Business 35904 **John H. Cochrane**

Problem Set 8

Time series, cross section and all that.

(This is a much shortened and updated version of Ch. 13 problem 3.)

We're going to program up time-series and cross sectional regressions. Along the way, I'll show you some common traps and we'll learn something about Fama-French and the CAPM.

Get the 25 Fama French portfolios and the Fama French factors from the class website. (The original data are from Ken French's website. http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html) Eliminate the data prior to 193201 (i.e. 67:end) to avoid the missing data in the earliest years of the sample. (Make sure you don't use any -99 data). Subtract the risk free rate rf from the 25 test assets to make them excess returns. The factors are already excess returns.

Write a program to run time-series and cross-sectional regressions for the CAPM and Fama French 3 factor models. By writing a program, you can write the code once to do CAPM, FF3F, and subsamples. Let the program take as input an arbitrary $T \times N$ matrix of test asset excess returns, and a $T \times K$ matrix of factors. The program should do the following

1. Report mean returns like Fama and French's Table 1a. Report all 25 long vectors in 5 $\times$5 matrices in small and value dimensions like Fama and French.
2. Run OLS time-series regressions to find parameter estimates of $\hat{\alpha}, \hat{\beta}, \hat{\lambda}$.
3. Find standard errors of $\hat{\alpha}, \hat{\beta}, \hat{\lambda}$using
   (a) Classic iid formulas
   (b) GMM, using 4 lags and Newey-West weighting.
   Note: the point of asset pricing models is to see if the *spread* in average return is matched by the *spread* of β, so we should be testing whether expected returns and betas are *different from each other.* Nobody does this, so I won't make you do it, but we should!
4. Compute the $\hat{\alpha}' V^{-1} \hat{\alpha}$ tests: Compute the traditional χ^2 test, the GRS F test, and the GMM test. Compute the test statistic, the 5% critical value, and the probability value (chance of seeing a statistic as high or higher than observed in sample. The cdf function will come in useful here). As a diagnostic, also present the root mean square and average absolute alphas.
5. Now, run OLS cross-sectional regressions.
   (a) Run the cross sectional regression with a constant, $E(R^{ei}) = \lambda_0 + \beta_i \lambda + \alpha_i$. There are standard error etc. formulas for this case, but I'll spare you programming them up. Report
      i. Estimates $(\hat{\lambda}_0), \hat{\lambda}$,
      ii. Root mean square and mean absolute pricing errors $\hat{\alpha}$, and the R^2 of actual vs. predicted mean returns.
   (b) Run it with no constant, $E(R^{ei}) = \beta_i \lambda + \alpha_i$, and include the factors (market) as a test asset.
      i. Collect estimates $\hat{\alpha}, \hat{\lambda}$,
      ii. Compute standard errors of $\hat{\lambda}$ and $\hat{\alpha}$ with traditional iid formulas. You do not have to present the $\hat{\alpha}$ standard errors, but you'll need them next. (The tables are getting big!)
      iii. Compute standard errors of $\hat{\lambda}$ and $\hat{\alpha}$ with GMM formulas.

iv. Compute the $\hat{\alpha}' V^{-1} \hat{\alpha}$ test with statistic, 5% critical value, p-value, again both from iid and GMM formulas.

v. Again compute the root mean square and mean absolute pricing errors, and compute the R^2 of actual vs. predicted mean returns. (See note below on cross-sectional R^2)

6. Finally, present average returns vs. betas for all three methods. Include dots for the test assets and factors, and include the lines fit by each of the three methods.

Let's take this out for a spin and do some tests. The actual problem set:

1. Evaluate the CAPM, starting in 194701 and ending now. This fills a niche; Fama and French did not show in Table 1 that the CAPM does not work. Perhaps the high average returns are associated with high CAPM betas

   (a) Do you see a pattern that high average returns are associated with high betas?

   (b) Is the market premium $\hat{\lambda}$ significant in the time series estimate?

   (c) Is the CAPM rejected in the time-series estimate?

   (d) How does the estimate of $\hat{\lambda}$ and the $\hat{\alpha}$ in the cross sectional regression with a free constant differ from that of the time-series regression? Track down where this difference comes from – there is an important lesson here. (Hint: look at the expected return vs beta graphs.)

   (e) How does the estimate of $\hat{\lambda}$ when we force the intercept to zero and price the factors compare?

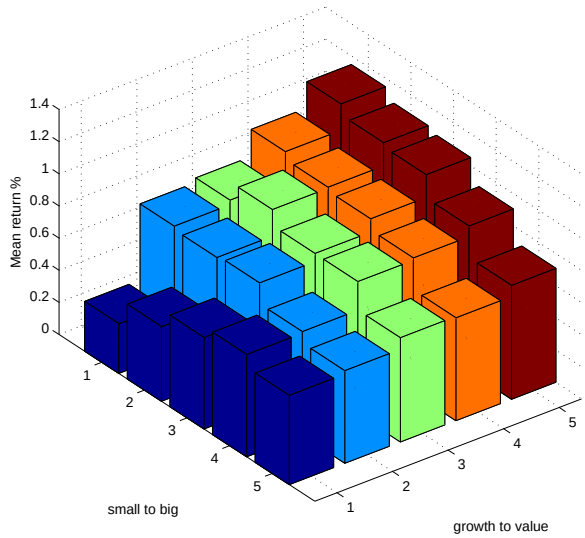
2. Repeat, using the sample from 193201 to 196212, i.e. before the Fama and French sample. You should see a big and surprising difference in the performance of the CAPM. (You don't need to reproduce all the numbers. Just reproduce what's important for evaluating the performance of the CAPM. Include the graphs.)

Hints. My output has a big set of 5x5 tables with means, alphas, alpha standard errors, then betas, beta standard errors, and R2. Then I have a comparison of λ estimates and test statistics like this

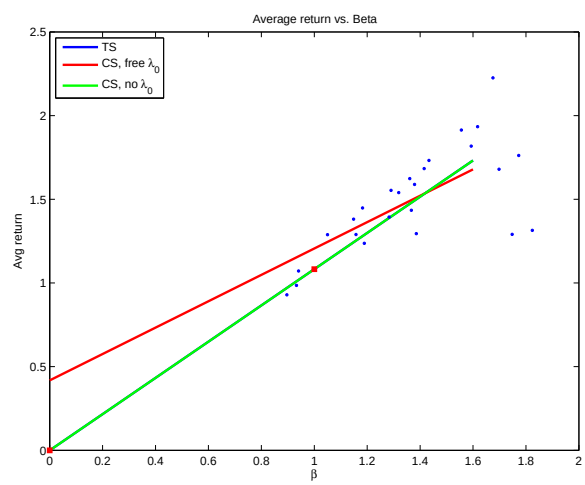
```
capm
                lambda0  lambda  chi2/F      5%      1%     % p
 Time Series
        E(f)      0.00   (fill in numbers)
   iid se/GRS     0.00
     iid chi2     0.00
      gmm         0.00

 Cross section, free constant, factors not priced
    estimate
       iid se
       gmm se

 Cross section, no constant, factors priced
     estimate     0.00
       iid se
       gmm se
```

```
alpha diagnostics
               rmse a  mean|a|   R2_1    R2_2    R2_3
   ts
   cs
   cs price f
Note: R2_1 =  var(beta*lambda)/var(Ere); R2_2 = 1 - var(alpha,1)/var(mean(rx),1); R2_3 = 1 - mean(alpha
```

To help understand the table, I made two graphs. First I used the bar3(reshape(x, 5,5)')) command to make plots out of the 5 x 5 tables, for example mean returns:



Second, I made graphs of mean vs. beta. This is NOT what your graph will look like because it's for the full sample.



Note on cross-sectional R^2 The cross-sectional R^2 is a natural measure of how well an asset pricing model is doing to capture mean returns. Do the points in your $E(R^e)$ vs β graph all line up

around the line or are they widely spread? In a regression $y = a + xb + \varepsilon$ there are two definitions of R^2

$$R^2 = \frac{var(xb)}{var(y)};\ R^2 = 1 - \frac{var(\varepsilon)}{var(y)}$$

These suggest for our cross-sectional regression $E(R^e) = \gamma + \beta\lambda + \alpha$,the measures

$$R^2 = \frac{var(\beta\lambda)}{var(E(R^e))};\ R^2 = 1 - \frac{var(\alpha)}{var(E(R^e))}$$

(*var here is across assets not over time*). However, the constant is important to us as well. If all the alphas are 1 then $var(\alpha) = 0$, and $R^2 = 1$ computed either way, but it's obviously a terrible model since the points are all on a line that is too high up. That suggests a third measure,

$$R^2 = 1 - \frac{E(\alpha^2)}{var(E(R^e))}$$

Compute all three measures of R^2 for both the time-series and cross-sectional estimates. Make sense of the differences between these measures! Though I emphasize these nice graphs of average returns vs. betas, this problem will show some of the dangers of the technique. Also, one can reform the portfolios to make the graphs and these cross-sectional R^2 measures look as pretty or bad as one wants.

Computer hint: this will work better if you use $var(z) = \frac{1}{N}\sum(z_i - \bar{z})^2$ to compute variance, not $var(z) = \frac{1}{N-1}\sum(z_i - \bar{z})^2$. The matlab var(z) uses $1/(N-1)$, while var(z,1) uses $1/N$.)