine the four moment conditions into only two by*

$$A\, \mathrm{E}\, g_t(\beta_1, \beta_2) = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathrm{E}\begin{bmatrix} R_{1t}^e - \beta_1 f_t \\ R_{2t}^e - \beta_2 f_t \\ f_t(R_{1t}^e - \beta_1 f_t) \\ f_t(R_{2t}^e - \beta_2 f_t) \end{bmatrix} = \mathbf{0}_{2\times 1}.$$

Remark 9 *(Test of overidentifying assumption in GMM) When the GMM estimator solves the quadratic loss function $\bar{g}(\beta)' S_0^{-1} \bar{g}(\beta)$ (or is exactly identified), then the J test statistic is*

$$T \bar{g}(\hat{\beta})' S_0^{-1} \bar{g}(\hat{\beta}) \overset{d}{\to} \chi^2_{q-k},$$

where q is the number of moment conditions and k is the number of parameters.

Remark 10 *(Distribution of GMM, more general results) When GMM solves $\min_b \bar{g}(b)' W \bar{g}(b)$ or $A\bar{g}(\hat{\beta}) = \mathbf{0}_{k\times 1}$, the distribution of the GMM estimator and the test of overidentifying assumptions are different than in Remarks 4 and 9.*

3.3.3 Size and Power of the CAPM 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).

3.3.4 Choice of Portfolios

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.

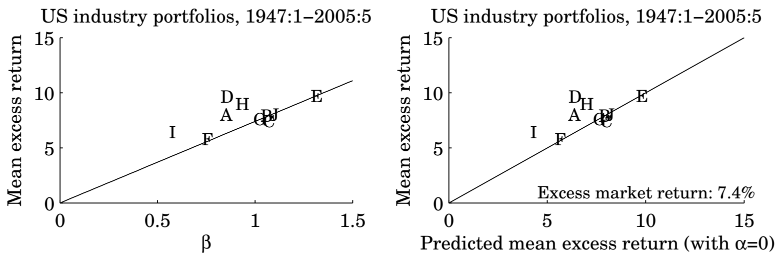
It sometimes makes economic sense to sort the assets according to a characteristic (size or perhaps book/market)—and then test if the model is true for these portfolios. Rejection of the CAPM for such portfolios may have an interest in itself.

3.3.5 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 $\mathrm{E}\, R^e_{it} = \beta_i \mathrm{E}\, R^e_{mt}$). 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).



	alpha	pval	StdErr
all	NaN	0.02	NaN
A (NoDur)	1.99	0.11	8.11
B (Durbl)	0.36	0.82	11.82
C (Manuf)	−0.25	0.77	6.08
D (Enrgy)	3.58	0.04	12.88
E (HiTec)	0.21	0.90	12.03
F (Telcm)	0.45	0.73	9.78
G (Shops)	0.32	0.80	8.59
H (Hlth)	2.33	0.11	10.95
I (Utils)	2.33	0.10	10.35
J (Other)	0.10	0.90	6.04

CAPM
Factor: US market
test statistics use Newey–West std
test statistics ~ $\chi^2(n)$, n = no. of test assets
α and StdErr are in annualized %

Figure 3.2: CAPM, US industry portfolios

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

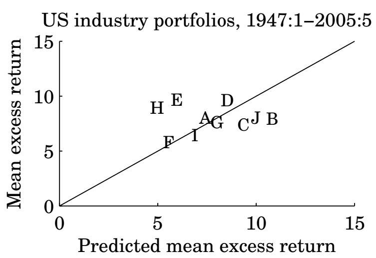
Reference: Cochrane (2001) 12.1; Campbell, Lo, and MacKinlay (1997) 6.2.1

3.4.1 A Multi-Factor Model

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

$$R^e_{it} = \alpha_i + \beta_i' f_t + \varepsilon_{it}, \text{ where } \mathrm{E}\,\varepsilon_{it} = 0 \text{ and } \mathrm{Cov}(f_t, \varepsilon_{it}) = \mathbf{0}_{K\times 1}. \tag{3.31}$$

and β_i is now an $K \times 1$ vector. The CAPM regression is a special case when the market excess return is the only factor. In other models like ICAPM (see Cochrane (2001) 9.2),



	alpha	pval	StdErr
all	NaN	0.00	NaN
A (NoDur)	0.95	0.44	7.98
B (Durbl)	−2.54	0.10	11.12
C (Manuf)	−1.61	0.04	5.79
D (Enrgy)	1.46	0.38	12.01
E (HiTec)	4.06	0.00	10.30
F (Telcm)	0.57	0.68	9.63
G (Shops)	−0.03	0.98	8.49
H (Hlth)	4.38	0.00	10.40
I (Utils)	−0.25	0.85	9.46
J (Other)	−1.69	0.03	5.49

Fama−French model
Factors: US market, SMB (size), and HML (book−to−market)
test statistics use Newey−West std
test statistics ~ $\chi^2(n)$, n = no. of test assets
α and StdErr are in annualized %

Figure 3.3: Three-factor model, US industry portfolios

we typically have several factors. We stack the returns for n assets to get

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

$$R^e_t = \alpha + \beta f_t + \varepsilon_t, \text{ where } \mathrm{E}\,\varepsilon_t = \mathbf{0}_{n\times 1} \text{ and } \mathrm{Cov}(f_t, \varepsilon_t) = \mathbf{0}_{K\times n}, \qquad (3.32)$$

where α is $n \times 1$ and β is $n \times K$. Notice that β_{ij} shows how the ith asset depends on the jth factor.

This is, of course, very similar to the CAPM (one-factor) model—and both the LS and GMM approaches are straightforward to extend. I will elaborate on the GMM approach.

3.4.2 Multi-Factor Model: GMM

The moment conditions are

$$\mathrm{E}\, g_t(\alpha, \beta) = \mathrm{E}\left(\begin{bmatrix} 1 \\ f_t \end{bmatrix} \otimes \varepsilon_t\right) = \mathrm{E}\left(\begin{bmatrix} 1 \\ f_t \end{bmatrix} \otimes (R^e_t - \alpha - \beta f_t)\right) = \mathbf{0}_{n(1+K)\times 1}. \qquad (3.33)$$

Note that this expression looks similar to (3.15)—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.

When the system is exactly identified, the GMM estimator solves

$$\bar{g}(\alpha, \beta) = \mathbf{0}_{n(1+K)\times 1}, \qquad (3.34)$$

which is the same as LS equation by equation. Instead, when we restrict $\alpha = \mathbf{0}_{n\times 1}$ (overidentified system), then we either specify a weighting matrix W and solve

$$\min_\beta \bar{g}(\beta)' W \bar{g}(\beta), \qquad (3.35)$$

or we specify a matrix A to combine the moment conditions and solve

$$A_{nK\times n(1+K)}\bar{g}(\beta) = \mathbf{0}_{nK\times 1}. \qquad (3.36)$$

For instance, to get the classical LS estimator without intercepts we specify

$$A = \begin{bmatrix} 0_{nK\times n} & I_{nK} \end{bmatrix} \mathrm{E}\left(\begin{bmatrix} 1 \\ f_t \end{bmatrix} \otimes (R^e_t - \beta f_t)\right). \qquad (3.37)$$

Example 11 *(Moment condition with two assets and two factors) The moment conditions for $n = 2$ and $K = 2$ are*

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

Restricting $\alpha_1 = \alpha_2 = 0$ gives the moment conditions for the overidentified case.

3.4.3 Empirical Evidence

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).

3.4.4 Coding of the GMM Problem

This section describes how the GMM problem can be programmed. We treat the case with n assets and K Factors (which are all excess returns). The moments are of the form

$$g_t = \left(\begin{bmatrix} 1 \\ f_t \end{bmatrix} \otimes (R_t^e - \alpha - \beta f_t) \right) \tag{3.38}$$

$$g_t = \left(\begin{bmatrix} 1 \\ f_t \end{bmatrix} \otimes (R_t^e - \beta f_t) \right) \tag{3.39}$$

for the exactly identified and overidentified case respectively

We want to write the moments on the form

$$g_t = z_t \left(y_t - x_t' b \right), \tag{3.40}$$

to make it easy to use matrix algrebra in the calcuation of the estimate. In that case we could let

$$\Sigma_{zy} = \frac{1}{T} \sum_{t=1}^{T} z_t y_t \text{ and } \Sigma_{zx} = \frac{1}{T} \sum_{t=1}^{T} z_t x_t', \text{ so } \frac{1}{T} \sum_{t=1}^{T} g_t = \Sigma_{zy} - \Sigma_{zx} b. \tag{3.41}$$

In the exactly identified case, we then have

$$\bar{g}_t = \Sigma_{zy} - \Sigma_{zx} b = \mathbf{0}, \text{ so } \hat{b} = \Sigma_{zx}^{-1} \Sigma_{zy}. \tag{3.42}$$

(It is straightforward to show that this can also be calculated equation by equation.) In the

overidentified case with a weighting matrix, the loss function can be written

$$\bar{g}' W \bar{g} = (\Sigma_{zy} - \Sigma_{zx} b)' W (\Sigma_{zy} - \Sigma_{zx} b), \text{ so}$$
$$\Sigma_{zx}' W \Sigma_{zy} - \Sigma_{zx}' W \Sigma_{zx} \hat{b} = \mathbf{0} \text{ and } \hat{b} = (\Sigma_{zx}' W \Sigma_{zx})^{-1} \Sigma_{zx}' W \Sigma_{zy}. \tag{3.43}$$

In the overidentified case when we premultiply the moment conditions by A, we get

$$A\bar{g} = A\Sigma_{zy} - A\Sigma_{zx} b = \mathbf{0}, \text{ so } b = (A\Sigma_{zx})^{-1} A\Sigma_{zy}. \tag{3.44}$$

In practice, we never perform an explicit inversion—it is typically much better (in terms of both speed and precision) to let the software solve the system of linear equations instead.

It is straightforward to show that this works nice if we write the moment conditions as

$$g_t = \underbrace{\left(\begin{bmatrix} 1 \\ f_t \end{bmatrix} \otimes I_n \right)}_{z_t} \left(R_t^e - \underbrace{\left(\begin{bmatrix} 1 \\ f_t \end{bmatrix}' \otimes I_n \right)}_{x_t'} b \right), \text{ with } b = \operatorname{vec}(\alpha, \beta) \tag{3.45}$$

$$g_t = \underbrace{\left(\begin{bmatrix} 1 \\ f_t \end{bmatrix} \otimes I_n \right)}_{z_t} \left(R_t^e - \underbrace{(f_t' \otimes I_n)}_{x_t'} b \right), \text{ with } b = \operatorname{vec}(\beta) \tag{3.46}$$

for the exactly identified and overidentified case respectively. Clearly, z_t and x_t are matrices, not vectors. (z_t is $n(1+K) \times n$ and either of the same dimension or has one row less.)

Example 12 *(Rewriting the moment conditions) For the moment conditions in Example 11 we have*

$$g_t(\alpha, \beta) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ f_{1t} & 0 \\ 0 & f_{1t} \\ f_{2t} & 0 \\ 0 & f_{2t} \end{bmatrix} \left(\begin{bmatrix} R_{1t}^e \\ R_{2t}^e \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ f_{1t} & 0 \\ 0 & f_{1t} \\ f_{2t} & 0 \\ 0 & f_{2t} \end{bmatrix}' \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta_{11} \\ \beta_{21} \\ \beta_{12} \\ \beta_{22} \end{bmatrix} \right).$$

Proof. (of 3.45))From the properties of Kronecker products, we know that *(i)* $\operatorname{vec}(ABC) = (C' \otimes A)\operatorname{vec}(B)$; and *(ii)* if a is $m \times 1$ and c is $n \times 1$, then $a \otimes c = (a \otimes I_n)c$. The first rule allows to write

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

The second rule allows us two write

$$\begin{bmatrix} 1 \\ f_t \end{bmatrix} \otimes (R_t^e - \alpha - \beta f_t) \text{ as } \underbrace{\left(\begin{bmatrix} 1 \\ f_t \end{bmatrix} \otimes I_n \right)}_{z_t} (R_t^e - \alpha - \beta f_t).$$

(For the exactly identified case, we could also use the fact $(A \otimes B)' = A' \otimes B'$ to notice that $z_t = x_t$.) ■

Remark 13 *(*Quick matrix calculations of Σ_{zx} and Σ_{zy}) Although a loop wouldn't take too long time to calculate Σ_{zx} and Σ_{zy}, there is a quicker way. Put $\begin{bmatrix} 1 \\ f_t \end{bmatrix}'$ in row t of the matrix $Z_{T\times(1+K)}$ and $R_t^{e\prime}$ in row t of the matrix $R_{T\times n}$. For the exactly identified case, let $X = Z$. For the overidentified case, put f_t' in row t of the matrix $X_{T\times K}$. Then, calculate*

$$\Sigma_{zx} = (Z'X/T) \otimes I_n \text{ and } vec(R'Z/T) = \Sigma_{zy}.$$

3.5 Testing Multi-Factor Models (General Factors)

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

3.5.1 GMM Estimation with General Factors

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

$$\mathrm{E}\, R_{it}^e = \beta_i'\lambda, \text{ where } \lambda \text{ is } K \times 1, \text{ for } i = 1, \dots n, \text{ or} \tag{3.47}$$

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

$$\mathrm{E}\, R_t^e = \beta\lambda, \tag{3.48}$$

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 (3.32) (equation by equation); second, use $\hat{\beta}_i$ as regressors in a regression equation of the type (3.47) with a residual added

$$\Sigma_{t=1}^T R_{it}^e / T = \hat{\beta}_i'\lambda + u_i. \tag{3.49}$$

It is then tested if $u_i = 0$ for all assets $i = 1, \dots, n$. 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)* it gives a downwards bias of the estimates of λ and an upward bias of the mean of the fitted residuals; and *(ii)* invalidates the standard expression of the test of λ.

A way to handle these problems is to combine the moment conditions for the regression function (3.33) (to estimate β) with (3.48) (to estimate λ) to get a joint system

$$\mathrm{E}\, g_t(\alpha, \beta, \lambda) = \mathrm{E} \begin{bmatrix} \begin{bmatrix} 1 \\ f_t \end{bmatrix} \otimes (R_t^e - \alpha - \beta f_t) \\ R_t^e - \beta\lambda \end{bmatrix} = \mathbf{0}_{n(1+K+1)\times 1}. \tag{3.50}$$

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. Notice that this is a non-linear estimation problem, since the parameters in β multiply the parameters in λ. From the GMM estimation using (3.50) 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 are priced (each of them or perhaps all jointly) by using a Wald test.

Example 14 *(Two assets and one factor) we have the moment conditions*

$$\mathrm{E}\, g_t(\alpha_1,\alpha_2,\beta_1,\beta_2,\lambda) = \mathrm{E}\begin{bmatrix} R^e_{1t}-\alpha_1-\beta_1 f_t \\ R^e_{2t}-\alpha_2-\beta_2 f_t \\ f_t(R^e_{1t}-\alpha_1-\beta_1 f_t) \\ f_t(R^e_{2t}-\alpha_2-\beta_2 f_t) \\ R^e_{1t}-\beta_1\lambda \\ R^e_{2t}-\beta_2\lambda \end{bmatrix} = \mathbf{0}_{6\times 1}.$$

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.

3.5.2 Traditional Cross-Sectional Regressions as Special Cases

Instead of specifying a weighting matrix, we could combine the moment equations so they become equal to the number of parameters, for instance by specifying a matrix A and combine as $A\,\mathrm{E}\,g_t=\mathbf{0}$. This does not generate any overidentifying restrictions, but it still allows us to test hypothesis about λ. One possibility is to let the upper left block of A be an identity matrix and just combine the last n moment conditions, $R^e_t-\beta\lambda$, to just K moment conditions

$$\begin{bmatrix} I_{n(1+K)} & \mathbf{0}_{n(1+K)\times n} \\ \mathbf{0}_{K\times n(1+K)} & \theta_{K\times n} \end{bmatrix} \mathrm{E}\begin{bmatrix} \begin{bmatrix} 1 \\ f_t \end{bmatrix} \otimes (R^e_t-\alpha-\beta f_t) \\ R^e_t-\beta\lambda \end{bmatrix} = \mathbf{0}_{n(1+K)\times 1} \quad (3.51)$$

$$\mathrm{E}\begin{bmatrix} \begin{bmatrix} 1 \\ f_t \end{bmatrix} \otimes (R^e_t-\alpha-\beta f_t) \\ \theta(R^e_t-\beta\lambda) \end{bmatrix} = \quad (3.52)$$

In this case, we can estiumate α and β with LS equation by equation—as a standard time-series regression of a factor model. To estimate the $K\times 1$ vector λ, notice that we can solve the 2nd set of moment conditions as

$$\theta\,\mathrm{E}(R^e_t-\beta\lambda) = \mathbf{0}_{K\times 1} \text{ or } \lambda = (\theta\beta)^{-1}\theta\,\mathrm{E}\,R^e_t, \quad (3.53)$$

which is just like a cross-sectional instrumental variables regression of $\mathrm{E}\,R^e_t=\beta\lambda$ (with β being the regressors, θ the instruments, and $\mathrm{E}\,R^e_t$ the dependent variable).

With $\theta=\beta'$, we get the traditional cross-sectional approach (3.47). The only difference is we here take the uncertainty about the generated betas into account (in the testing). With Alternatively, let Σ be the covariance matrix of the residuals from the time-series estimation of the factor model. Then, using $\theta=\beta'\Sigma$ gives a traditional GLS cross-sectional approach.

Example 15 *(LS cross-sectional regression) With the moment conditions in Example (14) and the weighting vector $\theta=[\beta_1,\beta_2]$ we get*

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

which has as many parameters as moment conditions.

3.5.3 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 (3.32) imply that

$$\mathrm{E}\, R_t^e = \alpha + \beta\, \mathrm{E}\, f_t. \tag{3.54}$$

Equate the expected returns in the two formulations (3.54) and (3.47) to get

$$\alpha = \beta(\lambda - \mathrm{E}\, f_t), \tag{3.55}$$

which is another way of summarizing the restrictions that the linear factor model gives. We can then rewrite the moment conditions (3.50) as (substitute for α and skip the last set of moments)

$$\mathrm{E}\, g_t(\beta, \lambda) = \mathrm{E}\left[\begin{bmatrix} 1 \\ f_t \end{bmatrix} \otimes (R_t^e - \beta(\lambda - \mathrm{E}\, f_t) - \beta f_t)\right] = \mathbf{0}_{n(1+K)\times 1}. \tag{3.56}$$

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).

Example 16 *Two assets and one factor) Use the restrictions (3.55) in the moment conditions for that case (compare with (3.19)) to get*

$$\mathrm{E}\, g_t(\beta_1, \beta_2, \lambda) = \mathrm{E}\begin{bmatrix} R_{1t}^e - \beta_1(\lambda - \mathrm{E}\, f_t) - \beta_1 f_t \\ R_{2t}^e - \beta_2(\lambda - \mathrm{E}\, f_t) - \beta_2 f_t \\ f_t[R_{1t}^e - \beta_1(\lambda - \mathrm{E}\, f_t) - \beta_1 f_t] \\ f_t[R_{2t}^e - \beta_2(\lambda - \mathrm{E}\, f_t) - \beta_2 f_t] \end{bmatrix} = \mathbf{0}_{4\times 1}.$$

This gives 4 moment conditions, but only three parameters, so there is one overidentifying restriction to test—just as with (3.51).

3.5.4 What if the Factor is a Portfolio?

It would (perhaps) be natural if the tests discussed in this section coincided with those in Section 3.4 when the factors are in fact 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 3.4 *impose* $\lambda = \mathrm{E}\, f_t$. If we were to put this restriction on (3.56), then we are back to the LM test of the multifactor model where (3.56) 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).

3.5.5 Empirical Evidence

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. Breeden, Gibbons, and Litzenberger (1989) and Lettau and Ludvigson (2001) estimate models where consumption growth is the factor—with very mixed results.

3.6 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 3.5. The method has three steps, described below.

- First, estimate the betas β_i ($i = 1, \ldots, n$) from (3.1) (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

$$R_{it}^e = \lambda_t' \hat{\beta}_i + \varepsilon_{it}, \tag{3.57}$$

where $\hat{\beta}_i$ are the regressors. (Note the difference to the traditional cross-sectional approach discussed in (3.14), where the second stage regression regressed E R_{it}^e on $\hat{\beta}_i$, while the Fama-French approach runs one regression for every time period.)

- Third, estimate the time averages

$$\hat{\varepsilon}_i = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{it} \text{ for } i = 1, \ldots, n, \text{ (for every asset)} \tag{3.58}$$

$$\hat{\lambda} = \frac{1}{T}\sum_{t=1}^{T}\hat{\lambda}_t. \tag{3.59}$$

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 (3.57) 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 can test the model by studying if $\varepsilon_i = 0$ (recall from (3.58) that ε_i is the time average of the residual for asset i, ε_{it}), by forming a t-test $\hat{\varepsilon}_i / \text{Std}(\hat{\varepsilon}_i)$. Fama and MacBeth (1973) suggest that the standard deviation should be found by studying the time-variation in $\hat{\varepsilon}_{it}$. In particular, they suggest that the variance of $\hat{\varepsilon}_{it}$ (not $\hat{\varepsilon}_i$) can be estimated by the (average) squared variation around its mean

$$\text{Var}(\hat{\varepsilon}_{it}) = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it} - \hat{\varepsilon}_i\right)^2. \tag{3.60}$$

Since $\hat{\varepsilon}_i$ is the sample average of $\hat{\varepsilon}_{it}$, the variance of the former is the variance of the latter divided by T (the sample size)—provided $\hat{\varepsilon}_{it}$ is iid. That is,

$$\text{Var}(\hat{\varepsilon}_i) = \frac{1}{T}\text{Var}(\hat{\varepsilon}_{it}) = \frac{1}{T^2}\sum_{t=1}^{T}\left(\hat{\varepsilon}_{it} - \hat{\varepsilon}_i\right)^2. \tag{3.61}$$

A similar argument leads to the variance of $\hat{\lambda}$

$$\text{Var}(\hat{\lambda}) = \frac{1}{T^2}\sum_{t=1}^{T}(\hat{\lambda}_t - \hat{\lambda})^2. \tag{3.62}$$

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.

Bibliography

Breeden, D. T., M. R. Gibbons, and R. H. Litzenberger, 1989, “Empirical Tests of the Consumption-Oriented CAPM,” *Journal of Finance*, 44, 231–262.

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., 2003, *Econometric Analysis*, Prentice-Hall, Upper Saddle River, New Jersey, 5th 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.

Lettau, M., and S. Ludvigson, 2001, “Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying,” *Journal of Political Economy*, 109, 1238–1287.

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