7 Predictive Regression and Why We Use Newey-West

7.1 How Testing the Efficient Markets Hypothesis Led to an Amazing Discovery

1. Statement of Efficient Markets
   (a) Current price of asset reflects all currently available public information.
   (b) Implies you shouldn't be able to systematically predict future returns and earn abnormal profits based on publically available information.
   (c) When there is news, it gets incorporated into asset prices immediately.
2. Let I_t be information set at t. Let y_t be the one-period holding return on a stock or portfolio.
$$E(y_{t+1}|I_t) = 0$$
3. Testable implications. Let x_t be a vector of publically available information. Then β in the regression should be zero.
$$y_{t+1} = \alpha + \beta' x_t + \epsilon_{t+1}$$
   ϵ_{t+1} is uncorrelated (orthogonal) to I_t
4. What to use for x_t? How about past returns? Yes. But even better is the dividend yield.
5. Fama, for many years (in the 70s and 80s, the biggest proponent of efficient markets hypothesis.
6. Then Fama himself uncovered violations of the hypothesis, with multiperiod returns.
7. What about multiperiod returns? If it's true for one-period ahead, it must be true for multiperiod returns.

7.2 Overlapping observations and serial correlation in error

Let $y_{t,t+k}$ be the holding return from t to $t + k$, P_t be the stock (portfolio) price and

$$x_t = \frac{d_t}{P_t}$$

be the dividend yield. The efficient markets idea is the current price reflects all relevant publically available information, so the future return should not be predictable with such information.

1. Run the regression (ignoring the constant)

$$y_{t,t+1} = \beta x_t + \epsilon_{t,t+1}$$

and do a t-test for $\hat{\beta} = 0$.

2. We might want to see how today's dividend yield predicts the 2-period holding return. In this case,

$$y_{t,t+2} = bx_t + \epsilon_{t,t+2}$$

3. Why stop at $k = 2$? Illustrate with monthly Shiller data in **lh_dy.wf1**.

   - Run long-horizon regressions at various horizons
   - Plot x_{t-k} against $y_{t-k,t}$ for various k

4. This is a famous regression in finance

$$y_{t,t+k} = \alpha + \beta_k x_t + \epsilon_{t,t+k} \tag{22}$$

where $y_{t,t+k}$ is the $k-$ period ahead future return on the stock market, x_t is the current period dividend yield on the market.

   - Fama and French, Campbell and Shiller ran these regressions long ago. There are two points
   - First, that returns are predictable
   - Second, because the dividend yield moves around, the risk premium (on the market) varys over time.

5. **Econometric issue.** The t-test for $\hat{b} = 0$ at horizons beyond 1 need adjusting. Suppose the dividend yield follows an AR(1) process (ignoring the constant)

$$x_t = \rho x_{t-1} + u_t$$

Substituting these into the 2-period return regression gives

$$
\begin{aligned}
y_{t,t+2} &= \underbrace{\beta x_t + \epsilon_{t,t+1}}_{r_{t+1}} + \underbrace{\beta x_{t+1} + \epsilon_{t+1,t+2}}_{r_{t+2}} \\
&= \beta x_t + \epsilon_{t,t+1} + \beta\Big(\underbrace{\rho x_t + u_{t+1}}_{x_{t+1}}\Big) + \epsilon_{t+1,t+2} \\
&= \beta\,(1+\rho)\,x_t + \underbrace{\epsilon_{t,t+1} + \epsilon_{t+1,t+2} + \beta u_{t+1}}_{\epsilon_{t,t+2}} \\
\epsilon_{t,t+2} &= \epsilon_{t,t+1} + \epsilon_{t+1,t+2} + \beta u_{t+1}
\end{aligned}
$$

If we're going to use all the observations, the next observation will be $y_{t+1,t+3}$ and it's error term is

$$\epsilon_{t+1,t+3} = \epsilon_{t+1,t+2} + \epsilon_{t+2,t+3} + \beta u_{t+2}$$

$\epsilon_{t+1,t+2}$ is common to both error terms $\epsilon_{t,t+2}$ and $\epsilon_{t+1,t+3}$, so they are serially correlated. This violates the *iid* assumption on the error term when we compute the standard error. The error term has an MA structure with 1 for all the coefficients. Hence, we can't (or wouldn't want to) estimate an MA model for the error term.

We also could reduce the frequency to make the observations not overlap. But then we'd be throwing away data. We don't like to do that.

6. The solution is to adjust the formula for computing the standard error by accounting for serial correlation in the error term. This is called the **Newey-West** standard error. Newey and West also allows for the variance of the error term to vary over time, which is something that we see in the data (say between growth and recession periods).

7. **Newey-West**.

$$
\begin{aligned}
\hat{\beta} - \beta &= \frac{1}{\sum x_t^2} \sum x_t \epsilon_t \\
\operatorname{var}\left(\hat{\beta}\right) &= \left(\frac{1}{\sum x_t^2}\right)^2 \operatorname{var}\left(\sum x_t \epsilon_t\right) \\
\operatorname{var}\left(\sum x_t \epsilon_t\right) &= \operatorname{var}\left(x_1\epsilon_1 + x_2\epsilon_2 + x_3\epsilon_3 + \cdots + x_T\epsilon_t\right)
\end{aligned}
$$

We treat the $x's$ as exogenous constants. When the ϵ are i.i.d., then the variance of sums is sum of variances,

$$\begin{aligned}\text{var}\left(\sum x_t\epsilon_t\right) &= x_1^2\sigma_\epsilon^2 + x_2^2\sigma_\epsilon^2 + x_3^2\sigma_\epsilon^2 + \cdots + x_T^2\sigma_\epsilon^2 \\ &= \sigma_\epsilon^2 \sum x_t^2\end{aligned}$$

Hence,

$$\text{var}\left(\hat{\beta}\right) = \left(\frac{1}{\sum x_t^2}\right)^2 \sigma_\epsilon^2 \sum x_t^2 = \frac{\sigma_\epsilon^2}{\sum x_t^2}$$

But here, the ϵ are **dependent** and there are covariances to account for. To **fix ideas**, assume $T = 4$.

$$\begin{aligned}\text{var}\left(\sum x_t\epsilon_t\right) &= \text{E}\left(x_1\epsilon_1 + x_2\epsilon_2 + x_3\epsilon_3 + x_4\epsilon_4\right)\left(x_1\epsilon_1 + x_2\epsilon_2 + x_3\epsilon_3 + x_4\epsilon_4\right) \\ &= \text{E}\{\left(x_1^2\epsilon_1^2 + x_1x_2\epsilon_1\epsilon_2 + x_1x_3\epsilon_1\epsilon_3 + x_1x_4\epsilon_1\epsilon_4\right) \\ &\quad + \left(x_2x_1\epsilon_2\epsilon_1 + x_2^2\epsilon_2^2 + x_2x_3\epsilon_2\epsilon_3 + x_2x_4\epsilon_2\epsilon_4\right) \\ &\quad + \left(x_3x_1\epsilon_3\epsilon_1 + x_3x_2\epsilon_3\epsilon_2 + x_3^2\epsilon_3^2 + x_3x_4\epsilon_3\epsilon_4\right) \\ &\quad + \left(x_4x_1\epsilon_4\epsilon_1 + x_4x_2\epsilon_4\epsilon_2 + x_4x_3\epsilon_4\epsilon_3 + x_4^2\epsilon_4^2\right)\} \\ &= \text{E}\left(\sum_{t=1}^{4} x_t^2\epsilon_t^2\right) + 2\text{E}\left(\sum_{t=j+1}^{4} \epsilon_t\epsilon_{t-j}\left(x_tx_{t-j}\right)\right) \\ &= \sum_{t=1}^{4} x_t^2 E(\epsilon_t^2) + 2\sum_{t=j+1}^{4}(x_tx_{t-j})E\left(\epsilon_t\epsilon_{t-j}\right)\end{aligned}$$

So in the general case, replace 4 with T,

$$\text{var}\left(\sum x_t\epsilon_t\right) = \sum_{t=1}^{T} x_t^2 E\left(\epsilon_t^2\right) + 2\sum_{t=j+1}^{T}(x_tx_{t-j})E\left(\epsilon_t\epsilon_{t-j}\right)$$

But here's thing. We may not know exactly the extent of the dependence. The error term could be serially correlated at higher orders. Also, here, we assumed that the error terms are homoskedastic (constant variance over time). Maybe that's not true. Newey-West is a general formula to compute the standard error for arbitrary serial dependence and heteroskedasticity. Here's the formula for the simple regression. Usually we set $p = T^{1/4}$.

$$\begin{aligned}
\text{var}\left(\hat{\beta}\right) &= \left(\frac{1}{\sum x_t^2}\right)^2 S_T \\
S_T &= S_0 + \frac{1}{T}\sum_{j=1}^{p} w(j) \sum_{t=j+1}^{T} 2\hat{\epsilon}_t\hat{\epsilon}_{t-j}\left(x_t x_{t-j}\right) \\
S_0 &= \sum \hat{\epsilon}_t^2 x_t^2 \\
w(j) &= 1 - \frac{j}{p+1}
\end{aligned}$$

8. Don't worry. There is an option in Eviews to do this computation and automatically get Newey-West t-ratios. But here's the rule: **In time-series regression always do Newey-West.**

   Go back to lh_dy.wf1

7.3 Dividend yield as predictor of future return

Optional: We may not do this section

P is the stock price (not log price), d is the dividend (not log). $\beta = \frac{1}{1+\rho}$ is the subjective discount factor, ρ is the discount rate. Present value model,

$$P_t = E_t \sum_{j=0}^{\infty} \beta^j d_{t+j}$$

Let's say we expect dividends to grow at rate δ each period, so that

$$\begin{aligned}
E_t d_{t+1} &= (1+\delta)\, d_t \\
E_t d_{t+j} &= (1+\delta)^j\, d_t
\end{aligned}$$

and $\rho > \delta$. Then

$$P_t = E_t \sum_{j=0}^{\infty} \beta^j d_{t+j} = \sum_{j=0}^{\infty} \left(\frac{1+\delta}{1+\rho}\right)^j d_t = \left(\frac{1+\rho}{\rho-\delta}\right) d_t$$

$$E_t P_{t+1} = \left(\frac{1+\rho}{\rho-\delta}\right)(1+\delta)\, d_t$$

$$E_t P_{t+2} = \left(\frac{1+\rho}{\rho-\delta}\right)(1+\delta)^2\, d_t$$

$$E_t \frac{P_{t+1}}{P_t} = \left(\frac{1+\rho}{\rho-\delta}\right)(1+\delta)\, \frac{d_t}{P_t}$$

$$E_t \frac{P_{t+2}}{P_t} = \left(\frac{1+\rho}{\rho-\delta}\right)(1+\delta)^2\, \frac{d_t}{P_t}$$

Must be the case that

$$E_t \frac{P_{t+k}}{P_t} = \left(\frac{1+\rho}{\rho-\delta}\right)(1+\delta)^k\, \frac{d_t}{P_t}$$

Dividend yield predicts future returns. Slope gets bigger as horizon increases.

- Before running this regression, we want to know if the dividend yield is stationary, or if it has a unit root. **Let's do it! with lh_dy.wf1.**
- Note also that the predictive regression for $k > 1$ will induce serial correlation in regression error terms. Standard errors need to be estimated by Newey-West.
- The results reject the random walk hypothesis for stock prices.

7.4 Additional thoughts on time series regression in finance

Optional. We may not do this section

1. Time-varying risk premia.
   (a) **Program: LHret_DY_plot.m** Look at S&P Price and returns.
      i. Log prices look like a random walk.
      ii. Returns look random. 2 or 3 dots up followed by 2 or 3 dots down. We used to think that stock returns were random.
   (b) Finance is not about taking things for granted. It is about taking a hard look at the data and figuring out how things work. It is, fundamentally, an empirical field.
2. Can we predict returns? Run regression, view fitted value as the conditional expectation.

$$R_{t+1} = \alpha + \beta R_t + \epsilon_{t+1}$$

Nothing in this regression. ($\beta = 0$). Constant expected return. This was the old view. Let's do it now **LHret_DY_plot.m** The slope is 0.2796 and Newey-West t-ratio is 10.13.

3. New view. There are variables that forecast stock returns.

$$R^e_{t+k,t} = \alpha + \beta \left(\frac{d_t}{P_t} \right) + \epsilon_{t+k}$$

Let's see a plot of the dividend yield against the 7 year return. If you bought stocks in 1980, great! Bought stocks in 2000–not so good. **Code: LHret_DY_plot.m**.

4. The lesson is that $E_t R^e_{t+k,t}$ varies a lot. Expected excess returns are large and **variable**. Why?

   The 1970s economy was terrible. Since 2008, the economy has been terrible. In the bad state, people have low tolerance for risk.

   You might think P is high because people forecast high dividend growth. Empirical says that's wrong. Prices are high relative to dividends because people expect low future returns. It's prices that move around dividends. Dividends don't move around prices.

   Prices are high (dividend yield low) when people are willing to take the risk (expansion/boom). Prices are low (dividend yield high) when people are unwilling to bear risk (recession).