Section 11: Asymptotic Results

JJ Chen

November 7, 2014

Review of Week 11

Key Results: 3 theorems

- So far in class we have proved three theorems:
  - Thm1: $\operatorname{plim} \hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$
  - Thm2: $\operatorname{plim} \hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$
  - Thm3: $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathrm{N}(\mathbf{O}, \sigma^2 \boldsymbol{Q}^{-1})$
- based on the following assumptions:
  - a: $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{\varepsilon}$
  - b: $\{\boldsymbol{x}_i', \varepsilon_i\}$ be an i.i.d. sequence
  - ci: $\mathrm{E}(\varepsilon_i \mid \boldsymbol{x}_i) = 0$
  - cii: $\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}) = \sigma^2 \boldsymbol{I}, \sigma^2 < \infty$
  - di: $\mathrm{E}(\mid x_{ij} \mid^2) < \infty, j = 1, \cdots, k$
  - dii: $\mathrm{E}(\boldsymbol{x}_i \boldsymbol{x}_i') = \boldsymbol{Q}$, positive definite

Key Results: Est.Asym.Var of $\hat{\boldsymbol{\beta}}$

- Thm 3 (asymtotic normality) "implies" that $\hat{\boldsymbol{\beta}} \xrightarrow{d} \mathrm{N}(\boldsymbol{\beta}, \frac{\sigma^2}{n}\boldsymbol{Q}^{-1})$, which allows us to see the asymtotic covariance matrix of $\hat{\boldsymbol{\beta}}$ as $\frac{\sigma^2}{n}\boldsymbol{Q}^{-1}$
    - Note that $\boldsymbol{Q} = \mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$.
    - A consistent estimator is $\frac{1}{n}\sum\limits_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'$ (based on di, dii, and LLN)
    - So that a consistent estimator for $\boldsymbol{Q}^{-1}$ is $n(\sum\limits_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i')^{-1}$
    - In homework you will prove s^2 is a consistent estimator for σ^2
- Together we get the est. asym. cov. matrix for $\hat{\boldsymbol{\beta}}$ as
    - $\frac{s^2}{n}n(\sum\limits_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i')^{-1} = s^2(\sum\limits_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i')^{-1} = s^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

Asymtotic Distribution of the Wald-Stat

Proof 1: Notation

- Recall that under $\mathrm{H}_0 : \boldsymbol{R}\hat{\boldsymbol{\beta}} = q$, thus based on Thm3
  - $\sqrt{n}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}) \xrightarrow{d} \mathrm{N}(\mathbf{O}, \sigma^2 \boldsymbol{R}\boldsymbol{Q}^{-1}\boldsymbol{R}')$
- Define $\boldsymbol{z} = \sqrt{n}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q})$ and $\boldsymbol{P} = \sigma^2 \boldsymbol{R}\boldsymbol{Q}^{-1}\boldsymbol{R}'$, so that
  - $\boldsymbol{z} \xrightarrow{d} \mathrm{N}(\mathbf{O}, \boldsymbol{P})$
- Since $\boldsymbol{P}$ is positive definite (we've imposed $\boldsymbol{Q}$ to be positive definite in assumption dii of Thm3), $\boldsymbol{P}^{-1}$ exists so does its sqrt matrix $\boldsymbol{P}^{-\frac{1}{2}}$. Note that
  - $\boldsymbol{P}^{-\frac{1}{2}}\boldsymbol{P}^{-\frac{1}{2}} = \boldsymbol{P}^{-1}$

Proof 2: Quadratic Form and χ^2

- Since $\boldsymbol{z} \xrightarrow{d} \mathrm{N}(\mathbf{O}, \boldsymbol{P})$
  - $\boldsymbol{P}^{-\frac{1}{2}}\boldsymbol{z} \xrightarrow{d} \mathrm{N}(\boldsymbol{P}^{-\frac{1}{2}}\mathbf{O}, \boldsymbol{P}^{-\frac{1}{2}}\boldsymbol{P}\boldsymbol{P}^{-\frac{1}{2}})$, or
  - $\boldsymbol{P}^{-\frac{1}{2}}\boldsymbol{z} \xrightarrow{d} \mathrm{N}(\mathbf{O}, \boldsymbol{I})$
- Thus
  - $(\boldsymbol{P}^{-\frac{1}{2}}\boldsymbol{z})'(\boldsymbol{P}^{-\frac{1}{2}}\boldsymbol{z}) = \boldsymbol{z}'\boldsymbol{P}^{-1}\boldsymbol{z} \xrightarrow{d} \chi^2(J)$
- Note that $\boldsymbol{z}'\boldsymbol{P}^{-1}\boldsymbol{z}$ is a quadratic form of a J-dimensional asymptotically normal random vector with the inverse of its covariance matrix $\boldsymbol{P}^{-1}$, this is distributed as a χ^2 with J degrees of freedom. (J is the $\#$ of linear restrictions.)

Proof 3: Plug in for What $\boldsymbol{z}$ and $\boldsymbol{P}^{-1}$ are

- $\boldsymbol{z}'\boldsymbol{P}^{-1}\boldsymbol{z} \xrightarrow{d} \chi^2(J)$
  - $\implies \sqrt{n}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q})^{-1}[\sigma^2\boldsymbol{R}\boldsymbol{Q}^{-1}\boldsymbol{R}']^{-1}\sqrt{n}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}) \xrightarrow{d} \chi^2(J)$
  - $\implies (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q})^{-1}n[\sigma^2\boldsymbol{R}\boldsymbol{Q}^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}) \xrightarrow{d} \chi^2(J)$
  - $\implies (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q})^{-1}[\frac{1}{n}\sigma^2\boldsymbol{R}(\frac{\boldsymbol{X}'\boldsymbol{X}}{n})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}) \xrightarrow{d} \chi^2(J)$
  - $\implies (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q})^{-1}[\sigma^2\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q}) \xrightarrow{d} \chi^2(J)$
- Thus the Wald statistic is asymptotically distributed as $\chi^2(J)$.
  - Note that $W = (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q})^{-1}[\sigma^2\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{q})$
  - Note also that above we use the fact $(\frac{\boldsymbol{X}'\boldsymbol{X}}{n})^{-1} \xrightarrow{p} \boldsymbol{Q}^{-1}$
  - Since $s^2 \xrightarrow{p} \sigma^2$, we can replace the σ^2 with s^2 in W and the result still holds

Asymtotic Relationship between F and χ^2 distribution

- Recall that $F = \frac{(\boldsymbol{R}\hat{\boldsymbol{\beta}}-\boldsymbol{q})^{-1}[\sigma^2\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}}-\boldsymbol{q})/J}{\frac{\boldsymbol{e}'\boldsymbol{e}}{\sigma^2}/n-k}$
- We can rewrite it as
  - $F = \frac{\chi^2_J/J}{\chi^2_{n-k}/n-k}$
- Note that as $n \to \infty$, the denominator "converge" to 1 because
  - $\mathrm{E}(\frac{\chi^2_{n-k}}{n-K}) = \frac{n-k}{n-k} = 1$
  - $\mathrm{V}(\frac{\chi^2_{n-k}}{n-k}) = \frac{2(n-k)}{(n-k)^2} = \frac{2}{n-k} \to 0$
  - Thus the distribution degenerate to a constant 1 (its mean)
- Thus $F \overset{a}{\sim} \chi^2(J)/J$. This is the basis for large sample hypothesis testing, which is based on the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ rather than the exact normal distribution of ε (A6).

Comments and Notation Alerts

Comments and Notation Alert 1

- It might be helpful to define asymptotic bias of $\hat{\boldsymbol{\beta}}$ as $\operatorname{plim} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$. Thus consistency can be treated as "asymptotic unbiasedness".
- It might be helpful to write $\operatorname{plim}_{n\to\infty} \hat{\boldsymbol{\beta}}_n = \boldsymbol{\beta}$ at the beginning. It reminds me that $\{\hat{\boldsymbol{\beta}}_n\}$ is also a sequence of random vector.
- Sometimes we write "$\xrightarrow{d}$" as "$\overset{a}{\sim}$" to link to the exact distribution notation "$\sim$".

Comments and Notation Alert 2

- Sometimes we switch to observation specific notation "$\boldsymbol{x}_i$". Different authors have difference preference for the exact definition of $\boldsymbol{x}_i$:
  - some prefer to use it as a row vector coming directly from the design matrix $\boldsymbol{X}$
  - some prefer to keep it as a column vector.
  - Here we treat it as a $k \times 1$ column vector, but try to distinguish it from to the characteristic specific column vector $\boldsymbol{x}_k$, which is $n \times 1$.
  - You can think of our $\boldsymbol{x}_i$ as a **transpose** of a generic row vector from our matrix $\boldsymbol{X}$

Comments and Notation Alert 3

- We can rewrite the familiar matrix notation of $\hat{\boldsymbol{\beta}}$ as
  - $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = [\sum_i \boldsymbol{x}_i\boldsymbol{x}_i']^{-1}\sum_i \boldsymbol{x}_i y_i$
  - Note that y_i is not a vector
- Consider a simple example of $\boldsymbol{X}'\boldsymbol{y}$:
  - $\boldsymbol{X}'\boldsymbol{y} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}' \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_{11}y_1 + x_{21}y_2 \\ x_{12}y_1 + x_{22}y_2 \end{pmatrix}$
  - $\sum_i^2 \boldsymbol{x}_i y_i = \boldsymbol{x}_1 y_1 + \boldsymbol{x}_2 y_2 = \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix} y_1 + \begin{pmatrix} x_{21} \\ x_{22} \end{pmatrix} y_2$

Comments and Notation Alert 4

- To see one advantage of the observation specific notation, rewrite the OLS estimator again:
  - $\hat{\boldsymbol{\beta}} = [\frac{1}{n}\sum_i \boldsymbol{x}_i\boldsymbol{x}_i']^{-1}[\frac{1}{n}\sum_i \boldsymbol{x}_i y_i]$
- We've introduced this notation before and interpret it as a sample analog
- Treating OLS estimators as functions of sample average is the key to most asymptotic results: it allows us to apply the Law of Large Number and Central Limit Theorem
  - There are many versions of LLN and CLT. If you are proving asymptotic results in your problem set, try to include the specific versions of LLN or CLT you are using, so that both you and me know clearly which assumptions are necessary for your proof.