12

Regression-Based Tests of Linear Factor Models

THE NEXT FOUR chapters study the question, how should we estimate and evaluate linear factor models: models of the form $p = E(mx)$, $m = b'f$ or equivalently $E(R^e) = \beta'\lambda$? These models are by far the most common in empirical asset pricing, and there is a large literature on econometric techniques to estimate and evaluate them. Each technique focuses on the same questions: how to estimate parameters, how to calculate standard errors of the estimated parameters, how to calculate standard errors of the pricing errors, and how to test the model, usually with a test statistic of the form $\hat{\alpha}' V^{-1} \hat{\alpha}$.

I start in this chapter with simple and long-standing time-series and cross-sectional regression tests. In Chapter 13, I pursue the GMM approach to the model expressed in $p = E(mx)$, $m = b'f$ form. Chapter 14 summarizes the principle of maximum likelihood estimation and derives maximum likelihood estimates and tests. Finally, Chapter 15 compares the different approaches.

As always, the theme is the underlying unity. All of the techniques come down to one of two basic ideas: time-series regression or cross-sectional regression. Time-series regression turns out to be a limiting case of cross-sectional regression. The GMM, $p = E(mx)$ approach turns out to be almost identical to cross-sectional regressions. Maximum likelihood (with appropriate statistical assumptions) justifies the time-series and cross-sectional regression approaches. The formulas for parameter estimates, standard errors, and test statistics are all strikingly similar.

229

12.1 Time-Series Regressions

When the factor is also a return, we can evaluate the model

$$E(R^{ei}) = \beta_i E(f)$$

by running OLS *time-series regressions*

$$R^{ei}_t = \alpha_i + \beta_i f_t + \varepsilon^i_t, \qquad t = 1, 2, \ldots, T,$$

for each asset. The OLS distribution formulas (with corrected standard errors) provide standard errors of α and β.

With errors that are i.i.d. over time, homoskedastic, and independent of the factors, the asymptotic joint distribution of the intercepts gives the model test statistic,

$$T\left[1 + \left(\frac{E_T(f)}{\hat{\sigma}(f)}\right)^2\right]^{-1} \hat{\alpha}' \widehat{\Sigma}^{-1} \hat{\alpha} \sim \chi^2_N.$$

The Gibbons–Ross–Shanken test is a multivariate, finite-sample counterpart to this statistic, when the errors are also normally distributed,

$$\frac{T - N - K}{N}\left(1 + E_T(f)' \widehat{\Omega}^{-1} E_T(f)\right)^{-1} \hat{\alpha}' \widehat{\Sigma}^{-1} \hat{\alpha} \sim F_{N, T-N-K}.$$

I show how to construct the same test statistics with heteroskedastic and autocorrelated errors via GMM.

I start with the simplest case. We have a factor pricing model with a single factor. The factor is an excess return (for example, the CAPM, with $R^{em} = R^m - R^f$), and the test assets are all excess returns. We express the model in expected return-beta form. The betas are defined by regression coefficients

$$R^{ei}_t = \alpha_i + \beta_i f_t + \varepsilon^i_t \tag{12.1}$$

and the model states that expected returns are linear in the betas:

$$E(R^{ei}) = \beta_i E(f). \tag{12.2}$$

Since the factor is also an excess return, the model applies to the factor as well, so $E(f) = 1 \times \lambda$.

Comparing the model (12.2) and the expectation of the time-series regression (12.1), we see that the model has one and only one implication for the data: *all the regression intercepts α_i should be zero*. The regression intercepts are equal to the pricing errors.

Given this fact, Black, Jensen, and Scholes (1972) suggested a natural strategy for estimation and evaluation: Run time-series regressions (12.1) for

each test asset. The estimate of the factor risk premium is just the sample mean of the factor,

$$\hat{\lambda} = E_T(f).$$

Then, use standard OLS formulas for a distribution theory of the parameters. In particular, you can use t-tests to check whether the pricing errors α are in fact zero. These distributions are usually presented for the case that the regression errors in (12.1) are uncorrelated and homoskedastic, but the formulas in Section 11.4 show easily how to calculate standard errors for arbitrary error covariance structures.

We also want to know whether all the pricing errors are *jointly* equal to zero. This requires us to go beyond standard formulas for the regression (12.1) taken alone, as we want to know the joint distribution of α estimates from separate regressions running side by side but with errors correlated across assets ($E(\varepsilon_t^i \varepsilon_t^j) \neq 0$). (We can think of (12.1) as a panel regression, and then it is a test whether the firm dummies are jointly zero.) The classic form of these tests assume no autocorrelation or heteroskedasticity. Dividing the $\hat{\alpha}$ regression coefficients by their variance-covariance matrix leads to a χ^2 test,

$$T \left[1 + \left(\frac{E_T(f)}{\hat{\sigma}(f)} \right)^2 \right]^{-1} \hat{\alpha}' \widehat{\Sigma}^{-1} \hat{\alpha} \sim \chi_N^2, \tag{12.3}$$

where $E_T(f)$ denotes sample mean, $\hat{\sigma}^2(f)$ denotes sample variance, $\hat{\alpha}$ is a vector of the estimated intercepts,

$$\hat{\alpha} = \begin{bmatrix} \hat{\alpha}_1 & \hat{\alpha}_2 & \cdots & \hat{\alpha}_N \end{bmatrix}'.$$

$\widehat{\Sigma}$ is the residual covariance matrix, i.e., the sample estimate of $E(\varepsilon_t \varepsilon_t') = \Sigma$, where

$$\varepsilon_t = \begin{bmatrix} \varepsilon_t^1 & \varepsilon_t^2 & \cdots & \varepsilon_t^N \end{bmatrix}'.$$

As usual when testing hypotheses about regression coefficients, this test is valid asymptotically. The asymptotic distribution theory assumes that $\sigma^2(f)$ (i.e., $X'X$ and Σ have converged to their probability limits; therefore, it is asymptotically valid even though the factor is stochastic and Σ is estimated, but it ignores those sources of variation in a finite sample. It does not require that the errors are normal, relying on the central limit theorem so that $\hat{\alpha}$ is normal. I derive (12.3) below.

Also as usual in a regression context, we can derive a finite-sample F distribution for the hypothesis that a set of parameters are jointly zero,

$$\frac{T - N - 1}{N} \left[1 + \left(\frac{E_T(f)}{\hat{\sigma}(f)} \right)^2 \right]^{-1} \hat{\alpha}' \widehat{\Sigma}^{-1} \hat{\alpha} \sim F_{N,\, T-N-1}. \tag{12.4}$$

This is the Gibbons, Ross, and Shanken (1989) or "GRS" test statistic. The F distribution recognizes sampling variation in $\widehat{\Sigma}$, which is not included in (12.3). This distribution requires that the errors ε are normal as well as uncorrelated and homoskedastic. With normal errors, the $\hat{\alpha}$ are normal and $\widehat{\Sigma}$ is an independent Wishart (the multivariate version of a χ^2), so the ratio is F. This distribution is exact in a finite sample.

Tests (12.3) and (12.4) have a very intuitive form. The basic part of the test is a quadratic form in the pricing errors, $\hat{\alpha}'\widehat{\Sigma}^{-1}\hat{\alpha}$. If there were no βf in the model, then the $\hat{\alpha}$ would simply be the sample mean of the regression errors ε_t. Assuming i.i.d. ε_t, the variance of their sample mean is just $1/T\Sigma$. Thus, if we knew Σ, then $T\hat{\alpha}'\Sigma^{-1}\hat{\alpha}$ would be a sum of squared sample means divided by their variance-covariance matrix, which would have an asymptotic χ_N^2 distribution, or a finite-sample χ_N^2 distribution if the ε_t are normal. But we have to estimate Σ, which is why the finite-sample distribution is F rather than χ^2. We also estimate the β, and the second term in (12.3) and (12.4) accounts for that fact.

Recall that a single-beta representation exists if and only if the reference return is on the mean-variance frontier. Thus, the test can also be interpreted as a test whether f is ex ante mean-variance efficient—whether it is on the mean-variance frontier using *population* moments—after accounting for sampling error. Even if f is on the true or ex ante mean-variance frontier, other returns will outperform it in sample due to luck, so the return f will usually be inside the ex post mean-variance frontier—i.e., the frontier drawn using *sample* moments. Still, it should not be too far inside the sample frontier. Gibbons, Ross, and Shanken show that the test statistic can be expressed in terms of how far inside the ex post frontier the return f is,

$$\frac{T-N-1}{N} \frac{\left(\mu_q/\sigma_q\right)^2 - \left(E_T(f)/\hat{\sigma}(f)\right)^2}{1 + \left(E_T(f)/\hat{\sigma}(f)\right)^2}. \tag{12.5}$$

$\left(\mu_q/\sigma_q\right)^2$ is the Sharpe ratio of the ex post tangency portfolio (maximum ex post Sharpe ratio) formed from the test assets plus the factor f. The last term in the numerator is the Sharpe ratio of the factor, so the numerator expresses how for the factor is inside the ex-post frontier.

If there are many factors that are excess returns, the same ideas work, with some cost of algebraic complexity. The regression equation is

$$R^{ei} = \alpha_i + \beta_i'f_t + \varepsilon_t^i.$$

The asset pricing model

$$E(R^{ei}) = \beta_i'E(f)$$

again predicts that the intercepts should be zero. We can estimate α and β with OLS time-series regressions. Assuming normal i.i.d. errors,

the quadratic form $\hat{\alpha}'\widehat{\Sigma}^{-1}\hat{\alpha}$ has the distribution

$$\frac{T-N-K}{N}\left(1 + E_T(f)'\widehat{\Omega}^{-1}E_T(f)\right)^{-1}\hat{\alpha}'\widehat{\Sigma}^{-1}\hat{\alpha} \sim F_{N,\,T-N-K}, \qquad (12.6)$$

where

$$N = \text{number of assets},$$

$$K = \text{number of factors},$$

$$\widehat{\Omega} = \frac{1}{T}\sum_{t=1}^{T}\left[f_t - E_T(f)\right]\left[f_t - E_T(f)\right]',$$

$$\widehat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t\,\hat{\varepsilon}_t'.$$

The main difference is that the Sharpe ratio of the single factor is replaced by the natural generalization $E_T(f)'\widehat{\Omega}^{-1}E_T(f)$.

Derivation of The χ^2 Statistic, and Distributions with General Errors

I derive (12.3) as an instance of GMM. This approach allows us to generate straightforwardly the required corrections for autocorrelated and heteroskedastic disturbances. (MacKinlay and Richardson [1991] advocate GMM approaches to regression tests in this way.) It also serves to remind us that GMM and $p = E(mx)$ are not necessarily paired; one can do a GMM estimate of an expected return-beta model, too. The mechanics are only slightly different than what we did to generate distributions for OLS regression coefficients in Section 11.4, since we keep track of N OLS regressions simultaneously.

Write the equations for all N assets together in vector form,

$$R_t^e = \alpha + \beta f_t + \varepsilon_t.$$

We use the usual OLS moments to estimate the coefficients,

$$g_T(b) = \begin{bmatrix} E_T(R_t^e - \alpha - \beta f_t) \\ E_T[(R_t^e - \alpha - \beta f_t)f_t] \end{bmatrix} = E_T\left(\begin{bmatrix} \varepsilon_t \\ f_t\varepsilon_t \end{bmatrix}\right) = 0.$$

These moments exactly identify the parameters α, β, so the a matrix in $ag_T(\hat{b}) = 0$ is the identity matrix. Solving, the GMM estimates are of course the OLS estimates,

$$\hat{\alpha} = E_T(R_t^e) - \hat{\beta}E_T(f_t),$$

$$\hat{\beta} = \frac{E_T[(R_t^e - E_T(R_t^e))f_t]}{E_T[(f_t - E_T(f_t))f_t]} = \frac{\text{cov}_T(R_t^e, f_t)}{\text{var}_T(f_t)}.$$

The d matrix in the general GMM formula is

$$d \equiv \frac{\partial g_T(b)}{\partial b'} = -\begin{bmatrix} I_N & I_N E(f_t) \\ I_N E(f_t) & I_N E(f_t^2) \end{bmatrix} = -\begin{bmatrix} 1 & E(f_t) \\ E(f_t) & E(f_t^2) \end{bmatrix} \otimes I_N,$$

where I_N is an $N \times N$ identity matrix. The S matrix is

$$S = \sum_{j=-\infty}^{\infty} \begin{bmatrix} E(\varepsilon_t \varepsilon'_{t-j}) & E(\varepsilon_t \varepsilon'_{t-j} f_{t-j}) \\ E(f_t \varepsilon_t \varepsilon'_{t-j}) & E(f_t \varepsilon_t \varepsilon'_{t-j} f_{t-j}) \end{bmatrix}.$$

Using the GMM variance formula (11.4) with $a = I$, we have

$$\text{var}\left(\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}\right) = \frac{1}{T} d^{-1} S d^{-1'}. \tag{12.7}$$

At this point, we are done. The upper left-hand corner of $\text{var}(\hat{\alpha}\ \hat{\beta})$ gives us $\text{var}(\hat{\alpha})$ and the test we are looking for is $\hat{\alpha}'\text{var}(\hat{\alpha})^{-1}\hat{\alpha} \sim \chi^2_N$.

The standard formulas make this expression prettier by assuming that the errors are uncorrelated over time and not heteroskedastic. These assumptions simplify the S matrix, as for the standard OLS formulas in Section 11.4. If we assume that f and ε are independent as well as orthogonal, $E(f\varepsilon\varepsilon') = E(f)E(\varepsilon\varepsilon')$ and $E(f^2\varepsilon\varepsilon') = E(f^2)E(\varepsilon\varepsilon')$. If we assume that the errors are independent over time as well, we lose all the lead and lag terms. Then, the S matrix simplifies to

$$S = \begin{bmatrix} E(\varepsilon_t \varepsilon'_t) & E(\varepsilon_t \varepsilon'_t)E(f_t) \\ E(f_t)E(\varepsilon_t \varepsilon'_t) & E(\varepsilon_t \varepsilon'_t)E(f_t^2) \end{bmatrix} = \begin{bmatrix} 1 & E(f_t) \\ E(f_t) & E(f_t^2) \end{bmatrix} \otimes \Sigma. \tag{12.8}$$

Now we can plug into (12.7). Using $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ and $(A \otimes B)(C \otimes D) = AC \otimes BD$, we obtain

$$\text{var}\left(\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}\right) = \frac{1}{T}\left(\begin{bmatrix} 1 & E(f_t) \\ E(f_t) & E(f_t^2) \end{bmatrix}^{-1} \otimes \Sigma\right).$$

Evaluating the inverse,

$$\text{var}\left(\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}\right) = \frac{1}{T}\frac{1}{\text{var}(f)}\begin{bmatrix} E(f_t^2) & -E(f_t) \\ -E(f_t) & 1 \end{bmatrix} \otimes \Sigma.$$

We are interested in the top left corner. Using $E(f^2) = E(f)^2 + \text{var}(f)$,

$$\text{var}(\hat{\alpha}) = \frac{1}{T}\left(1 + \frac{E(f)^2}{\text{var}(f)}\right)\Sigma.$$

This is the traditional formula (12.3). Though this formula is pretty, there is now no real reason to assume that the errors are i.i.d. or independent of the factors. By simply calculating (12.7), we can easily construct standard errors and test statistics that do not require these assumptions.

12.2 Cross-Sectional Regressions

> We can fit
> $$E(R^{ei}) = \beta_i'\lambda + \alpha_i$$
> by running a *cross-sectional* regression of average returns on the betas. This technique can be used whether the factor is a return or not.
>   I discuss OLS and GLS cross-sectional regressions, I find formulas for the standard errors of λ, and a χ^2 test whether the α are jointly zero. I derive the distributions as an instance of GMM, and I show how to implement the same approach for autocorrelated and heteroskedastic errors. I show that the GLS cross-sectional regression is the same as the time-series regression when the factor is also an excess return, and is included in the set of test assets.

Start again with the *K* factor model, written as

$$E(R^{ei}) = \beta_i'\lambda, \qquad i = 1, 2, \ldots, N.$$

The central economic question is why average returns vary *across* assets; expected returns of an asset should be high if that asset has high betas or a large risk exposure to factors that carry high risk premia.

Figure 12.1 graphs the case of a single factor such as the CAPM. Each dot represents one asset i. The model says that average returns should be proportional to betas, so plot the sample average returns against the betas. Even if the model is true, this plot will not work out perfectly in each sample, so there will be some spread as shown.

Given these facts, a natural idea is to run a *cross-sectional regression* to fit a line through the scatterplot of Figure 12.1. First find estimates of the betas from time-series regressions,

$$R_t^{ei} = a_i + \beta_i'f_t + \varepsilon_t^i, \qquad t = 1, 2, \ldots, T \quad \text{for each } i. \tag{12.9}$$

Then estimate the factor risk premia λ from a regression across assets of average returns on the betas,

$$E_T(R^{ei}) = \beta_i'\lambda + \alpha_i, \qquad i = 1, 2, \ldots, N. \tag{12.10}$$

As in the figure, β are the right-hand variables, λ are the regression coefficients, and the cross-sectional regression residuals α_i are the pricing errors.
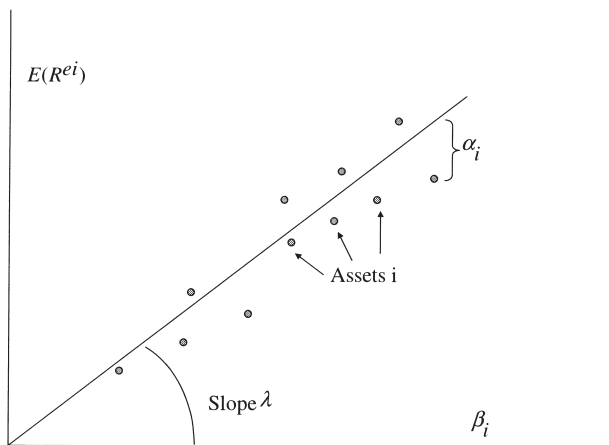
Figure 12.1.   *Cross-sectional regression.*

This is also known as a *two-pass* regression estimate, because one estimates first time-series and then cross-sectional regressions.

You can run the cross-sectional regression with or without a constant. The theory says that the constant or zero-beta excess return should be zero. You can impose this restriction or estimate a constant and see if it turns out to be small. The usual trade-off between efficiency (impose the null as much as possible to get efficient estimates) and robustness applies.

OLS Cross-Sectional Regression

It will simplify notation to consider a single factor; the case of multiple factors looks the same with vectors in place of scalars. I denote vectors from 1 to N with missing sub or superscripts, i.e., $\varepsilon_t = \begin{bmatrix} \varepsilon_t^1 & \varepsilon_t^2 & \cdots & \varepsilon_t^N \end{bmatrix}'$, $\beta = \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_N \end{bmatrix}'$, and similarly for R_t^e and α. For simplicity take the case of no intercept in the cross-sectional regression. With this notation OLS cross-sectional estimates are

$$\hat{\lambda} = \left(\beta' \beta \right)^{-1} \beta' E_T(R^e),$$

$$\hat{\alpha} = E_T(R^e) - \hat{\lambda} \beta. \tag{12.11}$$

Next, we need a distribution theory for the estimated parameters. The most natural place to start is with the standard OLS distribution formulas. I start with the traditional assumption that the true errors are i.i.d. over time, and independent of the factors. This will give us some easily interpretable formulas, and we will see most of these terms remain when we do the distribution theory right later on.

In an OLS regression $Y = X\beta + u$ and $E(uu') = \Omega$, the variance of the β estimate is $(X'X)^{-1}X'\Omega X(X'X)^{-1}$. The residual covariance matrix is $(I - X(X'X)^{-1}X')\Omega(I - X(X'X)^{-1}X')'$.

To apply these formulas we need $\text{cov}(\alpha, \alpha')$, the error covariance in the cross-sectional regression. With the traditional assumption that the factors and errors are i.i.d. over time, the answer is $\text{cov}(\alpha, \alpha') = \frac{1}{T}\left(\beta\Sigma_f\beta' + \Sigma\right)$, where $\Sigma_f \equiv \text{cov}(f_t, f_t')$ and $\Sigma = \text{cov}(\varepsilon_t\varepsilon_t')$. To see this, start with $\alpha = E_T(R^e) - \beta\lambda$. With $R_t^e = a + \beta f_t + \varepsilon_t$, we have $E_T\left(R_t^e\right) = a + \beta E_T\left(f_t\right) + E_T\left(\varepsilon_t\right)$. Under the null that the model is correct, so $E(R^e) = a + \beta E(f) = \beta\lambda$, then, we have $\text{cov}(\alpha, \alpha') = \text{cov}\left[E_T(R^e), E_T(R^e)'\right] = \frac{1}{T}\left(\beta\Sigma_f\beta' + \Sigma\right)$. (Don't confuse this covariance with the covariance of the *estimated* α in the cross-sectional regression. Like a residual covariance vs. an error covariance, there are additional terms in the covariance of the estimated α, which I develop below. Yes, we want the covariance of $E_T(R^e)$, not of $E(R^e)$, which is a number and has no covariance, or of R_t^e. $E_T(R^e)$ is the y variable in the cross-sectional regression.)

Then, the conventional OLS formulas for the covariance matrices of OLS estimates and residuals, accounting for correlated errors, give

$$\sigma^2(\hat{\lambda}) = \frac{1}{T}\left[\left(\beta'\beta\right)^{-1}\beta'\Sigma\beta(\beta'\beta)^{-1} + \Sigma_f\right] \tag{12.12}$$

$$\text{cov}(\hat{\alpha}) = \frac{1}{T}\left[I - \beta(\beta'\beta)^{-1}\beta'\right]\Sigma\left[I - \beta(\beta'\beta)^{-1}\beta'\right]' \tag{12.13}$$

The correct formulas, (12.19) and (12.20), which account for the fact that β are estimated, are straightforward generalizations. (The Σ_f term cancels in (12.13).)

We could test whether all pricing errors are zero with the statistic

$$\hat{\alpha}'\,\text{cov}(\hat{\alpha})^{-1}\hat{\alpha} \sim \chi^2_{N-1}. \tag{12.14}$$

The distribution is χ^2_{N-1}, not χ^2_N, because the covariance matrix is singular. The singularity and the extra terms in (12.13) result from the fact that the λ coefficient was estimated along the way, and means that we have to use a generalized inverse. (If there are K factors, we obviously end up with χ^2_{N-K}.)

A test of the residuals is unusual in OLS regressions. We do not usually test whether the residuals are "too large," since we have no information other than the residuals themselves about how large they should be. In this case, however, the first-stage time-series regression gives us some independent information about the size of $\text{cov}(\alpha\alpha')$, information that we could not get from looking at the cross-sectional residual α itself.

GLS Cross-Sectional Regression

Since the residuals in the cross-sectional regression (12.10) are correlated with each other, standard textbook advice is to run a GLS cross-sectional regression rather than OLS, using $E(\alpha\alpha') = \frac{1}{T}(\Sigma + \beta\Sigma_f\beta')$ as the error covariance matrix:

$$\hat{\lambda} = (\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}E_T(R^e),$$
$$\hat{\alpha} = E_T(R^e) - \hat{\lambda}\beta.$$

(12.15)

(The GLS formula is

$$\hat{\lambda} = \left[\beta'\left(\beta\Sigma_f^{-1}\beta' + \Sigma\right)^{-1}\beta\right]^{-1}\beta'(\beta\Sigma_f^{-1}\beta' + \Sigma)^{-1}E_T(R^e).$$

However, it turns out that we can drop the $\beta\Sigma_f^{-1}\beta'$ term.[1])

The standard regression formulas give the variance of these estimates as

$$\sigma^2(\hat{\lambda}) = \frac{1}{T}\left[(\beta'\Sigma^{-1}\beta)^{-1} + \Sigma_f\right],$$

(12.16)

$$\mathrm{cov}(\hat{\alpha}) = \frac{1}{T}\left[\Sigma - \beta(\beta'\Sigma^{-1}\beta)^{-1}\beta'\right].$$

(12.17)

---

[1] Here's the algebra. Let
$$A = I + \beta'\Sigma^{-1}\beta\Sigma_f^{-1}.$$

Then,

$$\hat{\lambda} = \left[\beta'\left(\beta\Sigma_f^{-1}\beta' + \Sigma\right)^{-1}\beta\right]^{-1} A^{-1}A\beta'\left(\beta\Sigma_f^{-1}\beta' + \Sigma\right)^{-1}E_T(R^e)$$
$$= \left[A\beta'\left(\beta\Sigma_f^{-1}\beta' + \Sigma\right)^{-1}\beta\right]^{-1} A\beta'\left(\beta\Sigma_f^{-1}\beta' + \Sigma\right)^{-1}E_T(R^e).$$

Now,

$$A\beta' = \left(I + \beta'\Sigma^{-1}\beta\Sigma_f^{-1}\right)\beta'$$
$$= \beta'\left(I + \Sigma^{-1}\beta\Sigma_f^{-1}\beta'\right)$$
$$= \beta'\Sigma^{-1}\left(\Sigma + \beta\Sigma_f^{-1}\beta'\right).$$

Thus,

$$\hat{\lambda} = (\beta'\Sigma^{-1}\beta)^{-1}\beta'\Sigma^{-1}E_T(R^e).$$

The comments of Section 11.5 warning that OLS is sometimes much more robust than GLS apply in this case. The GLS regression should improve efficiency, i.e., give more precise estimates. However, Σ may be hard to estimate and to invert, especially if the cross section N is large. One may well choose the robustness of OLS over the asymptotic statistical advantages of GLS.

A GLS regression can be understood as a transformation of the space of returns, to focus attention on the statistically most informative portfolios. Finding (say, by Choleski decomposition) a matrix C such that $CC' = \Sigma^{-1}$, the GLS regression is the same as an OLS regression of $CE_T(R^e)$ on $C\beta$, i.e., of testing the model on the portfolios CR^e. The statistically most informative portfolios are those with the lowest residual variance Σ. But this asymptotic statistical theory assumes that the covariance matrix has converged to its true value. In most samples, the ex post or sample mean-variance frontier still seems to indicate lots of luck, and this is especially true if the cross section is large, anything more than $1/10$ of the time series. The portfolios CR^e are likely to contain many extreme long-short positions.

Again, we could test the hypothesis that all the α are equal to zero with (12.14). Though the appearance of the statistic is the same, the covariance matrix is smaller, reflecting the greater power of the GLS test. As with the J_T test, (11.10), we can develop an equivalent test that does not require a generalized inverse:

$$T\hat{\alpha}'\Sigma^{-1}\hat{\alpha} \sim \chi^2_{N-1}. \tag{12.18}$$

To derive (12.18), I proceed exactly as in the derivation of the J_T test (11.10). Define, say by Choleski decomposition, a matrix C such that $CC' = \Sigma^{-1}$. Now, find the covariance matrix of $\sqrt{T}C'\hat{\alpha}$:

$$\text{cov}\left(\sqrt{T}C'\alpha\right) = C'\left((CC')^{-1} - \beta(\beta'CC'\beta)^{-1}\beta'\right)C = I - \delta(\delta'\delta)^{-1}\delta',$$

where

$$\delta = C'\beta.$$

In sum, $\hat{\alpha}$ is asymptotically normal so $\sqrt{T}C'\hat{\alpha}$ is asymptotically normal, $\text{cov}(\sqrt{T}C'\hat{\alpha})$ is an idempotent matrix with rank $N-1$; therefore $T\hat{\alpha}'CC'\hat{\alpha} = T\hat{\alpha}'\Sigma^{-1}\hat{\alpha}$ is χ^2_{N-1}.

Correction for the Fact that β Are Estimated, and GMM Formulas that Do Not Need i.i.d. Errors

In applying standard OLS formulas to a cross-sectional regression, we assume that the right-hand variables β are fixed. The β in the cross-sectional regression are not fixed, of course, but are estimated in the time-series regression. This turns out to matter, even asymptotically.

In this section, I derive the correct asymptotic standard errors. With the simplifying assumption that the errors ε are i.i.d. over time and independent of the factors, the result is

$$\sigma^2(\hat{\lambda}_{\text{OLS}}) = \frac{1}{T}\left[(\beta'\beta)^{-1}\beta'\Sigma\beta(\beta'\beta)^{-1}(1+\lambda'\Sigma_f^{-1}\lambda) + \Sigma_f\right],$$

$$\sigma^2(\hat{\lambda}_{\text{GLS}}) = \frac{1}{T}\left[(\beta'\Sigma^{-1}\beta)^{-1}(1+\lambda'\Sigma_f^{-1}\lambda) + \Sigma_f\right],$$

(12.19)

where Σ_f is the variance-covariance matrix of the factors. This correction is due to Shanken (1992b). Comparing these standard errors to (12.12) and (12.16), we see that there is a multiplicative correction $(1+\lambda'\Sigma_f^{-1}\lambda)$.

The asymptotic variance-covariance matrix of the pricing errors is

$$\text{cov}(\hat{\alpha}_{\text{OLS}}) = \frac{1}{T}(I_N - \beta(\beta'\beta)^{-1}\beta')\Sigma(I_N - \beta(\beta'\beta)^{-1}\beta')$$
$$\times (1+\lambda'\Sigma_f^{-1}\lambda)$$

(12.20)

$$\text{cov}(\hat{\alpha}_{\text{GLS}}) = \frac{1}{T}(\Sigma - \beta(\beta'\Sigma^{-1}\beta)^{-1}\beta')(1+\lambda'\Sigma_f^{-1}\lambda).$$

(12.21)

Comparing these results to (12.13) and (12.17), we see the same multiplicative correction.

We can form the asymptotic χ^2 test of the pricing errors by dividing pricing errors by their variance-covariance matrix, $\hat{\alpha}\,\text{cov}(\hat{\alpha})^{-1}\hat{\alpha}$. Following (12.18), we can simplify this result for the GLS pricing errors resulting in

$$T(1+\lambda'\Sigma_f^{-1}\lambda)\hat{\alpha}_{\text{GLS}}'\Sigma^{-1}\hat{\alpha}_{\text{GLS}} \sim \chi^2_{N-K}.$$

(12.22)

Is the correction important relative to the simple-to-derive regression formulas (12.12), (12.13), (12.16), (12.17)? In the CAPM, $\lambda = E(R^{em})$ so $\lambda^2/\sigma^2(R^{em}) \approx (0.08/0.16)^2 = 0.25$ in annual data. In annual data, then, the multiplicative term is too large to ignore. However, the mean and variance both scale with horizon, so the Sharpe ratio scales with the square root of horizon. Therefore, for a monthly interval $\lambda^2/\sigma^2(R^{em}) \approx 0.25/12 \approx 0.02$, which is quite small, so ignoring the multiplicative term makes little difference.

Suppose the factor is in fact a return. Then the factor risk premium is $\lambda = E(f)$, and we use Σ_f/T, the standard error of the factor mean, as the standard error of λ. The terms in β correct for the small differences between cross-sectional and time-series estimates. They are therefore likely to be small, and the Σ_f/T term is likely to be the most important term.

Comparing (12.22) to the GRS tests for a time-series regression, (12.3), (12.4), (12.6), we see the same statistic. The only difference is that

by estimating λ from the cross section rather than imposing $\lambda = E(f)$, the cross-sectional regression loses degrees of freedom equal to the number of factors.

Though these formulas are standard classics, I emphasize that we do not have to make the severe assumptions on the error terms that are used to derive them. As with the time-series case, I derive a general formula for the distribution of $\hat{\lambda}$ and $\hat{\alpha}$, and only at the last moment make classic error term assumptions to make the spectral density matrix pretty.

Derivation and Formulas that Do Not Require i.i.d. Errors

The easy and elegant way to account for the effects of "generated regressors" such as the β in the cross-sectional regression is to map the whole thing into GMM. Then, we treat the moments that generate the regressors β at the same time as the moments that generate the cross-sectional regression coefficient λ, and the covariance matrix S between the two sets of moments captures the effects of generating the regressors on the standard error of the cross-sectional regression coefficients. Comparing this straightforward derivation with the difficulty of Shanken's (1992b) paper that originally derived the corrections for $\hat{\lambda}$, and noting that Shanken did not go on to find the formulas (12.20) that allow a test of the pricing errors is a nice argument for the simplicity and power of the GMM framework.

To keep the algebra manageable, I treat the case of a single factor. The moments are

$$g_T(b) = \begin{bmatrix} E(R_t^e - a - \beta f_t) \\ E[(R_t^e - a - \beta f_t)f_t] \\ E(R^e - \beta \lambda) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{12.23}$$

The top two moment conditions exactly identify a and β as the time-series OLS estimates. (Note a not α. The time-series intercept is not necessarily equal to the pricing error in a cross-sectional regression.) The bottom moment condition is the asset pricing model. It is in general overidentified in a sample, since there is only one extra parameter (λ) and N extra moment conditions. If we use a weighting vector β' on this condition, we obtain the OLS cross-sectional estimate of λ. If we use a weighting vector $\beta'\Sigma^{-1}$, we obtain the GLS cross-sectional estimate of λ. To accommodate both cases, use a weighting vector γ', and then substitute $\gamma' = \beta'$ or $\gamma' = \beta'\Sigma^{-1}$ at the end. However, once we abandon i.i.d. errors, the GLS cross-sectional regression weighted by Σ^{-1} is no longer the optimal estimate. Once we recognize that the errors do not obey classical assumptions, and if we want efficient estimates, we might as well calculate the correct and

fully efficient estimates. Having decided on a cross-sectional regression, the efficient estimates of the moments (12.23) are $d'S^{-1}g_T(a, \beta, \lambda) = 0$.

The standard errors for $\hat{\lambda}$ come straight from the general GMM standard error formula (11.4). The $\hat{\alpha}$ are not parameters, but are the last N moments. Their covariance matrix is thus given by the GMM formula (11.5) for the sample variation of the g_T. All we have to do is map the problem into the GMM notation.

The parameter vector is

$$b' = \begin{bmatrix} a' & \beta' & \lambda \end{bmatrix}.$$

The a matrix chooses which moment conditions are set to zero in estimation,

$$a = \begin{bmatrix} I_{2N} & 0 \\ 0 & \gamma' \end{bmatrix}.$$

The d matrix is the sensitivity of the moment conditions to the parameters,

$$d = \frac{\partial g_T}{\partial b'} = \begin{bmatrix} -I_N & -I_N E(f) & 0 \\ -I_N E(f) & -I_N E(f^2) & 0 \\ 0 & -\lambda I_N & -\beta \end{bmatrix}.$$

The S matrix is the long-run covariance matrix of the moments,

$$S = \sum_{j=-\infty}^{\infty} E\left(\begin{bmatrix} R_t^e - a - \beta f_t \\ (R_t^e - a - \beta f_t)f_t \\ R_t^e - \beta\lambda \end{bmatrix} \begin{bmatrix} R_{t-j}^e - a - \beta f_{t-j} \\ (R_{t-j}^e - a - \beta f_{t-j})f_{t-j} \\ R_{t-j}^e - \beta\lambda \end{bmatrix}'\right)$$

$$= \sum_{j=-\infty}^{\infty} E\left(\begin{bmatrix} \varepsilon_t \\ \varepsilon_t f_t \\ \beta(f_t - Ef) + \varepsilon_t \end{bmatrix} \begin{bmatrix} \varepsilon_{t-j} \\ \varepsilon_{t-j}f_{t-j} \\ \beta(f_{t-j} - Ef) + \varepsilon_{t-j} \end{bmatrix}'\right).$$

In the second expression, I have used the regression model and the restriction under the null that $E(R_t^e) = \beta\lambda$. In calculations, of course, you could simply estimate the first expression.

We are done. We have the ingredients to calculate the GMM standard error formula (11.4) and formula for the covariance of moments (11.5).

With a vector f, the moments are

$$\begin{bmatrix} I_N \otimes I_{K+1} & \\ & \gamma' \end{bmatrix} \begin{bmatrix} E_T(R^e - a - \beta f) \\ E_T[(R^e - a - \beta f) \otimes f] \\ E_T(R^e - \beta\lambda) \end{bmatrix} = 0,$$

where $\beta_i = N \times 1$, and $\gamma' = \beta'$ for OLS and $\gamma' = \beta'(\Sigma^{-1})$ for GLS.

Note that the GLS estimate is *not* the "efficient GMM" estimate when returns are not i.i.d. The efficient GMM estimate is $d'S^{-1}g_T = 0$. The d matrix is

$$
d = \frac{\partial g_T}{\partial\left[\alpha'\ \beta'_1\ \beta'_2\ \lambda'\right]} = -\left[\begin{bmatrix} 1 & E(f') \\ E(f) & E(ff') \\ 0 & \lambda' \end{bmatrix} \otimes I_N \begin{bmatrix} 0 \\ 0 \\ \beta \end{bmatrix}\right].
$$

We can recover the classic formulas (12.19), (12.20), (12.21) by adding the assumption that the errors are i.i.d. and independent of the factors, and that the factors are uncorrelated over time as well. The assumption that the errors and factors are uncorrelated over time means we can ignore the lead and lag terms. Thus, the top left corner of S is $E(\varepsilon_t\varepsilon'_t) = \Sigma$. The assumption that the errors are independent from the factors f_t simplifies the terms in which ε_t and f_t are multiplied: $E(\varepsilon_t(\varepsilon'_t f_t)) = E(f)\Sigma$ for example. The result is

$$
S = \begin{bmatrix} \Sigma & E(f)\Sigma & \Sigma \\ E(f)\Sigma & E(f^2)\Sigma & E(f)\Sigma \\ \Sigma & E(f)\Sigma & \beta\beta'\sigma^2(f) + \Sigma \end{bmatrix}.
$$

Multiplying a, d, S together as specified by the GMM formula for the covariance matrix of parameters (11.4), we obtain the covariance matrix of all the parameters, and its $(3, 3)$ element gives the variance of $\hat{\lambda}$. Multiplying the terms together as specified by (11.5), we obtain the sampling distribution of the $\hat{\alpha}$, (12.20). The formulas (12.19) reported above are derived the same way with a vector of factors f_t rather than a scalar; the second-moment condition in (12.23) then reads $E\left[(R_t^e - a - \beta f_t) \otimes f_t\right]$. The matrix multiplication is not particularly enlightening.

Once again, there is really no need to make the assumption that the errors are i.i.d. and especially that they are conditionally homoskedastic—that the factor f and errors ε are independent. It is quite easy to estimate an S matrix that does not impose these conditions and calculate standard errors. They will not have the pretty analytic form given above, but they will more closely report the true sampling uncertainty of the estimate. Furthermore, if one is really interested in efficiency, the GLS cross-sectional estimate should use the spectral density matrix as weighting matrix applied to all the moments rather than Σ^{-1} applied only to the pricing errors.

Time Series vs. Cross Section

How are the time-series and cross-sectional approaches different?

Most importantly, you can run the cross-sectional regression when the factor is not a return. The time-series test requires factors that are also

returns, so that you can estimate factor risk premia by $\hat{\lambda} = E_T(f)$. The asset pricing model does predict a restriction on the intercepts in the time-series regression. Why not just test these? If you impose the restriction $E(R^{ei}) = \beta_i'\lambda$, you can write the time-series regression (12.9) as

$$R_t^{ei} = \beta_i'\lambda + \beta_i'\left(f_t - E(f)\right) + \varepsilon_t^i, \qquad t = 1, 2, \ldots, T \quad \text{for each } i.$$

Thus, the intercept restriction is

$$a_i = \beta_i'(\lambda - E(f)). \tag{12.24}$$

This restriction makes sense. The model says that mean returns should be proportional to betas, and the intercept in the time-series regression controls the mean return. You can also see how $\lambda = E(f)$ results in a zero intercept. Finally, however, you see that without an estimate of λ, you cannot check this intercept restriction. If the factor is not a return, you will be forced to do something like a cross-sectional regression.

When the factor is a return, so that we can compare the two methods, time-series and cross-sectional regressions are not necessarily the same. The time-series regression estimates the factor risk premium as the sample mean of the factor. Hence, the factor receives a zero pricing error in each sample. Also, the predicted zero-beta excess return is also zero. Thus, the time-series regression describes the cross section of expected returns by drawing a line as in Figure 12.1 that runs through the origin and through the factor, ignoring all of the other points. The OLS cross-sectional regression picks the slope and intercept, if you include one, to best fit all the points: to minimize the sum of squares of all the pricing errors.

If the factor is a return, *the GLS cross-sectional regression, including the factor as a test asset, is identical to the time-series regression.* The time-series regression for the factor is, of course,

$$f_t = 0 + 1f_t + 0,$$

so it has a zero intercept, beta equal to one, and zero residual in every sample. The residual variance-covariance matrix of the returns, including the factor, is

$$E\left(\begin{bmatrix} R^e - a - \beta f \\ f - 0 - 1f \end{bmatrix}[\cdot]'\right) = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}.$$

Since the factor has *zero* residual variance, a GLS regression puts all its weight on that asset. Therefore, $\hat{\lambda} = E_T(f)$ just as for the time-series regression. The pricing errors are the same, as is their distribution and the χ^2 test. (You gain a degree of freedom by adding the factor to the cross-sectional regression, so the test is a χ_N^2.)

Why does the "efficient" technique ignore the pricing errors of all of the other assets in estimating the factor risk premium, and focus only on the mean return? The answer is simple, though subtle. In the regression model

$$R_t^e = a + \beta f_t + \varepsilon_t,$$

the average return of each asset in a sample is equal to beta times the average return of the factor in the sample, plus the average residual in the sample. An average return carries *no* information about the mean of the factor that is not already observed in the sample mean of the factor. A signal plus noise carries no additional information beyond that in the *same* signal. Thus, an "efficient" cross-sectional regression wisely ignores all the information in the other asset returns and uses only the information in the factor return to estimate the factor risk premium.

––––––––––

12.3 Fama–MacBeth Procedure

> I introduce the Fama–MacBeth procedure for running cross-sectional regression and calculating standard errors that correct for cross-sectional correlation in a panel. I show that, when the right-hand variables do not vary over time, Fama–MacBeth is numerically equivalent to pooled time-series, cross-section OLS with standard errors corrected for cross-sectional correlation, and also to a single cross-sectional regression on time-series averages with standard errors corrected for cross-sectional correlation. Fama–MacBeth standard errors do not include corrections for the fact that the betas are also estimated.

Fama and MacBeth (1973) suggest an alternative procedure for running cross-sectional regressions, and for producing standard errors and test statistics. This is a historically important procedure, it is computationally simple to implement, and is still widely used, so it is important to understand it and relate it to other procedures.

First, you find beta estimates with a time-series regression. Fama and MacBeth use rolling 5-year regressions, but one can also use the technique with full-sample betas, and I will consider that simpler case. Second, instead of estimating a single cross-sectional regression with the sample averages, we now run a cross-sectional regression *at each time period*, i.e.,

$$R_t^{ei} = \beta_i' \lambda_t + \alpha_{it}, \qquad i = 1, 2, \ldots, N \quad \text{for each } t.$$

I write the case of a single factor for simplicity, but it is easy to extend the model to multiple factors. Then, Fama and MacBeth suggest that

we estimate λ and α_i as the average of the cross-sectional regression estimates,

$$\hat{\lambda} = \frac{1}{T} \sum_{t=1}^{T} \hat{\lambda}_t, \qquad \hat{\alpha}_i = \frac{1}{T} \sum_{t=1}^{T} \hat{\alpha}_{it}.$$

Most importantly, they suggest that we use the standard deviations of the cross-sectional regression estimates to generate the sampling errors for these estimates,

$$\sigma^2(\hat{\lambda}) = \frac{1}{T^2} \sum_{t=1}^{T} \left(\hat{\lambda}_t - \hat{\lambda} \right)^2, \qquad \sigma^2(\hat{\alpha}_i) = \frac{1}{T^2} \sum_{t=1}^{T} \left(\hat{\alpha}_{it} - \hat{\alpha}_i \right)^2.$$

It is $1/T^2$ because we are finding standard errors of sample means, σ^2/T.

This is an intuitively appealing procedure once you stop to think about it. Sampling error is, after all, about how a statistic would vary from one sample to the next if we repeated the observations. We cannot do that with only one sample, but why not cut the sample in half, and deduce how a statistic would vary from one full sample to the next from how it varies from the first half of the sample to the next half? Proceeding, why not cut the sample in fourths, eighths, and so on? The Fama–MacBeth procedure carries this idea to its logical conclusion, using the variation in the statistic $\hat{\lambda}_t$ over time to deduce its variation across samples.

We are used to deducing the sampling variance of the sample mean of a series x_t by looking at the variation of x_t through time in the sample, using $\sigma^2(\bar{x}) = \sigma^2(x)/T = \frac{1}{T^2} \sum_t (x_t - \bar{x})^2$. The Fama–MacBeth technique just applies this idea to the slope and pricing error estimates. The formula assumes that the time series is not autocorrelated, but one could easily extend the idea to estimates $\hat{\lambda}_t$ that are correlated over time by using a long-run variance matrix, i.e., estimate

$$\sigma^2(\hat{\lambda}) = \frac{1}{T} \sum_{j=-\infty}^{\infty} \text{cov}_T(\hat{\lambda}_t, \hat{\lambda}_{t-j}).$$

One should of course use some sort of weighting matrix or a parametric description of the autocorrelations of $\hat{\lambda}$, as explained in Section 11.7. Asset return data are usually not highly correlated, but accounting for such correlation could have a big effect on the application of the Fama–MacBeth technique to corporate finance data or other regressions in which the cross-sectional estimates are highly correlated over time.

It is natural to use this sampling theory to test whether all the pricing errors are jointly zero as we have before. Denote by α the vector of

pricing errors across assets. We could estimate the covariance matrix of the sample pricing errors by

$$\hat{\alpha} = \frac{1}{T} \sum_{t=1}^{T} \hat{\alpha}_t,$$

$$\text{cov}(\hat{\alpha}) = \frac{1}{T^2} \sum_{t=1}^{T} \left(\hat{\alpha}_t - \hat{\alpha} \right) \left(\hat{\alpha}_t - \hat{\alpha} \right)',$$

(or a general version that accounts for correlation over time) and then use the test

$$\hat{\alpha}' \, \text{cov}(\hat{\alpha})^{-1} \hat{\alpha} \sim \chi^2_{N-1}.$$

Fama–MacBeth in Depth

The GRS procedure and the analysis of a single cross-sectional regression are familiar from any course in regression. We will see them justified by maximum likelihood below. The Fama–MacBeth procedure seems unlike anything you have seen in any econometrics course, and it is obviously a useful and simple technique that can be widely used in panel data in economics and corporate finance as well as asset pricing. Is it truly different? Is there something different about asset pricing data that requires a fundamentally new technique not taught in standard regression courses? Or is it similar to standard techniques? To answer these questions it is worth looking in a little more detail at what it accomplishes and why.

It is easier to do this in a more standard setup, with left-hand variable y and right-hand variable x. Consider a regression

$$y_{it} = \beta' x_{it} + \varepsilon_{it}, \qquad i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T.$$

The data in this regression have a cross-sectional element as well as a time-series element. In corporate finance, for example, you might be interested in the relationship between investment and financial variables, and the data set has many firms (N) as well as time-series observations for each firm (T). In an expected return-beta asset pricing model, the x_{it} stands for the β_i and β stands for λ.

An obvious thing to do in this context is simply to stack the i and t observations together and estimate β by OLS. I will call this the *pooled time-series cross-section estimate*. However, the error terms are not likely to be uncorrelated with each other. In particular, the error terms are likely to be cross-sectionally correlated at a given time. If one stock's return is unusually high this month, another stock's return is also likely to be high;

if one firm invests an unusually great amount this year, another firm is also likely to do so. When errors are correlated, OLS is still consistent, but the OLS distribution theory is wrong, and typically suggests standard errors that are much too small. In the extreme case that the N errors are perfectly correlated at each time period, there is really only one observation for each time period, so one really has T rather than NT observations. Therefore, a pooled time-series cross-section estimate must include corrected standard errors. People often ignore this fact and report OLS standard errors.

Another thing we could do is first take time-series averages and then run a *pure cross-sectional* regression of

$$E_T(y_{it}) = \beta' E_T(x_{it}) + u_i, \qquad i = 1, 2, \ldots, N.$$

This procedure would lose any information due to variation of the x_{it} over time, but at least it might be easier to figure out a variance-covariance matrix for u_i and correct the standard errors for residual correlation. (You could also average cross-sectionally and then run a single time-series regression. We will get to that option later.) In either case, the standard error corrections are just applications of the standard formula for OLS regressions with correlated error terms.

Finally, we could run the Fama–MacBeth procedure: run a cross-sectional regression at each point in time, average the cross-sectional $\hat{\beta}_t$ estimates to get an estimate $\hat{\beta}$, and use the time-series standard deviation of $\hat{\beta}_t$ to estimate the standard error of $\hat{\beta}$.

It turns out that the Fama–MacBeth procedure is another way of calculating the standard errors, corrected for cross-sectional correlation.

Proposition: *If the x_{it} variables do not vary over time, and if the errors are cross-sectionally correlated but not correlated over time, then the Fama–MacBeth estimate, the pure cross-sectional OLS estimate, and the pooled time-series cross-sectional OLS estimates are identical. Also, the Fama–MacBeth standard errors are identical to the cross-sectional regression or stacked OLS standard errors, corrected for residual correlation. None of these relations hold if the x_{it} vary through time.*

Since they are identical procedures, whether one calculates estimates and standard errors in one way or the other is a matter of taste.

I emphasize one procedure that is incorrect: pooled time-series and cross-section OLS with no correction of the standard errors. The errors are so highly cross-sectionally correlated in most finance applications that the standard errors so computed are often off by a factor of 10.

The assumption that the errors are not correlated over time is probably not so bad for asset pricing applications, since returns are close to independent. However, when pooled time-series cross-section regressions are used in corporate finance applications, errors are likely to be as severely correlated over time as across firms, if not more so. The "other factors" (ε) that

cause, say, company i to invest more at time t than predicted by a set of right-hand variables is surely correlated with the other factors that cause company j to invest more. But such factors are especially likely to cause company i to invest more at time $t + 1$ as well. In this case, any standard errors must also correct for serial correlation in the errors; the GMM-based formulas in Section 11.4 can do this easily.

The Fama–MacBeth standard errors also do not correct for the fact that $\hat{\beta}$ are generated regressors. If one is going to use them, it is a good idea to at least calculate the Shanken correction factors outlined above, and check that the corrections are not large.

Proof: We just have to write out the three approaches and compare them. Having assumed that the x variables do not vary over time, the regression is

$$y_{it} = x_i'\beta + \varepsilon_{it}.$$

We can stack up the cross sections $i = 1, \ldots, N$ and write the regression as

$$y_t = x\beta + \varepsilon_t.$$

x is now a matrix with the x_i' as rows. The error assumptions mean $E(\varepsilon_t\varepsilon_t') = \Sigma$.

Pooled OLS: To run pooled OLS, we stack the time series and cross sections by writing

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}, \qquad X = \begin{bmatrix} x \\ x \\ \vdots \\ x \end{bmatrix}, \qquad \epsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix}$$

and then

$$Y = X\beta + \epsilon,$$

with

$$E(\epsilon\epsilon') = \Omega = \begin{bmatrix} \Sigma & & \\ & \ddots & \\ & & \Sigma \end{bmatrix}.$$

The estimate and its standard error are then

$$\hat{\beta}_{\text{OLS}} = (X'X)^{-1}X'Y,$$
$$\text{cov}(\hat{\beta}_{\text{OLS}}) = (X'X)^{-1}X'\Omega X(X'X)^{-1}.$$

Writing this out from the definitions of the stacked matrices, with $X'X = Tx'x$,

$$\hat{\beta}_{\mathrm{OLS}} = (x'x)^{-1}x'E_T(y_t),$$

$$\mathrm{cov}(\hat{\beta}_{\mathrm{OLS}}) = \frac{1}{T}(x'x)^{-1}(x'\Sigma x)(x'x)^{-1}.$$

We can estimate this sampling variance with

$$\widehat{\Sigma} = E_T(\hat{\varepsilon}_t\hat{\varepsilon}_t'), \qquad \hat{\varepsilon}_t \equiv y_t - x\hat{\beta}_{\mathrm{OLS}}.$$

Pure cross-section: The pure cross-sectional estimator runs one cross-sectional regression of the time-series averages. So, take those averages,

$$E_T\left(y_t\right) = x\beta + E_T\left(\varepsilon_t\right),$$

where $x = E_T(x)$ since x is constant. Having assumed i.i.d. errors over time, the error covariance matrix is

$$E\left[E_T\left(\varepsilon_t\right)E_T\left(\varepsilon_t'\right)\right] = \frac{1}{T}\Sigma.$$

The cross-sectional estimate and corrected standard errors are then

$$\hat{\beta}_{\mathrm{XS}} = (x'x)^{-1}x'E_T(y_t),$$

$$\sigma^2(\hat{\beta}_{\mathrm{XS}}) = \frac{1}{T}(x'x)^{-1}x'\Sigma x(x'x)^{-1}.$$

Thus, the cross-sectional and pooled OLS estimates and standard errors are exactly the same, in each sample.

Fama–MacBeth: The Fama–MacBeth estimator is formed by first running the cross-sectional regression at each moment in time,

$$\hat{\beta}_t = \left(x'x\right)^{-1}x'y_t.$$

Then the estimate is the average of the cross-sectional regression estimates,

$$\hat{\beta}_{FM} = E_T\left(\hat{\beta}_t\right) = \left(x'x\right)^{-1}x'E_T\left(y_t\right).$$

Thus, the Fama–MacBeth estimator is also the same as the OLS estimator, in each sample. The Fama–MacBeth standard error is based on the time-series standard deviation of the $\hat{\beta}_t$. Using cov_T to denote sample covariance,

$$\mathrm{cov}\left(\hat{\beta}_{FM}\right) = \frac{1}{T}\mathrm{cov}_T\left(\hat{\beta}_t\right) = \frac{1}{T}(x'x)^{-1}x'\,\mathrm{cov}_T(y_t)x(x'x)^{-1},$$