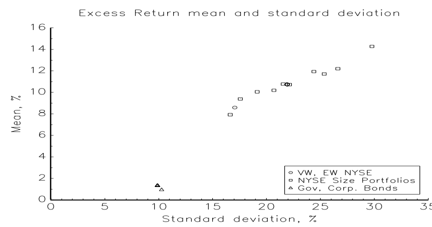
Fama and French

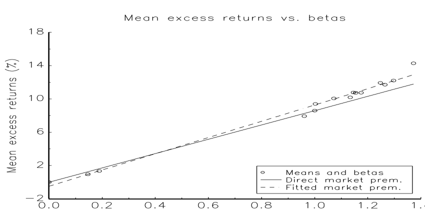
Multifactor explanations

1. Background: CAPM, example 1, size

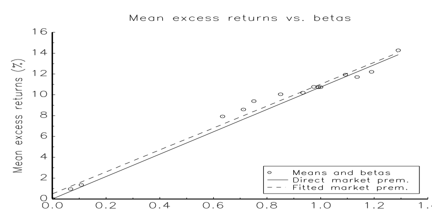
   (a) Expected returns



   (b) Betas

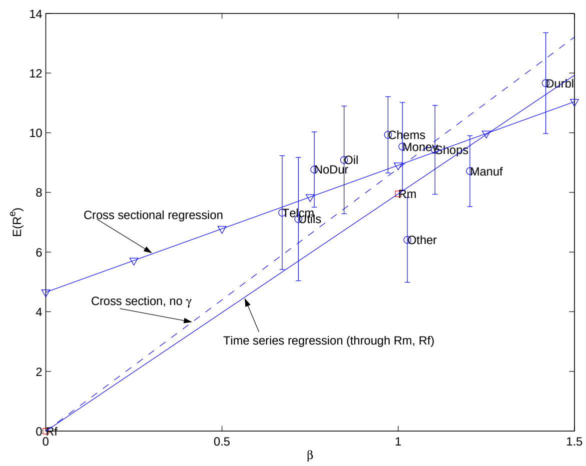


VW market



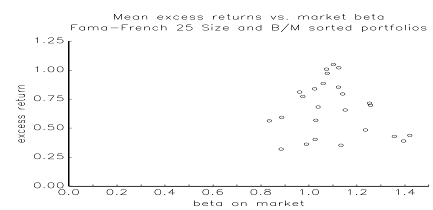
EW market

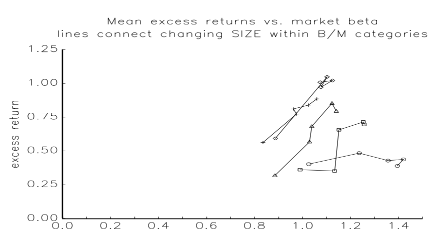
2. CAPM Example 2: industry portfolios

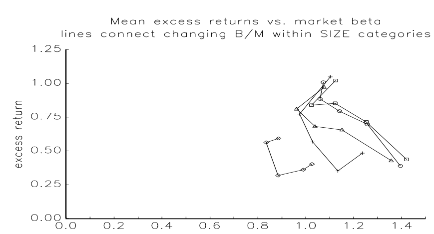


3. FF: What about *book/market* sorted portfolios? (See paper too)

   (a) Facts: There is a big spread in average returns. But market beta is a disaster. *Really the FF puzzle is not that value stocks earn high returns, it is that they do not have high betas. It's a beta puzzle.*







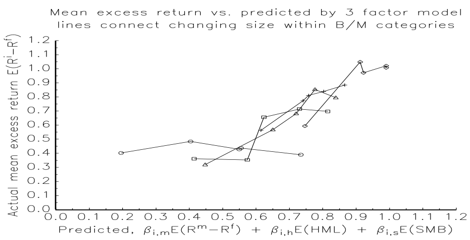
4. Fama-French solution:

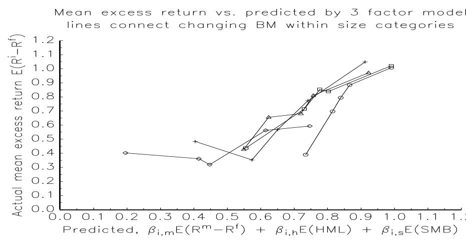
   (a) Run *time series regressions* that include additional *factors* (portfolios of stocks) SMB, HML

   $$R_t^{ei} = \alpha_i + b_i R_t^{em} + s_i SMB_t + h_i HML_t + \varepsilon_t^i;\ t = 1, 2...T \text{ for each } i = 1, 2..N.$$

   (b) Look across stocks

   $$E(R^{ei}) = \alpha_i + b_i E(R^{em}) + s_i E(SMB) + h_i E(HML)$$



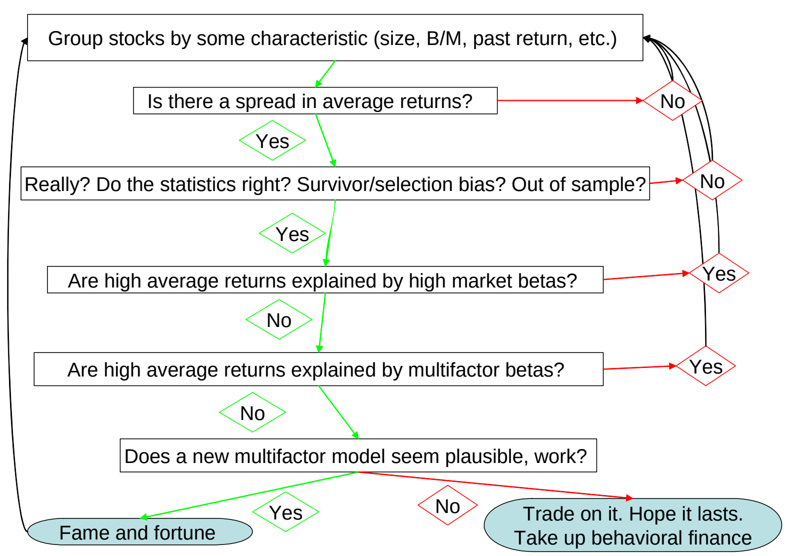


   (c) See Table 1. *Table 1 is* **data** *for an ocular cross-sectional regression*

5. FF other tables. No surprise momentum factor explains momentum portfolios

6. *The point of FF was to produce something that "worked like the capm" allowed "risk ajustment of other puzzles,"* We see this in the other tables.

Empirical Asset Pricing Flowchart



Comments on Fama French

1. Distinguish the uses of time-series regression

$$R_t^{ei} = \alpha_i + b_i R_t^{em} + s_i SMB_t + h_i HML_t + \varepsilon_t^i;\ t = 1, 2...T \text{ for each } i = 1, 2..N.$$

   (a) From the implied cross-sectional regression

$$E(R^{ei}) = \alpha_i + b_i E(R^{em}) + s_i E(SMB) + h_i E(HML)$$

   Read FF Table 1 as a table of data for an ocular regression of this relationhip. The text empasizes that you should look for high betas where you see high average returns.

   (b) FF don't really care about the time-series regression per se – its R[2], beta t-stats, etc. They only care about means. (See the time-series vs. cross-sectional graph). Example: CAPM's greatest success is when $R^2 = 0$ – an asset has high idiosyncratic varinace, no beta and no expected return.

2. What's wrong with $E(R^{ei}) = (size_i)\lambda_s + (b/m_i)\lambda_{B/M}$? ("characteristic adjustment") call it an "explanation.")

   (a) A good *description*. Not a good *explanation*

   (b) If so, you could make money. Find small/value stocks [$(size_i)$, (b/m_i)] that have no betas, (or form a portfolio with no beta). Make huge $ with no risk!

   (c) It's easy to change *characteristic*. Small companies can merge. A portfolio of small companies is a "large" company. This does not change betas.

   (d) Really, what we want is $E(R^{ei}) = (size_i)\lambda_s + (b/m_i)\lambda_{B/M} = s_i\lambda_s + h_i\lambda_m$, the characteristics should be proxies for betas.

3. Is it a tautology to "explain" 25 B/M, size portfolios by 2 B/M, size portfolios? (No. ABC example. Independent stocks example, in which Sharpe ratios rise)

4. Why test with portfolios and characteristics rather than just look at stocks?

   (a) Individual stocks have $\sigma = 40 - 80\%$, so $\sigma/\sqrt{T}$ makes it nearly impossible to accurately measure $E(R)$.Portfolios have lower σ by diversification, so $\sigma/\sqrt{T}$ is not so bad.

   (b) Betas are badly measured too, and vary over time.

   (c) You need an *interesting alternative*. Group stocks together that might have a violation, this gives much more power.

   (d) This is what people do to (try to) make money. Thus keep tests and practice close.

   (e) Most of all, *we have a hunch that $E(R), cov(R, f)$ attach to the characteristic (size, b/m) not to the individual security name.* Thus, forming portfolios avoids having to model $E_t(R_{t+1}^i)$, $cov_t(R_{t+1}^i, ..)$ that vary over time as the characteristics vary.

5. The CAPM seemed fine (and still does) until stocks were grouped by B/M. The CAPM still works fine for some *groupings* (size), not others (value)! Similarly FF3F works well for some groupings not others.

6. Implied cross-sectional regression II. What are FF factors? What are the right factors to use? A: It Depends on what question you want to ask.

7. (a) Purpose/interpretation 1: Fight about "rational" vs. "irrational." Then you want proxies for state variables $proj(m|X)$. $0 = E(mR^e) = E(proj(m|X)R^e)$. "State variables of concern to investors" p. 77 But they left for us, what's the m.

   (b) 1. Story 1: outside income Pepole lose jobs in recessions. *Gven* market beta, they avoid stocks that go down in recessions→drive down prices→drive up expected returns. Now expected returns depend on tendency to go down in recessions as well as market beta. "state variable for distress."
   2. Independent evidence? HML doesn't do that badly in recessions. The search for earnings etc. factors has not been conclusive.
   3. Story 2: intertemporal hedging (ICAPM "state variables.") There is some work showing that B/M innovations predict the market as a whole, but still weak.
   4. Lots of work on "what is hml a proxy for?" (Lettau ludvigson conditional CAPM is one example.)

8. Purpose/interpretation 2: APT: A "Minimalist interpretation." p. 76.

   (a) The central finding of Table 1 paper is that size, B/M portfolios *move together,* high R^2 This fact survives even if the value premium disappears. A,B,C portfolios would not have huge R^2 on the AL/MZ portfolio. *There are three dominant eigenvalues in the FF 25 portfolios*, as you found in the problem set.

   (b) (Interestingly, people who did factor analysis on individual stock returns did not find this. However, they did not realize that the point we're all after is characteristic-based characterization. They tried to look at $cov_t(R^i, R^j)$ where i, j vary across all individual stocks. They should have looked at $cov(R^i, R^j|c^j, c^j)$ where c is a vector of characteristics. FF are saying that size$_i$ and BM$_i$ are important characteristics here.)

   (c) This bears a bit on point #1. "Irrational pricing" stories can describe why mean returns of BM stocks are high, but why should they all move together on news?

   (d) APT "minimialist interpretation" "*where there is mean there must be common movement, or huge sharpe ratios will emerge.*" And lo, there it is.

   1. Basic APT

$$
\begin{aligned}
R^{ei} - b_i rmrf - h_i hml - s_i smb &= \alpha_i + \varepsilon_i \\
SR &= a/\sigma(\varepsilon)
\end{aligned}
$$

   where R^2 is high α is low or there is a big sharpe ratio.

2. More generally

$$R^e = \alpha + \beta f + \varepsilon$$

$$R^{ep}_{t+1} = w' R^e_{t+1} - v' f_{t+1}$$

Solve the problem: max Sharpe ratio of such portfolios

$$\max_{\{w,v\}} \frac{E\left(R^{ep}\right)}{\sigma\left(R^{ep}\right)}$$

Answer:

$$SR^2 = \underbrace{E(f)'cov(f,f')^{-1}E(f)}_{\text{max } SR \text{ from factors alone}} + \underbrace{\alpha' cov(\varepsilon,\varepsilon')^{-1}\alpha}_{\text{extra SR from exploiting } \alpha}$$

3. To the extent that the FF model is successful, then, its lesson is, *you only need to price the 3, not the 25.* "pricing the 25" is a bit silly. In this *the point of FF is to reduce information in 25 portfolios to that of 3 factors – why is E(hml) high?* Hilariously, most of the literature misses this and tests on the 25 portfolios. You don't need a macro model to price the 25, you need only price hml! Pricing the 25 is throwing out the great summary FF gave you. "horse races" with FF only assess if you can price small growth better, which may be a bit silly.
4. Thus, the central puzzle is that HML seems mispriced by CAPM. Given HML mispricing other portfolios follow by APT logic

(e) APT/ICAPM?

(f) Will it still work on non size-B/M portfolios? APT says only if they still give high R^2 A "real" model works on all sorts, even those that give low R2.

(g) Here is where explaining *other* sorts becomes important. Success? Yes (sales) but no (portfolios still have high R^2) In some sense the sales growth regressions are a success for the FF3F *model of variance*, not their model of means!

(h) This highights what I think is Fama-French's real **purpose**, which is neither. They want *to replace the CAPM as all-purpose easy risk adjustment in empirical work.* In this, it is remarkably successful. Every anomaly paper uses three (or four, with momentum) factor adjustment to document new anomalies.

(i) Variance, time series regressions and risk management

(j) Sometimes we do care about the time series regression. It expresses an analysis of *variance*

$$var\left(R^{ei}_t\right) = b_i^2 var\left(R^{em}_t\right) + s_i var\left(SMB_t\right) + ... + var\left(\varepsilon^i_t\right)$$

For risk management, etc. this is important. The time-series regression tells you for example what "hege ratios" to use to minimize the variance of an investment by shorting factor portfolios. Regression coefficients are portfolio weights! Just as with the APT

$$R^{ep} = R^{ei} - b_i rmrf - h_i hml - s_i smb = \alpha_i + \varepsilon_i$$

is a tradeable portfolio!

(k) For this case *non-priced factors are just as interesting.* Example: industry portfolios (further correlation of the errors).

(l) Include nonpriced factors or not? Yes for variance, no for expected returns An example of use: hedging outside income.(Note including nonpriced factors can also reduce $\sigma(\varepsilon)$ hence $\sigma(\hat{\alpha}), \sigma(\hat{\beta})$

(m) Why do we spend so much time on searching only for priced factors? These are only important for the one last mean-variance investor?

(n) A final purpose: "Performance attribution" Carhart mutual funds for example, uses momentum factor. In this case the *question* is merely "can I earn this return with passive strategies without paying you fees?" For this purpose it does not matter if the factors are "real" or "irrational." So the standards are much lower, and momentum factors make more sense.

9. FF other tables.

   (a) Sales growth

   (b) Reversal

   (c) Momentum. Note negative betas / disaster. Note momentum and value are negatively correlated.

   (d) Why not momentum factor? It depends on your purpose.

Dissecting anomalies

- Where we are. 1) The CAPM worked 2) Many anomalies sprang up. The world was confused. 3) Anomalies all explained by size and B/M. The world is simple again! 4) Now, there are many more anomalies. 5) It's time to unite again under a few factors! FF 2 is the first step.
- (Style: questions to ask the class to provoke discussion)

1. The point of this paper is to look at momentum, and a bunch of additional variables that appeared since the size and B/M work.

   (a) Are they real?

   (b) Are they subsumed in size, B/M?

   (c) Are they all independent, or are some subsumed by others? (1654 "which have information about average returns that is missed by the others.")

   (d) The paper also looks at what anomalies are there in big stocks, vs. what is just a feature of microcaps, which can't really be exploited. Take some time with this paper to really digest Table 2 and 4.

   (e) Finally, the paper agonizes about functional form. Is expected return really related to a firm's B/M, to the log of B/M, to which decile of B/M a firm is in, etc.? This is vital for our methodological query.

2. Are the average returns in Table II raw, excess, or adjusted somehow? Do they represent returns, or alphas, or something else? (Discussion on ff procedure)

   A: they are "characteristic-adjusted", explained 1658 below II. sorts. This means, find the portfolio of 25 size/book/market whose size and B/M are closest, and subtract off that return. The text says that true size and book/market alphas gives similar results, though since there are some big alphas (small/growth) separating average returns and betas in the 25, I'm not alltogether convinced. OTOH, FF argue that individual-stock hml, smb betas are measured badly and wander over time. Thus, they say, the characteristic is a better measure of beta than beta itself. Anyway, read the table as FF's ideas about alphas *after* controlling for size and b/m.

3. Explain the first row of Net stock isues (Market) in Table II. What do the numbers mean?

   A: this just leads to a discussion to make sure we understand table construction.

4. Why are the t- statistics for the High-Low portfolio so much better than for the individual portfolios?

   A: We're really not that interested in whether portfolio excess returns are different from zero. We want to know if they're different *from each other.* If all averages were equal to each other but different from zero, it wouldn't be that interesting. Each portfolio could be within a standard error of zero, but if the long-short portfolio is significant, you have a trading strategy/anomaly.

5. Which anomalies produce strong average hedge returns for all three size groups? What numbers in Table 2 document your answer? (Hint: start by reading the H-L returns, then the H-L t stats, then look at the remaining columns)

   A: read 1662 pp3. Issues, accruals, and momentum. Look at the High-Low number. Look for consistency across 4 size groupings, and consistency across VW and EW results.

6. Note: FF are really interested in what goes on in the microcap range. I'll focus on the results that survive in the big ranges. Which anomalies seem only to work in tiny stocks in Table II?

   A: Asset growth. Top of p. 9, look at numbers

7. Which anomaly gives the highest Sharpe ratio in Table 2? (Help, there are no Sharpe ratios in Table 2! Hint: how is a t statistic computed? You can translate from t to Sharpe ratios.)

   A: $t = E(R)/\left(\sigma(R)/\sqrt{T}\right); E(R)/\sigma(R) = t/\sqrt{T}$. To annualize$E(R)/\sigma(R)_{annual} = \sqrt{12} \times t/\sqrt{T} = t/\sqrt{T_{years}}$, $\sqrt{T_{years}} = \sqrt{42.5} = 6.52$. Thus, a $t = 3.26$ translates to the market Sharpe ratio 0.5, and a t=6.52 translates to a Sharpe ratio of 1. Hedge funds think they can find Sharpes of 2 or more – good luck. Most of the ts are between 3 and 5, especially if you only look at big firms.

8. The Profitability sort seems not to work in Table 2. (Point to numbers). How did people think it was there? (Hint: 1663 pp2)

   A: 1663 pp2 *With controls for cap and B/M.* There is a profitability effect on its own, but size and B/M pick it up. This is a good instance of the point of the paper – what works *in the presence of the others*, what has *marginal* power, what is the *multiple* regression forecast of returns, not each variable at a time.

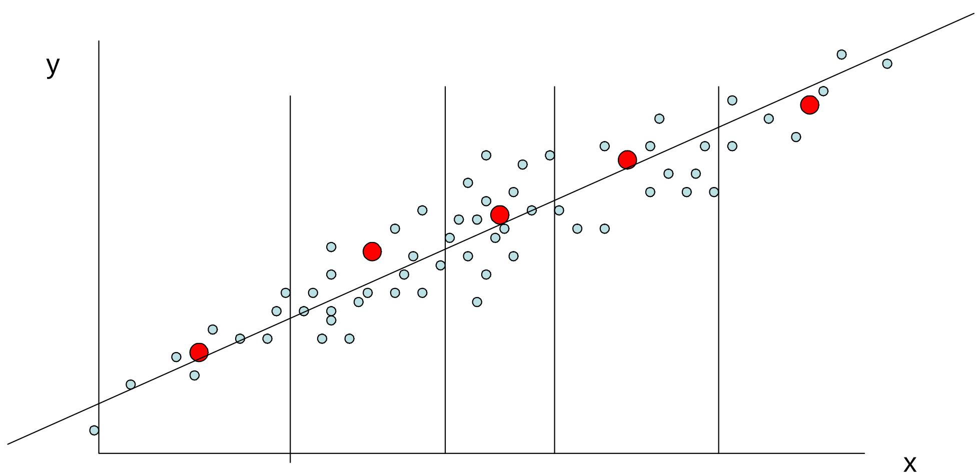
9. Explain why the numbers in Table III jump so much between 4 and high.

   A: Graph in class. The 1/5 of extreme values of any distribution is way spread out.

10. **Note** on the way to Table 4. Table 4 has Fama-MacBeth regressions,

$$E\left(R^i\right) = a + b_1 \log(MC_i) + b_2 \log(B/M_i) + b_3 Mom_i + .. + \varepsilon_i;\ i = 1, 2..N.$$

   If all works well, this regression gives the same information as splitting things into 5 groups and looking at group means. But the paper is all about the pitfalls of each method vs. the other. One reason for doing regressions is there is no way to split things into groups based on 2,3,4, etc. variables, to see whether, for example, momentum is still important after accounting for size and B/M. 1666 below III, "which anomalies are distinct and which have little marginal ability to predict returns?" But regressions need to take more of a stand on functional form, which FF worry about a lot. ("pervasive" is also about functional form though. It's only "pervasive" if expected returns are linear in the portfolio numbr.)



11. Explain what the top left 4 numbers mean in Table 4.

    A: you're seeing the basic size and B/M effects in expected returns. Larger size means smaller ER, Larger B/M means larger ER.

12. "The novel evidence is that these results [size effect] draw much of their power from tiny stocks" What numbers in Table 4 are behind this conclusion?

    A: This is the disappearance of the size coefficient in the other groups in the top left part of Table 4. Note size is also much weaker post 1979 – when the effect was published and small stock funds started. (not in this paper)

13. "The relation between average returns and B/M is more robust" What numbers in Table 4 are behind this conclusion?

    A: the B/M coefficients down the second column in Table 4.

14. What is a "good" pattern of results in Table 4? Which variables have it, and which do not?

    A: We're looking for a large coefficient and t stat, and we want the coefficient to be consistent in the size groupings. Issues, momentum, and positive accruals do.

15. Asset growth and profitability have nice big t stats in the top rows of Table 4. Yet FF dump on them Why?

    A: look at the size groupings – it only works for tiny stocks.

16. Do any of the anomaly variables drive the other ones out, or does each seem to give a separate piece of information about expected returns?

    A: Table 4 and bottom. The ones that were individually significant all seem to survive, Except Zero NS. This is sad, *A central point of the paper – I was hoping that some anomalies would drive other sout.*

17. In the conclusions, FF say "The evidence..is consistent with the standard valuation equation which says that controlling for B/M, higher expected net cashflows...imply higher expected stock returns" and "Holding the current book-to-market ratio fixed, firms with high expected future cash flows must have high expected returns" Isn't this the fallacy that "profitable companies have higher stock returns" , or "confusing good companies with good stocks"?

    A: Note the crucial "holding B/M fixed." Holding price fixed, anything that forecasts cashflows must also forecast returns. Go back to our linearized present value formula,

$$p_t - d_t = E_t \sum_{j=1}^{\infty} \rho^{j-1} \Delta d_{t+j} - E_t \sum_{j=1}^{\infty} \rho^{j-1} r_{t+j}$$

    The fallacy is that high Δd_{t+j} means nothing about r, because it just means high $p - d$. But *holding p-d constant* high Δd must also come with high r. In that sense, the cashflow variables can be thought of as "cleaning up" B/M for the fact that B/M forecasts both cashflows and returns.

    Note: If you can't directly answer the following questions from the paper, at least think about what else you need to know in order to figure out the answer.

18. Do these new average returns correspond to new dimensions of common movement across stocks, as B/M and size corresponded to B/M and size factors?

    A: This paper does not go on to do the next obvious question: do we now have 5 or 6 factors? Do we need a "new issues factor?" *Low-hanging fruit: ok, we know there are new dimensions of ER not captured by size, b/m and momentum. Do these correspond to known* $cov(R, f)$ *or do we need a new-issues HML factor etc.? Are we going to have new factors for every anomaly?*

19. What is the highest Sharpe ratio you can get from exploiting one of these anomalies? (Choose any one). / Do these correspond to new factors or is one extra factor enough?

    A: We don't know. That also depends on the covariance structure. If the stocks or portfolios sorted on a new anomaly are independent, Sharpe ratios go through the roof. If there are also common factors, the sharpe ratios from diversified portfolios that load on new anomalies is not so large. We know the individual portfolio sharpe ratios are 0.5-1.0 from the t stats, though, and we know these are uncorrelated from the market, so we know there are some interesting Sharpe ratios in here, even if there are new common factors! *Q1 is there a "new issues factor?" Q2 do we need a new issues factor to price stocks? Is the new issues factor priced by the other factors? (*reminder: you need a new factor iff $f_t = \alpha + b_f rmrf + h_f hml + s_f smb + \varepsilon$ has $\alpha \neq 0$)

20. What is the highest Sharpe ratio you can get from combining all these anomalies and exploiting them as much as possible?

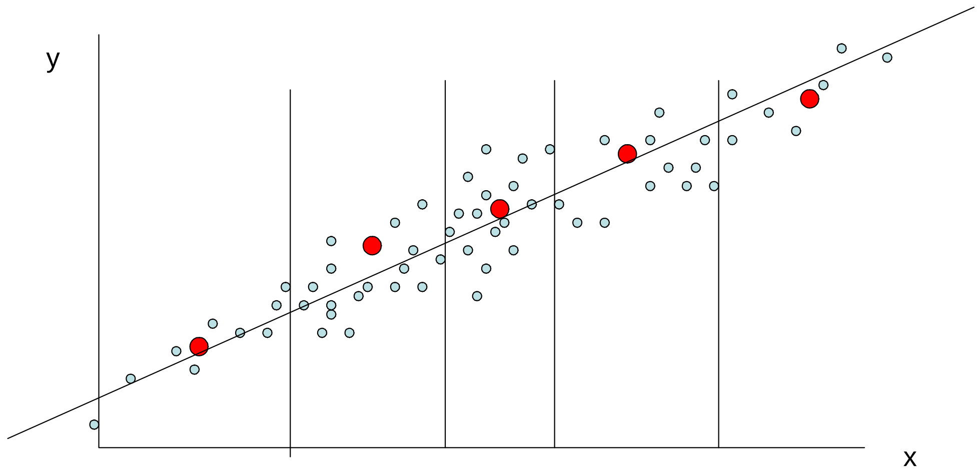
    A: Again, we don't know without knowing a) are there new common factors b) how correlated are the new common factors. Low hanging fruit!

Challenges for empirical work going forward

- Let c_{it} denote a characteristic of asset i – interest differential, B/M, etc.

$$\begin{aligned} R^i_{t+1} &= a + bc_{it} + \varepsilon_{i,t+1} \\ E(R^i_{t+1}|c_{it}) &= a + bc_{it} \end{aligned}$$

Now let's think about what FF do. *By sorting into portfolios and taking means, they are running a non-parametric cross-sectional regression.* They are characterizing $E(R|c)$.



- *So many techniques are all delivering the same idea*

1. Time series regression

$$R^i_{t+1} = a_i + b_i c_{it} + \varepsilon_{i,t+1} \; t = 1, 2, ..T \forall \iota$$

2. Cross-sectional regression

$$E(R^i_{t+1}) = a + bc_{it} + \varepsilon_i$$

3. Fama-MacBeth, pooled time-series cross section

$$R^i_{t+1} = a + bc_{it} + \varepsilon_{i,t+1} \; t = 1, 2, ..T \forall \iota$$

with many variations $b_i = b$;a$_t$ vs. a_i, etc.

4. Portfolio means

- The central idea: *Expected returns (and later, covariances) attach to a* **characteristic** *such as size, B/M, i-i* not to the security (country, firm, etc.)* Of course those are characteristics too, just not ones that we think matter.

- The brilliance of FF portfolios. They turn a hard-looking *dynamic* asset pricing problem – characterize $E_t(R^i_{t+1})$ – into a *static* one – characterize $E(R_{t+1})$ in portfolio i.

1. But, as with CP2, studying the parameter b *does the same thing.*
2. Also, you've seen this many times before, i.e. managed portfolios $E_t(m_{t+1}R_{t+1}|z_t)$ becomes $E(m_{t+1}R_{t+1}\otimes z_t) = E(m_{t+1}(R_{t+1}\otimes z_t))$

- Portfolio means are brilliant because *they are a lens for seeing the economic significance of small* R^2. Example 1: Momentum portfolios. $\rho = 0.1, R^2 = 0.01$, yet huge returns. Why? $R_t = 100\%$ implies $E_t R_{t+1} = 10\%$. More generally, what matters is (say) $E(R_{t+1})/\sigma(\varepsilon)$ sharpe ratio, not $\sigma^2(bc_{it})/\sigma^2(R)$ (R^2) *you can have huge Sharpe ratios with tiny* R^2

- Why high-low? (or the usual 1-10 spread)

1. Because we're really interested in "are expected returns different from each other" not "are they different from zero?" *Develop a statistic for "are expected returns different from each other.* It's not as easy as "are they different by two standard errors" of course because returns are correlated. $\sigma^2(A-B) = \sigma^2(A) + \sigma^2(B) - 2\sigma(A,B)$ But why ignore the evidence in all the other portfolios?
2. Looked at this way, how do you test for difference in means better than the 1-10 portfolio? A: Test the slope coefficient in the cross-sectional regression! (Or nonparametric counterpart).
3. It seems we get better returns and higher t statistics the finer we chop portfolios. Can you make anything look good by making 100 portfolios and then looking at the 1-100 spread? A Yes and No.
4. Yes: you can make *means* look better! That's why we should look at t stats or E/σ
5. No, really. the variance goes up as well, so the sharpe ratio E/σ and the t statistic $E/\left(\sigma\sqrt{T}\right)$ should stabilize as you get more extreme.

$$\frac{E(R^{ei}-R^{ej})}{\sigma(R^{ei}-R^{ej})} = \frac{\left(\beta^i-\beta^j\right)\lambda}{\sqrt{\left(\beta^i-\beta^j\right)^2\sigma_f^2+\sigma^2(\varepsilon^i-\varepsilon^j)}} \rightarrow \frac{\left(\beta^i-\beta^j\right)\lambda}{\left(\beta^i-\beta^j\right)\sigma_f} = \frac{\lambda}{\sigma_f} = sr_f$$

This issue is worth digging more. Why not estimate the limit? Why not (and how) use information in all securities to estimate this quantity?

6. Fama - French: 1-10 focuses on microcaps. Especially if equally weighted inside the portfolios)

- Having found unity, *What's the right way?* (Avoid 30 years of "because that's how Fama does it.") Well, think about the differences between the techniques!

1. LRV portfolios are based on f-s. Thus, they do *not* have a country-specific constant. They are similar to

$$R^i_{t+1} = a + bc_{it} + \varepsilon_{t+1} \text{ (no } a_i\text{)}$$

2. More generally, we want to characterize means by a panel regression

$$R^i_{t+1} = a + c_{it} + \varepsilon_{t+1}$$

Should we have time dummies a_t – being higher *than everyone else* at t is what gives you excess return, but bieng higher at t than at $t-1$ does not? Should we have "firm dummies" in some sense, a_i, so that being higher *on average* doesn't give any return, but being higher *than usual for you* does? (That is my speculation in the book, but LRV seem to say no). At a minimum, *time vs. firm dummies can give quite different results, and summarize quite different aspects of the data* (Denis Chaves thesis, time dummies vs. firm dummies.)

3. Portfolios, being a pure cross-sectional relation, thow away time-series information. If portfolio 6 c is higher at t than t-1 and this means higher returns at t+1, it is missed.

4. Does return correlated with the characteristic itself c_{it} or with the portfolio rank? FF search for "consistent" i.e. linear relation in the portfolios is a search for rank. But the characteristic is usually much further out for the extreme portfolios, so common s-shaped relations as a function of rank suggest the characteristic itself matters. (PICTURE)

5. Moral: *the functional form between characteristics and return matters a lot.* FF procedure hides a lot of assumptions.

6. We'll come back and think about this in other contexts.

- Now, as with FF it seems we have translated a dynamic problem into a static one. How do we translate the covariance statement back to equations?

1.
$$\text{General: } E_t(R^{ei}_{t+1}) = cov_t(R^{ei}_{t+1}, f_{t+1})\lambda_t$$

2. FF / LVR model $E_t(R_{t+1})$ as a function of characteristics, prespecify factors, and then use constant market prices of risk. I think they're doing this.

$$\text{FF,LVR:} E(R^{ei}_{t+1}|c_{it}) = cov(R^{ei}_{t+1}, f_{t+1}|c_{it})\lambda$$

(Make graph.) Ideally, we should do this parametrically/nonparametrically and then test for equality of the *functions*

3. But this ignores that f are formed on the same basis. Really we're saying something about the covariance matrix of returns $cov_t(R^{ei}_t R^{ej}_t) = f(c_i, c_j)$

4. What did we do?

$$CP2 : E(R_{t+1}|c_{it}) = cov(R_{t+1}, f_{t+1}|c_{it})\lambda$$

5. Reconciliation of CPII and LRV

Value and momentum everywhere

- Uniting asset classes (markets) both on ER and Cov dimensions
- "latest" on value, momentum
- Table 1. Value and momentum "factors" in individual markets.
  1. Focus onE/σ to avoid scale issues.
  2. Negative correlation means R_{t+1} of the value and momentum portfolios are negatively correlted. Thus value and momentum together do even better. (AQR)
  3. Imperfectly correlated across countries / asset classes so "everywhere" strategies do better still. Bottom right number (8.5% 10.31 1.97) is the champ. Annualized
- Table 2. Correlations of value across markets, momentum across marekts, and value with momentum. A global "value factor" of course, limits their maginicent sharpe ratio. Note correlations are different at monthly, quarterly horizon so returns are not iid (measurement?)
- Figure 1. A level up value down momentum factor. *The central point. The covariance matrix.* This seems to unite value and momentum to a single factor loading.
- Table 3 Cross section using global value and momentum. How about global first factor? coming up.

Notes on empirical methods

Statistics of time series and cross sectional regressions

1. Time Series Regression (Fama-French).
   (a) *Method:* Run and interpret
   $$R_t^{ei} = \alpha_i + \beta_i' f_t + \varepsilon_t^i \quad t = 1, 2...T. \text{ for each } i$$
   (b) *Estimates*:
      1. $\hat{\alpha}, \hat{\beta}$: OLS TS regression.
      $$R_t^{ei} = \alpha_i + \beta_i' f_t + \varepsilon_t^i \quad t = 1, 2...T \text{ for each } i.$$
      2. $\hat{\lambda}$: Mean of the factor,
      $$\hat{\lambda} = \frac{1}{T}\sum_{t=1}^{T} f_t = \bar{f}.$$
   (c) *Standard errors: If* ε_t^i are independent *over time.*
      1. OLS standard errors $\hat{\alpha}_i, \hat{\beta}_i$.

2. $\hat{\lambda}$:

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

(d) *Test* α are *jointly* zero?

1. Answer: look at

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

Precise forms,

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

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

Intuition. $R_t^{ei} = \alpha_i + \beta_i' f_t + \varepsilon_t^i$ means that $\hat{\alpha}_i \approx \alpha + \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t^i$ (except for beta fitting). Thus $cov(\hat{\alpha}) \approx \frac{1}{T}\Sigma$. The other terms correct for beta fitting. As usual χ^2 is asymptotic for any iid distribution, F is finite-sample for normal ε.

2. Cross-sectional regression, two steps

(a) *Procedure*

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

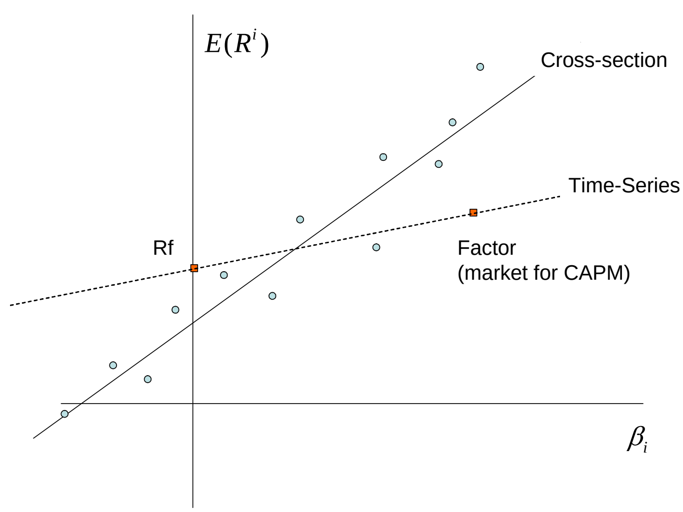
1. TS (over time for each asset) to get β_i,

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

2. Run XS (across assets) to get λ.

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

3. TS vs. OLS CS.



(b) *Estimates*:

1. $\hat{\beta}$ from TS.
2. $\hat{\lambda}$ slope coefficient in CS.
3. $\hat{\alpha}$ 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.*

(c) *Standard errors.*

1. $\sigma(\hat{\beta})$ from TS, OLS formulas.
2. $\sigma(\hat{\lambda})$. You can't use OLS formulas. Errors $\hat{\alpha}$ are correlated, β are estimated. Answer: With no intercept in XS,

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

3. $cov(\hat{\alpha})$

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

(d) *Test*

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

Warning $cov(\hat{\alpha})$ is singular; use pinv or eigenvalue decompose and only invert the nonzero eigenvalues.

(e) *Formulas with a free intercept*

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

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

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

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

3. GLS cross sectional regression.

(a) Formulas

$$\hat{\lambda} = \left[\beta' cov(\alpha)^{-1}\beta\right]^{-1}\beta' cov(\alpha)^{-1}E(R^e)$$
$$\hat{\lambda} = \left[\beta'\Sigma^{-1}\beta\right]^{-1}\beta'\Sigma^{-1}E(R^e)$$

see *Asset pricing* for $\sigma(\hat{\lambda}), cov(\hat{\alpha})$

(b) Theorem. If you include f_t and R^f (0) as test assets, then Σ is singular in just the right places and GLS CS = Time series

4. Fama-MacBeth Procedure

(a) Run TS to get betas.

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

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

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

(c) *Estimates* of λ, α are the *averages across time*

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

(d) *Standard errors* use our friend $\sigma^2(\bar{x}) = \sigma^2(x)/T$

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

This is one main point. These standard errors are easy to calculate.

(e) *Test*

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

Why don't people do this?

5. GMM. Of *course* you should be doing GMM and not assuming ε_t are iid over time, right?

6. Testing one model vs. another (see longer description below)

(a) Example. FF3F.

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

Drop size?

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

(b) Solution: "Orthogonalized factor",

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

We can drop smb from the three factor model if and only α_{smb} is zero. .

(c) Equivalently, we are forming an "orthogonalized factor"

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

and drop if $E(smb^*) = 0$

(d) This must be equivalent to a proper test whether $\alpha'\alpha$ declines, (*not* comparing GRS statistics!)

Testing whether a factor is redundant.

Smb looks pretty marginal. Can we get by in our explanation of *mean* returns (not necessarily in our understanding of the *return covariance matrix)* with a simpler model that only uses hml and rmrf? (Fama French might be perfectly right that smb is an important factor in covariances, as industry factors are, but exposure to smb might not be important to explain *mean* returns)

Lots of people think that we test for the importance of smb by looking at s_i t statistics, mistaking the time-series regression that measures decomposition of *variance* for the implied cross-sectional relation which measures the extent to which the factor model explains *means.* Lots more people mistakenly think that we test for dropping smb by testing whether $E(smb) = \lambda_{smb} = 0$, "is the factor priced."

This is wrong however. Why? Suppose $smb_t = \frac{1}{2}rmrf_t + \frac{1}{2}hml_t$. Then, obviously, $E(smb) > 0$, but just as obviously, we can drop smb from the right hand side with no harm at all. We need to test whether smb is useful to *price other things*, not whether it *is priced.* Equivalently, when you drop smb from the regression the other coefficients change, and they may change just enough to still explain expected returns, even if $E(smb) > 0$.

We can do the right test very simply by running a regression of smb_t on $rmrf_t$ and hml_t.

$$smb_t = \alpha_s + b_s rmrf_t + h_s hml_t + \varepsilon_t$$

If $\alpha_s = 0$, then smb is priced by other two factors, and this *is* the test whether we can drop smb from the model.

Why? Think about defining an orthogonalized factor $smb_t^* = \alpha_s + \varepsilon_t = smb_t - b_s rmrf_t - h_s hml_t$ Rewriting the original model in terms of smb^*,

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

and hence

$$E(R^{ei}) = \alpha_i + (b_i + s_i b_s) E(rmrf_t) + (h_i + s_i h_s) E(hml_t) + s_i E(smb_t^*)$$

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. And $E(smb^*) = 0$ is exactly the same as $\alpha_s = 0$. (This is equivalent for a test $b_s = 0$ in $m = a - b_m rmrf - b_h hml - b_s smb$, $0 = E(mR^e)$ as advocated in the λ vs. b Chapter 13.4 of *Asset pricing.* This must be equivalent to a test whether $\alpha' \Omega^{-1} \alpha$ has risen for some well defined common Ω, but neither I nor anyone else has worked that out.)

Eigenvalue factor decompositions

Summary of the procedure

Given an $N \times 1$ vector of random variables y, with $cov(y, y') = \Sigma$, we form $Q\Lambda Q' = cov(y)$ by the eigenvalue decomposition. If Y is a $T \times N$ matrix of data on y, then [Q,L] = eig(cov(y)) in matlab. Λ is diagonal and Q is orthonormal, $QQ' = Q'Q = I$.

We form "factors" by $x = Q'y$.The columns of Q thus express how to construct factors x from the data on y.

We can then write $y = Qx$, i.e. $y_t = q_1 x_t^{(1)} + q_2 x_t^{(2)} + ...$where $q_1 = Q(:,1)$ denotes the first column of Q. The columns of y thus also give "loadings" that describe how each y moves if one of the factors x moves.

We have $cov(x, x') = Q'Q\Lambda Q'Q = \Lambda$, i.e. the x are uncorrelated with each other.

If some of the diagonals Λ are zero, then we express all movements in y by reference to only a few underlying factors. For example, if only the first Λ is nonzero, then we can express $x = q_1 z_1$ In practice, we often find that many of the diagonals of Λ are very small, so setting them to zero and fitting y with only a few factors leads to an excellent approximation.

Since the factors are uncorrelated, if we ignore some factors, the loadings on the remaining ones are the same as if we ran regressions,

$$\begin{aligned} y_t &= q_1 x_t^{(1)} + q_2 x_t^{(2)} + [q_3 x_t^{(3)}] \\ y_t &= q_1 x_t^{(1)} + q_2 x_t^{(2)} + \varepsilon_t \end{aligned}$$

Derivation

Factors constructed in this way solve in turn the question "what linear combinations of y have maximum variance, subject to the constraint that the sum of squared weights is one and each linear combination is orthogonal to the previous ones?" In equations, each column q_i of Q satisfies

$$\max\left[var\left(q_i' y\right)\right] \text{ s.t. } q_i' q_i = 1,\ q_i' q_j = 0,\ j < 1$$

Why is this an interesting question? "What linear combination $q' y_t$ of the y_t has the highest variance?" would be too easy a question. Just make q big. The constraint $q'q = 1$ means the sum of squared q must be equal to one, so you can't boost variance by making q big. You have to find the right *pattern* of q across the elements of y.

Why is this the answer? Let's look at the first maximization – what linear combination of the y has the largest variance, if you constrain the sum of squared weights to be less than one?

$$\max_{\{q\}} var(q' y_t) \text{ s.t. } q'q = 1$$

Forming a Lagrangian,

$$L = q'\Sigma q - \lambda q' q$$

The first order condition $(\partial/\partial q)$ is

$$\Sigma q = \lambda q$$

This is an eigenvalue problem! The answer to this question is then, *choose q as an eigenvector of the matrix Σ*. Now, which one? Let's see what variance of x we get out of all this

$$var(x_t) = var(q'y) = q'\Sigma q = q'\lambda q = \lambda q' q \lambda$$

Aha! *The eigenvalue gives us the the variance of* x. So, the answer to our maximization is, *choose the eigenvector* q *corresponding to the largest eigvenvalue* λ.

In sum, to find the linear combination of y with largest variance, and sum of squared weights equal one, we choose as weights q the eigenvector corresponding to the largest eigenvalue of the covariance matrix of y. Q is a matrix of eigenvectors, so you're done. Eigenvectors are orthogonal $q_i'q_j = 0$ so the eigenvectors corresponding to successively smaller eigenvalues answer the question for the remaining factors.

Now, what does this have to do with R^2? Suppose we leave out some factors

$$y_t = q_1 x_t^{(1)} + q_2 x_t^{(2)} + [q_3 x_t^{(3)}]$$

Since $x^{(1)}$ and $x^{(2)}$ have *maximum* variance, this means $x^{(3)}$ and beyond have *minimum* variance. In short *we have found a factor model that for each choice of how many factors to use maximizes the* R^2 *in these regressions.*

Rotation

There are lots of equivalent ways to write any factor model. If we have

$$y_t = Qx_t$$

then if R is any orthogonal (rotation) matrix, i.e. any matrix with $RR' = R'R = I$, we can define new factors $z_t = R'x_t$. Then $x_t = Rz_t$

$$y_t = Q(Rz_t) = (QR)z_t$$

Now the columns of QR give us new loadings on the new factors, which are constructed by $z_t = R'x_t = R'Q'y_t = (QR)'\,y_t$

This works even better if we use unit variance shocks.

$$\begin{aligned} y_t &= \left(Q\Lambda^{\frac{1}{2}}\right)\left(\Lambda^{-\frac{1}{2}}x_t\right) = \left(Q\Lambda^{\frac{1}{2}}\right)z_t \\ cov(z_t, z_t') &= \Lambda^{-\frac{1}{2}}\Lambda\Lambda^{-\frac{1}{2}} = I \end{aligned}$$

Now if we rotate,

$$w_t = Rz_t, z_t = R'w_t$$

we have

$$y_t = \left(Q\Lambda^{\frac{1}{2}}\right)z_t = \left(Q\Lambda^{\frac{1}{2}}\right)R'w_t$$

$$cov(w_t, w_t') = cov(Rz_t, z_t'R') = RR' = I$$

In sum, *if we rotate unit-variance factors, they are still uncorrelated with each other and still have unit variance.* You're free to recombine factors in any way you want to make them look pretty.