Review Session 1: Regression Review, Return Predictability

Jiahui Shui

November 8, 2024

1 Finite Sample

1.1 Important Properties of Conditional Expectation

Consider $\mathbb{E}[X|Y]$.

- If X and Y are independent, then $\mathbb{E}[X|Y] = \mathbb{E}X$
- If X is $\sigma(Y)$-measurable, then $\mathbb{E}[X|Y] = X$. In particular, $\mathbb{E}[f(Y)|Y] = f(Y)$
- $\mathbb{E}[Xf(Y)|Y] = f(Y)\mathbb{E}[X|Y]$ (pulling out known factors)
- $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$. Law of Iterated Expectation

1.2 Assumptions

A multiple regression model is:

$$y_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_K x_{i,K} + \varepsilon_i, \quad i = 1, \cdot, n \tag{1}$$

- Vector Form: $y_i = x_i'\beta + \varepsilon_i, \quad i = 1, \cdots, n$
- Matrix Form: $y = X\beta + \varepsilon$

In finite sample theory, we have 4 assumptions:

- (**H1**) Linear Model: $y = X\beta + \varepsilon$
- (**H2**) Strict Exogeneity: $\mathbb{E}[\varepsilon_i|X] = 0$
  - $\mathbb{E}[y|X] = X\beta$
  - $\mathbb{E}[\varepsilon_i] = 0$
  - $\mathbb{E}[\varepsilon_i x_{j,k}] = 0, \forall j, k$
  - $\text{Cov}(\varepsilon_i, x_{j,k}) = 0, \forall j, k$

- (**H3**) No Multicollinearity: $\text{rank}(X) = K$
- (**H4**) Spherical disturbance: $\text{Var}(\varepsilon|X) = \sigma^2 I_n$
  - Conditional Homoskedasticity: $\mathbb{E}[\varepsilon_i^2|X] = \sigma^2, \forall i$
  - No correlation: $\mathbb{E}[\varepsilon_i \varepsilon_j|X] = 0, \forall i \neq j$

1.3 OLS Estimation

We want to find $\hat{\beta}$ that minimize sum of squared residuals

$$\text{SSR}(\hat{\beta}) = \sum_{i=1}^{n}(y_i - x_i'\hat{\beta})^2 = (y - X\hat{\beta})'(y - X\hat{\beta}) \tag{2}$$

The first order condition gives us

$$\hat{\beta}^{OLS} = (X'X)^{-1}X'y \tag{3}$$

Also, it can be written as

$$\hat{\beta}^{OLS} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} x_i y_i\right) = S_{xx}^{-1} S_{xy}$$

1.4 Properties

Define OLS residual

$$e = y - \hat{y} = y - X\hat{\beta} \tag{4}$$

Proposition 1. *The residual is orthogonal to explanatory variable, i.e. $X'e = 0$. When the explanatory variables include a constant term, $\sum_{i=1}^{n} e_i = 0$.*

Define

$$P = X(X'X)^{-1}X', \quad M = I_n - P \tag{5}$$

then

- $PX = X$
- $Pe = 0$
- $MX = 0$
- P and M are both symmetric and idempotent
- $\hat{y} = X\hat{\beta}^{OLS} = X(X'X)^{-1}X'y = Py$
- $e = y - \hat{y} = y - Py = My = M\varepsilon$

Theorem 1. *Under assumptions H1-H4, we have*

1. *$\hat{\beta}^{OLS}$ is unbiased, i.e.* $\mathbb{E}[\hat{\beta}^{OLS}|X] = \beta$
2. *Efficiency:* $\text{Var}(\hat{\beta}^{OLS}|X) = (X'X)^{-1}\sigma^2$. *For any linear unbiased estimator $\hat{\beta}$,* $\text{Var}(\hat{\beta}|X) \geq \text{Var}(\hat{\beta}^{OLS}|X)$
3. $s^2 = e'e/(n-K)$ *is an unbiased estimator for* σ^2.

2 Large Sample Theory/Asymptotic Theory

2.1 Assumptions

- (A1): Linear model. $y = X\beta + \varepsilon$
- (A2): Ergodic stationarity: $\{y_t, x_t\}$ is stationary and ergodic. (If two processes are far enough, say x_k and x_{t+k} as $t \to \infty$, then they can be thought as "independent")
- (A3): (Predetermined regressors): $\mathbb{E}[x_{t,i}\varepsilon_t] = 0$, $\forall i, t$. Define $g_t = x_t\varepsilon_t$, then $\mathbb{E}[g_t] = 0$.
- (A4): $\mathbb{E}[x_t x_t']$ has full rank.
- (A5): $\mathbb{E}[g_t g_t'] < \infty$ and g_t is a martingale difference sequence. Also, $\mathbb{E}[g_t g_t']$ has full rank.

2.2 Properties

Under A1-A4 $\Rightarrow \text{plim}_{T\to\infty} \hat{\beta} = \beta$.

Additionally, if we have A5, then

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \text{Avar}(\hat{\beta})) \tag{6}$$

where

$$\text{Avar}(\hat{\beta}) = \Sigma_{xx}^{-1} S \Sigma_{xx}^{-1}, \quad \Sigma_{xx} = \mathbb{E}[x_t x_t'], \quad S = \mathbb{E}[g_t g_t'] \tag{7}$$

If x_t is a scalar, then

$$\Sigma_{xx} = \sigma_x^2, \quad S = \mathbb{E}[x_t^2 \varepsilon_t^2], \quad \text{Avar}(\hat{\beta}) = \frac{\mathbb{E}[x_t^2 \varepsilon_t^2]}{\sigma_x^4}$$

Remark: In lecture 3, Professor Valkanov wrote "intuitively" $\text{Avar}(\hat{\beta}) = \sigma_\varepsilon^2/\sigma_x^2$. This requires additional assumption that $\mathbb{E}[x_t^2\varepsilon_t^2] = \mathbb{E}[x_t^2]\mathbb{E}[\varepsilon_t^2]$

3 Campbell and Shiller (1988)

The relationship starts from log stock returns

$$r_{t+1} = \log\left(\frac{P_{t+1} + D_{t+1}}{P_t}\right) = \Delta p_{t+1} + \log\left(1 + \frac{D_{t+1}}{P_{t+1}}\right) \tag{8}$$

where $\Delta p_{t+1} = \log(P_{t+1}/P_t)$. Let $dp_t = \log(D_t/P_t)$. Then we have approximation

$$\log\left(1 + \frac{D_{t+1}}{P_{t+1}}\right) \approx \kappa + (1-\rho)dp_{t+1} \tag{9}$$

This can be down from the following Taylor expansion:

$$\log(1+e^x) \approx \log(1+e^{x_0}) + \frac{e^{x_0}}{1+e^{x_0}}(x - x_0)$$

We apply this to LHS of equation (9) at $\overline{dp}$, then we know that

$$\kappa = \log(1+e^{\overline{dp}}) - \frac{e^{\overline{dp}}}{1+e^{\overline{dp}}}\overline{dp}, \quad \rho = \frac{1}{1+e^{\overline{dp}}}$$

Then we have

$$\begin{aligned} r_{t+1} &= \kappa - \rho dp_{t+1} + \Delta d_{t+1} + dp_t \\ dp_t &= -\kappa + r_{t+1} - \Delta d_{t+1} + \rho dp_{t+1} \end{aligned} \tag{10}$$

Iterate this forward we can get

$$dp_t = -\frac{\kappa}{1-\rho} + \sum_{j=0}^{\infty} \rho^j r_{t+1+j} - \sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} \tag{11}$$

The present-value relationship holds ex-post as well as exante:

$$dp_t = -\frac{\kappa}{1-\rho} + \mathbb{E}_t\left[\sum_{j=0}^{\infty} \rho^j r_{t+1+j}\right] - \mathbb{E}_t\left[\sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j}\right] \tag{12}$$

Hence, movements in prices can be attributed to fluctuations in expected growth rates, expected returns, or both.

Shiller (1981) provides the first evidence that prices appear to move more than what is implied by expected dividends, even realized dividends. This is the celebrated excess volatility puzzle.

Empirical Findings:

- Evidence of return predictability in the post-war sample period, but weaker before the second world war.
- Dividend growth is predictable by the price-dividend ratio before the second world war, not thereafter. Potential explanation: changes in dividend smoothing.
- Return predictability tends to be stronger at longer horizons

4 Econometric issues in return predictability

A large econometric literature is concerned with correct inference as many variables, including the price-dividend ratio, are highly persistent:

Consider the following system

$$\begin{aligned} y_{t+1} &= \alpha + \beta x_t + \varepsilon_{t+1} \\ x_{t+1} &= \mu_x + \phi x_t + u_{t+1} \end{aligned} \tag{13}$$

In this system, x_t is highly persistent ($\phi \approx 1$), $\beta > 0$. Now we suppose that $\mathbb{E}[x_{t-1}\varepsilon_t] = 0$ by construction. $\hat{\phi}$ tends to be downward biased. This is standard issue in OLS.

$$\mathbb{E}[\hat{\phi}] = \phi - \frac{1+3\phi}{T} + O(1/T^2) \tag{14}$$

And $\hat{\beta}$ is upward biased, which means that we reject the null of no predictability too often.

$$\mathbb{E}[\hat{\beta} - \beta] = \frac{\sigma_{\varepsilon u}}{\sigma_u^2}\mathbb{E}[\hat{\phi} - \phi] = -\frac{\sigma_{\varepsilon u}}{\sigma_u^2}\frac{1+3\phi}{T} \tag{15}$$

Note that here the covariance $\sigma_{\varepsilon u}$ between between the shocks of the predictor and the shocks to returns is large and negative, around -0.9. The reason is that a positive dp shock usually has no news about dividends, so means a negative p shock and a negative r shock.