1 Advanced Empirical Asset Pricing

- Class 3: Predictive Regressions and Time Varying Expected Returns
  - Statistical Issues
    - ∗ Bias: Stambaugh (JFE, 1999); Nelson and Kim (1993)
    - ∗ Standard Errors (Hansen-Hodrick (1988); Newey West, Valkanov (JFE, 2003))
    - ∗ Out-of-Sample: Goyal Welch (RFS, 2007), Campbell and Thompson (RFS, 2007), Pettenuzzo, Timmermann, Valkanov (JFE, 2014)
  - Economics
    - ∗ Fama and French (JFE 1989)
    - ∗ Shiller (AER 1981)
    - ∗ Campbell and Shiller (1988)

2 Statistical Issues

2.1 Bias: Stambaugh (1986, JFE 1999)

- The predictive regression is

$$\begin{aligned} r_{t+1} &= \alpha + \beta x_t + u_{t+1} \\ x_{t+1} &= c + \phi x_t + v_{t+1} \\ cov\left(\varepsilon_{t+1} u_{t+1}\right) &= \sigma_{u,v} \neq 0 \end{aligned}$$

- Recall example from last lecture for a reason why $\sigma_{u,v} < 0$
- Stambaugh showed that

$$E\left(\hat{\beta} - \beta\right) = \frac{\sigma_{u,v}}{\sigma_v^2} E\left(\hat{\phi} - \phi\right)$$

- From previous lecture, we know that

$$E\left(\hat{\phi} - \phi\right) = -\frac{(1 + 3\phi)}{T} + O(T^2)$$

- Or

$$\begin{aligned} E\left(\hat{\beta} - \beta\right) &= -\frac{\sigma_{u,v}}{\sigma_v^2} \frac{(1 + 3\phi)}{T} \\ E\left(\hat{\beta} - \beta\right) &= BIAS \end{aligned}$$

- If $\sigma_{u,v} < 0$

$$BIAS > 0$$

- Result: In most cases, we will overstate the predictability in returns!
- In sum, we will overstate the predictability
  - The higher is the persistence of x_t (high ϕ)
  - The lower the volatility of shocks to x_t (low σ_v^2)
  - The higher is the correlation $\sigma_{u,v}$
  - The smaller the sample size (T)
- Stambaugh correction
  - Regress r_{t+1} on x_t to get $\hat{\beta}^{OLS}$, $\hat{u}_{t+1}$
  - Regress x_{t+1} on x_t to get $\hat{\phi}$, $\hat{v}_{t+1}$
  - Compute $\hat{\sigma}_v^2$ and $\hat{\sigma}_{u,v}$
  - Form estimate of bias

$$\widehat{BIAS} = -\frac{\hat{\sigma}_{u,v}}{\hat{\sigma}_v^2}\frac{\left(1+3\hat{\phi}\right)}{T}$$

$$\hat{\beta}^u \approx \hat{\beta}^{OLS} - \widehat{BIAS}$$

Stambaugh (1999): Results

- Data (Table 1)
- Bias Correction (Figure 1)
- Aside: Cavanaugh, Elliott, and Stock (ET, 1995)– frequentist way of deriving the same result (LTU) asymptotics

2.2 Standard Errors: Inference

- Recall the OLS estimator in our context minimizes the following objective function
$$Min_{\beta} E\left[(y_{t+1} - \beta' x_t)^2\right]$$
- We can view OLS as a particular case of GMM (or extremum) estimator, where we minimize
$$J(\beta) = \left(\frac{1}{T}\sum_{t=1}^{T} g_{t+1}(\beta)'\right) W \left(\frac{1}{T}\sum_{t=1}^{T} g_{t+1}(\beta)\right)$$
- In our case, $W = I$ and
$$g_{t+1}(\beta) = (y_{t+1} - \beta' x_t) = \epsilon_{t+1}$$
whose FOC yields the familiar OLS orthogonality condition
$$E\left[(y_{t+1} - \beta' x_t) x_t'\right] = \mathbf{0}$$
- The estimator of the OLS estimator variance depends on the assumption about the lead–lag autocorrelation structure of error terms
- If the errors are i.i.d, then
$$\begin{aligned} E[\epsilon_{t+1}|x_t, x_{t-1}, ..., \epsilon_t, \epsilon_{t-1}, ...] &= 0 \\ E[\epsilon_{t+1}^2|x_t, x_{t-1}, ..., \epsilon_t, \epsilon_{t-1}, ...] &= \sigma_{\epsilon}^2 \end{aligned}$$

and the estimator is

$$\widehat{var(\widehat{\beta})} = \widehat{\sigma}_{\epsilon}^{2} \left[\sum_{t=1}^{T} (x_t x_t') \right]^{-1} \qquad \widehat{\sigma}_{\epsilon}^{2} = \frac{1}{T} \sum_{t=1}^{T} \widehat{\epsilon}_{t+1}^{2}$$

- The population variance of the OLS estimator is

$$var(\widehat{\beta}) = \frac{1}{T} E[x_t x_t']^{-1} \left[\sum_{j=-\infty}^{\infty} E(\epsilon_{t+1} x_t x_{t-j}' \epsilon_{t+1-j}') \right] E[x_t x_t']^{-1}$$

- If errors are i.i.d with variance σ_ϵ^2, then we are back to the usual OLS formula

$$var(\widehat{\beta}) = \frac{1}{T} E[x_t x_t']^{-1} \sigma_\epsilon^2$$

- The estimator of the OLS estimator variance depends on the assumption about the lead–lag autocorrelation structure of error terms

- Indeed, if the errors are i.i.d, then

$$\begin{aligned} E[\epsilon_{t+1} | x_t, x_{t-1}, ..., \epsilon_t, \epsilon_{t-1}, ...] &= 0 \\ E[\epsilon_{t+1}^2 | x_t, x_{t-1}, ..., \epsilon_t, \epsilon_{t-1}, ...] &= \sigma_\epsilon^2 \end{aligned}$$

- If the errors are correlated, then the general formulation applies

$$var(\widehat{\beta}) = \frac{1}{T}E[x_t x_t']^{-1}\left[\sum_{j=-\infty}^{\infty} E(\epsilon_{t+1}x_t x_{t-j}'\epsilon_{t+1-j}')\right]E[x_t x_t']^{-1}$$

- Hansen–Hodrick propose the following estimator

$$\widehat{var(\widehat{\beta})} = \frac{1}{T}\left[\sum_{t=1}^{T}(x_t x_t')\right]^{-1}\Gamma\left[\sum_{t=1}^{T}(x_t x_t')\right]^{-1}$$

$$\begin{aligned}\Gamma &= \Gamma_0 + \sum_{j=1}^{k}(\Gamma_j + \Gamma_j') \\ \Gamma_0 &= \sum_{t=1}^{T}(\epsilon_{t+1}\epsilon_{t+1}') \\ \Gamma_j &= \sum_{t=1}^{T}(\widehat{\epsilon}_{t+1}x_t x_{t-j}'\widehat{\epsilon}_{t+1-j}')\end{aligned}$$

- Hansen–Hodrick VCV matrix is not guaranteed to be positive semidefinite
- The problem is that the VCV at longer lags is badly (and not consistently) estimated as we have fewer and fewer observations to use
- Newey and West propose the following estimator

which is always positive semidefinite

$$\widehat{var(\widehat{\beta})} = \frac{1}{T}\left[\sum_{t=1}^{T}(x_t x_t')\right]^{-1} \Gamma \left[\sum_{t=1}^{T}(x_t x_t')\right]^{-1}$$

$$\Gamma = \Gamma_0 + \sum_{j=1}^{q}\left(1 - \frac{j}{q+1}\right)(\Gamma_j + \Gamma_j')$$

$$\Gamma_0 = \sum_{t=1}^{T}(\epsilon_{t+1}\epsilon_{t+1}')$$

$$\Gamma_j = \sum_{t=1}^{T}(\widehat{\epsilon}_{t+1} x_t x_{t-j}' \widehat{\epsilon}_{t+1-j}')$$

- Under the null of no predictability, returns are generated by

$$r_{t+1} = \alpha + \epsilon_{t+1}$$

- Now, assume you aggregate returns in an overlapping way.
- This is the same as passing a moving average filter to a white noise series.
- Namely, you create the series

$$\begin{aligned} r_{t+1:t+2} &= 2\alpha + \epsilon_{t+1} + \epsilon_{t+2} \\ r_{t+2:t+3} &= 2\alpha + \epsilon_{t+2} + \epsilon_{t+3} \\ r_{t+3:t+4} &= 2\alpha + \epsilon_{t+3} + \epsilon_{t+4} \\ &\cdots \end{aligned}$$

- This means that under the null, long–horizon overlapping returns ARE autocorrelated even if one–period returns are truly uncorrelated
- In particular, k–horizon returns follow an MA($k-1$) under the null
- In practice, people try to account for this spurious autocorrelation by using a value of $q \geq k-1$ in the Newey–West formula
- On top of this spurious autocorrelation in residuals, the persistence of the regressor also plays a role

- Valkanov(2003) analyzes the properties of β in

$$\begin{aligned} r_{t+1:t+k} &= \alpha + \beta x_t + \epsilon_{t+1} \\ x_{t+1} &= \mu + \phi x_t + v_{t+1} \end{aligned}$$

as a function of k.

- Main point: as you increase k, the LHS series becomes more an more persistent. Hence, the variables tend be cointegrated.

- By aggregating returns across multiple horizons, we are changing the nature of the process toward I(1)

- All standard asymptotic theory breaks down, must use theory of nonstationary processes

- In this context, Valkanov shows that standard t-ratios and R^2 do not have a well behaved limiting distribution as we increase k

- Instead, the t-ratios obtained as

$$t(k) = \frac{t}{\sqrt{k}}$$

where t is the t ratio from i.i.d. errors have a nice and tractable limiting distribution, which can be tabulated.

- In fact, people run two types of long-horizon regressions

$$
\begin{aligned}
r_{t+1:t+k} &= \alpha + \beta x_t + \epsilon_{t+1} \\
r_{t+1:t+k} &= \alpha + \beta x_{t,t-k} + \epsilon_{t+1}
\end{aligned}
$$

- The properties of $\hat{\beta}$ and its t-statistic depend on which specification is used.
- Table 1
- Table 2
- Empirical Result: Evidence of predictability is not all that strong

2.3 Out of Sample Evidence: Welch and Goyal (2007)

- GW07 provide a comprehensive look at the empirical performance of return predictability.
- They look at the IS and OOS performance of the most popular predictors on the same sample period and same metrics.
- They conclude that many predictors perform poorly even IS, and largely underperform a model based on the unconditional average.
- The success of some variables seems to be uniquely driven by the Oil Shock 1973-1975 period.

- Dividend–Price ratio (d/p), Dividend Yield (d/y), Earnings–Price ratio (e/p), Dividend Payout ratio (d/e)
- Stock Variance (svar)
- Cross–sectional premium (csp), Book-to-Market ratio (b/m)
- Net Equity Expansion (ntis) and Percent Equity Issuing (equis)
- Treasury bill rate (tbl), Long Term Rate of Return (ltr), Term Spread (tms), Default Yeild Spread (dfy), Default Return Spread (dfr), Inflation (infl)
- Investment to Capital Ratio (i/k)
- Cay (cay)

- Univariate regressions, model selection (select the best combination up to time t and use it for OOS) as well as all regression (includes all variables but cay)

- OOS statistics are:

$$
\begin{aligned}
R^2 &= 1 - \frac{MSE_A}{MSE_N} \\
\overline{R}^2 &= R^2 - (1 - R^2) \times \left(\frac{T-k}{T-1}\right) \\
\Delta MSE &= \sqrt{MSE_N} - \sqrt{MSE_A} \\
MSE - F &= (T - h + 1) \times \left(\frac{\sqrt{MSE_N} - \sqrt{MSE_A}}{\sqrt{MSE_A}}\right)
\end{aligned}
$$

- where MSE_A is under the alternative (predictability) and MSE_N is under the null (no predictability)

- If $MSE_N < MSE_A$, $\frac{MSE_A}{MSE_N} > 1$, $R^2 < 0$

- If $R^2 < 0$, uncondtional mean does better than the conditional mean

- Test the null that the model does not improve the OOS performance with respect to the historical mean, or $\Delta MSE = 0$

- Bootstrapped p–value for the F-test

- Compare IS and OOS performance. If a model has poor IS performance, the OOS performance is not very interesting.
- Use OOS performance as a diagnostic statistic to check for model usefulness and stability.
- GW plot: for IS, the difference between the cumulative squared demeaned equity premium and the cumulative squared regression residual; for OOS, the cumulative squared prediction errors of the prevailing mean minus the cumulative squared prediction error of the predictive variable from the linear historical regression.
- Four desirable features of a model:
  - both significant IS and reasonably good OOS performance
  - generally, an upward drift of the line
  - an upward drift not concentrated in one short period (e.g. Oil Shock)
  - an upward drift that remains positive during the recent, say, 30 years

- Results

  Figure 1: Expanding $\Delta MSE = \sqrt{MSE_N} - \sqrt{MSE_A}$

  Table 1: Sample Results (Annual frequency)

  $*$ Main Point: Difference between in sample and out of sample results

- Cochrane (2008): The Dog that Did not Bark
- Campbell and Thompson (2007)

  Campbell and Thompson (RFS, 2007) GT07 challenge GW07 results by including economic restrictions in the predictive regressions.

  Use economic theory and models to impose constraints on the sign and magnitude of the predictive coefficients.

  For example, positive risk aversion implies that the predicted risk premium cannot be negative.

  Imposing these restrictions improves the real time OOS performance of many predictors, and implies that investors could have profited by using market timing strategies.

- Start from a complete set of potential predictors.
- Look at the in-sample performance.
- Look at the OOS performance without any restriction (as in GW07).
- Look at the restrictions (prior) economic theory places on some of these variables.
- Apply these restrictions to OOS regressions.

∗ The OOS R^2 is defined as

$$R^2_{OOS} = 1 - \frac{\sum_{t=1}^{T}(r_t - \widehat{r}_t)^2}{\sum_{t=1}^{T}(r_t - \overline{r}_t)^2}$$

where $\widehat{r}_t$ and $\overline{r}_t$ use just information up to time $t-1$.

∗ When R^2_{OOS} is greater than zero, then the predictive regression has lower average mean-square prediction error than the historical mean.

∗ If returns are truly unpredictable, the expected value of this statistic is negative.

∗ The historical mean is calculated using all data from 1871.

Reasoning: Regressions which use a limited amount of data may yield implausible forecasts as they are highly influenced by a few movements and their coefficients are subject to severe sampling error.

In practice, an investor would use prior knowledge when using the output of a regression such as:

∗ Set the coefficient to zero whenever it has the "unexpected" sign

∗ Set the forecasted equity premium to zero whenever it is negative

Imposing these two simple restrictions enhances the OOS performance of the predictors.

Results: Table 1

Fama and French (JFE, 1989)

Two research questions

- Do the expected returns on bonds and stocks move together? In particular, do the same variables forecast bond and stock returns?
- Is the variation in expected bond returns related to business conditions? Are the relations consistent with intuition, theory, and existing evidence on the exposure of different assets to changes in business conditions?

Their answer to both questions is YES!

Data

Testing portfolios (r_{t+1})

* VW and EW CRSP Indexes, useful to check the impact of size
* Portfolio on 100 bonds
* Portfolios of Aaa, Aa, A, Baa, LG rated bonds

Conditioning variables (x_t)

* D/P: sum of past monthly dividends during year [t - 1;t] divided by price
* TERM: difference between the yield on the Aaa bond portfolio and the one-month bill rate
* DEF: difference between the bond portfolio and the Aaa

Properties of Conditioning Variables

Autocorrelations of D/P, DEF and more rapidly for TERM decay to zero as lag increases => stationary but persistent variables.

TERM shows little correlation versus D/P and DEF (■ 0.2), while D/P and DEF are strongly correlated (■ 0.7)

* Interpretation: D/P, DEF, and TERM are all measures of business cycle conditions but TERM tracks short-term movements, at business cycle frequency as measured by NBER
* D/P and DEF show some swings at business cycle frequency but mostly variations that go beyond that – they capture lower frequency, long-term shocks that span several measured business cycles, periods of persistently poor or good times
* All these variables are high in bad times and low in good times (counter-cyclical risk premia)
* For example, the TERM is low around peaks and high near troughs

Fama and French (1989), Table 1

Main Predictive Results (Table 2)

Bivariate regressions: Regress r_{t+1} on $(D/P, TERM)$, $(DEF, TERM)$, and $(D/P, DEF)$.

Summary of Main Results

* The same predictive variables explain risk premia in bond and stocks.
* The variables are countercyclical, thus risk premia are high when times are poor and low when times are good
* However, these variables have different persistence (DEF and D/P versus TERM) and capture different frequencies of changing economic conditions.
* This suggests that there is a mix of components in risk premia that relate to long- and short- term aspects of business conditions As in FF88, R2s are increasing with the holding period return
* DEF and D/P slopes are higher for stocks versus bonds
* The loadings on TERM don't show much cross-sectional variation =>TERM captures a component of risk premium that is common across all long-term securities.

Bigger Picture

- ∗ The same variables capture variation in risk premia across different assets is a somewhat comforting result from an economic viewpoint.
- ∗ The explanatory power is similar across assets.
- ∗ Evidence of business–cycle fluctuations in risk premia

Shiller (1981)

Excess volatility

Idea: Compare the properties of returns and fundamentals

Why: Movement in returns must come from fluctuations in fundamentals (cash flows)–assuming a constant discount rate of 4.8%

In the plot, for $\gamma = 0.954$, he constructs

$$p_t^* = \Sigma_{k=1}^{\infty} \gamma^{k+1} d_{t+1+k}$$

Argument: Fundamentals are too smooth when compared to the volatility of realized returns

Campbell and Shiller (RFS, 1988)–log-linearizing returns

- Idea: If dividend volatility is not sufficient to account for price volatility, we must look for framework that allows us to disentangle discount rates news from cash ﬂow news
- Start from the definition of log return:

$$r_{t+1} = \log\left(\frac{P_{t+1} + D_{t+1}}{P_t}\right) = \log(P_{t+1} + D_{t+1}) - \log(P_t)$$

- Idea: can we write a linear relationship between returns, log prices, and log dividends?
- Problem: the above equation is non-linear, due to Jensen's inequality. It involves the log of the sum.
- Solution: get an *approximate* linear relationship via first-order Taylor series expansion. Replace the (weighted average) of the logs.

Use a first order Taylor expansion of returns to get

$$r_{t+1} \cong k + \rho p_{t+1} + (1-\rho) d_{t+1} - p_t$$

where we defined

$$\rho \equiv \frac{1}{1 + e^{\overline{dp}}} \qquad k \equiv -\log(\rho) - (1-\rho)\log(1/\rho - 1)$$

The return is written as a weighted average of of the log price and log dividend, with weights that sum up to 1.

These coefficients depend on the average dp ratio. The average dp ratio from above implies a value for $\rho = 0.965$.

Thus, the log price gets a much larger weight than the log dividend.

How accurate is the approximation?

In general, it holds OK as long as the average dp ratio is constant and the dp does not wander too far from its average – see Lettau Van Nieuwerburgh (RFS, 2008).
∗ General: monthly data

On annual data, however, the precision deteriorates as the DP experiences sharper fluctuations and is less persistent.

From Returns to Prices

We can view the above expression as a recursive relationship in log prices:

$$p_t = k + \rho p_{t+1} + (1-\rho)d_{t+1} - r_{t+1}$$

Iterate this expression forward for J periods to get

$$p_t = (k+\rho k+\rho^2 k+...+\rho^J k)+\sum_{j=0}^{J} \rho^j[(1-\rho)d_{t+j}-r_{t+j}]+\rho^J p_{t+J}$$

No rational bubble condition:

$$\lim_{j\to\infty} \rho^j p_{t+j} = 0$$

Note: if $\rho < 1$, as is in our case,

$$\sum_{j=0}^{\infty} \rho^j = \frac{1}{1-\rho}$$

Therefore, the current price can be written as

$$p_t = \frac{k}{1-\rho} + \sum_{j=0}^{\infty} \rho^j[(1-\rho)d_{t+1+j} - r_{t+1+j}]$$

Here, we are not using any economic model. It is a *dynamic accounting identity*. A higher price today *must* reflect either future higher dividends, lower returns, or a combination of the two.

If we take expectations on both sides based on time-t information to get

$$p_t = \frac{k}{1-\rho} + E_t\left[\sum_{j=0}^{\infty}\rho^j[(1-\rho)d_{t+1+j} - r_{t+1+j}]\right]$$

A higher price today *must* reflect either *expectations* of future higher dividends, lower returns, or a combination of the two.

- We can re-write the previous identity as an expression for the (log) dp ratio:

$$d_t - p_t = \frac{k}{1-\rho} + E_t\left[\sum_{j=0}^{\infty} \rho^j [r_{t+1+j} - \Delta d_{t+1+j}]\right]$$

- This tells us that:
  - A higher dp ratio today *must* reflect either expectations of higher returns, lower dividend growth, or a combination of the two.
  - The fact that the dp ratio moves over time means that one or both are time-varying.
  - If returns and dividend growth are stationary, so is the dp ratio. Nice framework to work empirically.
  - The log dp ratio should be a linear predictor of of future dividend growth rates and/or discount rates
- Dynamic generalization to the Gordon growth model when expected dividend growth and expected returns are allowed to change over time.

The effect of changes in discount rates and expected dividend growth now depends on the time series properties of $E_t[d_{t+1+j}]$ and $E_t[r_{t+1+j}]$.

Two alternative empirical approaches are followed to proceed further:

- use regression-type tools, univariate (just focus on returns) and multivariate (VAR)
- make additional statistical assumptions on the time-series behavior of expected returns and expected dividend-growth in the context of a fully-fledged economic model, and then calibrate it/estimate it on the observed data