Empirical Methods in Finance

A Note on Linear Factor Models:

Time series and cross-sectional tests

Lecturer: Lars A. Lochstoer*
UCLA Anderson School of Management

January 28, 2017

*Let me know of any errors, please. Contact email: lars.lochstoer@anderson.ucla.edu

0.1 Linear Factor Models

In this note, we discuss testing and intepretations of linear factor models. We start with the simplest case of a single factor and then consider multifactor models. The notation follows that of the slides.

0.2 Time-Series test of single factor model

We are testing whether $\alpha_i = 0$ in

$$R^e_{i,t+1} = \alpha_i + \beta_i F_{t+1} + \varepsilon_{i,t+1} \tag{1}$$

In other words, we are testing whether the pricing errors are simply a product of "normal" sample variation or in fact a result of a mis-specified model. From standard OLS standard errors (try writing it out yourself to get the first equality below):

$$Est.Var\left(\hat{\alpha}_{iT}\right) = \frac{\hat{s}^2_{iT}}{T \times Var_T\left(F_t\right)}\overline{F^2_t} = \left(1 + \frac{\overline{F^2_t} - Var_T\left(F_t\right)}{Var_T\left(F_t\right)}\right)\frac{\hat{s}^2_{iT}}{T} = \left(1 + \frac{\overline{F_t}^2}{Var_T\left(F_t\right)}\right)\frac{\hat{s}^2_{iT}}{T} \tag{2}$$

Let's think about this expression a bit. $\hat{\alpha}$ is the estimate of the sample mean of the residuals. The variance of a sample mean is $\frac{\hat{s}^2_{iT}}{T}$. Why is the term $\frac{\overline{F_t}^2}{Var_T(F_t)}$ present? The reason is that the estimation of β introduces additional uncertainty in the sample average estimation - we have to *estimate* the residuals. This uncertainty is less if there is a stronger signal (high variance of the factor), but higher if the mean of the factor is high (in this case, the factor mean has a greater impact on the sample mean of the residuals).

What is an appropriate test for the null hypothesis $\alpha_i = 0$? From the regression theory

$$\frac{\hat{\alpha}^2_{iT}}{Est.Var\left(\hat{\alpha}_{iT}\right)} \sim F\left(1, T-2\right) \tag{3}$$

$$\frac{T\hat{\alpha}^2_{iT}}{\left(1 + \frac{\overline{F_t}^2}{Var_T(F_t)}\right)\hat{s}^2_{iT}} \sim F\left(1, T-2\right) \tag{4}$$

Our factor is the excess return to a portfolio. Thus, $\hat{\theta}^2_p \equiv \frac{\overline{F_t}^2}{Var_T(F_t)}$ is the sample squared Sharpe ratio of this portfolio. Rewrite as

$$\frac{T}{\left(1 + \hat{\theta}^2_p\right)}\frac{\hat{\alpha}^2_{iT}}{\hat{s}^2_{iT}} \sim F\left(1, T-2\right) \tag{5}$$

If we know the volatility of the residuals, this statistic is distributed chi-squared:

$$\frac{T}{\left(1+\hat{\theta}_p^2\right)}\frac{\hat{\alpha}_{iT}^2}{\sigma_\varepsilon^2} \sim \chi^2\left(1\right) \tag{6}$$

Thus, the higher the sample Sharpe ratio of the factor portfolio (which under the null is mean-variance efficient, i.e. a maximal Sharpe ratio portfolio), the harder it is to reject the null for a given pricing error.

0.2.1 Joint test of multiple assets

We could run the test for each asset, but statistically it makes more sense to run a joint test using all our test portfolios. Let's say we observe one portfolio with a high alpha and one with a low alpha. If the estimated alpha's of these portfolios are negatively correlated, should we treat these observations as independent? Should we put more or less weight on this fact in our test statistic? We should put less weight on these observations. Given that one is positive (say, because of pure sample error), we know that the other is likely to be negative. In other words, we need to take into account the covariances of the α estimates when we perform a joint test of the model.

Here is how to do it. Run N separate regressions for each asset i. Get the sample vector of residuals ε_{iT}. Calculate the sample covariance matrix of the residuals $\hat{\Sigma}$. The covariance matrix of the $\alpha's$ is proportional to the covariance matrix of the sample means of the residuals. Thus, the covariance matrix of the α's is propotional to $\hat{\Sigma}/T$. In particular, if we know the residual covariance matrix, the multivariate test is

$$\frac{T}{\left(1+\hat{\theta}_p^2\right)}\hat{\alpha}_T'\Sigma^{-1}\hat{\alpha}_T \sim \chi^2\left(N\right) \tag{7}$$

Gibbons, Ross, and Shanken (1987) show us that when we have to estimate the covariance matrix, the joint test of the alphas equal to zero takes the form

$$\frac{T-N-1}{N}\frac{1}{1+\hat{\theta}_p^2}\hat{\alpha}_T'\hat{\Sigma}^{-1}\hat{\alpha}_T \sim F\left(N, T-N-1\right) \tag{8}$$

Here $\hat{\Sigma}$ is the estimated covariance matrix of the residuals without adjusting for the degrees of freedom. Estimating the covariance matrix is hard when there are many assets N relative to the number of observations T. The noise introduced in the estimation of $\hat{\Sigma}$ makes it harder to reject the model. Since the residuals are estimated by minimizing their variance, in-sample spurious correlation makes them too small on average, and that's why the small sample adjustment comes about. As $T \to \infty$, the statistic converges to the chi-squared version stated for the case with known covariance matrix. More on this in a couple of slides.

0.3 Interpreting the test as a test of mean-variance efficiency

We've established that the factor portfolio under the null hypothesis is mean-variance efficient. An intuitive test of the model is then to test whether a measure of the distance between the sample Sharpe ratio of the factor portfolio and the maximal sample Sharpe ratio attainable, using all the test assets, is "too big". Intuitively, if the factor portfolio is "too far" from being in-sample mean-variance efficient, we can reject the model. This is exactly the interpretation of the GRS test-statistic.

Consider the problem of finding the maximal Sharpe ratio portfolio:

$$\min_w w'\hat{V}w \quad s.t. \;\; w'\overline{R^e} = m \tag{9}$$

where w is a vector of portfolio weights and $\bar{R}^e$ is a vector of expected *excess* returns. The first order conditions yield $w = k\hat{V}^{-1}\overline{R^e}$ and $w'\overline{R^e} = m$, where the Lagrange multiplier is $2k$. The variance is $k^2\overline{R^e}'\hat{V}^{-1}\overline{R^e}$, and the expected excess return is $k\overline{R^e}'\hat{V}^{-1}\overline{R^e}$. The maximal squared Sharpe ratio is then

$$\max SR^2 \equiv (\theta^*)^2 = \overline{R^e}'\hat{V}^{-1}\overline{R^e} \tag{10}$$

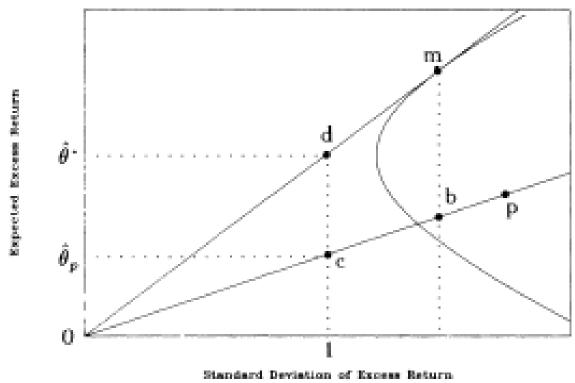
Next, consider the set of zero-investment strategies that go long one unit of each asset i, financed by borrowing, and short β_i units of the factor portfolio, also financed by borrowing. For intuition, think of these portfolios as a mutual fund managers active deviations from holding the market. The maximal Sharpe ratio of these "active portfolios" is then analogously $\hat{\alpha}_T'\hat{\Sigma}^{-1}\hat{\alpha}_T$. Since the factor portfolio and the active portfolios are uncorrelated (from the identifying restriction of regression), the squared maximal Sharpe ratio in the economy is simply the sum of the Sharpe ratio of the factor portfolio and the maximal Sharpe ratio of the active portfolios:

$$(\theta^*)^2 = \theta_p^2 + \hat{\alpha}_T'\hat{\Sigma}^{-1}\hat{\alpha}_T \tag{11}$$

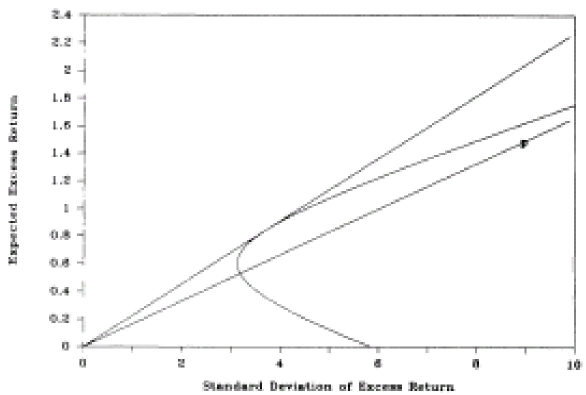
$$\frac{1+(\theta^*)^2}{1+\theta_p^2} - 1 = \frac{\hat{\alpha}_T'\hat{\Sigma}^{-1}\hat{\alpha}_T}{1+\theta_p^2} \tag{12}$$

The right hand side is the test statistic of GRS (before scaling by T), and the left hand side is the ratio of the maximal in-sample Sharpe ratio and the in-sample Sharpe ratio of the factor portfolio. This ratio cannot be below 1, obviously. Thus, the test-statistic is in fact a measure of the distance between the Sharpe ratio of the factor portfolio and the maximal Sharpe ratio portfolio. If this distance is too big, we reject that the factor portfolio is mean-variance efficient. Gibbons, Ross, and Shanken illustrate this graphically:

PORTFOLIO EFFICIENCY 1127



1a.) Geometric intuition for W. Note the distance Oc is $\sqrt{1+\hat{\theta}_p^2}$, and the distance Od is $\sqrt{1+\hat{\theta}^{*2}}$.



1b.) *Ex post* efficient frontier based on 10 beta-sorted portfolios and the CRSP Equal-Weighted Index using monthly data, 1931-1965. Point p represents the CRSP Equal-Weighted Index.

FIGURE 1.—Various plots of ex post mean variance efficient frontiers.

Why is GRS an Important Paper?

- *not* just because it derived statistical test of CAPM using multiple assets.
  - That's an advanced econometric class assignment.
- because it gave the profession a good reason to use the joint test . . .
  - . . . and because it told us how to think about the test in an economically intuitive way.
  - Richard Roll: Only testable hypothesis of CAPM is that market portfolio is M-V efficient.
  - Geometry of GRS test shows this.

- Long-run influence is that it focused attention on maximally mispriced portfolio.
  - Once we know that statistical tests effectively use this portfolio, we can debate whether this is good or not.

Results from GRS test:

- Using the GRS test statistic, there is no (longer) statistically significant evidence against the CAPM using the 10 size-sorted portfolios over the 1926–2006 period.

1 Interrim Summary

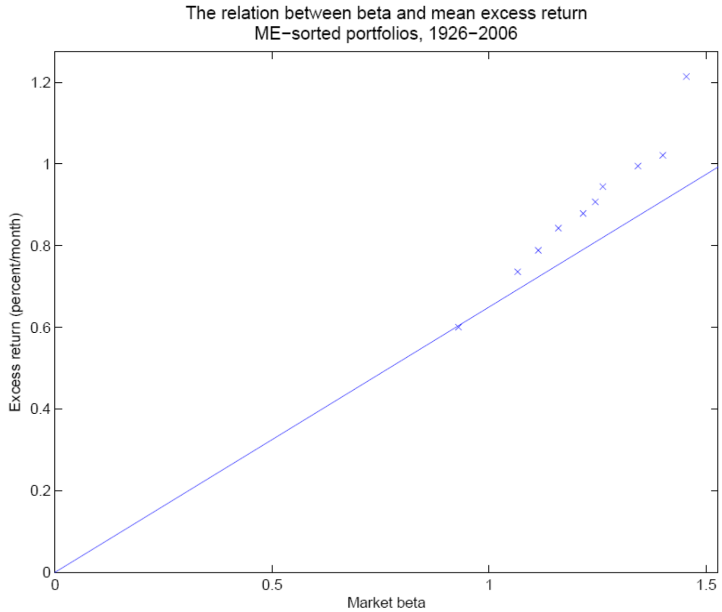
Up until now we have considered time-series tests of linear factor models and the GRS test.

For this test, we assumed:

1. Factor portfolio(s) are known portfolios of traded assets (e.g., the market portfolio).
2. The risk-free rate is observed.
3. Constant conditional betas, so our unconditional time-series tests are correctly specified.

The test was a test based on model fit: if the intercept (alpha) of a set of test assets are jointly significantly different from zero, we reject the model.

- However, we are often interested in linear factor models where assumptions (1) and (2) above do not hold.
  - How do we proceed in these cases?
  - What about testing alternative hypotheses?
- Cross-sectional regressions resolve these issues:
  - Consider the figure:



- If the model is right, the points should fall on a line from the origin
- This is a cross-sectional regression of mean returns on betas!

2 Cross-sectional Asset Pricing: Cross-sectional Tests

If the factor is not a portfolio of traded assets, we cannot apply the time-series tests. The reason is that it is no longer the case that the null hypothesis implies that α in the time-series regression is zero. To see this, consider the following one factor model

$$E_t\left[R^e_{it+1}\right] = \beta_{it}\lambda, \qquad \beta_i = \frac{Cov\left(F_{t+1}, R^e_{i,t+1}\right)}{Var\left(F_{t+1}\right)} \tag{13}$$

Here, F_{t+1} is the factor, e.g. aggregate consumption growth, GDP growth. λ_t is the conditional risk premium (the price of risk) associated with "1 risk unit" exposure to the factor. Let's consider the time-series regression

$$R^e_{i,t+1} = \alpha_i + \beta_i F_{t+1} + \varepsilon_{i,t+1} \tag{14}$$

This gives

$$\begin{aligned} E\left[R^e_{i,t+1}\right] &= \beta_i\lambda = \alpha_i + \beta_i E\left[F_{t+1}\right] \\ \alpha_i &= \beta_i\left(\lambda - E\left[F_{t+1}\right]\right) \end{aligned} \tag{15}$$

In the case of the excess return of a traded asset $\lambda - E[F_{t+1}] = 0$. Check this yourself by noting that the β-coefficient now trivially is equal to 1. However, this will not be the case in general for a non-traded factors like consumption or GDP growth. In this case, we need to estimate the factor price of risk λ. This is done in cross-sectional tests. There are two main test procedures: Two pass regression and Fama-MacBeth. These procedures can of course also be used when the factor(s) are traded assets.

2.1 Two Pass Regression

Assume F_t has a constant conditional mean. That means that the innovations are uncorrelated over time.

1. First Pass:

Estimate the time-series regression

$$R^e_{i,t+1} = \hat{a}_i + \hat{\beta}_i F_{t+1} + e_{i,t+1} \tag{16}$$

for each asset i. Save the $\hat{\beta}_i$'s, the average excess returns $\overline{R}^e_i$ and the residuals $\{e_{i,t}\}_{t=1}^T$.

2. Second Pass:

Estimate the price(s) of risk (λ) in the cross-sectional regression:

$$E_T[R^e_i] = \lambda\hat{\beta}_i + \alpha_i \tag{17}$$

using the N sample average asset returns and the estimated $\beta_i's$, where $E_T[\cdot]$ denotes the sample average. Note that the regression has no intercept, which is what the null hypothesis prescribes. The residuals from the cross-sectional regression (α_i) are the pricing errors of each asset. We can then test, as in the time series case, the null hypothesis that these are jointly zero. We can also test whether the factor is "priced", i.e. $\hat{\lambda} \neq 0$.

Sometimes, researchers run the second pass regression including an intercept and test if the intercept is different from zero. This is fine - it is certainly an implication of the model! - but not as strong as testing whether *all* the pricing errors are zero. A different approach is test whether all $\alpha's$ are zero, but to include an intercept in the regression, which means we allow for average mispricing, in the name of robustness and economic significance. This may be more robust, since all our models are misspecified. The question is then whether the model can capture the *variation* in excess returns. For instance, we may be interested in whether aggregate consumption-betas can explain the variation in average excess risky returns, even if we cannot simultaneously price the risk-free asset (i.e., if the intercept is significantly different from zero). In that case, researchers have tended to focus mainly on whether the cross-sectional R^2 of the second pass regression is high. At times, they do not even report the test of whether this intercept is zero.

2.1.1 Test statistics

The test statistics depend as usual on what we assume about the properties about the residuals, e_i. Here, we will assume that the errors and factors are i.i.d. over time. Thus, we allow for contemporaneous cross-correlation among stock returns, which from an empirical standpoint it is vital to do, but no cross-correlations over time nor any heteroskedasticity. It is relatively straightforward to allow for more complicated residual dynamics (see Cochrane, Ch 12).

Remember from standard OLS where $Y = X\beta + \varepsilon$ and $E[\varepsilon\varepsilon'] = \Omega$, that $var\left(\hat{\beta}\right) = (X'X)^{-1} X'\Omega X (X'X)^{-1}$ and that $var(\hat{\varepsilon}) = \left(I - X(X'X)^{-1} X\right) \Omega \left(I - X(X'X)^{-1} X\right)'$. Here Ω is a general (positive definite) variance-covariance matrix, so as to allow for correlated errors across assets. This is important when running cross-sectional regressions as averages of realized asset returns indeed tend to be correlated across assets.

Now, from the first pass regression, assuming the errors are i.i.d. over time (i.e., no autocorrelation or cross-correlation over time – only contemporaneous correlation), compute the variance covariance matrix of the residuals:

$$\Sigma = \frac{1}{T} \sum_{t=1}^{T} e_t e_t'. \tag{18}$$

Further, denote the variance-covariance matrix of the factors, which also are assumed to be i.i.d. over time, in the first pass regression as Σ_f. Finally, for now assume that the β's from the first pass regression are estimated without any error.

In this case, the true variance-covariance matrix of the α's in the second pass regressions (these are the residuals in the second pass regression and so their covariance matrix corresponds to Ω in the above OLS notation) is $\frac{1}{T}\left(\beta\Sigma_f\beta' + \Sigma\right)$. To see this, start with

$$\alpha = E_T[R^e] - \beta\lambda, \tag{19}$$

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

$$E_T[R_t^e] = a + \beta E_T[f_t] + E_T[\varepsilon_t]. \tag{20}$$

Under the null that the model is correct (we always develop our standard test statistics under the null), we have that:

$$E[R^e] = a + \beta E[f] = \beta\lambda. \tag{21}$$

Then, the variance-covariance matrix of the true α's is,

$$\begin{aligned}
cov(\alpha,\alpha) &= cov\left(E_T\left[R^e\right]-\beta\lambda, E_T\left[R^e\right]-\beta\lambda\right) \\
&= cov\left(a+\beta E_T\left[f_t\right]+E_T\left[\varepsilon_t\right]-\beta\lambda, a+\beta E_T\left[f_t\right]+E_T\left[\varepsilon_t\right]-\beta\lambda\right) && (22)\\
&= cov\left(\beta E_T\left[f_t\right],\beta E_T\left[f_t\right]\right)+cov\left(E_T\left[\varepsilon_t\right],E_T\left[\varepsilon_t\right]\right) && (23)\\
&= \frac{1}{T}\left(\beta\Sigma_f\beta'+\Sigma\right). && (24)
\end{aligned}$$

(remember, the factors and the errors are uncorrelated under the null).

Recall, the α's are the residuals of the cross-sectional regression. Thus, the covariance matrix of the errors and the prices of risk are then (simply using the above standard OLS formulas):

$$\begin{aligned}
cov\left(\hat{\lambda}\right) &= \left(\beta'\beta\right)^{-1}\beta' cov(\alpha,\alpha)\,\beta\left(\beta'\beta\right)^{-1} && (25)\\
&= \left(\beta'\beta\right)^{-1}\beta'\left(\frac{1}{T}\left(\beta\Sigma_f\beta'+\Sigma\right)\right)\beta\left(\beta'\beta\right)^{-1} && (26)\\
&= \frac{1}{T}\left(\beta'\beta\right)^{-1}\beta'\beta\Sigma_f\beta'\beta\left(\beta'\beta\right)^{-1}+\frac{1}{T}\left(\beta'\beta\right)^{-1}\beta'\Sigma\beta\left(\beta'\beta\right)^{-1} && (27)\\
&= \frac{1}{T}\left[\Sigma_f+\left(\beta'\beta\right)^{-1}\beta'\Sigma\beta\left(\beta'\beta\right)^{-1}\right], && (28)\\
cov(\hat{\alpha}) &= \frac{1}{T}\left(I_N-\beta\left(\beta'\beta\right)^{-1}\beta'\right)\Sigma\left(I_N-\beta\left(\beta'\beta\right)^{-1}\beta'\right)', && (29)
\end{aligned}$$

where I_N is the $N\times N$ identity matrix. We can now test whether the pricing errors (the α's)are all zero. Note that we needed the input from the first-stage regression of Σ to be able to apply this test. We cannot usually test whether the residuals are zero as we estimate the residuals' variance-covariance matrix from the residuals themselves... Here, however, we have extra information from the first pass regression. The test is:

$$\hat{\alpha}' cov(\hat{\alpha})^{-1}\hat{\alpha}\sim\chi^2_{N-K}, \tag{30}$$

where N is the number if test assets (portfolios) and K is the number of factors. Since we have to estimate the λ's along the way, the covariance matrix is singular (which is why the degrees of freedom has a -K term) and we have to use a generalized inverse.

Shanken (1992) worked out how to account for the generated regressors in the second stage regression (the $\hat{\beta}$'s), and the formulas given this correction are (here, hats for estimated quantities are suppressed in the formulas):

$$\begin{aligned}
cov\left(\hat{\lambda}\right) &= \frac{1}{T}\left[\left(\beta'\beta\right)^{-1}\beta'\Sigma\beta\left(\beta'\beta\right)^{-1}\times\left(1+\lambda'\Sigma_f^{-1}\lambda\right)+\Sigma_f\right], && (31)\\
cov(\hat{\alpha}) &= \frac{1}{T}\left(I_N-\beta\left(\beta'\beta\right)^{-1}\beta'\right)\Sigma\left(I_N-\beta\left(\beta'\beta\right)^{-1}\beta'\right)'\times\left(1+\lambda'\Sigma_f^{-1}\lambda\right). && (32)
\end{aligned}$$

The astute reader may have noticed that since we have a measure of the covariance matrix of the residuals from the first stage regression, we can run GLS regressions in the second pass. This is correct, of course, and more efficient, in principle. However, as with all GLS regressions, there may be a sacrifice in terms of robustness given that the residual covariance matrix is still estimated with error. Most studies only apply the OLS test. The formulas for the GLS case are given in Cochrane (2004; Ch. 12).

2.1.2 Discussion of Two Pass Regression

The two pass regression can be applied also when the factor is a (portfolio of) traded asset(s). In that case, the time-series tests and the two pass regression tests are not in general the same. The time-series tests forces zero pricing error on the zero-beta and the factor portfolio. The second pass regression simply finds the best fit, allowing pricing error in both the factor itself and the zero-beta portfolio (if we include an intercept). If we run the second stage as a GLS regression and include the factor itself as one of the test assets, the time-series and two-pass tests are identical.

The two pass regression has two advantages over the time series test, in addition to allow for non-asset factors.

1. It allows for tests of the null against specific alternatives

    (a) By adding a constant and additional factors in the cross-sectional regression (such as idiosyncratic volatility from market regression), we can test the null against the alternative that the additional factors explain returns. You simply test whether the price of risk of the additional factors is zero. See Cochrane (2004) Ch. 13.4. Size and book-to-market are other characteristics that may be of interest.

        This point is important enough to emphasize.

        - In the second pass regression (across portfolios i), include J characteristics you want to test for. Put these into the vector X_i.
        - Let the first element of X_i be 1, so you allow for an intercept in the regression:

        $$R_i^e = \hat{\lambda}' \begin{pmatrix} X_i \\ \hat{\beta}_i \end{pmatrix} + \hat{\alpha}_i \tag{33}$$

        - Next, you need to adjust the formulas for the distribution of $\hat{\lambda}$: (the notation here becomes a bit messy)

            i. In the formulas, replace $\hat{\beta}$ with

            $$\begin{pmatrix} \hat{X} \\ \hat{B} \end{pmatrix} \tag{34}$$

where $\hat{B}$ is the estimates of beta from the first pass. Thus, if you have K pricing factors and L additional characteristics, the new, augmented $\hat{\beta}$ used in the equation for calculating $cov(\lambda)$ is an $N \times (K+L)$ matrix.

ii. Where $\hat{\Sigma}_F$ is added to covariance matrix, augment this matrix with first J rows and columns of zeros (no factor uncertainty affects these coefficients)

iii. Where $\hat{\Sigma}_F$ enters multiplicatively, drop the first J elements of $\hat{\lambda}$.

- The test that $\hat{\alpha}'s$ are jointly zero is a test that chosen regressors pick up all cross-sectional variation in expected excess returns (not a test of the overall model - instead, a specification test).

2. It allows for testing when no risk free rate exists or it is unobserved (zero beta approach)

   The intuition here is to let the return on assets (or portfolios of assets) with a zero beta be the equivalent of the risk free rate.

$$
\begin{aligned}
E\left[R_{i,t}^{e}\right] &= \beta_i \lambda &(35)\\
E\left[R_{i,t}\right] &= E\left[R_{zerobeta}\right] + \beta_i \lambda \\
&= \lambda_0 + \beta_i \lambda &(36)
\end{aligned}
$$

   Run first pass regression with raw (not excess) returns. Then run the cross-sectional regression using mean raw returns and add an intercept in the cross-sectional regression. The estimate of the intercept is the estimated zero-beta rate.

3. The disadvantage is that it throws away information in mean excess returns to the market.

2.2 Fama-MacBeth

- The Fama-MacBeth procedure was developed before the two pass regression test.

- It is easier and allows for time-varying betas. However, it does not correct (as one should) for the generated regressor problem, but otherwise it gets the OLS two-pass regression statistics about right.

- The Fama-MacBeth procedure is also used in corporate finance applications.

2.2.1 Constant Betas

Assume K factors. The first pass is the same as before:

$$R_{it}^{e} = \hat{\alpha}_i + \hat{\beta}_i' F_t + e_{it} \tag{37}$$

Save the $N \times K$ vector of β's and the residuals. Next estimate T cross-sectional regressions (the constant term is not needed).

$$R^e_{it} = \lambda_{0t} + \hat{\beta}_i \lambda_{1t} + \alpha_{it}, \qquad \hat{\lambda}_t = \left(\hat{\lambda}_{0t} \ \hat{\lambda}_{1t} \right)' \tag{38}$$

Now, you have a time-series of parameter estimates. Fama-MacBeth propose a very intuitive method for calculating the variance-covariance matrix: they simply calculate the variance of the estimates and uses the sample mean estimate as the estimate of the prices of risk.

$$\hat{\lambda} = \frac{1}{T} \sum_{t=1}^{T} \hat{\lambda}_t \tag{39}$$

The variance-covariance of the sample mean is then $\frac{1}{T}$ times the sample variance of $\hat{\lambda}_t$:

$$Est.Asy.Var\left(\hat{\lambda}\right) = \frac{1}{T^2} \sum_{t=1}^{T} \left(\hat{\lambda}_t - \hat{\lambda}\right) \left(\hat{\lambda}_t - \hat{\lambda}\right)' \tag{40}$$

Note that this ignores serial correlation in the λ estimates. Not so much a problem for asset returns, but could be in other (e.g., corporate finance applications). Note that the tests of the prices of risk λ do not suffer from the curse of dimensionality. As long as the number of factors is small, we can have many assets (N) relative to observations (T). Thus, there is no need to focus on testing only portfolios of assets for this purpose, except for the fact that individual stock β's have a lot of noise in them.

The test of the residuals α_t (a vector in i) being sufficiently close to zero requires mean α's and, importantly, their covariance matrix:

$$\hat{\alpha} = \frac{1}{T} \sum_{t=1}^{T} \hat{\alpha}_t \tag{41}$$

$$Est.Asy.Var\left(\hat{\alpha}\right) = \frac{1}{T^2} \sum_{t=1}^{T} \left(\hat{\alpha}_t - \hat{\alpha}\right) \left(\hat{\alpha}_t - \hat{\alpha}\right)' \tag{42}$$

Now, we are back to the usual problem: the estimated covariance matrix will be singular if there are two few time-series observations relative to test assets. The joint test of the pricing errors is then

$$\hat{\alpha}' \left[\frac{1}{T^2} \sum_{t=1}^{T} \left(\hat{\alpha}_t - \hat{\alpha}\right) \left(\hat{\alpha}_t - \hat{\alpha}\right)' \right]^{-1} \hat{\alpha} \sim \chi^2 \left(N - K - 1\right) \tag{43}$$

There are K factors and a constant, which accounts for the degrees of freedom. However, since we've estimated the $\lambda's$, we lose degrees of freedom and we must use a pseudo-inverse in place of the inverse covariance matrix.

2.2.2 Why does it work?

To see how this relates to a more standard two-pass regression, define

$$R_t^e \equiv \begin{bmatrix} R_{1t}^e \\ R_{2t}^e \\ \vdots \\ R_{Nt}^e \end{bmatrix}, \qquad \hat{B} = \begin{pmatrix} 1 & \hat{\beta} \end{pmatrix} \tag{44}$$

Then each period's parameter vector is

$$\hat{\lambda}_t = \left(\hat{B}'\hat{B}\right)^{-1} \hat{B}' R_t^e \tag{45}$$

and the price of risk estimate is

$$\hat{\lambda} = \frac{1}{T} \sum_{t=1}^{T} \hat{\lambda}_t = \left(\hat{B}'\hat{B}\right)^{-1} \hat{B}' \overline{R_t^e} \tag{46}$$

Thus, the estimates are what we would get from a single OLS second pass regression. Assuming no serial correlation in excess returns, ignoring the fact that betas are estimated and applying the Delta Method, we have

$$Asy.Var\left(\hat{\lambda}\right) = \left(\hat{B}'\hat{B}\right)^{-1} \hat{B}' \frac{Var\left(R_t^e\right)}{T} \hat{B} \left(\hat{B}'\hat{B}\right)^{-1} \tag{47}$$

The Fama-MacBeth estimate of the variance is asymptotically the same (write it out from above definition and see for yourself)

$$Est.Asy.Var\left(\hat{\lambda}\right) = \left(\hat{B}'\hat{B}\right)^{-1} \hat{B}' \left(\frac{1}{T^2} \sum_{t=1}^{T} \left(R_t^e - \overline{R^e}\right)\left(R_t^e - \overline{R^e}\right)'\right) \hat{B} \left(\hat{B}'\hat{B}\right)^{-1} \tag{48}$$

The FM procedure automatically adjusts for the uncertainty in sample mean of factors. However, it ignores the uncertainty in the estimated β's. You can scale up the variance of $\hat{\alpha}$ by $\left(1 + \hat{\lambda}_1' \Sigma_F^{-1} \hat{\lambda}_1\right)$ to adjust for this. There is no straightforward way to adjust for this in the variance of $\hat{\lambda}$.

2.2.3 Time-varying β's.

The Fama-MacBeth procedure can accommodate time-varying β's. Here's how.

1. Create a time-series of estimated β's for each security. FM use rolling regressions to construct

one-month ahead estimates of betas.

2. Run cross-sectional regression for each time t

$$R^e_{it} = \lambda_{0t} + \lambda_{1t}\beta_{it} + \alpha_{it} \tag{49}$$

where the β estimate is based on data up until date $t-1$ (why?).

3. Proceed as in the constant β case to obtain price of risk estimates and their uncertainty.

- There is no direct comparison between this procedure and the two pass regression since the β's are time-varying.
- The FM procedure can also easily handle a time-varying number of firms, for which there is no easy comparison with two-pass regression statistics.

3 Cross-sectional Asset Pricing: The Attack on the Static CAPM

In this section, we go over some seminal empirical papers that provide evidence against the CAPM.

3.1 Sorting assets into portfolios (Black, Jensen and Scholes test of CAPM)

- The test statistics have shown us that we would like low volatility of residuals.
- Because of difficulty of estimating covariance matrices, we would like the residuals to have low correlation.
  - The first and, to some extent, the second are achieved by sorting into portfolios, which reduces the idiosyncratic risk.
  - However, sorting into portfolios can dampen the α's. To increase our power, likelihood of rejecting the model when the null is false, we would like sort on α.
  - Therefore, portfolios should be sorted on characteristics that theory (alternative hypothesis) suggests may be associated with α. Historically, portfolios have been sorted on

1. Market cap
2. Estimated beta (Black, Jensen, and Scholes)
3. Dividend/Price
4. Industries

5. Return variance

- Typically, we do this by sorting on characteristics prior to time t.
- Then, we investigate the resulting portfolio returns until time $t+j$, where we redo the sort, etc.
- (It is a bit strange to think about the rationale for doing this: We assumed constant $\beta's$ to derive our tests. The factor methodology is therefore a bit ad hoc.)

Question: Why is a sort on in-sample betas inappropriate?

Answer: Betas lie roughly between 0 and 2, with the average around 1. Because of sample error in the beta estimates, some stocks with beta close to one will end up in a high or low beta decile. Some high or low beta stocks will end up in the middle beta deciles. Thus, the true betas of the estimated high and low beta deciles are smaller and larger, respectively, than their estimated values. The estimated price of beta risk will therefore be too low.

- We can show this for both the time-series and cross-sectional case using the now familar formula for the bias in OLS estimates. At the same time we might gain some intuition for when biases can arise in your own future econometric specifications. Remember that if

$$y_t = \alpha + \beta x_t + \varepsilon_t \tag{50}$$

the bias is

$$E\begin{bmatrix} \hat{\alpha}_T \\ \hat{\beta}_T \end{bmatrix} - \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = E\left[\frac{1}{Var_T\left(x_t\right)}\begin{pmatrix} \overline{x_t^2}\overline{\varepsilon_t} - \bar{x}_t\overline{x_t\varepsilon_t} \\ \overline{x_t\varepsilon_t} - \bar{x}_t\bar{\varepsilon}_t \end{pmatrix}\right] \tag{51}$$

- What happens when we sort into portfolios and use the in-sample sorted portfolios as test assets?
  - The high-beta portfolio has truly a lower beta (note: the estimated betas in the second stage regression are the $x-$variables here).
  - Thus, this portfolio has a low residual (the model now overestimates the expected return; i.e. in the time-series regression, these portfolios will tend to have negative intercepts).
  - The low-beta portfolio has truly a higher beta, and its residual is high (i.e., positive intercept in time-series regression). Thus, there is a negative correlation between the x-variable and the residuals in the second stage cross-sectional regression.
  - The term $\overline{x_t\varepsilon_t}$ is therefore less than $\bar{x}_t\overline{\varepsilon_t}$, and the term $\bar{x}_t\overline{x_t\varepsilon_t}$ is less than $\overline{x_t^2}\overline{\varepsilon_t}$.
  - Thus, the intercept in the cross-sectional regression is in this case positively biased, while the slope is negatively biased. This intuition was first noted in Black, Jensen, and Scholes (1972).

3.1.1 Early tests of the CAPM (BJS cont'd)

Pre-sort assets into portfolios based on beta (or other characteristic). Run second stage regression

$$\overline{R_i^e} = \lambda_0 + \lambda_1 \beta_i + \alpha_i \tag{52}$$

The null hypothesis is that $\lambda_0 = 0$ and $\alpha_i = 0$, and that $\lambda_1 = E\left[R_M^e\right]$. Black, Jensen and Scholes (and later studies) find that $\lambda_0 > 0$ and $0 < \lambda_1 < \overline{R_M^e}$. You don't need the intercept, but adding it gives us a more intuitive interpretation of the results: Market risk is priced ($\lambda_1 > 0$), but it is not the only risk factor ($\lambda_0 > 0$). If the market is not the right risk factor, the time-series estimate of the compensation for one unit market risk, $\overline{R_m^e}$, reflects the correlation of the market with the true risk factor. However, λ_1 need not be equal to $\overline{R_m^e}$ as individual asset β's with the market in this case need not be related to their β with respect to the true factor. Without λ_0, λ_1 would pick up average excess returns as well as the effect of the market beta.

BJS conclude that a zero-beta CAPM (Black, 1970), which is a two-factor model, may explain these results. This evidence pointing at multiple (or other) risk factors made multifactor models the natural next step.

3.2 Fama-French (1992) - The Death of the CAPM?

Prior to Fama-French (1992), standard view was

1. Positive relation between betas and expected excess returns to stocks

   A little sensitive to the choice of portfolios

2. Other variables also captured some variation in expected returns

   - Firm characteristics ME, BE/ME, E/P
   - Macro factors (Chen, Roll, and Ross (1986))
   - But some deviations expected because aggregate stock market is only a proxy for market portfolio

3. We do not have any better practical theory to determine stocks' expected returns.

After Fama-French (1992), standard view has become

- Static CAPM viewed as almost useless for explaining cross-sectional variations in expected returns to stocks

- Yet no proper statistical test of the CAPM in the paper!

- Main contributions

  1. A clever method for constructing portfolios
  2. Changing the terms of the debate about models of expected returns
     Focus on R^2 of cross-sectional variations in expected returns, not on formal statistical tests of α's (i.e. focus on economic significance and not statistical significance).

3.2.1 The Fama-French Portfolios

- Typical sorts are on ME, but FF note ME, beta are highly negatively correlated
- FF use double sort: 10 deciles of ME, then each into 10 pretest beta deciles
  - Within a given ME range, pre- and post-test betas can vary widely - reduces correlation
- Portfolio rebalanced yearly, 1963–1989. Beginning date determined by availability of Compustat data on book value of equity, earnings
- Table of 100 mean portfolio returns is main intuitive evidence against CAPM:

Table I

Average Returns, Post-Ranking βs and Average Size For Portfolios Formed on Size and then β: Stocks Sorted on ME (Down) then Pre-Ranking β (Across): July 1963 to December 1990

Portfolios are formed yearly. The breakpoints for the size (ME, price times shares outstanding) deciles are determined in June of year t (t = 1963–1990) using all NYSE stocks on CRSP. All NYSE, AMEX, and NASDAQ stocks that meet the CRSP-COMPUSTAT data requirements are allocated to the 10 size portfolios using the NYSE breakpoints. Each size decile is subdivided into 10 β portfolios using pre-ranking βs of individual stocks, estimated with 2 to 5 years of monthly returns (as available) ending in June of year t. We use only NYSE stocks that meet the CRSP-COMPUSTAT data requirements to establish the β breakpoints. The equal-weighted monthly returns on the resulting 100 portfolios are then calculated for July of year t to June of year $t + 1$.

The post-ranking βs use the full (July 1963 to December 1990) sample of post-ranking returns for each portfolio. The pre- and post-ranking βs (here and in all other tables) are the sum of the slopes from a regression of monthly returns on the current and prior month's returns on the value-weighted portfolio of NYSE, AMEX, and (after 1972) NASDAQ stocks. The average return is the time-series average of the monthly equal-weighted portfolio returns, in percent. The average size of a portfolio is the time-series average of monthly averages of ln(ME) for stocks in the portfolio at the end of June of each year, with ME denominated in millions of dollars.

The average number of stocks per month for the size-β portfolios in the smallest size decile varies from 70 to 177. The average number of stocks for the size-β portfolios in size deciles 2 and 3 is between 15 and 41, and the average number for the largest 7 size deciles is between 11 and 22.

The All column shows statistics for equal-weighted size-decile (ME) portfolios. The All row shows statistics for equal-weighted portfolios of the stocks in each β group.

	All	Low-β	β-2	β-3	β-4	β-5	β-6	β-7	β-8	β-9	High-β
					Panel A: Average Monthly Returns (in Percent)						
All	1.25	1.34	1.29	1.36	1.31	1.33	1.28	1.24	1.21	1.25	1.14
Small-ME	1.52	1.71	1.57	1.79	1.61	1.50	1.50	1.37	1.63	1.50	1.42
ME-2	1.29	1.25	1.42	1.36	1.39	1.65	1.61	1.37	1.31	1.34	1.11
ME-3	1.24	1.12	1.31	1.17	1.70	1.29	1.10	1.31	1.36	1.26	0.76
ME-4	1.25	1.27	1.13	1.54	1.06	1.34	1.06	1.41	1.17	1.35	0.98
ME-5	1.29	1.34	1.42	1.39	1.48	1.42	1.18	1.13	1.27	1.18	1.08
ME-6	1.17	1.08	1.53	1.27	1.15	1.20	1.21	1.18	1.04	1.07	1.02
ME-7	1.07	0.95	1.21	1.26	1.09	1.18	1.11	1.24	0.62	1.32	0.76
ME-8	1.10	1.09	1.05	1.37	1.20	1.27	0.98	1.18	1.02	1.01	0.94
ME-9	0.95	0.98	0.88	1.02	1.14	1.07	1.23	0.94	0.82	0.88	0.59
Large-ME	0.89	1.01	0.93	1.10	0.94	0.93	0.89	1.03	0.71	0.74	0.56

Table I—*Continued*

	All	Low-β	β-2	β-3	β-4	β-5	β-6	β-7	β-8	β-9	High-β
					Panel B: Post-Ranking βs						
All		0.87	0.99	1.09	1.16	1.26	1.29	1.35	1.45	1.52	1.72
Small-ME	1.44	1.05	1.18	1.28	1.32	1.40	1.40	1.49	1.61	1.64	1.79
ME-2	1.39	0.91	1.15	1.17	1.24	1.36	1.41	1.43	1.50	1.66	1.76
ME-3	1.35	0.97	1.13	1.13	1.21	1.26	1.28	1.39	1.50	1.51	1.75
ME-4	1.34	0.78	1.03	1.17	1.16	1.29	1.37	1.46	1.51	1.64	1.71
ME-5	1.25	0.66	0.85	1.12	1.15	1.16	1.26	1.30	1.43	1.59	1.68
ME-6	1.23	0.61	0.78	1.05	1.16	1.22	1.28	1.36	1.46	1.49	1.70
ME-7	1.17	0.57	0.92	1.01	1.11	1.14	1.26	1.24	1.39	1.34	1.60
ME-8	1.09	0.53	0.74	0.94	1.02	1.13	1.12	1.18	1.26	1.35	1.52
ME-9	1.03	0.58	0.74	0.80	0.95	1.06	1.15	1.14	1.21	1.22	1.42
Large-ME	0.92	0.57	0.71	0.78	0.89	0.95	0.92	1.02	1.01	1.11	1.32

3.2.2 Statistical Tests in Fama-French (1992)

- Run Fama-MacBeth regressions on individual stocks (not 100 portfolios)
  - Include ME and BE/ME in cross-sectional regression

$$R_{it} = b_0 + b_1 \hat{\beta}_{it} + b_2 \log(ME_{it}) + b_3 \log(BE/ME_{it}) + \alpha_{it} \tag{53}$$

  - Do not estimate stock betas; assign them portfolio betas
  - Impossible to do this with two-pass regression (changing number of stocks per month, too many assets to estimate)
  - Cannot test whether mean α's are close to zero (stock mix changes over time)
- Test whether individual mean cross-sectional coefficients are zero
- No test of whether coef on market equals time-series mean excess market return

Table III

Average Slopes (*t*-Statistics) from Month-by-Month Regressions of Stock Returns on β, Size, Book-to-Market Equity, Leverage, and E/P: July 1963 to December 1990

Stocks are assigned the post-ranking β of the size-β portfolio they are in at the end of June of year t (Table I). BE is the book value of common equity plus balance-sheet deferred taxes, A is total book assets, and E is earnings (income before extraordinary items, plus income-statement deferred taxes, minus preferred dividends). BE, A, and E are for each firm's latest fiscal year ending in calendar year $t-1$. The accounting ratios are measured using market equity ME in December of year $t-1$. Firm size ln(ME) is measured in June of year t. In the regressions, these values of the explanatory variables for individual stocks are matched with CRSP returns for the months from July of year t to June of year $t+1$. The gap between the accounting data and the returns ensures that the accounting data are available prior to the returns. If earnings are positive, E(+)/P is the ratio of total earnings to market equity and E/P dummy is 0. If earnings are negative, E(+)/P is 0 and E/P dummy is 1.

The average slope is the time-series average of the monthly regression slopes for July 1963 to December 1990, and the t-statistic is the average slope divided by its time-series standard error.

On average, there are 2267 stocks in the monthly regressions. To avoid giving extreme observations heavy weight in the regressions, the smallest and largest 0.5% of the observations on E(+)/P, BE/ME, A/ME, and A/BE are set equal to the next largest or smallest values of the ratios (the 0.005 and 0.995 fractiles). This has no effect on inferences.

β	ln(ME)	ln(BE/ME)	ln(A/ME)	ln(A/BE)	E/P Dummy	E(+)/P
0.15						
(0.46)						
	−0.15					
	(−2.58)					
−0.37	−0.17					
(−1.21)	(−3.41)					
		0.50				
		(5.71)				
			0.50	−0.57		
			(5.69)	(−5.34)		
					0.57	4.72
					(2.28)	(4.57)
	−0.11	0.35				
	(−1.99)	(4.44)				
	−0.11		0.35	−0.50		
	(−2.06)		(4.32)	(−4.56)		
	−0.16				0.06	2.99
	(−3.06)				(0.38)	(3.04)
	−0.13	0.33			−0.14	0.87
	(−2.47)	(4.46)			(−0.90)	(1.23)
	−0.13		0.32	−0.46	−0.08	1.15
	(−2.47)		(4.28)	(−4.45)	(−0.56)	(1.57)

- Interpretation of statistical evidence

  1. Loading on market return not helpful in explaining cross-sectional variation in mean returns to their chosen portfolios
  2. ME, BE/ME have strong explanatory power, both statistically and economically
  3. Therefore from the practical perspective of determining expected returns, ME and BE/ME dominate beta

- Important unanswered question: Do ME, BE/ME proxy for unobserved economic factors which investors want to avoid?

FF '92 suggest BE/ME proxies for financial distress, which may be a priced risk.

The enduring influence of Fama and French (1992)

- Started debate: should models of discount rates be evaluated primarily with statistical tests, "practical value" (e.g., cross-sectional R2), or some other metric?
- Changed set of assets used to evaluate asset-pricing models
  - Emphasis on cross-sectional dispersion that allows researcher to distinguish among beta, ME, and BE/ME
- (Along with FF (1993)) asked a question that generated hundreds of papers:
  - why are mean returns to stocks with high BE/ME ratios so large?

Where do we go after FF (1992)?

- One branch of the literature studies metrics for evaluating models of discount rates (esp. Hansen and coauthors)
- Another branch pursues atheoretic, "practical value" models of discount rates (e.g., Fama-French 3-factor model)
- Others relax assumptions that underlie CAPM and build alternative models
  - Alternative proxies for the return to total wealth
  - Nonseparable utility (consumption goods, states of world)
  - Time-varying investment opportunity set
  - Limitations on trading assets (esp. human capital)
  - Non-rational investors