Linear Regression Models

Jin-Lung Lin

1 Simple Linear regression Models

$$\begin{aligned}
y_t &= \beta_0 + \beta_1 x_{1t} + \epsilon_t, \quad t = 1, 2, \cdots, T \\
E(\epsilon_t) &= 0, \quad \forall t \\
V(\epsilon_t) &= \sigma^2, \quad \forall t \\
E(\epsilon_t \epsilon_s) &= 0, \quad \forall t \neq s \\
E(x_{1t}\epsilon_t) &= 0
\end{aligned}$$

Take expectation of both sides to obtain

$$\mu_y = \beta_0 + \beta_1 \mu_x$$

and subtract it from the regression model,

$$y_t = \mu_y = \beta_1(x_{1t} - \mu_x) + \epsilon_t, \quad t = 1, 2, \cdots, T$$

2 Multiple Linear regression Models

$$\begin{aligned}
y_t &= \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \epsilon_t, \quad t = 1, 2, \cdots, T \\
E(\epsilon_t) &= 0, \quad \forall t \\
V(\epsilon_t) &= \sigma^2, \quad \forall t \\
E(\epsilon_t \epsilon_s) &= 0, \quad \forall t \neq s \\
E(x_{it}\epsilon_t) &= 0, \quad i = 1, 2
\end{aligned}$$

3 What does this equation means?

1. y_t is the caused (dependent) variable while x_{1t}, x_{2t} are the causal (independent) variables.

2. The relation between y_t, x_{1t}, x_{2t} is linear and no *important* causal variables are left out.

3. β_1 measures the effect of one unit of change of x_{1t} on y_t with x_{2t} being fixed. Similarly, β_2 measures the effect of one unit of change of x_{2t} on y_t with x_{1t} being fixed.

4. Note that

$$
\begin{aligned}
\beta_1 &= \frac{\partial y_t}{\partial x_{1t}} \\
\beta_2 &= \frac{\partial y_t}{\partial x_{2t}} \\
\frac{dy_t}{dx_{1t}} &= \frac{\partial y_t}{\partial x_{1t}} + \frac{\partial y_t}{\partial x_{2t}} \frac{dx_{2t}}{dx_{1t}}
\end{aligned}
$$

4 Estimating $\beta's$

Least square estimates are derived from minimizing the squared residuals

$$
\begin{aligned}
\text{Min SSE} &= \sum_{t=1}^{T} (y_t - \beta_0 - \beta_1 x_{1t} - \beta_2 x_{2t})^2 \\
\hat{\beta}_1 &= \frac{\sum_{t=1}^{T}(x_{1t}-\bar{x}_1)(y_t-\bar{y})\sum_{t=1}^{T}(x_{2t}-\bar{x}_2)^2 - \sum_{t=1}^{T}(x_{1t}-\bar{x}_1)(x_{2t}-\bar{x}_2)\sum_{t=1}^{T}(x_{2t}-\bar{x}_2)(y_t-\bar{y})}{\sum_{t=1}^{T}(x_{1t}^2-\bar{x})\sum_{t=1}^{T}(x_{2t}^2-\bar{x}_2) - (\sum_{t=1}^{T}(x_{1t}-\bar{x}_1)(x_{2t}-\bar{x}))^2} \\
\hat{\beta}_2 &= \frac{\sum_{t=1}^{T}(x_{1t}-\bar{x}_1)^2\sum_{t=1}^{T}(x_{2t}-\bar{x}_2)(y_t-\bar{y}) - \sum_{t=1}^{T}(x_{1t}-\bar{x}_1)(y_t-\bar{y})\sum_{t=1}^{T}(x_{1t}-\bar{x}_1)(x_{2t}-\bar{x}_2)}{\sum_{t=1}^{T}(x_{1t}-\bar{x}_1)^2\sum_{t=1}^{T}(x_{2t}-\bar{x}_2)^2 - (\sum_{t=1}^{T}(x_{1t}-\bar{x}_1)(x_{2t}-\bar{x}_2))^2} \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1\bar{x}_1 - \hat{\beta}_2\bar{x}_2
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\hat{y}_t &= \hat{\beta}_0 + \hat{\beta}_1 x_{1t} + \hat{\beta}_2 x_{2t} \\
\hat{y}_t - \bar{y} &= \hat{\beta}_1(x_{1t}-\bar{x}_1) + \hat{\beta}_2(x_{2t}-\bar{x}_2) \\
\hat{\epsilon}_t &= y_t - \hat{y}_t
\end{aligned}
$$

5 How to interpret least square estimates

$$
\begin{aligned}
y_t &= \hat{y}_t + \hat{\epsilon}_t \\
&= \hat{\beta}_0 + \hat{\beta}_1 x_{1t} + \hat{\beta}_2 x_{2t} + \hat{\epsilon}_t
\end{aligned}
$$

$\hat{y}_t \perp \hat{\epsilon}_t$ since

$$
\begin{aligned}
\sum_{t=1}^{T} \hat{y}_t \hat{\epsilon} &= (\hat{\beta}_0 + \hat{\beta}_1 x_{1t} + \hat{\beta}_2 x_{2t}) \hat{\epsilon}_t \\
&= \hat{\beta}_0 \sum_{t=1}^{T} (y_t - \hat{\beta}_1 x_{1t} - \hat{\beta}_2 x_{2t}) \\
&+ \hat{\beta}_1 \sum_{t=1}^{T} x_{1t} (y_t - \hat{\beta}_1 x_{1t} - \hat{\beta}_2 x_{2t}) \\
&+ \hat{\beta}_2 \sum_{t=1}^{T} x_{2t} (y_t - \hat{\beta}_1 x_{1t} - \hat{\beta}_2 x_{2t}) \\
&= 0
\end{aligned}
$$

by the normal equations. Further,

$$
\sum_{t=1}^{T} (y_t - \bar{y})^2 = \sum_{t=1}^{T} (\hat{y} - \bar{y})^2 + \sum_{t=1}^{T} (y_t - \hat{y}_t)^2
$$

Note that we are decomposition y_t into the sum of two terms $\hat{y}_t$ and $\hat{\epsilon}_t$. It is the orthogonality condition that makes the decomposition unique. To sum up, regressing y_t on $(x_{1t}, x_{2t}$ is equivalent to projecting y_t onto the space spanned by x_{1t}, x_{2t}.

6 Another look at LSE

To solve $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_0$, one needs to know $(\sum_{t=1}^{T} x_{1t}^2, \sum_{t=1}^{T} x_{2t}^2, \sum_{t=1}^{T} x_{1t} x_{2t}$ and $(\sum_{t=1}^{T} x_{1t} y_t, \sum_{t=1}^{T} x_{2t} y_t$, that is the covariance among x_{it} and covariance between y_t and x_{it}. The requirement is obviously as in computing $\partial y_t / \partial x_{1t}, \partial y_t$ one needs to disentangle the covariance among x_{it}.

7 General cases

Generally speaking,

$$
Y = X\beta + \epsilon
$$

where

$$
Y = (y_1, y_2, \cdots, y_T)', X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{T1} & x_{T2} & \vdots & x_{Tk} \end{bmatrix}
$$

$$\hat{\beta} = (X'X)^{-1}X'Y$$

where

$$X'X = \begin{bmatrix} T & \sum_{t=1}^{T} x_{t1} & \sum_{t=1}^{T} x_{t2} & \cdots & \sum_{t=1}^{T} x_{tk} \\ \sum_{t=1}^{T} x_{t1} & \sum_{t=1}^{T} x_{t1}^2 & \sum_{t=1}^{T} x_{t1}x_{t2} & \cdots & \sum_{t=1}^{T} x_{t1}x_{tk} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum_{t=1}^{T} x_{tk} & \sum_{t=1}^{T} x_{tk}x_{t1} & \sum_{t=1}^{T} x_{tk}x_{t2} & \vdots & \sum_{t=1}^{T} x_{tk}^2 \end{bmatrix}$$

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

8 Testing hypotheses on β

Typically all elements of $(X'X)$ diverge to ∞ as $T \to \infty$ and $(X'X)^{-1}$ converges to a all zeros matrix. It is often the case that $\frac{1}{T}(X'X)^{-1} \xrightarrow{p} \Sigma_x^{-1}$ as $T \to \infty$. It can be shown that LSE $\hat{\beta}$ is the MLE. Under the null hypothesis $H_0 : \beta = \beta_0$

$$\sqrt{T}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \sigma^2\Sigma_x^{-1})$$

With this asymptotic normality, we can test any hypothesis on β.

$$T(\hat{\beta} - \beta_0)'\frac{1}{\hat{\sigma}^2}(\frac{X'X}{T})^{-1}(\hat{\beta} - \beta_0) \xrightarrow{d} \chi^2(k+1)$$

Let the hypothesis be expressed as $R\beta = r$ where R is a $m \times (k+1)$ matrix. Then,

$$T(R\hat{\beta} - r)'\frac{1}{\hat{\sigma}^2}(R(\frac{X'X}{T})^{-1}R)^{-1}(R\hat{\beta} - r) \xrightarrow{d} \chi^2(m)$$

In particular, if the hypothesis only involves a single parameter, ie. $R = (0, \cdots, 0, 1, 0, \cdots, 0)$, or only one single element R is 1 and the rest are zeros, it can be transformed into a t-test. For the hypothesis like $\beta_{g+1}, \cdots, \beta_k = 0$, a different F test is also feasible. Let SSE_R, SSE_C be the residual sum of squares under the null and the alternative hypothesis. Then,

$$F = \frac{(SSE_R - SSE_C)/(k-g)}{(SSE_C)/(n-k-1)} \sim F_{k,n-k-1}$$

9 A closer look at the assumptions

1. $E(\epsilon_t) = 0$ is usually fine as long as an intercept is included. If not, just subtract the mean from the equation.

2. $E(\epsilon_t \epsilon_s) = 0$ could be violated for time series regression. Need to diagnostically check this assumption.

3. $V(\epsilon_t) = \sigma^2 \quad \forall t$ homoscedasticity assumption could be violated for cross-section regression or for financial time series.

4. $E(x_{it}\epsilon_t) = 0$ could be violated due to simultaneity. Need to use instrument variables.

10 Empirical examples

Estimating market β for Taiwanese stock market