

## 14 Week 5 Empirical methods notes

1. Motivation and overview
2. Time series regressions
3. Cross sectional regressions
4. Fama MacBeth
5. Testing factor models.

### 14.1 Motivation and Overview

1. Expected return, beta model, as in CAPM, Fama-French APT.

1. TS regression, define  $\beta$  :  $R_t^{ei} = \alpha_i + \beta_i' f_t + \varepsilon_t^i \quad t = 1, 2, \dots, T$ .

2. Model:  $E(R^{ei}) = \beta_i' \lambda (+\alpha_i)$

( $+\alpha_i$ ) because the point of the model is that  $\alpha$  should = 0.

2. 1 is useful to *understand variation over time in a given return*, for example, to find good hedges, reduce variance. 2 is for when we want to *understand average returns in relation to betas (cross section)*
3. How do we take this to the data? Objectives:
  - (a) *Estimate* parameters.  $\hat{\alpha}, \hat{\beta}, \hat{\lambda}$ .
  - (b) *Standard errors* of parameter estimates. How do  $\hat{\alpha}, \hat{\beta}, \hat{\lambda}$  vary if you draw new data and try again?
  - (c) *Test* the model. Does  $E(R^{ei}) = \beta_i' \lambda$ ? Are the  $\alpha = 0$ ? Are the  $\hat{\alpha}$  that we see due to bad luck rather than real failure of the model?
  - (d) *Test* one model *vs. another*. Can we drop a factor, e.g. size?
4. Statistics reminder. There is a true value of  $\alpha, \beta, \lambda$ , which we don't know.
  - (a) We see a sample drawn from this truth, generating estimates  $\hat{\alpha}, \hat{\beta}, \hat{\lambda}$ . These sample values vary from the truth, but they're our best guess.
  - (b) "Standard errors:" If we ran history over and over again, with different luck, how would the estimates  $\hat{\alpha}, \hat{\beta}, \hat{\lambda}$  vary across these alternative universes. "
  - (c) Test" If the true alpha (say) were zero, how often (in how many alternate universes) would we see an alpha as big as the  $\hat{\alpha}$  we measure in this sample? If not many it's unlikely that the true  $\alpha$  is zero.

## 14.2 Statistics review

$x_t$  is a random variable (say, stock return) drawn from a distribution with mean  $\mu$  and variance  $\sigma^2$ . You see a sample  $\{x_t\}$ , meaning  $(x_1, x_2, \dots, x_T)$ . You do not see the true  $\mu, \sigma$ . Your job is to learn what you can about them from the sample.

*Estimate*

You might firm the *sample mean*

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T x_t$$

as a good guess of the mean. This is the *estimate*. We don't have to use this estimate – we could use the median if we wanted to. If so, we'd get different formulas for what follows.

The sample mean  $\hat{\mu}$  is a number to you. But it is also a random variable – it could have come out differently. It's like flipping a coin just once. *Quiz: do you really understand the difference between  $\mu$  and  $\hat{\mu}$ ?*

*Standard error*

The *standard error of the mean*

$$\sigma(\hat{\mu})$$

is your best guess of how differently  $\hat{\mu}$  might have come out if you run history all over again. If you will, it's your best guess of how much the sample mean varies across all alternative universes which are just like ours but different luck of the draw.

How can we guess such a thing, given that we only see one  $\hat{\mu}$ ? You can't take the variance with a single observation! Answer: with assumptions. Start with

$$\sigma^2(\hat{\mu}) = \sigma^2 \left( \frac{1}{T} \sum_{t=1}^T x_t \right) = \frac{1}{T^2} \left[ T\sigma^2(x_t) + 2(T-1)\text{cov}(x_t, x_{t-1}) + \dots \right]$$

(What the... Yes, look at  $T = 3$ , and remember  $\text{cov}(x_1, x_2) = \text{cov}(x_2, x_3) = \text{cov}(x_3, x_2) = \text{cov}(x_t, x_{t-1})$ ,

$$\sigma^2(\hat{\mu}) = \sigma^2 \left( \frac{1}{3} (x_1 + x_2 + x_3) \right) = \frac{1}{9} \left[ 3\sigma^2(x_t) + 2\text{cov}(x_t, x_{t-1}) + 1\text{cov}(x_t, x_{t-2}) \right]$$

Now, let's add an assumption that the sample  $x_t$  is also drawn *independently*. Then all the cov terms are zero, and we get the familiar formula

$$\sigma^2(\hat{\mu}) = \frac{\sigma^2(x)}{T}; \sigma(\hat{\mu}) = \frac{\sigma(x)}{\sqrt{T}}$$

Make sure you understand this. The *standard error of the (sample) mean*, your best guess of how the one sample mean you see would vary if you could run history all over again is related to the *standard deviation of  $x$* , how much  $x$  varies *over time* in *the one sample you do see*. It's rather amazing that we can connect these two quantities!

(What if  $x$  is correlated over time, you ask? Actually, that's easy to solve. Just use the formula with all the cov's in it. This is the "generalized method of moments" formula, which *Asset Pricing* explains in great detail. It's important to recognize: the standard formula  $\sigma/\sqrt{T}$  only works for  $x$  that are uncorrelated over time. There is a better formula that corrects for this correlation. As an

extreme example, if all the  $x$  are perfectly correlated, then you're really seeing one observation not  $T$  observations.)

### *Hypothesis test*

The next thing we might want to do is a *hypothesis test* to characterize our uncertainty about the true  $\mu$ . What we have so far is that  $\hat{\mu}$  is our best guess of the true  $\mu$  and  $\sigma(\hat{\mu})$  measures our uncertainty about that guess. The hypothesis test puts that observation another way: suppose you saw  $\hat{\mu} = 10$ . What is the chance of seeing  $\hat{\mu} = 10$  or larger, just by chance, if the true  $\mu = 0$ ?

To answer that question, let's add another assumption (not really needed, but easier for now), that the  $x_t$  are *normally distributed*. Now, we can rely on a theorem from statistics (which is just algebra) that the sum of normally distributed random variables is also normal. Thus, *if the true values are  $\mu$  and  $\sigma$ , then the sample mean has the distribution*

$$\hat{\mu} \sim N(\mu, \sigma(\hat{\mu}) = \sigma/\sqrt{T})$$

Equivalently, *the sample mean divided by its standard error has a standard normal distribution*

$$\frac{\hat{\mu} - \mu}{\sigma(\hat{\mu})} = \frac{\hat{\mu} - \mu}{\sigma/\sqrt{T}} = \sim N(0, 1)$$

Now you're ready to do a hypothesis test: If the true value  $\mu$  really was zero, form the number  $\hat{\mu}/\sigma(\hat{\mu}) = \hat{\mu}/(\sigma/\sqrt{T})$ . Compare the number you have to the normal (0,1) distribution (table or in matlab) that tells you the chance of seeing something this big or bigger. (A number of about 2 corresponds to about 2.5% chance of seeing a bigger number.)

A fly in the ointment: What do you use for  $\sigma$ ? You don't know that. You could try using your best guess, the *sample standard deviation*

$$\hat{\sigma}^2 = \frac{1}{T} \sum (x_t - \hat{\mu})^2$$

(or  $1/(T-1)$ , the difference doesn't matter here. ) *Quiz: Do you understand the difference between  $\hat{\sigma}$  and  $\sigma$ ?. Using that guess, you'd form the number*

$$\frac{\hat{\mu} - \mu}{\hat{\sigma}/\sqrt{T}}$$

However, looking across samples, (alternate universes)  $\hat{\sigma}$  now varies too, so this number will vary by more than the last number we created. Fortunately, a smart statistician figured out the distribution of this number – the sample mean divided by the *sample* standard deviation. It's called the  $t$  distribution. So we write

$$\frac{\hat{\mu} - \mu}{\hat{\sigma}/\sqrt{T}} \sim t$$

And you compare the number you create on the left hand side to this distribution, slightly fatter than the normal, to figure out what the chance is of seeing a  $\hat{\mu}$  this big (10) *if* the true  $\mu$  were zero.

The  $t$  distribution is exactly correct if the  $x$  are normal. As  $T$  rises, the  $t$  distribution looks more and more like the normal distribution. The  $\hat{\sigma}$  gets close to  $\sigma$  quickly. Even if the underlying  $x$  are not normally distributed, this number still converges to a normal distribution (the “central limit theorem”).

Similarly, statisticians have worked out the distribution of

$$\frac{(\hat{\mu} - \mu)^2}{\hat{\sigma}^2/T}$$

If the  $x$  are normally distributed, this has an  $F$  distribution. (That's just a name for a function, the way "normal distribution" is a name for  $f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2}$ . The form of the function doesn't matter, as your computer has it.) Even if the  $x$  are not normally distributed, for large  $T$  this number has a  $\chi^2$  distribution. This mirrors the facts for the  $t$  and normal distribution.

Why square it? Squaring it allows you to generalize and do multiple tests together. The basic idea: suppose  $x$  and  $y$  are independent of each other with a standard  $N(0, 1)$  distribution. Then

$$x^2 + y^2 \sim \chi_2^2$$

(read as "chi-squared with two degrees of freedom.") That just means that statisticians have tabulated the distribution of a sum of two normals.

Here's how we use that. Suppose  $x$  and  $y$  are both iid with mean  $\mu = \begin{bmatrix} \mu_x & \mu_y \end{bmatrix}$  and covariance matrix  $\Sigma$ . Then

$$\begin{bmatrix} \hat{\mu}_x - \mu_x & \hat{\mu}_y - \mu_y \end{bmatrix}' \left( \frac{\Sigma}{T} \right)^{-1} \begin{bmatrix} \hat{\mu}_x - \mu_x \\ \hat{\mu}_y - \mu_y \end{bmatrix}$$

has a  $\chi_2^2$  (chi-squared with two degrees of freedom) distribution for large  $T$ . If  $x$  and  $y$  are also normally distributed then this number has a  $F$  distribution, which is exact for any  $T$ .

### 14.3 Time-Series Regression

1. Setup:  $N$  test assets.  $K$  factors ( $K = 1$  for CAPM,  $K = 3$  for FF).  $T$  time periods. The factors are also excess returns like  $\text{rmrf}$ ,  $\text{hml}$ ,  $\text{smb}$ , not real variables like  $\Delta c_t$ ,
2. \* It's useful to think of data in vector, matrix format

$$T \text{ time} \begin{bmatrix} & \begin{matrix} N \text{ assets} \\ R_1^1 & R_1^2 & \dots & R_1^N \\ R_2^1 & R_2^2 & \dots & R_2^N \\ & & \vdots & \\ R_T^1 & R_T^2 & \dots & R_T^N \end{matrix} \end{bmatrix}$$

There are  $N$ -vectors *across assets*, e.g.

$$R_t = \begin{bmatrix} R_t^1 & R_t^2 & \dots & R_t^N \end{bmatrix}'$$

$$E(R) = E \begin{bmatrix} R^1 & R^2 & \dots & R^N \end{bmatrix}'$$

and  $K$ -vectors *across factors* e.g.

$$f_t = \begin{bmatrix} f_t^1 & f_t^2 & \dots & f_t^K \end{bmatrix}'$$

$$E(f) = \begin{bmatrix} E(f^1) & E(f^2) & \dots & E(f^K) \end{bmatrix}'$$

(I'll use  $K = 1$  wherever possible). Vectors are column vectors.

Time Series Regression (Fama-French)

1. *Method:*

- (a) *Run* the time series regression for each portfolio.

$$R_t^{ei} = \alpha_i + \beta_i' f_t + \varepsilon_t^i \quad t = 1, 2, \dots, T. \text{ for each } i$$

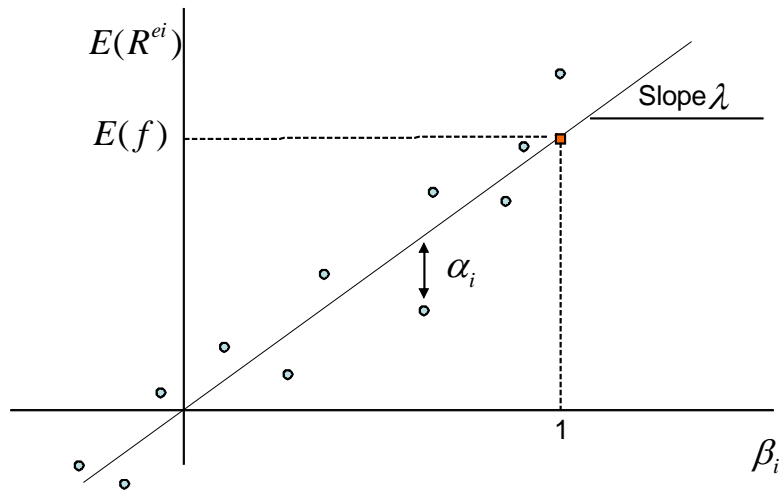
- (b) *Interpret* the regression as a description of the cross section

$$E(R_t^{ei}) = \beta_i' E(f_t) + \alpha_i \quad i = 1, 2, \dots, N$$

2. *Estimates:*  $\lambda$  How do we estimate  $\lambda$  in  $E(R^{ei}) = (\alpha_i) + \beta_i' \lambda$ ?

- (a) A: If the factor is an excess return, then we should see  $\lambda = E(f)$ . Estimate it that way!  
 (b) Why is  $\lambda = E(f)$ ? When  $f$  is an excess return, the model also applies to  $f$ .  $\beta = 1$  of course, so

$$\text{Model: } E(f) = 1 \times \lambda$$



3. *Estimates:*

- (a)  $\hat{\alpha}, \hat{\beta}$  : OLS time-series regression.

$$R_t^{ei} = \alpha_i + \beta_i' f_t + \varepsilon_t^i \quad t = 1, 2, \dots, T \text{ for each } i.$$

- (b)  $\hat{\lambda}$  : Mean of the factor,

$$\hat{\lambda} = \frac{1}{T} \sum_{t=1}^T f_t = \bar{f}.$$

- (c) Stop and admire. *We are here to estimate the cross-sectional relationship between  $E(R^{ei})$  and  $\beta_i$ . We don't actually run any cross sectional regressions. We estimate the slope of the cross-sectional relationship by finding the mean of the factor. We estimate the cross-sectional **error**  $\alpha$  as the time-series **intercept**.*

4. *Standard errors:*

- (a) Reminder of the standard error question: There are true parameters  $\alpha, \beta, \lambda$  which we don't know. In our sample, we produce estimates  $\hat{\alpha}, \hat{\beta}, \hat{\lambda}$ . Due to good or bad luck, these will be different from true values. If we could rewind and run history over and over again, we'd get different values of  $\hat{\alpha}, \hat{\beta}, \hat{\lambda}$ . How much would these vary across the different samples? If we thought  $\alpha = 0$ , how likely is it that the  $\hat{\alpha}$  we see is just due to luck? To answer these questions we need to know what  $\sigma(\hat{\alpha}), \sigma(\hat{\beta}), \sigma(\hat{\lambda})$  are. Watch it – this is *how much the estimates would vary across “alternate universes”*. These quantities are called standard errors.
- (b) Suppose  $\varepsilon_t^i$  are independent *over time*. For each *asset*,  $E(\varepsilon_t^i \varepsilon_{t+s}^i) = 0$ . I do not assume  $E(\varepsilon_t^i \varepsilon_t^j) = 0$  – that would be a big mistake – returns are correlated with each other at a point in time!
- (c)  $\Rightarrow$ (statistics) Then you can use OLS standard errors  $\hat{\alpha}_i, \hat{\beta}_i$ .
- (d)  $\Rightarrow$ (statistics) For  $\hat{\lambda}$ , our old friend: If  $x_t$  are uncorrelated over time,

$$\sigma(\bar{x}) = \frac{\sigma(x)}{\sqrt{T}}$$

Applying it to  $\hat{\lambda}$ ,

$$\sigma(\hat{\lambda}) = \frac{\sigma(f_t)}{\sqrt{T}}$$

5. *Test*

- (a) Reminder of t statistic.  $\hat{\alpha}/\sigma(\hat{\alpha}) = t$ . This shouldn't be too big if the true alpha is zero. “due to luck” it will only be over 2 in 5% of the samples.
- (b) We want to know if all  $\alpha$  are *jointly* zero. We know how to do  $t(\hat{\alpha}_i), \sigma(\hat{\alpha}_i), \sigma(\hat{\alpha}_j)$  – from the i and jth time series regression – but how do you test all the  $\alpha$  *together*?
- (c) Answer: look at

$$\hat{\alpha}' \text{cov}(\hat{\alpha}, \hat{\alpha}')^{-1} \hat{\alpha}$$

This is a number. If the true  $\alpha$  is zero, this should not be “too big.” Precise forms,

$$\hat{\alpha}' \text{cov}(\hat{\alpha})^{-1} \hat{\alpha} = T \left[ 1 + \bar{f}' \Sigma_f^{-1} \bar{f} \right]^{-1} \hat{\alpha}' \Sigma^{-1} \hat{\alpha} \sim \chi_N^2$$

$$\frac{T - N - K}{N} \left[ 1 + \bar{f}' \Sigma_f^{-1} \bar{f} \right]^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} \sim F_{N, T-N-K}$$

GOAL: Understand use this formula, know where to look it up. Not memorize or derive it!

- (d) “ $\sim$ ” means “is distributed as,” meaning the number of the left follows the distribution given on the right. Procedure: compute the *number* on the left. Compare that number to the distribution on the right, which tells you how likely it is to see a number this large, if the true  $\alpha$  are all zero. (Note  $K$  = number of factors only; i.e. 1 for the CAPM. Don't add another  $K$  for the constant. Also the F test wants you to use  $1/T$  in computing the covariance matrices. )

(e) Conceptually, it's just like a t test,

$$\frac{\hat{b}}{\sigma(\hat{b})} \sim t; \frac{\hat{b}^2}{\sigma(\hat{b})^2} \sim \chi^2, F,$$

You compute the *number* on the left, and compare to the t distribution on the right. If the number is bigger than 2, you say there is less than 5% chance of seeing a number this big if the true x is zero. Our test is just the vector version of the squared t test – estimate divided by standard error.

(f) What *is* all this stuff?

i.  $\hat{\alpha}$  = all the intercepts together.

$$\hat{\alpha} = \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \dots \\ \hat{\alpha}_N \end{bmatrix}$$

ii.  $\Sigma$  = covariance matrix of regression residuals.

$$\begin{aligned} \Sigma &= \begin{bmatrix} \sigma^2(\varepsilon_1) & \sigma(\varepsilon_1\varepsilon_2) & \sigma(\varepsilon_1\varepsilon_N) \\ \sigma(\varepsilon_1\varepsilon_2) & \sigma^2(\varepsilon_2) & \sigma(\varepsilon_2\varepsilon_N) \\ & & \ddots \\ & & & \sigma^2(\varepsilon_N) \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{T} \sum_{t=1}^T \varepsilon_{1t}^2 & \frac{1}{T} \sum_{t=1}^T \varepsilon_{1t}\varepsilon_{2t} \\ \frac{1}{T} \sum_{t=1}^T \varepsilon_{1t}\varepsilon_{2t} & \frac{1}{T} \sum_{t=1}^T \varepsilon_{2t}^2 \\ & & \ddots \end{bmatrix} \end{aligned}$$

Most simply,

$$\begin{aligned} \varepsilon_t &= \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \dots \\ \varepsilon_{Nt} \end{bmatrix} \\ \Sigma &= E(\varepsilon_t\varepsilon_t') \end{aligned}$$

$\Sigma$  is *not* diagonal. Returns and residuals are typically correlated *across assets* even if not *over time*.  $E(\varepsilon_t^1\varepsilon_{t-j}^1) = 0$ , but  $E(\varepsilon_t^1\varepsilon_t^2) \neq 0$ . (Industry, other nonpriced factors).

iii.  $\hat{\alpha}'\Sigma^{-1}\hat{\alpha}$  is a *number*. (See “quadratic form” in the matrix notes)

iv.  $\bar{f}$  = mean of the factors ( $\lambda$ );  $\Sigma_f$  = covariance matrix of the factors. .For CAPM,

$$\bar{f} = E(R^{em}); \Sigma_f = \sigma^2(R^{em})$$

For FF3F,

$$\begin{aligned} \bar{f} &= \begin{bmatrix} E(rmr f) \\ E(hml) \\ E(smb) \end{bmatrix} \\ \Sigma_f &= \begin{bmatrix} \sigma^2(rmr f) & \sigma(rmr f, hml) & \sigma(rmr f, smb) \\ \sigma(rmr f, hml) & \sigma^2(hml) & \sigma(hml, smb) \\ \sigma(rmr f, smb) & \sigma(hml, smb) & \sigma^2(smb) \end{bmatrix} \end{aligned}$$

- v.  $\left[1 + \bar{f}'\Sigma_f^{-1}\bar{f}\right]$  is a number.
- vi. Why isn't this in regression class? We are running  $K$  regressions side by side and asking for the joint distribution of coefficients across regressions. In regression class, you only run regressions one at a time.

(g) Interpreting the formula

- i. Even if the true  $\alpha = 0$ , we will see  $\hat{\alpha} \neq 0$  in sample. Firm  $i$  may get lucky, have a high average return *in sample*. But  $\hat{\alpha}$  should not be "too large."
- ii. Negative  $\hat{\alpha}$  is as bad as positive.
- iii. We want *one* number to summarize if the  $N$   $\hat{\alpha}$  are, on average, "too big." How about the sum of squared  $\hat{\alpha}_i$ ?

$$\hat{\alpha}_1^2 + \hat{\alpha}_2^2 + \dots + \hat{\alpha}_N^2 = \hat{\alpha}'\hat{\alpha}$$

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

iv.

$$\hat{\alpha}'\text{cov}(\hat{\alpha}, \hat{\alpha}')^{-1}\hat{\alpha}$$

is a *weighted* sum of squares. For example, if  $\text{cov}(\hat{\alpha}, \hat{\alpha}')$  is diagonal,

$$\begin{aligned} & \begin{bmatrix} \hat{\alpha}_1 & \hat{\alpha}_2 & \dots & \hat{\alpha}_N \end{bmatrix} \begin{bmatrix} \sigma_1^{-2} & & & \\ & \sigma_2^{-2} & & \\ & & \ddots & \\ & & & \sigma_N^{-2} \end{bmatrix} \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \vdots \\ \hat{\alpha}_N \end{bmatrix} \\ & \begin{bmatrix} \hat{\alpha}_1 & \hat{\alpha}_2 & \dots & \hat{\alpha}_N \end{bmatrix} \begin{bmatrix} \frac{\hat{\alpha}_1}{\sigma_1^2} \\ \frac{\hat{\alpha}_2}{\sigma_2^2} \\ \vdots \\ \frac{\hat{\alpha}_N}{\sigma_N^2} \end{bmatrix} \\ & = \left( \frac{\hat{\alpha}_1^2}{\sigma_1^2} + \frac{\hat{\alpha}_2^2}{\sigma_2^2} + \dots + \frac{\hat{\alpha}_N^2}{\sigma_N^2} \right) \end{aligned}$$

This pays more attention to alphas corresponding to assets with low  $\sigma$ . Those  $\hat{\alpha}$  are better measured. For example, if  $\sigma_i = 0$ , then  $\hat{\alpha}$  is perfectly measured – we should really pay attention to that one, and reject the model if it is even 0.00000001!

v. Fact:

$$\text{cov}(\hat{\alpha}, \hat{\alpha}') = \frac{1}{T} \left[ 1 + E(f)' \Sigma_f^{-1} E(f) \right] \Sigma$$

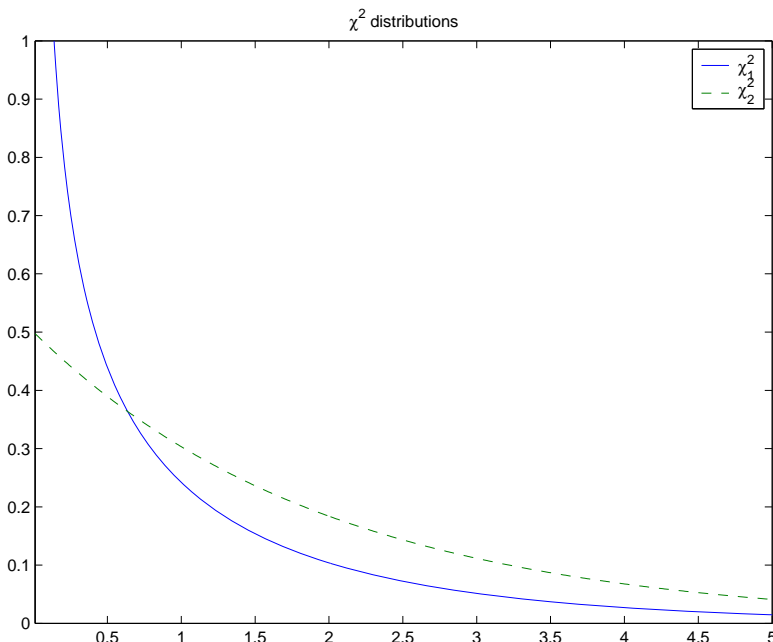
Thus our statistics *are* based on the size of

$$\hat{\alpha}'\text{cov}(\hat{\alpha}, \hat{\alpha}')^{-1}\hat{\alpha}$$

- vi. If the  $\alpha$  are really zero,  $\hat{\alpha}$  will not be zero due to luck, and  $\hat{\alpha}'\text{cov}(\hat{\alpha}, \hat{\alpha}')^{-1}\hat{\alpha} > 0$ . How large will  $\hat{\alpha}'\text{cov}(\hat{\alpha}, \hat{\alpha}')^{-1}\hat{\alpha}$  typically be? How often should we see 3? 5?



- vii. Statistics to the rescue. Statisticians have figured out the distribution of  $\hat{\alpha}' cov(\hat{\alpha}, \hat{\alpha}')^{-1} \hat{\alpha}$  assuming  $\hat{\alpha}$  is normal, i.e. “how large it should be.” If  $x$  is normal with  $E(x) = 0$ , then  $\frac{x^2}{\sigma^2}$  is  $\chi_1^2$ . If  $x_1, x_2$  are independent normal, then  $\frac{x_1^2}{\sigma_1^2} + \frac{x_2^2}{\sigma_2^2} \sim \chi_2^2$ . If  $x$  (N dimensional vector) is normal, then  $x' cov(x, x')^{-1} x$  is  $\chi_N^2$



- (h) Putting this all together,

$$\hat{\alpha}' cov(\hat{\alpha}, \hat{\alpha}')^{-1} \hat{\alpha} \sim \chi_N^2$$

or, explicitly,

$$T \left[ 1 + E(f)' \Sigma_f^{-1} E(f) \right]^{-1} \hat{\alpha}' \Sigma^{-1} \hat{\alpha} \sim \chi_N^2$$

( $\sim$  means “is distributed as”)

- (i) This works for large samples. If  $\varepsilon$  are also normal (not such a good assumption) then a refinement is valid in small samples.

$$\frac{T - N - K}{N} \left[ 1 + \bar{f}' \hat{\Sigma}_f^{-1} \bar{f} \right]^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} \sim F_{N, T-N-K}$$

The difference between the formulas: This one takes into account the fact that  $\Sigma, \Sigma_f$  vary across samples. This is the “GRS Test.” (Note you should use  $1/T$  for  $\Sigma$  and  $\Sigma_f$  here, not  $1/(T-1)$  or something else.  $\Sigma = \frac{1}{T} \sum_{t=1}^{\infty} \varepsilon_t \varepsilon_t'$  and  $\Sigma_f = \frac{1}{T} \sum_{t=1}^{\infty} (f_t - \bar{f})(f_t - \bar{f})'$ .)

- (j) Notice that  $\hat{\alpha}' \Sigma^{-1} \hat{\alpha}$  is the squared Sharpe ratio you can get by trading in the errors. (Remember the APT). In reality, this Sharpe ratio should be zero as the errors should be zero. However, good and bad luck will make it seem that you could have made money, “I should have bought Microsoft in 1982.” But if the alphas are really zero, there is a limit to how good this 20/20 hindsight Sharpe ratio can be, and that’s the basis of the GRS test. “How much do I beat the market by if I take the sample  $\hat{\alpha}$  as real?”
- (k) Motivating example – why do we need to consider all the  $\alpha$  together? Suppose

$$\hat{\alpha}_1 = 1, \hat{\alpha}_2 = 1$$

and

$$\sigma(\hat{\alpha}_1) = \sigma(\hat{\alpha}_2) = 1$$

Then

$$t = \frac{\hat{\alpha}}{\sigma(\hat{\alpha})} = 1$$

so each  $\alpha$  alone is not significant. Is the model right? Maybe the alpha of a equal weight portfolio of the two assets is significant?  $\hat{\alpha}_1 + \hat{\alpha}_2$  is significant? Let's check. The  $t$  for  $\hat{\alpha}_1 + \hat{\alpha}_2$  is

$$t = \frac{\hat{\alpha}_1 + \hat{\alpha}_2}{\sigma(\hat{\alpha}_1 + \hat{\alpha}_2)}$$

Now,

$$\sigma^2(\hat{\alpha}_1 + \hat{\alpha}_2) = \sigma^2(\hat{\alpha}_1) + \sigma^2(\hat{\alpha}_2) + 2cov(\hat{\alpha}_1, \hat{\alpha}_2)$$

Suppose

$$cov(\hat{\alpha}_1, \hat{\alpha}_2) = -0.9$$

Then

$$\sigma^2(\hat{\alpha}_1 + \hat{\alpha}_2) = 1 + 1 - 2 \times 0.9 = 0.2$$

$$\sigma(\hat{\alpha}_1 + \hat{\alpha}_2) = \sqrt{0.2} = 0.44$$

$$t = \frac{\hat{\alpha}_1 + \hat{\alpha}_2}{\sigma(\hat{\alpha}_1 + \hat{\alpha}_2)} = \frac{2}{0.44} = 4.54$$

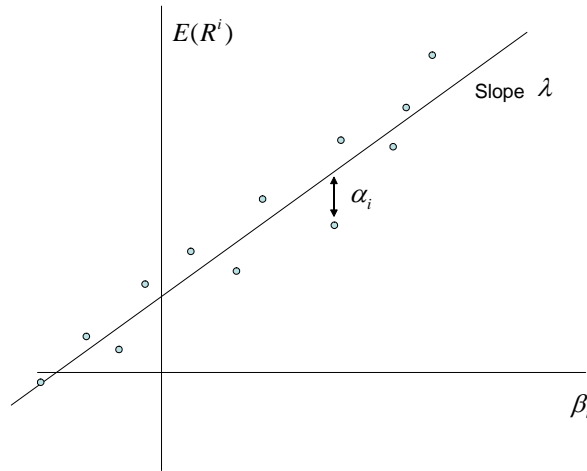
This *is* significant.

- (l) What happened? The model can be rejected for the *portfolio*  $R^{e1} + R^{e2}$  even though not for the individual assets. This *portfolio* alpha is much *better measured* than the individual  $\alpha$  because of the negative covariance. If we had done the initial test on the portfolio  $R^{e1} + R^{e2}$  rather than on the two  $R^e$  separately, we would have gotten a rejection. We don't want our test to depend on which portfolios we happen to choose! The test looks over all possible portfolios and makes sure we don't miss even the worst one. This portfolio also has a lot less variance. Well, standard error is  $\sigma/\sqrt{T}$ , so the minimum-variance portfolios are the best measured ones.
  - (m) Why do we need a new formula? Why wasn't this in regression class? Answer: In regression class, you learned how to do the standard errors of a *single* regression,  $R_t^{ei} = \alpha_i + \beta_i f_t + \varepsilon_t^i$ , for one  $i$  at a time. You learned "joint" distributions and tests like  $\alpha_i$  and  $\beta_i$  together. "Joint" meant multiple right hand variables with a single left hand variable. Here we are running multiple regressions – many left hand variables – next to each other. We need to think about the joint distribution of  $\alpha_i$  and  $\alpha_j$  when we're running two regressions at a time. Econometrics books just never thought of doing that. The methods and formulas are exactly the same as in your OLS regression classes though, once we ask that new question.
6. *BIG PICTURE*: We have a procedure for estimating the parameters  $\alpha, \beta, \lambda$ , for standard errors  $\sigma(\hat{\alpha}), \sigma(\hat{\beta}), \sigma(\hat{\lambda})$  and a test for "are all  $\alpha = 0$ " based on the idea that the sum of squared  $\alpha$  should not be "too big."
  7. Read Fama and French, "Multifactor Anomalies" p. 57 last paragraph, "the F test of..." This is what they're talking about! High  $R^2$  means small  $\Sigma$  so small  $\alpha$  are "significant."

8. All these statistics, and especially the GRS test, are derived from the classical assumptions that returns are independent over time, correlated across assets, and (for finite sample results) normally distributed. There are more modern formulas that account for correlation of asset returns over time, conditional heteroskedasticity (returns are more volatile at some times than others) and non-normality. They aren't that hard – they're all instances of GMM. Asset Pricing explains in detail. Though people should use these formulas, they largely don't, alas.

## 14.4 Cross-sectional regression

1. Another idea. The main point of the model is to understand average returns *across assets*,



$$E(R^{ei}) = \beta_i' \lambda + \alpha_i \quad i = 1, 2, \dots, N$$

Why not fit this as a *cross sectional regression*, rather than just infer  $\lambda$  from the factor mean? (More motivation, and time -series vs. cross section comparison at the end.)

2. *Method:* This is a two step procedure

- (a) Run TS (over time for each asset) to get  $\beta_i$ ,

$$R_t^{ei} = a_i + \beta_i f_t + \varepsilon_t^i \quad t = 1, 2, \dots, T \text{ for each } i.$$

( $a$  not  $\alpha$ . Let's reserve  $\alpha$  for the pricing error  $E(R^e) = \alpha + \beta\lambda$ , and it's not clear yet what  $\alpha$  will be. )

- (b) Run CS (across assets) to get  $\lambda$ .

$$E(R^{ei}) = (\gamma) + \beta_i \lambda + \alpha_i \quad i = 1, 2, \dots, N$$

$\gamma$  is optional.

3. Don't get confused.  $E(R^i)$  is the  $y$  variable.  $\beta_i$  is the  $x$  variable.  $\lambda$  is the slope (not  $\beta$ ).  $\alpha_i$  is the error term. (Potential confusion:  $\alpha_i$ , the *intercept* of the time series regression, is the *error* of the cross-sectional regression.)

4. *Estimates:*

- (a)  $\hat{\beta}$  from TS.
- (b)  $\hat{\lambda}$  slope coefficient in CS.
- (c)  $\hat{\alpha}$  from error in CS:  $\hat{\alpha} = \frac{1}{T} \left( \sum_{t=1}^T R_t^e \right) - \hat{\lambda} \hat{\beta}$ .  $\hat{\alpha} \neq a$  is not the intercept from the time series regression any more.

5. Standard errors.

- (a)  $\sigma(\hat{\beta})$  from TS, OLS regression formulas.
- (b)  $\sigma(\hat{\lambda})$ . With no intercept in CS,

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

- (c)  $cov(\hat{\alpha})$

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

- (d) Yuk yuk yuk... why is this so bad?

- i. The usual regression is  $Y_i = X_i b + \varepsilon_i$ . The standard error formulas assume the  $\varepsilon_i$  are not correlated with each other. Our regression is  $E(R^{ei}) = \beta_i \lambda + \alpha_i$ . The  $\alpha_i$  are correlated with each other. If Ford has an unluckily low alpha, then GM is also likely to have an unluckily low alpha. We need standard errors that correct for correlation across assets.
- ii. This is a pervasive problem, so important to think about. *Most regressions you run in finance have this problem.* Corporate example: does investment respond to market values or to profits (“internal cash”)?

$$\text{investment}_{it} = a + b \times \text{Book/Market}_{it} + c \times \text{profits}_{it} + \varepsilon_{it}$$

Whatever other shocks  $\varepsilon_{it}$  make Ford likely to invest more, also probably affect GM. Fama’s rule: *standard errors and t statistics that don’t correct for this are off by about a factor of 10.*

- iii. We also need to correct for the fact that  $\beta_i$  are also estimated, in the same sample.
- (e) Now, look at the formula. It’s really not that bad.

- i. Ingredient #1. Before we had  $\sigma^2(\hat{\lambda}) = \sigma^2 \left( \frac{1}{T} \sum_{t=1}^T f_t \right) = \Sigma_f / T$ . We have that again! (Last term)
- ii. Ingredient #2. Run a classic regression *with correlated errors*. (OLS regression with correlated errors, not GLS regression.)

$$\begin{aligned} Y &= Xb + u; \quad E(u u') = \Omega \\ \hat{b} &= (X'X)^{-1} X'Y \\ \hat{b} &= (X'X)^{-1} X' (Xb + u) \\ E(\hat{b}) &= b! \text{ (unbiased)} \\ \sigma^2(\hat{b}) &= E \left[ (X'X)^{-1} X' u u' X (X'X)^{-1} \right] \\ &= (X'X)^{-1} X' \Omega X (X'X)^{-1} \end{aligned}$$

That’s just what we have, with  $X = \beta$  (remember,  $\beta$  are right hand variables!).

- iii. This is a really useful tool! It's the basis for "clustered standard errors" and other ways to correct standard errors for correlation.
- iv. Similarly,  $cov(\hat{\alpha})$  formula is also the classic regression formula for covariance of the residuals ( $\alpha$  are *errors* in the CS regression) For math nerds who really want to see it:

$$\begin{aligned}\hat{u} &= Y - X\hat{\beta} \\ &= Y - X(X'X)^{-1}X'Y \\ &= (I - X(X'X)^{-1}X')Y \\ &= (I - X(X'X)^{-1}X')(X\beta + u)\end{aligned}$$

$$\begin{aligned}E(\hat{u}) &= (I - X(X'X)^{-1}X')X\beta \\ &= (X - X(X'X)^{-1}X'X)\beta = 0\end{aligned}$$

$$\begin{aligned}cov(\hat{u}, \hat{u}') &= E\left[(I - X(X'X)^{-1}X')u u' (I - X'(X'X)^{-1}X)\right] \\ &= (I - X(X'X)^{-1}X')\Omega(I - X'(X'X)^{-1}X)\end{aligned}$$

- (f) Yucky formulas, but look them up when you need them. Point: understand them well enough to look them up and calculate them.
- (g) With an intercept  $\gamma$ , ( $E(R^{ei}) = \gamma + \beta_i\lambda + \alpha_i$ ) we need  $\sigma(\hat{\gamma})$  too. This will work just like OLS with a constant Let

$$X = \begin{bmatrix} 1 & | \\ 1 & \beta \\ 1 & | \end{bmatrix}$$

i.e the right hand variables in the cross sectional regression. Then,

$$\sigma^2 \left( \begin{bmatrix} \hat{\gamma} \\ \hat{\lambda} \end{bmatrix} \right) = \frac{1}{T} \left[ (X'X)^{-1} X' \Sigma X (X'X)^{-1} (1 + \lambda' \Sigma_f^{-1} \lambda) + \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_f \end{bmatrix} \right]$$

$$cov(\hat{\alpha}) = \frac{1}{T} (I - X(X'X)^{-1}X') \Sigma (I - X(X'X)^{-1}X') (1 + \lambda' \Sigma_f^{-1} \lambda)$$

#### 6. Test

$$\hat{\alpha}' cov(\hat{\alpha}, \hat{\alpha}')^{-1} \hat{\alpha} \sim \chi_{N-K-1}^2.$$

- 7. Point: understand the question these formulas are answering, a bit of why they look so bad (and why OLS formulas are wrong/inadequate.) Know where to go look them up.

### 14.5 Fama - MacBeth procedure.

1. A relative of cross - sectional regression.
2. Procedure, applied to asset pricing models:

(a) Run TS to get betas.

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

(b) Run a cross sectional regression *at each time period*,

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

(Again,  $\beta$  is the  $x$  and  $\lambda_t$  is the slope. You can add a constant if you want.)

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

(c) Then, *estimates* of  $\lambda, \alpha$  are the *averages across time*

$$\hat{\lambda} = \frac{1}{T} \sum_{t=1}^T \hat{\lambda}_t; \quad \hat{\alpha}_i = \frac{1}{T} \sum_{t=1}^T \hat{\alpha}_{it}$$

(d) *Standard errors* use our friend  $\sigma^2(\bar{x}) = \sigma^2(x)/T$  (Assumes uncorrelated over time – ok for returns but not for many other uses!)

$$\begin{aligned} \sigma^2(\hat{\lambda}) &= \frac{1}{T} \text{var}(\hat{\lambda}_t) = \frac{1}{T^2} \sum_{t=1}^T (\hat{\lambda}_t - \hat{\lambda})^2 \\ \text{cov}(\hat{\alpha}) &= \frac{1}{T} \text{cov}(\hat{\alpha}_t) = \frac{1}{T^2} \sum_{t=1}^T (\hat{\alpha}_{it} - \hat{\alpha}_i)(\hat{\alpha}_{jt} - \hat{\alpha}_j) \end{aligned}$$

*This one main point. These standard errors are easy to calculate.* Back in 1972, we didn't have the huge formulas above, and this was the only way. Now you can do it either way. It's a mind blowing idea – use  $\sigma/\sqrt{T}$  applied to the monthly cs regression coefficients! But it works.

(e) *Test*

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

3. Intuition: To assess  $\sigma(\hat{\lambda})$  you could split sample in half. Or in quarters? ... Or go all the way.
4. Fact: If the  $\beta$  are the same over time, the FMB estimates are *identical* to the cross-sectional regression. The CS standard error formulas are just a little bit better (they include error of estimating  $\beta$ ), but the difference is very small in practice.
5. One advantage of FMB: it's easy to do models in which the betas *change* over time. That's harder to incorporate in time-series and cross-sectional regressions. (Still it's perfectly possible)
6. Other applications of Fama-MacBeth: any time you have a big cross-section, which may be correlated with each other. (If the errors are uncorrelated across time and space, just stack it up to a huge OLS regression, as on the problem set)

(a)

$$R_{t+1}^{ei} = a + b \ln \text{size}_{it} + c \ln \text{beme}_{it} + \varepsilon_{t+1}^i$$

This is an obvious candidate, since the errors are correlated across stocks but not across time. FF used FMB regressions in the second paper we read.

- (b) Corporate finance does this all the time.

$$\text{investment}_{it} = a + b \times \text{Book/Market}_{it} + c \times \text{profits}_{it} + \varepsilon_{it}$$

- (c) *You can't just stack it all up and run a big OLS regression.* Well, you can – the estimates are ok – but you can't use the standard errors, because *the  $\varepsilon$  are correlated with each other.* Think of the limit that the  $\varepsilon$  are all the same. Now you really have  $T$  data points, not  $T \times N$  data points. FMB is one way to correct the standard errors. “Clustering” which is like the cross-sectional regression formulas above is another.
- (d) *This is a big deal* in a recent survey Kellogg's Mitch Petersen found that more than half of all published papers in the journal of finance ignored the fact, and that standard errors are typically off by a factor of 10. Half the stuff in the JF is wrong!

## 14.6 Testing one model vs. another

1. Example. FF3F.

$$E(R^{ei}) = \alpha_i + b_i \lambda_{rmrf} + h_i \lambda_{hml} + s_i \lambda_{smb}$$

Do we really need the size factor? Or can we write

$$E(R^{ei}) = \alpha_i + b_i \lambda_{rmrf} + h_i \lambda_{hml}$$

and do as well? ( $\alpha$  will rise, but will they rise “much”)

2. A common misconception: Measure  $\lambda_{smb} = E(smb)$ . If  $\lambda_{smb} = 0$  (and “small”) we can drop it. Why is this wrong? Because if you drop *smb* from the regression,  $b_i$  and  $h_i$  also change!

- (a) Example 2: CAPM works well for size portfolios, but shows up in FF model

$$E(R^{ei}) = 0 + \beta_i \lambda_{rmrf} \leftarrow \text{works, } \beta_i \text{ higher with higher } E(R^{ei})$$

$$E(R^{ei}) = 0 + b_i \lambda_{rmrf} + s_i \lambda_{sml} \leftarrow \text{works too, all } b_i = 1 \text{ and } \lambda_{sml} > 0$$

- (b) Example 3. Suppose the CAPM is right, but you try a “Microsoft” factor.

$$E(R^{ei}) = \beta_i E(R^{em}) + \beta_{i,ms} E(R^{ms})$$

$E(R^{ms}) = \beta_{ms} \lambda_m > 0$  so it looks like we need the new factor!

- (c) Example 4: Suppose

$$smb_t = \frac{1}{2} rmrf_t + \frac{1}{2} hml_t$$

Now, obviously, we can drop *smb* as a factor. But just as obviously,  $\lambda_{smb} = E(smb_t) = \frac{1}{2} E(rmrf_t) + \frac{1}{2} E(hml_t) > 0$ .

3. Reminder: Multiple regression coefficients don't change if right hand variables are not correlated. Conclusion:  $\lambda_{smb} = 0$  would be right *if* *rmrf*, *hml*, *smb* were not correlated. They are correlated alas.

- (a) In my example,  $\beta > 0$  and correlation of microsoft with the market was crucial. In example 4, you can see the correlation of *smb* and the other factors.

(b) If we orthogonalize the right hand variables

$$\begin{aligned} R_t^{ms} &= \alpha_{ms} + \beta_{ms}R_t^{em} + \varepsilon_t^{ms} \\ R_t^{ei} &= \alpha_i + \beta_i R_t^{em} + \beta_{i,ms} (R_t^{ms} - \beta_{ms}R_t^{em}) + \varepsilon_t^i \\ E(R_t^{ei}) &= \alpha_i + \beta_i E(R_t^{em}) + \beta_{i,ms} E(R_t^{ms} - \beta_{ms}R_t^{em}) \end{aligned}$$

“Orthogonalize” means that by construction  $R_t^{em}$  and  $(R_t^{ms} - \beta_{ms}R_t^{em})$  are uncorrelated. In this case it’s clear that the CAPM holds if the mean of the new factor  $f_t = R_t^{ms} - \beta_{ms}R_t^{em}$  is zero.

4. Solution:

(a) “First run a regression of  $smb_t$  on  $rmrf_t$  and  $hml_t$  and take the residual,

$$smb_t = \alpha_{smb} + b_s rmrf_t + h_s hml_t + \varepsilon_t$$

Now, *we can drop smb from the three factor model if and only if  $\alpha_{smb}$  is zero.* Intuitively, if the other assets are enough to price smb, then they are enough to price anything that smb prices.

(b) “Drop smb” means the 25 portfolio alphas are the same with or without smb

(c) \*Equivalently, we are forming an “orthogonalized factor”

$$smb_t^* = \alpha_{smb} + \varepsilon_t = smb_t - b_s rmrf_t - h_s hml_t$$

This is a version of smb purged of its correlation with rmrf and hml. Now it is ok to drop  $smb$  if  $E(smb^*)$  is zero, because the  $b$  and  $h$  are not affected if you drop  $smb^*$

(d) \*Why does this work? Think about rewriting the original model in terms of  $smb^*$ ,

$$\begin{aligned} R_t^{ei} &= \alpha_i + b_i rmrf_t + h_i hml_t + s_i smb_t + \varepsilon_t^i \\ &= \alpha_i + (b_i + s_i b_s) rmrf_t + (h_i + s_i h_s) hml_t + s_i (smb_t - b_s rmrf_t - h_s hml_t) + \varepsilon_t^i \\ &= \alpha_i + (b_i + s_i b_s) rmrf_t + (h_i + s_i h_s) hml_t + s_i smb_t^* + \varepsilon_t^i \end{aligned}$$

The other factors would now get the betas that were assigned to smb merely because smb was correlated with the other factors. *This part of the smb premium can be captured by the other factors, we don’t need smb to do it. The only part that we need smb for is the last part. Thus average returns can be explained without smb if and only if  $E(smb^*) = 0$ .*

5. \*Other solutions (equivalent)

(a) Drop smb, redo, test if  $\alpha' cov(\alpha)^{-1} \alpha$  rises “too much.”

(b) Express the model as  $m = a - b_1 rmrf - b_2 hml - b_3 smb$ ,  $0 = E(mR^e)$ . A test on  $b_x$  is a test of “can you drop the extra factor.”

6. Why do FF keep smb?

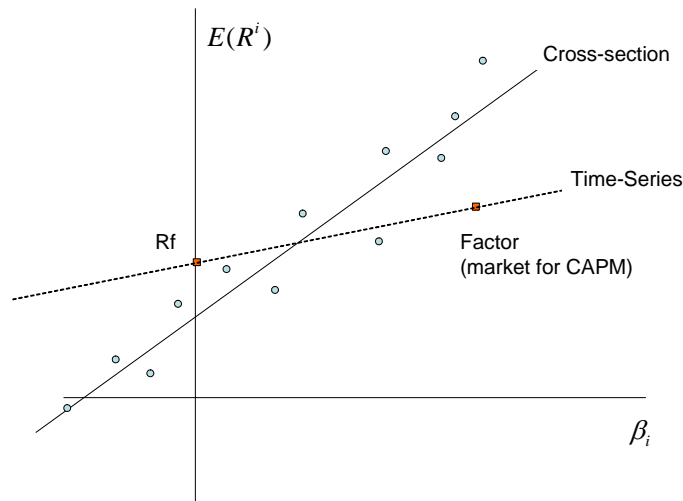
(a) Because they want a good factor “model of *returns* and average returns.” Smb can be very useful *to explain variance* even if its addition does not help *to explain means*.



- (b) Because adding smb reduces  $\sigma(\varepsilon)$  and  $\Sigma$ , making all measurements better and standard errors smaller
- (c) Example: Suppose  $R^2 = 100\%$  in the time series regression with smb, but 90% without it. Then knowing that all your test assets are different portfolios of smb, and all their pricing comes down to arbitrage pricing with smb is very useful!
- (d) Moral: what factors you want to use depends on what you want to use the model for!

### 14.7 Time series vs. cross section

1. What is the difference between TS and CS/FMB? *The cross sectional relation implied by the time series regression goes through the factor and  $R^f$  exactly, ignoring the other assets. The cross sectional regression minimizes the sum of squared  $\alpha$  across all assets.*



- (a) Why is this the “time-series” line? Run the time series regression for  $f$  and for  $R^f$ . *These assets have zero alpha. Thus, the cross sectional line implied by the time series regression goes through  $f$  and  $R^f$*

$$\text{TS} : f_t = 0 + 1 \times f_t + 0$$

$$\text{TS} : R_t^f - R_t^f = 0 + 0 \times f_t + 0$$

Take the CAPM as an example

$$E(R^{ei}) = \beta_{im} \lambda_m$$

If we run a time series regression and then take its mean, you get the cross-sectional implications of the time series regression,

$$\begin{aligned} R_{t+1}^{ei} &= \alpha_i + \beta_{im} R_{t+1}^{em} + \varepsilon_{t+1}^i \\ E(R^{ei}) &= \alpha_i + \beta_{im} E(R^{em}) \end{aligned}$$

Look:

- i.  $E(R^{em})$  is the slope  $\lambda$ .

- ii.  $\alpha_i$ , the *intercept* of the time series regression is now the *error* of the cross-sectional relation
  - iii. This cross-sectional line goes right the  $R_f$  and  $R_m$  with zero alpha;  $E(R^{em}) = 0 + 1 \times E(R^{em})$  and  $R^{ef} = R^f - R^f$  has  $0 = 0 + 0 \times E(R^{em})$
- (b) Issue 1: do you want to estimate the factor risk premium  $\lambda$  from the mean of the factor, or to best fit the cross section of returns?
- (c) Issue 2: Do you want to force the intercept to go through the risk-free rate? (Run cross-section regression with or without free constant  $\gamma$ )
- (d) Equivalent: Do you want to set  $\alpha = 0$  for the market and riskfree, or accept some bigger  $\alpha$  there to make other  $\alpha$  small?
- (e) Note: for a good model, it shouldn't really matter.
- (f) "Right answer?" Sorry! Econometrics is a lot of tools and you have to use judgement!
2. Why might you use a cross-sectional regression? Art!
- (a) It's closer to heart of the asset pricing model. "Average returns are higher if beta is higher."
- (b) It's useful to add characteristics and test if they are zero.

$$E(R^{ei}) = (\gamma) + \beta_i \lambda + c(\text{size}_i) + d(b/m_i) + f(\sigma_i^2) + \alpha_i \quad i = 1, 2, \dots, N$$

$c, d, f$  should be zero. Do betas drive out characteristics?

- (c) Fact: If the model is absolutely right, if the factor is an excess return (like CAPM, FF3F), the market proxy is perfect, if the interest rate proxy is the real risk free rate, and if returns are normally distributed, the time series regression (run the CAPM line through  $R^m$  and  $R^f$ , ignore other assets;  $\lambda_{hml} = E(hml)$ ,  $\lambda_{smb} = E(smb)$ , etc. ) is statistically optimal.
- (d) \*Fact: The time series regression is exactly the same as a *generalized least squares* GLS estimate of the cross sectional regression.  $\lambda_{GLS} = (\beta' \Sigma^{-1} \beta)^{-1} \beta' \Sigma^{-1} E(R^e)$ .  $\Sigma_{ii} = 0$  for the factors and risk free rate, since there is no error in their time-series regressions. (Same fact, GLS is statistically optimal)

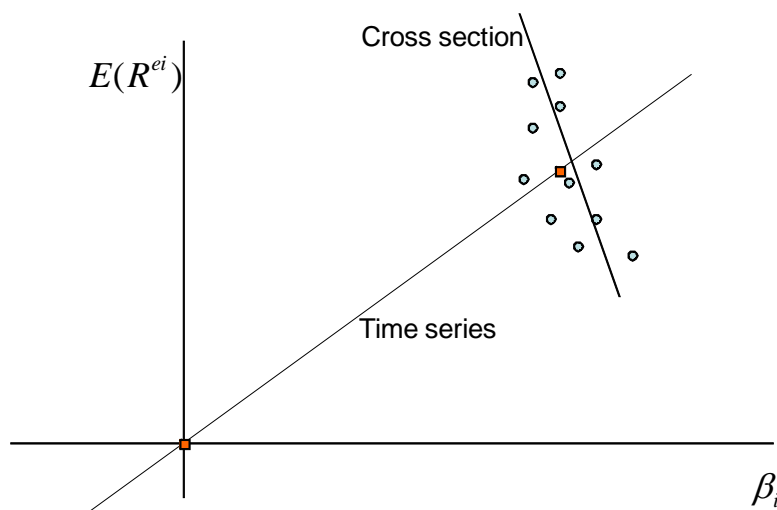
Why should you ignore all the information in the other alphas?

$$R_t^{ei} = a_i + \beta_i f_t + \varepsilon_t^i$$

$E(R^{ei})$  has no extra information about  $E(f)$  that  $f$  does not already have. The same measurement plus error does not refine a measurement.

- (e) But... Advice 1: when models are not perfect, it's often a good idea to run OLS (with correct standard errors) rather than "statistically optimal" GLS. Here: is your market proxy the real wealth portfolio? Do people borrow and lend freely at the 3 month T bill rate? Etc.
- (f) But... betas are not perfectly measured. Mis-measured right hand variables leads to "too flat" regression lines ("downward bias") – just what we see.  $\beta = 1$  on market is perfectly measured!

- (g) Advice 2: It should not make much difference. If it does, it's a sign of trouble with the model. Example: Negative premiums. very high or very negative cross-sectional intercept. The “small alphas” are very dubious here. No fancy black box principle is going to help you here:



3. Cross-sectional regressions are often used when the factor is not a return (consumption, labor income, “liquidity.” etc.) *In this case the intercept in the time series regression need not be zero.*

- (a) (Review and clarification). How is the “model” different from the average of the TS regression?

$$\begin{aligned} \text{TS regression:} \quad R_t^{ei} &= \alpha_i + \beta_i' f_t + \varepsilon_t^i \\ E(\cdot): \quad E(R_t^{ei}) &= \alpha_i + \beta_i' E(f_t) \end{aligned}$$

- (b) *If the factor is also an excess return (rmrf, hml, smb), then the model should apply to the factor as well,*

$$E(R_t^{ei}) = \beta_i' \lambda \rightarrow E(f) = 1 \times \lambda \rightarrow E(R_t^{ei}) = \beta_i' E(f)$$

In this case, the slope  $\lambda$  should equal the factor premium = factor mean. In this case, the force of the model is that time-series alphas should be zero.

Intuition: *Since  $f$  is a return, then  $R^{ei} - \beta_i f$  is also a return with mean  $\alpha_i$*

- (c) *If the factor is not also an excess return ( $\Delta c$ , etc.) then this intuition does not work. Then, you cannot conclude that  $E(f) = \lambda$ , so the intercept in the time-series regression need not be zero. We leave the model at  $E(R^{ei}) = \beta_i' \lambda$*
- (d) Important point: *the time-series regression intercept alpha is only a measure of mispricing if the right hand variable is an (excess) return. Intuition: how do you profit from alpha? Buy  $R_t^{ei}$ , short  $\beta_i f_t$ , then your portfolio return is  $\alpha_i + \varepsilon_t^i$ , with mean  $\alpha_i$  and volatility  $\sigma(\varepsilon_t^i)$ . But you can't short  $\Delta c$ , you can only short things that are returns!*

- (e) \*Technical note: You can still run a time series regression when the factor is not a return. This is a technical note because nobody ever does it, mostly because I don't think they know they can do it. Write the time series regression

$$R_t^{ei} = a_i + \beta_i f_t + \varepsilon_{it}$$

( $a$  not  $\alpha$  because this intercept is not necessarily a pricing error.) The model is

$$E(R^{ei}) = \alpha_i + \beta_i \lambda$$

and  $\alpha$  should be zero if the model is right. Contrasting the model with the expected value of the time series regression, we have

$$E(R^{ei}) = a_i + \beta_i E(f) = \alpha_i + \beta_i \lambda$$

$$\alpha_i = a_i + \beta_i [E(f) - \lambda]$$

The intercepts are not free, but they aren't zero either. If the model is right, we should see high intercepts  $a_i$  where we see high  $\beta_i$ . You can see that in the case the factor is a return, then  $E(f) = \lambda$ , and the alpha and intercept are the same thing. It's easy enough to measure alphas this way and test if they are zero, but for reasons I don't understand models with non-traded factors tend to be estimated by the cross-sectional method described below instead.

## 14.8 Summary of empirical procedures

1. The model is

$$E(R^{ei}) = \beta_i' \lambda + \alpha_i$$

(the  $\alpha_i$  should be zero) where the  $\beta$  are defined from time series regressions

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

2. Any procedure needs to estimate the parameters,  $\hat{\beta}, \hat{\lambda}, \hat{\alpha}$ ; give us standard errors of the estimates  $\sigma(\hat{\beta}), \sigma(\hat{\lambda}), \sigma(\hat{\alpha})$ , and an overall test for the model, whether the  $\alpha$  are all zero together, based on weighted sum of squares of the alphas,  $\hat{\alpha}' \text{cov}(\hat{\alpha})^{-1} \hat{\alpha}$ .
3. Time series regression:
  - (a)  $\hat{\alpha}, \hat{\beta}, \sigma(\hat{\alpha}), \sigma(\hat{\beta})$  from OLS time series regressions for each asset.
  - (b) Factor risk premium from mean of the factor,  $\hat{\lambda} = \bar{f}$ . Standard error from our old friend  $\sigma(\hat{\lambda}) = \sigma(f)/\sqrt{T}$ .
  - (c)  $\chi^2, F$  tests for  $\hat{\alpha}' \Sigma^{-1} \hat{\alpha}$  to test all  $\alpha$  together.
  - (d) Note:  $f$  must be a return or excess return for this to work. (True for CAPM, FF3F.)
4. Cross sectional regression:
  - (a)  $\hat{\beta}, \sigma(\hat{\beta})$  from OLS time series regressions for each asset

- (b)  $\hat{\gamma}, \hat{\lambda}$  from cross sectional regression

$$E(R^{ei}) = \gamma + \beta'_i \lambda + \alpha_i. \quad i = 1, 2, \dots, N$$

Ugly formulas for standard errors  $\sigma(\hat{\lambda})$  and  $\sigma(\hat{\alpha})$ .

- (c)  $\hat{\alpha}' \text{cov}(\hat{\alpha})^{-1} \hat{\alpha}$  to test all  $\alpha$  together.  
 (d) Can use if  $f$  is not a return ( $\Delta c$  for example)

#### 5. Fama -MacBeth

- (a)  $\hat{\beta}, \sigma(\hat{\beta})$  from OLS time series regressions for each asset.  
 (b) Cross sectional regressions at each time period

$$R_t^{ei} = \gamma_t + \beta'_{ti} \lambda_t + \alpha_{it} \quad i = 1, 2, \dots, N \text{ for each } t$$

Then  $\hat{\lambda}, \hat{\gamma}, \hat{\alpha}$  from averages of  $\hat{\lambda}_t, \hat{\gamma}_t, \hat{\alpha}_t$ . Standard errors from our old friend,  $\sigma(\hat{\lambda}) = \frac{1}{\sqrt{T}} \sigma(\hat{\lambda}_t)$  similarly for  $\hat{\alpha}$

- (c)  $\hat{\alpha}' \text{cov}(\hat{\alpha})^{-1} \hat{\alpha}$  to test all  $\alpha$  together.

## 14.9 Comments

### 1. Summary of regressions.

- (a) Forecasting regressions (return on D/P)

$$R_{t+1} = a + bx_t + \varepsilon_{t+1} \quad t = 1, 2, \dots, T$$

- (b) Time series regressions (CAPM)

$$R_{t+1}^i = a_i + \beta_i f_{t+1} + \varepsilon_{t+1}^i \quad t = 1, 2, \dots, T$$

- (c) Cross-sectional regressions

$$E(R_{t+1}^i) = \gamma + \beta_i \lambda + \alpha_i \quad i = 1, 2, \dots, N$$

- (d) (Fama-MacBeth cross-sectional)

$$R_t^i = \gamma_t + \beta_i \lambda_t + \alpha_{it} \quad i = 1, 2, \dots, N$$

- (e) These are *totally different regressions. Don't confuse them. In particular,*

- i. Time series regression is *not* about “forecasting returns”. It merely measures covariance, the tendency of  $R^i$  to go up when  $f$  goes up. Once  $f$  has moved, it's too late.
- ii. The  $R^2$  in the time series regression is not relevant to the CAPM. We look for more factors based on  $\alpha_i$ , not based on whether they improve  $R^2$  in time series. (High time series  $R^2$  does improve standard errors, and makes small  $\alpha$  less likely by APT logic of course)

iii. Cross sectional regressions aren't about forecasting returns either. They tell you whether the long run average return corresponds to more risk. (If not...buy!)

2. When is a factor "important"? "Important" for explaining *variance* is different from "important" for explaining means.

$$R_t^{ei} = \alpha_i + b_i rmr f_t + h_i hml_t + \varepsilon_{it}$$

$t$  on  $h$  tells you if hml is important for explaining variance of  $R^{ei}$ .

$$\begin{aligned} hml_t &= \beta_i rmr f_t + hml_t^* \\ R_t^{ei} &= \alpha_i + b_i rmr f_t + h_i hml_t^* + \varepsilon_{it} \end{aligned}$$

$E(hml^*)$ , or

$$hml_t = \alpha_{hml} + \beta rmr f_t + \varepsilon_t$$

$\alpha_{hml} > 0$  tell you if adding hml helps to explain the *mean*  $E(R^e)$  of all the other returns

3. When is the intercept alpha? *only if the right hand variable is a return*

$$\begin{aligned} R_t^{ei} &= \alpha_i + b_i rmr f_t + \varepsilon_{it} \quad YES \\ R_t^{ei} &= \alpha_i + b_i \Delta GDP_t + \varepsilon_{it} \quad NO \\ R_t^{ei} &= \alpha_i + b_i (rmr f_t > 0) + c_i (rmr f_t < 0) + \varepsilon_{it} \quad NO \\ R_t^{ei} &= \alpha_i + b_i R_t^{option} + \varepsilon_{it} \quad YES \end{aligned}$$

(and then, only if you don't do a second stage cross-sectional regression) In the other cases either 1) convert to a return 2) use a cross-sectional regression to get risk premium, convert to

$$E(R^{ei}) = \alpha_i + b_i \lambda$$

4. "Cross-sectional Regression" vs. "Cross-sectional implications of time series regression"

$$\begin{aligned} E(R^{ei}) &= \alpha_i + \beta_i E(rmr f) \\ E(R^{ei}) &= (\gamma) + \beta \lambda + \alpha_i \end{aligned}$$

5. Why do we do asset pricing with portfolios? Why not just use stocks?

- Individual stocks have  $\sigma = 40\% - 100\%$ ; portfolios have much less (diversification). Thus  $\sigma/\sqrt{T}$  is much worse for individual stocks; you can't tell statistically that  $E(R)$  varies across individual stocks.  $\beta, \alpha$  etc. are worse measured.
- Equivalently, time series  $R^2$  are much lower for individual stocks.
- Stocks'  $\beta (b, s, h)$  vary over time. Portfolios can have more stable characteristics. (Implicit hope that  $\beta, E(R), \alpha, b, s, h$ , etc. attach to the *characteristics* such as B/M, size, industry, etc. better than to the individual stock.)
- There are ways to do this all using individual stock data – not yet common in practice (but I think they will be!)

6. \*GMM advertisement

- (a) Motivation 1. How do you do standard errors if errors are *not* independent over time; autocorrelated  $cov(\varepsilon_t \varepsilon_{t+1}) \neq 0$ ; heteroskedastic  $\sigma^2(\varepsilon_t) \neq \sigma^2(\varepsilon_{t+1})$  etc?
- (b) Motivation 2: Where do these standard errors / test statistic formulas come from?
- (c) Motivation 3: How can I derive standard errors for new procedures?
- (d) Answer 1: Simulation. (Monte Carlo, Bootstrap)
- (e) Answer 2: GMM does *everything* (almost)
- (f) Basic idea: *everything in statistics can be mapped in to*  $\sigma(\bar{x}) = \sigma(x)/\sqrt{T}$
- (g) Example: how to fix regression standard errors for any problem under the sun. Example: derivation of all the formulas we've seen so far. And fancier versions that correct for autocorrelated and heteroskedastic ( $\sigma^2$  varies over time) returns! See *Asset Pricing*