Lecture 9a

Asymptotic theory for least squares – a quick summary

Lectures in SDPE: *Econometrics I* on February 26, 2024

Markus Jäntti
Swedish Institute For Social Research
Stockholm University

Introduction

(see Hansen 2021, sec 7.2, 7.3, 7.5–7.7, 7.10 and 7.16)

- We have the standard linear regression model

$$Y = \boldsymbol{X}'\beta + e \text{ with } \mathrm{E}[\boldsymbol{X}e] = 0, \tag{1}$$

an iid sample $\{(Y_i, \boldsymbol{X}_i)\}_{i=1}^n$ and the linear projection coefficients

$$\beta = (\mathrm{E}[\boldsymbol{X}\boldsymbol{X}'])^{-1}\mathrm{E}[\boldsymbol{X}Y] = \boldsymbol{Q}_{\boldsymbol{X}\boldsymbol{X}'}^{-1}\boldsymbol{Q}_{\boldsymbol{X}Y}. \tag{2}$$

- The results apply also to the CEF model (with the assumption that $\mathrm{E}[e|\boldsymbol{X}] = 0$) but the LP model is broader.
- We assume finite second moments and a positive definite design matrix; for the variance matrix estimators to be consistent, we assume finite fourth moments:
  1. $\mathrm{E}[Y^2] < \infty$
  2. $\mathrm{E}[||\boldsymbol{X}||^2] < \infty$
  3. $\boldsymbol{Q}_{\boldsymbol{X}\boldsymbol{X}'} = \mathrm{E}[\boldsymbol{X}\boldsymbol{X}']$ is positive definite.
  4. $\mathrm{E}[Y^4] < \infty$ and $\mathrm{E}[||\boldsymbol{X}||^4] < \infty$
- We'll make heavy use of the weak law of large numbers (convergence in probability), central limit theorems and convergence in distribution, rules that apply to transformations of these, as well as the delta method.

Consistency of Least-Squares Estimation

- Start with the LS estimator expressed as a sum of a non-stochastic and stochastic component re-written into

$$\widehat{\beta} - \beta = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathbf{e} = \widehat{\boldsymbol{Q}}_{XX'}^{-1}\widehat{\boldsymbol{Q}}_{Xe}, \tag{3}$$

  with $\widehat{\boldsymbol{Q}}_{Xe} = n^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i e_i$.
- The weak law of large numbers and the fact that the error is orthogonal to $\boldsymbol{X}$ imply that

$$\widehat{\boldsymbol{Q}}_{Xe} \xrightarrow{p} \mathrm{E}[\boldsymbol{X}e] = \mathbf{0}. \tag{4}$$

- Since

$$\widehat{\boldsymbol{Q}}_{XX'} \xrightarrow{p} \boldsymbol{Q}_{XX'}, \tag{5}$$

  we know that

$$\widehat{\beta} - \beta = \widehat{\boldsymbol{Q}}_{XX'}^{-1}\widehat{\boldsymbol{Q}}_{Xe} \xrightarrow{p} \boldsymbol{Q}_{XX'}^{-1}\mathbf{0} = \mathbf{0} \tag{6}$$

  so

$$\widehat{\beta} \xrightarrow{p} \beta. \tag{7}$$

- Thus, $\widehat{\beta}$ is a consistent estimator of β.

Asymptotic Normality

- To show asymptotic normality, take eq. 6 and scale it by $\sqrt{n}$:

$$\sqrt{n}(\widehat{\beta} - \beta) = \sqrt{n}\widehat{\boldsymbol{Q}}_{\boldsymbol{X}\boldsymbol{X}'}^{-1}\widehat{\boldsymbol{Q}}_{\boldsymbol{X}e} = \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{X}_i e_i\right) \tag{8}$$

- Since $\mathrm{E}[n^{-1/2}\sum_{i=1}^{n}\boldsymbol{X}_i e_i] = 0$, the central limit theorem applies and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{X}_i e_i \xrightarrow{d} N(0, \boldsymbol{\Omega}), \tag{9}$$

  where $\boldsymbol{\Omega} = \mathrm{E}[\boldsymbol{X}\boldsymbol{X}'e^2]$.

- We require $\boldsymbol{\Omega} < \infty$. To show that, resort to the Expectation Inequality (B.15) and the Cauchy-Schwartz inequality (B.17)

$$||\boldsymbol{\Omega}|| \leq \mathrm{E}[||\boldsymbol{X}e||^2] = \mathrm{E}[||\boldsymbol{X}||^2 e^2] \leq \left(\mathrm{E}[||\boldsymbol{X}||^4]\right)^{1/2}(\mathrm{E}[e^4])^{1/2}. \tag{10}$$

  By the assumption of finite fourth moments (slide 1), this is finite.

Asymptotic Normality

- Finally, using 9 and 5, the random vector on the LHS of eq. 8 converges in distribution:

$$\begin{aligned}\sqrt{n}(\widehat{\beta}-\beta) &\xrightarrow{d} \boldsymbol{Q}_{XX'}^{-1} N(0,\boldsymbol{\Omega}) \\ &= N(0, \boldsymbol{Q}_{XX'}^{-1}\boldsymbol{\Omega}\boldsymbol{Q}_{XX'}^{-1}) \\ &= N(0, \boldsymbol{V}_{\beta})\end{aligned} \tag{11}$$

- Note that the asymptotic variance $\mathrm{Var}[\widehat{\beta}] = \boldsymbol{V}_{\beta}$ is different from the finite sample conditional variance $\boldsymbol{V}_{\widehat{\beta}}$ but the two are related by

$$n\boldsymbol{V}_{\widehat{\beta}} \xrightarrow{p} \boldsymbol{V}_{\beta}. \tag{12}$$

- With *homoscedastic projection error*

$$\mathrm{Cov}[\boldsymbol{XX'}, e^2] = \boldsymbol{0} \tag{13}$$

  (includes homoscedastic errors but is broader) get

$$\boldsymbol{\Omega} = \mathrm{E}[\boldsymbol{XX'}]\mathrm{E}[e^2] = \boldsymbol{Q}_{XX'}\sigma^2; \quad \boldsymbol{V}_{\beta} = \boldsymbol{Q}_{XX'}^{-1}\sigma^2 \equiv \boldsymbol{V}_{\beta}^{0} \tag{14}$$

  giving the homoscedastic asymptotic covariance matrix.

Consistency of Error Variance Estimators

- To show that the error variance estimators are consistent, start from the regression residual

$$\hat{e}_i = Y_i - \boldsymbol{X}'\widehat{\beta} = e_i + \boldsymbol{X}_i'\beta - \boldsymbol{X}_i'\widehat{\beta} = e_i - \boldsymbol{X}_i'(\widehat{\beta} - \beta) \tag{15}$$

- Squaring both sides of eq 15 gives

$$\hat{e}_i^2 = e_i^2 - 2e_i\boldsymbol{X}_i'(\widehat{\beta} - \beta) + (\widehat{\beta} - \beta)'\boldsymbol{X}\boldsymbol{X}'(\widehat{\beta} - \beta) \tag{16}$$

- Now sum across all sample units and divide by sample size

$$\begin{aligned}\hat{\sigma}^2 = &\underbrace{\frac{1}{n}\sum_{i=1}^{n} e_i^2}_{\xrightarrow{p}\sigma^2} - 2\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} e_i\boldsymbol{X}_i'\right)}_{\xrightarrow{p}\mathrm{E}[e\boldsymbol{X}']=\boldsymbol{0}}(\underbrace{\widehat{\beta}}_{\xrightarrow{p}\beta} - \beta) + \\ &(\underbrace{\widehat{\beta}}_{\xrightarrow{p}\beta} - \beta)'\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'\right)}_{\xrightarrow{p}\boldsymbol{Q}_{XX'}}(\underbrace{\widehat{\beta}}_{\xrightarrow{p}\beta} - \beta)\end{aligned} \tag{17}$$

- Thus, by the WLLN, we have

$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2 \tag{18}$$

i.e., $\text{plim}\, \hat{\sigma}^2 = \sigma^2$.

- $n/(n-k) \to 1$ as $n \to \infty$ so

$$s^2 = \frac{n}{n-k} \hat{\sigma}^2 \xrightarrow{p} \sigma^2. \tag{19}$$

- We know $\text{plim}\, \widehat{\boldsymbol{Q}}_{XX'}^{-1} = \boldsymbol{Q}_{XX'}^{-1}$ so we know the estimator of $\text{Var}[\widehat{\beta}] = \boldsymbol{V}_{\beta}^{0}$ under homoscedasticity

$$\widehat{\boldsymbol{V}}_{\beta}^{0} = \widehat{\boldsymbol{Q}}_{XX'}^{-1} s^2 \tag{20}$$

is consistent.

- We have seen $\mathrm{Var}[\widehat{\beta}] = \boldsymbol{V}_\beta = \boldsymbol{Q}_{\boldsymbol{XX}'}^{-1}\boldsymbol{\Omega}\boldsymbol{Q}_{\boldsymbol{XX}'}^{-1}$.
- One "plug-in" or moment estimator is constructed using

$$\widehat{\boldsymbol{V}}_\beta = \widehat{\boldsymbol{Q}}_{\boldsymbol{XX}'}^{-1}\widehat{\boldsymbol{\Omega}}\widehat{\boldsymbol{Q}}_{\boldsymbol{XX}'}^{-1} = \widehat{\boldsymbol{Q}}_{\boldsymbol{XX}'}^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'\hat{e}_i^2\right)\widehat{\boldsymbol{Q}}_{\boldsymbol{XX}'}^{-1} \tag{21}$$

- We know $\mathrm{plim}\,\widehat{\boldsymbol{Q}}_{\boldsymbol{XX}'}^{-1} = \boldsymbol{Q}_{\boldsymbol{XX}'}^{-1}$ so the consistency of an estimator of $\boldsymbol{V}_\beta$ depends on the estimator of $\boldsymbol{\Omega}$.
- Examine

$$\widehat{\boldsymbol{\Omega}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'\underbrace{\hat{e}_i^2}_{=\hat{e}_i^2+e_i^2-e_i^2} = \underbrace{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'e_i^2}_{\xrightarrow{p}\mathrm{E}[\boldsymbol{XX}'e^2]=\boldsymbol{\Omega}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i'(\hat{e}_i^2 - e_i^2)}_{\xrightarrow{p}\boldsymbol{0}?}. \tag{22}$$

Functions of Parameters

- For many purposes, including hypothesis testing, our object of interest is not β but some function $r : \mathbb{R}^k \longrightarrow \mathbb{R}^q$ of it:

$$\widehat{\theta} = r(\widehat{\beta}). \tag{23}$$

- By the *continuous mapping theorem* and $\operatorname{plim} \widehat{\beta} = \beta$, $\operatorname{plim} \widehat{\theta} = \theta$ ($r()$ needs to be continuous at least in the vicinity of β).
- By the Delta method for a vector, if $r()$ is differentiable (at least at the true β) and the matrix of partial derivatives $\boldsymbol{R}_\beta = \partial r(\beta)'/\partial\beta$ has full rank (q), $\widehat{\theta}$ is asymptotically normal:

$$\sqrt{n}(\widehat{\theta} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{R}_\beta' \boldsymbol{V}_\beta \boldsymbol{R}_\beta) = N(\mathbf{0}, \boldsymbol{V}_\theta) \tag{24}$$

- A special case is the linear $r()$, $\boldsymbol{R}'\beta$ for a $k \times q$ matrix $\boldsymbol{R}$.
- A "selector matrix"

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{I}_{k_1} \\ \boldsymbol{0}_{k_2} \end{pmatrix} \tag{25}$$

applied to a conformably partitioned $\beta' = (\beta_1' \; \beta_2')$ so that $\boldsymbol{R}'\beta = \beta_1$ has

$$\boldsymbol{V}_{\boldsymbol{\theta}} = \begin{pmatrix} \boldsymbol{I}_{k_1} & \boldsymbol{0}_{k_2} \end{pmatrix} \boldsymbol{V}_{\beta} \begin{pmatrix} \boldsymbol{I}_{k_1} \\ \boldsymbol{0}_{k_2} \end{pmatrix} = \boldsymbol{V}_{11} \tag{26}$$

(see Hansen 2021, p. 175)

- The result is that

$$\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{V}_{11}) \tag{27}$$

- Typical non-linear functions might involve ratios of elements of β (see e.g. Hansen 2021, p. 175).

- The *estimator* of $\boldsymbol{V}_{\boldsymbol{\theta}}$ is
$$\widehat{\boldsymbol{V}}_{\boldsymbol{\theta}} = \widehat{\boldsymbol{R}}_{\beta}' \widehat{\boldsymbol{V}}_{\beta} \widehat{\boldsymbol{R}}_{\beta}. \tag{28}$$
- $\widehat{\boldsymbol{V}}_{\beta}$ can be replaced by the relevant consistent alternative.
- $\widehat{\boldsymbol{V}}_{\boldsymbol{\theta}}$ is consistent, because $\widehat{\boldsymbol{V}}_{\beta}$ is and, by the continuity of the partial derivatives and consistency of $\widehat{\beta}$,
$$\widehat{\boldsymbol{R}}_{\beta} = \frac{\partial}{\partial \beta} r(\widehat{\beta})' \xrightarrow{p} \frac{\partial}{\partial \beta} r(\beta)' = \boldsymbol{R}_{\beta}. \tag{29}$$
- The asymptotic standard errors (=estimated standard deviations of the distribution of the estimator) for $\widehat{\beta}$ or $\widehat{\theta}$ are the square roots of the main diagonal diagonal elements of $\widehat{\boldsymbol{V}}_{\widehat{\beta}}$ and $\widehat{\boldsymbol{V}}_{\widehat{\theta}}$.

References

Hansen, Bruce E (2021). *Econometrics*. Madison, WI: University of Wisconsin.