Econometrics III (module 5, 2023–2024)

Stanislav Anatolyev

Assignment 1

Problem 1 "Normal regression", 10 points

Let a random triple $(y, x_1, x_2)'$ be distributed as

$$\mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{y1} & \rho_{y2} \\ \rho_{y1} & 1 & \rho_{12} \\ \rho_{y2} & \rho_{12} & 1 \end{pmatrix}\right),$$

where ρ_{y1}, ρ_{y2}, ρ_{12} are unknown, but it is known that they are not zero. A researcher claims that $\mathbb{E}[y|x_1, x_2] = \alpha_1 x_1 + \alpha_2 x_2$ is the true regression, while $\mathbb{E}[y|x_1] = \beta_1 x_1$ and $\mathbb{E}[y|x_2] = \beta_2 x_2$ are not because of the omitted variables bias. Please comment on this (provocative) claim.

Problem 2 "Precision of asymptotic approximation", 40 points

Monte-Carlo study is a simulation exercise designed to shed light on the small-sample properties of estimators. The general idea is to:

(1) model the data-generating process,

(2) generate several sets of artificial data,

(3) employ these "data" and the estimator to create several estimates,

(4) use these estimates to gauge the sampling distribution properties of that estimator.

Suppose we have the following linear mean regression:

$$y = x_1\beta_1 + x_2\beta_2 + e, \quad \mathbb{E}[e|x_1, x_2] = 0.$$

There is random sample $\{(x_{1i}, x_{2i}, y_i)'\}_{i=1}^n$. Denote by $(\hat\beta_1, \hat\beta_2)'$ the OLS estimator of $(\beta_1, \beta_2)'$, and

$$\hat Q_{xx} = \frac{1}{n}\sum_{i=1}^n x_i x_i', \quad \hat V_{xe} = \frac{1}{n}\sum_{i=1}^n x_i x_i' \hat e_i^2,$$

where $x_i = (x_{1i}, x_{2i})'$ and $\hat e_i = y - x_1\hat\beta_1 - x_2\hat\beta_2$ for all $i = 1, ..., n$.

You will evaluate the quality of approximation of the exact distributions of the estimator $\hat\theta = \hat\beta_1/\hat\beta_2$ of the parameter $\theta = \beta_1/\beta_2$, and its t-statistic

$$t_\theta = \frac{\sqrt{n}\left(\hat\theta - \theta\right)}{\sqrt{\left(1/\hat\beta_2, -\hat\beta_1/\hat\beta_2^2\right)\hat Q_{xx}^{-1}\hat V_{xe}\hat Q_{xx}^{-1}\left(1/\hat\beta_2, -\hat\beta_1/\hat\beta_2^2\right)'}}$$

by corresponding asymptotic distributions.

When generating artificial data, use the following information: e is distributed as $\mathcal{N}(0, \sigma^2)$ independently of $(x_1, x_2)'$ which is distributed as $\mathcal{N}(0, I_2)$; the parameters are $\sigma^2 = 3$, $\beta_1 = 1$, $\beta_2 = 0.5$. Of course, this information cannot be used when finding the approximations or constructing statistics of interest, because a researcher does not know it. It is possible to show that under the circumstances,

$$\sqrt{n}\left(\hat{\theta} - 2\right) \xrightarrow{d} \mathcal{N}(0, 60)$$

and

$$t_2 \xrightarrow{d} \mathcal{N}(0, 1).$$

1. Generate a sample of $n = 4$ observations of $(x_{1i}, x_{2i}, y_i)'$ according to their joint distribution. Calculate $\hat{\theta}$ and t_2 for this sample. Repeat this $R = 10,000$ times. Now you have a collection of R values for $\hat{\theta}$ and t_2. These are simulated exact distributions.

2. Plot the asymptotic and simulated exact cumulative distributions on the same graph. How well does the asymptotic distribution match the simulated exact distribution? Does your answer differ for $\hat{\theta}$ and t_2?

3. Repeat what you did in parts 1–2 with $n = 20$. Discuss the role of sample size in the quality of approximations.

4. Comment thoroughly on the following claim. "The quality of asymptotic approximation depends only on the degree of nonlinearity of the transformation used in the Delta method."