GA2001 Econometrics
Solution to Problem Set 5

Junbiao Chen*

October 17, 2025

Problem 1.

1. Solution The probability limit of the OLS estimator $\hat{\beta}_n$ is

$$\beta_* = \left(\mathbb{E}[XX']\right)^{-1}\mathbb{E}[XY]$$

2. Solution If we have more parameters than observations, we can perfectly fit the data and achieves an $R^2 = 1.0$, which further leads to an over-fitting issue.

3. Solution Given data $\{x_1, x_2, \ldots, x_n\}$, define

$$\Psi(X) = \begin{bmatrix} x_1 & x_1^2 & \ldots & x_1^k \\ x_2 & x_2^2 & \ldots & x_2^k \\ \vdots & & & \\ x_n & x_n^2 & \ldots & x_n^k \end{bmatrix}$$

As such, the estimator of a polynomial regression is given by

$$\tilde{\beta}_n = \left(X'X\right)^{-1}X'Y$$

Therefore, the conditional variance is given by

$$\begin{aligned}
\mathrm{Var}\left(\tilde{\beta}_n|X_1,\ldots,X_n\right) &= \mathrm{Var}[\left(X'X\right)^{-1}X'Y|X] \\
&= \left(X'X\right)^{-1}X'\mathrm{Var}[Y|X](\left(X'X\right)^{-1}X')' \\
&= \sigma^2\left(X'X\right)^{-1}X'X\left(X'X\right)^{-1} \\
&= \sigma^2\left(X'X\right)^{-1}
\end{aligned}$$

*E-mail: jc14076@nyu.edu.

Problem 2.

1. Solution Given the data $\{Y_i, X_i\}_{i=1}^n$, the likelihood is given by

$$L = \prod_{i=1} \left(\frac{\exp(X_i'\theta)}{1+\exp(X_i'\theta)} \right)^{Y_i} \left(\frac{1}{1+\exp(X_i'\theta)} \right)^{1-Y_i}$$

It follows that the log-likelihood is given by

$$\ln L = \sum_{i=1} Y_i X_i'\theta - \ln\bigl(1+\exp\bigl(X_i'\theta\bigr)\bigr)$$

Thus,

$$Q_n = \frac{1}{n} \sum_{i=1} Y_i X_i'\theta - \ln\left[1+\exp\bigl(X_i'\theta\bigr)\right]$$

2. Solution The score is given by

$$S(\theta) = \frac{\partial Q_n}{\partial \theta} = \sum_i Y_i X_i - \bigl[1+\exp\bigl(X_i'\theta\bigr)\bigr]^{-1} X_i,$$

which is a $k \times 1$ vector (assuming $\theta \in \mathbb{R}^k$.) And Hessian is given by

$$\begin{aligned} H(\theta) &= \frac{\partial^2 \ln L}{\partial \theta \partial \theta'} \\ &= \sum_i \frac{X_i \exp(X_i'\theta) X_i'}{(1+\exp(X_i\theta))^2}, \end{aligned}$$

which is a $k \times k$ matrix.

3. Solution Under correctly specified model,

$$\mathbb{E}[Y|X] = \frac{\exp(X'\theta_0)}{1+\exp(X'\theta_0)}$$

Thus,

$$Q_\infty(\theta) = \mathbb{E}\left[YX'\theta\right] - \mathbb{E}[\ln\bigl(1+\exp\bigl(X'\theta\bigr)\bigr)]$$

By the Law of Iterated Expectation, we have

$$\begin{aligned} \mathbb{E}[YX'\theta] &= \mathbb{E}[\mathbb{E}[YX'\theta|X]] \\ &= \mathbb{E}\left[\mathbb{E}[Y|X]X'\theta\right] \\ &= \mathbb{E}\left[\frac{\exp(X'\theta_0)}{1+\exp(X'\theta_0)} X'\theta\right] \end{aligned}$$

Therefore,

$$Q_\infty(\theta) = \mathbb{E}\left[\frac{\exp(X'\theta_0)}{1+\exp(X'\theta_0)}X'\theta - \ln\left(1+\exp\big(X'\theta\big)\right)\right]$$

4. Solution Recall that the Lemma on page 28 of the lecture:

Lemma: Suppose (i) θ is compact, (ii) Q_∞ is continuous on θ, and (iii) Identification: θ_* is the unique maximizer of Q_∞ on θ, then the separation condition is satisfied.

Proof Since $\theta \in \mathbb{R}^k$, it is compact. (ii) By a convergence-argument, we would have the continuity of Q_∞. (iii) By the Lemma on page 34 of the lecture notes, we note that If $\{f(\dot{|};\theta : \theta \in \Theta)\}$ is correctly specified and one-to-one parameterized, then θ_0 is the unique solution. Thus, the separation condition is well-established.

5. Solution The Maximum Likelihood Estimator is given by

$$\hat{\theta}_n = \arg\max_{\theta\in\Theta} Q_n(\theta).$$

Therefore, we have

$$Q_n(\hat{\theta}_n) \geq Q_n(\theta_0).$$

It follows that

$$Q_n(\theta_0) - Q_n(\hat{\theta}_n) \leq 0.$$

Let $\delta_n = \sup_{\theta\in\Theta}|Q_\infty(\theta) - Q_n(\theta)|$. Since $Q_n(\hat{\theta}_n) \geq Q_n(\theta_0)$, and $Q_n(\theta_0) \geq Q_\infty(\theta_0) - \delta_n$, we have

$$Q_\infty(\theta_0) - \delta_n \leq Q_n(\hat{\theta}_n) \leq Q_\infty(\hat{\theta}_n) + \delta_n.$$

Rearranging terms, we get:

$$Q_\infty(\theta_0) - \delta_n \leq Q_\infty(\hat{\theta}_n) + \delta_n.$$

6. Solution Based on the Master Asymptotic Normality Theorem, the asymptotic distribution of the MLE is

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, V),$$

where the asymptotic variance V is the inverse of the Fisher Information matrix. The null hypothesis is given by $H_0 : \theta_{0,2} = \cdots = \theta_{0,k} = 0$. Under the null hypothesis, we can construct a Wald statistic W that follows a χ^2-distribution. We reject the null hypothesis if the calculated W statistic exceeds the critical value from the χ^2 distribution.