Empirical Asset Pricing:
Problem Set 1

March 18, 2024

1. Time-series predictability of returns and dividend growth.

   (a) Download monthly returns with and without dividends from CRSP in WRDS for the period 1945.01-2023.12. You will need two series: the values weighted index of returns on stocks in CRSP universe including dividends (VWRETD) and excluding dividends (VWRETX).

   (b) Compute monthly dividends.

   (c) Aggregate dividends within a year (from January to December) by investing them into cash (i.e., at the risk-free rate) or investing them into the aggregate stock market.

   (d) Construct non-overlapping annual returns, annual dividend growth, and the log price-dividend ratio for cash-invested and for market-invested dividends.

   Compute returns and dividend growth in geometric terms.

   Report the mean and volatility of dividend growth from both reinvestment methods. Explain the difference.

   (e) Continue with cash-invested dividends.

   Predict returns (in logs) and dividend growth (in logs) using the lagged log price-dividend ratio:

$$
\begin{aligned}
r_{t+1} &= a_r + b_r pd_t + \epsilon^r_{t+1}, &(1)\\
\Delta d_{t+1} &= a_d + b_d pd_t + \epsilon^d_{t+1}. &(2)
\end{aligned}
$$

   Report the coefficients and the R-squared values. Repeat this exercise for the sub-samples from 1945-1990 and 1990-2021. Try to explain the differences.

(f) Start from the Campbell-Shiller identity and estimate the equation

$$pd_{t+1} = a_{pd} + \phi pd_t + \epsilon^{pd}_{t+1}.$$

We can obtain a variance decomposition of the log price-dividend ratio via:

$$Var(pd_t) = Cov\left(\sum_{s=1}^{\infty} \rho^{s-1} E_t\left[\Delta d_{t+s}\right], pd_t\right) + Cov\left(-\sum_{s=0}^{\infty} \rho^{s-1} E_t\left[r_{t+s}\right], pd_t\right).$$

Divide both sides by $Var(pd_t)$ so that we can estimate how much of the variation in the log price-dividend ratio is due to discount rate news and cash flow news.

Comment on the economic interpretation of the results.

(g) The present-value identity implies restrictions between b_r, b_d, and ϕ. Derive the connection between the coefficients.

(h) Predict cumulative returns (in logs) and cumulative dividend growth (in logs) using the lagged log price-dividend ratio:

$$\sum_{j=1}^{n} r_{t+j} = a_{r,n} + b_{r,n} pd_t + \epsilon^{r}_{t+1}, \tag{3}$$

$$\sum_{j=1}^{n} \Delta d_{t+j} = a_{d,n} + b_{d,n} pd_t + \epsilon^{d}_{t+1}. \tag{4}$$

for $n = \{1, \cdots, 5\}$. Report the coefficients and the R-squared values. Is there more predictability at longer horizons?

(i) Assume that the log pd ratio still follows an AR(1) as above. Derive the connection between the coefficients $b_{r,n}$, $b_{d,n}$, and ϕ.

2. The Mankiw-Shapiro and Stambaugh bias.

   (a) Start from the predictive system that you estimated in the previous question for the sample from 1945-2023.

$$\begin{aligned} r_{t+1} &= a_r + b_r pd_t + \epsilon^r_{t+1}, \\ pd_{t+1} &= a_{pd} + \phi pd_t + \epsilon^{pd}_{t+1}. \end{aligned}$$

   We want to simulate from this model. Assume that the errors are normally distributed. Estimate and report the coefficients, including the covariance matrix of the shocks.

   (b) Simulate 10,000 samples of 69 observations and estimate the model for each of the samples. Importantly, fix the seed for the subsequent exercises at the beginning of each of the 10,000 runs. Compute the average estimate of b_r and ϕ and compare them to the population values.

   (c) Now repeat the exercise but setting the covariance

$$Cov(\epsilon^r_{t+1}, \epsilon^{pd}_{t+1}) = 0.$$

   Comment on the difference.

   (d) Set the covariance back to the sample estimate, but now plot the bias in b_r and ϕ as a function of ϕ where we vary $\phi = \lambda\widehat{\phi}$ for $\lambda = 0, 0.1, \ldots, 0.9, 1$. $\widehat{\phi}$ denotes the OLS estimate obtained in part 2.a. Comment on the results.