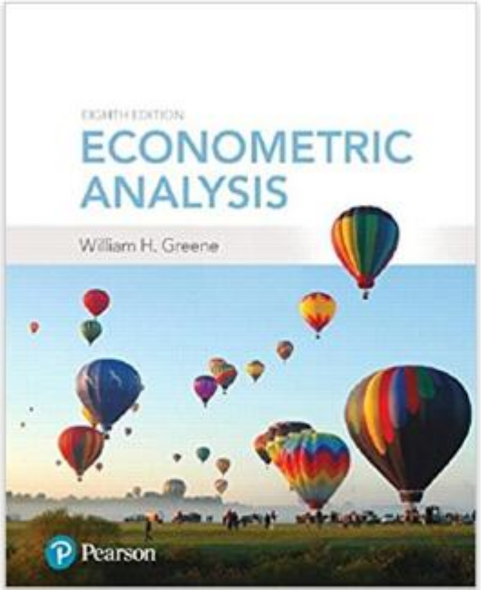
Econometrics I

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:

$$\text{Ordinary limit: } \frac{n(n+1)/2+3n+5}{n^2+2n+1} = \frac{\frac{1}{2}n^2+3\frac{1}{2}n+5}{n^2+2n+1} \xrightarrow{\quad} ? \quad \frac{1}{2}$$

(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.)

$$\overline{x}_n = \tfrac{1}{n}\Sigma_{i=1}^{n} x_i,\ E[\overline{x}_n] = \mu \to \mu,\ \text{Var}[\overline{x}_n] = \sigma^2 / n \to 0$$

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.

$(1/n)\Sigma_i g(z_i)$ converges to $E[g(z_i)]$.

Extending the Law of Large Numbers

Suppose x has mean μ and finite variance σ^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 $Var[z] = E[x^{2P}] - \{E[x^P]\}^2$. The LLN applies to $\bar{z}$ 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

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

$(P-1)!!$ = product of odd numbers up to P-1.

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

Probability Limit

Let θ be a constant, ε 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 θ. (A definition.)

In words, the probability that the difference between b_n and θ is larger than ε for any ε goes to zero. b_n becomes arbitrarily close to θ.

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

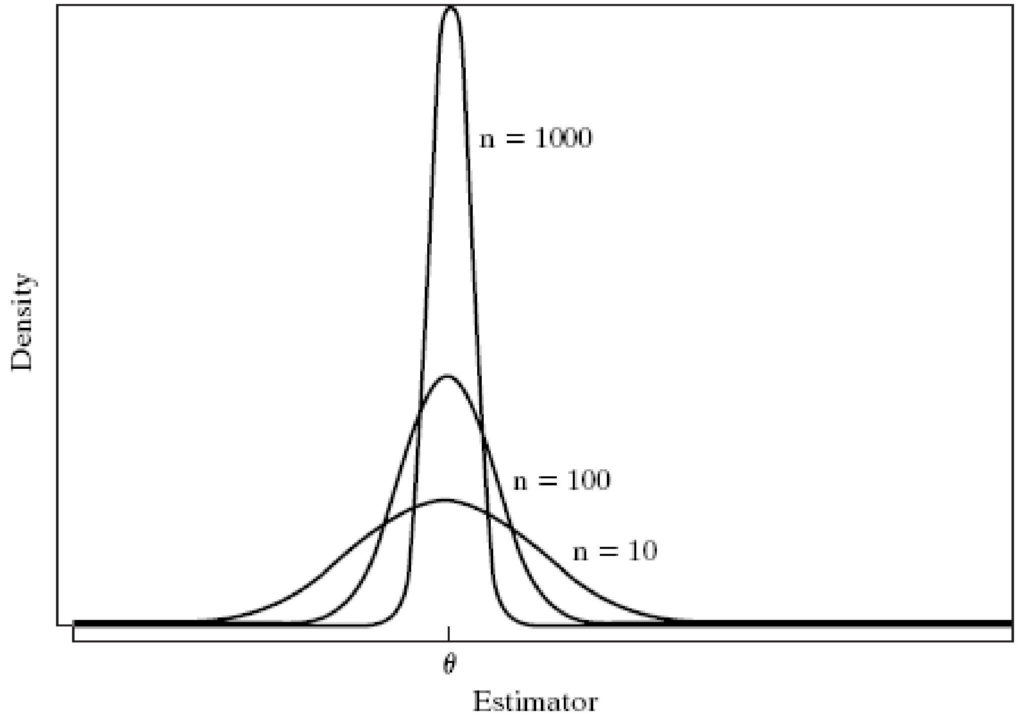


FIGURE D.1 Quadratic Convergence to a Constant, θ.

Probability Limits and Expecations

What is the difference between $E[b_n]$ and 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 θ.

Estimators can be inconsistent for θ 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 plim $b_n = \theta$.

For now, we assume θ is a constant.

g(.) is a continuous function with continuous derivatives.

g(.) is not a function of n.

Conclusion: Then plim[g(b_n)] = g[plim(b_n)] assuming g[plim(b_n)] exists. (VVIR!)

Works for probability limits. Does not work for expectations.

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

Slutsky Corollaries

x_n and y_n are two sequences of random variables with probability limits θ and μ.

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 plim$\mathbf{A}_n$ = $\mathbf{A}$ and plim$\mathbf{B}_n$ = $\mathbf{B}$ (element by element), then

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

and

$$\text{plim}(\mathbf{A}_n\mathbf{B}_n) = \text{plim}\mathbf{A}_n \text{plim } \mathbf{B}_n = \mathbf{AB}$$

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 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

$$x_n \xrightarrow{d} x \iff F_n(x_n) \xrightarrow[n\to\infty]{} F(x)$$

A Limiting Distribution

$x_1, x_2, ..., x_n$ = a random sample from $N[\mu, \sigma^2]$

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

$$t_{n-1} = \frac{\bar{x}_n}{\left(s_n / \sqrt{n}\right)}, \text{ where } s_n^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}_n\right)^2}{n-1}$$

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

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

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 $n \rightarrow \infty$, the distribution and the random variable converge to standard normal, which is written $t_{n-1} \xrightarrow{d} N[0,1]$.

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.

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

$t_n^2 \xrightarrow{d} [N(0,1)]^2$ = chi-squared[1].

An Extension of the Slutsky Theorem

If $x_n \xrightarrow{d} x$ (x_n has a limiting distribution) and

θ 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 θ)

and $\text{plim } \hat{\theta}_n = \theta$, then $g(x_n, \hat{\theta}_n) \xrightarrow{d} g(x_n, \theta)$

(replacing θ 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

$$F=\frac{(\mathbf{e}^{*\prime}\mathbf{e}^{*}-\mathbf{e}'\mathbf{e})/J}{\mathbf{e}'\mathbf{e}/(N-K)}=\frac{\left.\dfrac{(\mathbf{e}^{*\prime}\mathbf{e}^{*}-\mathbf{e}'\mathbf{e})}{J\sigma^2}\right\}\xrightarrow{d}\dfrac{\chi^2[J]}{J}}{\left.\dfrac{\hat{\sigma}^2}{\sigma^2}\right\}\xrightarrow{p}1}$$

Therefore, $JF\xrightarrow{d}\chi^2[J]$ 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 $x_1,...,x_n$ are a random sample from a population with finite mean μ and finite variance σ^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 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 *less than* n^δ) iff $n^{-\delta} h_n \to 0$.

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

'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: $\Sigma_i x_i^2$ 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?

$\text{Var}[\bar{x}_n] = \sigma^2/n$ which is $O(1/n)$. Most estimators are $O(1/n)$

Cornwell and Rupert Panel Data

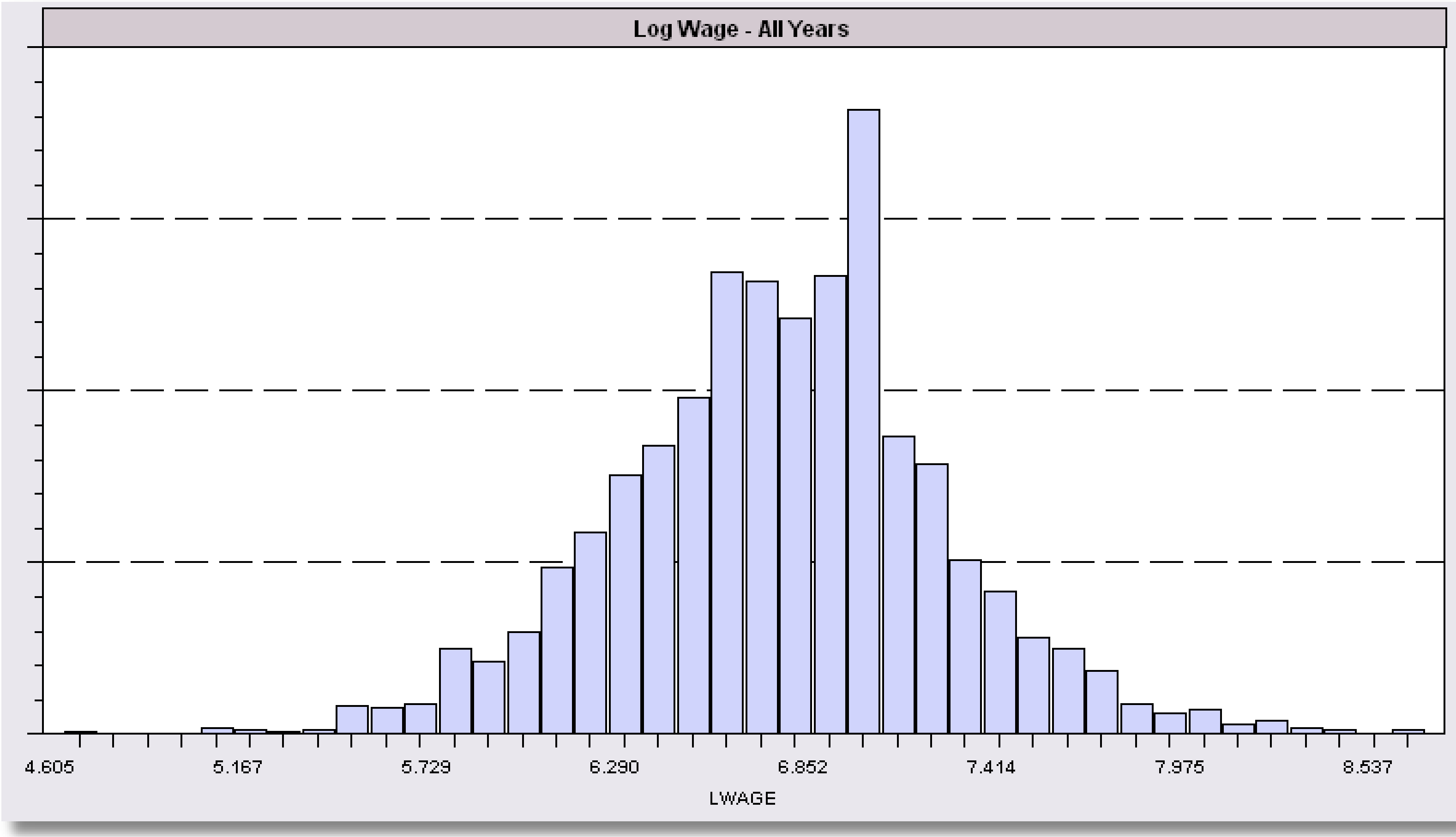
Cornwell and Rupert Returns to Schooling Data, 595 Individuals, 7 Years
Variables in the file are

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

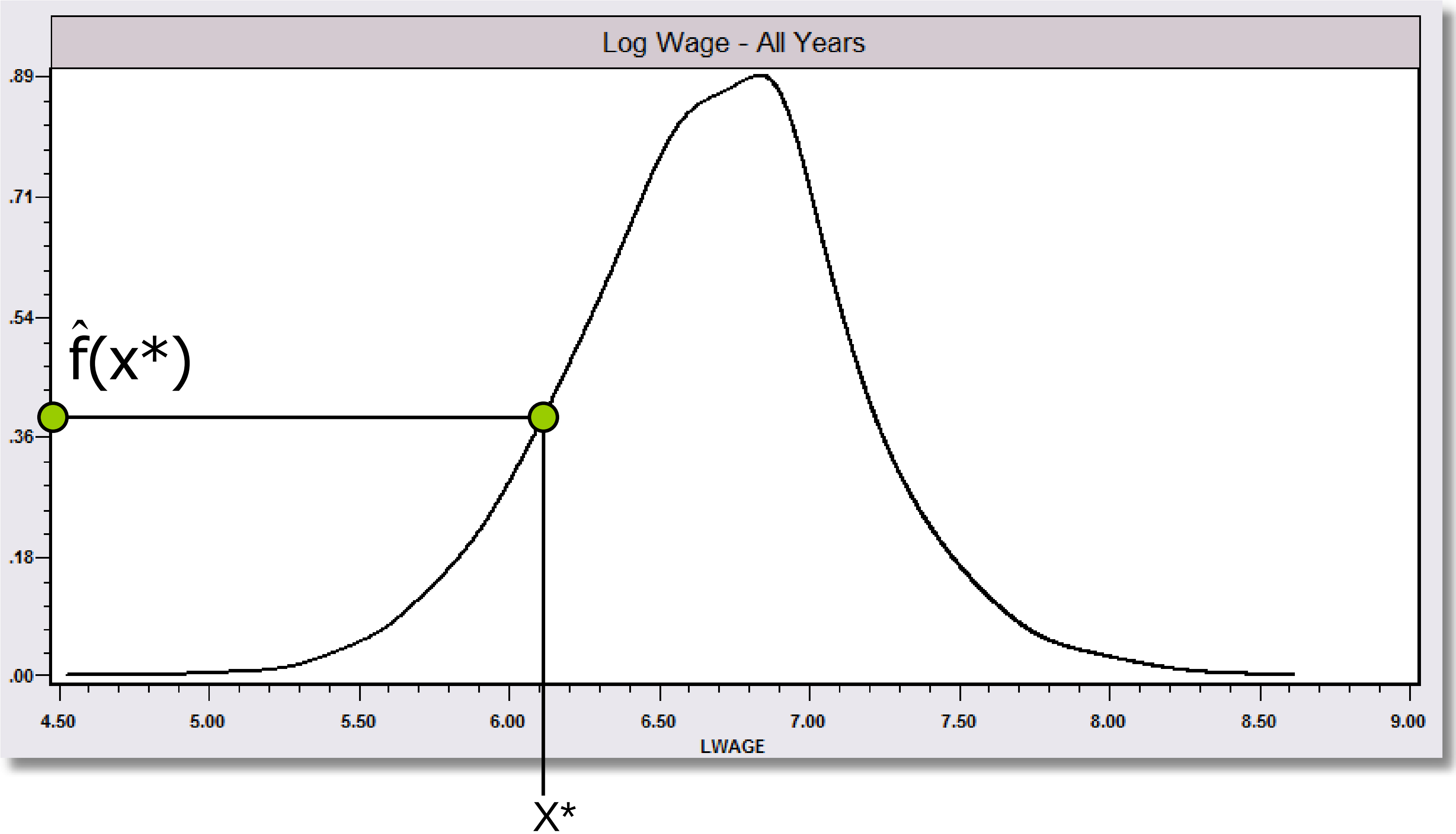
Variable	Mean	Std.Dev.
EXP	19.85378	10.96637
WKS	46.81152	5.129098
OCC	.511164	.499935
IND	.395438	.489003
SOUTH	.290276	.453944
SMSA	.653782	.475821
MS	.814406	.388826
FEM	.112605	.316147
UNION	.363986	.481202
LWAGE	6.676346	.461512
YEAR	4.0	2.000240

These data were analyzed in Cornwell, C. and Rupert, P., "Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variable Estimators," Journal of Applied Econometrics, 3, 1988, pp. 149-155.

Histogram for LWAGE



Kernel Estimator for LWAGE



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 $f(x^*)$

$$\frac{1}{n}\sum_{i=1}^{n}Q(x_i \mid x^*) = \bar{Q}(x^*).$$

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

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 $n^a x_n$ has a finite, nonzero variance.

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

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.

$\sigma^2/n =$ 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 plim $x_n = \mu$ but $\sqrt{n}\,(x_n - \mu)$ is asymptotically normally distributed (μ, σ^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, 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 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

$f(x_n) \xrightarrow{a} N\{f(\mu), [f'(\mu)]^2 \sigma^2/n\}$

Food Policy 50 (2015) 11–19

ELSEVIER

Contents lists available at ScienceDirect

Food Policy

journal homepage: www.elsevier.com/locate/foodpol

FOOD POLICY

Does SNAP improve your health? ☆

CrossMark

Christian A. Gregory [a,*], Partha Deb [b,c]

[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 2
Parameter estimates from ordered and count models.

	SAH
Female	0.034 (0.021)
Black	0.346*** (0.028)
Hispanic	−0.018 (0.029)
Other Race	0.021 * (0.051)
Married	−0.217*** (0.024)

One Vehicle Exempt per Adult	0.116** (0.049)
$tanh(\rho)$ / λ	0.305*** (.047)
$ln(\delta)$	
χ^2_{IV}	17.87*** (0.000)

tanh(ρ)	0.305*** (.047)

The parameters ρ and λ 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 ρ–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 atanh(ρ) = 0.305 (0.047). Note a typo in the paper. The label in the table of results is tanh(ρ). (Hyperbolic tangent.) Stata actually reports atanh(ρ), the hyperbolic arctangent. The difference is substantive, but this is an obvious typo.

The estimate of ρ is never reported. Is the claim true? We use the

delta method to find out. Write τ = atanh(ρ). 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 τ (0.305), we get the estimate of ρ, 0.296.

To get the estimated standard error, we need $|d\rho/d\tau|$ times the estimated standard error of τ (which is 0.047). Doing the differentiation the hard way,

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

$$= 4\exp(2\tau)/[\exp(2\tau)+1]^2,$$ which evaluates to 0.912.

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

Delta Method – More than One Parameter

If $\hat{\theta}_1, \hat{\theta}_1, ..., \hat{\theta}_K$ are K consistent estimators of K parameters $\theta_1, \theta_2, ..., \theta_K$

with asymptotic covariance matrix $\mathbf{V} = \begin{bmatrix} v_{11} & v_{12} & ... & v_{1K} \\ v_{21} & v_{22} & ... & v_{2K} \\ ... & ... & ... & ... \\ v_{K1} & v_{K2} & ... & v_{KK} \end{bmatrix}$,

and if $f(\hat{\theta}_1, \hat{\theta}_1, ..., \hat{\theta}_K)$ = a continuous function with continuous derivatives,

then the asymptotic variance of $f(\hat{\theta}_1, \hat{\theta}_1, ..., \hat{\theta}_K)$ is

$$\mathbf{g'Vg} = \begin{bmatrix} \frac{\partial f(.)}{\partial \theta_1} & \frac{\partial f(.)}{\partial \theta_2} & ... & \frac{\partial f(.)}{\partial \theta_K} \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} & ... & v_{1K} \\ v_{21} & v_{22} & ... & v_{2K} \\ ... & ... & ... & ... \\ v_{K1} & v_{K2} & ... & v_{KK} \end{bmatrix} \begin{bmatrix} \frac{\partial f(.)}{\partial \theta_1} \\ \frac{\partial f(.)}{\partial \theta_2} \\ ... \\ \frac{\partial f(.)}{\partial \theta_K} \end{bmatrix}$$

$$= \sum_{k=1}^{K} \sum_{l=1}^{K} \frac{\partial f(.)}{\partial \theta_k} \frac{\partial f(.)}{\partial \theta_l} V_{kl}$$

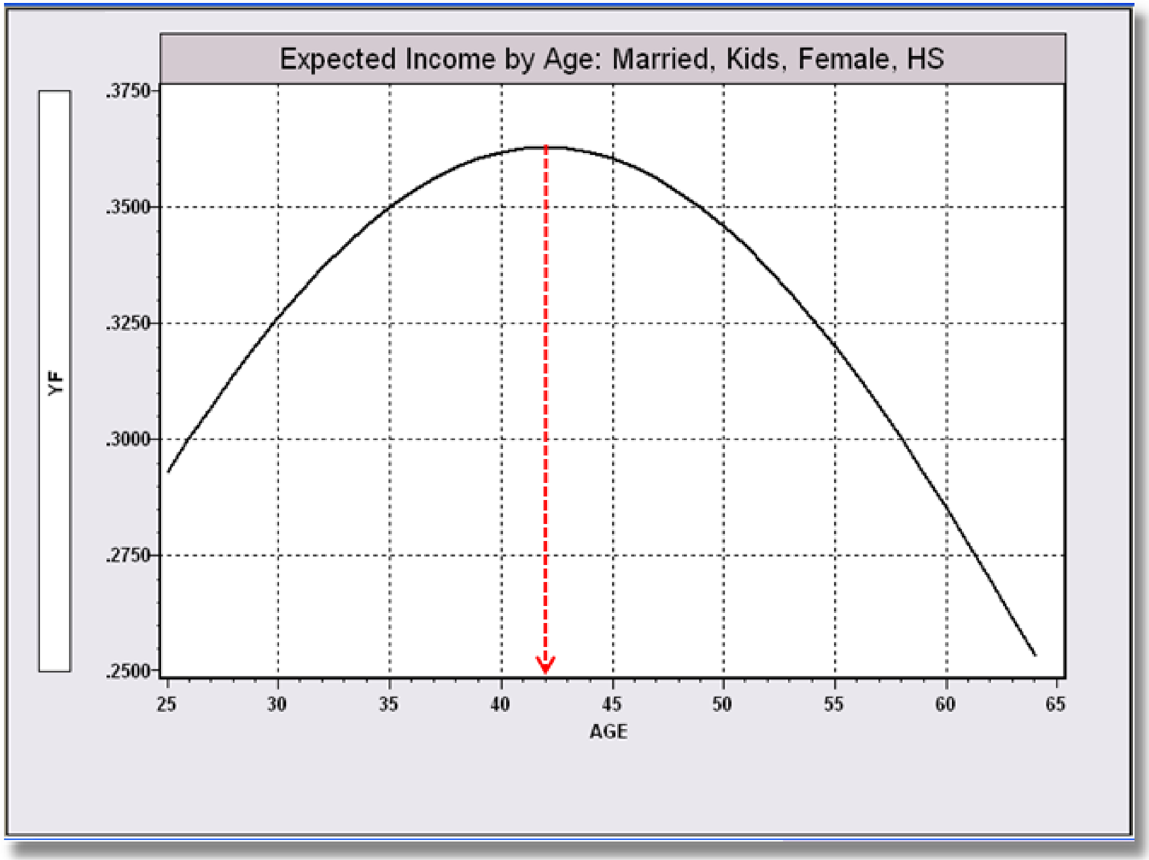
Log Income Equation

```
----------------------------------------------------------------------
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***     .242482D-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
--------+-------------------------------------------------------------
```

Estimated Cov[b1,b2]

	1	2	
1	4.54799e-006	-5.1285e-008	-9
2	-5.1285e-008	5.87973e-010	9.9
3	-9.0034e-005	9.91407e-007	[illegible]

Age-Income Profile: Married=1, Kids=1, Educ=12, Female=1



Application: Maximum of a Function

AGE|	.06225***	.00213	29.189	.0000	43.5272
AGESQ|	-.00074***	.242482D-04	-30.576	.0000	2022.99

	1	2	
1	4.54799e-006	-5.1285e-008	-9
2	-5.1285e-008	5.87973e-010	9.9
3	-9.0034e-005	9.91407e-007	[illegible]

$$\log Y = \beta_1 Age + \beta_2 Age^2 + ...$$

At what age does log income reach its maximum?

$$\frac{\partial \log Y}{\partial Age} = \beta_1 + 2\beta_2 Age = 0 \text{ => Age* } = \frac{-\beta_1}{2\beta_2} = \frac{-.06225}{2(-.00074)} = 42.1$$

$$\frac{\partial Age*}{\partial \beta_1} = \frac{-1}{2\beta_2} = \text{g}_1 = \frac{-1}{2(-.00074)} = 675.68$$

$$\frac{\partial Age*}{\partial \beta_2} = \frac{\beta_1}{2\beta_2^2} = g_2 = \frac{.06225}{2(-.00074)^2} = 56838.9$$

Delta Method Using Visible Digits

	1	2	
1	4.54799e-006	-5.1285e-008	-9
2	-5.1285e-008	5.87973e-010	9.9
3	-9.0034e-005	9.91407e-007	[illegible]

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

$$= .0366952$$

standard error = square root = .1915599

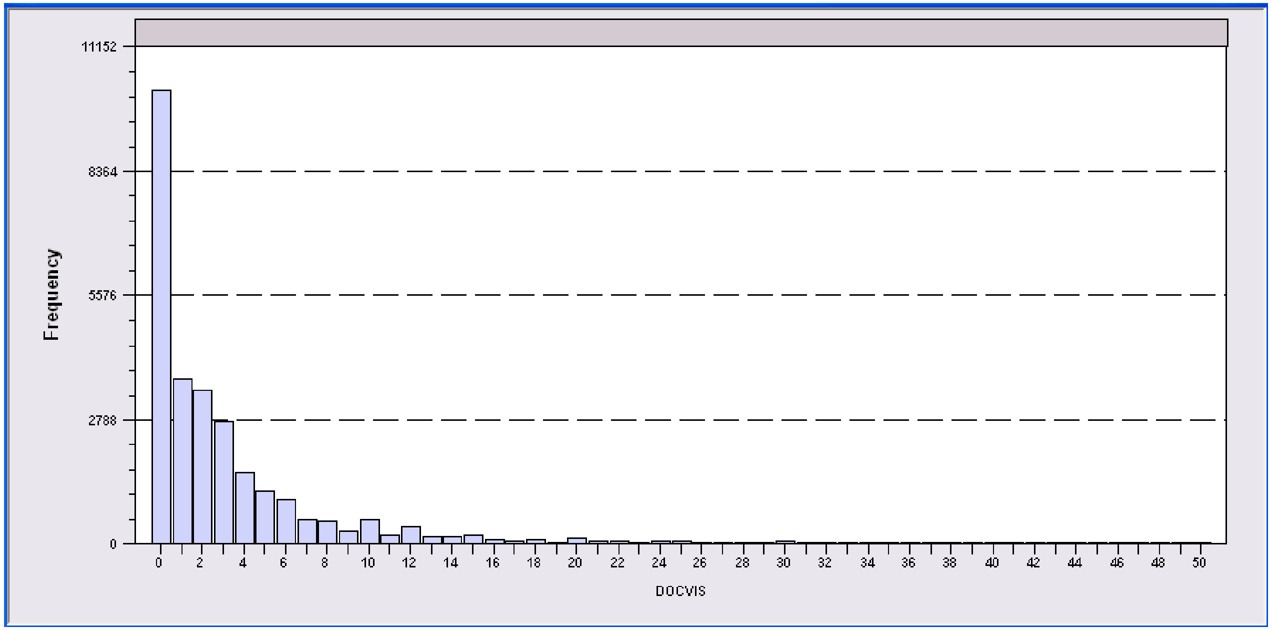
Delta Method Results Built into Software

```
-----------------------------------------------------------
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|income] = **exp**(1.412 - .0745*income)
  - Linear regression: g*(income)=3.917 - .208*income



A Nonlinear Model

$$E[docvis \mid x] = \exp(\beta_1 + \beta_2 AGE + \beta_3 EDUC ...)$$

```
pois ; if[year=1994]
     ; Lhs=docvis
     ; Rhs = one,age,educ,married,female,hhninc $
-----------------------------------------------------------------------------
Poisson Regression
Dependent variable                DOCVIS
Estimation based on N =    3377, K =    6
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient       Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|      .77733***      .06596    11.79  .0000      .64806     .90661
     AGE|      .02006***      .00079    25.42  .0000      .01851     .02160
    EDUC|     -.03078***      .00434    -7.09  .0000     -.03928    -.02228
 MARRIED|     -.02985         .02127    -1.40  .1604     -.07153     .01183
  FEMALE|      .39938***      .01820    21.94  .0000      .36370     .43505
  HHNINC|     -.39489***      .04773    -8.27  .0000     -.48843    -.30135
--------+--------------------------------------------------------------------
```

Interesting Partial Effects

Estimate Effects at the Means of the Data

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

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

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

...

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

Necessary Derivatives (Jacobian)

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

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

$$\frac{\partial f_{AGE}}{\partial b_2} = b_2 \frac{\partial \exp(...)}{\partial b_2} + \exp(...) \frac{\partial b_2}{\partial b_2} = b_2 \exp(...) \times \overline{AGE} \quad + e\,xp(...)1$$

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

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

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

$$\frac{\partial f_{AGE}}{\partial b_6} = \frac{\partial b_2 \exp(b_1 + b_2\overline{AGE} + b_3\overline{EDUC}...)}{\partial b_6} = b_2 \exp(...) \times \overline{HHNINC}$$

```
|-> simulate ; if [year=1994] ; means $

------------------------------------------------------------------------
Model Simulation Analysis for Exponential Regression Function
------------------------------------------------------------------------
Simulations are computed at sample means of all variables
------------------------------------------------------------------------
User Function           Function    Standard
(Delta method)           Value        Error       |t|   95% Confidence Interval
------------------------------------------------------------------------
Func. at means          3.54795       .03334   106.40       3.48259      3.61330
```

```
|-> partials ; if[year=1994] ; effects: age ; means $

------------------------------------------------------------------------
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
------------------------------------------------------------------------
df/dAGE                 Partial     Standard
(Delta method)          Effect        Error       |t|   95% Confidence Interval
------------------------------------------------------------------------
PE.Func(means)           .07116       .00274    25.93        .06578       .07654
```

Partial Effects at Means vs. Mean of Partial Effects

Partial Effects at the Means

$$\delta(\boldsymbol{\beta},\overline{\mathbf{x}}) = \frac{\partial f(\boldsymbol{\beta}\,|\,\overline{\mathbf{x}})}{\partial\overline{\mathbf{x}}} = \frac{\partial f\left(\boldsymbol{\beta}\,|\,\frac{1}{n}\Sigma_{i=1}^{n}\mathbf{x}_i\right)}{\partial\overline{\mathbf{x}}}$$

Mean of Partial Effects

$$\overline{\delta}(\boldsymbol{\beta},\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial f(\boldsymbol{\beta}\,|\,\mathbf{x}_i)}{\partial\mathbf{x}_i}$$

Makes more sense for dummy variables, d:

$$\Delta_i(\boldsymbol{\beta},\mathbf{x}_i,d) = f(\boldsymbol{\beta}\,|\,\mathbf{x}_i,d=1)\ -\ f(\boldsymbol{\beta}\,|\,\mathbf{x}_i,d=0)$$

$$\overline{\Delta}(\boldsymbol{\beta},\mathbf{X},d) \text{ makes more sense than } \delta(\boldsymbol{\beta},\overline{\mathbf{x}},\overline{d})$$

Partial Effect for a Dummy Variable?

```
|-> Partials ; if[year = 1994 ] ; Effects: Female $
--------------------------------------------------------------------
Partial Effects  Analysis for Exponential Regression Function
--------------------------------------------------------------------
Effects on function with respect to FEMALE
Results are computed by average over sample observations
Partial effects for binary var FEMALE   computed by first difference
--------------------------------------------------------------------
df/dFEMALE           Partial    Standard
(Delta method)       Effect      Error       |t|   95% Confidence Interval
--------------------------------------------------------------------
APE. Function        1.50212     .06856    21.91      1.36775      1.63649
|-> Partials ; if[year = 1994 ] ; Effects: Female ; means$
--------------------------------------------------------------------
Partial Effects  Analysis for Exponential Regression Function
--------------------------------------------------------------------
Effects on function with respect to FEMALE
Results are computed at sample means of all variables
Partial effects for continuous FEMALE   computed by differentiation
Effect is computed as derivative      = df(.)/dx
--------------------------------------------------------------------
df/dFEMALE           Partial    Standard
(Delta method)       Effect      Error       |t|   95% Confidence Interval
--------------------------------------------------------------------
PE.Func(means)       1.41696     .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

```
Random-effects probit regression                Number of obs      =     41164
Group variable: IDENT                           Number of groups   =      9578

Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       4.3
                                                               max =         7

                                                Wald chi2(7)       =   8594.46
Log likelihood  = -18180.523                    Prob > chi2        =    0.0000
```

COOPERA	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
coop_1	1.413728	.0206948	68.31	0.000	1.373167	1.454289
Prod_1	-.0220752	.0100407	-2.20	0.028	-.0417545	-.0023958
ri_1	.0996992	.0071557	13.93	0.000	.0856744	.1137241
lsize	.1086429	.0085745	12.67	0.000	.0918372	.1254487
GRUPO	.2196952	.024657	8.91	0.000	.1713683	.2680221
funds	.6237657	.0193544	32.23	0.000	.5858318	.6616996
Foreign	.0069793	.0291385	0.24	0.811	-.0501311	.0640896
_cons	-2.331907	.1321667	-17.64	0.000	-2.590949	-2.072865
/lnsig2u	-1.123706	.0715548			-1.26395	-.983461
sigma_u	.5701517	.0203985			.5315408	.6115672
rho	.2453245	.0132477			.2202946	.2722056

Target of estimation is $\rho = \dfrac{\sigma^2}{1+\sigma^2}$

Estimation strategy:

(1) Estimate $\alpha = \log\sigma^2$ → /lnsig2u

(2) Estimate $\sigma = \exp(\alpha/2)$ → sigma_u

(3) Estimate $\rho = \sigma^2/(1+\sigma^2)$ → rho

Delta Method

/lnsig2u	-1.123706	.0715548
sigma_u	.5701517	.0203985
rho	.2453245	.0132477

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

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

$$\hat{g} = d\hat{\sigma}/d\hat{\alpha} = \tfrac{1}{2}\exp(\hat{\alpha}/2) = \tfrac{1}{2}\hat{\sigma} = .2850758$$

$$\hat{g}^2 = \left(d\hat{\sigma}/d\hat{\alpha}\right)^2 = .08126821$$

$$\left(d\hat{\sigma}/d\hat{\alpha}\right)^2 V_{\hat{\alpha}} = .08126821(.00512009) = .0004161$$

Estimated Standard Error for $\hat{\sigma} = \sqrt{.0004161} = .02039854$

Delta Method

Continuing the previous example, there are two approaches implied for estimating ρ:

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

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

Use the delta method to estimate a standard error for each of the two estimators of ρ. Do you obtain the same answer?

Confidence Intervals?

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

$\hat{\sigma} \in .5701517 \pm 1.96(.0203985) = .5301707$ to $.6101328$??

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)!

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,
AAAAAA