Linear Model (Master Level) *

Shane Xinyang Xuan
ShaneXuan.com

Department of Political Science
University of California, San Diego

March 16, 2018

1 Classical regression model

Classical regression assumptions include

1.) Linearity

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \varepsilon_i,\ i = 1, 2, ..., n \tag{1}$$

$$y_i = {x_i}'_{(1\times k)}\beta_{(k\times 1)} + \varepsilon_i,\ i = 1, 2, ..., n \tag{2}$$

We can also write this model in matrix form

$$y_{(T\times 1)} = \mathbf{X}_{(T\times k)}\beta_{(k\times 1)} + \varepsilon_{(T\times 1)} \tag{3}$$

2.) Strict exogeneity: $E[\varepsilon_i|\mathbf{X}] = 0,\ i = 1, 2, ..., n$

3.) No perfect collinearity: In the sample (and therefore population), none of the explanatory variables is constant, and there are no exact linear relationships among the explanatory variables; that is, the rank of the $T \times k$ matrix $\mathbf{X}$ is k with probability 1

4.) Spherical error variance:

- Homoskedasticity: $\mathbb{E}[\varepsilon_i^2|\mathbf{X}] = \sigma^2 > 0,\ i = 1, 2, ..., n$
- No serial correlation in the error term: $\mathbb{E}[\varepsilon_i\varepsilon_j|\mathbf{X}] = 0,\ i \neq j$

Under these assumptions, the least squares coefficients are (1) linear functions of the data, (2) unbiased estimators of the population regression coefficients, (3) the most efficient unbiased estimators, and (4) maximum likelihood estimators.

*Please send your thoughts/advice to xxuan@ucsd.edu, or comment on ShaneXuan.com. Thank you so much.

1.1 Notes on Assumption (2.)

1.1.1 The unconditional mean of the error term is zero

The unconditional mean of the error term is zero. The law of total expectations states that

$$\mathbb{E}[\mathbb{E}(y|\boldsymbol{x})] = \mathbb{E}[y] \tag{4}$$

Since $\mathbb{E}[\varepsilon_i|\mathbf{X}] = 0$, we know that

$$\mathbb{E}[\varepsilon_i] = \mathbb{E}[\mathbb{E}(\varepsilon_i|\mathbf{X})] = 0 \tag{5}$$

1.1.2 The regressors are orthogonal to the error term

We first apply the law of iterated expectations:

$$\mathbb{E}[\mathbf{x}_i\varepsilon_i] = \mathbb{E}[\mathbb{E}(\mathbf{x}_i\varepsilon_i|\mathbf{x}_i)] \tag{6}$$

It follows that

$$\begin{aligned}\mathbb{E}[\mathbf{x}_i\varepsilon_i] &= \mathbb{E}[\mathbf{x}_i\mathbb{E}(\varepsilon_i|\mathbf{x}_i)] \\ &= 0\end{aligned}$$

Hence, we have shown that $\mathbb{E}[\mathbf{x}_i\varepsilon_i] = 0$ for every observation.

1.2 Notes on Assumption (4.)

We can write Assumption (4.) in a more compact way:

$$\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}] = \sigma^2\boldsymbol{I}_T \tag{7}$$

$$\equiv \operatorname{var}(\boldsymbol{\varepsilon}|\mathbf{X}) \tag{8}$$

For example, the (i, j) element of the $T \times T$ matrix $\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'$ is $\varepsilon_i\varepsilon_j$, and $\mathbb{E}[\varepsilon_i\varepsilon_j] = 0$ because the (i, j) element is on the off-diagonal of matrix $\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'$. In sum, Equation (8) is a compact way that assumes both homoskedasticity, and no serial correlation in the error term. This assumption will be relaxed in certain circumstances.

2 Finite sample properties of b

Unbiased Under Assumptions (1.)–(3.), $\mathbb{E}[\boldsymbol{b}|\mathbf{X}] = \boldsymbol{\beta}$

Proof. To prove this property, we just need to show $\mathbb{E}[\boldsymbol{b} - \boldsymbol{\beta}|\mathbf{X}] = 0$. Note that

$$\begin{aligned}
\mathbb{E}[\boldsymbol{b} - \boldsymbol{\beta}|\mathbf{X}] &= \mathbb{E}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}|\mathbf{X}] \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}[\boldsymbol{\varepsilon}|\mathbf{X}] \\
&= 0
\end{aligned}$$

∎

Variance Under Assumptions (1.)–(4.), $\text{var}(\boldsymbol{b}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$

Proof.

$$\begin{aligned}
\text{var}(\boldsymbol{b}|\mathbf{X}) &= \mathbb{E}[(\boldsymbol{b} - \boldsymbol{\beta})^2|\mathbf{X}] && (9) \\
&= \mathbb{E}[(\boldsymbol{b} - \boldsymbol{\beta})(\boldsymbol{b} - \boldsymbol{\beta})'|\mathbf{X}] && (10) \\
&= \mathbb{E}[\boldsymbol{A}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{A}'|\mathbf{X}], \quad \boldsymbol{A}' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' && (11) \\
&= \boldsymbol{A}\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}]\boldsymbol{A}' && (12) \\
&= \sigma^2\boldsymbol{A}\boldsymbol{A}' && (13) \\
&= \sigma^2 \underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'}_{\boldsymbol{I}_k}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} && (14) \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} && (15)
\end{aligned}$$

∎