1 Intro to Linear Factor Models

1.1 What's a Linear Factor Model?

- For our central asset pricing equation

$$p = E(mx),$$

we have a *linear* factor model if we can express the SDF as

$$m = b'f.$$

1.2 Review: E(R)-Beta Representations

- Recall: A linear model for the discount factor is equivalent to an expected-return beta representation...
- The model:
  - Expected Return-Beta Representation of linear factor pricing models

$$E\left(R^i\right) = \gamma + \beta_{i,a}\lambda_a + \beta_{i,b}\lambda_b + \cdots,\ i = 1\ldots N$$

- The general idea:
  - Explain the variation in average returns across assets due to each assets exposure - as measured by the betas - to various risks, as priced by the lambdas.
- Recall how we can use our central asset pricing equation to get a basic pricing equation for returns:

$$\begin{aligned}
1 &= E\left[mR^i\right] \\
1 &= E(m)E\left(R^i\right) + cov\left(m, R^i\right) \\
E\left(R^i\right) &= R^f - R^f cov\left(m, R^i\right) \\
E\left(R^i\right) &= R^f + \left(\frac{cov\left(R^i, m\right)}{var(m)}\right)\left(-\frac{var(m)}{E(m)}\right)
\end{aligned}$$

- Or, equivalently

$$E\left(R^i\right) = R^f + \beta_{i,m}\lambda_m$$

- This derivation gave us an expression for an expected return-beta model...

$$E\left(R^i\right) = R^f + \beta_{i,m}\lambda_m$$

- In the consumption-based model, we figured out that assets whose returns covary positively with consumption growth must have higher expected returns as compensation for risk
  - How can we think about/interpret λ_m?

  - How can we think about/interpret the $\beta_{i,m}$'s?

- So what can we do now?
- The idea is that we're trying to find a proxy or proxies for λ_m
  - It doesn't necessarily have to be consumption growth!
- What we can see now is that if we can find "proxies" for consumption growth, good times/bad times, etc., we will have a "linear factor model" in expected return-beta form!

1.3 A Beta Pricing Model

- Generally speaking, an asset pricing model of the form

$$E\left(R^i\right) = R^f + \beta_{i,m}\lambda_m$$

- Says an expected return should be proportional to its beta - i.e., the regression coefficient $\beta_{i,m}$ of an asset's return R^i on risk factors λ_m
- What is λ_m?

- What is $\beta_{i,m}$?

1.4 Introduction to Evaluating Linear Factor Models

- Estimating and evaluating these types of models are what we will look at next...
- Linear factor models are the most common in empirical work.
  - How should we estimate / evaluate them?
- Next-Up:
  - Times-series and Cross-section Regressions (Yay!)

1.5 Lots of Econometric Techniques

- Same questions:
  - How do we estimate parameters?
    - How do we calculate their standard errors?
  - How can we calculate standard errors of the pricing errors?
  - How can we test the model?
- Recall, we've already addressed these issues in a GMM Framework.
  - Now, we're going to look at them through times-series and cross-section regressions (Yay!)

2 Time-Series Regressions

2.1 Simple One Factor Model

- Let's start with an example of the simplest model - a linear (1) factor model
- We can evaluate the model

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

by running OLS time-series regressions

$$R_t^{ei} = \alpha_i + \beta_i f_t + \varepsilon_t^i, \qquad t = 1, 2, ..., T.$$

- If the factor is also a return - like the excess return on the market portfolio, $f_t = R_t^{em} - R_t^f$ - we have the familiar CAPM.
- The theory says that

$$E\left(R_t^{ei}\right) = \beta_i E(f_t)$$

2.2 Using Excess Returns

- Since the theory says

$$E\left(R_t^{ei}\right) = \beta_i E(f_t)$$

the implications for

$$R_t^{ei} = \alpha_i + \beta_i f_t + \varepsilon_t^i, \qquad t = 1, 2, ..., T.$$

are that **all the regression intercepts α_i should be zero!**

- The regression intercepts here are equivalent to the "pricing errors"
- If the theory is correct, then α_i should $= 0$. If that condition is "true", then our model correctly prices that asset.

2.2.1 Why can we do this? (1)

- For our simple model, we have a factor pricing model with a single factor
  - The factor is an excess return, e.g., $R^{em} = R^m - R^f$
  - And all the test assets are excess returns
- Recall, first: If we are using excess returns

$$E\left(R^i\right) = \gamma + \beta_{i,a}\lambda_a + \beta_{i,b}\lambda_b + \cdots,\ i = 1\ldots N$$

  but

$$E\left(R^{ei}\right) = \beta_{i,a}\lambda_a + \beta_{i,b}\lambda_b + \cdots$$

- I.e., the intercept drops out.

2.2.2 Why can we do this? (2)

- Recall, second: If the factors themselves are returns, like $f = R^{em} = R_t^m - R_t^f$ for the CAPM, the model should apply to the factor as well.
- Then the time-series regression for the factor would be

$$R_t^{em} = \alpha + \beta_{i,m} R_t^{em} + \varepsilon_t^i$$

- So $\beta_{i,m} = 1$ - i.e., the factor has a beta of one on itself
- And we can write an estimate of the factor risk premium as

$$\widehat{\lambda} = E_T(f) = E_T\left(R^{em}\right)$$

2.3 Time-Series Procedure

- Really Easy
  - 1. Estimate the factor risk premium by the sample mean of the factor, e.g.,

$$\widehat{\lambda} = E_T(f) = E_T\left(R^{em}\right)$$

  - 2. Run OLS time-series regressions

$$R_t^{ei} = \alpha_i + \beta_i f_t + \varepsilon_t^i, \qquad t = 1, 2, ..., T$$

    for each test asset.
- If you have 10 test assets, how many regressions do you run? How many β's do you have?

2.3.1 Testing A Single Pricing Error

- What can you use to test whether a pricing error is zero?

- If the regression errors are uncorrelated and homoskedastic, you can use the standard distributions.
  - What if they are not? Do you know how to handle that?

2.3.2 Testing Many Pricing Errors

- Can you anticipate how to test that *all pricing errors are jointly zero*?

- The form is

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

- Very intuitive vis-a-vis what we've seen already
  - The meat of the test $\widehat{\alpha}'\widehat{\Sigma}^{-1}\widehat{\alpha}$ is just a quadratic form of the pricing errors
  - (Details are in the book, if you're interested in the derivation.)

2.3.3 Another Kind of Test

- We've already talked about the fact that the CAPM won't work well if we don't have a good proxy for R^{em}.
- Generally speaking, a single-beta representation (a one factor model!) exists iff the reference return is on the M-V frontier.
- So a test of this model can also be found through a statistic interpreted as a test of whether f is actually *ex ante* mean-variance efficient.

  - Why not a test of f on the *ex post* m-v frontier?

2.3.4 Why a Multi-Factor Model?

- One factor may be insufficient. Why?
- If we have k factors (still excess returns), we can just write

$$E\left(R_t^{ei}\right) = \beta_i' E\left(f_t\right)$$

and use regression equations

$$R_t^{ei} = \alpha_i + \beta_i' f_t + \varepsilon_t^i, \qquad t = 1, 2, ..., T.$$

- The basic difference is that we now have f and β as $k \times 1$ vectors and we get a little bit more algebra.

3 Cross-Sectional Regressions

3.1 Again: The Expected Return-Beta Representation

- Let's again think about an Expected Return-Beta Representation of linear factor pricing models

$$E\left(R^i\right) = \gamma + \beta_{i,a}\lambda_a + \beta_{i,b}\lambda_b + \cdots, \; i = 1 \ldots N$$

- Since we want to explain the variation in average returns **across assets** due to each assets exposure - as measured by the betas - to various risks, as priced by the lambdas...
  - We're now going to look at cross-sectional regressions...
- Think about a scatterplot of average returns verses betas in a single-factor model like the CAPM.
  - Average returns should be proportional to betas
- If we fit a line through the datapoints, the slope of that line would be our estimate of λ
  - Returns won't lie exactly on that line, of course, so we will have a *scatter* plot
  - Deviations from that line are the pricing errors (i.e., the residuals in a cross-sectional regression).

3.2 Cross-Sectional Regressions: Procedure

- A cross-sectional regression just fits a line through such a scatterplot.
- First step: We find estimates of the betas for each of our N assets from the time-series regressions

$$R_t^{ei} = \alpha_i + \beta_i f_t + \varepsilon_t^i, \qquad t = 1, 2, ..., T.$$

- Second step: We then estimate the factor risk premia λ from a regression across assets of average returns on the *betas we just estimated*

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

  - In the second step, what are we trying to estimate? Whare are the RHS variables? What are the pricing errors? Can / should you run this regression without a constant?

3.2.1 About the Estimates

- Estimates:
  - For

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

OLS cross-section estimates are

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

- Distribution Theory:
  - Note: Normally, when we use OLS in a cross-sectional regression, the assumption is that the RHS variables are fixed.
    - The betas (our RHS variables here) are not fixed - we estimated them from the time-series regressions, and this matters for the asymptotic distributions.
  - Your book has various derivations, but the correct asymptotic standard errors are given in Equations (12.19) and (12.20).
    - The correction is due to Shanken (1992), and basically incorporate the variance-covariance matrix of the factors into the correction.

3.2.2 Test of the Model?

- Again, a test of the model is a test of whether or not the pricing errors are close to zero.
  - Here, this can again be done with a chi-square test on the pricing errors - here, those pricing errors are the residuals.
- If that sounds strange, it probably should.
  - How can you test residuals in OLS regressions? What does it mean for the residuals to be zero?
  - Normally, we wouldn't have any information about the residuals, other than the residuals themselves.
  - However, now, the first-stage time-series regressions give us independent information about the size of $cov(\alpha\alpha')$ that we cannot get from looking at the cross-section residuals by themselves.

3.3 Time-Series v. Cross-Section

- How are the approaches different?
  - You can run cross-sectional regressions when the factors are not returns.
  - The time-series regressions require returns so you can estimate factor risk premia by $\widehat{\lambda} = E_T(f)$
- Our asset-pricing model predicts restrictions on the intercepts in the time-series regressions (that we can test, of course).
  - If we impose the restriction $E\left(R^{ei}\right) = \beta_i'\lambda$ we can write the time-series regression as

    $$R_t^{ei} = \beta_i'\lambda + \beta_i'(f_t - E(f)) + \varepsilon_t^i, \qquad t = 1, 2, ..., T,$$

    so the intercept restriction is

    $$a_i = \beta_i'(\lambda - E(f))$$
- This gives us our zero intercept condition (as expected).
  - Interpretation: Mean returns should be proportional to betas; the intercept controls the mean return.
- Even if factors are returns, the two approaches are still not necessarily the same.
- The time-series regression estimates the factor risk premium as the sample mean of the factor.
  - The factor receives a *zero pricing error* in each sample.
- The OLS cross-section regression picks the slope (estimate of the factor risk premium) and intercept to best fit *all the data points.*
- What about GLS?

- If the factor is a return, GLS cross-section is equivalent to the time-series regression since GLS puts all its weight on the asset with the lowest residual variance.
  - If the factor is included as a test asset, it has *zero residual variance.*
  - Interpretation: The "efficient" cross-section regression (GLS) ignores all information in other asset returns and uses only information in the factor returns to estimate the factor risk premium.

3.4 Fama-MacBeth

- Fama-MacBeth (1973) - historically important, computationally simple, and widely used.
- Procedure:
  - First, find estimates of the betas with time-series regressions.
  - Second, run a cross-sectional regression *at each time period.*

$$R_t^{ei} = \beta_i' \lambda_t + \alpha_{it}, \qquad i = 1, 2, ..., N$$

  - (This is done instead of estimating a *single* cross-sectional regression with sample averages.)
  - Third, estimate λ and a_i as averages of the cross-sectional regression estimates

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

- Fama-MacBeth (1973) sampling errors are then

$$\begin{aligned} \sigma^2 \left(\widehat{\lambda} \right) &= \frac{1}{T^2} \sum_{t=1}^{T} \left(\widehat{\lambda}_t - \widehat{\lambda} \right)^2 \\ \sigma^2 \left(\widehat{\alpha}_i \right) &= \frac{1}{T^2} \sum_{t=1}^{T} \left(\widehat{\alpha}_{it} - \widehat{\alpha}_i \right)^2 \end{aligned}$$

- Intuitively appealing, since sampling error is about how a statistic might vary from one sample to the next if observations were repeated.
- Fama-MacBeth uses variation in the statistic $\widehat{\lambda}_t$ over time to deduce its variation across samples.
- Note that this technique does not correct for a time-series that is autocorrelated, but your book gives standard formulas for how to do this.
  - Asset return data are usually not highly correlated
  - Corporate finance data or other regressions where the cross-sectional estimates are highly correlated over time would require this correction.
- Fama-MacBeth standard errors also do not correct for the fact that the betas are generated regressors. Your book also details the Shanken correction for these standard errors are well.
- Historical import: The FM procedure allows for changing betas, which a single unconditional cross-sectional regression or a time-series regression test cannot easily handle.