

Problem Set 1

Due in class, Wed week 2

1. Assemble data and verify basic regressions. We're going to forecast returns using dividend yields and using Lettau and Ludvigson's cay variable.

Get the CRSP value-weighted return with and without dividends from CRSP (via wrds is easiest). Get annual data (we'll keep it easy for now) as long as you can. I got all markets (nyse, amex, nasdaq), but this doesn't really matter. Also get the 3 month treasury bill return. Form the dividend price ratio D_t/P_t and dividend growth rate D_t/D_{t-1} from the returns. (vwretd = $(P_t + D_t)/P_{t-1} - 1$, vwretx = $P_t/P_{t-1} - 1$). Get cay from Sydney Ludvigson's website <http://www.econ.nyu.edu/user/ludvigsons/>. Convert cay to annual by taking the 4th quarter observation each year.

Plot dividend yield, stock return, and risk free rate to make sure your data are correct. Plot your data in percent. (It's always a good idea to check you got units and data transformations right!)

2. Reproduce the basic forecasting regressions

$$\begin{aligned} R_{t,t+j} &= a + b(D/P)_t + \varepsilon_{t+1} \\ R_{t,t+j} - R_{t,t+j}^f &= a + b(D/P)_t + \varepsilon_{t+1} \\ \Delta D_{t,t+j} &= a + b(D/P)_t + \varepsilon_{t+1} \end{aligned}$$

for $j = 1:10$, where $R_{t,t+j}$ represents the total return from time t to time $t+j$. Use overlapping annual observations for the full sample.

- (a) Tabulate and compare the coefficients and R^2 at the different horizons. Explain the patterns you see.
- (b) Are OLS coefficients or R^2 biased by the use of overlapping data?
- (c) For the first regression, compare
 - i. OLS standard errors and t statistics
 - ii. Standard errors and t statistics that correct for correlated residuals due to overlap (See "Hansen-Hodrick errors" in *Asset Pricing*." It's much better for you to program these up yourself than to use canned software.)
 - iii. Standard errors and t statistics from non-overlapping data.

To understand the effects of standard error calculations, use the same coefficients and compare t statistics using the three different standard errors.

- (d) Plot the actual and forecast 7 year returns, and actual and forecast 7 year dividend growth rates. Line up the forecast with the actual, i.e. plot $a + b(DP)_t$ together with $R_{t,t+7}$.

3. Now, let's play with a multivariate forecast. Using the shorter sample for which you have both series, add cay to the regressions, i.e., run

$$\begin{aligned} R_{t+j} - R_{t+j}^f &= a + b(D/P)_t + \varepsilon_{t+1} \\ R_{t+j} - R_{t+j}^f &= a + c \times cay_t + \varepsilon_{t+1} \\ R_{t+j} - R_{t+j}^f &= a + b(D/P)_t + c \times cay_t + \varepsilon_{t+1} \end{aligned}$$

Use overlapping data with HH corrected standard errors.

- (a) Start by comparing the DP-only results to the above. Are things different in this shorter sample? (Look at point estimates, not just t stats. t stats are always bigger in shorter samples.)
- (b) How do cay results differ from DP results as you increase horizon? Why?
- (c) Does one of DP or CAY seem to drive the other out in the multiple regressions? Is this different at different horizons?
- (d) You should find that adding cay makes DP much more significant, despite not changing its coefficient much. How is this possible?
- (e) Make a plot of actual vs. forecast returns at the one and 7 year horizons. How do DP and cay seem to complement each other in forecasting?