Empirical Asset Pricing

Francesco Franzoni

Swiss Finance Institute - University of Lugano

Testing Asset Pricing Models: Overview

Lecture Outline

1. Testing CAPM: historical development

2. Statistical description of the asset pricing tests

3. Statistical discipline and philosophy

Relevant readings:

- Cochrane: chapters 7, 12, 14, 15, 16

- Huang and Litzenberger: chapter 10

- Campbell, Lo, and MacKinlay: chapters 6 and 7

1. Testing CAPM: historical development

Quick Review: The CAPM

- The Sharpe-Lintner-Mossin (SLM) version of the model assumes that the risk-free rate is available. For any asset i :

$$E(R_i) = R_f + \beta_i \left(E(R_m) - R_f \right) \tag{1}$$

where R_m is the total wealth (i.e., market) portfolio and

$$\beta_i = \frac{Cov(R_i, R_m)}{Var(R_m)} \tag{2}$$

- The model states that the market return is mean-variance efficient

- That is: the expected return on any asset can be spanned by the expected return on the market and the risk free rate

- The Black (1972) version of the model is derived without assuming the existence of the risk-free rate:

$$E(R_i) = E(R_0) + \beta_i (E(R_m) - E(R_0)) \tag{3}$$

where R_0 is the return on a mean-variance efficient portfolio with zero covariance with R_m. It turns out that $E(R_m) - E(R_0) > 0$

Testable Predictions

- SLM version:

1. Consider the time-series regression

$$R_{it} - R_f = \alpha_i + \beta_i \left(R_{mt} - R_f \right) + \varepsilon_{it}. \tag{4}$$

   Take expectations of both sides. The model predicts

   (a) $\alpha_i = 0 \quad \forall i = 1 \ldots N$

2. Consider the cross-sectional regression

$$E_T \left(R_i^e \right) = a + \beta_i \lambda + d_i \gamma + u_i \qquad i = 1 \ldots N \tag{5}$$

   $E_T \left(R_i^e \right)$: time-series average excess return (it is an estimate of the expected excess return)
   d_i : vector of stock characteristics different from beta (e.g. size, dividend yield, volatility, etc.):

   (a) $\lambda = E\left(R_m\right) - R_f > 0$

   (b) $\alpha_i \equiv a + u_i = 0 \quad \forall i$ (zero pricing errors)

   (c) $\gamma = 0$

- Black version:

1. In the model

$$E(R_i) = \alpha_i + \beta_i E(R_m)$$

it has to be the case that

$$\alpha_i = E(R_0)(1 - \beta_i) \quad \forall i = 1 \ldots N.$$

This is a non-linear restriction on the parameters

Conceptual Issues in the Tests

1. The model predictions are stated in terms of *ex ante* expected returns and betas but tested on sample moments

   - The solution is to assume Rational Expectations: the realized returns are drawings from equilibrium distributions. Because of rational expectations, the *ex ante* distributions correspond to the equilibrium distributions. So, the sample moments converge to the ex-ante moments

2. CAPM holds period by period: it is a conditional model.
   How to handle non-stationarity in the moments of the *ex ante* distribution?

   - One solution is to assume that CAPM holds unconditionally too.
     Implications of this assumption:
     The conditional CAPM:

$$
\begin{aligned}
E_t\left(R^e_{i,t+1}\right) &= \beta_{it} E_t\left(R^e_{m,t+1}\right) \\
\beta_{it} &= \frac{Cov_t\left(R_{i,t+1}, R_{m,t+1}\right)}{Var_t\left(R_{m,t+1}\right)}
\end{aligned}
$$

Take unconditional expectations of both sides

$$\begin{aligned} E\left(E_t\left(R^e_{i,t+1}\right)\right) &= E\left(\beta_{it}E_t\left(R^e_{m,t+1}\right)\right) \\ E\left(R^e_{i,t+1}\right) &= Cov\left(\beta_{it}, E_t\left(R^e_{m,t+1}\right)\right) \\ &\quad +E\left(\beta_{it}\right)E\left(R^e_{m,t+1}\right) \end{aligned}$$

For the model to hold unconditionally, one needs to assume that

$$\begin{aligned} Cov\left(\beta_{it}, E_t\left(R^e_{m,t+1}\right)\right) &= 0 \\ E\left(\beta_{it}\right) &= \beta_i = \frac{Cov\left(R_i, R_m\right)}{Var\left(R_m\right)} \end{aligned}$$

This is not necessarily true. For example:

– Beta can change along the business cycle

– Beta can change along the firm's life-cycle

- The other solution is to model the conditional moments. We will discuss this strategy later on. For now, the focus is on unconditional tests

3. *Roll's (1977) critique*: the total wealth portfolio includes non-marketable assets (e.g. human capital, private businesses, etc.). Instead, the tests typically focus on a proxy of the market that is traded (e.g. the S&P 500, the CRSP index, etc.)

   - Consequence: measurement error in the market return.
     Rejection of CAPM may depend on use of incorrect market portfolio
   - Most tests ignore the unobservability and assume proxy is mean-variance efficient
   - Also: if the true market portfolio is sufficiently correlated with the proxy (above 70%), a rejection of the proxy implies a rejection of the true portfolio (Stambaugh (1982), Kandal and Stambaugh (1987), Shanken (1987))
   - Some authors have tried to compute a broader proxy by including the return on human capital (Jagannathan and Wang (1996))

Econometric Issues in the Tests

1. The errors in equation (5) are likely to be heteroskedastic and correlated across assets
    - OLS coefficients are unbiased but inefficient (OLS is not BLUE, GLS is)
    - OLS standard errors are biased
    - Possible Solutions (we will see them below in more detail):
    (a) Use feasible GLS
    (b) Use OLS to estimate the slopes and adjust the standard errors
    (c) Fama and MacBeth (1973) procedure
    - We will consider them in the next section

2. The betas in equation (5) are measured with error because they are estimates of true betas

- Measurement error causes $\hat{\gamma}$ to be biased towards zero and CAPM rejected

  True model:

$$E_T\left(R_i^e\right) = a + \beta_i \gamma + u_i$$

  But one observes:

$$\hat{\beta}_i = \beta_i + w_i$$

  The Plim of the cross-sectional estimate is:

$$\begin{aligned} P\lim \hat{\gamma} &= \frac{Cov\left(E_T\left(R_i^e\right), \hat{\beta}_i\right)}{Var\left(\hat{\beta}_i\right)} \\ &= \frac{Cov\left(a + \beta_i \gamma + u_i, \beta_i + w_i\right)}{Var\left(\beta_i + w_i\right)} \\ &= \gamma \frac{Var\left(\beta_i\right)}{Var\left(\beta_i\right) + Var\left(w_i\right)} < \gamma \end{aligned}$$

- Solutions:

(a) Group the data into portfolios

(b) Use instrumental variables

(c) Adjust the estimates for the bias

3. The betas in equation (4) are likely to change over time

   - Betas change over the life-cycle in a non-stationary way
   - Historical solution: form portfolios according to a stationary characteristic
   - Does it work?
     See Franzoni (2002): Where's beta going?

4. High volatility of individual stock returns

   - For individual stocks, cannot reject hypothesis that average returns are all the same
   - St. error of mean is $\frac{\sigma_i}{T^{1/2}}$. If σ_i between 40% and 80% and $T = 60$ months, very large confidence intervals
   - Solution: form portfolios

5. Non-synchronous trading

   - Small stocks react slowly to information

- Solution (Scholes and Williams, 1977, Dimson, 1979)

$$\hat{\beta} = \frac{\hat{\beta}^{+} + \hat{\beta} + \hat{\beta}^{-}}{1 + 2\rho}$$

where $\hat{\beta}^{+}, \hat{\beta}^{-}$, and $\hat{\beta}$ come from the simple regression of stock returns on leads, lags, and the contemporaneous market return:

$$\begin{aligned} R_{it}^{e} &= \alpha_i^{+} + \beta_i^{+} R_{m,t+1} + \varepsilon_t^{+} \\ R_{it}^{e} &= \alpha_i + \beta_i R_{m,t} + \varepsilon_t \\ R_{it}^{e} &= \alpha_i^{-} + \beta_i^{-} R_{m,t-1} + \varepsilon_t^{-} \end{aligned}$$

ρ is the 1st order auto-correlation of the market return.
Note that if $\rho = 1$, it is as if you were running three times the same regression. Then, you have to divide the sum of the betas by $1 + 2\rho = 3$
If $\rho = 0$, each of the three regressions provides independent information on the reaction of stock returns to market returns. Then, the sum of the three betas does not need to be normalized.

Portfolio Grouping Approach

- Consider forming G groups of L stocks
- The measured beta for group g is

$$\hat{\beta}_g = \frac{1}{L}\sum_{j=1}^{L}\hat{\beta}_j = \frac{1}{L}\sum_{j=1}^{L}\left(\beta_j + w_j\right) = \beta_g + \frac{1}{L}\sum_{j=1}^{L} w_j$$

- So the variance of the measurement error for the portfolio beta is

$$Var\left(\frac{1}{L}\sum_{j=1}^{L} w_j\right) = \frac{1}{L}\sigma_w^2$$

  as long as the measurement errors of the individual securities are uncorrelated

- As $L \to \infty$ the measurement error goes to zero and the estimates are consistent
- The grouping procedure can induce correlation in within-group measurement error
  E.g.: suppose grouping by estimated betas

Grouping and Efficiency

- Grouping reduces dispersion in betas relative to ungrouped data:

$$\underbrace{Var(\beta_i)}_{Ungrouped\ Variance} = \underbrace{Var(E(\beta|G))}_{Between-Group\ Variance} + \underbrace{E_G(Var(\beta_i|G))}_{Average\ Within-Group\ Variance}$$

- Hence, it reduces the efficiency of the cross-sectional estimates
- Need grouping procedure that maximizes between-group variation in betas and minimizes correlation with measurement error
- E.g.: group by non-contemporaneous estimates of betas from non-overlapping data
  - Measurement error is uncorrelated over time
  - Estimated betas are positively correlated with true betas over time

Instrumental Variable Approach

- Closely related to grouping approach
- Need to find instrument z_i for β_i such that

$$\begin{aligned} Cov\left(z_i, w_i\right) &= 0 \text{ uncorrelated with measurement error} \\ Cov\left(z_i, u_i\right) &= 0 \text{ uncorrelated with regression error} \\ Cov\left(z_i, \beta_i\right) &\neq 0 \text{ correlated with true beta} \end{aligned}$$

- Then, use Instrumental Variable (IV) estimator

$$\hat{\lambda}_{IV} = \frac{Cov\left(E_T\left(R_i^e\right), z_i\right)}{Cov\left(\hat{\beta}_i, z_i\right)}$$

- Or one can use 2SLS if multiple instruments
- One can show that grouping is equivalent to IV on ungrouped data with z_i equal to group rank, if group membership is constant over-time

Bias Adustment

- Remember the cross-sectional estimates with measurement error w_i

$$
\begin{aligned}
P \lim \hat{\gamma} &= \frac{Cov\left(E_T\left(R_i^e\right), \hat{\beta}_i\right)}{Var\left(\hat{\beta}_i\right)} \\
&= \gamma \frac{Var\left(\beta_i\right)}{Var\left(\beta_i\right) + Var\left(w_i\right)}
\end{aligned}
$$

These are cross-sectional population moments

- If one can find a consistent estimate of $Var\left(w_i\right)$, then one can compute consistent bias-adjusted estimate

$$
\hat{\gamma}_{adj} = \frac{Cov^s\left(E_T\left(R_i^e\right), \hat{\beta}_i\right)}{Var^s\left(\hat{\beta}_i\right) - \widehat{Var\left(w_i\right)}}
$$

where the superscript s denotes sample moments

- $\widehat{Var}(w_i)$ is provided by the average of estimated variances of $\hat{\beta}_i$ from time-series regressions of R^e_{it} on R^e_{mt}

$$\widehat{Var}(w_i) = \frac{1}{N}\sum_{j=1}^{N} \widehat{Var}\left(\hat{\beta}_j\right)$$

- To prove, use Law of Large Numbers for random variables with different mean:

$$\frac{1}{N}\sum_{j=1}^{N} X_i \underset{N\to\infty}{\longrightarrow} \frac{1}{N}\sum_{j=1}^{N} \mu_i$$

where μ_i is mean of X_i

2. Statistical description of AP tests

- General focus on multi-factor asset pricing models

$$E\left(R_i^e\right) = \beta_i' \lambda \tag{6}$$

  where β_i is a K-dimensional vector of multiple-regression slopes and λ is K-vector of factor risk premia

- For example: APT, ICAPM

- CAPM is the case with $K = 1$

- The methodology extends unambiguosly from the CAPM tests to tests of multifactor models

- Questions:
  - How to estimate parameters?
  - Standard errors?
  - How to test the model predictions?

Time-Series Regressions

- You can apply this approach only if factors $f_{1,...,}f_K$ are returns
- Then, the model (6) applies to factors as well:

$$E\left(f_k\right) = \lambda_k \quad k = 1...K \tag{7}$$

- So, one can re-write the model in (6) as:

$$E\left(R_i^e\right) = \beta_i' E\left(f\right) \tag{8}$$

where $E\left(f\right)$ is the vector of expected excess returns on the k factors

- Assume $K = 1$ for simplicity, consider the time-series regression:

$$R_{it}^e = \alpha_i + \beta_i f_t + \varepsilon_t \tag{9}$$

- Take expectation of each side. Then, the implication of (8) is

$$\alpha_i = 0 \quad i = 1...N$$

- You can estimate α_i by running regression (9) for each asset

- Then, use t-tests to test $\alpha_i = 0$

- But you want to test the hypothesis that all alphas are jointly zero

- The ε_{it} are correlated across assets with variance-covariance matrix $\boldsymbol{\Sigma} = E\left(\varepsilon_t \varepsilon_t'\right)$

- The asymptotically valid test for

$$H_o : \alpha = 0$$

  is

$$\hat{\alpha}' \left[\widehat{Var\left(\hat{\alpha}\right)}\right]^{-1} \hat{\alpha} \sim \chi^2_N$$

  where α is a N-dimensional vector of pricing errors

- Intuition: reject the model if the weighted sum of the squared errors is far from zero

- In the case $K = 1$, assuming no autocorrelation and stationarity, and noting that $\hat{\alpha}_i$ contains the average ε_{it}, the test becomes

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

where $E_T(f)$ is the sample mean of the factor (an estimate of the risk premium) and $\hat{\Sigma}$ is the sample vcov matrix of the residuals from the N regressions

Proof. (Sketch): $\widehat{Var(\hat{\alpha})}$ is $N \times N$ matrix

$$\widehat{Var(\hat{\alpha})}=\begin{bmatrix}\hat{\sigma}^2_{\hat{\alpha}_1} & \hat{\sigma}_{\hat{\alpha}_1\hat{\alpha}_2} & \cdots & \hat{\sigma}_{\hat{\alpha}_1\hat{\alpha}_N}\\ \hat{\sigma}_{\hat{\alpha}_2\hat{\alpha}_1} & \hat{\sigma}^2_{\hat{\alpha}_2} & & \\ \vdots & & \ddots & \\ \hat{\sigma}_{\hat{\alpha}_N\hat{\alpha}_1} & & & \hat{\sigma}^2_{\hat{\alpha}_N}\end{bmatrix}$$

Focus on:

$$\begin{aligned}\hat{\alpha}_i &= \bar{R}^e_i-\hat{\beta}_iE_T(f)\\ &= \alpha_i+\beta_iE_T(f)+\bar{\varepsilon}_i-\hat{\beta}_iE_T(f)\\ &= \alpha_i+\left(\beta_i-\hat{\beta}_i\right)E_T(f)+\bar{\varepsilon}_i\end{aligned}$$

So:

$$
\begin{aligned}
Var\left(\hat{\alpha}_i\right) &= Var\left(\hat{\beta}_i\right) E_T^2\left(f\right) + Var\left(\bar{\varepsilon}_i\right) \\
&= \frac{\sigma_{\varepsilon_i}^2}{\hat{\sigma}^2\left(f\right) T} E_T^2\left(f\right) + \frac{1}{T}\sigma_{\varepsilon_i}^2 \\
&= \frac{1}{T}\sigma_{\varepsilon_i}^2 \left[1 + \left(\frac{E_T\left(f\right)}{\hat{\sigma}\left(f\right)}\right)^2\right]
\end{aligned}
$$

Similarly:

$$
Cov\left(\hat{\alpha}_i, \hat{\alpha}_j\right) = \frac{1}{T}\sigma_{\varepsilon_i \varepsilon_j} \left[1 + \left(\frac{E_T\left(f\right)}{\hat{\sigma}\left(f\right)}\right)^2\right]
$$

Hence, it follows that:

$$
Var\left(\hat{\alpha}\right) = \frac{1}{T}\left[1 + \left(\frac{E_T\left(f\right)}{\hat{\sigma}\left(f\right)}\right)^2\right] \Sigma
$$

So:

$$\widehat{Var\left(\hat{\alpha}\right)} = \frac{1}{T}\left[1 + \left(\frac{E_T\left(f\right)}{\hat{\sigma}\left(f\right)}\right)^2\right]\hat{\Sigma}$$

Then:

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

You only need to prove that the distribution is a chi-squared. For that you can use standard theorems on the limit distribution of squared residuals. ■

- Gibbons, Ross, and Shanken (GRS, 1989) provide a small sample test assuming joint normality of the ε_{it}

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

- Using efficient set algebra, one can prove that

$$\left(\frac{\mu_q}{\sigma_q}\right)^2 = \left(\frac{E_T\left(f\right)}{\hat{\sigma}\left(f\right)}\right)^2 + \hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha} \qquad (10)$$

where $\frac{\mu_q}{\sigma_q}$ is the Sharpe ratio of the tangency portfolio and $\frac{E_T(f)}{\hat{\sigma}(f)}$ is the Sharpe ratio of the factor

- Then, the GRS statistic can be rewritten as

$$\frac{T-N-1}{N} \frac{\left(\mu_q/\sigma_q\right)^2 - \left(E_T(f)/\hat{\sigma}(f)\right)^2}{\left[1+\left(\frac{E_T(f)}{\hat{\sigma}(f)}\right)^2\right]^{-1}} \sim F_{N,\ T-N-1}$$

- In this form, the interpretation is: Reject the null if the factor is far from the tangency portfolio on the ex-post efficient frontier

- In other words: reject CAPM if the market is far from the ex-post efficient frontier

- In case of $K > 1$ factors, the GRS statistic is

$$\frac{T-N-K}{N} \frac{\left(\frac{\mu_q}{\sigma_q}\right)^2 - E_T(f)'\,\hat{\Omega}^{-1}E_T(f)}{\left[1+E_T(f)'\,\hat{\Omega}^{-1}E_T(f)\right]^{-1}} \sim F_{N,T-N-K}$$

where

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

is the sample vcov matrix of the factors and

$$E_T\left(f\right)' \hat{\Omega}^{-1} E_T\left(f\right)$$

is the multifactor equivalent of the squared Sharpe ratio of the factor

Proof of Equation (10)

- Let q be a portfolio on the mean variance efficient frontier. Let μ be the vector of expected excess returns on the $N+1$ assets in the market (i.e. N test assets and 1 factor). Hence, q has minimum variance for given expected return e

$$\begin{aligned} Min \quad & q'Vq \\ s.t.\ q'\mu &= e \end{aligned}$$

- The Lagrangean for this problem is

$$L = q'Vq + 2\lambda\left(e - q'\mu\right)$$

which gives first order conditions with respect to q and λ (the Lagrange multiplier)

$$\begin{aligned} 2Vq - 2\lambda\mu &= 0 \\ e - q'\mu &= 0 \end{aligned}$$

- Hence, the frontier portfolio q has to satisfy

$$q = \lambda V^{-1}\mu \tag{11}$$

- Using Equation (11), the variance of q is therefore

$$\sigma_q^2 = q'Vq = \lambda^2 \mu' V^{-1} \mu$$

- The squared expected excess return of q is

$$\mu_q^2 = \left(q'\mu\right)^2 = \lambda^2 \left(\mu' V^{-1} \mu\right)^2$$

- We can then compute the squared Sharpe ratio of q as

$$\frac{\mu_q^2}{\sigma_q^2} = \mu' V^{-1} \mu \tag{12}$$

- Now, let us remember that this market is composed of $N+1$ assets, where the first asset is the factor f

- The vector of returns of the N assets can be written as

$$R_t = \alpha + \beta f_t + \varepsilon_t,$$

where α and β are N-dimensional vectors capturing the alphas and betas of the N assets with respect to the factor f. Also, ε is the idiosyncratic component of returns, which has variance-covariance matrix equal to Σ

- Hence, the vector μ can be written as

$$\mu = \begin{bmatrix} \mu_f \\ \alpha + \beta \mu_f \end{bmatrix} \tag{13}$$

- For the same reason, the variance-covariance matrix V can be written as

$$V = \begin{bmatrix} \sigma_f^2 & \beta' \sigma_f^2 \\ \beta \sigma_f^2 & \beta \beta' \sigma_f^2 + \Sigma \end{bmatrix} \tag{14}$$

- Using the formula for the inverse of a partitioned matrix (see, e.g., Greene, Econometric Analysis), we can obtain the inverse of V as

$$V^{-1} = \begin{bmatrix} \frac{1}{\sigma_f^2} + \beta' \Sigma^{-1} \beta & -\beta' \Sigma^{-1} \\ -\Sigma^{-1} \beta & \Sigma^{-1} \end{bmatrix} \tag{15}$$

- Using Equations (13) and (15), after some tedious algebra, we can re-write Equation (12) as

$$\frac{\mu_q^2}{\sigma_q^2} = \frac{\mu_f^2}{\sigma_f^2} + \alpha' \Sigma^{-1} \alpha,$$

which completes the proof

Cross-Sectional Regressions

- This is the only approach available when the factors are not returns
- Two-pass methodology:

1. Estimate β_i from time-series regressions with multiple factors if $K > 1$

$$R^e_{it} = a_i + \beta_i' f_t + \varepsilon_{it} \quad t = 1...T \tag{16}$$

Note that if the AP model in (6) is correct

$$a_i = \beta_i' (\lambda - E(f)) \tag{17}$$

2. Regress average returns on estimated β_i

$$E_T(R^e_{it}) = \beta_i' \lambda + \alpha_i \quad i = 1...N \tag{18}$$

where E_T denotes a time-series average over the T observations in the sample.
Based on equations (16) and (17), under the null, the pricing errors α_i's probability limit is

(assume that $\hat{\beta} \to \beta$ and $\hat{\lambda} \to \lambda$):

$$\begin{aligned} \alpha_i &= E_T\left(R^e_{it}\right) - \beta_i'\lambda \\ &= a_i + \beta_i' E_T\left(f_t\right) + E_T\left(\varepsilon_{it}\right) - \beta_i'\lambda \\ &= \beta_i'\left(\lambda - E\left(f\right)\right) + \beta_i' E_T\left(f_t\right) + E_T\left(\varepsilon_{it}\right) - \beta_i'\lambda \\ &= E_T\left(\varepsilon_{it}\right) + \beta_i'\left(E_T\left(f_t\right) - E\left(f_t\right)\right) \end{aligned}$$

So, the vcov matrix of α, which is the $N \times 1$ vector of residuals from regression (18), is

$$E\left(\alpha\alpha'\right) = \frac{1}{T}\left(\Sigma + \beta\Omega\beta'\right)$$

where β is a $N \times K$ matrix of factor loadings for the N assets on the K factors, and Ω is the vcov of the K factors

- $\hat{\alpha}$ is the vector of fitted residuals, which under the null, have zero expectation

- One can test the AP model with

$$\hat{\alpha}'\left[Var\left(\hat{\alpha}\right)\right]^{-1}\hat{\alpha} \sim \chi^2_{N-K}$$

- Using standard OLS results ($Var(\hat{u}) = MVar(u)M$, where $\hat{u}$ are the fitted residuals and $M = I - X(X'X)^{-1}X'$ is the residual making matrix. In our case $u = \alpha$ and $X = \beta$), we can compute the vcov of $\hat{\alpha}$ as

$$Var(\hat{\alpha}) = \frac{1}{T}\left(I - \beta\left(\beta'\beta\right)^{-1}\beta'\right)\Sigma\left(I - \beta\left(\beta'\beta\right)^{-1}\beta'\right)$$

where $\beta\Omega\beta'$ cancels out from the vcov matrix

- Notice that these formulas assume that β are known. See below for corrections that take into account that β is estimated

GLS Cross-Sectional Regressions

- Since the residuals α_i are correlated, OLS is inefficient
- The BLUE estimator is GLS

$$
\begin{aligned}
\hat{\lambda} &= \left(\beta' E\left(\alpha\alpha'\right)^{-1}\beta\right)^{-1}\beta' E\left(\alpha\alpha'\right)^{-1} E_T\left(R^e\right) \\
\hat{\alpha} &= E_T\left(R^e\right) - \hat{\lambda}\beta
\end{aligned}
$$

- If you do Choleski decomposition

$$
E\left(\alpha\alpha'\right)^{-1} = CC'
$$

you can think of GLS as OLS performed on rotated data set

$$
\begin{aligned}
\beta^* &= C'\beta \\
E_T\left(R^{e*}\right) &= E_T\left(C'R^e\right)
\end{aligned}
$$

which implies testing the model on a different set of portfolios: $C'R^e$

- Unfortunately, Σ is not known

- Use an estimate $\hat{\Sigma}$ to do feasible GLS
- However, feasible GLS can be largely inefficient in small samples
- So, many authors prefer to do OLS and to report standard errors that account for correlation

$$\begin{aligned} \sigma^2\left(\hat{\lambda}_{ols}\right) &= \frac{1}{T}\left[\left(\beta'\beta\right)^{-1}\beta'\Sigma\beta\left(\beta'\beta\right)^{-1}+\Omega\right] \\ Var\left(\hat{\alpha}_{ols}\right) &= \frac{1}{T}\left(I-\beta\left(\beta'\beta\right)'^{-1}\beta'\right)' \\ &\quad \Sigma\left(I-\beta\left(\beta'\beta\right)'^{-1}\beta'\right) \end{aligned} \tag{19}$$

- With GLS estimates, you can test the AP model using

$$T\hat{\alpha}'_{gls}\hat{\Sigma}^{-1}\hat{\alpha}_{gls}\sim\chi^2_{N-K}$$

- To derive distribution, use Choleski decomposition and consider that asymptotically

$$T^{1/2}C'\hat{\alpha}\sim N\left(0,I-\delta\left(\delta'\delta\right)^{-1}\delta'\right)$$

where $\delta=C'\beta$ and $I-\delta\left(\delta'\delta\right)^{-1}\delta'$ is idempotent matrix with rank $N-K$

Shanken Correction

- In fact, β_i are estimates, not the true parameters
- Hence, need to account for sampling error in β_i, when computing standard errors
- Shanken's (1992) correction :

$$\begin{aligned}\sigma^2\left(\hat{\lambda}_{ols}\right) &= \frac{1}{T}\left[\left(\beta'\beta\right)^{-1}\beta'\Sigma\beta\left(\beta'\beta\right)^{-1}\left(1+\lambda'\Omega^{-1}\lambda\right)\right] \\ &\quad +\frac{1}{T}\Omega \\ \sigma^2\left(\hat{\lambda}_{gls}\right) &= \frac{1}{T}\left[\left(\beta'\Sigma^{-1}\beta\right)^{-1}\left(1+\lambda'\Omega^{-1}\lambda\right)+\Omega\right]\end{aligned} \tag{20}$$

  where Ω is the vcov of the factors
- You have a multiplicative $\left(1+\lambda'\Omega^{-1}\lambda\right)$ correction and an additive correction $\frac{1}{T}\Omega$
- For the test on the pricing errors:

$$T\left(1+\lambda'\Omega^{-1}\lambda\right)^{-1}\hat{\alpha}'_{gls}\hat{\Sigma}^{-1}\hat{\alpha}_{gls}\sim\chi^2_{N-K}$$

- Is the multiplicative correction important?

- Consider CAPM. In annual data: $(\lambda_m/\sigma_m)^2 = (0.08/0.16)^2 = 0.25$
  The correction is important.
  In monthly data $(\lambda_m/\sigma_m)^2 = 0.25/12 \approx 0.02$.
  The multiplicative correction is not important

- The additive correction is likely to be important

- $\frac{1}{T}\Omega$ is the standard error of the mean of the factor, which is non-negligible in case f is, e.g., the market return,

Time-Series vs. Cross-Section

- How are two approaches different?

1. TS can be applied only if factors are returns
   - In CAPM, it fits a line through the pricing errors by forcing a zero pricing error on R_m
   - Test: $\alpha_i = 0$ for all i
2. CS only alternative when factors are not returns
   - In CAPM, it minimizes all pricing errors, by allowing some error on R_m
   - Historically CS has been used to test for characteristics in the cross-section of returns (especially using Fama-MacBeth approach)

Fama-MacBeth (1973) Methodology

- It is a three-pass procedure:

  1. Obtain β_{it} from time-series regressions, using data up to $t-1$
  2. At each date t, run a cross sectional regression

     $$R^e_{it} = \beta'_{it}\lambda_t + \alpha_{it} \qquad i = 1...N$$

     and obtain a time series of $\hat{\lambda}_t$ and $\hat{\alpha}_{it}$ $\quad t = 1...T$
  3. Finally, obtain full-sample estimates as time series means

$$\begin{aligned} \hat{\lambda} &= \frac{1}{T}\sum_{\iota=1}^{T}\hat{\lambda}_t \\ \hat{\alpha}_i &= \frac{1}{T}\sum_{\iota=1}^{T}\hat{\alpha}_{it} \end{aligned}$$

and use the standard error of the mean

$$\sigma\left(\hat{\lambda}\right) = \frac{1}{T^{1/2}}\left(\underbrace{\frac{\sum_{t=1}^{T}\left(\hat{\lambda}_t-\hat{\lambda}\right)^2}{T-1}}_{\widehat{Var(\hat{\lambda}_t)}}\right)^{1/2}$$

$$\sigma\left(\hat{\alpha}_i\right) = \frac{1}{T^{1/2}}\left(\underbrace{\frac{\sum_{t=1}^{T}\left(\hat{\alpha}_{it}-\hat{\alpha}_i\right)^2}{T-1}}_{\widehat{Var(\hat{\alpha}_{it})}}\right)^{1/2}$$

- There are two main advantages to this approach:
  1. It allows for time-varying betas
  2. It computes standard errors by getting around the problem of heteroskedastic and correlated errors: exploit sample variation in $\hat{\lambda}_t$ and $\hat{\alpha}_{it}$
- You may still want to do a Shanken correction because betas are estimated

- If beta are estimated on overlapping data, you need to correct st. errors for time-series correlation (use Newey-West with q=#overlapping observations)

- The method is often used to test for the cross-sectional explanatory power of characteristics z_i

- You can test for zero pricing errors

$$\hat{\alpha}' \left[Var\left(\hat{\alpha}\right)\right]^{-1} \hat{\alpha} \sim \chi^2_{N-1}$$

  where you use sample vcov of the errors

$$Var\left(\hat{\alpha}\right) = \frac{1}{T^2} \sum_{i=1}^{T} \left(\hat{\alpha}_t - \hat{\alpha}\right)\left(\hat{\alpha}_t - \hat{\alpha}\right)'$$

- If β_{it} and z_{it} are constant over time and the errors are uncorrelated over time, one can show that Fama-MacBeth is equivalent to:

  1. OLS cross-sectional regressions with st. errors corrected for correlation as in (19)
  2. Pooled time-series and cross-sectional OLS regressions with st. errors corrected for correlation

Maximum Likelihood

- ML provides justification for tests seen above
- You need to make assumptions on distribution of errors (typically normality)
- The principle is to choose parameters θ to maximize likelihood function $L\left(\{x_t\};\theta\right)$

$$\hat{\theta} = \arg\max_{\theta} L\left(\{x_t\};\theta\right)$$

- ML estimators are asymptotically normal and asymptotically most efficient
- But: properties hold only if economic and statistical models are correctly specified
- Not robust to model misspecification!
- One can show that:
  1. If factors are returns, ML prescribes time-series regressions
  2. If factors are not returns, ML prescribes GLS cross-sectional regressions (with Shanken correction for st. errors): iterative procedure

3. Statistical Discipline and Philosophy

ML is often ignored

- ML prescribes GLS instead of OLS
- The historical development saw OLS regregressions coming first
- ML was used later as a motivation of OLS in some contexts
- However, if ML prescriptions differ from OLS, people continue to use OLS
- Why? Trade off: Efficiency vs. Robustness
- ML is the most (asymptotically) efficient if the model is correct (statistically and economically)
- In small samples, it can be way off the target
- ML is constructed to price a linear combination of the original portfolios. It transforms the original portfolios to something less 'interpretable' (less liquid, implausible long-short positions, etc.)

- Consider a model that works a reasonably well on a ‘reasonable’ set of assets
- OLS is more robust to model mispecification: it is constructed to price the original set of assets
- If you use GLS, you are likely to reject the model, because the model is unlikely to price an illiquid combination of the original set of assets (This is Cochrane’s point of view. Nagel and Singleton (2011, JF) disagree with this point of view)
- If you use OLS, you are less likely to reject a model
- An AP is bound to be imperfect. The question is: how well does it price a ‘reasonable’ set of portfolios (=feasible trading strategies)?
- Then, OLS answers this question

- From efficient set mathematics: Mean variance efficiency of a factor implies a $E(R) = \beta' E(R_q)$ representation, where β are betas on the factor and $E(R_q)$ is the mean return on the factor q, where q is a portfolio on the mean-variance frontier constructed from the (sample or population) moments of the N assets

Proof. Take a portfolio x of the N assets. You have that

$$\beta_{x,q} = \frac{x'Vq}{q'Vq} \tag{21}$$

where x is the vector of weights for the portfolio, q is the vector of weight for a mean-variance efficient portfolio, and V is the variance-covariance matrix of the N assets

To obtain a mean-variance efficient portfolio, it has to be the case that q solves the problem

$$\begin{aligned} & Min\ q'Vq \\ s.t.\quad : \quad & q'\mu = e \end{aligned}$$

where e is a fixed level of expected excess return and μ is the vector of expected excess returns for the N assets. The Lagrangean for the problem is

$$\mathcal{L} = q'Vq + 2\lambda\left(e - q'\mu\right)$$

and the first order conditions are

$$\begin{aligned} Vq - \lambda\mu &= 0 \\ q'\mu &= e \end{aligned}$$

from the first order conditions we obtain

$$Vq = \lambda\mu$$

which we can use in equation (21):

$$\begin{aligned} \beta_{x,q} &= \frac{x'Vq}{q'Vq} \\ &= \frac{\lambda x'\mu}{\lambda q'\mu} \end{aligned}$$

Rearranging and simplifying λ:

$$x'\mu = \beta_{x,q}q'\mu$$

or:

$$E\left(R_x\right) = \beta_{x,q} E\left(R_q\right)$$

■

- This theorem holds whatever distribution one uses to compute moments: objective, subjective, or ex-post (sample) distribution
- Hence: in the sample, there exists an $E(R) = \beta' E\left(R_q\right)$ representation based on sample moments
- That is: there is an ex-post mean-variance efficient portfolio that prices all assets
- Danger of "*Fishing for Factors*"! Finding the portfolio that in-sample works
- Limit to possibility of out-of-sample tests: international data dirty; need to wait for 30 years
- Discipline: a factor needs to be economically motivated

- Modern Approach to AP: Stochastic Discount Factor

$$E_t\left(m_{t+1}R^e_{i,t+1}\right)=0$$

where

$$m_{t+1}=\frac{u'\left(c_{t+1}\right)}{u'\left(c_t\right)}=1-b'f_{t+1}$$

- Different AP models can be derived from different specifications of SDF

- In any case, sensible factors must be related to consumption growth. E.g.: return on wealth portfolio (CAPM), state variables of hedging concern (ICAPM)

Data Snooping

- Related problem
- If you search the data for significant relationships over and over again, you will find something (with 5% chance)
- In principle, you should test model on different data set
- Objectively difficult
- Need to go to the data with economically founded models

Statistical Philosophy (Cochrane, ch. 16)

- The way knowledge has progressed differs from the pure prescriptions of econometric theory
- Historically, it took another model to beat a sensible model. A rejection of $H_0 : \alpha = 0$ is usually not enough
- For example: CAPM was finally rejected only when another model (based on size and B/M) explained returns (Fama and French, 1992 and 1993)
- The most efficient statistical procedure (ML) is not convincing, if it is not transparent
- There are more relevant questions in evaluating a theory than statistical significance:
  - Is the underlying theory sensible?
  - Do the assumptions make sense?
  - Do the empirical variables reflect the variables in the model?

- Robustness to: simplifying assumptions, measurement error, data snooping

- For example: Fama and French 3-factor model is statistically rejected by GRS test, but it is still widely used as a benchmark

- Classical statistics requires that nobody ever looked at the data. But we all use the same data sets (CRSP, Compustat)

- Bayesian statistics could incorporate prior results, but in practice people start with uninformative priors

- Bottom line: what has convinced people historically is not necessarily what was mostly statically sensible, but what looked more 'reasonable'

- Epistemology should not be normative (Popper, Friedman), it should be positive (Kuhn, McCloskey)

- Focus on how ideas convinced people in the past rather than on how ideas should convince people from a statistical point of view!