FNCE 921
Professor Nick Roussanov
Spring 2019
Assignment #1

Due Monday February 18^{th}*, 2019*

This problem set is mostly empirical in nature–you should work in groups of about 2 and at most 3 people. The goals are: you get a sense of the basic stylized facts of the data, familiarize yourself with the methods of empirical analysis, and build a toolbox for conducting future research.

1. Download the VWRET (total value weighted return, i.e. cum-dividend) and VWRETX (return based only on price changes, i.e. ex-dividend) from WRDS for the period 1929-2016. Use quarterly and annual data (you can repeat your analysis using both for extra credit - note that results might be different). Use these two returns to construct a series for dividend growth and for the price dividend ratio. Use WRDS to also download the return to 3-month and 12-month Treasury Bill and CPI. Use this data to construct the real risk free rate subtracting an expected inflation measure from the Treasury Bill yield. The expected inflation measure should be the AR(1) fitted value of inflation. See some description on data sources and conventions at the end of this assignment.

   (a) Describe the mean and standard deviation of the excess return on the market, dividend growth, the risk free rate, and the price dividend ratio.

   (b) Plot dividend growth and assess whether the series displays seasonality (for quarterly series)?

   (c) If so, one way to treat seasonality is to use D_t/D_{t-4} –that is the growth rate over a year.

   (d) Test for a structural break in dividend growth in 1973.2 —see Lettau and Van Nieuwerburgh "Reconciling the Return Predictability Evidence", RFS, 2007, pages – 1608-1652.

   (e) Plot the log of D/P ratio to make sure that it is OK (i.e. does not explode).

2. Run a regression of 1,2,3,4,5,10 year returns on log D/P. Are returns forecastable? Contrast the subsamples 1926-1990, 1947-1990, 1947-now and 1973-now with the full sample.

3. Calculate one return at the end of the full sample that would bring D/P back to its historical average. (Assume no change in dividends, $\frac{D_{T+1}}{D_T} = 1$.) What would the coefficient look like if we see this one last crash?

4. Let $X_t = [d_t - p_t, d_t - d_{t-1}, r_{ft}]'$ where small letters refer to the log of a variable and r_{ft} is the risk free rate. If you are using quarterly data and chose to use annual growth to eliminate seasonality in dividend growth treat $d_t - d_{t-1}$ in X_t as a stand in for $d_t - d_{t-4}$. Run the following VAR(1)

$$X_{t+1} = A_0 + A_1 X_t + \epsilon_t$$

   where A_0 is a 3×1 vector and A_1 is a 3×3 matrix.

(a) First report the A_0 and A_1 coefficients and R^2s. Construct the standard errors (HAC-robust).

(b) Is dividend growth predictable using the lagged dividend yield?

(c) Is dividend yield predictable by lagged dividend yield?

5. Given the VAR above we can also infer return predictability using the the Campbell Shiller approximation:

$$r_{t+1} \approx \kappa_0 + \kappa_1(p_{t+1} - d_{t+1}) + (d_{t+1} - d_t) - (p_t - d_t)$$

see notes for derivations.

(a) To first assess this approximation compare the return implied by this formula with the actual returns you downloaded. Plot the two series. Report root mean square error (RMSE) of the approximation.

(b) Now use the above equation and the VAR estimates above to infer the VAR coefficients for predicting the one period head return by the dividend yield. Does the dividend yield predict the one period ahead return? Do all three variables predict the one period ahead return?

6. Using the class notes report the multi-period predictability coefficients and R^2. Report the implied slopes for predicting cumulative dividend growth $\Delta d_{t,t+\tau} \equiv d_{t+\tau} - d_t = \Delta d_{t+1} + \Delta d_{t+2} + ... + \Delta d_{t+\tau}$ and for the multiperiod return $r_{t,t+\tau} = r_{t+1} + r_{t+2} + ... + r_{t+\tau}$ by the dividend yield. How do the VAR implied and actual (direct) regression coefficients compare?

(a) Specifically, report $\beta_{r,\tau}$ for predicting $r_{t,t+\tau}$ by $d_t - p_t$ and the corresponding $R^2_{r,\tau}$, and $\beta_{\Delta d,\tau}$ for predicting $\Delta d_{t,t+\tau}$ on $d_t - p_t$ and the corresponding $R^2_{\Delta d,\tau}$. Compute these for horizons of $\tau = 4, 12, 20$ (that is 1,3 and 5 years). Compute the standard errors of these quantities.

(b) Can you infer how much of the variation of the price dividend ratio is due to variation in cashflows and how much is due to variation in future returns, at each of these horizons?

7. Plot the response of this system to dividend growth $\epsilon_{1,t}$ and dividend yield $\epsilon_{2,t}$ shocks. Plot the response of the *level* of dividends, not their growth rate. Include responses of returns and prices to your shocks. Include responses of returns and price level to your two shocks.

8. Interpret the following

(a) Looking now at the shape of the responses, can we still label one shock an "expected return" shock and the other a "cashflow" shock?

(b) Let $\epsilon_{r,t}$ be the shock to the return forecasted by $d_t - p_t$, which you can again approximate by $\epsilon_{r,t} = \kappa\epsilon_{1,t} - \epsilon_{2,t}$. How correlated are the shocks $\epsilon_{1,t}$, $\epsilon_{r,t}$, and $\epsilon_{2,t}$? Does the fact that the $\epsilon_{1,t}$ and $\epsilon_{2,t}$ correlation is not exactly zero matter to the impulse-response interpretation?

(c) How would you produce shocks that are really orthogonal, and so that one shock is an expected return shock and the other shock is a cashflow shock. First, answer the question for $a_2 = 0$. Second (harder) can you do it for $a_2 \neq 0$?

(d) Suppose you started with a VAR composed of return and dividend yield. What would responses to return shocks look like? Run this, although you can infer it. How would you form "expected return" and "cashflow" shocks of this VAR?

(e) Bootstrap the first equation of this restricted VAR - one year log returns on lagged D/P, using the estimated VAR as the null model. Specifically, simulate the VAR a large number (e.g. $N = 1000$) times, by drawing the residuals from their empirical distribution (with replacement). Estimate the VAR every time. Plot the distribution of the estimated coefficient. Is the OLS estimate biased? Are OLS standard errors misleading? [Bonus: do this under the null of no predictability and dividend yield being a random walk.]

9. Compute and plot the terms of the approximate price-dividend identity

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

The plot lets you decompose "how much of the p-d on date t was due to a change in expected dividend growth, and how much was due to a change in expected returns"?

Bonus Use earnings growth (can get them off Robert Shiller website or any other source you'd like) either as an additional regressor (or instead of dividend growth) in the VAR. Again you can run the return regression or imply it based on the identity. Compare the coefficients and R^2 for forecasting returns and dividend growth. Given the present value identity, how could it be that other regressors both help predict dividend growth rate and returns. If dividend growth is more predictable, doesn't that mean returns should be less predictable?

10. Compute and plot the the terms of the approximate return decomposition:

$$\begin{aligned} r_t &= E_{t-1}(r_t) + \epsilon_t^r \\ \epsilon_t^r &= (E_t - E_{t-1}) \left(\Delta d_t + \sum_{j=1}^{\infty} \kappa^{j-1} \Delta d_{t+j} - \sum_{j=1}^{\infty} \kappa^{j-1} r_{t+j} \right) \end{aligned}$$

Do the plots differ from the p-d decomposition and if so why?

11. Welch and Goyal "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction" (RFS, 2007) challenge the presence of return predictability. They argue that there is a look-ahead bias and if one looks at out of sample evidence the support for return predictability is weak. Repeat part (a) above but using rolling VAR. That is use 20 quarters to estimate the VAR, then predict the return next period. Then update the data by one observation and run the VAR again and predict the subsequent return. Continue until the end of the sample. Then provide the RMSE between the predicted and realized return series and implied R^2.

Data Convention and Sources

Unless stated otherwise:

- All variables are as of the end of the time period. E.g. 2011 stock price refers to the price level on December 31, 2011; the level of consumption in Q1 of 2011 measures the consumption from January up to March of 2011.

- Most variables are in logs. E.g., log consumption growth; log market return; log price-dividend ratio; bond yield (negative log price), etc.

Macro Variables

- Real consumption growth. Annual (from 1929) and quarterly (from 1947) per capita chained data are available at the Bureau of Economic Analysis, NIPA Table 7.1. Typically, real consumption level that we use in the models is equal to the real expenditure on non-durable goods and services, i.e. sum of rows 16 and 17. Convert into log growth rate. You can also find monthly data on personal consumption expenditure (see Table 2.8.5 for nominal data from 1959). The data is oftern considered to be of lower quality due to potential measurement errors. Annual data going far back in time can be found, e.g., in Robert Barro datasets.

- Inflation Rate. There are multiple sources for price level and inflation data. CRSP at WRDS provides CPI index, though it is not seasonally adjusted. FRED (Federal Reserve Bank of St. Louis) dataset contains seasonally-adjusted CPI data at monthly frequency, but post war. I am not aware of good monthly or quarterly, seasonally adjusted, pre war inflation data, so a few times I had to remove seasonality myself on the web you can find X11 filters from Census; EVIEWS also has one build in.

  Sometimes it might be appropriate to use GDP or consumption deflators as a price index. This corresponds to the ratio between nominal and chained GDP (consumption) from the BEA tables.

Financial Data

- Aggregate Stock Market: Returns, Dividends, Price-Dividend Ratio. A standard source is CRSP at WRDS. Its typical to go for monthly index (the broadest one) data and download value-weighted return including dividends, $vwridt + 1 = (P_{t+1} + D_{t+1})/P_t - 1$, and excluding dividends, $vwredt + 1 = (P_{t+1})/P_t - 1$. From here you can obtain the series of prices and dividends. Initialize $P_0 = 1$, and recursively update $P_t = P_{t-1}(1 + vwredt)$, $D_t = P_{t-1}(vwridt - vwredt)$. For lower-frequency data (e.g., annual), sum up dividends levels from January till December over the corresponding year. Your annual (arithmetic) return is equal to price as of December 31 plus the annual dividend, divided by the price level as of December of previous year. From here you can obtain log dividend growth rate and price-dividend ratio. Take logs of return and dividend growth and price-dividend ratio to express them in continuously compounded form. Note that dividend data has seasonalities it can be tricky to work with it at frequencies higher than annual. For the same reason, price-dividend ratio is typically defined as a current price over the sum of past 12 months of dividends, even on quarterly or monthly frequency. Also note that the return and dividend data is nominal. Youd need to subtract inflation rate (all in logs) to obtain real quantities. This data starts from 1926 and goes till the end of the previous year or so; see quarterly and annual updates for more recent data. You can find earlier data on prices, earnings and dividends (from 1870s) on Robert Shiller website. The daily return data is also available on CRSP.

- Interest Rates. Long data sample (from 1920) of interest rates is available at Global Financial Database: ticker ITUSA3D for 90-day Tbill (short rate), and IGUSA10D for 10-year yield (long rate), close prices. You can also find a risk- free rate data on Fama-French website. FRED also has a short-term interest rate data from 30s or so. Note that all these rates are nominal. CRSP has more recent data on prices and yields of zero-coupon bonds; follow a link to Fama-Bliss Discount Bond Prices (or Yields). This data is for 1 to 5 year to maturity bond yields and it starts from 1952. Another source of nominal discount bond yield data is based on the work of Refet S. Gurkaynak, Brian Sack, and Jonathan H. Wright The U.S. Treasury Yield Curve: 1961 to the Present, and it is available on Fed website http://www.federalreserve.gov/econresdata/researchdata.htm. The advantage of this data is that it contains longer-term yields (e.g., 10-30 year), though, the sample is more recent (e.g., 10 year bond yield data starts in 1971).