Microecononometrics Cheatsheet

J.T. Cho [joncho@seas.upenn.edu]

The Basics

PDF of Normal Distribution

$$X \sim \mathcal{N}(\mu, \sigma^2) \ : \ f(X) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(- \frac{(X-\mu)^2}{2\sigma^2} \right)$$

Variance of a Random Variable

$$\text{Var}(X) = E[(X-\mu)^2] = E[X^2] - E^2[X]$$

Independence of R.V.'s

$$\begin{aligned} f(X,Y) =& f(X)f(Y) \\ & \implies E[X|Y] = E[X] \\ & \implies \text{Cov}(X,Y) = 0 \end{aligned}$$

Correlated R.V.'s

$$\text{Cov}(X,Y) = E[XY] - \mu_X \mu_Y$$

If $\text{Cov}(X,Y) = 0$, X, Y are uncorrelated.

Properties of Estimators

(i) unbiasedness, $E[\hat{\mu_y}] = \mu_y$

(ii) consistency, $\forall \epsilon > 0, \ \Pr(\mid \hat{\mu_y} - \mu_y \mid > \epsilon) \to 0$ as $n \to \infty$

Ordinary Least Squares (Simple)

Identifying Assumptions

- nonstochastic x_i, s.t. only source of random variation in y_i is ϵ_i
- mean-zero; $E[\epsilon] = \mathbf{0}$
  - $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$
- homoskedasticity; $E[\epsilon_i^2] = \sigma^2$
- $\text{Cov}(\epsilon_i, \epsilon_j) = 0 \ \forall \ i \neq j$

Derivation of Matrix-formulation of OLS (Stochastic Regressors)

$$\hat{\beta} = \text{argmin}_{\beta}\epsilon'\epsilon = \text{argmin}_{\beta}(Y - X\beta)'(Y - X\beta)$$

$$\begin{aligned}
\frac{\partial}{\partial\beta} &= 2X'(Y - X\beta) = 0 \\
X'Y &- X'X\beta = 0 \\
X'Y &= X'X\beta \\
\beta &= (X'X)^{-1}X'Y
\end{aligned}$$

Since $\epsilon = Y - X\beta$, the F.O.C. implies that $X'\epsilon = 0$. If X has a column of 1's corresponding to a constant term, then X' in the expression implies that

$$\sum_{i=1}^{n} \epsilon_i = 0$$

Then, the mean of the residuals is 0.

Consistency of OLS Estimator

We give the asymptotic distribution of $\hat{\beta} - \beta$.

$$\begin{aligned}
\hat{\beta} - \beta &= (X'X)^{-1}X'Y - \beta \\
&= (X'X)^{-1}X'(X\beta + \epsilon) - \beta \\
&= \beta + (X'X)^{-1}X'\epsilon - \beta \\
&= (X'X)^{-1}X'\epsilon
\end{aligned}$$

Since $\epsilon \sim \mathcal{N}(0, \sigma^2 I)$, observe,

$$\begin{aligned}
\hat{\beta} - \beta &\sim \mathcal{N}(0, (X'X)^{-1}X'\sigma^2 I X(X'X)^{-1}) \\
&\sim \mathcal{N}(0, \sigma^2(X'X)^{-1}X'X(X'X)^{-1}) \\
&\sim \mathcal{N}(0, \sigma^2(X'X)^{-1})
\end{aligned}$$

<u>Estimator for σ^2 for the variance-covariance matrix</u>

$$\hat{\sigma}^2 = \frac{1}{n-k}\sum_{i=1}^{n}\hat{\epsilon}_i^2$$

Heteroskedasticity

Relax assumption that $E[\epsilon_i^2] = \sigma^2$ for all i. Then, let $E[\epsilon\epsilon'] = \Omega$,

$$\Omega = \begin{pmatrix} \sigma_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_n^2 \end{pmatrix}$$

The robust var-cov estimator is

$$\hat{\Sigma}_{robust} = (\sum_{i=1}^{n} x_i x_i')^{-1} \sum_{i=1}^{n} x_i x_i' \hat{\epsilon}_i^2 (\sum_{i=1}^{n} x_i X_i')^{-1}$$

Generally Useful Techniques

Convergence

Suppose we have the sequence of random variables $\{X_i\}_{i=1}^n$ with sample mean $\frac{1}{n}\sum_{i=1}^n X_i$.

Convergence in probability

For all $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr(\mid X_n - X \mid < \epsilon) = 1$$

Alternatively, is written $\text{plim } X_n = X$.

Implies convergence in distribution.

Convergence in mean-square

$$\lim_{n\to\infty} E[(X_n - X)^2] = 0$$

Implies convergence in probability.

Convergence almost surely

For all $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr(\mid X_m - X \mid < \epsilon \;\forall\; m \geq n) = 1$$

In other words, X_m converges to X for some n and stays small for all larger n.

Implies convergence in probability.

Convergence in distribution

$$\lim_{n\to\infty} F_n(X_n) = F(X) \text{ for all } X \text{ such that } F(X) \text{ is continuous}$$

Let $F_i(x)$ be the CDF of X.

Slutsky's Theorem

If $X_n \to_D X$ and $A_n \to_P a$ for constant a, then

(i) $A_n X_n \to_D aX$

(ii) $A_n + X_n \to_D a + X$

(iii) $\frac{X_n}{A_n} \to_D \frac{X}{a}$

Mann-Wald/Continuous-Mapping Theorem

Let g be a continuous function. Then,

(a) $X_n \to_D X \implies g(X_n) \to_D g(X)$

(b) $X_n \to_P X \implies g(X_n) \to_P g(X)$

(c) $X_n \to_{A.S.} X \implies g(X_n) \to_{A.S.} g(X)$

In other words, continuous functions are *limit preserving* even if arguments are sequences of R.V.'s.

Chebyshev's Weak Law of Large Numbers

If $\{Y_i\}_i^n$ are i.i.d. with $E[Y_i] = \mu_Y, \text{Var}(Y_i) = \sigma_Y^2 < \infty$, then $\bar{Y} \to_P \mu_Y$.

Lindeberg-Levy Central Limit Theorem For i.i.d. R.V.'s $\{X_i\}_i^n$ with mean μ and variance σ^2, consider the sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$E[\bar{X}_n] = \mu$$

$$\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$$

Then, defining the Z statistic, Z_n,

$$Z_n = \frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

The LLCLT states that the R.V.'s $Z_n \to_D \mathcal{N}(0,1)$

Law of Iterated Expectations

$$E[X'\epsilon] = E[E[X'\epsilon]] = E[X'E[\epsilon \mid X]]$$

OLS with Stochastic Regressors

Identifying Assumptions

Allow X to be random. ϵ_i are i.i.d., with

$$\begin{aligned} E[\epsilon_i \mid x_i] = 0, E[\epsilon_i^2 \mid x_i] = \sigma^2 \\ E[x_i x_i'] < \infty, E[x_i x_i']^{-1} \text{ exists} \\ E[\epsilon_i \epsilon_j \mid X] = 0 \; \forall \; i \neq j \end{aligned}$$

Unbiasedness

Observe that expectations are taken over x_i and ϵ_i.

$$\begin{aligned} E[\hat{\beta}] &= E[(X'X)^{-1}X'Y] \\ &= E[(X'X)^{-1}X'(X\beta + \epsilon)] \\ &= E[\beta + (X'X)^{-1}X'\epsilon] \\ &= \beta + E[(X'X)^{-1}X'\epsilon] \\ &= \beta + E[(X'X)^{-1}X'E[\epsilon \mid X]] \\ &= \beta \end{aligned}$$

Asymptotic Distribution

Consider the R.V. sequence $\{x_i\epsilon_i\}_{i=1}^n$.

Observe that $E[x_i\epsilon_i] = 0$, applying the L.I.E.

$$\begin{aligned} \mathrm{Var}(x_i\epsilon_i) &= E[x_i^2\epsilon_i^2] \\ &= E[x_i^2 E[\epsilon_i^2 \mid x_i]] \\ &= E[x_i x_i' \cdot \sigma^2] \\ &= E[x_i x_i']\sigma^2 \end{aligned}$$

$$
\begin{aligned}
\hat{\beta} - \beta &= (X'X)^{-1}(X'\epsilon) \\
&= [\sum_{i=1}^{n} x_i x_i']^{-1}[\sum_{i=1}^{n} x_i \epsilon_i] \\
&= [\frac{1}{n}\sum_{i=1}^{n} x_i x_i']^{-1}[\frac{1}{n}\sum_{i=1}^{n} x_i \epsilon_i]
\end{aligned}
$$

Observe that for any constant c, $(cX)^{-1} = \frac{1}{c}X^{-1}$.

Consider the following expression:

$$
\sqrt{n}(\hat{\beta} - \beta) = [\frac{1}{n}\sum_{i=1}^{n} x_i x_i']^{-1}[\frac{1}{n} \cdot \sqrt{n}\sum_{i=1}^{n} x_i \epsilon_i]
$$

Applying the LLCLT to the second term gives

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^{n} x_i \epsilon_i \sim \mathcal{N}(0, \sigma^2 E[x_i x_i'])
$$

Since x_i' is a $1 \times k$ vector, the quantity $E[x_i x_i']$ is a matrix. Noting that by assumption plim $\frac{1}{n}\sum_{i=1}^{n} x_i x_i' = E[x_i x_i']$ and applying Slutsky's theorem,

$$
\begin{aligned}
\sqrt{n}(\hat{\beta} - \beta) &\sim \mathcal{N}(0, \sigma^2 E[x_i x_i']^{-1} E[x_i x_i'] E[x_i x_i']^{-1}) \\
&\sim \mathcal{N}(0, \sigma^2 E[x_i x_i']^{-1})
\end{aligned}
$$

(It is helpful to remember that convergence in probability implies convergence in distribution, which is necessary for the LLCLT).

Consistency

From the above expression for $\hat{\beta} - \beta$, it immediately follows from the WLLN that the second term $\frac{1}{n}\sum_{i=1}^{n} x_i \epsilon_i \to_P 0$. Then, applying Slutsky's theorem, the product converges in probability to 0.