Math Review Notes—Linear Regression

Gregory Faletto

Contents

Last updated November 21, 2018

1 Linear Regression

These notes are based on my notes from *Time Series and Panel Data Econometrics* (1st edition) by M. Hashem Pesaran and coursework for Economics 613: Economic and Financial Time Series I at USC. I also borrowed from some other sources which I mention when I use them.

1.1 Chapter 1: Linear Regression

1.1.1 Preliminaries

Suppose the true model is $y_i = \alpha + \beta x_i + \epsilon_i$. Classical assumptions:

(i) $\mathbb{E}(\epsilon_i) = 0$

(ii) $\text{Var}(\epsilon_i \mid x_i = \sigma^2$ (constant)

(iii) $\text{Cov}(\epsilon_i, \epsilon_j) = 0$ if $i \neq j$

(iv) ϵ_i is uncorrelated to x_i, or $\mathbb{E}(\epsilon_i \mid x_j) = 0$ for all i, j.

1.1.2 Estimation

$$\hat{\beta} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} = \frac{\sum_{i=1}^n x_i y_i - n \overline{xy}}{\sum_{i=1}^n x_i^2 - n\overline{x}^2}$$

$$\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}$$

or

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} = \frac{S_{XY}}{S_{XX}}$$

or

$$\hat{\beta} = r \frac{S_{YY}}{S_{XX}}$$

where r is the correlation coefficient.

Let

$$w_i = \frac{x_i - \overline{x}}{\sum_{i=1}^n (x_i - \overline{x})^2}$$

so that

$$\hat{\beta} = \sum_{i=1}^n w_i(y_i - \overline{y}) = \sum_{i=1}^n w_i y_i - \overline{y}\frac{\sum_{i=1}^n x_i - \overline{x}}{\sum_{i=1}^n (x_i - \overline{x})^2} = \sum_{i=1}^n w_i y_i$$

since $\sum_{i=1}^n x_i - \overline{x} = 0$. Then a simple expression for $\mathrm{Var}(\hat{\beta})$ is

$$\mathrm{Var}(\hat{\beta}) = \sum_{i=1}^n w_i^2 \mathrm{Var}(y_i \mid x_i) = \sum_{i=1}^n w_i^2 \mathrm{Var}(\epsilon \mid x_i) = \sigma^2 \sum_{i=1}^n w_i^2 = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \overline{x})^2} = \frac{\sigma^2}{S_{XX}}$$

We can estimate these quantities as follows:

$$\hat{\sigma}^2 = \frac{1}{n-2} \cdot \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta} x_i)^2$$

Note that

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{t=1}^T (y_t - \hat{\alpha} - \hat{\beta}x_t)^2 = \frac{1}{n-2}\sum_{t=1}^T \left[(y_t - (\overline{y} - \hat{\beta}\overline{x}) - \hat{\beta}x_t)^2 \right] = \frac{1}{n-2}\sum_{t=1}^T (y_t - \overline{y} - \hat{\beta}(x_t - \overline{x}))^2$$

$$= \frac{1}{n-2}\sum_{t=1}^T (y_t - \overline{y})^2 - 2\hat{\beta}(x_t - \overline{x})(y_t - \overline{y}) + \hat{\beta}^2 (x_t - \overline{x})^2$$

In the case where there is no intercept, we have

$$\hat{\sigma}^2 = \frac{1}{T-1}\sum_{t=1}^T (y_t - \hat{\beta}x_t)^2 = \frac{1}{T-1}\sum_{t=1}^T \left(y_t^2 - 2r\frac{S_{YY}}{S_{XX}} x_t y_t + r^2 \frac{S_{YY}^2}{S_{XX}^2} x_t^2 \right)$$

Also,

$$\widehat{\mathrm{Var}}(\hat{\beta}) = \frac{\hat{\sigma}^2}{S_{XX}} = \frac{1}{n-2} \cdot \frac{\sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2}{\sum_{i=1}^n (x_i - \overline{x})^2}$$

Correlation coefficient:

$$r^2 = \frac{\left(\sum_{t=1}^T x_t y_t\right)^2}{\sum_{t=1}^T x_t^2 \sum_{t=1}^T y_t^2}$$

$$r = \frac{1}{T-1}\frac{S_{XY}}{\sqrt{S_{XX} S_{YY}}}$$

Remark. The formulas for the coefficients in univariate OLS can also be derived by considering (x, y) as a bivariate normal distribution and calculating the conditional expectation of y given x. (See Theorem (**??**).)

1.2 Chapter 2: Multiple Regression

General OLS:

$$\hat{\beta} = (\boldsymbol{X}'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + u) = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u = \beta + (X'X)^{-1}X'u$$

$$\mathrm{Var}(\hat{\beta}) = \mathrm{Var}(\beta + (X'X)^{-1}X'u) = \mathrm{Var}(\beta) + \mathrm{Var}((X'X)^{-1}X'u) = 0 + \mathbb{E}[(X'X)^{-1}X'uu'X(X'X)^{-1}]$$

$$= \mathbb{E}\big[(X'X)^{-1}X'\mathbb{E}(uu' \mid X)X(X'X)^{-1}\big] = \sigma^2\mathbb{E}\big[(X'X)^{-1}X'I_TX(X'X)^{-1}\big] = \sigma^2\mathbb{E}\big[(X'X)^{-1}\big]$$

$$= \sigma^2(X'X)^{-1}$$

$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}}{T-k}$$

1.3 Chapter 3: Hypothesis testing in regression

In this section, I borrow from C. Flinn's notes "Asymptotic Results for the Linear Regression Model," available online at http://www.econ.nyu.edu/user/flinnc/notes1.pdf.

Lemma 1.

$$\frac{1}{n} \cdot X'\epsilon \xrightarrow{p} 0$$

Proof. Note that $\mathbb{E}\frac{1}{n} \cdot X'\epsilon = 0$ for any n. Then we have

$$\mathrm{Var}\left(\frac{1}{n} \cdot X'\epsilon\right) = \mathbb{E}\left(\frac{1}{n} \cdot X'\epsilon\right)^2 = n^{-2}\mathbb{E}(X'\epsilon\epsilon'X) = n^{-2}\mathbb{E}(\epsilon\epsilon')X'X = \frac{\sigma^2}{n}\frac{X'X}{n}$$

implying that $\lim_{n\to\infty} \mathrm{Var}\left(\frac{1}{n} \cdot X'\epsilon\right) = 0$. Therefore the result follows from Chebyshev's Inequality (Theorem **??**). □

Lemma 2. If ϵ is i.i.d. with $E(\epsilon_i) = 0$ and $\mathbb{E}(\epsilon_i^2) = \sigma^2$ for all i, the elements of the matrix X are uniformly bounded so that $|X_{ij}| < U$ for all i and j and for U finite, and $\lim_{n\to\infty} X'X/n = Q$ is finite and nonsingular, then

$$\frac{1}{\sqrt{n}}X'\epsilon \xrightarrow{d} \mathcal{N}(0, \sigma^2 Q)$$

Proof. If we have one regressor, then $n^{-1/2}\sum_{i=1}^n X_i\epsilon_i$ is a scalar. Let G_i be the cdf of $X_i\epsilon_i$. Let

$$S_n^2 = \sum_{i=1}^n \text{Var}(X_i\epsilon_i) = \sigma^2\sum_{i=1}^n X_i^2$$

In this scalar case, $Q = \lim_{n\to\infty} n^{-1}\sum_i X_i^2$. By the Lindberg-Feller Theorem, a necessary and sufficient condition for $Z_n \to \mathcal{N}(0\sigma^2 Q)$ is

$$\lim_{n\to\infty}\frac{1}{S_n^2}\sum_{i=1}^n\int_{|\omega|>\nu S_n}\omega^2 dG_i(\omega) = 0$$

for all $\nu > 0$. Now $G_i(\omega) = F(\omega/|X_i|)$. Then rewrite the above equation as

$$\lim_{n\to\infty}\frac{n}{S_n^2}\sum_{i=1}^n\frac{X_i^2}{n}\int_{|\omega/X_i|>\nu S_n/|X_i|}\left(\frac{\omega}{X_i}\right)^2 dF(\omega/|X_i|) = 0$$

Since $\lim_{n\to\infty} S_n^2 = \lim_{n\to\infty} n\sigma^2\sum_{i=1}^n X_i^2/n = n\sigma^2 Q$, we have $\lim_{n\to\infty} n/S_n^2 = (\sigma^2 Q)^{-1}$, which is a finite and nonzero scalar. Then we need to show

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n X_i^2\delta_{i,n} = 0$$

where

$$\delta_{i,n} = \int_{|\omega/X_i|>\nu S_n/|X_i|}\left(\frac{\omega}{X_i}\right)^2 dF(\omega/|X_i|)$$

But $\lim_{n\to\infty}\delta_{i,n} = 0$ for all i and any fixed ν since $|X_i|$ is bounded while $\lim_{n\to\infty} X_n = \infty$, so the measure of the set $\{|\omega/X_i| > \nu S_n/|X_i|\}$ goes to 0 asymptotically. Since $\lim_{n\to\infty} n^{-1}\sum_i X_i^2$ is finite and $\lim_{n\to\infty}\delta_{i,n} = 0$ for all i, $\lim_{n\to\infty} n^{-1}\sum_i X_i^2\delta_{i,n} = 0$, so $\frac{1}{n}\cdot X'\epsilon \xrightarrow{p} 0$.

□

Theorem 3. Under the conditions of Lemma 2 (ϵ is i.i.d. with $E(\epsilon_i) = 0$ and $\mathbb{E}(\epsilon_i^2) = \sigma^2$ for all i, the elements of the matrix X are uniformly bounded so that $|X_{ij}| < U$ for all i and j and for U finite, and $\lim_{n\to\infty} X'X/n = Q$ is finite and nonsingular),

$$\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} \mathcal{N}(0, \sigma^2 Q^{-1})$$

Proof.

$$\sqrt{n}(\hat{\beta}-\beta) = \left(\frac{X'X}{n}\right)^{-1}\frac{1}{\sqrt{n}}X'\epsilon$$

Since $\lim_{n\to\infty}(X'X/n)^{-1} = Q^{-1}$ and by Lemma 2

$$\frac{1}{\sqrt{n}} X' \epsilon \xrightarrow{d} \mathcal{N}(0, \sigma^2 Q)$$

then

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \sigma^2 Q^{-1} Q Q^{-1}) = \mathcal{N}(0, \sigma^2 Q^{-1})$$

□

t-test statistic:

$$t = \frac{\hat{\beta} - 0}{s.e.(\hat{\beta})}$$

F-test statistic:

$$F = \left(\frac{T - k - 1}{r} \right) \left(\frac{SSR_R - SSR_U}{SSR_U} \right)$$

Since

$$R^2 = \frac{\sum_t (y_t - \overline{y})^2 - \sum_t (y_t - \hat{y}_t)^2}{\sum_t (y_t - \overline{y})^2} = \frac{\sum_t (y_t - \overline{y})^2 - SSR_U}{\sum_t (y_t - \overline{y})^2}$$

we have

$$SSR_U = \sum_t (y_t - \overline{y})^2 - R^2 \sum_t (y_t - \overline{y})^2 = (1 - R^2) \sum_t (y_t - \overline{y})^2$$

yielding

$$F = \left(\frac{T - k - 1}{r} \right) \left(\frac{\sum_t (y_t - \overline{y})^2 - (1 - R^2) \sum_t (y_t - \overline{y})^2}{(1 - R^2) \sum_t (y_t - \overline{y})^2} \right) = \left(\frac{T - k - 1}{r} \right) \left(\frac{R^2}{1 - R^2} \right)$$

Confidence interval for sums of coefficients. (Two coefficient case.) Suppose we want to test $H_0 : \beta_1 + \beta_2 = k$. Let $\delta = \beta_1 + \beta_2 - k$, $\hat{\delta} = \hat{\beta}_1 + \hat{\beta}_2 - k$. Note that under the null hypothesis $\delta = 0$. We can construct a t-statistic

$$t_{\hat{\delta}} = \frac{\hat{\delta} - 0}{\sqrt{\hat{\text{Var}}(\hat{\delta})}} = \frac{\hat{\beta}_1 + \hat{\beta}_2 - k}{\sqrt{\hat{\text{Var}}(\hat{\delta})}}$$

where

$$\hat{\text{Var}}(\hat{\delta}) = \hat{\text{Var}}(\hat{\beta}_1) + \hat{\text{Var}}(\hat{\beta}_2) + 2\hat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_2)$$

This means that a 95% confidence interval for δ can be constructed in the following way:

$$\hat{\delta} \pm t^* \sqrt{\hat{\text{Var}}(\hat{\delta})}$$

where t^* is the 95% critical value for the t-distribution.

1.4 Chapter 4: Heteroskedasticity

Under heteroskedasticity, the OLS estimator $\hat{\beta} = (X'X)^{-1}X'y$ is unbiased, but the true covariance matrix of $\hat{\beta}$ no longer matches the OLS formula. For instance, suppose we have

$$y_t = \sum_{i=1}^{K} \beta_i x_{ti} + u_t$$

where $\text{Var}(u_t) = \sigma^2 z_t^2$.

$$\hat{\beta} = (X'X)^{-1}X'y = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u = \beta + (X'X)^{-1}X'u$$

$$\implies \mathbb{E}(\hat{\beta}) = \mathbb{E}[\beta] + (X'X)^{-1}X'\mathbb{E}[u] = \beta$$

since $\mathbb{E}(u)$ is still 0. However,

$$\text{Var}(\hat{\beta}) = \mathbb{E}\big[\big(\hat{\beta} - \mathbb{E}(\hat{\beta})\big)\big(\hat{\beta} - \mathbb{E}(\hat{\beta})\big)'\big] = \mathbb{E}\big[\big(\beta + (X'X)^{-1}X'u - \beta\big)\big(\beta + (X'X)^{-1}X'u - \beta\big)'\big]$$

$$= \mathbb{E}\big[\big((X'X)^{-1}X'u\big)\big((X'X)^{-1}X'u\big)'\big] = \mathbb{E}\big[(X'X)^{-1}X'uu'X\big((X'X)^{-1}\big)'\big]$$

$$= (X'X)^{-1}X'\mathbb{E}[uu' \mid X]X(X'X)^{-1}$$

$$= (X'X)^{-1}X' \begin{bmatrix} \sigma^2 z_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma^2 z_2^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma^2 z_3^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma^2 z_T^2 \end{bmatrix} X(X'X)^{-1}$$

$$= \sigma^2 (X'X)^{-1}X' \begin{bmatrix} z_1^2 & 0 & 0 & \dots & 0 \\ 0 & z_2^2 & 0 & \dots & 0 \\ 0 & 0 & z_3^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & z_T^2 \end{bmatrix} X(X'X)^{-1}$$

which is different from the OLS estimator of the covariance matrix $\sigma^2(X'X)^{-1}$. Therefore the estimate of the variances of $\hat{\beta}$ will be biased if the OLS formulas are used, and the usual t and F tests for $\hat{\beta}$ will be invalid.

1.5 Chapter 5: Autocorrelated disturbances

Generalized least squares model:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}$$

where

$$\mathbb{E}(\boldsymbol{u} \mid \boldsymbol{X}) = 0 \ \ \forall\, t$$

$$\mathbb{E}(\boldsymbol{uu}' \mid \boldsymbol{X}) = \boldsymbol{\Sigma}$$

where $\boldsymbol{\Sigma}$ is a positive definite matrix.

$$\hat{\beta}_{GLS} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$$

$$\text{Var}(\hat{\beta}_{GLS}) = (X'\Sigma^{-1}X)^{-1}$$