

Problem Set 5

Due Sat 12:00, week 5

Part 1 A. Simple very short readings questions

1. Suppose returns follow a factor model

$$R_{t+1}^{ei} = \alpha_i + \beta_{i1}f_{t+1}^1 + \beta_{i2}f_{t+1}^2 + \varepsilon_{t+1}^i.$$

What is the portfolio long the asset and short the factors with minimum variance? What is the Sharpe ratio of this portfolio?

2. Can utility be negative?
3. Suppose people are completely risk neutral, $u(c) = kc$. What is the value p_t of a payoff x_{t+1} in this case?
4. We approximate around $\delta = 0$ and $\Delta c = 0$,

$$m_t = e^{-\delta} e^{-\gamma \Delta c_{t+1}} \approx 1 - \delta - \gamma \Delta c_{t+1}$$

Verify this Taylor approximation, and find the corresponding *second* order approximation, i.e. including a Δc_{t+1}^2 terms? (Don't bother with the second order term in δ) Reminder

$$f(x) \approx f(0) + f'(0)x + \frac{1}{2}f''(0)x^2$$

5. If $\sigma(\Delta c) = 0.02$ (2%) and risk aversion is less than 10, what is the maximum Sharpe ratio we should observe?

Part 1 B simple theory problems

1. This problem is designed to help you relate the “theory” we do here to oranges / apples theory you did in microeconomics class. An investor has utility function

$$U(c_t, c_{t+1}) = \ln(c_t) + 0.95 \ln(c_{t+1})$$

- (a) Plot indifference curves $U(c_t, c_{t+1}) = \text{constant}$ for this investor (c_t on the x axis and c_{t+1} on the y axis). (A sketch is enough. If you want to plot with a computer, you're looking for $\ln(c_t) + 0.95 \ln(c_{t+1}) = k$ for a few different utility levels k . Solve for c_{t+1} on the y axis and plot.)
- (b) Find the slope $\partial c_{t+1} / \partial c_t$ at the 45 degree point $c_{t+1} = c_t$ by taking the total derivative of $\ln(c_t) + 0.95 \ln(c_{t+1}) = k$, and indicate that on your plot.
- (c) Find the optimal choice $\{c_t, c_{t+1}\}$ for this consumer if he/she has \$1000 of initial wealth and the interest rate is 10%. (Figure out the budget constraint, what c_t, c_{t+1} he/she can buy with $W_t = \$1000$, and then the best c_t, c_{t+1} . If you're having trouble, you might want to look at the maximization review in the Notes. The final answer isn't a round number.) Plot this maximum on the graph; show the budget constraint, the optimal consumption point. (Again, sketch is enough, but I welcome computer graphs. You'll learn more by having to program up all the various pieces.)

- (d) Does your consumer act patiently, consuming more at $t + 1$ than at t ? If so, why, given the 0.95 that says in the utility function that he prefers current consumption to waiting?

2. This problem is designed to help you think about risk aversion.

- (a) First, we'll find out what a "risk aversion coefficient" $-cu''(c)/u'(c)$ is. Suppose you only care about consumption tomorrow, so the utility function is simply $E[u(c)]$. You have enough income to support consumption \bar{c} for sure.
- Suppose you are forced to gamble a small fraction of ε of future consumption, thus the gamble is $\pm \varepsilon \bar{c}$ with 50% chance of each and your new consumption level will be $c = \bar{c} + \varepsilon \bar{c}$ or $c = \bar{c} - \varepsilon \bar{c}$ with equal probability. How much utility do you lose? Use a second order Taylor approximation to show that the answer is

$$Eu(c) - u(\bar{c}) = \frac{1}{2}u''(\bar{c})\bar{c}^2\varepsilon^2$$

(Hint: In case you forgot, the Taylor approximation is $f(x+\varepsilon) \approx f(x) + f'(x)\varepsilon + \frac{1}{2}f''(x)\varepsilon^2$.)

- Similarly, find out how much better off you are if you are given a gift $\delta \bar{c}$. (You only need a first order expansion here, $f(x + \varepsilon) \approx f(x) + f'(x)\varepsilon$.)
 - Now, for the important question: How much gift $\delta \bar{c}$ to I have to give you in order to compensate you for taking on the risk ε ? Equivalently, what compensation would be enough to bribe you to take the risk? Equate your answer in i and ii, solve for δ .
- (b) For the power utility function, $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$,
- Calculate the risk aversion coefficient $-\frac{cu''(c)}{u'(c)}$ by taking derivatives of the utility function
 - How is this "risk aversion coefficient" related to the bribe per unit of variance, δ/ε^2 from the last problem?

Part II. Empirical questions.

This problem helps you to think through what Fama and French did in the "Dissecting Anomalies" paper. It also gives you practice with a technique – cross-sectional regressions – that I think will take over from this business of looking at portfolio averages, because you can't slice portfolios more than two ways. ("Discount Rates" goes on about this point, and you can see FF struggling with a 5 way sort in "Five factor model.") The objective is simply to characterize the means of FF's 25 portfolios in the same way that FF did in Table IV of "dissecting anomalies."

Load the size and beme data from the class website. These are the actual size (market equity = me) and book to market equity $beme$ of each portfolio at each point in time. Use returns in the 196301-now subsample. Our objective is to see how average returns of these 25 portfolios $E(R^{ei})$ correlate with the *characteristics*, size and book/market. We want to do this with regressions, not just portfolio means as Fama and French 1996 did in Table 1 panel A.

In regression classes you learned always to plot the data first, and adjust the units (levels, logs, square roots, ratios) and think about functional forms (linear, quadratic, log, etc.) I did this for you and it turns out you want to transform the data as follows. Use log of size as a fraction of total market value,

$$size_{it} = \log \left(\frac{me_{it}}{\sum_{i=1}^N me_{it}} \right)$$

and use the log of the book/market ratios B/M

$$bm_{it} = \log(beme_{it})$$

Create these transformed variables. (The appendix on transformations below explains in some detail why you want to make these transformations, and see the programming hints even if you're not using matlab.)

Part II A: Cross sectional regressions of returns on characteristics

1. Run a *cross-sectional* regression to characterize average returns in terms of size and book/market, i.e.

$$E(R^{ei}) = a + bE[size_{it}] + cE[bm_{it}] + \varepsilon_i; \quad i = 1, 2, \dots, 25$$

This is a regression with 25 data points, one for each portfolio. Report the coefficients a, b and c and the R^2 . Do these coefficients make sense? Are size and bm related to average returns the way you think they should be? Does this look something like Dissecting Anomalies Table IV? (Don't bother with t statistics yet. The standard OLS t statistics are wrong. We'll learn how to do standard errors for this kind of regression later. Just interpret coefficients.)

2. To see how well the regressions capture the pattern of average returns in the 25 portfolios, make a table or bar plot of average returns (like my 5x5 bar plot last week) with the fitted value from these models. I.e. state or plot the 25 values of

$$a + bE[size_{it}] + cE[bm_{it}]$$

Compare this to the actual 25 mean excess returns $E(R^{ei})$. You should see that this regression does a pretty good job of capturing the mean returns of the 25 portfolios. It won't be perfect of course: we are capturing 25 numbers with 3 free parameters, and there is no real reason the relation should be so nicely log-linear.

3. You ought to find a decent fit, but as usual not so good among the growth firms with small growth an outlier. As you look at the pattern of average returns and the pattern of $a + bE[size_{it}] + cE[bm_{it}]$, the plots should suggest to you that we need a cross term. We'll fit average returns a lot better with

$$E(R^{ei}) = a + bE[size_{it}] + cE[bm_{it}] + d(E[size_{it}] \times E[bm_{it}]) + \varepsilon_i; \quad i = 1, 2, \dots, 25$$

Run this regression, and compare the actual average returns with their fitted values. Does the model with a cross term fit better, especially in the small growth dimension?

4. Are you allowed to put cross terms in here? Hint: Do these regressions constitute an *explanation* of average returns? If we were fitting the FF model, and the fit were better with a cross term

$$E(R^{ei}) = \alpha_i + b_i \lambda_m + s_i \lambda_s + h_i \lambda_h + (s_i \times h_i) \lambda_{sh} + \varepsilon_i; \quad i = 1, 2, \dots, 25$$

where b_i , s_i and h_i are slope coefficients in time series regressions

$$R_t^{ei} = \alpha_i + b_i r_{mr} f_t + s_i smb_t + h_i hml_t + \varepsilon_i; \quad i = 1, 2, \dots, 25$$

would that be a good way to improve the model in the same way to better fit the small growth stocks?

5. The stuff we're doing here is related to predictability. Given what we did week 1, if someone said "maybe high book to market ratios tell you good returns," you might have been tempted to implement that idea by running forecasting regressions

$$R_{t+1}^{ei} = a + b \times size_{it} + c \times bm_{it} + \varepsilon_{it+1}. \quad T = 1, 2, \dots, T \tag{1}$$

But if you did this, then, taking the expectation of both sides, we have

$$E(R_{t+1}^{ei}) = a + b \times E(size_{it}) + c \times E(bm_{it})!$$

and this is exactly the cross-sectional regression. The time series forecasts and these averages of returns based on size and bm forecasts are basically the same thing.

Let's try it. Run the forecasting regression (1). Rather than run 25 different forecasting regressions, just stack up the forecasts for the 25 portfolios in one regression. This is called a *pooled time-series cross-sectional regression*. See hints. Be glad you're not using excel. Compare your coefficients b and c here to those of part 2. Roughly the same? Different?

6. To diagnose why you're getting slightly different answers, let's break up each portfolio's size and bm into its *average value*, and its *variation over time around that average*. Mathematically,

$$size_{it} = E(size_{it}) + [size_{it} - E(size_{it})]$$

where $E()$ is the average over time. So, first, run the forecasting regression

$$R_{t+1}^{ei} = a + b \times E(size_{it}) + c \times E(bm_{it}) + \varepsilon_{it+1}.$$

and then run the forecasting regression

$$R_{t+1}^{ei} = a + b \times [size_{it} - E(size_{it})] + c \times [bm_{it} - E(bm_{it})] + \varepsilon_{it+1}.$$

Run both of these regressions pooled over both i and t . Can you see why the regression (1) gave different coefficients from the cross-sectional regression? There is an important lesson here: Fama and French left a lot of forecast power on the table! (The "Discount rates" appendix reading takes up this issue. You're running a "panel data regression with portfolio dummies" here.)

Part II B. Standard errors and Fama MacBeth regressions

1. Now, let's think about standard errors. To keep it simple, let's use the simple specification of returns on size and beme. Refresh question one, a simple cross-sectional regression of average returns on average size and bm,

$$E(R^{ei}) = a + bE[size_{it}] + cE[bm_{it}] + \varepsilon_i; \quad i = 1, 2, \dots, 25 \quad (2)$$

Report the coefficients a,b,c; conventional standard errors of a,b,c, and t statistics.

2. Now, these standard errors are wrong. The formula for the standard error assumes that ε_i are uncorrelated with each other. They are not. If the (2,3) portfolio has an unusually high average return, then the (2,4) portfolio is also likely to have an unusually high return. That's really FF's point – the three factors drive almost all return variation, so the portfolios don't go off their own way. We really only have 3 data points not 25. So we need to *correct for cross-sectional correlations of the error terms*.
3. The Fama-MacBeth technique is one way of producing correct standard errors, among other things. And we read so much about FMB regressions that I want you to run one anyway. Here's how.

Run a separate cross-sectional regression at each date (in a loop!)

$$R_{t+1}^{ei} = a_t + b_t size_{it} + c_t bm_{it} + \varepsilon_i; \quad i = 1, 2, \dots, 25 \text{ at each date } t. \quad (3)$$

Keep track of $T \times 1$ vectors a_t, b_t, c_t . Your estimates are the average of these individual cross sectional regression estimates

$$\hat{a} = E(a_t) = \frac{1}{T} \sum_{t=1}^T a_t; \text{ etc.}$$

More importantly, you can use the variation of the $\hat{a}_t, \hat{b}_t, \hat{c}_t$ over time to measure the standard errors of the coefficients.

$$\sigma(\hat{a}) = \frac{\sigma(a_t)}{\sqrt{T}}; \sigma(\hat{b}) = \frac{\sigma(b_t)}{\sqrt{T}}; \text{ etc.}$$

here on the left, $\sigma(\hat{b})$ means standard error, the thing you're looking for, and $\sigma(b_t)$ means the standard deviation over time, i.e. if bt is a $T \times 1$ vector of b_t estimates, $\sigma(b_t) = \text{stdev}(bt)$. Yes, our friend σ/\sqrt{T} can apply to regression coefficients too!

Add these Fama-MacBeth coefficient estimates, standard errors and t statistics to your table. The coefficients should be about the same as you found before. The standard errors and t statistics should tell you that the regular formula in part a was much too optimistic.

4. We can also correct standard errors directly: there is a formula for the correct standard errors. The little note below explains the technique. Writing the regression (2) as

$$\begin{aligned} E(R^e) &= xb + \varepsilon \\ (25 \times 1) &= (25 \times 3)(3 \times 1) + (25 \times 1) \end{aligned}$$

the standard formula is

$$\sigma^2(\hat{b}) = (x'x)^{-1}\sigma^2(\varepsilon)$$

If you used a regression package in the first case, start by computing these standard errors to make sure you're on the right track. (They may not be exactly the same, as the regression package will use a fancier estimate of $\sigma(\varepsilon)$. Just get close.) If you computed standard errors using this formula already, you're done.

5. Now, form residuals *at each date*, and form the covariance matrix of these residuals.

$$\begin{aligned} \varepsilon_{t+1} &= R_{t+1}^e - x\hat{b} \\ (25 \times 1) &= (25 \times 1) - (25 \times 3)(3 \times 1) \end{aligned}$$

i.e.,

$$\varepsilon_{i,t+1} = E(R^{ei}) - \hat{a} - \hat{b}E[size_{it}] - \hat{c}E[bm_{it}]$$

Then form

$$\Sigma = \text{cov}(\varepsilon_{t+1}, \varepsilon'_{t+1}) \quad (25 \times 25)$$

The correct formula for standard errors is

$$\sigma^2(\hat{b}) = (x'x)^{-1} \frac{\Sigma}{T} (x'x)^{-1}$$

Compute these standard errors, and put them in your table. They should be close, though not exactly the same, as FmB standard errors.

Congratulations! You have run a Fama-MacBeth regression, and produced standard errors that correct for cross-sectional correlation. This situation happens all the time in finance.

Note on standard error corrections.

This note explains why my standard errors work.

Think of the regression

$$y = xb + \varepsilon$$

where y is a $N \times 1$ vector of y data, x is a $N \times K$ vector including a column of ones, b is a $K \times 1$ vector and ε is an $N \times 1$ vector of shocks. The OLS estimate of the regression coefficient is

$$\hat{b} = (x'x)^{-1} x'y$$

The standard error is the square root of the variance of the estimate \hat{b} . And

$$\hat{b} = (x'x)^{-1} x'(xb + \varepsilon) = b + (x'x)^{-1} x'\varepsilon.$$

So,

$$cov(\hat{b}) = (x'x)^{-1} x'cov(\varepsilon, \varepsilon')x(x'x)^{-1}.$$

(I'm using $cov(A\varepsilon, (A\varepsilon)') = Acov(\varepsilon, \varepsilon')A'$.)

Conventionally, we assume errors have the same variance and are uncorrelated with each other, so

$$cov(\varepsilon, \varepsilon') = \sigma^2(\varepsilon)I$$

where I is the identity matrix. Then the $x'x$ cancel and we have

$$cov(\hat{b}) = (x'x)^{-1} \sigma^2(\varepsilon).$$

So, you calculate the variance of the residuals, multiply by $(x'x)^{-1}$, look at the diagonals, and take the square root. `b=x\y; e=y-x*b; stdb = diag(inv(x'*x)*var(e)).^0.5;`

We don't want to make that assumption. Alas, from just looking at the residuals we don't know what their covariance is. But the structure of our asset pricing problem lets us measure the covariance of the cross sectional errors! We're running

$$E(R^e) = xb + \varepsilon$$

If we write

$$R_{t+1}^e = xb + \varepsilon_{t+1}$$

then

$$E(R^e) = E(R_{t+1}^e) = \frac{1}{T} \sum_{t=1}^T R_{t+1}^e,$$

$$\varepsilon = E(\varepsilon_{t+1}) = \frac{1}{T} \sum_{t=1}^T \varepsilon_{t+1},$$

and most of all

$$cov(\varepsilon) = cov\left(\frac{1}{T} \sum_{t=1}^T \varepsilon_{t+1}\right) = \frac{1}{T} cov(\varepsilon_{t+1})$$

(I use $cov(\varepsilon)$ and $cov(\varepsilon, \varepsilon')$ equivalently.) I am assuming here that returns are uncorrelated *over time*, though correlated *across portfolios*, which is the key to the magic. You have to assume something.

So we have a new formula for standard errors. In sum, after running

$$E(R^e) = x\hat{b} + e,$$

(hat means estimate, e is residual where ε is true error) form a residual at each date,

$$e_{t+1} = R_{t+1}^e - x\hat{b},$$

take its covariance

$$\Sigma = \text{cov}(e_{t+1})$$

and form standard errors from

$$\text{cov}(\hat{b}) = (x'x)^{-1} x' \frac{\Sigma}{T} x (x'x)^{-1}.$$

Programming hints

To run the huge forecast with 25 portfolios and many time periods, you have to get the size and beme data to match your returns data. Size is easy: use December 1962 size for the Jan 1963 return, etc. Beme is harder, because you only have annual data. You want to use the 1962 value of beme for the whole year 1963. To expand the annual beme data to monthly you can use the kron command. Try

```
kron([1;1],[3;4]); kron([3;4],[1;1])
```

and you'll see how this nice function works. You can also do it in a loop: for Jan 1963 use the 1962 value, for Feb 1963 use the 1962 value...for Dec 1963 use the 1962 value, then for Jan 1964 use the 1963 value etc.)

Both mean and sum have a second argument. If you have a $T \times N$ matrix x ,

```
mean(x), mean(x,1)
```

produce the same thing, a $1 \times N$ row vector with the column means of x in each element. Sum works the same way. But

```
sum(x,2) mean(x,2)
```

take the sums and means horizontally, producing a $T \times 1$ column vector of the row means and sums of x .

You can use this fact to take out the market size for each month in one step:

```
logmc2 = log(mc./(sum(mc,2)*ones(1,25)));
```

Now, how does that work? mc is a $T \times 25$ matrix. $sum(mc,2)$ produces the row sum, which is $T \times 1$ column of market sizes at each date. Now, I multiply by $ones(1,25)$ to produce $T \times 25$ where the rows are all the same. $./$ then takes each mc and divides by the market size at that date.

To put all the data in one big regression, you can use

```
y = reshape(rx,T*25,1)
```

and similarly for the x variables. This produces a big vector with 25 time t variables vertically, then 25 time $t+1$ variables, and so on.

The same trick I used for $logmc2$ will give you the time averages and differences from time averages

```
logmc4 = ones(T,1)*mean(logmc2);  
logbeme4 = ones(T,1)*mean(log(beme));  
x = [ones(T*25,1) reshape(logmc4,25*T,1) reshape(logbeme4,T*25,1)];
```

The first two lines create a row vector of means for each portfolio, and then the $ones(T,1)*$ operation uses that mean in place of the variable everywhere. Then I use reshape to stack them up again. To do the differences from means

```
logmc4 = logmc2 - ones(T,1)*mean(logmc2);
```


For Fama MacBeth, here is what I did

```
btv = zeros(T,3); % set up room for bt
for t = 1:T;
    y = rx(t,:)' ; % data is set up so this is r_t+1
    x = [ones(25,1) logmc2(t,:)' log(beme(t,:))'] ;
    [bt,bse,btstat,r2,F] = regress_jc(y,x,0);
    btv(t,:)=bt' ; % save this time's regression
end;

bfb = mean(btv)' ; % FMB coefficients
sefb = std(btv)'/T^0.5; % FMB standard errors
```

Appendix on transformations.

You don't have to read this. It documents how I fished around to decide what regression you should run, why logs etc.

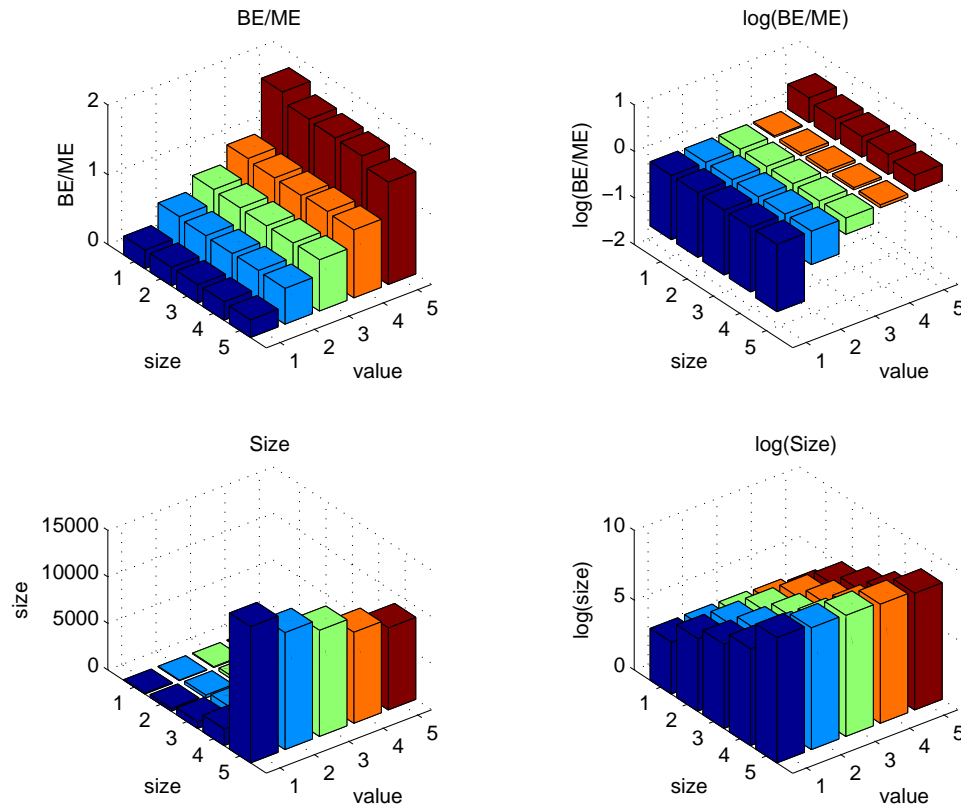
I asked you to transform variables before starting, to

$$size_{it} = \log \left(\frac{me_{it}}{\sum_{i=1}^N me_{it}} \right)$$

and to use log of B/M

$$bm_{it} = \log(beme_{it})$$

Why? FF found that average returns are basically linear across the portfolio numbers. But the large stocks are much, much, bigger than everyone else. Size and beme are not at all linear across portfolios. Here are plots of the actual and log size and beme in the 25 portfolios.



As we might have expected, size (especially) and beme are skewed across portfolios. There are a lot of small firms, “most” firms are very small. This is an independently important point – *almost all the value of the NYSE is in the largest-cap portfolios*. PAY ATTENTION TO THAT. This is why FF spend so much time on value vs equal weighting, and distinguishing which effects are limited to dusty corners of tiny stocks.

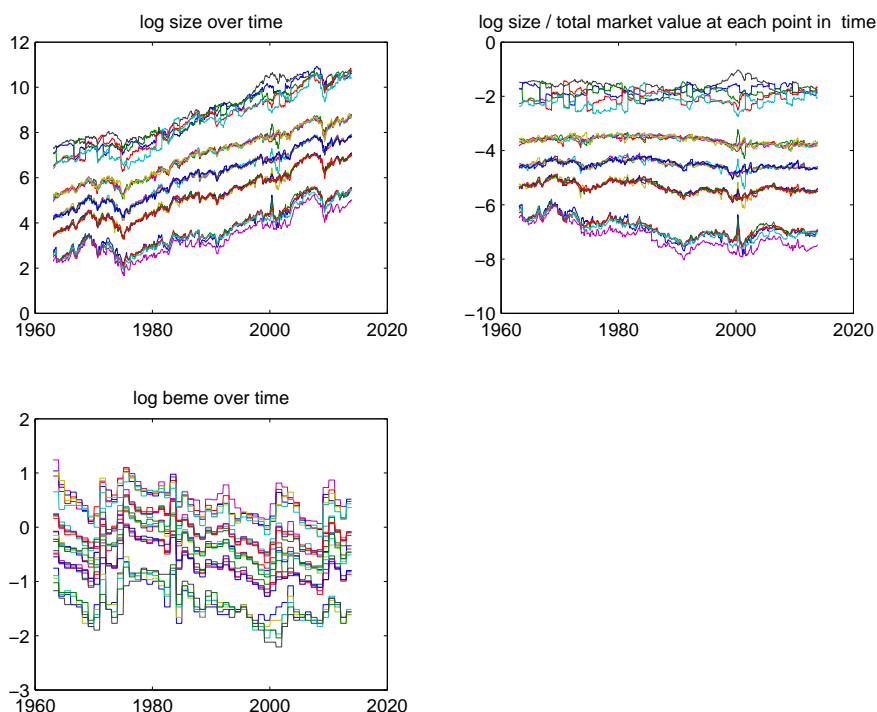
BEME is also not evenly distributed; the “value” portfolios with near zero me have much higher beme. Logs (right panel) is much smoother across portfolios. If average returns were also skewed across portfolios this wouldn't be a problem – we're looking for average return = linear function(characteristic).

We know the ER we're trying to model vary smoothly across portfolios, so the right hand variables should also vary smoothly. For that purpose, logs look like the right units.

This is the “functional form” question you studied in regression class. They taught you to think about units, and functional form by making plots before running regressions. We're doing that here.

The second reason for transforming size is that me trends up through time; a “large” firm in 1960 dollars is a “puny” firm today. The FF portfolio assignment treats the largest 1/10 of firms in 1960 as behaving the same as the largest 1/10 firms today. Taking out the mean at each date accomplishes that.

Here are plots of log beme and log me over time.



You can see that size trends up. It makes no sense to correlate the same *absolute* value of size in 1962 with returns in 1963 as the size in 2009 that we correlate with returns in 2010. The second graph shows that my transformation takes out most of the trend. There is less trend in B/M, which makes sense in that it is a ratio.

These transformations do not take out all the trend in size and beme, and actually I think that's good. FF's portfolios take out all variation in beme over time. Suppose (week 1) prices rise so expected returns on everything is lower. Well, prices rose for the top 1/10 as well as for the bottom 1/10 so though everyone's beme lower, and the portfolio returns are lower, we don't see that information. Regressions can link the behavior of beme over time to subsequent returns, and portfolios throw that out. (You can also throw it out by adding time-dummies to the regression, so that it does not use any information over time. That would have been a separate and maybe easier way to handle the size trend.)

Lesson: *Units matter in cross-sectional regressions!* Actually, units matter in everything. It's not enough to say “b/m is associated with returns,” you have to think about the functional form, and you have to think whether time-series or cross-sectional variation in b/m have different relations with following returns.