Problem Set 7

Table 4 of Fama-French "dissecting anomalies" uses cross-sectional regressions of returns on characteristics. Here we're going to explore this technique and start thinking about the project to replace 1-10 portfolio sorts with cross-sectional regressions, as in "Discount rates."

Load the size and beme data. (It's in the 25 portfolio data, or download from the class website. I used the value weighted average be/me, as it looked a little better in plots (over time) than sum be/sum me.) We'll use data from 1947 on. The beme data are for July of the indicated year. Thus, use kron to expand them to monthly data.

Our goal is to understand expected returns as a function of characteristics visible ex ante,

$$E\left(R_{t+1}^{ei}|beme_t^i, me_t^i, ..\right)$$

We want a relation that is stable across time and across assets. For January 1947 - Dec 1947 returns use the (July) 1946 value of beme. Size is updated monthly so for Jan 1947 return we can use the Dec 1946 value of size; for Feb 1947 return we can use Jan 1947 size, etc.

1. We want to find the best way of representing the average return by a cross-sectional regression of averages of the right hand variables.

$$E\left(R_{t+1}^{ei}\right) = a + b \times E\left[f\left(beme_t^i\right)\right] + c \times E\left[g\left(me_t^i\right)\right] + ... + \varepsilon_i \;\; i = 1, 2, ..N$$

Your task is to figure out good functional form and run this regression.

(a) Plot $\log(beme_{it})$ and $\log\left(size_{it}\right)$ over time. You will see a decided trend in size. Of course, market cap is not a stationary variable, we cannot compare the size of GM in 1932 dollars to the size of...well maybe I chose a bad example, but you get the point. I solved this by using $me^* = me_{it}/\sum_i me_{it}$, i.e. fraction of total market capitalization, as my basic size variable. . beme also has a bit of a trend, as does d/p, but I decided to ignore this trend. Maybe a slow decline in beme does track a slow decline in average returns, and this is real.

(b) We're looking for a linear relationship between returns and characteristic. The average returns are (in this case) roughly linear functions of portfolio number, (if you don't remember this bar plot, make a plot of $E(R^{ei})$ in the 25 portfolios) so if we transform the right hand variables so portfolio means roughly linear in portfolio number we will end up with average returns linear in the characteristic. Make a bar plot or table of $E(beme)$ and $E(me^*)$, and a similar one for $E(\log(beme))$ and $E(\log(me^*))$ This should convince you that the raw values of these numbers are not a good idea, and you should use $\log(beme)$ and $\log(me^*)$ instead.

(c) OK, now we can try to fit the FF Table 1 Panel A facts with cross-sectional regressions

$$E\left(R_{t+1}^{ei}\right) = a + b \times E\left[\log\left(beme_t^i\right)\right] + c \times E\left[\log\left(me_t^{*i}\right)\right] + ... + \varepsilon_i \;\; i = 1, 2, ..N$$

Run the regressions and compare actual average returns $E\left(R_{t+1}^{ei}\right)$ with the fitted values. (I found bar plots of $E(R^{ei})$, the fitted value of the regressions, and ε_i to be useful. I also computed root mean square and average absolute errors. Remember that the FF model fits with about 10 bp errors.) An "actual vs. predicted" graph works well too.

Remember, we're not trying to "explain" average returns here – we're trying to "describe"' average returns. The fit should be quite good for this purpose. However, including $\sigma(E(R^e)) = \sigma(R^e)/\sqrt{T}$ gives you a sense of which patterns might just be sampling errors.

(d) You should find a good fit, except this model can't really capture the small growth problem. Try a cross term as well

$$E\left(R_{t+1}^{ei}\right) = a + b \times E\left[\log\left(beme_t^i\right)\right] + c \times E\left[\log\left(me_t^{*i}\right)\right] + d \times E\left[\log\left(me_t^{*i}\right)\right] \times E\left[\log\left(beme_t^i\right)\right] \ldots + \varepsilon_i \ \ i = 1, 2, ..N$$

This will fit the small growth better, but then overfit the large value. I ask you to do it to emphasize that there is nothing that limits us to linear functions on the right hand side. If we really want to match the pattern of sample expected returns, we clearly need a function that has a different size coefficient for deep growth stocks, but I don't want to pursue that here.

(e) *Note*: How to calculate standard errors: Consider for example a pure cross sectional regression of average returns on log beme

$$E\left(R_{t+1}^{ei}\right) = a + bE(\log(beme_t)) + \varepsilon_i; \ \ i = 1, 2...N$$

Now, the ε_i are correlated across portfoios, so standard OLS regression t statistics are wrong. The right answer is to plug it all into GMM, but I didn't do that and I don't expect you to do it either. Here are two simple approaches which assume residuals uncorrelated across time. I did both and they give the same answer.

   i. Define ε_{it} by
$$R_{t+1}^{ei} = a + bE(\log(beme_t)) + \varepsilon_{it+1}$$

   (yes, that's E and a,b are also constant over time on the right hand side) then we have

$$cov(\varepsilon\varepsilon') = cov(\varepsilon_t\varepsilon_t')/T$$

   Then we can use the formula

$$cov(\beta) = (X'X)^{-1} X'cov(\varepsilon\varepsilon')X(X'X)^{-1}$$

   ii. You can also Fama-MacBeth the standard errors. We're doing a pure cross sectional regression right now, not a Fama-MacBeth regression, and they are not the same thing since $beme_{it}$ changes over time. To use FmB to calculate the standard errors of a pure cross-section, you run

$$R_{t+1}^{ei} = a_t + b_t E(\log(bm_{it})) + \varepsilon_{it+1}; \ \ i = 1, 2, ..N \ \ \forall t$$

   again the right hand variable is the full sample mean here, not $beme_{it}$

$$\hat{b} = E(\hat{b}_t), \sigma(\hat{b}) = \sigma(\hat{b}_t)/\sqrt{T}$$

   Intuition: When you're all done, you've achieved a cross-sectional regression interpretation of FF Table 1 panel A, you've described the pattern of mean returns across 25 portfolios by a cross sectional regression. Furthermore, your cross sectional regression depends on *actual characteristics* not on *portfolio number* or *rank*.

2. Now, the heart of any asset pricing model is whether

$$E(R^{ei}|C_i) = \beta(R^{ei}, f|C_i)\lambda$$

We're trying to think about how the *function* on the left is matched by the *function* on the right, e.g. if we ran

$$
\begin{aligned}
E(R^{ei}|bm_i, me_i) &= a + c_{bm}E(\ln bm_i) + c_{me}E(\ln me_i) \\
\beta(bm_i, me_i) &= d + e_{bm}E(\ln bm_i) + e_{me}E(\ln me_i)
\end{aligned}
\tag{1}
$$

then we'd want to know whether $a = d\lambda$ and $c_i = e_i\lambda$. This isn't vacuous because there are three equations in one unknown λ, and we could add a fourth, $\lambda = E(hml)$.

Here is a very simple attempt, which will help us all to think about the concept. Do this at least for the indicated two-variable model; if you got a better characteristic model for means in part 1, then do that one as well[1].

(a) Let's just run a cross-sectional regression of FF3F betas on characteristics. We run each of the three betas separately on the characteristics, i.e. if you had bm and me in your cross sectional regressions it would be

$$
\begin{aligned}
a_i &= d_a + e_{a,bm}E(\ln bm_i) + e_{a,me}E(\ln me_i) + error \quad (2)\\
b_i &= d_b + e_{b,bm}E(\ln bm_i) + e_{b,me}E(\ln me_i) + error \\
h_i &= d_h + e_{h,bm}E(\ln bm_i) + e_{h,me}E(\ln me_i) + error \\
s_i &= d_s + e_{s,bm}E(\ln bm_i) + e_{s,me}E(\ln me_i) + error
\end{aligned}
$$

(Computing standard errors here needs us to develop new formulas, so let's not even try.) You should come up with a cross-sectional regression coefficient interpretation of FF's (Table 1) observation that market betas are about the same, hml betas rise with beme, and smb betas rise with (small) size.

(b) What about λ? Let's try the "time series regression approach" in which we use the means of the factors to estimate the factor risk premia. So, given the *coefficients* by which betas depend on characteristics, we can predict what the expected return *coefficients* on characteristics should be by comparing (1) and the weighted sum of the last three rows of (2)

$$
\begin{aligned}
a &= ? = d_b E\left(rmrf\right) + d_h E(hml) + d_s E(smb) \\
c_{bm} &= ? = e_{b,bm}E\left(rmrf\right) + e_{h,bm}E(hml) + d_{n,me}E(smb).
\end{aligned}
$$

So compute from the beta coefficients in a what the expected return coefficients *should be*. Compare the two: is $\sum \beta_i \lambda_i$ *as a function of characteristics* about the same as $E(R^{ei})$ *as a function of characteristics,* as revealed by the coefficients. Is there an interpretable difference?

(c) Well, let's compute the "alpha" *as a function of characteristics*, i.e. show its loadings on the right hand variables. Subtracting the coefficients in part b from the coefficients in (1). How

---

[1] The cross-sectional regression for the mean function is justified by the definition of mean; if

$$E(R^{ei}|C_i) = a + bC_i$$

then

$$R^{ei}_{t+1} = a + bC_i + \varepsilon_{it+1}; \quad E(\varepsilon_{it+1}) = 0$$

and

$$\frac{1}{T}\sum_{t=1}^{T} R^{ei}_{t+1} = a + bC_i + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it+1};$$

the cross-sectional error $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it+1}$ is mean zero and uncorrelated with the right hand variable. A similar argument justifies a cross-sectional regression for the covariance function.

$$
\begin{aligned}
cov(R^{ei}, F) &= d + eC_i \\
E\left[R^{ei}\left(F - E(F)\right)\right] &= d + eC_i \\
R^{ei}_t\left(F_t - E(F)\right) &= d + eC_i + \varepsilon_{it} \\
\frac{1}{T}\sum_{t=1}^{T} R^{ei}_t\left(F_t - E(F)\right) &= d + eC_i + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}
\end{aligned}
$$

So again, running the sample covariance on C_i is ok by OLS. The error is not likely to be normal, but that's not necessary. Betas are just a scale transformation. I like the idea of using covariances rather than betas, so that the regressions are single regression coefficients and do not depend on the number of factors, and so factor risk premiums λ are the same as b tests for whether to include the factor. But let's not innovate too much here.

do these implied alphas load on the characteristics? Is this the same or different from how the direct regression of alphas on characteristics worked in (2)?

(d) Make a plot to display the workings of this model. I did a bar chart with the predicted 5x5 values of $E(R^{ei})$, $b'\lambda$, $E(R^{ei}) - \beta'\lambda$ and the actual (not a fitted function of characteristics) α_i. I encourage graphical innovation however. Ideally, these should be functions of the characteristics not portfolio numbers, but again I wimped out of that.

(e) Now, can we do better by picking better factor risk premia?....A question for another day.

(This is all very obviously not the "right" thing to do. Under the null that betas are a linear function of characteristics, that null should inform beta estimation too. We should be using a kernel estimate. But that's also for another day. We did achieve a reinterpretation of all of FF's points in terms of cross-sectional regressions, and cross-sectional regressions on *characteristics*, not *portfolio numbers*)

3. Next, we want to get away from just using averages as the characteristic. We want a robust model, for example

$$E_t(R^{ei}_{t+1}) = a + b\ln(beme_{it}).$$

Doing this emphasizes that *Expected returns in Fama-French portfolios are exactly the same thing as forecasting regressions.* We want to make that connection.

This model means that variation of a *given* portfolio's beme *over time* generates the same variation in expected returns that you see by looking at the *cross-sectional* difference between one portfolio's beme and another's. Is it true? These are the two issues we'll consider in this problem.

Throughout the problem, ignore standard errors. We all know how to correct standard errors in panel data regressions.

(a) Let's start by running forecasting regressions and using only *beme.*Thus, as background, start with the cross sectional regression you ran last time,

$$E(R^{ei}) = a + bE(\ln(beme_i)) + \varepsilon_i \ i = 1, 2, ..N$$

(b) Now, let's run forecasting regressions.

   i. Run

   $$R^{ei}_{t+1} = a + b\ln(beme_{it}) + \varepsilon^i_{t+1}.$$

   This is a *pooled time-series cross-sectional regression.* You run it by stacking $R^{e1}_1 R^{e2}_1 R^{e3}_1 ... R^{e1}_2 R^{e2}_2$ etc. into one big vector. (Do it this way, not by T. You'll see why below.)

   ii. Run it with time dummies (equivalently, taking out the average of each right hand variable across portfolios at each moment in time).

   $$R^{ei}_{t+1} = a_t + b\ln(beme_{it}) + \varepsilon^i_{t+1}$$

   $$R^{ei}_{t+1} = a + b\left[\ln(beme_{it}) - \frac{1}{N}\sum_{i=1}^{N}\ln(beme_{it})\right] + \varepsilon^i_{t+1}$$

   This answers the question, "if portfolio 1 has more bm than portfolio 2 at any moment in time, does portfolio 1 have more subsequent return" but ignores the question "if portfolio 1 has more bm than it usually does is its return temporarily higher."

   iii. Run it with portfolio dummies (equivalently, taking out the average of each right hand variable across time for each portfolio)

   $$R^{ei}_{t+1} = a_i + b\ln(beme_{it}) + \varepsilon^i_{t+1}$$

   $$R^{ei}_{t+1} = a + b\left[\ln(beme_{it}) - \frac{1}{T}\sum_{t=1}^{T}\ln(beme_{it})\right] + \varepsilon^i_{t+1}$$

4

This answers the opposite question.

iv. So, is there a big difference between time and portfolio effects? Are we justified in writing

$$E(R_{t+1}^{ei}|beme_{it}) = a + b\ln(beme_{it})?$$

v. Obviously, we'd like a specification that does not refer to special characteristics of a portfolio.

Note: I did standard errors. These are not required, but you might be interested later. You can't use OLS standard errors, since the errors are cross-correlated. By stacking the vectors this way, however, you can use the standard error formula

$$\sigma^2(\hat{\beta}) = (X'X)^{-1}X'\Omega X(X'X)^{-1}$$

Now,

$$\Omega = \begin{bmatrix} \Sigma & & & \\ & \Sigma & & \\ & & \Sigma & \\ & & & \ddots \end{bmatrix}$$

where $\Sigma = cov(\varepsilon_t^{1..N})$, and denoting

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \end{bmatrix}$$

(each x_i is $N \times 1$) then

$$X'X = \sum_{t=1}^{T} x_i'x_i$$

$$X'\Omega X = \sum_{t=1}^{T} x_i'\Sigma x_i$$

This makes the program go much faster than constructing huge matrices. There must be a clever way to Fama MacBeth the same standard errors too.

Note 2: As you have probably guessed, time and firm dummies give different results. As advocated in "discount rates" I've been playing wit dynamic formulas, for example

$$R_{t+1}^{ei} = a + b\left[\ln\left(beme_{it}\right)\right] + c\left[\ln\left(beme_{it}\right) - \ln\left(beme_{it-1}\right)\right] + \varepsilon_{it+1}$$

The goal here is to get a specification for which time and firm dummies (and pooled) all give the same answer. We cannot use firm dummies when we go to individual stocks. I haven't quite gotten there yet... All for another day....

Readings questions

Give *short* (1-2 sentence) answers. Page numbers, table numbers and quoted numbers are appropriate.

Dissecting anomalies

1. New anomaly variables might just be proxying for value. For example, high B/M companies should have low investment (Q theory) so if investment forecasts returns, that fact might just be a proxy for the value effect. In Table II, do Fama and French control for this somehow?

2. Why are the t- statistics for the High-Low portfolio so much better than for the individual portfolios? Is this cheating?

3. Which anomalies produce strong average returns for all three size groups? What numbers in Table 2 document your answer?

4. Which anomaly gives the highest Sharpe ratio in Table 2?

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

6. On p. 1163, FF ask "Which anomalies are present in all size groups and produce returns that vary systematically from the low to the high ends of the sorts?" and they note that stock issues do not satisfy that criterion. Why is this important? How does Table III illuminate the issue?

7. p. 1667. "The novel evidence is that the market cap (MC) result draws much of its power from microcaps." What numbers are behind this conclusion?

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

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

10. P. 1675, FF say "the evidence from the sorts and the regressions is consistent with the standard valuation equation that says that controlling for B/M, higher expected net cash flows (earnings minus investment, per dollar of book value) imply higher expected stock returns." Isn't this the fallacy that "profitable companies have higher stock returns" or "confusing good companies with good stocks"?

11. 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? (Warning, this is a bit of a trick question)

12. What is the highest Sharpe ratio you can get from combining all these anomalies and exploiting them as much as possible? (Also a trick question)

Value and Momentum Everywhere

1. For stocks, value is price/book. But there's no "book" for commodities. How to Asness et al measure "value" for commodities, currencies, and bonds?

2. Is the alpha in Table 1 with respect to the CAPM or the FF 3 factor model?

3. What is the RW factor in Table 1?

4. Asness et al link the poor performance of momentum in Japan to the strong performance of value. How?

5. How does the "combo" portfolio do so much better than value and momentum alone? After all, the mean return of a portfolio is the average of the mean returns of its constituents.

6. p. 15 says "The last row of Panel A of Table 1 shows the power of combinbing value and momentum portfolios everywhere" and is proud of the larger Sharpe ratios. Are these Sharpe ratios indeed larger than what you'd expect if the individual strategies were uncorrelated?

7. What's the point of Figure 1?

8. Does the single "long value short momentum" factor explain average returns?

Lamont and Thaler

1. How are there two ways to buy Palm stock at different prices? What are the two prices?

2. Do L&T claim to have found an exploitable arbitrage opportunity?

3. If there is no way to make money, how can they say markets are inefficient?

4. Is a short Palm / long 3Com position riskfree at a monthly horizon?

5. How is "real world" shorting different from our frictionless textbook? What are the extra costs and risks?

6. Why do they say a short constraint lead to overpricing? Would it ever make sense to buy a stock that you *know* is overpriced, and there is no chance that the price will rise further (no "greater fool")?

7. Was there a lot of shorting in the "overpriced" subsidiaries? Was there more or less than in the parents? Did the sub shorting increase or decrease over time?

8. What do you learn from Fig. 5, 6?

9. If we can't short, let's buy November (data of spinoff) puts, or create a synthetic short position in options markets. Will this work and if not why not?

10. How do turnover and institutional ownership of Palm compare to that of 3Com? What conclusions do L&T draw from these facts?

11. p. 261 Is Palm more or less liquid than 3Com?

12. What happened to 3Com price during this episode? What conclusions to L&T draw?

Cochrane stocks as money questions and answers

1. According to Cochrane, how are money and bonds like 3Com/Palm ?

2. How does the monetary/convenience yield view say overpricing is associated with

   (a) Turnover
   (b) Supply
   (c) Short sales constraints

    (d) "Specialness" of the security (Palm, money); presence of substitutes

3. Is turnover associated with "overpricing" for 3Com/Palm?

4. How much does a typical Palm investor lose by holding Palm, not 3Com? Is this "a lot" or "not much"?

5. What's the point of Figure 5?

6. Wait, monetary theory says you are willing to put up with low returns on money because there is no substitute. If you want to bet on Palm, why not buy 3Com or use options instead? (Point to evidence here in Table 1, Figure 7.)

7. To Cochrane, the fact that 3Com *fell* is explained. How? What do Lamont and Thaler say about it?

8. What evidence does Cochrane give that this "money" story might apply more broadly?

9. Cochrane admits a big hole in the money comparison. What is it?

Brandt and Kavajecz questions

1. Why might price and signed volume be correlated, beyond the simple "price pressure" "downward sloping demand" "theory?"

2. How do B&K measure "orderflow?" You see a trade; how do you know if it's a "buy" or a "sell?" ("The market went up on a wave of buying" is a classic *fallacy* – for every buy, someone sold!)

3. Table IV: Central table. ***What does the number -0.72 in the top left corner of table IV mean? (This is a question about units – if x moves by what, what happens to y)

4. (Overall, how much of the unexpected daily change in yields is accounted for (notice I'm not saying "caused by"!) orderflow?)

5. There is a pattern in the coefficients of Table IV – which orderflows are most important for explaining each kind of yield change?

6. Is the orderflow effect stronger or weaker on days with big macro announcements? Why do we care?

7. How well would you do forecasting yield changes with one "order flow in all maturities" variable, rather than separate order flows?

8. What is the "inventory premium" view of the correlation between orders and price changes?

9. (The most important question) Overall, what three pieces of evidence lead Brandt to a "price discovery" view of the impact of order flow on prices, rather than the simpler view that "selling pressure does reduce prices after all" or there is an "inventory premium". (Hint: Tables IV VI and VII matter here as well as Brandt's discussion.)

10. Note p. 2641 "price discovery where substitutes are present tends to take place in the market that is most liquid." I.e. Palm vs. 3com