FHS Microeconometrics Notes

Harry Folkard, Keble College

2022

Abstract

These are my Microeconometrics notes made for my finals in 2022. They cover all of the topics. Feel free to use these notes and pass them on to others. Please note, however, that these have just been made by a student and not checked over. They likely contain errors, so it will be worth checking things for yourself. Thanks to Kevin Sheppard and Vanessa Berenguer Rico - these notes are just my interpretation of their lectures and tutorials.

Contents

Need to Know Theorems

Chebyshev iid LLN

Theorem *(Law of Large Numbers by Chebyshev)*

For $i = 1, ..., n$ let x_i be independent and identically distributed with finite mean, μ, and variance σ^2. Then, as $n \to \infty$,

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i \xrightarrow{P} \mu.$$

Lindeberg-Levy iid CLT

Theorem *(Central Limit Theorem by Lindeberg-Levy)*

For $i = 1, ..., n$ let x_i be independent and identically distributed with finite mean, μ, and variance σ^2. Then, as $n \to \infty$,

$$\frac{\sqrt{n}(\bar{x}_n - \mu)}{\sigma} \xrightarrow{D} N(0, 1)$$

Multivariate iid CLT

Theorem *(Multivariate Lindeberg-Levy CLT)*

Let Z_i for $i = 1, ..., n$ be independent and identically distributed m-dimensional random vectors with finite mean vector $\mu_Z = E[Z_i]$, and finite positive definite covariance matrix $\Sigma_Z = E[(Z_i - \mu_Z)(Z_i - \mu_Z)']$. Then

$$\sqrt{n}(\bar{Z}_i - \mu_Z) \xrightarrow{D} N(0_m, \Sigma_Z)$$

where $\bar{Z}_n = n^{-1} \sum_{i=1}^{n} Z_i$ and $N(0_m, \Sigma_Z)$ is multivariate normal.

Slutsky's Theorem

Theorem *(Slutsky Theorem)*

Let $Y_n \xrightarrow{P} c$ and $X_n \xrightarrow{D} X$, then:

(a) $Y_n + X_n \xrightarrow{D} c + X$
(b) $Y_n X_n \xrightarrow{D} cX$
(c) $Y_n^{-1} X_n \xrightarrow{D} c^{-1} X$ if $c \neq 0$
(d) If $c = 0$ then $Y_n X_n \xrightarrow{P} 0$

Scalar OLS Algebraic Facts

Model

$$y_i = \beta_1 + \beta_2 x_i + u_i$$

OLS Problem & FOCs

$$argmin_{(\beta_1,\beta_2)\in\mathbb{R}^2} \sum_{i=1}^{n}(y_i - \beta_1 - \beta_2 x_i)^2$$
$$\sum_{i=1}^{n}(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i) = 0$$
$$\sum_{i=1}^{n}(y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i)x_i = 0$$

Solutions

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$$
$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

From which, with some basic algebra, we get,

$$\begin{aligned}
\hat{\beta}_1 &= \bar{y} - \hat{\beta}_2 \bar{x} \\
&= (\beta_1 + \beta_2 \bar{x} + \bar{u}) - \hat{\beta}_2 \bar{x} \\
&= \beta_1 + \bar{u} - (\hat{\beta}_2 - \beta_2)\bar{x}
\end{aligned}$$

$$\begin{aligned}
\hat{\beta}_2 &= \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\
&= \frac{\sum_{i=1}^{n}(\beta_1 + \beta_2 x_i + u_i - \beta_1 - \beta_2 \bar{x} - \bar{u})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\
&= \frac{\sum_{i=1}^{n}\beta_2(x_i - \bar{x})^2 + (u_i - \bar{u})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\
&= \beta_2 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})(u_i - \bar{u})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\
&= \frac{\sum_{i=1}^{n}\beta_2(x_i - \bar{x})^2 + (u_i - \bar{u})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\
&= \beta_2 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}
\end{aligned}$$

Predicted Values

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$$

Residual

$$\hat{u}_i = y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i$$

Hence, by construction (FOCs),

$$\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i = 0$$

$$\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i x_i = 0$$

Mean

$$\bar{y} = \beta_1 + \beta_2\bar{x} + \bar{u}$$

and

$$y = \hat{\beta}_1 + \hat{\beta}_2\bar{x}$$

Other Useful Facts

$$\sum_{i=1}^{n}(x_i - \bar{x})(u_i - \bar{u}) = \sum_{i=1}^{n}(x_i - \bar{x})u_i$$

$$\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i = \bar{y}$$

$$(y_i - \bar{y}) = (u_i - \bar{u}) + \beta_2(x_i - \bar{x})$$

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2$$

Proof,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}(x_i^2 + \bar{x}^2 - 2x_i\bar{x}) &= \frac{1}{n}[\sum_{i=1}^{n}x_i^2 + n\bar{x}^2 - 2\bar{x}\sum_{i=1}^{n}x_i] \\
&= \frac{1}{n}\sum_{i=1}^{n}x_i^2 + \bar{x}^2 - 2\bar{x}\frac{1}{n}\sum_{i=1}^{n}x_i \\
&= \frac{1}{n}\sum_{i=1}^{n}x_i^2 + \bar{x}^2 - 2\bar{x}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2
\end{aligned}$$

Scalar Proofs

Introduction

These six types of proofs cover a large part of Part A, and to an extent many of the skills require for Part B, in the exam. It is therefore really worth your while getting to properly understand them as well as being able to do them for many different models.

The required proofs are:

(1) ***Finding OLS Estimators***

(2) ***Showing Unbiasedness***

(3) ***Finding (Conditional) Variance***

(4) ***Proving Consistency***

- By Mean Square Convergence (MSC) *and* directly. Proving by MSC is a huge time saver in exams, especially if earlier questions have already asked to you find the expectation and variance - why waste time trying to find the right scaling factor when you can just show that in the limit the variance is zero.

(5) ***Finding Asymptotic Distribution***

(6) ***Showing Asymptotic Normality of t-statistic***

For all of these proofs I give general pointers - what we are trying to achieve/show - and then an example proof.

Question for Examples:

These example proofs use a model taken for the 2021 Oxford Econometrics paper, unless stated otherwise. They focus on the model and distribution,

$$y_i = \beta + u_i \text{ where } u_i \sim \text{independent } N(0, \{1/2\}^i) \;\; i = 1, ..., n$$

Find OLS Estimator

The OLS estimator is the solution to minimising the sum of *squared* error.

Example Proof:

$$\hat{\beta} = argmin \sum_{i=1}^{n}(y_i - \beta)^2$$

FOC,

$$0 = -2\sum_{i=1}^{n}(y_i - \hat{\beta})$$

$$0 = \sum_{i=1}^{n} y_i - n\hat{\beta}$$

Hence,

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Unbiased

Take expectations, looking to show that expectation of estimator is the population parameter.

Properties of Expectations:

$$E[a + bX + cY] = a + bE[X] + cE[Y]$$

$$E[\bar{z}] = E[\frac{1}{n}\sum_{i=1}^{n} z_i] = \frac{1}{n}\sum_{i=1}^{n} E[z_i] = \mu_z$$

$$E[y_i] = E[E[y_i \mid x_i]]$$

hence, usefully, $E[x_i u_i] = E[E[x_i u_i \mid x_i]] = E[x_i E[u_i \mid x_i]]$

Example Proof:

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n}(\beta + u_i) = \beta + \frac{1}{n}\sum_{i=1}^{n} u_i$$

$$E[\hat{\beta}] = E[\beta] + E[\frac{1}{n}\sum_{i=1}^{n} u_i]$$

$$= \beta + \frac{1}{n}\sum_{i=1}^{n} E[u_i] \text{ where } E[u_i] = 0$$

$$E[\hat{\beta}] = \beta$$

(Conditional) Variance

Find the variance of an estimator, recall that an estimator is a random variable.

Properties of Variance (and Covariance):

$$Var(a + bX) = b^2 Var(X)$$
$$Var(\sum_{i=1}^{n} Var(x_i)) = \sum_{i=1}^{n} Var(x_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Cov(x_i, x_j)$$
$$Cov(aX + b, cY + dZ) = aCov(X, Y) + adCov(X, Z) + c\underbrace{Cov(b, Y)}_{=0} + d\underbrace{Cov(b, Z)}_{=0}$$

Example Proof:

$$\begin{aligned}
Var(\hat{\beta}) &= Var(\frac{1}{n}\sum_{i=1}^{n} y_i) \\
&= \frac{1}{n^2} Var(n\beta + \sum_{i=1}^{n} u_i) \\
&= \frac{1}{n^2} \underbrace{Var(n\beta)}_{=0} + Var(\sum_{i=1}^{n} u_i) \\
&= \frac{1}{n^2}[\sum_{i=1}^{n} Var(u_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \underbrace{Cov(u_i, u_j)}_{=0}] \\
&= \frac{1}{n^2}\sum_{i=1}^{n}(\frac{1}{2} + \frac{1}{2^2} + ... + \frac{1}{2^n}) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{2^i} \\
&= \frac{1}{n^2}(1 - (\frac{1}{2})^n)
\end{aligned}$$

Use online calculators for simplifications of geometric series: symbolab calculator

Consistency

Proving that, in the limit, the estimator tends in probability to the population parameter.

Types of Convergence

It will be useful for this section to understand, to some level, three types of convergence:

(I) AS (Almost Sure)

- A sequence of random scalars $\{x_n\}$ **converges almost surely** to μ if,

$$Prob(\lim_{n\to\infty} x_n = \mu) = 1$$

(II) MS (Mean Square)

- A sequence of random scalars $\{x_n\}$ **converges in mean square** to μ if,

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

- The idea behind mean squared convergence is that if x_i (a random variable) has an expectation μ, then the sample $\{x_n\}$ converges to μ in the limit if the variance of $\{x_n\}$ tends to zero. That is if in the limit $\{x_n\}$ tends to a constant, that constant must be μ.

(III) P (Probability)

- A sequence of random scalars $\{x_n\}$ **converges in probability** to μ if, for any $\epsilon > 0$,

$$\lim_{n\to\infty} Prob(|x_n - \mu| > \epsilon) = 0$$

These types of convergence descend in strength, which means that they each imply the ones below, such that,

$$ASC \Rightarrow MSC \Rightarrow PC$$

In a question on consistency all you need to show is **convergence in probability**, hence if you can show MSC (or ASC), you get your answer for free!

Why is this useful?

Because MSC, in my opinion, is often easier to show. All we need to show it for an estimator $\hat{\beta}$ is that,

(1) $E[\hat{\beta}] = \beta$, and;

(2) $\lim_{n\to\infty} Var(\hat{\beta}) = 0$.

Especially if, in prior parts of a question, you have already found the expectation and variance of some $\hat{\beta}$, all you need to do is state these facts you've already proved and show that in the limit the variance is zero.

Example Proof 1: Mean Square Convergence

Recall from earlier that,

$$E[\hat{\beta}] = \beta \text{ and } Var(\hat{\beta}) = \frac{1}{n^2}(1 - \frac{1}{2^n})$$

Prove by showing $\lim\limits_{n\to\infty} Var(\hat{\beta}) = 0$,

$$\lim_{n\to\infty} \frac{1}{n^2} = 0$$

$$\lim_{n\to\infty} 1 - \frac{1}{2^n} = 1$$

Hence,

$$\lim_{n\to\infty} Var(\hat{\beta}) = \lim_{n\to\infty} \frac{1}{n^2}(1 - \frac{1}{2^n}) = 0$$

Which implies by MSC that,

$$\hat{\beta} \xrightarrow{P} \beta$$

Example Proof 2: Convergence in Probability

This example considers the estimator of the error variance, again from the 2021 paper. We are asked to prove the consistency $\hat{\sigma}^2$, where,

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(u_i - \bar{u})^2 - \frac{1}{n-2}\frac{\{\sum_{i=1}^{n}(x_i - \bar{x})(u_i - \bar{u})\}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \text{ and } x_i \sim iid\ N(0, \sigma_x^2)\ ,\ u_i \sim iid\ N(0, \sigma^2)$$

Consider, to begin, the first term:

Multiply by $\frac{n}{n}$

$$\frac{1}{n-2}\frac{n}{n}\sum_{i=1}^{n}(u_i - \bar{u})^2 = \frac{n}{n-2}\frac{1}{n}\sum_{i=1}^{n}(u_i^2 + \bar{u}^2 - 2u_i\bar{u})$$

Notice that,

$$\frac{1}{n}\sum_{i=1}^{n}(u_i^2 + \bar{u}^2 - 2u_i\bar{u}) = \frac{1}{n}\sum_{i=1}^{n}u_i^2 - \bar{u}^2$$

Taking limits of each term,

$$\lim_{n\to\infty} \frac{n}{n-2} = 1$$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}u_i^2 = E[u_i^2] = \sigma^2$$

$$\lim_{n\to\infty} \bar{u}^2 = \lim_{n\to\infty} (\frac{1}{n}\sum_{i=1}^{n}u_i)^2 = (E[u_i])^2 = 0$$

Overall,

$$\frac{1}{n-2}\frac{n}{n}\sum_{i=1}^{n}(u_i - \bar{u})^2 \xrightarrow{P} \sigma^2$$

Now the second term,

$$\frac{1}{n-2}\frac{n}{n}\frac{\{\sum_{i=1}^{n}(x_i-\bar{x})(u_i-\bar{u})\}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{n}{n-2}\frac{\frac{1}{n^2}\{\sum_{i=1}^{n}(x_i-\bar{x})(u_i-\bar{u})\}^2}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{n}{n-2}\frac{\{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(u_i-\bar{u})\}^2}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

Denominator,

Notice the denominator must tend to some nonzero number such that the fraction is well defined.

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2-\bar{x}^2$$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}x_i^2 = E[x_i^2] = \sigma_x^2$$

$$\lim_{n\to\infty}\bar{x}^2 = (\frac{1}{n}\sum_{i=1}^{n}x_i)^2 = (E[x_i])^2 = 0$$

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 \xrightarrow{P} \sigma_x^2$$

USEFUL to note if x_i is *not* mean zero with some specified variance (which often it isn't),

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}x_i^2 = E[x_i^2]$$

$$\lim_{n\to\infty}\bar{x}^2 = (\frac{1}{n}\sum_{i=1}^{n}x_i)^2 = (E[x_i])^2$$

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 \xrightarrow{P} E[x_i^2]-(E[x_i])^2 = Var(x_i)$$

Numerator,

$$\{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(u_i-\bar{u})\}^2 = \{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})u_i\}^2 = \{\frac{1}{n}\sum_{i=1}^{n}(x_iu_i-\bar{x}u_i)\}^2$$

$$\frac{1}{n}\sum_{i=1}^{n}(x_iu_i-\bar{x}u_i) = \frac{1}{n}\sum_{i=1}^{n}x_iu_i-\bar{x}\frac{1}{n}\sum_{i=1}^{n}u_i$$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}x_iu_i = E[x_iu_i] \underset{(x_j,u_i \text{ iid})}{=} E[x_i]E[u_i] = 0$$

$$\lim_{n\to\infty}\bar{x}\frac{1}{n}\sum_{i=1}^{n}u_i = \bar{x}E[u_i] = 0$$

$$\{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(u_i-\bar{u})\}^2 \xrightarrow{P} \{0\}^2 = 0$$

USEFUL for products of RVs,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}x_iu_i = E[x_iu_i] = E[E[x_iu_i \mid x_i]] = E[x_iE[u_i \mid x_i]] = 0$$

Overall,

$$
\begin{aligned}
\frac{1}{n-2}\frac{\{\sum_{i=1}^{n}(x_i-\bar{x})(u_i-\bar{u})\}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} &= \frac{n}{n-2}\frac{\{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(u_i-\bar{u})\}^2}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2} \\
\lim_{n\to\infty}\frac{n}{n-2} &= 1 \\
\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 &= \sigma_x^2 \\
\lim_{n\to\infty}\{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})u_i\}^2 &= 0 \\
\frac{1}{n-2}\frac{\{\sum_{i=1}^{n}(x_i-\bar{x})(u_i-\bar{u})\}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} &\xrightarrow{P} 1\times\frac{0}{\sigma_x^2} = 0
\end{aligned}
$$

Final Answer,

$$
\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(u_i-\bar{u})^2 - \frac{1}{n-2}\frac{\{\sum_{i=1}^{n}(x_i-\bar{x})(u_i-\bar{u})\}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \xrightarrow{P} \sigma^2 - 1\times\frac{0}{\sigma_x^2} = \sigma^2
$$

Asymptotic Distribution

Finding distribution in the limit – they tend to be normally distributed in some way or another!

Convergence in Distribution

Let $\{x_n\}$ be a sequence of random scalars with CDF F_n, and let X be an RV. If $F_n(x) \longrightarrow F(x)$ as $n \to \infty$ for every continuity point x, where F is the CDF of the random variable X, then x_n **converges in distribution** to the random variable X, denoted,

$$x_n \xrightarrow{D} X$$

Useful Understanding from Kevin Sheppard (2022)

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\epsilon_i - \bar{\epsilon})^2 \\
\hat{\sigma}^2 - \sigma^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(\epsilon_i - \bar{\epsilon})^2 - \sigma^2 \text{ Subtract } \sigma^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\epsilon_i^2 - \frac{n}{n-1}\bar{\epsilon}^2 - \sigma^2 \text{ Expand square} \\
&= \frac{1}{n-1}\frac{n}{n}\sum_{i=1}^{n}\epsilon_i^2 - \frac{n}{n-1}\bar{\epsilon}^2 - \sigma^2 \text{ Multiply by } \frac{n}{n} \\
&= \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2 - \frac{n}{n-1}\bar{\epsilon}^2 - \sigma^2 \text{ Reorder} \\
&= \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(\epsilon_i^2 - \sigma^2) - \frac{n}{n-1}\bar{\epsilon}^2 - \frac{1}{n-1}\sigma^2 \text{ Move } \sigma^2 \text{ inside and track extra part} \\
\sqrt{n}(\hat{\sigma}^2 - \sigma^2) &= \underbrace{\frac{n}{n-1}}_{\text{Goes to 1}}\underbrace{\left\{\sqrt{n}\frac{1}{n}\sum_{i=1}^{n}(\epsilon_i^2 - \sigma^2)\right\}}_{CLT} - \underbrace{\sqrt{n}\frac{n}{n-1}\bar{\epsilon}^2}_{\text{Goes to 0}} - \underbrace{\sqrt{n}\frac{1}{n-1}\sigma^2}_{\text{Goes to 0}} \\
\sqrt{n}\frac{n}{n-1}\bar{\epsilon}^2 &= \underbrace{\frac{n}{n-1}}_{\text{Goes to 1}}\underbrace{\left\{n^{-1/2}\sum_{i=1}^{n}\epsilon\right\}}_{CLT}\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\epsilon_i\right)}_{\text{Goes to 1}}
\end{aligned}$$

Example Proof 1:

$$\hat{\beta} = \beta + \frac{1}{n}\sum_{i=1}^{n} u_i$$

$E[\hat{\beta}] = \beta$ and $Var(\hat{\beta}) = \frac{1}{n^2}(1 - \frac{1}{2^n})$

Sum of normally distributed RVs is normally distributed, hence,

$$\hat{\beta} \sim N(\beta, \frac{1}{n^2}(1 - \frac{1}{2^n}))$$

$$\frac{\hat{\beta} - \beta}{\sqrt{\frac{1}{n^2}(1 - \frac{1}{2^n})}} = \frac{n(\hat{\beta} - \beta)}{\sqrt{(1 - \frac{1}{2^n})}} \sim N(0, 1)$$

$$n(\hat{\beta} - \beta) \sim N(0, [1 - \frac{1}{2^n}])$$

$$n(\hat{\beta} - \beta) \xrightarrow{D} N(0, 1)$$

Example Proof 2: OLS

This is a very standard proof of the asymptotic distribution of the OLS estimator and can be found in most textbooks.

$$\sqrt{n}(\hat{\beta} - \beta) = \sqrt{n}\frac{n^{-1}\sum_{i=1}^{n}(x_i - \bar{x})u_i}{n^{-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{n^{-1/2}\sum_{i=1}^{n}(x_i - \bar{x})u_i}{n^{-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Denominator,

$$n^{-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \xrightarrow{P} Var(x_i)$$

Numerator,

(Use the fact that $E[x_i] = \mu_x$)

$$\begin{aligned}
n^{-1/2}\sum_{i=1}^{n}(x_i - \bar{x})u_i &= n^{-1/2}\sum_{i=1}^{n}(x_i - \mu_x - \bar{x} + \mu_x)u_i \\
&= n^{-1/2}\sum_{i=1}^{n}(x_i - \mu_x - \bar{x} + \mu_x)u_i \\
&= n^{-1/2}\sum_{i=1}^{n}([x_i - \mu_x]u_i - [\bar{x} - \mu_x]u_i) \\
&= n^{-1/2}\sum_{i=1}^{n}(x_i - \mu_x)u_i - (\bar{x} - \mu_x)n^{-1/2}\sum_{i=1}^{n}u_i
\end{aligned}$$

Second term,

$$(\bar{x} - \mu_x)n^{-1/2}\sum_{i=1}^{n}u_i$$

$$\lim_{n\to\infty}(\bar{x} - \mu_x) = \left\{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}x_i\right\} - \mu_x = 0$$

$$n^{-1/2}\sum_{i=1}^{n}u_i \xrightarrow{D} N(0, \sigma^2)$$

$$(\bar{x} - \mu_x)n^{-1/2}\sum_{i=1}^{n}u_i \xrightarrow{P} 0$$

First term,

$$
\begin{aligned}
&n^{-1/2}\sum_{i=1}^{n}(x_i-\mu_x)u_i \\
&\qquad E[(x_i-\mu_x)u_i]=0 \\
&\qquad Var((x_i-\mu_x)u_i)=E[(x_i-\mu_x)^2u_i^2]=E[(x_i-\mu_x)^2E[u_i^2\mid x_i]]=\sigma^2Var(x_i) \\
&n^{-1/2}\sum_{i=1}^{n}(x_i-\mu_x)u_i \xrightarrow{D} N(0,\sigma^2Var(x_i))
\end{aligned}
$$

Overall,

$$\sqrt{n}(\hat{\beta}-\beta)=\frac{n^{-1/2}\sum_{i=1}^{n}(x_i-\bar{x})u_i}{n^{-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}\xrightarrow{D}\frac{N(0,\sigma^2Var(x_i))}{Var(x_i)}=N(0,\sigma^2Var(x_i)^{-1})$$

$$\sqrt{n}(\hat{\beta}-\beta)\xrightarrow{D}N(0,\sigma^2Var(x_i)^{-1})$$

Asymptotic Normally of t-statistic

Just generally a t-statistic is the estimator minus the mean divided by its standard error.

Also note that in finite samples do not use t_∞, rather use t_{n-k} where k is the number of regressors.

Example Proof: OLS

$$H_0 : \beta_2 = b$$

$$t_{\beta_2} = \frac{\hat{\beta}_2 - b}{s.e.(\hat{\beta}_2)} = \frac{\hat{\beta}_2 - b}{\sqrt{\frac{\hat{\sigma}_u^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} = \frac{\sqrt{n}(\hat{\beta}_2 - b)}{\sqrt{n}\sqrt{\frac{\hat{\sigma}_u^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} = \frac{\hat{\beta}_2 - b}{\sqrt{\frac{\hat{\sigma}_u^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} = \frac{\sqrt{n}(\hat{\beta}_2 - b)}{\sqrt{\frac{\hat{\sigma}_u^2}{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2}}}$$

$$\sqrt{n}(\hat{\beta}_2 - b) \xrightarrow{D} N(0, \sigma_u^2 Var(x_i)^{-1})$$

$$\hat{\sigma}_u^2 \xrightarrow{P} \sigma_u^2$$

$$\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2 \xrightarrow{P} Var(x_i)$$

$$t_{\beta_2} = \frac{\sqrt{n}(\hat{\beta}_2 - b)}{\sqrt{\frac{\hat{\sigma}_u^2}{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2}}} \xrightarrow{D} \frac{N(0, \sigma_u^2 Var(x_i)^{-1})}{\sigma_u \sqrt{Var(x_i)^{-1}}} = N(0,1)$$

Matrix OLS Algebraic Facts

Model

$Y = X\beta + U$

Where,

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \ X = \begin{pmatrix} x_{11} & ... & x_{k1} \\ \vdots & \ddots & \vdots \\ x_{1n} & \vdots & x_{kn} \end{pmatrix}, \ \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \ U = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$$

Or equivalently,

$y_i = \beta_1 x_{11} + \beta_2 x_{2i} + ... + \beta_k x_{ki} + u_i \ \ \forall i = 1, 2, ..., n$

OLS Problem & FOCs

$$\begin{aligned} argmin_{\beta \in \mathbb{R}^k} \ U(\beta)'U(\beta) &= argmin_{\beta \in \mathbb{R}^k} \ (Y - X\beta)'(Y - X\beta) \\ &= argmin_{\beta \in \mathbb{R}^k} \ (Y' - \beta' X')(Y - X\beta) \\ &= argmin_{\beta \in \mathbb{R}^k} \ Y'Y - \beta' X' Y - Y' X \beta + \beta' X' X \beta \\ \text{Note that } \beta' X' Y &= Y' X \beta \text{ therefore,} \\ &= argmin_{\beta \in \mathbb{R}^k} \ Y'Y - 2Y'X\beta + \beta' X' X \beta \end{aligned}$$

$$\frac{\partial U(\beta)'U(\beta)}{\partial \beta} = 2X'Y + 2X'X\hat{\beta} = 0$$

Here we have used the rules: $\dfrac{\partial (Ax)}{\partial x} = A'$ and $\dfrac{\partial (x'Ax)}{\partial x} = (A + A')x$

Solutions

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

Predicted Values

$\hat{Y} = X\hat{\beta}$

Therefore,

$\hat{U} = Y - \hat{Y}$

Residual

$$\hat{U} = Y - X\hat{\beta}$$

Hence, by construction (FOCs),

$$X'Y + X'X\hat{\beta} = 0$$

$$X'(Y + X\hat{\beta}) = 0$$

$$X'\hat{U} = 0$$

Other Useful Facts

Projection matrix,

$P_{n\times n} = X(X'X)^{-1}X'$

Symmetric $P = P'$ and idempotent $P = P^2$

$PX = X$

Annihilator matrix,

$M_{n\times n} = I_n - P$

Symmetric $M = M'$ and idempotent $M = M^2$

$MX = 0$

Matrix Proofs

Find OLS Estimator

$$argmin_{\beta\in\mathbb{R}^k}\ U(\beta)'U(\beta) = argmin_{\beta\in\mathbb{R}^k}\ (Y - X\beta)'(Y - X\beta)$$

Use the rules: $\dfrac{\partial(Ax)}{\partial x} = A'$ and $\dfrac{\partial(x'Ax)}{\partial x} = (A + A')x$ where A is symmetric

Unbiased

Take expectations, looking to show that the expectation of the estimator is the population parameter.

Example Proof: Estimator of the Error Variance

$$\begin{aligned}
\hat{\sigma}^2_U &= \frac{\hat{U}'\hat{U}}{n-k} \\
\text{Where,}& \\
\hat{U} &= Y - X\hat{\beta} = (I_n + X(X'X)^{-1}X')Y = MY \\
&= MY = M(X\beta + U) = MU \text{ since } MX = 0 \\
\text{So then,}& \\
\hat{\sigma}^2_U &= \frac{U'M'MU}{n-k} = \frac{U'MU}{n-k} \text{ recall M is idempotent and symmetric} \\
\text{Where,}& \\
\hat{U}'\hat{U} &= U'MU = tr(U'MU) = tr(MUU') \\
E[\hat{U}'\hat{U} \mid X] &= E[tr(MUU') \mid X] = tr(ME[UU' \mid X]) = \sigma^2_U \\
tr(M) &= tr(I_n - X(X'X)^{-1}X') = tr(I_n) - tr((X'X)^{-1}X'X) = tr(I_n) - tr(I_k) = n - k \\
\text{So,}& \\
E[\hat{U}'\hat{U}] &= E[E[\hat{U}'\hat{U} \mid X]] = \sigma^2_U(n-k) \\
\text{Overall,}& \\
E[\hat{\sigma}^2_U] &= \frac{E[\hat{U}'\hat{U}]}{n-k} = \frac{\sigma^2_U(n-k)}{n-k} = \sigma^2_U
\end{aligned}$$

Conditional Variance

Take variance or use expectations, both proofs are given below.

Example Proof: OLS Estimator Method 1 - Variance

$$\begin{aligned}
Var(\hat{\beta} \mid X) &= Var(\hat{\beta} - \beta \mid X) \text{ since } \beta \text{ is not random} \\
&= Var((X'X)^{-1}X'U \mid X) \\
&= (X'X)^{-1}X'Var(U \mid X)\{(X'X)^{-1}X'\}' \\
&= (X'X)^{-1}X'E[UU' \mid X]X(X'X)^{-1} \\
&= (X'X)^{-1}X'X(X'X)^{-1}\sigma_U^2 I_n \\
&= (X'X)^{-1}\sigma_U^2 I_n
\end{aligned}$$

$$Var(\hat{\beta} \mid X) = \sigma_U^2(X'X)^{-1}$$

Example Proof: OLS Estimator Method 2 - Expectations

$$\begin{aligned}
Var(\hat{\beta} \mid X) &= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)' \mid X] \\
&= E[(X'X)^{-1}X'U\{(X'X)^{-1}X'U\}' \mid X] \\
&= E[(X'X)^{-1}X'UU'X(X'X)^{-1} \mid X] \\
&= (X'X)^{-1}X'E[UU' \mid X]X(X'X)^{-1} \\
&= (X'X)^{-1}X'X(X'X)^{-1}E[UU' \mid X] \\
&= (X'X)^{-1}E[UU' \mid X]
\end{aligned}$$

$$Var(\hat{\beta} \mid X) = \sigma_U^2(X'X)^{-1}$$

Consistency

Proving that, in the limit, the estimator tends in probability to the population parameter.

Useful Fact

$$Q = \underset{n\to\infty}{p\lim}\frac{X'X}{n} = E[X_i X_i']$$

Example Proof: OLS Estimator

$$\hat{\beta} = \beta + (X'X)^{-1}X'U = \beta + \left(\frac{X'X}{n}\right)^{-1}\frac{X'U}{n}$$

$$\left(\frac{X'X}{n}\right)^{-1} \xrightarrow{P} Q^{-1}$$

$$\frac{X'U}{n} \xrightarrow{P} 0$$

Where,

$$\frac{X'U}{n} = \begin{pmatrix} n^{-1}\sum_{i=1}^{n} x_{1i}u_i \\ n^{-1}\sum_{i=1}^{n} x_{2i}u_i \\ \vdots \\ n^{-1}\sum_{i=1}^{n} x_{ki}u_i \end{pmatrix} = n^{-1}\sum_{i=1}^{n} X_i u_i$$

$$E[X_i u_i] = E[X_i E[u_i \mid X_i] = 0$$

$$Var(X_i u_i) = E[X_i X_i E[u_i^2 \mid X_i] = \sigma_U^2 E(X_i X_i) < \infty$$

$$\hat{\beta} \xrightarrow{P} \beta + Q^{-1} \times 0 = \beta$$

Asymptotic Normality

These are, in my opinion, the hardest proofs

Example Proof: OLS Estimator

$$\sqrt{n}(\widehat{\beta}-\beta)=\left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'U}{\sqrt{n}}\right)$$

$$\left(\frac{X'X}{n}\right)^{-1}\xrightarrow{P}Q^{-1}$$

$$\frac{X'U}{\sqrt{n}}=\begin{pmatrix} n^{-1/2}\sum_{i=1}^{n}x_{1i}u_i \\ n^{-1/2}\sum_{i=1}^{n}x_{2i}u_i \\ \vdots \\ n^{-1/2}\sum_{i=1}^{n}x_{Ki}u_i \end{pmatrix}=n^{-1/2}\sum_{i=1}^{n}X_iu_i \text{ where } X_i=\begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ni} \end{pmatrix}$$

If we can show that the assumptions of the multivariate Lindeberg-Levy CLT hold for $X_iu_i=Z_i$, assuming that X_iu_i is iid, then we can apply it,

$$E\left(X_iu_i\right)=E\left[E\left(X_iu_i\mid X\right)\right]=E\left[X_iE\left(u_i\mid x\right)\right]=0$$

$$\operatorname{Var}\left(X_iu_i\right)=E\left[X_iu_iu_iX_i'\right]=E\left[{u_i}^2X_iX_i'\right]=E\left[X_iX_i'E\left({u_i}^2\mid X\right)\right]={\sigma_U}^2E\left(X_iX_i'\right)=\sigma_U^2Q$$

Given this, then,

$$\left(\frac{X'X}{n}\right)^{-1}\xrightarrow{P}Q^{-1}$$

$$\left(\frac{X'U}{\sqrt{n}}\right)\xrightarrow{D}N\left\{\mathbf{0}_K\ ,\ \sigma_U^2Q\right\}$$

Meaning, overall,

$$\sqrt{n}(\widehat{\beta}-\beta)\xrightarrow{D}N\left\{\mathbf{0}_K\ ,\ Q^{-1}\sigma_U^2QQ^{-1}\right\}$$

$$\sqrt{n}(\widehat{\beta}-\beta)\xrightarrow{D}N\left\{\mathbf{0}_K,\sigma_U^2Q^{-1}\right\}$$

$$\hat{\beta}\sim N\left(\beta,\frac{\sigma_U^2}{n}Q^{-1}\right)$$

Asymptotic Normality of t-statistic

Example Proof: OLS Estimator

$$\frac{\widehat{\beta}_k-\beta_k^0}{s.e.\left(\widehat{\beta}_k\right)}\ ,\ s.e.\left(\widehat{\beta}_k\right)=\sqrt{\hat{\sigma}_U^2\left(X'X\right)_{kk}^{-1}}$$

$$\frac{\widehat{\beta}_k-\beta_k^0}{s.e.\left(\widehat{\beta}_k\right)}=\frac{\sqrt{n}(\widehat{\beta}_k-\beta_k^0)}{\sqrt{n\hat{\sigma}_U^2(X'X)_{kk}^{-1}}}=\frac{\sqrt{n}(\widehat{\beta}_k-\beta_k^0)}{\sqrt{\hat{\sigma}_U^2\frac{(X'X)_{kk}^{-1}}{n}}}$$

$$\sqrt{n}\left(\widehat{\beta}_k-\beta_k^0\right)\xrightarrow{D}N\left\{0\ ,\ \sigma_U^2Q_{kk}^{-1}\right\}\quad,\quad\left(\frac{X'X}{n}\right)_{kk}^{-1}\xrightarrow{P}Q_{kk}^{-1}\ \ ;,\ \widehat{\sigma}_U{}^2\xrightarrow{P}\sigma_U^2$$

$$\frac{\widehat{\beta}_k-\beta_k^0}{s.e.\left(\widehat{\beta}_k\right)}=\frac{\sqrt{n}\left(\widehat{\beta}_k-\beta_k^0\right)}{\sqrt{\widehat{\sigma}_U^2\left(\frac{X'X}{n}\right)_{kk}^{-1}}}\xrightarrow{D}\frac{N\left\{0,\sigma_U^2Q_{kk}^{-1}\right\}}{\sqrt{\sigma_U^2Q_{kk}^{-1}}}=N(0,1)$$

Multiple Hypothesis Testing: The F-test

Null Hypothesis

Testing j null hypotheses,

$$\begin{gathered} r_{11}\beta_1 + r_{12}\beta_2 + ... + r_{1k}\beta_k = q_1 \\ r_{21}\beta_1 + r_{22}\beta_2 + ... + r_{2k}\beta_k = q_2 \\ \vdots \\ r_{j1}\beta_1 + r_{j2}\beta_2 + ... + r_{jk}\beta_k = q_j \end{gathered}$$

Which we simplify to,

$$R = \begin{pmatrix} r_{11} & r_{12} & ... & r_{1k} \\ r_{21} & r_{22} & ... & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{j1} & r_{j2} & ... & r_{jk} \end{pmatrix}, \quad q = \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_j \end{pmatrix}$$

Giving,

$$H_0 : R\beta = q$$

For example,

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + u_i$$

Where,

$$H_0 : \beta_2 + \beta_3 = 1 \text{ and } \beta_4 = \beta_5$$

Hence $j = 2$, and,

$$R = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix} \text{ and } q = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Where,

$$H_0 : R\beta = q$$

For completeness recall that,

$$R\beta = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = \begin{pmatrix} 0\beta_1 + 1\beta_2 + 1\beta_3 + 0\beta 4 + 0\beta_5 \\ 0\beta_1 + 0\beta_2 + 0\beta_3 + 1\beta_4 - 1\beta_5 \end{pmatrix} = \begin{pmatrix} \beta_2 + \beta_3 \\ \beta_4 - \beta_5 \end{pmatrix}$$

And so we get the same null hypothesis, that is,

$$R\beta = \begin{pmatrix} \beta_2 + \beta_3 \\ \beta_4 - \beta_5 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = q$$

Or rather that,

$$\beta_2 + \beta_3 = 1 \text{ and } \beta_4 = \beta_5$$

Test Statistic (Assuming Normality)

Assumptions,

- Full rank (no perfect multicollinearity)
  - X is and $n \times k$ matrix with rank k
- (Conditionally) normal errors

Test Statistic,

$$F = \frac{(R\hat{\beta} - q)'\{\hat{\sigma}_U^2 R(X'X)^{-1}R'\}^{-1}(R\hat{\beta} - q)}{j}$$

$$F \sim F_{j,n-k}$$

Test Statistic (Asymptotic)

Assumptions,

- $\{X, u\}$ is iid
- Mean independence,
  - $E[U \mid X] = 0$
- Q has full rank k
  - $Q = \underset{n\to\infty}{p\lim} \frac{X'X}{n} = E[X_i X_i']$ has rank k
- Homoskedasticity and no autocorrelation
  - $E[UU' \mid X] = cI_n$ where c is a constant

Test Statistic,

$$W = (R\hat{\beta} - q)'\{\hat{\sigma}_U^2 R(X'X)^{-1}R'\}^{-1}(R\hat{\beta} - q) = jF \xrightarrow{D} \mathcal{X}_j^2$$

Gauss-Markov Theorem

Scalar GMT

Assumptions

1. $E[u_i \mid x_1, ..., x_n] = 0$
2. $Var(u_i \mid x_1, ..., x_n) = \sigma_u^2 < \infty$
3. $E[u_i u_j \mid x_1, ..., x_n] = 0$ for $i \neq j$

Theorem

OLS estimators $\hat{\beta}_1, \hat{\beta}_2$ given these properties are BLUE (Best Linear (conditionally) Unbiased Estimators). That is they have the lowest variance (are the most efficient) out of all other possible estimators.

Start of the Proof

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} y_i$$
$$\hat{\beta}_2 = \sum_{i=1}^n w_i y_i \text{ where } w_i = \frac{\sum_{i=1}^n (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Now we will consider some other estimator $\tilde{\beta}_2$ where $\tilde{\beta}_2 = \sum_{i=1}^n c_i y_i$ with $c_i = w_i + d$ where d is some non-zero difference between c_i and w_i.

From here we need to,

1. Show the conditions for it to be such that $\tilde{\beta}_2$ is unbiased.
2. Calculate the variance of $\tilde{\beta}_2$.
3. Show that $Var(\tilde{\beta}_2 \mid x_1, ..., x_n) \geq Var(\hat{\beta}_2 \mid x_1, ..., x_n)$.

Matrix GMT

Assumptions

1. No perfect multicollinearity: X is an $n \times k$ martic with rank k (k columns).
   - This is important since $\hat{\beta} = (X'X)^{-1}X'Y$ and $Var(\hat{\beta}) = \sigma^2(X'X)^{-1}$, and multicollineary implies that $(X'X)$ is singular (determinant is zero), hence it's inverse doesn't exits. So OLS fails.
   - Notice that even if there is no perfect multicollinearity, but instead a close linear relationship among the predictors then $(X'X)^{-1}$ will be huge, hence variance will blow up. We therefore would be unable to say anything precise about coefficients or do any good inference.
2. Mean independence: $E[U \mid X] = 0$.
3. No heteroskedascity/autocorrelation: $E[UU' \mid X] = \sigma_n^2 I_n$.

Theorem

OLS estimator $\hat{\beta}$, given these properties, is BLUE (Best Linear (conditionally) Unbiased Estimator).

Start of the Proof

$\hat{\beta} = AY$, where $A = (X'X'^{-1}X'$.

Suppose another linear unbiased estimator $\tilde{\beta} = CY$, where $C = A + D$ and C is a function of X.

$Var(\tilde{\beta} \mid X) = CC'\sigma^2$

$Var(\tilde{\beta} \mid X) = \sigma^2(X'X)^{-1} + \sigma_u^2 DD'$

$Var(\tilde{\beta} \mid X) = Var(\hat{\beta} \mid X) + \sigma_u^2 DD'$

Then show that DD' is positive semi-definite to get the appropriate inequality. (A simple proof that for some matrix A, AA' is PSD from online should suffice.)

Heteroskedasticity

Suppose that we allow,

$$E[UU' \mid X] = \sigma_U^2 \Omega = \Sigma$$

$$\sigma_U^2 \Omega = \begin{pmatrix} \sigma_1^2(X) & 0 & ... & 0 \\ 0 & \sigma_2^2(X) & ... & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & ... & \sigma_n^2(X) \end{pmatrix}$$

(We have assumed heteroskedasticity but still no autocorrelation here.)

Consequences (finite sample)

- Unbiased
- Different variance: $Var(\hat{\beta} \mid X) = \sigma_U^2 (X'X)^{-1} X'\Omega X (X'X)^{-1}$
- No longer BLUE (not efficient)
- $\hat{\sigma}_U^2$ is biased.
- Inference using $\hat{\sigma}^2 (X'X)^{-1}$ may be misleading.

Consequences (large sample)

- $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{D} N(0, \sigma_U^2 Q^{-1} Q^* Q^{-1})$
- $AVar(\hat{\beta}) = \frac{\sigma_U^2}{n} Q^{-1} Q^* Q^{-1}$
- Therefore it is consistent and asymptotically normal, but with new variance
- Where, $Q^* = p \lim_{n \longrightarrow \infty} \frac{X'\Omega X}{n}$ and $Q = p \lim_{n \longrightarrow \infty} \left(\frac{X'X}{n}\right) = E[X_i X_i']$

Testing: White's Test

(1) Estimate the Model by OLS and compute OLS residuals
   - $\hat{u}_i = y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i}$

(2) Run the auxiliary regression
   - $\hat{u}_i^2 = \alpha_1 + \alpha_2 x_{2i} + \alpha_3 x_{3i} + \gamma_2 x_{2i}^2 + \gamma_3 x_{3i}^2 + \gamma_1 x_{2i} x_{3i} + e_i$
   - Here essentially we are asking if the variability of the OLS residual can be explained by the regressors, if it can then there is heteroskedasticity.

(3) Test the null (homoscedastic) hypothesis,
   - $H_0 \ : \ \alpha_2 = \alpha_3 = \gamma_1 = \gamma_2 = \gamma_3 = 0$
   - $nR^2 \overset{a}{\sim} \chi_{p-1}$
   - n = sample size, R^2 = coefficient of determination (of auxiliary regression), p = number of parameters of the auxiliary regression (in our case $p = 6$, so $p - 1 = 5$)

To run an auxiliary regression one needs all the cross products, so for a bigger model this would be more like:

$$\hat{u}_i^2 = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{2i} + \alpha_3 x_{3i} + \gamma_1 x_{1i}^2 + \gamma_2 x_{2i}^2 + \gamma_3 x_{3i}^2 + \delta_1 x_{1i} x_{2i} + \delta_1 x_{1i} x_{3i} + \delta_1 x_{2i} x_{3i} + e_i$$
$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \gamma_1 = \gamma_2 = \gamma_3 = \delta_1 = \delta_2 = \delta_3$$

Dealing with it: Heteroskedastic Robust Standard Errors

$$AVar(\hat{\beta}) = \frac{\sigma_U^2}{n} Q^{-1} Q^{*} Q^{-1}$$
$$AV\hat{a}r(\hat{\beta}) = \frac{1}{n}\left(\frac{X'X}{n}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i X_i X_i'\right)\left(\frac{X'X}{n}\right)^{-1}$$

Instrumental Variables

Problem: Endogeneity

Caused by,

- Measurement errors,
- Omitted Variables,
- Simultaneity.

Means that,

- $E[U \mid X] \neq 0$ (X tells us something about U)
- Hence OLS estimator is,
  - Biased,
  - Inconsistent.

Solution (L=K)

This is the solution when Z is an $n \times L$ matrix, X is $n \times K$ AND $L = K$.

Conditions

(1) Exogeneity - The instruments are uncorrelated with the error term.

(2) Relevance - The instruments are correlated with the endogenous independent variable.

- Relevance implies the invertibility of the matrix $Z'X$.

IV Estimator

Suppose that x_{3i} is exogenous, but we have an instrument for it, IV_i. Therefore let's replace,

$$X_i = \begin{pmatrix} 1 \\ x_{2i} \\ x_{3i} \end{pmatrix}$$

with,

$$Z_i = \begin{pmatrix} 1 \\ x_{2i} \\ IV_i \end{pmatrix}$$

Exploiting the exogeneity assumption we know that $E[Z_i u_i] = 0$ and that $u_i = y_i - X_i'\beta$.

This means that,

$$E[Z_i(y_i - X_i'\beta)] = 0$$
$$E[Z_i y_i] = \beta E[Z_i X_i']$$
$$\beta = \left\{E[Z_i X_i']\right\}^{-1} E[Z_i y_i]$$

Giving,

$$\hat{\beta}_{IV} = (Z'X)^{-1}(Z'Y)$$

Asymptotics

Assumptions,

(i) (X_i', Z_i', u_i') are iid with finite fourth moments.

(ii) Exogeneity: $E[U \mid Z] = 0$

(iii) Relevance: $Q_{ZX} = p\lim_{n\longrightarrow\infty} n^{-1}Z'X$ is a finite $L \times X$ matrix with rank K.

(iv) $Q_{XX} = p\lim_{n\longrightarrow\infty} n^{-1}X'X$ is a finite positive definite matrix with rank K.

(v) $Q_{ZZ} = p\lim_{n\longrightarrow\infty} n^{-1}Z'Z$ is a finite positive definite matrix with rank L.

(vi) $E[UU' \mid Z] = \sigma_U^2 I_n$

Results,

$$\begin{aligned}
\sqrt{n}(\hat{\beta}_{IV} - \beta) &\xrightarrow{D} N(0, \sigma_U^2 Q_{ZX}^{-1} Q_{ZZ} Q_{XZ}^{-1}) \\
AVar(\hat{\beta}_{IV}) &= \frac{\sigma_U^2}{n} Q_{ZX}^{-1} Q_{ZZ} Q_{XZ}^{-1}
\end{aligned}$$

$$\begin{aligned}
AV\hat{a}r(\hat{\beta}_{IV}) &= \hat{\sigma}_{U_{IV}}^2 (Z'X)^{-1}(Z'Z)(X'Z)^{-1} \\
\hat{\sigma}_{U_{IV}}^2 &= \frac{\hat{U}_{IV}'\hat{U}_{IV}}{n}
\end{aligned}$$

Solution: 2SLS (L>K)

Now $Z'X$ is not invertible, since it is not a square matrix (it's an $L \times K$ matrix).

$$\begin{aligned}
Y &= X\beta + U \\
X &= Z\alpha + V
\end{aligned}$$

$$\begin{aligned}
\hat{\alpha} &= (Z'Z)^{-1}Z'X \\
\hat{X} &= Z\hat{\alpha} = Z(Z'Z)^{-1}Z'X
\end{aligned}$$

$$\hat{\beta}_{2SLS} = (\hat{X}'\hat{X})^{-1}(\hat{X}'Y) = \left\{X'Z(Z'Z)^{-1}Z'X\right\}^{-1}\left\{X'Z(Z'Z)^{-1}Z'Y\right\}$$

This is most efficient IV estimator

Using same assumptions,

$$\begin{aligned}
\sqrt{n}(\hat{\beta}_{2SLS} - \beta) &\xrightarrow{D} N(0, \sigma_U^2 (Q_{XZ} Q_{ZZ}^{-1} Q_{ZX})^{-1}) \\
AVar(\hat{\beta}_{2SLS}) &= \frac{\sigma_U^2}{n} (Q_{XZ} Q_{ZZ}^{-1} Q_{ZX})^{-1}
\end{aligned}$$

$$\begin{aligned}
AV\hat{a}r(\hat{\beta}_{2SLS}) &= \hat{\sigma}_{U_{2SLS}}^2 \left\{(X'Z)(Z'Z)^{-1}(Z'X)\right\}^{-1} \\
\hat{\sigma}_{U_{2SLS}}^2 &= \frac{\hat{U}_{2SLS}'\hat{U}_{2SLS}}{n}
\end{aligned}$$

Non-Linear in Variables

You only **can't use OLS** if a model is non-linear in *parameters*. Therefore,

(A) $y_i = \theta g(x_i) + u_i$ - θ can be estimated by OLS.

- Example: $y_i = Ak_i^{\theta} e^{u_i}$ - take logs then estimate θ by OLS.

(B) $y_i = g(x_i, \theta) + u_i$ - θ cannot be estimated by OLS.

- Example: $y_i = Ak_i^{\theta} + u_i$ - cannot estimate θ by OLS.

Polynomials

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + ... + \beta_r X^r + u$$

We can test how many polynomials we need by using the F-test to test the null of linearity against the r^{th} degree polynomial. The r at which we reject the null is how many we need,

$$H_0 : \beta_2 = ... = \beta_r = 0$$

Suppose $r = 2$ For small changes the causal effect of X on Y is given by,

$$\frac{\partial Y}{\partial X} = \beta_1 + 2\beta_2 X$$

Logarithms

(1) Linear-Log:

$$Y = \beta_0 + \beta_1 logX + u$$
$$\Delta Y = \beta_1 \Delta logX \approx \beta_1 \frac{\Delta X}{X}$$

"A 1% increase in X has a $0.01 \times \beta_1$ effect on Y"

(2) Log-Linear

$$logY = \beta_0 + \beta_1 X + u$$
$$\Delta logY = \beta_1 \Delta X \Rightarrow \frac{\Delta Y}{Y} = \beta_1 \Delta X$$

"A unit increase of X increases Y by $\beta_1 \times 100\%$". Neat rule: $(e^{\hat{\beta}_1} - 1) \times 100$ gives the percentage change in Y given a change in X.

(3) Log-Log

$$logY = \beta_0 + \beta_1 logX + u$$
$$\Delta logY = \beta_1 \Delta logX \Rightarrow \frac{\Delta Y}{Y} = \beta_1 \frac{\Delta X}{X} \Rightarrow \beta_1 = \frac{\Delta Y/Y}{\Delta X/X}$$

"β_1 is the elasticity of Y wrt to X"

Dummies

(1) Constant

$$Y = \beta_0 + \beta_1 X + \beta_2 D + u$$

If $D = 1$ then the constant is $\beta_0 + \beta_2$, or β_0 when $D = 0$.

(2) Slope

$$Y = \beta_0 + \beta_1 X + \beta_3 X \cdot D + u$$

Where $\frac{\partial Y}{\partial X} = \beta_1 + \beta_3 D = \begin{cases} \beta_1 \ , \ D = 0 \\ \beta_1 + \beta_3 \ , \ D = 1 \end{cases}$.

(3) Constant & Slope

$$Y = \beta_0 + \beta_1 X + \beta_2 D + \beta_3 X \cdot D + u$$

In this case the interaction term affects both the constant and the slope.

Non-Linear in Parameters

We **can't use OLS** to estimate coefficients when a model is non-linear in *parameters*. Below are examples of such models.

Logistic Curve

$$y_i = \frac{1}{1 + e^{-(\beta_1 + \beta_2 x_i)}} + u_i$$

Box-Cox Transformation

$$y_i = \alpha + \beta \frac{x_i^\lambda - 1}{\lambda} + u_i$$

When,

$$\begin{aligned} \lambda = 1 &: y_i = \alpha^* + \beta x_i + u_i \\ \lambda = -1 &: y_i = \tilde{\alpha} + \tilde{\beta}\frac{1}{x_i} + u_i \\ \lambda = 0 &: y_i = \alpha + ln x_i + u_i \end{aligned}$$

CES Production Function

$$\begin{aligned} y &= \gamma \left\{ \delta k^{-\rho} + (1-\delta) l^{-\rho} \right\}^{-\frac{v}{\rho}} e^{\epsilon} \\ lny &= ln\gamma - \frac{v}{\rho} ln(\delta k^{-\rho} + (1-\delta) l^{-\rho}) + \epsilon \end{aligned}$$

Coefficient Interpretation

Consider $y = g(x, \theta)$ with $E[y \mid x] = g(x, \theta)$.

- Marginal effect: $\frac{\partial E[y \mid x]}{\partial x}$
- Response of average individual: $\left.\frac{\partial E[y \mid x]}{\partial x}\right|_{\bar{x}}$
- Response of individual with $x = x*$: $\left.\frac{\partial E[y \mid x]}{\partial x}\right|_{x^*}$
- Average response of all individuals: $\frac{1}{n}\sum_{i=1}^{n} \frac{\partial E[y \mid x]}{\partial x_i}$

Non-Linear Least Squares (NLLS)

NLLS

Model,

$$y_i = g(x_i, \theta) + u_i$$

Where θ is a $k \times 1$ vector of unknown parameters.

Problem,

$$\hat{\theta} = \underset{\theta \in \Theta}{argmin}\ Q_n(\theta)$$

$$Q_n(\theta) = \sum_{i=1}^{n} \{y_i - g(x_i, \theta)\}^2$$

$$\hat{\theta} = \underset{\theta \in \Theta}{argmin}\ \sum_{i=1}^{n} \{y_i - g(x_i, \theta)\}^2$$

FOC,

$$\dot{Q}_n(\theta) = \frac{\partial Q_n(\theta)}{\partial \theta} = -2 \sum_{i=1}^{n} \Big\{ y_i - g(x_i, \theta) \Big\} \dot{g}(x_i, \theta) = 0$$

$$\text{Where } \dot{g}(x_i, \theta) = \frac{\partial g(x_i, \theta)}{\partial \theta}$$

Gauss-Newton Preliminaries

Taylor's Theorem

An m^{th} order taylor approximation of $f(x)$ about x_0 and for some c between x and x_0.

$$f(x) \approx f(x_0) + \frac{\dot{f}(x_0)}{1!}(x - x_0) + \frac{\ddot{f}(x_0)}{2!}(x - x_0)^2 + ... + \frac{f^(m)(x_0)}{m!}(x - x_0)^m + \frac{f^{(m+1)}(c)}{(m+1)!}(x - x_0)^{(m+1)}$$

When the remainder is small then,

$$f(x) \approx f(x_0) + \frac{\dot{f}(x_0)}{1!}(x - x_0) + \frac{\ddot{f}(x_0)}{2!}(x - x_0)^2 + ... + \frac{f^{(m)}(x_0)}{m!}(x - x_0)^m$$

Example Taylor Approximation

First order taylor approximation of $f(x) = ln(1 + x)$ about $x_0 = 0$.

Taylor's Theorem for first order approximation

$$f(x) \approx f(x_0) + \frac{\dot{f}(x_0)}{1!}(x - x_0)$$

Hence

$$ln(1 + x) \approx ln(1 + 0) + \frac{1}{1!}\frac{1}{1 + 0}$$

$$ln(1 + x) \approx x$$

This is true when the remainder is small, which is the case when x is small since the remainder is given by,

$$R_m(x) = \frac{1}{2!}\Bigg\{ -\frac{1}{(1 + c)^2}x^2 \Bigg\} = -\frac{x^2}{2(1 + c)^2}$$

And when x is small x^2 is very small.

Gauss-Newton Method

Consider,

$$y_i = g(x_i, \theta) + u_i$$

Where x_i and θ are scalars (for simplicity).

Begin with a first order taylor approximation about some initial estimate of θ, call it $\hat{\theta}_{(1)}$.

$$g(x_i, \theta) \approx g(x_i, \hat{\theta}_{(1)}) + \dot{g}(x_i, \hat{\theta}_{(1)})(\theta - \hat{\theta}_{(1)})$$

Which implies that,

$$y_i \approx g(x_i, \hat{\theta}_{(1)}) + \dot{g}(x_i, \hat{\theta}_{(1)})(\theta - \hat{\theta}_{(1)}) + u_i$$

Now we can use NLLS since *theta* enters linearly!

$$\hat{\theta}_{(2)} = \underset{\theta \in \Theta}{argmin}\ \tilde{Q}_n(\theta)$$

$$\hat{\theta}_{(2)} = \underset{\theta \in \Theta}{argmin}\ \sum_{i=1}^{n} \{y_i - g(x_i, \hat{\theta}_{(1)}) - \dot{g}(x_i, \hat{\theta}_{(1)})(\theta - \hat{\theta}_{(1)})\}^2$$

FOC,

$$\frac{\partial \tilde{Q}_n(\theta)}{\partial \theta} = -2 \sum_{i=1}^{n} \Big\{ y_i - g(x_i, \hat{\theta}_{(1)}) - \dot{g}(x_i, \hat{\theta}_{(1)})(\hat{\theta}_{(2)} - \hat{\theta}_{(1)}) \Big\} \dot{g}(x_i, \hat{\theta}_{(1)}) = 0$$

Notice now that $\hat{\theta}_{(2)}$ is our new estimator of θ.

Simplifying the FOC,

$$0 = -2 \sum_{i=1}^{n} \Big\{ y_i - g(x_i, \hat{\theta}_{(1)}) - \dot{g}(x_i, \hat{\theta}_{(1)})(\hat{\theta}_{(2)} - \hat{\theta}_{(1)}) \Big\} \dot{g}(x_i, \hat{\theta}_{(1)})$$

$$0 = \sum_{i=1}^{n} \Big\{ y_i - g(x_i, \hat{\theta}_{(1)}) - \dot{g}(x_i, \hat{\theta}_{(1)})(\hat{\theta}_{(2)} - \hat{\theta}_{(1)}) \Big\} \dot{g}(x_i, \hat{\theta}_{(1)})$$

$$0 = \sum_{i=1}^{n} \Big\{ y_i - g(x_i, \hat{\theta}_{(1)}) \Big\} \dot{g}(x_i, \hat{\theta}_{(1)}) - \hat{\theta}_{(2)} \sum_{i=1}^{n} \dot{g}(x_i, \hat{\theta}_{(1)})^2 + \hat{\theta}_{(1)} \sum_{i=1}^{n} \dot{g}(x_i, \hat{\theta}_{(1)})^2$$

$$\hat{\theta}_{(2)} = \hat{\theta}_{(1)} \frac{\sum_{i=1}^{n} \dot{g}(x_i, \hat{\theta}_{(1)})^2}{\sum_{i=1}^{n} \dot{g}(x_i, \hat{\theta}_{(1)})^2} + \frac{\sum_{i=1}^{n} \Big\{ y_i - g(x_i, \hat{\theta}_{(1)}) \Big\} \dot{g}(x_i, \hat{\theta}_{(1)})}{\sum_{i=1}^{n} \dot{g}(x_i, \hat{\theta}_{(1)})^2}$$

$$\hat{\theta}_{(2)} = \hat{\theta}_{(1)} + \frac{\sum_{i=1}^{n} \Big\{ y_i - g(x_i, \hat{\theta}_{(1)}) \Big\} \dot{g}(x_i, \hat{\theta}_{(1)})}{\sum_{i=1}^{n} \dot{g}(x_i, \hat{\theta}_{(1)})^2}$$

Which leaves us with,

$$\hat{\theta}_{(2)} = \hat{\theta}_{(1)} + \Big\{ \sum_{i=1}^{n} \dot{g}(x_i, \hat{\theta}_{(1)})^2 \Big\}^{-1} \sum_{i=1}^{n} \Big\{ y_i - g(x_i, \hat{\theta}_{(1)}) \Big\} \dot{g}(x_i, \hat{\theta}_{(1)})$$

Then iterate until convergence,

$$\hat{\theta}_{(p+1)} = \hat{\theta}_{(p)} + \Big\{ \sum_{i=1}^{n} \dot{g}(x_i, \hat{\theta}_{(p)})^2 \Big\}^{-1} \sum_{i=1}^{n} \Big\{ y_i - g(x_i, \hat{\theta}_{(p)}) \Big\} \dot{g}(x_i, \hat{\theta}_{(p)})$$

Maximum Likelihood (ML)

The likelihood function describes the joint probability distribution of the data. It tells us what the probability is that the distribution parameters are in fact θ given that we have observed $X_1 = x, ..., X_n = x_n$ and $Y_n = y_1, ..., Y_n = y_n$.

$$\hat{\theta}_{ML} = \underset{\theta}{argmax}\ l_{y_1,...,y_n}(\theta)$$

$$l_{y_1,...,y_n}(\theta) = ln(L_{y_1,...,y_n}(\theta))$$

$$l_{y_1,...,y_n}(\theta) = ln \prod_{i=1}^{n} f_\theta(y_i) = \sum_{i=1}^{n} ln f_\theta(y_i)$$

Example: Bernoulli Random Variable

We are trying to estimate what θ is, given that we have our set of observations of y, $\{y_i\}_{i=1}^n$. To do this we maximise the probability that θ equals some θ_0 given our observations.

Setting up the problem,

$$(y_1, ..., y_n) \text{ iid}$$

$$P(y_i = 1) = \theta\ ,\ P(y_i = 0) = (1 - \theta)$$

$$\theta \in (0, 1)$$

PDF,

$$f_\theta(y_i) = \theta^{y_i}(1 - \theta)^{1-y_i}$$

Recall that the PDF = probability density function. It gives the probability of a certain event occurring. For example if we are rolling a 6-sided dice then P(2)=1/6, P(6)=1/6, etc.

Joint distribution and Log-likelihood function,

Since the y's are iid then,

$$f_\theta(y_1, ..., y_n) = f_\theta(y_1) \times ... \times f_\theta(y_n) = \prod_{i=1}^{n} f_\theta(y_i) = \prod_{i=1}^{n} \theta^{y_i}(1 - \theta)^{1-y_i}$$

Which can be simplified to,

$$f_\theta(y_1, ..., y_n) = \prod_{i=1}^{n} \theta^{y_i}(1 - \theta)^{1-y_i} = \theta^{\sum_{i=1}^{n} y_i}(1 - \theta)^{\sum_{i=1}^{n}(1-y_i)}$$

$$f_\theta(y_1, ..., y_n) = \theta^{n\bar{y}}(1 - \theta)^{n(1-\bar{y})}$$

Notice this is the joint distribution of the data, if in an exam they give you joint distribution then you can just use that straight away.

Then,

$$L_{y_1,...,y_n}(\theta) = \{\theta^{\bar{y}}(1 - \theta)^{(1-\bar{y})}\}^n$$

$$l_{y_1,...,y_n}(\theta) = n\{\bar{y} ln\theta + (1 - \bar{y}) ln(1 - \theta)\}$$

So the likelihood function exactly is the joint distribution of the data. The only difference is that we know the data values (the y's) and we don't know the parameter θ.

Maximum Likelihood Problem,

$$\hat{\theta}_{ML} = \underset{\theta \in (0,1)}{argmax}\ n\big\{\bar{y} ln\theta + (1-\bar{y}) ln(1-\theta)\big\}$$

FOC,

$$\frac{\partial l_{y_1,\ldots,y_n}(\theta)}{\partial\theta} = n\left\{\frac{\bar{y}}{\hat{\theta}_{ML}} - \frac{1-\bar{y}}{1-\hat{\theta}_{ML}}\right\} = 0$$

$$\frac{\bar{y}}{\hat{\theta}_{ML}} = \frac{1-\bar{y}}{1-\hat{\theta}_{ML}}$$

$$\bar{y} - \hat{\theta}_{ML}\bar{y} = \hat{\theta}_{ML} - \hat{\theta}_{ML}\bar{y}$$

Giving the solution,

$$\hat{\theta}_{ML} = \bar{y}$$

SOC,

$$\frac{\partial^2 l_{y_1,\ldots,y_n}(\theta)}{\partial\theta^2} = -n\left\{\frac{\bar{y}}{\hat{\theta}_{ML}^2} - \frac{1-\bar{y}}{(1-\hat{\theta}_{ML})^2}\right\}$$

Which for $\theta \in (0,1)$ and $0 \leq \bar{y} \leq 1$ implies that,

$$\frac{\partial^2 l_{y_1,\ldots,y_n}(\theta)}{\partial\theta^2} < 0$$

So it is certainly a maximum.

Example: Poisson Random Variable

Poisson variables take values $0, 1, 2, 3 \ldots$ and a so are useful for distribution of ‘number of times a person has been arrested’ or ‘number of children a family has’, etc. For more info on Poisson RVs see the next section below.

Density,

$$f_\theta(y) = \frac{e^{-\theta}\theta^y}{y!}\ ,\ \theta > 0\ ,\ y = 0, 1, 2, 3, \ldots$$

Joint-distribution and Log-likelihood

Assuming iidness,

$$f_\theta(y_1, \ldots, y_n) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^y}{y!}$$

$$l_{y_1,\ldots,y_n}(\theta) = ln \prod_{i=1}^{n} \frac{e^{-\theta}\theta^y}{y!} = \sum_{i=1}^{n} ln \frac{e^{-\theta}\theta^y}{y!}$$

$$l_{y_1,\ldots,y_n}(\theta) = \sum_{i=1}^{n} \big\{ -\theta ln e + y_i ln(\theta) - ln(y_i!)\big\}$$

$$l_{y_1,\ldots,y_n}(\theta) = -n\theta + ln(\theta)\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} ln(y_i!)$$

Maximum Likelihood Problem,

$$\hat{\theta}_{ML} = \underset{\theta>0}{argmax}\ l_{y_1,...,y_n}(\theta)$$

FOC,

$$\frac{\partial l_{y_1,...,y_n}(\theta)}{\partial \theta} = -n + \frac{1}{\hat{\theta}_{ML}} \sum_{i=1}^{n} y_i$$
$$n = \frac{1}{\hat{\theta}_{ML}} \sum_{i=1}^{n} y_i$$

Solution,

$$\hat{\theta}_{ML} = \bar{y}$$

SOC,

$$\frac{\partial^2 l_{y_1,...,y_n}(\theta)}{\partial \theta^2} = -\frac{\bar{y}}{\hat{\theta}^2_{ML}}$$

Which is negative for any positive mean of y and any value of θ, hence we have a maximum.

Example: Conditional Poisson Distribution

Model,

$$y_i = e^{x_i'\beta} + u_i$$

Where x and β are $P \times 1$ vectors.

Probability Density Function,

$$f_{\theta_i(\beta)}(y_i \mid x_i) = \frac{e^{-\theta_i(\beta)}\{\theta_i(\beta)\}^{y_i}}{y_i!}\ ,\ y_i = 0, 1, 2, ...$$

Given that y is the Poisson distributed variable and that in the case of the explanation of the Poisson Distribution below θ is equal to the expectation of the Poisson variable y, we can take the conditional expectation here to find θ.

$$E[y_i \mid x_i] = \theta_i(\beta) = e^{-x_i'\beta}$$

Finding Joint-distribution and Likelihood function,

$$
\begin{aligned}
f_{\theta_i(\beta)}(y_1, ..., y_n \mid x_1, ..., x_n) &= \prod_{i=1}^{n} \frac{e^{-\theta_i(\beta)}\{\theta_i(\beta)\}^{y_i}}{y_i!} \\
l_{y_1,...,y_n \mid x_1,...,x_n}(\theta_i(\beta)) &= ln \prod_{i=1}^{n} \frac{e^{-\theta_i(\beta)}\{\theta_i(\beta)\}^{y_i}}{y_i!} \\
&= \sum_{i=1}^{n} ln \frac{e^{-\theta_i(\beta)}\{\theta_i(\beta)\}^{y_i}}{y_i!} \\
&= \sum_{i=1}^{n} \Big\{ -\theta_i(\beta) lne + y_i ln\theta_i(\beta) - ln(y_i!) \Big\} \\
l_{y_1,...,y_n \mid x_1,...,x_n}(\beta) &= \sum_{i=1}^{n} \Big\{ -e^{x_i'\beta} + y_i ln(e^{x_i'\beta} - ln(y_i!) \Big\} \\
&= \sum_{i=1}^{n} \Big\{ -e^{x_i'\beta} + y_i x_i'\beta - ln(y_i!) \Big\}
\end{aligned}
$$

FOCs,

$$
\frac{\partial l_{y_1,...,y_n \mid x_1,...,x_n}(\beta)}{\partial \beta} = \sum_{i=1}^{n} \Big\{ -x_i' e^{x_i'\beta} + y_i x_i' \Big\} = \sum_{i=1}^{n} \{ y_i - e^{x_i'\beta} \} x_i' = 0
$$

We would need to use the Gauss-Newton method to approximate a solution.

Poisson Random Variables

Let Y be a Poisson RV. The information below is just general info on Poisson RVs should it be needed.

Expectation

The expectation of Y is given by each value Y could take $(0, 1, 2, \dots)$ weighted by the probability of that value occurring.

$$E[Y] = \sum_{y=0}^{\infty} y f_\theta(y) = \sum_{y=0}^{\infty} y \frac{e^{-\theta}\theta^y}{y!} = \sum_{y=1}^{\infty} y \frac{e^{-\theta}\theta^y}{y!}$$

Notice,

$$\begin{aligned}
y! &= 1 \times 2 \times 3 \times ... \times y \\
y! &= 1 \times 2 \times 3 \times ... \times (y-1) \times y \\
y! &= \big[1 \times 2 \times 3 \times ... \times (y-1)\big] y \\
y! &= (y-1)! \times y
\end{aligned}$$

So then,

$$\begin{aligned}
E[Y] &= \sum_{y=1}^{\infty} \frac{e^{-\theta}\theta^y}{(y-1)!} \\
&= \theta e^{-\theta} \sum_{y=1}^{\infty} \frac{\theta^{(y-1)}}{(y-1)!} \\
&= \theta e^{-\theta} \sum_{y=0}^{\infty} \frac{\theta^y}{y!} \\
&= \theta e^{-\theta} \left[1 + \frac{\theta}{1!} + \frac{\theta^2}{2!} + ...\right]
\end{aligned}$$

The part in the square brackets is the definition of the exponential function, hence,

$$\begin{aligned}
E[Y] &= \theta e^{-\theta} e^{\theta} \\
E[Y] &= \theta
\end{aligned}$$

Variance

$$Var(Y) = E[Y^2] - E[Y]^2$$

Already knowing the expectation, consider,

$$E[Y^2] = \sum_{y=0}^{\infty} y^2 \frac{e^{-\theta}\theta^y}{y!} = \sum_{y=1}^{\infty} y^2 \frac{e^{-\theta}\theta^y}{y!} = \sum_{y=1}^{\infty} y^2 \frac{e^{-\theta}\theta^y}{(y-1)!y} = \sum_{y=1}^{\infty} y \frac{e^{-\theta}\theta^y}{(y-1)!}$$

We can change the limit since when $y = 0$ the term is zero, hence the sum is unaffected.

$$\begin{aligned}
E[Y^2] &= \sum_{y=1}^{\infty} y \frac{e^{-\theta}\theta^y}{(y-1)!} \\
&= e^{-\theta}\theta \sum_{y=1}^{\infty} y \frac{\theta^{(y-1)}}{(y-1)!} \\
&= e^{-\theta}\theta \sum_{y=1}^{\infty} \left\{ (y-1) \frac{\theta^{(y-1)}}{(y-1)!} + \frac{\theta^{(y-1)}}{(y-1)!} \right\} \\
&= e^{-\theta}\theta \left\{ \sum_{y=1}^{\infty} (y-1) \frac{\theta^{(y-1)}}{(y-1)!} + \sum_{y=1}^{\infty} \frac{\theta^{(y-1)}}{(y-1)!} \right\}
\end{aligned}$$

The first term is 0 when $y = 1$ hence change the limit to $y = 2$, also notice that $(y-1)! = (y-2)!(y-1)$,

$$E[Y^2] = e^{-\theta}\theta\left\{\theta\sum_{y=2}^{\infty}\frac{\theta^{(y-2)}}{(y-2)!} + \sum_{y=1}^{\infty}\frac{\theta^{(y-1)}}{(y-1)!}\right\}$$

Finally counting from zero,

$$\begin{aligned} E[Y^2] &= e^{-\theta}\theta\left\{\theta\sum_{y=0}^{\infty}\frac{\theta^{y}}{y!} + \sum_{y=0}^{\infty}\frac{\theta^{y}}{y!}\right\} \\ &= e^{-\theta}\theta\{\theta e^{\theta} + e^{\theta}\} \\ &= \theta^2 + \theta \end{aligned}$$

Putting it all together,

$$Var(Y) = \theta$$

Non-Linear Asymptotics

$$\hat{\theta} = \underset{\theta \in \Theta}{argmax}\; Q_n(\theta)$$

Assumptions

(A) The parameter space Θ is compact (closed and bounded) ($\theta_0 \in \Theta$)

(B) $Q_n(\theta)$ is continuous in $\theta \in \Theta$ (differentiable)

(C) For a non-stochastic function $Q(\theta)$, uniformly in θ,

$$Q_n(\theta) \xrightarrow{P} Q(\theta)$$

and $Q(\theta)$ achieves a unique global maximum (or minimu for NLLS) at θ_0.

Consistency

$$\hat{\theta}_n \xrightarrow{P} \theta_0$$

Justification by example: NLLS

$$y_i = (\theta_0 + x_i)^2 + u_i$$

Assume:

(i) $\theta_0 \in [a, b]$

(ii) $\underset{n\to\infty}{lim} \frac{1}{n}\sum_{i=1}^{n} x_i = q$

(iii) $\underset{n\to\infty}{lim} \frac{1}{n}\sum_{i=1}^{n} x_i^2 = p > q^2$

(iv) $u_i \sim iid\; N(0, 1)$

$$\begin{aligned}\hat{\theta}_n &= argmin\; \frac{1}{n}\sum_{i=1}^{n}\left\{y_i - (\theta + x_i)^2\right\}^2 \\ &= argmin\; \frac{1}{n}\sum_{i=1}^{n}\left\{u_i + (\theta_0 + x_i)^2 - (\theta + x_i)^2\right\}^2\end{aligned}$$

$$(\theta_0 + x_i)^2 - (\theta + x_i)^2 = 2x_i(\theta_0 - \theta) + (\theta_0^2 - \theta^2)$$

For condition C:

It can be shown that $Q_n(\theta) \xrightarrow{P} Q(\theta)$ uniformly in θ,

$$\begin{aligned}Q_n(\theta) &= \frac{1}{n}\sum_{i=1}^{n}\left\{u_i + 2x_i(\theta_0 - \theta) + (\theta_0^2 - \theta^2)\right\}^2 \\ Q_n(\theta) &\xrightarrow{P} Q(\theta) \\ Q(\theta) &= \sigma_u^2 + (\theta_0^2 - \theta^2)^2 + 4p(\theta_0 - \theta)^2 + 4q(\theta_0^2 - \theta^2)(\theta_0 - \theta)\end{aligned}$$

And it can be shown that$Q(\theta)$ has a unique global minimum (since this is NLLS and it is a min not a max) at θ_0,

$$Q(\theta) \geq \sigma_u^2 + (\theta_0^2 - \theta^2)(\theta_0 + \theta + 2q)$$

Hence $Q(\theta)$ is minimised at $\theta_0 = \theta$.

Normality

$$\sqrt{n}(\hat{\theta}_n - \theta_0)$$

The solution to optimisation problem are,

$$\dot{Q}_n(\theta) = \frac{\partial Q_n(\theta)}{\partial \theta} = 0$$
$$\dot{Q}_n(\hat{\theta}_n) = 0$$

Applying first order Taylor expansion

$$\dot{Q}_n(\hat{\theta}_n) \approx \dot{Q}_n(\theta_0) + \ddot{Q}_n(\theta_0)(\hat{\theta}_n - \theta_0)$$

Knowing this is equivalent to zero,

$$(\hat{\theta}_n - \theta_0) \approx \dot{Q}_n(\theta_0)\Big\{\ddot{Q}_n(\theta_0)\Big\}^{-1}$$
$$\sqrt{n}(\hat{\theta}_n - \theta_0) \approx \sqrt{n}\dot{Q}_n(\theta_0)\Big\{\ddot{Q}_n(\theta_0)\Big\}^{-1}$$

Then,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N\Big\{0, A(\theta_0)^{-1}B(\theta_0)A(\theta_0)^{-1}\Big\}$$

Justification by example: NLLS

$$\hat{\theta}_n = argmin\ Q_n^{NLLS}(\theta)$$
$$Q_n^{NLLS}(\theta) = \sum_{i=1}^{n}\Big\{y_i - g(x_i,\theta)\Big\}^2$$

FOCs

$$\dot{Q}_n^{NLLS}(\theta) = -2\sum_{i=1}^{n}\Big\{y_i - g(x_i,\hat{\theta}_n)\Big\}\dot{g}(x_i,\hat{\theta}_n) = 0$$

Taylor expansion,

$$\dot{Q}_n^{NLLS}(\theta) \approx \dot{Q}_n^{NLLS}(\theta_0) + \ddot{Q}_n^{NLLS}(\theta_0)(\hat{\theta}_n - \theta_0)$$

SOCs

$$\ddot{Q}_n^{NLLS}(\theta) = 2\sum_{i=1}^{n}\dot{g}(x_i,\theta)^2 - 2\sum_{i=1}^{n}\Big\{y_i - g(x_i,\theta)\Big\}\ddot{g}(x_i,\theta)$$

FOCs and SOCs evaluated at θ_0,

$$\dot{Q}_n^{NLLS}(\theta_0) = -2\sum_{i=1}^{n}u_i\dot{g}(x_i,\theta_0)$$
$$\ddot{Q}_n^{NLLS}(\theta_0) = 2\sum_{i=1}^{n}\dot{g}(x_i,\theta_0)^2 - 2\sum_{i=1}^{n}u_i\ddot{g}(x_i,\theta_0)$$

The first order Taylor expansion gave $(\hat{\theta}_n - \theta_0) \approx \dot{Q}_n(\theta_0)\Big\{\ddot{Q}_n(\theta_0)\Big\}^{-1}$,

$$\sqrt{n}(\hat{\theta} - \theta_0) \approx \frac{n^{-\frac{1}{2}}\sum_{i=1}^{n}u_i\dot{g}(x_i,\theta_0)}{n^{-1}\Big\{\sum_{i=1}^{n}\dot{g}(x_i,\theta_0)^2 - \sum_{i=1}^{n}u_i\ddot{g}(x_i,\theta_0)\Big\}}$$
$$\approx \frac{n^{-\frac{1}{2}}\sum_{i=1}^{n}u_i\dot{g}(x_i,\theta_0)}{n^{-1}\sum_{i=1}^{n}\dot{g}(x_i,\theta_0)^2} + \text{small}$$

(since the second derivative is small)

Giving the final result as,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \approx \frac{n^{-\frac{1}{2}} \sum_{i=1}^{n} u_i \dot{g}(x_i, \theta_0)}{n^{-1} \sum_{i=1}^{n} \dot{g}(x_i, \theta_0)^2} + \text{small} \xrightarrow{D} N(0, \sigma_u^2 C^{-1})$$

$$C = p \underset{n \to \infty}{lim} n^{-1} \sum_{i=1}^{n} \left\{ \dot{g}(x_i, \theta_0) \right\}^2 > 0$$

Binary Choice Models

Used for binary choice,

- Whether a mortgage application is accepted or denied,
- Participate in the labour market or not,
- Doing an MPhil or not.

Dependent variable is binary: $y = \begin{cases} 1 \text{ if event } A \text{ occurs} \\ 0 \text{ otherwise} \end{cases}$

We want to model the probability that $y = 1$, hence,

$$P_\theta(y = 1 \mid x) = G(x, \theta)$$
$$P_\theta(y = 0 \mid x) = 1 - G(x, \theta)$$

$$E[y \mid x] = 1 \times G(x, \theta) + 0 \times (1 - G(x, \theta)) = G(x, \theta)$$

(1) Linear Probability Model:

$$P_\theta(y = 1 \mid x) = G(x, \theta) = x'\theta = \theta_1 + \theta_2 x_2 + ... + \theta_k x_k$$

Unsurprisingly, given the name, $G(x, \theta)$ is a linear function.

And specifies,

$$y_i = \theta_1 + \theta_2 x_2 + ... + \theta_k x_k + u_i = x'\theta + u_i$$

This is linear in parameters; hence we can use OLS, but the problem with this is:

(1) The model error term is heteroskedastic by construction

- Notice that given y_i must be equal to zero or one, that means that $x_i'\theta + u_i$ must also always equal zero or one.
- $u_i = \begin{cases} -x_i'\theta \text{ and } P(u_i = -x_i'\theta) = G(x, \theta) \\ 1 - x_i'\theta \text{ and } P(u_i = 1 - u_i'\theta) = 1 - G(x, theta) \end{cases}$
- This means that $Var(u_i \mid x_i) = x_i'\theta(1 - x_i'\theta)$
- In other words, the error is conditionally heteroskedastic

(1.1) Gauss-Markov does not apply

(1.2) More efficient estimators exist

(2) The predicted probabilities may not belong to $[0, 1]$, that is $\hat{y}_i = x_i'\hat{\theta} \notin [0, 1]$.

Overcoming problem (2)

- This can be overcome by modelling the actual probabilities so that $\underset{x'\theta \to \infty}{lim} P_\theta(y = 1 \mid x) = 1$ and $\underset{x'\theta \to -\infty}{lim} P_\theta(y = 1 \mid x) = 0$.
- But this is achieved by choosing the function $P_\theta(y = 1 \mid x) = G(x, \theta)$ to be a distribution function.

That is exactly what probit and logit models do!

- They use normal distribution (probit)
- And logistic distribution (logit)

(2) Probit & Logit Model:

Generally: $P_\theta(y = 1 \mid x) = E[y \mid x] = G(x'\theta)$ where $G(x'\theta)$ is nonlinear.

Probit

$$P_\theta(y = 1 \mid x) = \int_{-\infty}^{x'\theta} \phi(t)dt = \Phi(x'\theta)$$

Where ϕ is the standard normal density (PDF) and Φ the normal distribution (CDF).

Logit

$$P_\theta(y = 1 \mid x) = \frac{e^{x'\theta}}{1 + e^{x'\theta}} = \Lambda(x'\theta)$$

Interpretation

(I) General case: Measuring the effect of regressors by marginal effect.

- Where x is a binary variable the effect of changing x_2 from zero to one is:

$$G(\theta_1 + \theta_2 + \theta_3 x_3 + ... + \theta_k x_k) - G(\theta_1 + \theta_3 x_3 + ... + \theta_k x_k)$$

- Where x is continuous the effect of a small change in x_j is:

$$\frac{\partial E[y \mid x]}{\partial x_j} = \frac{\partial G(\theta_1 + \theta_2 x_2 + \theta_3 x_3 + ... + \theta_k x_k)}{\partial x_j}$$

(II) Probit Case

$$\frac{\partial E[y \mid x]}{\partial x_j} = \frac{d\Phi(x'\theta)}{d(x'\theta)} \frac{\partial (x'\theta)}{\partial x_j} = \frac{d\Phi(x'\theta)}{d(x'\theta)} \theta_j = \phi(x'\theta)\theta_j$$

- $\phi(x'\theta)$ is known as the scale factor.
- This needs to be evaluated at a specific value of x,

$$\left. \frac{\partial E[y \mid x]}{\partial x_j} \right|_{\bar{x}} = \phi(\bar{x}'\theta)\theta_j$$

(III) Logit Case

$$\begin{aligned} \frac{\partial E[y \mid x]}{\partial x_j} &= \frac{d\Lambda(x'\theta)}{d(x'\theta)} \frac{\partial (x'\theta)}{\partial x_j} = \frac{d\Lambda(x'\theta)}{d(x'\theta)} \theta_j \\ &= \frac{e^{x'\theta}(1 + e^{x'\theta}) - e^{x'\theta} e^{x'\theta}}{(1 + e^{x'\theta})^2} \theta_j \\ &= \Lambda(x'\theta)(1 - \Lambda(x'\theta))\theta_j \end{aligned}$$

- $\Lambda(x'\theta)(1 - \Lambda(x'\theta))$ is known as the scale factor.
- This needs to be evaluated at a specific value of x,

$$\left. \frac{\partial E[y \mid x]}{\partial x_j} \right|_{\bar{x}} = \Lambda(\bar{x}'\theta)(1 - \Lambda(\bar{x}'\theta))\theta_j$$

Estimation

We estimate θ by Maximum Likelihood, not NLLS as it doesn't account for heteroskedasticity.

(I) General Case

$$G(x_i'\theta)$$

PDF and Joint Density Function,

$$P_\theta(y_i \mid x_i) = \left[G(x_i'\theta)\right]^{y_i}\left[1 - G(x_i'\theta)\right]^{1-y_i}$$
$$P_\theta(y_1, ..., y_n \mid x_1, ..., x_n) = \prod_{i=1}^{n}\left[G(x_i'\theta)\right]^{y_i}\left[1 - G(x_i'\theta)\right]^{1-y_i}$$

Log-Likelihood Function

$$l_{(y_1,...,y_n \mid x_1,...,x_n)}(\theta) = ln\prod_{i=1}^{n}\left[G(x_i'\theta)\right]^{y_i}\left[1 - G(x_i'\theta)\right]^{1-y_i}$$
$$l_{(y_1,...,y_n \mid x_1,...,x_n)}(\theta) = \sum_{i=1}^{n} y_i ln\left[G(x_i'\theta)\right] + \sum_{i=1}^{n}(1-y_i)ln\left[1 - G(x_i'\theta)\right]$$

Estimator

$$\hat{\theta}_{ML} = \underset{\theta\in\mathbb{R}^k}{argmax}\ l_{(y_1,...,y_n \mid x_1,...,x_n)}(\theta)$$

FOCs

$$\frac{\partial l_{(y_1,...,y_n \mid x_1,...,x_n)}(\theta)}{\partial\theta} = \sum_{i=1}^{n} y_i\frac{\dot{G}(x_i'\theta)x_i}{G(x_i'\theta)} + \sum_{i=1}^{n}(1-y_i)\frac{-\dot{G}(x_i'\theta)x_i}{1-G(x_i'\theta)} = 0$$
$$0 = \sum_{i=1}^{n}\left[y_i\frac{\dot{G}(x_i'\theta)}{G(x_i'\theta)} - (1-y_i)\frac{\dot{G}(x_i'\theta)}{1-G(x_i'\theta)}\right]x_i$$

(II) Probit Case

$$G(x_i\theta) = \Phi(x_i'\theta)$$

Hence,

$$\frac{\partial l_{(y_1,...,y_n \mid x_1,...,x_n)}(\theta)}{\partial\theta} = \sum_{i=1}^{n}\left[y_i\frac{\phi(x_i'\theta)}{\Phi(x_i'\theta)} - (1-y_i)\frac{\phi(x_i'\theta)}{1-\Phi(x_i'\theta)}\right]x_i = 0$$

Which gives,

$$\frac{\partial l_{(y_1,...,y_n \mid x_1,...,x_n)}(\theta)}{\partial\theta} = \sum_{i=1}^{n} w_i\left[y_i - \Phi(x_i'\theta)\right]x_i = 0$$

(III) Logit Case

$$G(x_i\theta) = \Lambda(x_i'\theta) = \frac{e^{x_i'\theta}}{(1+e^{x_i'\theta})}$$

Hence,

$$\frac{\partial l_{(y_1,...,y_n \mid x_1,...,x_n)}(\theta)}{\partial\theta} = \sum_{i=1}^{n}\left[y_i\frac{\frac{e^{x_i'\theta}}{(1+e^{x_i'\theta})^2}}{\frac{e^{x_i'\theta}}{(1+e^{x_i'\theta})}} - (1-y_i)\frac{\frac{e^{x_i'\theta}}{(1+e^{x_i'\theta})^2}}{1-\frac{e^{x_i'\theta}}{(1+e^{x_i'\theta})}}\right]x_i = 0$$

Which gives,

$$\frac{\partial l_{(y_1,...,y_n \mid x_1,...,x_n)}(\theta)}{\partial\theta} = \sum_{i=1}^{n}\left[y_i - \Lambda(x_i'\theta)\right]x_i = 0$$

Measure of Fit

$$R^2_{McFadden} = 1 - \frac{l_{(y_1,\ldots,y_n \mid x_1,\ldots,x_n)}(\hat{\theta}_{ML})}{l_{(y_1,\ldots,y_n \mid x_1,\ldots,x_n)}(\bar{y})}$$

$$R^2_{McFadden} = 1 - \frac{\sum_{i=1}^{n}\Big[y_i ln\big[G(x_i'\hat{\theta}_{ML})\big] + (1-y_i)ln\big[1 - G(x_i'\hat{\theta}_{ML})\big]\Big]}{n\big[y_i ln\bar{y} + (1-y_i)ln(1-\bar{y})\big]}$$

Asymptotic Properties

$$AV\hat{a}r(\hat{\theta}_{ML} = \Big[\sum_{i=1}^{n} \frac{\big\{G(x_i'\theta)\big\}^2 x_i x_i'}{G(x_i'\theta)(1-G(x_i'\theta))}\Big]^{-1}$$

Inference

$$H_0 : \theta_j = 0$$
$$H_1 : \theta_j \neq 0$$
$$t = \frac{\hat{\theta}_j}{s.e.(\hat{\theta}_j)}$$

Wald Test

$$H_0 : c(\theta) = q$$
$$H_1 : c(\theta) \neq q$$

(where $c(.)$ is possibily non-linear)

$$W = \big\{c(\hat{\theta}) - q\big\}'\Big[AV\hat{a}r\big\{c(\hat{\theta}) - q\big\}\Big]^{-1}\big\{c(\hat{\theta}) - q\big\}$$

If the null holds then $\big\{c(\hat{\theta}) - q\big\}$ should be close to zero. Further this is self-normalising since $\big\{c(\hat{\theta}) - q\big\}$ is divided by its standard deviation.

$$W \overset{a}{\sim} \mathcal{X}^2_J$$

Likelihood Ratio Test

$$H_0 : c(\theta) = q$$
$$H_1 : c(\theta) \neq q$$

$$LR = 2\big\{ln(\hat{L}_R) - ln(\hat{L}_{UR})\big\} = 2ln\left(\frac{\hat{L}_R}{\hat{L}_{UR}}\right)$$

Here the L's represent the values of the likelihood function evaluated at the restricted (takes into account the null) and unrestricted models. If the null holds then the restricted and unrestricted models will be close to one another and LR will be close to zero.

$$LR \overset{a}{\sim} \mathcal{X}^2_J$$

Lagrange Multiplier Test

$$H_0 : c(\theta) = q$$
$$H_1 : c(\theta) \neq q$$

$$LM = \left\{ \frac{\partial lnL(\hat{\theta}_R)}{\partial \hat{\theta}_R} \right\}' \left\{ I(\hat{\theta}_R) \right\}^{-1} \left\{ \frac{\partial lnL(\hat{\theta}_R)}{\partial \hat{\theta}_R} \right\}$$

Where $I(\hat{\theta}_R)$ is the information matrix (expected Hessian matrix of second derivatives).

This test works as if the null hypothesis holds then $\hat{\theta}_R$ will not be far from the point that maximises the log-likelihood and hence the score $\frac{\partial lnL(\hat{\theta}_R)}{\partial \hat{\theta}_R}$ will be close to zero.

$$LM \overset{a}{\sim} \mathcal{X}^2_J$$

Note all three tests are asymptotically equivalent (they have the same distribution!)

Count Data Models

Like Binary choice models except the the variable of interest is a count of the number of occurrences of a certain event,

- Number of children born in a family,
- Number of times a person has been arrested,
- Number of people surviving past age 100.

In all of these y takes the values 0,1,2,3,...

Poisson Regression Model

The Poisson regression model specifies,

$$E[y \mid x] = e^{\theta_1 + \theta_2 x_2 + ... + \theta_k x_k}$$

Predicted values for y are always positive since the exponential function is always positive.

Marginal effect

Binary regressor x_2:

$$E[y \mid x_2 = 1, ..., x_k] - E[y \mid x_2 = 0, ..., x_k] = exp(\theta_1 + \theta_2 + \theta_3 x_3 + ... + \theta_k x_k) - exp(\theta_1 + \theta_3 x_3 + ... + \theta_k x_k)$$

Continuous regressor x_j:

$$\frac{\partial E[y \mid x]}{\partial x_j} = exp(\theta_1 + \theta_2 x_2 + ... + \theta_k x_k)\theta_j$$

(needs to be evaluated at a given x)

Coefficient interpretation

$$ln(E[y \mid x]) = \theta_1 + \theta_2 x_2 + ... + \theta_k x_k$$

Then,

$$ln(E[y \mid x_1, ..., x_k]) - ln(E[y \mid x_1, ..., \tilde{x}_k]) = \theta_k (x_k - \tilde{x}_k) = \theta_k \Delta x_k$$

This means that $100 \times \theta_k$ is approximately the percentage change in given a one-unit increase in x_k.

In more detail let,

$$\begin{aligned}
\Delta_k E[y \mid x] &= \frac{E[y \mid x_1, ..., x_k] - E[y \mid x_1, ..., \tilde{x}_k]}{E[y \mid x_1, ..., \tilde{x}_k]} \\
&= \frac{exp(\theta_1 + \theta_2 x_2 + ... + \theta_k x_k) - exp(\theta_1 + \theta_2 x_2 + ... + \theta_k \tilde{x}_k)}{exp(\theta_1 + \theta_2 x_2 + ... + \theta_k \tilde{x}_k)} \\
&= \frac{exp(\theta_1 + \theta_2 x_2 + ... + \theta_k x_k)}{exp(\theta_1 + \theta_2 x_2 + ... + \theta_k \tilde{x}_k)} - 1 \\
&= exp(\theta_k \Delta x_k) - 1
\end{aligned}$$

First order Taylor expansion of the exponential function gives,

$$exp(\theta_k \Delta x_k) \approx exp(0) + \frac{exp(0)}{1!}(\theta_k \Delta x_k - 0) = 1 + \theta_k \Delta x_k$$

Subbing this back into the change in the conditional expectation gives,

$$\Delta_k E[y \mid x] \approx \theta_k \Delta x_k$$

ML Estimation

$$f_{\theta_i(\beta)}(y_i \mid x_i) = \frac{e^{-\theta_i(\beta)}\theta_i(\beta)^{y_i}}{y_i!} \text{ where } \theta_i(\beta) = e^{x_i'\beta}$$
$$E[y_i \mid x_i] = Var(y_i \mid x_i) = \theta_i(\beta) = e^{x_i'\beta}$$

And,

$$y_i = e^{x_i'\beta} + u_i$$

By the iid assumption,

$$f_{\theta_i(\beta)}(y_1, ..., y_n \mid x_1, ..., x_n) = \prod_{i=1}^{n} \frac{e^{-\theta_i(\beta)}\theta_i(\beta)^{y_i}}{y_i!}$$

Which means that,

$$\begin{aligned} l_{y_1,...,y_n \mid x_1,...,x_n}(\beta) &= \sum_{i=1}^{n} ln\frac{e^{-\theta_i(\beta)}\theta_i(\beta)^{y_i}}{y_i!} \\ &= \sum_{i=1}^{n} \Big\{ -\theta_i(\beta)lne + y_i ln\theta_i(\beta) - ln(y!) \Big\} \end{aligned}$$

Given that $\theta_i(\beta) = e^{x_i'\beta}$

$$l_{y_1,...,y_n \mid x_1,...,x_n}(\beta) = \sum_{i=1}^{n} \Big\{ -\theta_i(\beta) + y_i x_i'\beta - ln(y!) \Big\}$$

We know that

$$\hat{\beta}_{ML} = argmax\ l_{y_1,...,y_n \mid x_1,...,x_n}(\beta)$$

FOCs,

$$\frac{\partial l_{y_1,...,y_n \mid x_1,...,x_n}(\beta)}{\partial \beta} = \frac{1}{n}\sum_{i=1}^{n}(y_i - e^{x_i'\hat{\beta}}x_i' = 0$$

Use Gauss-Newton.

Asymptotics

The ML estimator is,

- Consistent,
- Normal,
- Efficient.

$$AV\hat{a}r(\hat{\beta}) = \Big\{ \sum_{i=1}^{n} exp(x_i'\beta)x_i x_i' \Big\}^{-1} \text{ which is a } k \times k \text{ matrix.}$$

Inference,

- Wald test,
- Likelihood ratio,
- Lagrange multiplier test.