Econometrics I

Professor William Greene

Stern School of Business

Department of Economics
Econometrics I

Part 8 – Asymptotic Distribution Theory
Asymptotics: Setting

Most modeling situations involve stochastic regressors, nonlinear models or nonlinear estimation techniques. The number of exact statistical results, such as expected value or true distribution, that can be obtained in these cases is very low. We rely, instead, on approximate results that are based on what we know about the behavior of certain statistics in large samples. Example from basic statistics: We know a lot about \( \bar{x} \). What can we say about \( 1/ \bar{x} \)?
Convergence

Definitions, kinds of convergence as $n$ grows large:

1. To a constant; **example**, the sample mean, $\bar{x}$ converges to the population mean.

2. To a random variable; **example**, a $t$ statistic with $n - 1$ degrees of freedom converges to a standard normal random variable
Convergence to a Constant

Sequences and limits.

Convergence of a sequence of constants, indexed by n:

Ordinary limit: \[ \lim_{n \to \infty} \frac{n(n+1)/2+3n+5}{n^2+2n+1} = \frac{1}{2} \frac{n^2 + \frac{3}{2} n + 5}{n^2 + 2n+1} \rightarrow \frac{1}{2} \]

(Note the use of the “leading term”)

Convergence of a sequence of random variables.

What does it mean for a random variable to converge to a constant? Convergence of the variance to zero. The random variable converges to something that is not random.
Convergence Results

Convergence of a sequence of random variables to a constant -

**Convergence in mean square:**
*Mean converges to a constant, variance converges to zero.*
(Far from the most general, but definitely sufficient for our purposes.)

\[
\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad E[\bar{x}_n] = \mu \rightarrow \mu, \quad \text{Var}[\bar{x}_n] = \sigma^2 / n \rightarrow 0
\]

A convergence theorem for sample moments.

**Sample moments converge in probability to their population counterparts.**

Generally the form of **The Law of Large Numbers**. (Many forms; see Appendix D in your text. This is the “weak” law of large numbers.)

Note the great generality of the preceding result.

\[
(1/n) \sum_i g(z_i) \text{ converges to } E[g(z_i)].
\]
Extending the Law of Large Numbers

Suppose $x$ has mean $\mu$ and finite variance $\sigma^2$ and $x_1, x_2, ..., x_n$ are a random sample. Then the LLN applies to $\bar{x}$.

Let $z_i = x_i^p$. Then, $z_1, z_2, ..., z_n$ are a random sample from a population with mean $E[z] = E[x^p]$ and $\text{Var}[z] = E[x^{2p}] - \{E[x^p]\}^2$. The LLN applies to $z$ as long as the moments are finite.

There is no mention of normality in any of this.

Example: If $x \sim N[0, \sigma^2]$, then

$$E[x^p] = \begin{cases} 0 & \text{if } P \text{ is odd} \\ \sigma^p (P - 1)!! & \text{if } P \text{ is even} \end{cases}$$

$(P - 1)!! = \text{product of odd numbers up to } P-1$.

No power of $x$ is normally distributed. Normality is irrelevant to the LLN.
Probability Limit

Let \( \theta \) be a constant, \( \varepsilon \) be any positive value, and \( n \) index the sequence. If \( \lim_{n \to \infty} \text{Prob}[|b_n - \theta| > \varepsilon] = 0 \) then, \( \text{plim} \ b_n = \theta \).

\( b_n \) *converges in probability* to \( \theta \). (A definition.)

In words, the probability that the difference between \( b_n \) and \( \theta \) is larger than \( \varepsilon \) for any \( \varepsilon \) goes to zero. \( b_n \) becomes arbitrarily close to \( \theta \).

Mean square convergence is sufficient (not necessary) for convergence in probability. (We will not require other, broader definitions of convergence, such as "almost sure convergence."
Mean Square Convergence

![Graph showing mean square convergence](image)

**Figure D.1** Quadratic Convergence to a Constant, $\theta$. 
Probability Limits and Expectations

What is the difference between $E[b_n]$ and $\text{plim } b_n$?

A notation

$\text{plim } b_n = \theta \iff b_n \xrightarrow{P} \theta$
Consistency of an Estimator

If the random variable in question, $b_n$ is an estimator (such as the mean), and if

$$\text{plim } b_n = \theta,$$

then $b_n$ is a consistent estimator of $\theta$. Estimators can be inconsistent for $\theta$ for two reasons:

1. They are consistent for something other than the thing that interests us.
2. They do not converge to constants. They are not consistent estimators of anything.

We will study examples of both.
The Slutsky Theorem

Assumptions: If
\( b_n \) is a random variable such that \( \text{plim} b_n = \theta \).
For now, we assume \( \theta \) is a constant.
g(.) is a continuous function with continuous derivatives.
g(.) is not a function of \( n \).

Conclusion: Then \( \text{plim}[g(b_n)] = g[\text{plim}(b_n)] \) assuming
\( g[\text{plim}(b_n)] \) exists. (VVIR!)
Works for probability limits. Does not work for expectations.

\[
E[\bar{x}_n] = \mu; \quad \text{plim}(\bar{x}_n) = \mu, \quad E[1/\bar{x}_n] = ?; \quad \text{plim}(1/\bar{x}_n) = 1/\mu
\]
Slutsky Corollaries

$x_n$ and $y_n$ are two sequences of random variables with probability limits $\theta$ and $\mu$.

Plim \((x_n \pm y_n) = \theta \pm \mu \) (sum)

Plim \((x_n \times y_n) = \theta \times \mu \) (product)

Plim \((x_n / y_n) = \theta / \mu \) (ratio, if $\mu \neq 0$)

Plim\[g(x_n, y_n)\] = \(g(\theta, \mu)\) assuming it exists and $g(.)$ is continuous with continuous partials, etc.
Slutsky Results for Matrices

Functions of matrices are continuous functions of the elements of the matrices. Therefore, if $\text{plim} A_n = A$ and $\text{plim} B_n = B$ (element by element), then

$$\text{plim}(A_n^{-1}) = [\text{plim} A_n]^{-1} = A^{-1}$$

and

$$\text{plim}(A_n B_n) = \text{plim} A_n \text{plim} B_n = AB$$
Limiting Distributions

Convergence to a kind of random variable instead of to a constant

$x_n$ is a random sequence with cdf $F_n(x_n)$. If $\text{plim } x_n = \theta$ (a constant), then $F_n(x_n)$ becomes a point. But, $x_n$ may converge to a specific random variable. The distribution of that random variable is the **limiting distribution of $x_n$**. Denoted

$$
\begin{align*}
  x_n \xrightarrow{d} x & \iff F_n(x_n) \xrightarrow{n \to \infty} F(x)
\end{align*}
$$
A Limiting Distribution

\( x_1, x_2, \ldots, x_n = \) a random sample from \( \text{N}[\mu, \sigma^2] \)

For purpose of testing \( H_0 : \mu = 0 \), the usual test statistic is

\[
t_{n-1} = \frac{\bar{x}_n}{\left( \frac{s_n}{\sqrt{n}} \right)}
\]

where

\[
s_n = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2}{n - 1}
\]

The exact density of the random variable \( t_{n-1} \) is \( t \) with \( n-1 \) degrees of freedom. The density varies with \( n \);

\[
f(t_{n-1}) = \frac{\Gamma(n/2)}{\Gamma((n-1)/2) \sqrt{(n-1)\pi}} \left[ 1 + \frac{t_{n-1}^2}{n-1} \right]^{-n/2}
\]

The cdf, \( F_{n-1}(t) = \int_{-\infty}^{t} f_{n-1}(x)dx \). The distribution has mean zero and variance \((n-1)/(n-3)\). As \( n \to \infty \), the distribution and the random variable converge to standard normal, which is written \( t_{n-1} \overset{d}{\to} \text{N}[0,1] \).
A Slutsky Theorem for Random Variables (Continuous Mapping Theorem)

If $x_n \xrightarrow{d} x$, and if $g(x_n)$ is a continuous function with continuous derivatives and does not involve $n$, then $g(x_n) \xrightarrow{d} g(x)$.

Example: $t_n$ = random variable with $t$ distribution with $n$ degrees of freedom.

$t_n^2$ = exactly, an $F$ random variable with $[1,n]$ degrees of freedom.

$t_n \xrightarrow{d} N(0,1),$

$t_n^2 \xrightarrow{d} [N(0,1)]^2 = \text{chi-squared}[1].$
An Extension of the Slutsky Theorem

If \( x_n \xrightarrow{d} x \) (\( x_n \) has a limiting distribution) and \( \theta \) is some relevant constant (estimator), and \( g(x_n, \theta) \xrightarrow{d} g \) (i.e., \( g_n \) has a limiting distribution that is some function of \( \theta \)) and \( \text{plim} \ \hat{\theta}_n = \theta \), then \( g(x_n, \hat{\theta}_n) \xrightarrow{d} g(x_n, \theta) \) (replacing \( \theta \) with a consistent estimator leads to the same limiting distribution).
Application of the Slutsky Theorem

Large sample behavior of the F statistic for testing restrictions

\[ F = \frac{(e^*e^* - e'e)}{J} \frac{(e^*e^* - e'e)}{J\sigma^2} \xrightarrow{d} \frac{\chi^2[J]}{J} \]

\[ \frac{\hat{\sigma}^2}{\sigma^2} \xrightarrow{p} 1 \]

Therefore, \( \frac{JF}{\sigma^2} \xrightarrow{d} \chi^2[J] \) as \( N \) increases

Establishing the numerator requires a central limit theorem.
We will come to that shortly.
Central Limit Theorems

Central Limit Theorems describe the large sample behavior of random variables that involve sums of variables. “Tendency toward normality.”

Generality: When you find sums of random variables, the CLT shows up eventually.

The CLT does not state that means of samples have normal distributions.
A Central Limit Theorem

Lindeberg-Levy CLT (the simplest version of the CLT)
If $x_1,\ldots,x_n$ are a random sample from a population with finite mean $\mu$ and finite variance $\sigma^2$, then

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

Note, not the limiting distribution of the mean, since the mean, itself, converges to a constant.

A useful corollary: if $\text{plim} \ s_n = \sigma$, and the other conditions are met, then

$$
\frac{\sqrt{n}(\bar{x} - \mu)}{s_n} \xrightarrow{d} N(0,1)
$$

Note this does not assume sampling from a normal population.
Lindeberg-Levy vs. Lindeberg-Feller

Lindeberg-Levy assumes random sampling – observations have the same mean and same variance.

Lindeberg-Feller allows variances to differ across observations, with some necessary assumptions about how they vary.

Most econometric estimators require Lindeberg-Feller (and extensions such as Lyapunov).
Order of a Sequence

Order of a sequence
‘Little oh’ o(.). Sequence \( h_n \) is \( o(n^\delta) \) (order \textit{less than} \( n^\delta \)) iff \( n^{-\delta} h_n \to 0 \).

Example: \( h_n = n^{1.4} \) is \( o(n^{1.5}) \) since \( n^{-1.5} h_n = 1/n^{1.1} \to 0 \).

‘Big oh’ \( O(.) \). Sequence \( h_n \) is \( O(n^\delta) \) iff \( n^{-\delta} h_n \to \) a finite nonzero constant.

Example 1: \( h_n = (n^2 + 2n + 1) \) is \( O(n^2) \).

Example 2: \( \sum_i x_i^2 \) is usually \( O(n^1) \) since this is \( n \times \) the mean of \( x_i^2 \)
and the mean of \( x_i^2 \) generally converges to \( E[x_i^2] \), a finite constant.

What if the sequence is a random variable? The order is in terms of the variance.

Example: What is the order of the sequence \( \bar{X}_n \) in random sampling?
\[ \text{Var}[\bar{X}_n] = \sigma^2/n \text{ which is } O(1/n). \] Most estimators are \( O(1/n) \)
Cornwell and Rupert Panel Data

Cornwell and Rupert Returns to Schooling Data, 595 Individuals, 7 Years

Variables in the file are

- EXP = work experience
- WKS = weeks worked
- OCC = occupation, 1 if blue collar,
- IND = 1 if manufacturing industry
- SOUTH = 1 if resides in south
- SMSA = 1 if resides in a city (SMSA)
- MS = 1 if married
- FEM = 1 if female
- UNION = 1 if wage set by union contract
- ED = years of education
- LWAGE = log of wage = dependent variable in regressions

Histogram for LWAGE
Kernel Estimator for LWAGE

\[ \hat{f}(x^*) \]
Kernel Density Estimator

The curse of dimensionality

\[ \hat{f}(x^*_m) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{B} K \left( \frac{x_i - x^*_m}{B} \right), \text{ for a set of points } x^*_m \]

B = "bandwidth"

K = the kernel function

x* = the point at which the density is approximated.

\( \hat{f}(x^*) \) is an estimator of \( f(x^*) \)

\[ \frac{1}{n} \sum_{i=1}^{n} Q(x_i \mid x^*) = \overline{Q}(x^*). \]

But, \( \text{Var}[\overline{Q}(x^*)] \neq \frac{1}{n} \times \text{Something} \). Rather, \( \text{Var}[\overline{Q}(x^*)] = \frac{1}{n^{3/5}} \times \text{Something} \)

i.e., \( \hat{f}(x^*) \) does not converge to \( f(x^*) \) at the same rate as a mean converges to a population mean.
Asymptotic Distributions

An asymptotic distribution is a finite sample approximation to the true distribution of a random variable that is good for large samples, but not necessarily for small samples.

Stabilizing transformation to obtain a limiting distribution. Multiply random variable $x_n$ by some power, $a$, of $n$ such that the limiting distribution of $n^ax_n$ has a finite, nonzero variance.

Example, $\bar{X}_n$ has a limiting variance of zero, since the variance is $\sigma^2/n$. But, the variance of $\sqrt{n} \ X_n$ is $\sigma^2$. However, this does not stabilize the distribution because $E[\sqrt{n} \ X]\neq \sqrt{n}\mu$.

The stabilizing transformation would be $\sqrt{n}(\bar{X} - \mu)$
Asymptotic Distribution

Obtaining an asymptotic distribution from a limiting distribution

Obtain the limiting distribution via a stabilizing transformation

Assume the limiting distribution applies reasonably well in finite samples

Invert the stabilizing transformation to obtain the asymptotic distribution

\[
\sqrt{n}(\bar{x} - \mu) / \sigma \xrightarrow{d} N[0,1]
\]

Assume holds in finite samples. Then,

\[
\sqrt{n}(\bar{x} - \mu) \xrightarrow{a} N[0, \sigma^2]
\]

\[
(\bar{x} - \mu) \xrightarrow{a} N[0, \sigma^2 / n]
\]

\[
\bar{x} \xrightarrow{a} N[\mu, \sigma^2 / n]
\]

Asymptotic distribution.

\[
\sigma^2 / n = \text{the asymptotic variance.}
\]

Asymptotic normality of a distribution.
Asymptotic Efficiency

- Comparison of asymptotic variances
- How to compare consistent estimators? If both converge to constants, both variances go to zero.
  - Example: Random sampling from the normal distribution,
    - Sample mean is asymptotically normal $[\mu, \sigma^2/n]$
    - Median is asymptotically normal $[\mu, (\pi/2)\sigma^2/n]$
    - Mean is asymptotically more efficient
The Delta Method

The **delta method** (combines most of these concepts)

**Nonlinear transformation of a random variable:**

$f(x_n)$ such that $\text{plim } x_n = \mu$ but $\sqrt{n} (x_n - \mu)$ is asymptotically normally distributed $(\mu, \sigma^2)$. What is the asymptotic behavior of $f(x_n)$?

**Taylor series approximation:** $f(x_n) \approx f(\mu) + f'(\mu) (x_n - \mu)$

**By the Slutsky theorem,**

$\text{plim } f(x_n) = f(\mu)$

$\sqrt{n}[f(x_n) - f(\mu)] \approx f'(\mu) [\sqrt{n} (x_n - \mu)]$

$\sqrt{n}[f(x_n) - f(\mu)] \rightarrow f'(\mu) \times N[\mu, \sigma^2]$

Large sample behaviors of the LHS and RHS sides are the same.
Large sample variance is $[f'(\mu)]^2$ times large sample $\text{Var}[\sqrt{n} (x_n - \mu)]$.
Delta Method
Asymptotic Distribution of a Function

If \( x_n \xrightarrow{a} N[\mu, \sigma^2 / n] \) and \( f(x_n) \) is a continuous and continuously differentiable function that does not involve \( n \), then
\[
\begin{align*}
f(x_n) & \xrightarrow{a} N\{f(\mu), [f'(\mu)]^2 \sigma^2 / n\}
\end{align*}
\]
Does SNAP improve your health?®

Christian A. Gregory, Partha Deb

a Diet, Safety and Health Economics Branch, Food Economics Division, Economic Research Service, USDA, Washington DC, United States
b Dept. of Economics, Hunter College, City University of New York, New York, United States


<table>
<thead>
<tr>
<th>Table 2</th>
<th>Parameter estimates from ordered and count models.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SAH</td>
</tr>
<tr>
<td>Female</td>
<td>0.034 (0.021)</td>
</tr>
<tr>
<td>Black</td>
<td>0.346* (0.028)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>−0.018 (0.029)</td>
</tr>
<tr>
<td>Other Race</td>
<td>0.021 * (0.051)</td>
</tr>
<tr>
<td>Married</td>
<td>−0.217*** (0.024)</td>
</tr>
</tbody>
</table>

One Vehicle Exempt per Adult

\[ \text{tanh}(\rho) / \lambda \]

<table>
<thead>
<tr>
<th>ln(\delta)</th>
<th>0.305*** (0.047)</th>
</tr>
</thead>
<tbody>
<tr>
<td>\chi^2_n</td>
<td>17.87*** (0.000)</td>
</tr>
</tbody>
</table>
The parameters $\rho$ and $\lambda$ represent the different measures of correlation between the unobservables in the selection equation and the outcome equation for self-assessed health and the count outcomes, respectively. The value of the parameter $\rho$—the correlation between bivariate normal errors in the two equations—indicates that SNAP participants are more likely to report worse health “before” entering SNAP—that is, selection is adverse rather than beneficial. This parameter is highly statistically significant. The
**Delta Method**

Author (Stata) reports \( \text{atanh}(\rho) = 0.305 \) (0.047). Note a typo in the paper. The label in the table of results is \( \tanh(\rho) \). (Hyperbolic tangent.) Stata actually reports \( \text{atanh}(\rho) \), the hyperbolic \text{arctangent}. The difference is substantive, but this is an obvious typo. The estimate of \( \rho \) is never reported. Is the claim true? We use the delta method to find out. Write \( \tau = \text{atanh}(\rho) \). This function is

\[
\tau = (1/2)\ln[(1 + \rho)/(1 - \rho)].
\]

You can solve this for

\[
\rho = [\exp(2\tau) - 1] / [\exp(2\tau) + 1]
\]

So, plugging in the value of \( \tau \) (0.305), we get the estimate of \( \rho \), 0.296.
To get the estimated standard error, we need $|d\rho/d\tau|$ times the estimated standard error of $\tau$ (which is 0.047). Doing the differentiation the hard way,

$$d\rho/d\tau = \frac{[\exp(2\tau) + 1]2\exp(2\tau) - [\exp(2\tau) - 1] 2(\exp(2\tau))}{[\exp(2\tau) + 1]^2}$$

$$= \frac{4\exp(2\tau)}{[\exp(2\tau) + 1]^2},$$

which evaluates to 0.912.

Finishing, the estimated standard error for the estimator of $\rho$ is $0.912 \times 0.047 = 0.043$. So, the claim is correct; the estimate of $\rho$ is statistically significant; $0.296/0.043 = 6.88 > 1.96$. 


Delta Method – More than One Parameter

If \( \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k \) are \( k \) consistent estimators of \( k \) parameters \( \theta_1, \theta_2, \ldots, \theta_k \)

with asymptotic covariance matrix

\[
V = \begin{bmatrix}
v_{11} & v_{12} & \cdots & v_{1k} \\
v_{21} & v_{22} & \cdots & v_{2k} \\
\vdots & \vdots & \ddots & \vdots \\
v_{k1} & v_{k2} & \cdots & v_{kk}
\end{bmatrix},
\]

and if \( f(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k) \) = a continuous function with continuous derivatives,

then the asymptotic variance of \( f(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k) \) is

\[
g'Vg = \begin{bmatrix}
\frac{\partial f(\cdot)}{\partial \theta_1} & \frac{\partial f(\cdot)}{\partial \theta_2} & \cdots & \frac{\partial f(\cdot)}{\partial \theta_k}
\end{bmatrix}
\begin{bmatrix}
v_{11} & v_{12} & \cdots & v_{1k} \\
v_{21} & v_{22} & \cdots & v_{2k} \\
\vdots & \vdots & \ddots & \vdots \\
v_{k1} & v_{k2} & \cdots & v_{kk}
\end{bmatrix}
\begin{bmatrix}
\frac{\partial f(\cdot)}{\partial \theta_1} \\
\frac{\partial f(\cdot)}{\partial \theta_2} \\
\vdots \\
\frac{\partial f(\cdot)}{\partial \theta_k}
\end{bmatrix}
\]

\[
= \sum_{k=1}^{K} \sum_{l=1}^{K} \frac{\partial f(\cdot)}{\partial \theta_k} \frac{\partial f(\cdot)}{\partial \theta_l} V_{kl}
\]
Log Income Equation

Ordinary least squares regression ............
LHS=LOGY
Mean = -1.15746
Standard deviation = .49149
Number of observs. = 27322
Model size Parameters = 7
Degrees of freedom = 27315
Residuals Sum of squares = 5462.03686
Standard error of e = .44717
Fit R-squared = .17237

| Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z] | Mean of X |
|----------|-------------|----------------|---------|----------|-----------|
| AGE      | .06225***   | .00213         | 29.189  | .0000    | 43.5272   |
| AGESQRD  | -.00074***  | .0242482D-04   | -30.576 | .0000    | 2022.99   |
| Constant | -3.19130*** | .04567         | -69.884 | .0000    |           |
| MARRIED  | .32153***   | .00703         | 45.767  | .0000    | .75869    |
| HHKIDS   | -.11134***  | .00655         | -17.002 | .0000    | .40272    |
| FEMALE   | -.00491     | .00552         | -.889   | .3739    | .47881    |
| EDUC     | .05542***   | .00120         | 46.050  | .0000    | 11.3202   |
Age-Income Profile:
Married=1, Kids=1, Educ=12, Female=1
Application: Maximum of a Function

\[ \log Y = \beta_1 \text{Age} + \beta_2 \text{Age}^2 + \ldots \]

At what age does \( \log \text{income} \) reach its maximum?

\[ \frac{\partial \log Y}{\partial \text{Age}} = \beta_1 + 2\beta_2 \text{Age} = 0 \Rightarrow \text{Age}^* = \frac{-\beta_1}{2\beta_2} = \frac{-0.06225}{2(-0.00074)} = 42.1 \]

\[ \frac{\partial \text{Age}^*}{\partial \beta_1} = \frac{-1}{2\beta_2} = g_1 = \frac{-1}{2(-0.00074)} = 675.68 \]

\[ \frac{\partial \text{Age}^*}{\partial \beta_2} = \frac{\beta_1}{2\beta_2} = g_2 = \frac{0.06225}{2(-0.00074)^2} = 56838.9 \]
Delta Method Using Visible Digits

\[
\begin{array}{c|c|c}
1 & 2 \\
\hline
1 & 4.54799\times10^{-6} & -5.1285\times10^{-8} \\
2 & -5.1285\times10^{-8} & 5.87973\times10^{-10} \\
3 & 9.8993\times10^{-8} & 9.91407\times10^{-10} \\
\end{array}
\]

\[
675.68^2(4.54799\times10^{-6}) + 56838.9^2(5.8797\times10^{-10}) + 2(675.68)(56838.9)(-5.1285\times10^{-8})
\]

\[
= 0.0366952
\]

standard error = square root = \(0.1915599\)
### Delta Method Results Built into Software

WALD procedure.

| Variable | Coefficient | Standard Error | b/St.Er. | P[|Z|>z]     |
|----------|-------------|----------------|---------|-------------|
| G1       | 674.399***  | 22.05686       | 30.575  | .0000       |
| G2       | 56623.8***  | 1797.294       | 31.505  | .0000       |
| AGESTAR  | 41.9809***  | .19193         | 218.727 | .0000       |

(Computed using all 17 internal digits of regression results)
Application: Doctor Visits

- German Individual Health Care data: n=27,236
- Simple model for number of visits to the doctor:
  - True $E[v|\text{income}] = \exp(1.412 - .0745*\text{income})$
  - Linear regression: $g^*(\text{income})=3.917 - .208*\text{income}$
A Nonlinear Model

\[ E[\text{docvis} \mid x] = \exp(\beta_1 + \beta_2 \text{AGE} + \beta_3 \text{EDUC} \ldots) \]
Interesting Partial Effects

Estimate Effects at the Means of the Data

\[ \hat{E}[\text{docvis} | x] = \exp(b_1 + b_2 \overline{AGE} + b_3 \overline{EDUC} + \ldots) \]

\[ \frac{\partial \hat{E}[\text{docvis} | x]}{\partial \text{AGE}} = \exp(b_1 + b_2 \overline{AGE} + b_3 \overline{EDUC} + \ldots)b_2 \]

\[ \frac{\partial \hat{E}[\text{docvis} | x]}{\partial \text{EDUC}} = \exp(b_1 + b_2 \overline{AGE} + b_3 \overline{EDUC} + \ldots)b_3 \]

\[ \ldots \]

\[ \frac{\partial \hat{E}[\text{docvis} | x]}{\partial \text{INCOME}} = \exp(b_1 + b_2 \overline{AGE} + b_3 \overline{EDUC} + \ldots)b_6 \]
Necessary Derivatives (Jacobian)

$$\frac{\partial \hat{E}[docvis | x]}{\partial \text{AGE}} = \exp(b_1 + b_2 \text{AGE} + b_3 \text{EDUC}...)b_2 = f_{\text{AGE}}(b_1, b_2, ..., b_6 | \text{AGE}, \text{EDUC})$$

$$\frac{\partial f_{\text{AGE}}}{\partial b_1} = \frac{\partial b_2 \exp(b_1 + b_2 \text{AGE} + b_3 \text{EDUC}...)}{\partial b_1} = b_2 \exp(...) \times 1$$

$$\frac{\partial f_{\text{AGE}}}{\partial b_2} = b_2 \frac{\partial \exp(\ldots)}{\partial b_2} + \exp(\ldots) \frac{\partial b_2}{\partial b_2} = b_2 \exp(...) \times \text{AGE} + \exp(...) 1$$

$$\frac{\partial f_{\text{AGE}}}{\partial b_3} = \frac{\partial b_2 \exp(b_1 + b_2 \text{AGE} + b_3 \text{EDUC}...)}{\partial b_3} = b_2 \exp(...) \times \text{EDUC}$$

$$\frac{\partial f_{\text{AGE}}}{\partial b_4} = \frac{\partial b_2 \exp(b_1 + b_2 \text{AGE} + b_3 \text{EDUC}...)}{\partial b_4} = b_2 \exp(...) \times \text{MARRIED}$$

$$\frac{\partial f_{\text{AGE}}}{\partial b_5} = \frac{\partial b_2 \exp(b_1 + b_2 \text{AGE} + b_3 \text{EDUC}...)}{\partial b_5} = b_2 \exp(...) \times \text{FEMALE}$$

$$\frac{\partial f_{\text{AGE}}}{\partial b_6} = \frac{\partial b_2 \exp(b_1 + b_2 \text{AGE} + b_3 \text{EDUC}...)}{\partial b_6} = b_2 \exp(...) \times \text{HHNINC}$$
```r
|-> simulate ; if [year=1994] ; means $ \\
\hline
Model Simulation Analysis for Exponential Regression Function \\
Simulations are computed at sample means of all variables \\
\hline
User Function | Function | Standard Error | |t| 95% Confidence Interval \\
(Delta method) | Value    |              |       |      |          |
\hline
Func. at means | 3.54795  | .03334       | 106.40| 3.48259| 3.61330 |
\hline
|-> partials ; if[year=1994] ; effects: age ; means $ \\
\hline
Partial Effects Analysis for Exponential Regression Function \\
Effects on function with respect to AGE \\
Results are computed at sample means of all variables \\
Partial effects for continuous AGE computed by differentiation \\
Effect is computed as derivative = df(.)/dx \\
\hline
df/dAGE | Partial Effect | Standard Error | |t| 95% Confidence Interval \\
(Delta method) |          |              |       |      |          |
\hline
PE.Func(means) | .07116   | .00274       | 25.93 | .06578| .07654 |
```
Partial Effects at Means vs. Mean of Partial Effects

Partial Effects at the Means

$$\delta(\beta, \bar{x}) = \frac{\partial f(\beta \mid \bar{x})}{\partial \bar{x}} = \frac{\partial f\left(\beta \mid \frac{1}{n} \sum_{i=1}^{n} x_i\right)}{\partial \bar{x}}$$

Mean of Partial Effects

$$\bar{\delta}(\beta, X) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial f(\beta \mid x_i)}{\partial x_i}$$

Makes more sense for dummy variables, d:

$$\Delta_i(\beta, x_i, d) = f(\beta \mid x_i, d=1) - f(\beta \mid x_i, d=0)$$

$$\bar{\Delta}(\beta, X, d)$$ makes more sense than $$\delta(\beta, \bar{x}, \bar{d})$$
### Partial Effect for a Dummy Variable?

```text
<table>
<thead>
<tr>
<th>Partial Effects Analysis for Exponential Regression Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effects on function with respect to FEMALE</td>
</tr>
</tbody>
</table>

Results are computed by average over sample observations

Partial effects for binary var FEMALE computed by first difference

| df/dFEMALE (Delta method) | Partial Effect | Standard Error | |t| 95% Confidence Interval |
|---------------------------|----------------|----------------|---|-------------------------|
| APE. Function             | 1.50212        | 0.06856        | 21.91 | 1.36775 | 1.63649 |
```

```text
<table>
<thead>
<tr>
<th>Partial Effects Analysis for Exponential Regression Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effects on function with respect to FEMALE</td>
</tr>
</tbody>
</table>

Results are computed at sample means of all variables

Partial effects for continuous FEMALE computed by differentiation

Effect is computed as derivative $\frac{df(.)}{dx}$

| df/dFEMALE (Delta method) | Partial Effect | Standard Error | |t| 95% Confidence Interval |
|---------------------------|----------------|----------------|---|-------------------------|
| PE.Func(means)            | 1.41696        | 0.06347        | 22.32 | 1.29256 | 1.54136 |
```
### Partial Effects for Exponential Regression Function

Partial Effects Averaged Over Observations

* => Partial Effect for a Binary Variable

| (Delta method) | Partial Effect | Standard Error | |t| | 95% Confidence Interval |
|----------------|----------------|----------------|-----|------------------------|
| AGE            | .06549         | .00100         | 65.61 | .06353 | .06745 |
| EDUC           | -.09123        | .00552         | 16.52 | -.10205 | -.08041 |
| INCOME         | -1.68502       | .06996         | 24.09 | -1.82213 | -1.54791 |
| * FEMALE       | .93019         | .02210         | 42.10 | .88688 | .97350 |
Delta Method, Stata Application

Target of estimation is \( \rho = \frac{\sigma^2}{1 + \sigma^2} \)

Estimation strategy:

1. Estimate \( \alpha = \log \sigma^2 \)
2. Estimate \( \sigma = \exp(\alpha/2) \)
3. Estimate \( \rho = \frac{\sigma^2}{1 + \sigma^2} \)
Delta Method

\[
\hat{\alpha} = -1.123706 \quad V_\hat{\alpha} = .0715548^2 = .00512009
\]

\[
\hat{\sigma} = \exp(\hat{\alpha} / 2) = \exp(-1.123706 / 2) = \exp(-.561853) = .5701517
\]

\[
\hat{g} = d\hat{\sigma} / d\hat{\alpha} = \frac{1}{2} \exp(\hat{\alpha} / 2) = \frac{1}{2} \hat{\sigma} = .2850758
\]

\[
\hat{g}^2 = (d\hat{\sigma} / d\hat{\alpha})^2 = .08126821
\]

\[
(d\hat{\sigma} / d\hat{\alpha})^2 V_\hat{\alpha} = .08126821(.00512009) = .0004161
\]

Estimated Standard Error for \( \hat{\sigma} = \sqrt{.0004161} = .02039854 \)
Continuing the previous example, there are two approaches implied for estimating $\rho$:

(1) $\rho = f(\sigma) = \frac{\sigma^2}{1 + \sigma^2}$

(2) $\rho = h(\alpha) = \frac{\exp(\alpha)}{1 + \exp(\alpha)}$

Use the delta method to estimate a standard error for each of the two estimators of $\rho$. Do you obtain the same answer?
Confidence Intervals?

The center of the confidence interval given in the table is .571554!
What is going on here?
The confidence limits given are \( \exp(-1.23695/2) \) to \( \exp(-.984361/2) \)!

| COOPERACA | Coef.    | Std. Err. | z       | P>|z| | [95% Conf. Interval] |
|-----------|----------|-----------|---------|-----|---------------------|
| /lnsig2u  | -1.123706| .0715548  |         |     | -1.26395 to -.983461|
| sigma_u   | .5701517 | .0203985  |         |     | .5315408 to .6115672|
| rho       | .2453245 | .0132477  |         |     | .2202946 to .2722056|

\( \hat{\sigma} \in .5701517 \pm 1.96(.0203985) = .5301707 \text{ to } .6101328 \)!!
Received October 6, 2012

Dear Prof. Greene,
I am AAAAAA, an assistant professor of Finance at the xxxxx university of xxxxx, xxxxx. I would be grateful if you could answer my question regarding the parameter estimates and the marginal effects in Multinomial Logit (MNL).
After running my estimations, the parameter estimate of my variable of interest is statistically significant, but its marginal effect, evaluated at the mean of the explanatory variables, is not. Can I just rely on the parameter estimates’ results to say that the variable of interest is statistically significant? How can I reconcile the parameter estimates and the marginal effects’ results?
Thank you very much in advance!
Best,
AAAAAAA