1 Advanced Empirical Asset Pricing

- Class 4: Predictive Regressions–Recent Topics
  - Structural Breaks–Lettau Van Nieuwerburgh (RFS, 2008)
  - Spurious Predictive Regressions–Ferson, Sarkissian, Simin (JF, 2003)
  - Economic Significance of Predictability (Campbell and Thompson, 2008)
  - Noisy Predictive Regressions–Pastor and Stambaugh (JF, 2010)

1.1 Structural Breaks

- Recall from last lecture that we can use log-linearization to write the (log) dp ratio as

$$d_t - p_t = \overline{dp} + E_t \left[\sum_{j=0}^{\infty} \rho^j \left[(r_{t+1+j} - \overline{r}) - \left(\Delta d_{t+1+j} - \overline{\Delta d} \right) \right] \right]$$

- This relationship implies 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.
- The log-linearization was taken around the "stead-state" log dp ratio, $\overline{d-p}$
  - Long-term average
- What if the long-term average changes over time
  - Structural break
  - Break is exogenous–because dp is taken as exogenous!

Questions: What if there is a permanent change in the average dp ratio?

- Where does it enter in our previous expressions? What will it happen to our regressions?
- The Goyal and Welch (2008) results might be interpreted as being driven by structural instability in predictive regressions
- What if we estimate the predictive regressions with a rolling (30-year) window
  LVN (2008), Figure 1
- The framework LVN (2008) have in mind is:

$$d_t - p_t = \overline{dp_t} + E_t \left[\sum_{j=0}^{\infty} \rho^j \left[(r_{t+1+j} - \overline{r}_t) - \left(\Delta d_{t+1+j} - \overline{\Delta d}_t \right) \right] \right]$$

- Hence, rather than using $d_t - p_t$ to forecast returns, the authors suggest the de-trended log dp ratio:

$$(d_t - p_t) - \overline{dp_t}$$

- There is a distinction between extremely persistent but stationary and permanent structural change.

- LVN (2008), Table 1
  - More evidence of predictability if sample is split into two or three subsamples
  - Chow test of structural break is significant
- LVN (2008), Figure 1
  - Visualizing the possible structural breaks in the dp ratio
- When we consider breaks, we have to think about
  - Is there a break at point t? Chow (F) test
  - How many breaks?
    - * How to test for number of breaks?
    - * (0 versus 1), (1 versus 2) or (0 versus 2)
  - Location of breaks
  - sup F test (Bai and Perron, EMA 1998)
    - * Joint test of location and number of breaks
    - * Have to simulate the distribution
  - Note location of 1 vs 2 breaks in the testing

- LVN (2008), Table 2 (Cont'd)
  - Strong evidence for 1 break
  - Evidence for two breaks
  - Thus the difference in means from Figure 1 are statistically significant
- How to proceed?
  - Create de-meaned series
  - Use this predictor instead of the original series
  - Panel B: The persistence of the de-meaned series is lower than that of the original series (alleviates Stambaugh (1999) bias)

- Instability in other valuation ratios (Table 4)
  - ep ratio
  - btm ratio
  - de ratio–payout ratio
  - dp - excluding NASDAQ firms
  - dp -adjusted for shares repurchases
- Idea: dp = ep + de: which one moved?
  - All three have structural breaks

- Out of Sample Results
- Have to estimate at t the likelihood of being in a new regime next period
- sup
- Hamilton (1989) regime switching model
  Specify the number of regimes
- Results–Figure 5
- All these results are at monthly horizons

Long-Horizon Regressions

Start from the predictability equations

$$\begin{aligned} r_{t+1} &= \kappa_r dp_t + \tau^r_{t+1} \\ \Delta d_{t+1} &= \kappa_d dp_t + \tau^d_{t+1} \end{aligned}$$

where all variables are demeaned and stationary.

The CS88 log-linearization implies that

$$r_{t+1} = dp_t - \rho dp_{t+1} + \Delta d_{t+1}$$

The previous equations **imply** that

$$\begin{aligned} dp_{t+1} &= \phi dp_t + \tau^{dp}_{t+1} \\ \tau^{dp}_{t+1} &= (\tau^d_{t+1} - \tau^r_{t+1})/\rho \\ 1 - \rho\phi &= \kappa_r - \kappa_d \end{aligned}$$

Equation says that the dp ratio follows an AR(1) model with root ϕ.

Equation says that the shock in the dp ratio is function of the shocks to dividend growth (positive) and to returns (negative).

Equation is a present-value constraint. It says that the persistence of the dp ratio is intimately related to the predictive coefficients. Moreover, if $\rho < 1$ as is our case and $|\phi| < 1$ to ensure stationarity of the dp, it MUST be the case that $\kappa_r - \kappa_d > 0$.

Thus, either returns must be forecastable ($\kappa_r \neq 0$), or dividend growth must be forecastable ($\kappa_d \neq 0$), or a combination of the two.

Cochrane (RFS, 2008) wrote a very good piece on this key concept: we cannot impose jointly the null that $\kappa_d = \kappa_r = 0$!!! We will get back to this...

Moreover, starting from these equations we can derive the following annualized H-period dividend growth and return forecasting equations:

$$\begin{aligned}\frac{1}{H}\sum_{j=1}^{H} r_{t+j} &= \kappa_r(H)dp_t + \tau^r_{t+1,t+H}\\ \frac{1}{H}\sum_{j=1}^{H} \Delta d_{t+j} &= \kappa_d(H)dp_t + \tau^d_{t+1,t+H}\end{aligned}$$

which are intimately linked to the one-period (one-year) regressions by the following relationship:

$$\begin{aligned}\kappa_r(H) &= \kappa_r\frac{1}{H}\left(\frac{1-\phi^H}{1-\phi}\right)\\ \kappa_d(H) &= \kappa_d\frac{1}{H}\left(\frac{1-\phi^H}{1-\phi}\right)\end{aligned}$$

This implies that we can use the long-horizon coefficients to improve our knowledge of the short-horizon coefficients by means of a single GMM estimation.

LVN use horizons of $H = \{1, 3, 5, 7, 10\}$.

Basically, they use in the GMM the full system of eight equations (4 horizons times 2 dependent variables) in two unknowns (κ_r, κ_d).

Notice that we are not using the AR process for the dp ratio in the system. Why so??

$$g(\kappa_r, \kappa_d) = \begin{pmatrix} \frac{1}{T-1}\sum_{t=1}^{T}(r_{t+1} - k_r dp_t)dp_t \\ \frac{1}{T-3}\sum_{t=1}^{T}(r_{t+1:t+3} - k_r(3)dp_t)dp_t \\ ... \\ \frac{1}{T-10}\sum_{t=1}^{T}(r_{t+1:t+10} - k_r(10)dp_t)dp_t \\ \frac{1}{T-1}\sum_{t=1}^{T}(\Delta d_{t+1} - k_d dp_t)dp_t \\ \frac{1}{T-3}\sum_{t=1}^{T}(\Delta d_{t+1:t+3} - k_d(3)dp_t)dp_t \\ ... \\ \frac{1}{T-10}\sum_{t=1}^{T}(\Delta d_{t+1:t+10} - k_d(10)dp_t)dp_t \end{pmatrix}$$

where we impose

$$\kappa_r(H) = \kappa_r \frac{1}{H}\left(\frac{1-\phi^H}{1-\phi}\right)$$

$$\kappa_d(H) = \kappa_d \frac{1}{H}\left(\frac{1-\phi^H}{1-\phi}\right)$$

and

$$1 - \rho\phi = \kappa_r - \kappa_d$$

We then find the GMM estimates $(\widehat{\kappa}_r, \widehat{\kappa}_d)$ by minimizing the objective function

$$J(\kappa_r, \kappa_d) = g'Wg$$

where W is the Identity (first-step) and then the optimal (second-step) weighting matrix.

These estimates are fully consistent in that they take into account the long-short and cross-equations restrictions.

Table 7: Results

Final Results:

Returns forecastability

Lack of dividend growth forecastability

Watch for stationarity of the regressor!

Link across forecasting equations and horizons.

1.2 Spurious Predictive Regressions-Ferson et al. (JF, 2003)

- Recall lecture 1–spurious predictive regressions
- Motivating Example: Table 1
- The return is generated by

$$r_{t+1} = \mu + Z_t^* + e_{t+1}$$

But we run the regression

$$r_{t+1} = \mu + \beta Z_t + e_{t+1}$$

where Z_t and Z_t^* are uncorrelated and are both persistent!

- This is exactly the spurious regression setup (Granger and Newbold (1974), Phillips (1986))
- Two issues:
  - Spuriousness (Table 2)
  - Data Mining (Table 3)
  - The critical values are much larger than the usual, normal t-stats
- If we revisit the results in Table 1, the evidence in favor of predictability is much weaker

- Interesting setup of the predictor being persistent
- Let

$$\begin{aligned} r_{t+1} &= \alpha + \beta x_t + u_{t+1} \\ x_{t+1} &= \phi x_t + \varepsilon_t \\ \phi &= 1 + \frac{c}{T} \end{aligned}$$

- If we look at the variance of the predictor,

$$var\left(x_{t+1}\right) = \frac{\sigma_\varepsilon^2}{1-\phi^2}$$

  or the more persistent is the predictor, the larger its variance must be
- We don't observe this in practice!
- Most predictors have vols that are orders of magnitude lower than the vol of returns
- Need to keep vol low as the process becomes persistent

- How about the following (Rubia, Moon, Valkanov (2009))

$$\begin{aligned} x_{t+1} &= \phi x_t + \varepsilon_{t+1} \\ \phi &= 1 + \frac{c}{T} \\ \varepsilon_t &= \sigma_u \sqrt{1-\rho} e_t \end{aligned}$$

  where $var(e_t) = 1$, or

$$\sigma_\varepsilon^2 = \sigma_u^2 \frac{c}{T}$$

- The variance of the predictor is local to zero.

1.3 Are Low R^2 Economically Relevant?

- CT07 provide a useful example which allows us to measure the economic significance of the R^2. Namely, why should a, say, monthly R^2 of 3% be relevant to begin with?
- Consider excess returns on a risky asset are time-varying and predicted by x_t

$$r_{t+1} = \mu + x_t + \epsilon_{t+1}$$

  where x_t and ϵ_{t+1} have mean-zero and constant variance σ_x^2 and σ_ϵ^2, respectively.
- Consider now a mean-variance investor with RRA equal to γ. She then solves

$$\max_\alpha \alpha E[r_p] - \frac{\gamma}{2}\alpha^2 V(r_p) \quad r_p = \alpha(r + r_f) + (1 - \alpha) r_f$$

- The question now is: how would she improve the Sharpe Ratio of her portfolio by having access to the conditioning variable x_t?

- If the investor doesn't know x_t, her optimal allocation is

$$\alpha_t^* = \alpha^* = \frac{1}{\gamma}\left(\frac{\mu}{\sigma_\epsilon^2 + \sigma_x^2}\right)$$

so it is constant.

- The corresponding average (expected) excess return she achieves is then

$$E[r_p^* - r_f] = \alpha^* E[r] = \frac{1}{\gamma}\left(\frac{\mu^2}{\sigma_\epsilon^2 + \sigma_x^2}\right) = \frac{S^2}{\gamma}$$

where $S \equiv E[r_p]/std(r_p)$ is the Sharpe Ratio.

- If the investor does know x_t, her optimal allocation is
$$\alpha_t^* = \frac{1}{\gamma}\left(\frac{\mu + x_t}{\sigma_\epsilon^2}\right)$$
which varies over time.

- The corresponding average (expected) excess return she achieves is then
$$E[r_p^* - r_f] = \alpha_t^* E[r] = \frac{1}{\gamma}\left(\frac{\mu^2 + \sigma_x^2}{\sigma_\epsilon^2}\right) = \frac{1}{\gamma}\left(\frac{S^2 + R^2}{1 - R^2}\right)$$
where $R^2 \equiv \sigma_x^2/(\sigma_x^2 + \sigma_\epsilon^2)$ is the predictable component in excess returns.

- The absolute difference in expected returns equals
$$\frac{1}{\gamma}\left(\frac{S^2+R^2}{1-R^2}\right)(1+S^2)>0$$
which implies a percentage gain with respect to the unconditional case of
$$\left(\frac{R^2}{1-R^2}\right)\left(\frac{1+S^2}{S^2}\right)>\frac{R^2}{S^2}$$
- The R^2 must be compared to the S^2 to gauge its economic magnitude.
- The advantage stems from the reduction in variance knowing x_t.
- With a monthly Sharpe Ratio of 0.108, a monthly OOS R^2 of 0.43% implies a gain of about 36%, or about 5% per year if $\gamma=1$.
- CT07 also measure the welfare gain of an investor with mean-variance utility function who has access to the information of x_t.
- Even with realistic short-selling constraints, this amounts to 1-2% per year for most valuation ratios.
- Would need to take into account transactions costs from trading.

1.4 Noisy Predictive Regressions–Pastor and Stambaugh (JF, 2010)

- We have a ssumed that we have a perfect instrument of expected returns, dp ratio, etc.
- Suppose we have

$$\begin{aligned} r_{t+1} &= \mu_t + u_{t+1} \\ x_{t+1} &= \theta + Ax_t + v_{t+1} \\ \mu_{t+1} &= \alpha + \beta\mu_t + w_{t+1} \end{aligned}$$

- The shocks $[u_t, v_t, w_t]$ are potentially correlated
- The conditional mean is unobservable but can be inferred from observable information
- The correlation $\rho_{u,w} < 0$
- The authors use Kalman filtering to show that

$$E_t\left(r_{t+1}|D_t\right) = c + \sum_{j=0}^{\infty} \lambda_j \left(r_{t-j} - E\left(r_{t-j}|D_{t-j-1}\right)\right) + \phi_s v_{t-j}$$

- To understand this result, we need to know Kalman Filtering
- Kalman filtering in finance: Conrad and Kaul (1988), Johannes, Polson, and Stroud (2002), Ang and Piazzesi (2003), Brandt and Kang (2004), Dangl and Halling (2006), Duffee (2006), and Rytchkov (2006).

- Digression: Kalman (1963) paper

1.5 Setup (Hamilton CH 13)

$$\begin{aligned} y_t &= A'x_t + H'z_t + w_t \\ z_t &= Fz_{t-1} + v_t \end{aligned}$$

where

- y_t is the observable variable (think "returns")

  The first equation, the y_t equation is called the "space" or the "observation" equation.

- z_t is the unobservable variable (think "condtional mean" or "state of the economy")

  The second equation, the z_t equation is called the "state" equation.

- x_t is a vector of exogenous (or predetermined) variables (we can set $x_t = 0$ for now).

- v_t and w_t are iid and assumed to be uncorrelated at all lags

$$E\left(w_t v_t'\right) = 0$$

- Also $E\left(v_t v_t'\right) = Q$, $E\left(w_t w_t'\right) = R$

- The system of equations is known as a state-space representation.

- Any time series can be written in a state-space representation.

- In standard engineering problems, it is assumed that we know the parameters A, H, F, Q, R.
- The problem is to give impulses x_t such that, given the states z_t, the missile is guided as closely to target as possible.
- In finance, we want to estimate the unknown parameters A, H, F, Q, R in order to understand where the system is going, given the states z_t. There is little attempt at guiding the system. In fact, we usually assume that $x_t = 1$ and $A = E(Y_t)$, or even that $x_t = 0$.

- Note: Any time series can be written as a state space.
- Example: AR(2): $Y_{t+1} - \mu = \phi_1 (Y_t - \mu) + \phi_2 (Y_{t-1} - \mu) + \varepsilon_{t+1}$
- State equation:

$$\begin{bmatrix} Y_{t+1} - \mu \\ Y_t - \mu \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} Y_t - \mu \\ Y_{t-1} - \mu \end{bmatrix} + \begin{bmatrix} \varepsilon_{t+1} \\ 0 \end{bmatrix}$$

- Observation equation:

$$y_t = \mu + \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} Y_{t+1} - \mu \\ Y_t - \mu \end{bmatrix}$$

- There are other state-space representations of Y_t. Can you write down another one?

- As a first step, we will assume that A, H, F, Q, R are known.
- Our goal would be to find a best linear forecast of the state (unobserved) vector z_t. Such a forecast is needed in control problems (to take decisions) and in finance (state of the economy, forecasts of unobserved volatility).
- The forecasts will be denoted by:
$$z_{t+1|t} = E\left(z_{t+1}|y_t..., x_t....\right)$$
and we assume that we are only taking linear projections of z_{t+1} on $y_t..., x_t.....$ Nonlinear Kalman Filters exist but the results are a bit more complicated.
- The Kalman Filter calculates the forecasts $z_{t+1|t}$ recursively, starting with $z_{1|0}$, then $z_{2|1}$, ...until $z_{T|T-1}$.
- Since $z_{t|t-1}$ is a forecast, we can ask how good of a forecast it is?
- Therefore, we define $P_{t|t-1} = E\left(\left(z_t - z_{t|t-1}\right)\left(z_t - z_{t|t-1}\right)\right)$, which is the forecasting error from the recursive forecast $z_{t|t-1}$.

- The Kalman Filter can be broken down into 5 steps

  Initialization of the recursion. We need $z_{1|0}$. Usually, we take $z_{1|0}$ to be the unconditional mean, or $z_{1|0} = E\left(z_1\right)$. (Q: how can we estimate $E\left(z_1\right)$?) The associated error with this forecast is $P_{1|0} = E\left(\left(z_{1|0} - z_1\right)\left(z_{1|0} - z_1\right)\right)$

Forecasting y_t (intermediate step)

The ultimate goal is to calculate $z_{t|t-1}$, but we do that recursively. We will first need to forecast the value of y_t, based on available information:

$$E\left(y_t | x_t, z_t\right) = A' x_t + H' z_t$$

From the law of iterated expectations,

$$E_{t-1}\left(E_t\left(y_t\right)\right) = E_{t-1}\left(y_t\right) = A' x_t + H' z_{t|t-1}$$

The error from this forecast is

$$y_t - y_{t|t-1} = H'\left(z_t - z_{t|t-1}\right) + w_t$$

with MSE

$$\begin{aligned} & E\left(y_t - y_{t|t-1}\right)\left(y_t - y_{t|t-1}\right)' \\ = \; & E\left[H'\left(z_t - z_{t|t-1}\right)\left(z_t - z_{t|t-1}\right)' H\right] + E\left[w_t w_t'\right] \\ = \; & H' P_{t|t-1} H + R \end{aligned}$$

Updating Step ($z_{t|t}$)

Once we observe y_t, we can update our forecast of z_t, denoting it by $z_{t|t}$, before making the new forecast, $z_{t+1|t}$.

We do this by calculating $E\left(z_t|y_t, x_t, ...\right) = z_{t|t}$

$$\begin{aligned} z_{t|t} &= z_{t|t-1} + E\left(\left(z_t - z_{t|t-1}\right)\left(y_t - y_{t|t-1}\right)\right) * \\ &\quad \left(E\left(y_t - y_{t|t-1}\right)\left(y_t - y_{t|t-1}\right)'\right)^{-1}\left(y_t - y_{t|t-1}\right) \end{aligned}$$

We can write this a bit more intuitively as:

$$z_{t|t} = z_{t|t-1} + \beta\left(y_t - y_{t|t-1}\right)$$

where β is the OLS coefficient from regressing $\left(z_t - z_{t|t-1}\right)$ on $\left(y_t - y_{t|t-1}\right)$.

The bigger is the relationship between the two forecasting errors, the bigger the correction must be.

It can be shown that

$$z_{t|t} = z_{t|t-1} + P_{t|t-1} H \left(H' P_{t|t-1} H + R \right)^{-1} \left(y_t - A' x_t - H' z_{t|t-1} \right)$$

This updated forecast uses the old forecast $z_{t|t-1}$,
and the just observed values of y_t and x_t.

Forecast $z_{t+1|t}$.

Once we have an update of the old forecast, we can produce a new forecast, the forecast of $z_{t+1|t}$

$$\begin{aligned} E_t\left(z_{t+1}\right) &= E\left(z_{t+1}|y_t, x_t, ...\right) \\ &= E\left(Fz_t + v_{t+1}|y_t, x_t, ...\right) \\ &= FE\left(z_t|y_t, x_t, ...\right) + 0 \\ &= Fz_{t|t} \end{aligned}$$

We can use the above equation to write

$$\begin{aligned} E_t\left(z_{t+1}\right) &= F\{z_{t|t-1} \\ &\quad +P_{t|t-1}H\left(H'P_{t|t-1}H + R\right)^{-1}\left(y_t - A'x_t - H'z_{t|t-1}\right) \\ &= Fz_{t|t-1} \\ &\quad +FP_{t|t-1}H\left(H'P_{t|t-1}H + R\right)^{-1}\left(y_t - A'x_t - H'z_{t|t-\right. \end{aligned}$$

We can also derive an equation for the error in forecast as a recursion

$$\begin{aligned} P_{t+1|t} &= F[P_{t|t} \\ &\quad -P_{t|t-1}H\left(H'P_{t|t-1}H + R\right)^{-1}H'P_{t|t-1}]F' \\ &\quad +Q \end{aligned}$$

Go to step 2, until we reach T. Then, we are done.

- Summary: The Kalman Filter produces

  The optimal forecasts of $z_{t+1|t}$ and $y_{t+1|t}$ (optimal within the class of linear forecasts)

  We need some initialization assumptions

  We need to know the parameters of the system, i.e. A, H, F, Q, R.

- Now, we need to find a way to estimate the parameters A, H, F, Q, R.
- By far, the most popular method is MLE.
- Aside: Simulations Methods–getting away from the restrictive assumptions of ε_t