Lecture 4
Binary Data

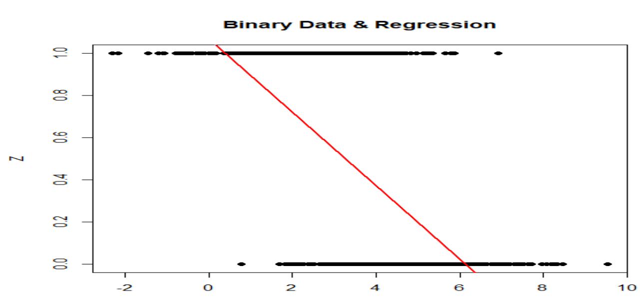
(for private use, not to be posted/shared online).

1

DCM: Binary Data – Intuition

• It is common to have binary {(1, 0); Yes/No)} data: Invest in stocks or not, visit a doctor or not, buy a Houston Texans ticket or not, etc.

• We want to model the binary choice decision as a function of some independent variables $\boldsymbol{x}_n$. Recall that we can do OLS, but …



DCM: Binary Data – Intuition

• A regression will not work well, we want to predict values between 0 and 1. If we interpret the values of y_n as probabilities, the line does not work either.

• Intuition: We need to transform the data from discontinuous to continuous, like the dependent variable. We think of y_n as probabilities, but, with the idea that $f(\boldsymbol{x}_n)$ should produce be very high or very low values. A candidate function: a CDF.

• This is the basic idea of all the models with discrete choice. For example, in a Probit model, we use the normal distribution as $f(\boldsymbol{x}_n)$. The S shape CDF will fit the data much better than a simple line.

DCM: Binary Data – Review

• Q: How can we model this choice decision?

- We can use an ad-hoc model, say: $y_n = \boldsymbol{x}_n{}'\boldsymbol{\beta} + \varepsilon_n$
- We can use economic theory. This is what we do.

• We use the McFadden approach, described in the previous lecture.

1) **We model a consumer's decision** (1=Yes or 0=No) by specifying the utility to be gained from 1 as a function of:

- characteristics of the decision, $\boldsymbol{z}_n$: price, quality, fun, etc.)
- characteristics of individuals, $\boldsymbol{w}_n$: age, sex, income, education, etc.):

$U_{n0} = \alpha_0 + \boldsymbol{z}_{n0}{}'\,\delta_0 + \boldsymbol{w}_n{}'\,\gamma_0 + \varepsilon_{n0}$ – utility from decision 0

$U_{n1} = \alpha_1 + \boldsymbol{z}_{n1}{}'\,\delta_1 + \boldsymbol{w}_n{}'\,\gamma_1 + \varepsilon_{n1}$ – utility from decision 1

We assume that the errors are *i.i.d.* $(0, \sigma^2)$.

DCM: Binary Data – Review

2) **We introduce the probability model**:

$$
\begin{aligned}
P_{n1}[Decision\ 1] &= P(U_{n1} - U_{n0} > 0) \\
&= P(\alpha_1 + \boldsymbol{z}_{n1}'\,\delta_1 + \boldsymbol{w}_n'\,\gamma_1 + \varepsilon_{n1} > \\
&\qquad \alpha_0 + \boldsymbol{z}_{n0}'\,\delta_0 + \boldsymbol{w}_n'\,\gamma_0 + \varepsilon_{n0}) \\
&= P(\boldsymbol{x}_n'\boldsymbol{\beta} - \xi_n > 0) \\
&= P(\,\xi_n < \boldsymbol{x}_n'\boldsymbol{\beta})
\end{aligned}
$$

where $\boldsymbol{x}_n$= [1 $\boldsymbol{z}_{n1}$ $\boldsymbol{z}_{n0}$ $\boldsymbol{w}_n$]' and $\xi_n = \varepsilon_{n0} - \varepsilon_{n1}$.

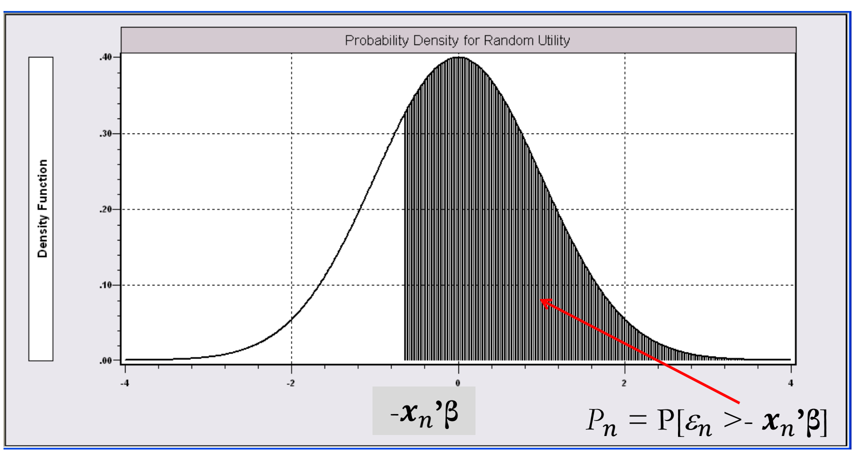
Then,

$$P_{n1} = \text{P}[y_n = 1\,] = \text{P}(\xi_n < \boldsymbol{x}_n\text{'}\beta)$$

Our problem is to estimate β given $\boldsymbol{z}_n$, $\boldsymbol{w}_n$ and y_n. To proceed with the estimation, we assume ε_n follows a specific distribution. Then, we can naturally estimate the model using ML.

DCM: Binary Data – Review

Probability Model for Choice Between Two Alternatives



$$\varepsilon_{n0} > -[\alpha_0 + \boldsymbol{z}_{n0}'\,\delta_0 + \boldsymbol{w}_n'\,\gamma_0\,]$$

DCM: Binary Data – Review

3) Estimation

If we assume we know $f(\varepsilon_n)$, ML is a natural way to proceed.

$$L(\beta) = \Pi_n (1 - P[y_n = 1 | \boldsymbol{x}, \boldsymbol{\beta}]) P[y_n = 1 | \boldsymbol{x}, \boldsymbol{\beta}]$$

$$\Rightarrow \text{Log } L(\beta) = \sum_{n(y=0)}^{N} \log(1 - F(\boldsymbol{x}_n{}'\boldsymbol{\beta})) + \sum_{n(y=1)}^{N} \log(F(\boldsymbol{x}_n{}'\boldsymbol{\beta}))$$

• If we have grouped data: p_i, $\boldsymbol{x}_i$, & n_i, then

$$\Rightarrow \text{Log } L(\beta) = \sum_{i=1}^{N} n_i [p_i \log[F(\boldsymbol{x}_n{}'\boldsymbol{\beta})] + (1 - p_i) \log\{(1 - F(\boldsymbol{x}_n{}'\boldsymbol{\beta})\}]$$

Note: NLLS is easy in this context. Since:

$$E[y_n | \boldsymbol{x}] = P[y_n = 1] = P(\xi_n < \boldsymbol{x}_n{}'\boldsymbol{\beta}) = g(\boldsymbol{x}_n'\boldsymbol{\beta})$$

$$\Rightarrow \quad y_n = E[y_n | \boldsymbol{x}] + v_n = g(\boldsymbol{x}_n'\boldsymbol{\beta}) + v_n \quad \text{-- with } E[v_n | \boldsymbol{x}] = 0.$$

(Bayesian and simulation methods can be used too.)

DCM: Binary Data – Review

• What do we do learn from this model?

- Are the z_n & w_n "relevant?"

- Can we predict behavior? Will a specific company do an M&A? How many mergers will happen next year?

- What is the effect of a change in $\boldsymbol{x}_n$ in y_n? Do traders trade more if they receive more research/information?

Binary Data – Latent Variable Interpretation

• In the previous theoretical setup, there is an unobservable variable, the difference in utilities. We can assume that this latent variable, y_n^*, is distributed around its mean: $E[y_n^* | \boldsymbol{X}_n] = \beta_0 + \boldsymbol{x}_n{}'\beta_1$

• We observed y's (choices) -in the binary case, (0, 1). We observe y_n=1 for that portion of the latent variable distribution above τ.

$$y_n = \begin{cases} 1 & if \quad y_n^* > \tau \\ 0 & if \quad y_n^* \leq \tau \end{cases}$$

Then, if $y_n^* = (U_{n1} - U_{n0}) > 0$, n selects 1. $\Rightarrow \tau = 0$.

• Given the conditional expectation assumed:

$$\Rightarrow \quad y_n^* = U_{n1} - U_{n0} = \beta_0 + \boldsymbol{x}_n{}'\beta_1 + \varepsilon_n, \qquad \varepsilon_n \sim D(0, 1)$$

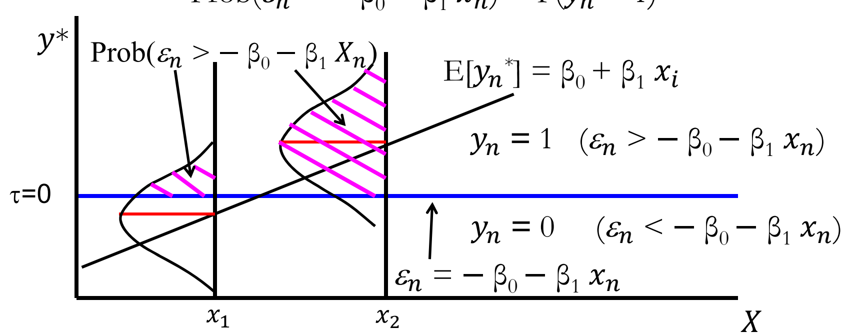
9

Binary Data – Latent Variable Interpretation

• Latent Variable Model –Goldeberg (1964):

$$y_n^* = U_{n1} - U_{n0} = \beta_0 + \beta_1\, x_i + \varepsilon_n, \qquad \varepsilon_n \sim \text{symmetric } D(0,1)$$

• $P(y_n^* > 0) = \text{Prob}(\beta_0 + \beta_1\, x_n + \varepsilon_n > 0)$
$= \text{Prob}(\varepsilon_n > -\beta_0 - \beta_1\, x_n) = P(y_n = 1)$



• Even if $E(y_n^* | x_n)$ is in the (pink) portion where $y_n = 1$, we can still observe a 0 if ε_n is large and negative.

10

Binary Data – Partial Effects

• In general, we are interested not in the probability per se, but on the expected (or other central) value and the effect of a change in x_i on the expected value. With binary (0, 1) data, the expected mean is:

$$E[y_n \mid \boldsymbol{x}] = 0 * P[y_n = 0] + 1\ P[y_n = 1] = P[y_n = 1] = g(\boldsymbol{x}_n'\boldsymbol{\beta})$$

• The predicted y_n is the predicted probability of Yes.

• Note that β_k is not the usual elasticity. It gives the change in the probability of Yes when x_k changes. Not very interesting.

• To make sense of the parameters, we calculate:

$$\text{Partial effect} = \text{Marginal effect} = \frac{\delta P(\alpha + \beta_1\, Income + \ldots)}{\delta x_k}$$

Binary Data – Partial Effects

• The partial effects will vary with level of $\boldsymbol{x}$. We usually calculate them at the sample means of the $\boldsymbol{x}$. For example:

$$\text{Estimated Marginal effect} = f(\alpha + \beta_1\, Mean[Income] + \ldots) * \beta_k$$

• The marginal effect can also be computed as the average of the marginal effects at every observation.

• Q: Computing effects at the data means or as an average?
- At the data means: easy and inference is well defined.
- As an average of the individual effects: Its appeal is based on the LLN, but the asymptotic standard errors are problematic.

Binary Data – Partial Effects & Dummies

• A complication in binary choice models arises when **x** includes dummy variables. For example, marital status, sex, MBA degree, etc.

• A derivative (with respect to a small change) is not appropriate to calculate partial effects. A jump from 0 to 1 is not small.

• We calculate the marginal effect as:

$$\text{Prob}[y_n = 1 \,|\, \boldsymbol{x}_n,\ d_n{=}1] - \text{Prob}[y_n = 1 \,|\, \boldsymbol{x}_n, d_n{=}0],$$

where $\boldsymbol{x}_n$ is computed at sample mean values.

Binary Data – Non-linear Regression

• From the expected mean, we can write:

$$y_n = \text{P}[y_n{=}1] + v_n = g(\boldsymbol{x}_n'\boldsymbol{\beta}) + v_n \quad \text{-- with E}[v_n|\boldsymbol{x}] = 0$$

⇒ we have a **non-linear regression** relation connecting the binary variable y_n and P($\boldsymbol{x}_n'\boldsymbol{\beta}$).

Usual P($\boldsymbol{x}_n'\boldsymbol{\beta}$): Normal, Logistic, and Gompertz distributions.

• In principle, NLLS is possible to estimate $\boldsymbol{\beta}$.

• But, note this model is not homoskedastic.

$$\begin{aligned}\text{Var}[y_n\,|\,\boldsymbol{x}] &= (0 - \text{E}[y_n\,|\,\boldsymbol{x}])^2 * \text{P}[y_n{=}0] + (1 - \text{E}[y_n\,|\,\boldsymbol{x}])^2 * \text{P}[y_n{=}1] \\ &= (-\text{P}[y_n{=}1])^2\,(1 - \text{P}[y_n{=}1]) + ((1 - \text{P}[y_n{=}1])^2\,\text{P}[y_n{=}1] \\ &= (1 - \text{P}[y_n = 1]) * \text{P}[y_n = 1]\end{aligned}$$

⇒ NLLS will not be efficient. A GLS adjustment is possible.

Linear Probability Model (LPM)

• A linear regression can also be used to estimate $\boldsymbol{\beta}$ by assuming

$$P[y_n = 1] = \boldsymbol{x}_n'\boldsymbol{\beta}$$

• This model is called the *linear probability model*. It delivers:

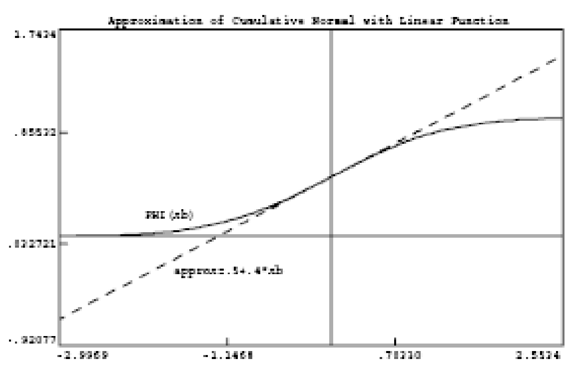
$$y_n = P[y_n = 1] + v_n = \boldsymbol{x}_n'\boldsymbol{\beta} + v_n$$

⇒ now, we can regress the binary data against the $\boldsymbol{x}_n'$ to get an estimator of $\boldsymbol{\beta}$. Very simple!

• We have constant partial effects: β_k.

• Difficult to obtain this model from utility maximization, but $\Phi(\boldsymbol{x}_n'\boldsymbol{\beta})$ & $\boldsymbol{x}_n'\boldsymbol{\beta}$ are closely related over much of the likely range of $\boldsymbol{x}_n'\boldsymbol{\beta}$.

LPM: Approximation



Amemiya notes that the linear probability model 0.5 + .4($\boldsymbol{x}_n'\boldsymbol{\beta}$) and $\Phi(\boldsymbol{x}_n'\boldsymbol{\beta})$ are reasonably close for values of Φ between .3 and .7.

LPM: Advantages & Disadvantages

• Advantage: Estimation!

• Potential problems:

- The probability can be outside [0, 1].

- In addition, we may estimate effects that imply a change in x changes the probability by more than +1 or –1. Nonsense!

- Model and data suggest heteroscedasticity. LPM ignores it.

- Partial effects. The linear model predicts constant marginal effects. But, we observe non-linear effects. For example, at very low level of income a family does not own a house; at very high level of income every one owns a house; the marginal effect of income is small.

- Non-normal errors. The errors are $(1 - \boldsymbol{x}_n'\boldsymbol{\beta})$ or $(- \boldsymbol{x}_n'\boldsymbol{\beta})$.

LPM: GLS

• This model is not homoskedastic. Problem for standard inferences.

$$\begin{aligned}\text{Var}[y_n = 1 | \boldsymbol{x}_n] &= (1 - P[y_n = 1 | \boldsymbol{x}_n])\, P[y_n = 1 | \boldsymbol{x}_n] \\ &= (1 - \boldsymbol{x}_n'\boldsymbol{\beta}) * \boldsymbol{x}_n'\boldsymbol{\beta} \\ &= \boldsymbol{x}_n'\boldsymbol{\beta} - (\boldsymbol{x}_n'\boldsymbol{\beta})^2\end{aligned}$$

⇒ the variance changes with the level of the regressors.

- $\boldsymbol{\beta}$ is still unbiased, but inefficient. We can transform the model to gain efficiency: A GLS transformation with sqrt$\{\boldsymbol{x}_n'\boldsymbol{\beta} - (\boldsymbol{x}_n'\boldsymbol{\beta})^2\}$.

<u>Additional Problem</u>: $\boldsymbol{x}_n'\boldsymbol{\beta} - (\boldsymbol{x}_n'\boldsymbol{\beta})^2$ may not be positive.

• Despite its drawbacks, it is a good place to start when y_n is binary
⇒ used to get a "feel" for the relationship between y_n and $\boldsymbol{x}_n$.

Binary Probit Model: Setup

We have data on whether individuals buy a ticket to see the Houston Rockets or the Houston Texans. We have various characteristics of the tickets/teams, $\boldsymbol{z}_i$, (price, team record, opponent's record, etc.) and the individuals who buy them, $\boldsymbol{w}_n$ (age, sex, married, children, income, education, etc.).

Steps to build a DCM:

1) Specifying the utility to be gained from attending a game as a function of $\boldsymbol{z}_i$ and $\boldsymbol{w}_n$:

$U_{n0} = \alpha_0 + \boldsymbol{z}_{n0}{}'\,\delta_0 + \boldsymbol{w}_n{}'\,\gamma_0 + \varepsilon_{n0}$ – utility from Rockets game

$U_{n1} = \alpha_1 + \boldsymbol{z}_{n1}{}'\,\delta_1 + \boldsymbol{w}_n{}'\,\gamma_1 + \varepsilon_{n1}$ – utility from Texans game

We assume that the errors are *i.i.d.* $(0, \sigma^2)$.

Binary Probit Model: Setup – Normal CDF

2) Probability model:

$$\begin{aligned} P_{n1}[Texans\ game] &= P(U_{n1} - U_{n0} > 0) \\ &= P(\boldsymbol{x}_n'\boldsymbol{\beta} - \xi_n > 0) \\ &= P(\xi_n < \boldsymbol{x}_n'\boldsymbol{\beta}) \end{aligned}$$

where $\boldsymbol{x}_n$=[1 $\boldsymbol{z}_{n1}$ $\boldsymbol{z}_{n0}$ $\boldsymbol{w}_n$]' and $\xi_{n0} = \varepsilon_{n0} - \varepsilon_{n1}$

Let $y_n = 1$ if a Texans game is chosen and 0 otherwise. Then,

$$P_{n1} = \text{P}[y_n = 1] = \text{P}(\xi_n < \boldsymbol{x}_n{}'\boldsymbol{\beta})$$

Our problem is to estimate $\boldsymbol{\beta}$ given $\boldsymbol{z}_{ni}$, $\boldsymbol{w}_n$ and y_n. We assume ε_n is normally distributed. Then, ξ_n is also normal.

$$\text{P}(\xi_n < \boldsymbol{x}_n{}'\beta) = \Phi(\boldsymbol{x}_n{}'\beta) = \int_{-\infty}^{x_n'\beta} \frac{1}{\sqrt{2\pi}}\, \boldsymbol{e}^{-\frac{1}{2}\varepsilon^2}\, d\varepsilon$$

Binary Probit Model: Setup – Identification

3) Identification:

Normalization: The variance of ξ_n is set to 1, as it is impossible to identify it.

Intuition of normalization: Several views:

1) One can do it formally by observing that the score for σ^2 would be zero for any value.

2) Another is to observe that $P(\xi_n < \boldsymbol{x}_n'\beta) = P[\sigma^{-1}\xi_n < \sigma^{-1}\boldsymbol{x}_n'\beta)]$, making $\boldsymbol{\beta}$ identifiable only up to a factor of proportionality.

3) More basic. The problem arises because the numbers in the data are arbitrary - i.e. we could have assigned the values (1, 2) instead of (0, 1) to y_n. It is possible to produce any range of values in y_n.

Binary Probit Model: Non-linear Regression

Now, we can formally write the integral:

$$P(\xi_n < \boldsymbol{x}_n'\beta) = \Phi(\boldsymbol{x}_n'\beta) = \int_{-\infty}^{x_n'\beta} \frac{1}{\sqrt{2\pi}} \boldsymbol{e}^{-\frac{1}{2}\varepsilon^2}\, d\varepsilon$$

This is the **Probit Model**.

• In this case, the Probit model estimates

$$y_n = \Phi(\boldsymbol{x}_n'\boldsymbol{\beta}) + v_n,$$

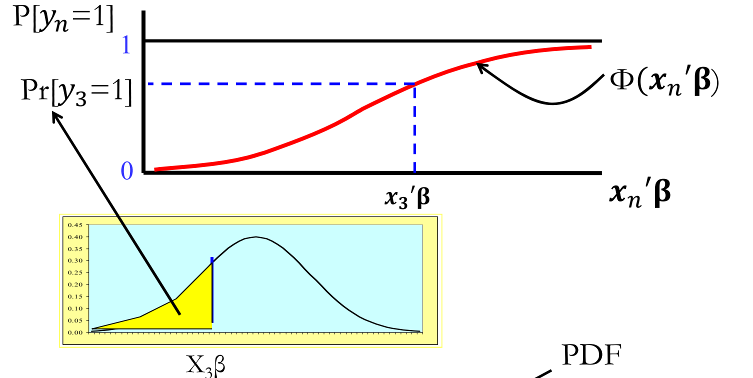
a non-linear regression model. NLLS is possible.

Note: We could alternatively have just begun with the proposition that the probability of buying a Texans ticket, $P[y_n=1] = \Phi(\boldsymbol{x}_n'\boldsymbol{\beta})$, is some function of a set of characteristics $\boldsymbol{x}_n$.

Binary Probit Model - Summary

• We derived : E[y_n |$\boldsymbol{x}$] = P[y_n = 1]. We assume a normal CDF.

$$\Rightarrow y_n = \Phi(\boldsymbol{x}_n{}'\boldsymbol{\beta}) + v_n$$



PDF

$$P(y_n = 1 \mid x_n) = \int_{-\infty}^{x_n'\beta} f(\varepsilon)\, d\varepsilon = \Phi(x_n{}'\beta)$$

$$P(y_n = 0 \mid x_n) = 1 - \Phi(x_n{}'\beta)$$

CDF

Binary Logit Model: Setup

• Usual setup. Suppose we are interested in whether an agent chooses to visit a physician or not. We have data on doctor's visits, y_n, and the agent's characteristics, X (age, sex, income, etc.).

• Dependent variable: y_n = 1, if agent visits a doctor at least one
= 0, if no visits.

• RUM: Net utility of visit at least once

$$U_{visit} = \alpha + \beta_1 \text{Age} + \beta_2 \text{Income} + \beta_3 \text{Sex} + \varepsilon$$

• Visit if net utility is positive: Net utility = $U_{visit} - U_{no\ visit} > 0$

$\Rightarrow$ Agent chooses to visit: $U_{visit} > 0$ (set $U_{no\ visit} = 0$)

$$\alpha + \beta_1 \text{Age} + \beta_2 \text{Income} + \beta_1 \text{Sex} + \varepsilon > 0$$
$$\varepsilon \geq -[\alpha + \beta_1 \text{Age} + \beta_2 \text{Income} + \beta_1 \text{Sex}]$$

Binary Logit Model: Setup

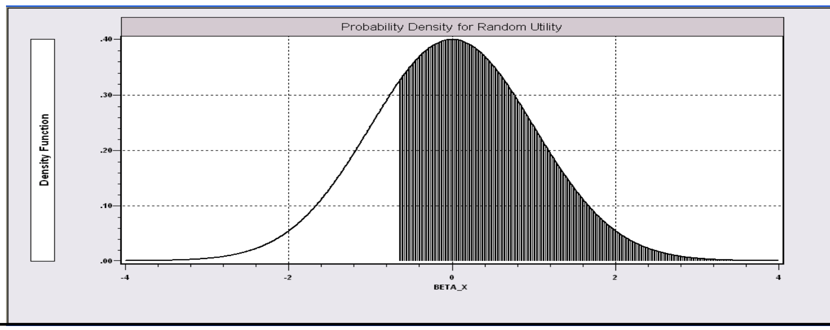
- Agent chooses to visit: $U_{visit} > 0$

  $\alpha + \beta_1 \text{ Age} + \beta_2 \text{ Income} + \beta_1 \text{ Sex} + \varepsilon > 0$

  $\varepsilon \;\geq\; -[\alpha + \beta_1 \text{ Age} + \beta_2 \text{ Income} + \beta_1 \text{ Sex}]$

- Add a probability model for ε

Prob[y_n = 1] = Prob[$\varepsilon > -(\alpha + \beta_1 \text{ Age} + \beta_2 \text{ Income} + \beta_3 \text{ Sex})$]



Binary Logit Model: Gumbel Distribution

- $$P_n\,[y_n = 1] = \int I[\boldsymbol{x}'_n\beta + \varepsilon_n > 0]\, f(\varepsilon)\, d\varepsilon$$
  $$= \int I[\varepsilon_n > -\boldsymbol{x}'_n\beta]\, f(\varepsilon)\, d\varepsilon$$
  $$= \int_{\varepsilon_n = -x'_n\beta}^{\infty} f(\varepsilon)\, d\varepsilon$$

<u>Assumption</u>: The error terms are *i.i.d.* and follow a *Gumbel distribution*. That is,

$$CDF: \quad F(\varepsilon_{nj}) = e^{-e^{-\varepsilon_{nj}}}$$

$$PDF: \quad f(\varepsilon_{nj}) = e^{-\varepsilon_{nj}}\, e^{-e^{-\varepsilon_{nj}}}$$

- The Gumbel distribution is the most common of the three types of Fisher-Tippett extreme value distributions, also referred as ***Type I distribution***. These are distributions of an extreme order statistic for a distribution of N elements. The term "Gumbel distribution" is used to refer to the distribution of the minimum.

Binary Logit Model: Gumbel Distribution

• The Gumbel distribution: General CDF and PDF

$$CDF: \quad F(x) = \exp[-e^{-(\frac{x-\lambda}{\beta})}]$$

$$PDF: \quad f(x) = \frac{1}{\beta} e^{-(\frac{x-\lambda}{\beta}) - e^{-(\frac{x-\lambda}{\beta})}}$$

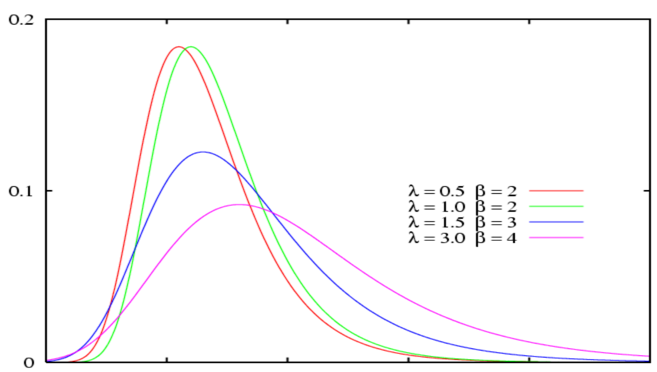
• Parameters: λ is the location, and β is the scale.

• Mean = λ + γ β (γ: Euler-Mascheroni constant ≈0.5772).
Variance = $\beta^2 \pi^2/6$

• Nice property: The difference of two Gumbel-distributed RVs has a logistic distribution.

Binary Logit Model: Gumbel Distribution

• Graph: Gumbel pdf



• Parameters used in the Logit Model: λ=0, β=1
Mean = γ ≈ 0.5772.
Variance = $\pi^2/6$ ≈ 1.65.

Binary Logit Model: Gumbel Distribution

- Assuming linearity for the RUM-model, we state the choice problem in terms of covariates:

$$P_n = \int_{\varepsilon_n = -x_n'\beta}^{\infty} f(\varepsilon)\, d\varepsilon \qquad f(\varepsilon_{nj}) = e^{-\varepsilon_{nj}} e^{-e^{-\varepsilon_{nj}}}$$

$$= 1 - F(-\boldsymbol{x}_n'\beta) \qquad F(\varepsilon_{nj}) = e^{-e^{-\varepsilon_{nj}}}$$

$$= 1 - 1/[1 + exp(\boldsymbol{x}_n'\beta)]$$

$$= exp(\boldsymbol{x}_n'\beta)/[1 + exp(\boldsymbol{x}_n'\beta)]$$ (This is the *logistic function*.)

- Technical Details:

$$\int_{-\infty}^{c} e^{-\varepsilon} e^{-e^{-\varepsilon}} d\varepsilon = F(c) = \exp(-\exp(-c))$$

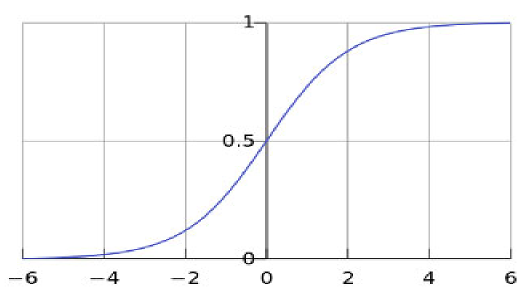
$$\int_{-\infty}^{\infty} e^{-\varepsilon} e^{-e^{-\varepsilon-c}} d\varepsilon = \int_{-\infty}^{\infty} e^{-\eta+c} e^{-e^{-\eta}} d\eta = e^{c} \int_{-\infty}^{\infty} e^{-\eta} e^{-e^{-\eta}} d\eta = \exp(c)$$

where we have used change of variables with c = -ln(1 + exp($\boldsymbol{x}_n'\beta$))

Binary Logit Model: Logistic Function

- We have the following expression for the logit choice probability:

$$\text{P}[y_n = 1] = \frac{exp(V_n)}{1 + exp(V_n)} = \frac{exp(x_n'\boldsymbol{\beta})}{1 + exp(x_n'\boldsymbol{\beta})}$$



- Properties:
  - Nonlinear effect of covariates on dependent variable
  - Logistic curve with inflection point at P=0.5

Binary Logit Model: Non-linear Regression

• The logit choice probability:

$$P[y_n = 1] = \frac{exp(V_n)}{1 + exp(V_n)} = \frac{exp(\boldsymbol{x_n}'\boldsymbol{\beta})}{1 + exp(\boldsymbol{x_n}'\boldsymbol{\beta})}$$

Note: We could alternatively have just begun with the proposition that the probability of visiting a doctor follows a logistic distribution, as a function of the set of characteristics $\boldsymbol{x_n}$:

$$P[y_n = 1] = \frac{exp(\boldsymbol{x_n}'\boldsymbol{\beta})}{1 + exp(\boldsymbol{x_n}'\boldsymbol{\beta})} = F(\boldsymbol{x_n}'\boldsymbol{\beta})$$

In this case, the logit model estimates:

$$y_n = \frac{exp(\boldsymbol{x_n}'\boldsymbol{\beta})}{1 + exp(\boldsymbol{x_n}'\boldsymbol{\beta})} + v_n,$$

another non-linear regression model. Again, NLLS can be used.

Binary Logit Model: Estimation – NLLS

• We can estimate the models using NLLS or MLE.

(1) NLLS. Use Gauss-Newton. Let's linearize P_n:

$P_n \approx F(-\boldsymbol{x_n}'\boldsymbol{\beta}_0) + \delta F/\delta\boldsymbol{\beta}_{(0)}\,(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$ $\boldsymbol{J}$:Jacobian $= \delta F(x_n;\boldsymbol{\beta})/\delta\boldsymbol{\beta}$.

$\Rightarrow y_n - F(-\boldsymbol{x_n}'\boldsymbol{\beta}_0) \approx \boldsymbol{J}_{n,(0)}\,(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \text{error}_n,$

• The update is a regression: $y_n - F(-\boldsymbol{x_n}'\boldsymbol{\beta}_0)$ against $\boldsymbol{J}_n$.

• Given the heteroscedasticity of the models, NLLS will not be efficient. Weighted NLLS can be used. Then, in the algorithm use:

- Dependent variable: $y_n - F(-x'_n\boldsymbol{\beta}_0)$ /sqrt{(1 - P[y_n=1]) * P[y_n=1]}
- Independent variable: $\boldsymbol{J}_{n,(0)}$/sqrt{(1 - P[y_n=1]) * P[y_n=1]}

Binary Logit Model: Estimation – MLE

(2) MLE. Since we specify a pdf, we can do MLE:

$L(\beta) = \prod_n (1 - P[y_n = 1 \mid \boldsymbol{x}_n, \boldsymbol{\beta}])\, P[y_n=1 \mid \boldsymbol{x}_n, \boldsymbol{\beta}]$

$\Rightarrow \text{Log } L(\beta) = \sum_{n\,(y=0)} log(1 - F(\boldsymbol{x}_n'\boldsymbol{\beta})) + \sum_{n\,(y=1)} log(F(\boldsymbol{x}_n'\boldsymbol{\beta}))$

Then, the k f.o.c. for maximization of the total sample Log L(β) are

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{T}\left[\frac{y_i f_i}{F_i} + (1\text{-}y_i)\frac{-f_i}{(1-F_i)}\right] X_i = 0$$

where f_i is the pdf $\equiv dF/d(Z_i)$, which are functions of $\boldsymbol{\beta}$ and **x**.

• Under most likely conditions this likelihood function is globally concave. ⇒ uniqueness of the ML parameter estimates

• In general, it can get complicated.

Binary Choice Models: Estimation - Review

• In general, we assume the following distributions:

– Normal: **Probit Model** = $\Phi(\boldsymbol{x}_n'\boldsymbol{\beta})$

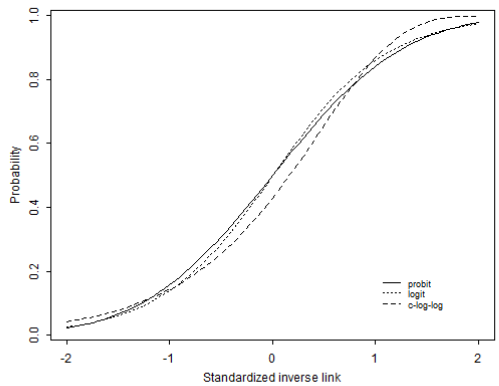
– Logistic: **Logit Model** = $\frac{exp(\boldsymbol{x}_n'\boldsymbol{\beta})}{1 + exp(\boldsymbol{x}_n'\boldsymbol{\beta})}$

– Gompertz: **Extreme Value Model** = $1 - exp[- exp(\boldsymbol{x}_n'\boldsymbol{\beta})]$

• Methods

- ML estimation (Numerical optimization)
- Bayesian estimation (MCMC methods)
- Simulation-assisted estimation

Comparison: Probit vs Logit



• Logit has fatter tails, but, in practica, it is difficutl to distinguish probabilities and fit of both models.

• The coefficients are not directly comparable.

• Signs and signficances are similar.

ML Estimation – Application I (Greene)

• Logit Model for doctor's visits as function of age, income and gender.

```
-----------------------------------------------------------------------
Binary Logit Model for Binary Choice
Dependent variable                DOCTOR
Log likelihood function      -2097.48109
Restricted log likelihood    -2169.26982
Chi squared [    3 d.f.]       143.57744
Significance level               .00000
McFadden Pseudo R-squared      .0330935
Estimation based on N =    3377, K =   4
Information Criteria: Normalization=1/N
              Normalized    Unnormalized
AIC              1.24458      4202.96219
Fin.Smpl.AIC     1.24459      4202.97405
Bayes IC         1.25184      4227.46116
Hannan Quinn     1.24718      4211.72150
Model estimated: Nov 04, 2009, 08:07:51
Hosmer-Lemeshow chi-squared =   26.50241
P-value=  .00086 with deg.fr. =        8
--------+--------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+--------------------------------------------------------------
        |Characteristics in numerator of Prob[Y = 1]
Constant|     -.42085***       .15810      -2.662   .0078
     AGE|      .02365***       .00328       7.205   .0000     42.6266
  INCOME|     -.44198***       .16936      -2.610   .0091      .44476
  FEMALE|      .63825***       .07551       8.453   .0000      .46343
--------+--------------------------------------------------------------
Note: ***, **, * = Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------
```

What do these mean?

ML Estimation – Application I (Greene)

• Logit Model for doctor's visits as function of age, income and sex.

	LOGIT		PROBIT		EXTREME VALUE	
Variable	Estimate	t-ratio	Estimate	t-ratio	Estimate	t-ratio
Constant	-0.42085	-2.662	-0.25179	-2.600	0.00960	0.078
Age	0.02365	7.205	0.01445	7.257	0.01878	7.129
Income	-0.44198	-2.610	-0.27128	-2.635	-0.32343	-2.536
Sex	0.63825	8.453	0.38685	8.472	0.52280	8.407
Log-L	-2097.48		-2097.35		-2098.17	
Log-L(0)	-2169.27		-2169.27		-2169.27	

Note: For now, ignore the *t-ratios*.

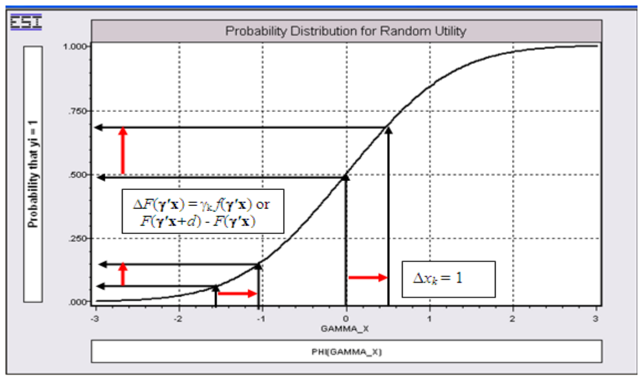
ML Estimation – Application II (Wooldridge)

• Labor participation of married women (Example 15.2).

Dependent Variable: *inlf*			
Independent Variable	LPM (OLS)	Logit (MLE)	Probit (MLE)
nwifeinc	−.0034 (.0015)	−.021 (.008)	−.012 (.005)
educ	.038 (.007)	.221 (.043)	.131 (.025)
exper	.039 (.006)	.206 (.032)	.123 (.019)
exper^2	−.00060 (.00019)	−.0032 (.0010)	−.0019 (.0006)
age	−.016 (.002)	−.088 (.015)	−.053 (.008)
kidslt6	−.262 (.032)	−1.443 (0.204)	−.868 (.119)
kidsge6	.013 (.013)	.060 (.075)	.036 (.043)
constant	.586 (.151)	.425 (.860)	.270 (.509)
Number of observations	753	753	753
Percent correctly predicted	73.4	73.6	73.4
Log-likelihood value	—	−401.77	−401.30
Pseudo *R*-squared	.264	.220	.221

Partial Effects (Greene)

• We want to study the effect on the predicted probability of an Increase in Age. We need to calculate partial effects.



α + β_1 (Age+1) + β_2 (Income) + β_3 Sex (β_1 is positive)

Partial Effects

• Recall the β are not marginal effects. We need to calculate them with the 1st derivative of P[.]. w.r.t. $\boldsymbol{x}$. For the Logit Model:

- Partial effects: $$\frac{\partial P[y_n = j \mid x_n]}{\partial x_{nk}} = P_{nj} * (1 - P_{nj}) * \beta_k$$

- Quasi-elasticity $$\frac{\partial P(y_n = 1 \mid x_n)}{\partial x_n} x_n = P_n(1 - P_n)\beta_k x_n$$

$\Rightarrow$ Both values depend on $\boldsymbol{x}_n$. We usually evaluate these effects using sample means for $\boldsymbol{x}_n$. We can also average the partial effects over individuals.

Partial Effects

- The partial effects vary with the models:

LOGIT: $E[y \mid \bar{\mathbf{x}}] = \exp\left(\hat{\beta}'\bar{\mathbf{x}}\right) / \left[1 + \exp\left(\hat{\beta}'\bar{\mathbf{x}}\right)\right] = \Lambda\left(\hat{\beta}'\bar{\mathbf{x}}\right)$

$$\hat{\delta} = \partial E[y \mid \bar{\mathbf{x}}] / \partial \bar{\mathbf{x}} = \left[\Lambda\left(\hat{\beta}'\bar{\mathbf{x}}\right)\right]\left[1 - \Lambda\left(\hat{\beta}'\bar{\mathbf{x}}\right)\right]\hat{\beta}$$

PROBIT $E[y \mid \bar{\mathbf{x}}] = \Phi\left(\hat{\beta}'\bar{\mathbf{x}}\right)$

$$\hat{\delta} = \partial E[y \mid \bar{\mathbf{x}}] / \partial \bar{\mathbf{x}} = \left[\phi\left(\hat{\beta}'\bar{\mathbf{x}}\right)\right]\hat{\beta}$$

EXTREME VALUE $E[y \mid \bar{\mathbf{x}}] = P_1 = \exp\left[-\exp\left(-\hat{\beta}'\bar{\mathbf{x}}\right)\right]$

$$\hat{\delta} = \partial E[y \mid \bar{\mathbf{x}}] / \partial \bar{\mathbf{x}} = P_1[-\log P_1]\hat{\beta}$$

Partial Effects – Summary (Greene)

- We start with Prob[Outcome] = some $F(\alpha+\beta_1\text{Income}\ldots) = F(x'_n\beta)$
- Partial effect = $\partial F(\alpha+\beta_1\text{Income}\ldots)/\partial x = f(\alpha+\beta_1\text{Income}\ldots) \times \beta_1$
  - Partial effects are derivatives (usually, evaluated at data means.)
  - Results vary with model:

–Logit: $\partial F(\alpha+\beta_1\text{Income}\ldots)/\partial\mathbf{x}$	=	Prob * (1-Prob) $\times \beta$
–Probit: $\partial F(\alpha+\beta_1\text{Income}\ldots)/\partial\mathbf{x}$	=	Normal density $\times \beta$
–Extreme Value: $\partial F(\alpha+\beta_1\text{Income}\ldots)/\partial\mathbf{x}$	=	Prob * (-log Prob) $\times \beta$

<u>Note</u>: Scaling usually erases model differences.

- Partial effects for a Dummy Variable:

If $F(\alpha+\beta_1\text{Income}\ldots+\beta_k\text{Sex}\ldots) = F(\boldsymbol{x}_n'\beta + \boldsymbol{d}_n'\gamma)$

$\Rightarrow$ Partial effect of d = $\text{Prob}[y_n=1 \mid \boldsymbol{x}_n, d_n=1] - \text{Prob}[y_n=1 \mid \boldsymbol{x}_n, d_n=0]$

Partial Effects – Summary (Greene)

• Partial effects with non-linearities

When the model has non-linear effects, squared terms, interactive terms, the partial effects have to incorporate them.

• Suppose we have:

Prob[Outcome] = F(α + β_1Income + β_2 Age + β_3 Age^2 …)

The usual partial effects, given by a computer software, will make no sense. The software treats Age and Age^2 as two different variables.

Partial effect = ∂F(α+β_1Income…)/∂Age

= f (α+β_1Income…) * (β_2 +2 β_3 Age)

Note: Similar problem for interactive terms –say, β_4 Income x Age.

Partial Effects – Summary (Greene)

• Partial effects with interaction effects

- The partial effect calculated as before, the partial derivative of F(.) w.r.t. x_k.

- There is also an *interaction effect*: The cross derivative w.r.t the two interacted variables. Careful, if $\boldsymbol{x}_n{}'\boldsymbol{\beta}$ = 0 the effect will be non-zero!

Example: A Probit Model

$$\text{Prob} = \Phi(\alpha + \beta_1\text{Age} + \beta_2\text{Income} + \beta_3\text{Age*Income} + \ldots)$$

$$\frac{\partial \text{Prob}}{\partial \text{Income}} = \phi(\alpha + \beta_1\text{Age} + \beta_2\text{Income} + \beta_3\text{Age*Income} + \ldots)(\beta_2 + \beta_3\text{Age})$$

The "interaction effect"

$$\frac{\partial^2 \text{Prob}}{\partial \text{Income}\partial \text{Age}} = -\boldsymbol{\beta}'\mathbf{x}\ \phi(\boldsymbol{\beta}'\mathbf{x})(\beta_1 + \beta_3\text{Income})(\beta_2 + \beta_3\text{Age}) + \phi(\boldsymbol{\beta}'\mathbf{x})\beta_3$$

$$= -(\boldsymbol{\beta}'\mathbf{x})\phi(\boldsymbol{\beta}'\mathbf{x})\beta_1\beta_2 \text{ if } \beta_3 = 0. \text{ Note, nonzero even if } \beta_3 = 0.$$

Partial Effects – Standard Errors (Greene)

• Q: How do we compute SE for the partial effects?

- *Delta method*: Variance of linear approximation to non-linear function.
- *Krinsky and Robb*: Variance of a sample of draws for the underlying population of function.
- Bootstrapping: Variance of a sample replicates the underlying estimates.

Partial Effects – SE: Delta Method

• We use the delta method to calculate the standard errors.

• Delta Method Review:

$$\hat{\boldsymbol{\delta}} = f\left(\hat{\boldsymbol{\beta}}, \mathbf{x}\right)$$

$$Est.Asy.Var\left[\hat{\boldsymbol{\delta}}\right] = \left[\mathbf{G}\left(\hat{\boldsymbol{\beta}}, \mathbf{x}\right)\right] \hat{\mathbf{V}} \left[\mathbf{G}\left(\hat{\boldsymbol{\beta}}, \mathbf{x}\right)\right]$$

$$\hat{\mathbf{V}} = Est.Asy.Var\left[\hat{\boldsymbol{\beta}}\right]$$

$$\mathbf{G}\left(\hat{\boldsymbol{\beta}}, \mathbf{x}\right) = \frac{\partial f\left(\hat{\boldsymbol{\beta}}, \mathbf{x}\right)}{\partial \hat{\boldsymbol{\beta}}'}$$

Partial Effects – SE: Delta Method

• For the Logit and Probit Models we compute:

• Logit
$$\mathrm{E}[y \mid \bar{\mathbf{x}}] = \exp\left(\hat{\beta}'\bar{\mathbf{x}}\right) / \left[1 + \exp\left(\hat{\beta}'\bar{\mathbf{x}}\right)\right] = \Lambda\left(\hat{\beta}'\bar{\mathbf{x}}\right)$$
$$\hat{\delta} = \partial \mathrm{E}[y \mid \bar{\mathbf{x}}] \big/ \partial \bar{\mathbf{x}} = \left[\Lambda\left(\hat{\beta}'\bar{\mathbf{x}}\right)\right]\left[1 - \Lambda\left(\hat{\beta}'\bar{\mathbf{x}}\right)\right]\hat{\beta}$$
$$\mathbf{G} = \left[\Lambda\left(\hat{\beta}'\bar{\mathbf{x}}\right)\right]\left[1 - \Lambda\left(\hat{\beta}'\bar{\mathbf{x}}\right)\right]\left\{\mathbf{I} + \left[1 - 2\Lambda\left(\hat{\beta}'\bar{\mathbf{x}}\right)\right]\hat{\beta}\bar{\mathbf{x}}'\right\}$$

• Probit
:
$$\mathrm{E}[y \mid \bar{\mathbf{x}}] = \Phi\left(\hat{\beta}'\bar{\mathbf{x}}\right)$$
$$\hat{\delta} = \partial \mathrm{E}[y \mid \bar{\mathbf{x}}] \big/ \partial \bar{\mathbf{x}} = \left[\phi\left(\hat{\beta}'\bar{\mathbf{x}}\right)\right]\hat{\beta}$$
$$\mathbf{G} = \left[\phi\left(\hat{\beta}'\bar{\mathbf{x}}\right)\right]\left\{\mathbf{I} - \left(\hat{\beta}'\bar{\mathbf{x}}\right)\hat{\beta}\bar{\mathbf{x}}'\right\}$$

Partial Effects – SE: Krinsky & Robb

• Estimate β by Maximum Likelihood with **b**

- Estimate asymptotic covariance matrix with **V**

- Draw R observations **b**(r) from the normal population, N(**b**, **V**):

b(r) = **b** + **C** * **v**(r),

v(r) drawn from N(0, **I**) and

C = Cholesky matrix, **V** = **CC'**

- Compute partial effects **d**(r) using **b**(r)

- Compute the sample variance of **d**(r), r = 1, 2, …, R

- Use the sample standard deviations of the R observations to estimate the sampling standard errors for the partial effects.

Partial Effects – SE: Bootstrapping

For R repetitions:

- Draw N observations with replacement
- Refit the model
- Recompute the vector of partial effects
- Compute the empirical standard deviation of the R observations on the partial effects.

Partial Effects – Application I (Greene)

Binary Logit Model for Binary Choice
Dependent variable DOCTOR
Log likelihood function −2097.48109

DOCTOR	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Characteristics in numerator of Prob[DOCTOR=1]					
Constant	−.42085***	.15810	−2.66	.0078	−.73072	−.11099
AGE	.02365***	.00328	7.21	.0000	.01722	.03008
INCOME	−.44198***	.16936	−2.61	.0091	−.77393	−.11003
FEMALE	.63825***	.07551	8.45	.0000	.49026	.78624

Partial derivatives of E[y] = F[*] with respect to the vector of characteristics
Average partial effects for sample obs.

DOCTOR	Partial Effect	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval		
AGE	.00510***	.00070	7.25	.0000	.00372	.00648	
INCOME	−.09531***	.03649	−2.61	.0090	−.16684	−.02378	
FEMALE	.13849***	.01603	8.64	.0000	.10707	.16992	#

Partial effect for dummy variable is E[y|x,d=1] − E[y|x,d=0]
Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

The Linear Probability Model vs. Parametric Logit Model

DOCTOR	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
Constant	.42012***	.03468	12.12	.0000	.35216	.48808
AGE	.00504***	.00069	7.27	.0000	.00368	.00640
INCOME	−.09273**	.03697	−2.51	.0121	−.16519	−.02027
FEMALE	.13837***	.01611	8.59	.0000	.10678	.16995

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

Partial Effects – Application I (Greene)

- Now, we calculate the partial effects after scaling.

	LOGIT		PROBIT		EXTREME VALUE	
	Estimate	t ratio	Estimate	t ratio	Estimate	t ratio
Age	.00527	7.235	.00527	7.269	.00506	6.291
Income	-.09844	-2.611	-.09897	-2.636	-.09711	-2.527
Female	.14026	8.663	.13958	8.264	.13539	8.747

Note: Scaling usually erases model differences.

Partial Effects – Application I (Greene)

- Average Partial Effects vs. Partial Effects at Data Means

$$\text{Probability} = P_i = F(\beta'\mathbf{x}_i)$$

$$\text{Partial Effect} = \frac{\partial P_i}{\partial \mathbf{x}_i} = \frac{\partial F(\beta'\mathbf{x}_i)}{\partial \mathbf{x}_i} = f(\beta'\mathbf{x}_i)\times\beta = \mathbf{d}_i$$

$$\text{Average Partial Effect} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{d}_i = \beta\left(\frac{1}{n}\sum_{i=1}^{n} f(\beta'\mathbf{x}_i)\right)$$

are estimates of $\delta = E[\mathbf{d}_i]$ under certain assumptions.

Variable	Mean	Std.Dev.	S.E.Mean
ME_AGE	.00511838	.00061147	.0000106
ME_INCOM	-.0960923	.0114797	.0001987
ME_FEMAL	.137915	.0109264	.000189

PROBIT	
Estimate	t ratio
.00527	7.269
-.09897	-2.636
.13958	8.264

Note: Similar results!.

Partial Effects – Application I (Greene) – K&R

```
WALD procedure. Estimates and standard errors
for nonlinear functions and joint test of
nonlinear restrictions.
Wald Statistic               =*************
Prob. from Chi-squared[ 5] =          .00000
Krinsky-Robb method used with 1000 draws
```

Variable	Coefficient	Standard Error	b/St.Er.
Fncn(1)	.42279***	.02273	18.599
Fncn(2)	.36483***	.00351	104.020
Fncn(3)	.00527***	.00071	7.467
Fncn(4)	-.09897***	.03829	-2.585
Fncn(5)	.14114***	.01642	8.597

Note: ***, **, * = Significance at 1%, 5%, 10% lev

Delta Method

PROBIT	
Estimate	t ratio
.00527	7.269
-.09897	-2.636
.13958	8.264

Partial Effects – Standard Errors: Bootstrapping

Untitled 1 *

Insert Name:

```
procedure $
logit   ; lhs=doctor
          ;rhs=one,age,income,female
          ; prob = plogit $
create ; g = plogit*(1-plogit) $
matrix ; ape = {dpdbx = xbr(g)}*b $
endproc $
execute ; n = 50 ; bootstrap = ape $
```

Delta Method

DOCTOR	Partial Effect	Standard Error
AGE	.00510***	.00070
INCOME	-.09531***	.03649
FEMALE	.13849***	.01603

```
Results of bootstrap estimation of model.
Model has been reestimated       50 times.
Coefficients shown below are the original
model estimates based on the full sample.
Bootstrap samples have 3377 observations.
Estimated parameter vector is APE      .
Estimated variance matrix saved as VARB.
```

BootStrp	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
APE002	.00510***	.00066	7.74	.0000	.00381	.00639
APE003	-.09531***	.03353	-2.84	.0045	-.16102	-.02960
APE004	.13764***	.01787	7.70	.0000	.10262	.17266

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

Partial Effects – Application I (Greene)

• Partial Effects for the Sex dummy.

```
----------------------------------------------------------------------
Partial derivatives of E[y] = F[*]  with
respect to the vector of characteristics
They are computed at the means of the Xs
Observations used for means are All Obs.
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]  Elasticity
--------+-------------------------------------------------------------
        |Index function for probability
Constant|    -.09186***        .03550       -2.588   .0097
     AGE|     .00527***        .00073        7.269   .0000       .33855
  INCOME|    -.09897***        .03755       -2.636   .0084      -.06632
--------+-------------------------------------------------------------
        |Marginal effect for dummy variable is P|1 - P|0.
  FEMALE|     .13958***        .01618        8.624   .0000       .09745
--------+-------------------------------------------------------------
Note: ***, **, * = Significance at 1%, 5%, 10% level.
Elasticity for a binary variable = marginal effect/Mean.
----------------------------------------------------------------------
```

Partial Effects – Application I (Greene)

• Partial Effects with Non-linear effects Φ

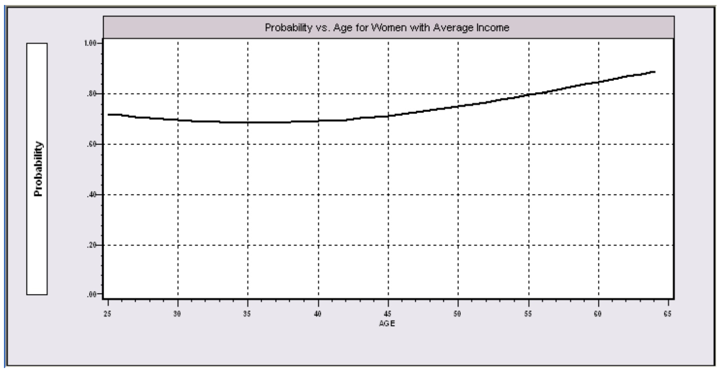
Now, we have the following Probit F, with Φ(Age; Age^2; Income; Sex)

```
----------------------------------------------------------------------
Binomial Probit Model
Dependent variable                  DOCTOR
Log likelihood function        -2086.94545
Restricted log likelihood      -2169.26982
Chi squared [   4 d.f.]          164.64874
Significance level                  .00000
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
        |Index function for probability
Constant|    1.30811***        .35673        3.667   .0002
     AGE|    -.06487***        .01757       -3.693   .0002     42.6266
   AGESQ|     .00091***        .00020        4.540   .0000     1951.22
  INCOME|    -.17362*          .10537       -1.648   .0994      .44476
  FEMALE|     .39666***        .04583        8.655   .0000      .46343
--------+-------------------------------------------------------------
Note: ***, **, * = Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------
```

Partial Effects – Application I (Greene)

• Partial Effects with Non-linear effects

The probability implied by the model is.



Partial Effects – Application I (Greene)

• Partial Effects with Non-linear effects

Partial derivatives of E[y] = F[*] with
respect to the vector of characteristics
They are computed at the means of the Xs
Observations used for means are All Obs.

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Elasticity
	Index function for probability				
AGE	-.02363***	.00639	-3.696	.0002	-1.51422
AGESQ	.00033***	.729872D-04	4.545	.0000	.97316
INCOME	-.06324*	.03837	-1.648	.0993	-.04228
	Marginal effect for dummy variable is P|1 - P|0.				
FEMALE	.14282***	.01620	8.819	.0000	.09950

<u>Note</u>: “Usual” partial (separate) effects for Age and Age^2 make no sense. They are not varying “partially.”

Partial Effects – Application I (Greene)

• Partial Effects with Interaction terms

We estimate a Probit model, with Φ(Age; Income; Income*Age; Sex)

Note: The software does not know that Age_Inc = Age*Income.

```
----------------------------------------------------------------------
Partial derivatives of E[y] = F[*]  with
respect to the vector of characteristics
They are computed at the means of the Xs
Observations used for means are All Obs.
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er.  P[|Z|>z]  Elasticity
--------+-------------------------------------------------------------
        |Index function for probability
Constant|    -.18002**         .07421       -2.426   .0153
     AGE|     .00732***        .00168        4.365   .0000       .46983
  INCOME|     .11681           .16362         .714   .4753       .07825
 AGE_INC|    -.00497           .00367       -1.355   .1753      -.14250
        |Marginal effect for dummy variable is P|1 - P|0.
  FEMALE|     .13902***        .01619        8.586   .0000       .09703
--------+-------------------------------------------------------------
```

Binary Data – Model Evaluation

• GoF Measures

- Calculated from the Log L.
  - "Pseudo R squared" = 1 – log L/log L0
  - LR tests
  - Information Criteria, especially for non-nested models.

• Forecasting accuracy/Model evaluation

- Predictions : $y_n = 1$ if $F(\boldsymbol{x}_n'\boldsymbol{\beta}) > c$ (e.g. 0.5)

  $y_n = 0$ if $F(\boldsymbol{x}_n'\boldsymbol{\beta}) \leq c$
- Compute *hit rate* = % of correct predictions
- Many measures: Cramer, Efron, Veall and Zimmerman.

Binary Data – Model Evaluation I (Greene)

• GoF Measures

```
----------------------------------------------------------------------
Binary Logit Model for Binary Choice
Dependent variable                DOCTOR
Log likelihood function      -2085.92452  <--  Full model         LogL
Restricted log likelihood    -2169.26982  <--  Constant term only LogL0
Chi squared [   5 d.f.]        166.69058
Significance level               .00000
McFadden Pseudo R-squared       .0384209  <--  1 - LogL/logL0
Estimation based on N =    3377, K =    6
Information Criteria: Normalization=1/N
              Normalized   Unnormalized
AIC              1.23892     4183.84905      -2LogL + 2K
Fin.Smpl.AIC     1.23893     4183.87398      -2LogL + 2K + 2K(K+1)/(N-K-1)
Bayes IC         1.24981     4220.59751      -2LogL + KlnN
Hannan Quinn     1.24282     4196.98802      -2LogL + 2Kln(lnN)
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
        |Characteristics in numerator of Prob[Y = 1]
Constant|    1.86428***      .67793        2.750   .0060
     AGE|    -.10209***      .03056       -3.341   .0008      42.6266
   AGESQ|     .00154***      .00034        4.556   .0000      1951.22
  INCOME|     .51206         .74600         .686   .4925       .44476
 AGE_INC|    -.01843         .01691       -1.090   .2756      19.0288
  FEMALE|     .65366***      .07588        8.615   .0000       .46343
--------+-------------------------------------------------------------
```

Binary Data – Model Evaluation

• Fit Measures Based on Predictions

- Cramer Fit Measure:

$\hat{F}$ = Predicted Probability

$$\hat{\lambda} = \frac{\Sigma_{i=1}^{N} y_i \hat{F}}{N_1} - \frac{\Sigma_{i=1}^{N}(1-y_i)\hat{F}}{N_0}$$

$$\hat{\lambda} = \left(\text{Mean } \hat{F} \mid \text{when } y=1\right) - \left(\text{Mean } \hat{F} \mid \text{when } y=0\right)$$

= reward for correct predictions minus
penalty for incorrect predictions

```
+----------------------------------------+
| Fit Measures Based on Model Predictions|
| Efron                       =    .04825|
| Ben Akiva and Lerman        =    .57139|
| Veall and Zimmerman         =    .08365|
| Cramer                      =    .04771|
+----------------------------------------+
```

Binary Data – Application III - PIM

Example (Bucklin and Gupta (1992)): Purchase Incidence Model

$$p_t^n \text{ (inc)} = \frac{exp(W_t^n)}{1 + exp(W_t^n)}$$

p_t^n (inc) = Probability that household n engages in a category purchase in the store on purchase occasion t.

W_t^n = Utility of the purchase option. Let W_t^n follow

$$W_t^n = \gamma_0 + \gamma_1 CR^n + \gamma_2 INV_t^n + \gamma_3 CV_t^n + \varepsilon_t^n$$

where

CR^n = rate of consumption for household *n*

INV_t^n = inventory level for household *n*, time *t*

CV_t^n = category value for household *n*, time *t*

Binary Data – Application III - PIM

• Goodness-of-Fit

Model	# param.	LogL	U^2 (pseudo R^2)	BIC
Null model	1	-13614.4	-	13619.6
Full model	4	-11234.5	.175	11255.2

• Parameter estimates

Parameter	Estimate (t-statistic)
Intercept γ_0	-4.521 (-27.70)
CR γ_1	.549 (4.18)
INV γ_2	-.520 (-8.91)
CV γ_3	.410 (8.00)

Binary Data – Application IV - Ketchup

• Binary Logit Model (Franses and Paap (2001): www.few.eur.nl/few/people/paap)

• Data
  – A.C.Nielsen scanner panel data
  – 117 weeks: 65 for initialization, 52 for estimation
  – 565 households: 300 selected randomly for estimation, remaining hh = holdout sample for validation
  – Data set for estimation: 30.966 shopping trips, 2275 purchases in the category (liquid laundry detergent)
  – Estimation limited to the 7 top-selling brands (80% of category purchases), representing 28 brand-size combinations (= level of analysis for the choice model)

Binary Data – Application IV - Ketchup

• ML Estimation

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.222121	0.668483	0.332277	0.7397
DISPLHEINZ	0.573389	0.239492	2.394186	0.0167
DISPLHUNTS	-0.557648	0.247440	-2.253674	0.0242
FEATHEINZ	0.505656	0.313898	1.610896	0.1072
FEATHUNTS	-1.055859	0.349108	-3.024445	0.0025
FEATDISPLHEINZ	0.428319	0.438248	0.977344	0.3284
FEATDISPLHUNTS	-1.843528	0.468883	-3.931748	0.0001
PRICEHEINZ	-135.1312	10.34643	-13.06066	0.0000
PRICEHUNTS	222.6957	19.06951	11.67810	0.0000

Binary Data – Application III - Ketchup

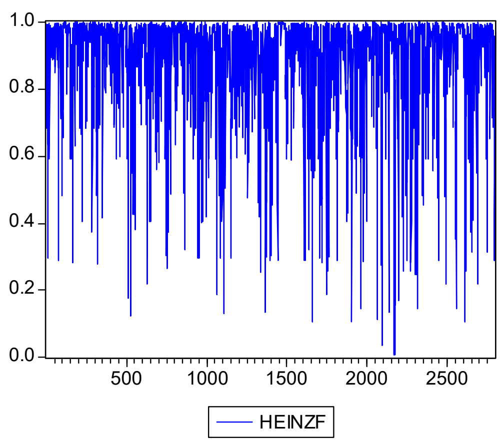
• Model Evaluation

Mean dependent var	0.890279	S.D. dependent var	0.312598
S.E. of regression	0.271955	Akaike info criterion	0.504027
Sum squared resid	206.2728	Schwarz criterion	0.523123
Log likelihood	-696.1344	Hannan-Quinn criter.	0.510921
Restr. log likelihood	-967.918	Avg. log likelihood	-0.248797
LR statistic (8 df)	543.5673	McFadden R-squared	0.280792
Probability(LR stat)	0.000000		

Obs with Dep=0	307	Total obs	2798
Obs with Dep=1	2491		

Binary Data – Application III - Ketchup

• Model Evaluation: Forecast Accuracy



Forecast: HEINZF
Actual: HEINZ
Forecast sample: 1 2798
Included observations: 2798

Root Mean Squared Error	0.271517
Mean Absolute Error	0.146875
Mean Abs. Percent Error	7.343760
Theil Inequality Coefficient	0.146965
Bias Proportion	0.000000
Variance Proportion	0.329791
Covariance Proportion	0.670209

Binary Data – Application III - Ketchup

• Model Evaluation: Aggregate Predictions
We judge the forecast accuracy based on the hit rate.

Classification Table[a]

			Predicted		
			HE		Percentage Correct
Observed			,00	1,00	
Step 1	HE	,00	81	226	26,4
		1,00	34	2457	98,6
	Overall Percentage				90,7

a. The cut value is ,500

Binary Data – Testing (Greene)

• ML estimation framework. Testing is based on the ML trilogy:
- LR Test, Wald Statistics and LM Tests.
• Different from regression (no residuals!). There is no F statistic.

Example: Base Model

```
---------------------------------------------------------------------
Binary Logit Model for Binary Choice
Dependent variable                 DOCTOR
Log likelihood function       -2085.92452
Restricted log likelihood     -2169.26982
Chi squared [   5 d.f.]         166.69058
Significance level                 .00000
McFadden Pseudo R-squared        .0384209
Estimation based on N =   3377, K =   6
--------+------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er.  P[|Z|>z]   Mean of X
--------+------------------------------------------------------------
        |Characteristics in numerator of Prob[Y = 1]
Constant|     1.86428***        .67793        2.750   .0060
     AGE|    -.10209***         .03056       -3.341   .0008      42.6266
   AGESQ|     .00154***         .00034        4.556   .0000      1951.22
  INCOME|     .51206            .74600         .686   .4925       .44476
 AGE_INC|    -.01843            .01691       -1.090   .2756      19.0288
  FEMALE|     .65366***         .07588        8.615   .0000       .46343
--------+------------------------------------------------------------
```

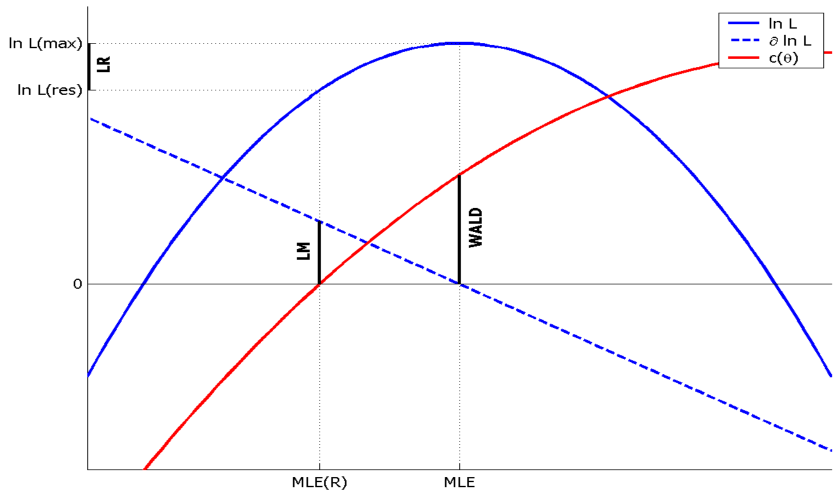
H_0: Age is not a significant determinant of Prob(Doctor = 1)

H_0: $\beta_2 = \beta_3 = \beta_5 = 0$

H_1: at least one β_2 , β_3 and/or $\beta_5 \neq 0$

Binary Data – Testing (Greene)

Testing is based on the ML trilogy: LR, Wald and LM Tests.



Binary Data – LR Test (Greene)

• LR Test: Based on Unrestricted and Restricted models

- LR-test statistic = 2 (LogL|H_1 - Unrestricted model – LogL|H_0 - Restrictions) $\geq 0 \sim \chi^2_3$

UNRESTRICTED MODEL	
Binary Logit Model for Binary Choice	
Dependent variable	DOCTOR
Log likelihood function	-2085.92452
Restricted log likelihood	-2169.26982
Chi squared [5 d.f.]	166.69058
Significance level	.00000
McFadden Pseudo R-squared	.0384209
Estimation based on N =	3377, K = 6

RESTRICTED MODEL	
Binary Logit Model for Binary Choice	
Dependent variable	DOCTOR
Log likelihood function	-2124.06568
Restricted log likelihood	-2169.26982
Chi squared [2 d.f.]	90.40827
Significance level	.00000
McFadden Pseudo R-squared	.0208384
Estimation based on N =	3377, K = 3

LR-test = 2[-2085.92452 - (-2124.06568)] = 77.46456

$\Rightarrow$ reject H_0

Binary Data – Wald Testing (Greene)

- Wald Test: Based on Unrestricted model
  - Discrepancy: $\mathbf{q} = \mathbf{Rb} - \mathbf{m}$ (or $\mathbf{r}(\mathbf{b},\mathbf{m})$ if nonlinear) is computed
  - Variance of discrepancy is estimated
  - Wald Statistic is $\mathbf{q}'[\mathrm{Var}(\mathbf{q})]^{-1}\mathbf{q} \sim \chi^2_{df=3}$

```
-----------------------------------------------------------------------------
Binary Logit Model for Binary Choice
Dependent variable               DOCTOR
Log likelihood function     -2085.92452
Restricted log likelihood   -2169.26982
Chi squared [   5 d.f.]       166.69058
Significance level               .00000
McFadden Pseudo R-squared      .0384209
Estimation based on N =   3377, K =   6
Inf.Cr.AIC  =   4183.8 AIC/N =    1.239
Wald test of        3 linear restrictions
Chi-squared =   69.05, P value =   .0000
--------+--------------------------------------------------------------------
        |                Standard            Prob.       95% Confidence
  DOCTOR|  Coefficient     Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
        |Characteristics in numerator of Prob[DOCTOR=1]
Constant|    1.86428***    .67793    2.75  .0060      .53557   3.19299
     AGE|    -.10209***    .03056   -3.34  .0008    -.16199   -.04219
   AGESQ|     .00154***    .00034    4.56  .0000     .00088    .00220
  INCOME|     .51206       .74600     .69  .4925    -.95008   1.97420
 AGE_INC|    -.01843       .01691   -1.09  .2756    -.05157    .01470
  FEMALE|     .65366***    .07588    8.61  .0000     .50494    .80237
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Binary Data – Testing (Greene)

- Wald Test – Matrix Computation

Matrix - B [6, 1]

	1
1	1.86428
2	-0.102093
3	0.00154004
4	0.512059
5	-0.018433
6	0.653659

Matrix - VARB [6, 6]

	1	2	3	4	5	6
1	0.459584	-0.019886	0.000198178	-0.231031	0.00556395	0.00125442
2	-0.019886	0.000934022	-9.98629e-006	0.00516338	-0.00014392	-0.000160159
3	0.000198178	-9.98629e-006	1.14272e-007	-1.03054e-006	2.83849e-007	1.56007e-006
4	-0.231031	0.00516338	-1.03054e-006	0.556523	-0.0122686	-0.000805883
5	0.00556395	-0.00014392	2.83849e-007	-0.0122686	0.000285847	2.77535e-005
6	0.00125442	-0.000160159	1.56007e-006	-0.000805883	2.77535e-005	0.00575729

Matrix - R [3, 6] Cell: 0

	1	2	3	4	5	6
1	0	1	0	0	0	0
2	0	0	1	0	0	0
3	0	0	0	0	1	0

Matrix - Q [3, 1] Cell: -0.102093

	1
1	-0.102093
2	0.00154004
3	-0.018433

Matrix - VQ [3, 3] Cell: 0.000934022

	1	2	3
1	0.000934022	-9.98629e-006	-0.00014392
2	-9.98629e-006	1.14272e-007	2.83849e-007
3	-0.00014392	2.83849e-007	0.000285847

Matrix - WALDSTAT [1, 1] Cell: 69.0541

	1
1	69.0541

Wald Test = 69.0541 $\Rightarrow$ reject H_0

Binary Data – Test (Greene)

• LM Test: Based on Restricted model
  - Derivatives of unrestricted model and variances of derivatives are computed at restricted estimates
  - Wald test of whether derivatives are zero tests the restrictions
  - Usually hard to compute – difficult to program the derivatives and their variances.

Example: Computation for Logit Model

- Compute $\mathbf{b}_0$ subject to H_0 –i.e., with zeros in appropriate positions.
- Compute $P_i(\mathbf{b}_0)$ for each observation.
- Compute $e_i(\mathbf{b}_0) = [x_i - P_i(\mathbf{b}_0)]$
- Compute $g_i\ (\mathbf{b}_0) = x_i e_i$ using full x_i vector

$$LM = [\Sigma_i \mathbf{g}_i(\mathbf{b}_0)]'[\Sigma_i \mathbf{g}_i(\mathbf{b}_0)\mathbf{g}_i(\mathbf{b}_0)]^{-1}[\Sigma_i \mathbf{g}_i(\mathbf{b}_0)] \sim \chi^2_{df=3}$$

Binary Data – LM Test (Greene)

• LM Test: Compuations

```
Matrix DERIV    has  6 rows and  1 columns.
        +-------------+
       1|  .2393443D-05  <-  zero from FOC
       2| 2268.60186
       3|  .2122049D+06
       4|  .9683957D-06  <-  zero from FOC
       5|  849.70485
       6|  .2380413D-05  <-  zero from FOC
        +-------------+
```

```
Matrix LM        has  1 rows and  1 columns.
                1
        +-------------+
       1|   81.45829  |
        +-------------+
```

=> reject H_0

Summary: LM test = 81.45829

LR-test = 77.46456

Wald test = 69.0541

Binary Data – Chow Test I (Greene)

• Health Satisfaction: Panel Data – 1984,85,86,87,88,1991,1994

To test parameter constancy over time, we do a Chow test: 1) Fit a model for each year (7 years total); 2) Fit a pooled model. 3) Do a LR test. LR ~ $\chi^2_{df=36}$

Healthy(0/1) = f(1, Age, Educ, Income, Married(0/1), Kids(0.1)

Year	Log Likelihood Function	Sample Size
1984	-2395.137	3874
1985	-2375.090	3794
1986	-2387.602	3792
1987	-2337.835	3666
1988	-2890.288	4483
1991	-2769.375	4340
1994	-2168.998	3377
Pool	-17365.76	27326

The log likelihood for the pooled sample is -17365.76. The sum of the log likelihoods for the seven individual years is -17324.33. Twice the difference is 82.87. The degrees of freedom is 6×6 = 36. The 95% critical value from the chi squared table is 50.998, so the pooling hypothesis is rejected.

Binary Data – Chow Test II (Greene)

• Determinants of Doctor's visits.

To test parameter constancy over groups (male, female), we do a Chow test: 1) Fit a model for males & another for females; 2) Fit a pooled model. 3) Do a LR test. LR ~ $\chi^2_{df=5}$

```
-----------------------------------------------------------------
Dependent variable                 DOCTOR
Pooled Log likelihood function    -2123.84754
--------+--------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]
--------+--------------------------------------------------------
Constant|    1.76536***        .67060         2.633    .0085
     AGE|    -.08577***        .03018        -2.842    .0045
   AGESQ|     .00139***        .00033         4.168    .0000
  INCOME|     .61090           .74073          .825    .4095
 AGE_INC|    -.02192           .01678        -1.306    .1915
--------+--------------------------------------------------------
Male Log likelihood function      -1198.55615
--------+--------------------------------------------------------
Female Log likelihood function     -885.19118
--------+--------------------------------------------------------
 LR-test = 2[-885.19118+(-1198.55615)-(-2123.84754] = 80.2004
```

Specification Issues

• Main issues
- Neglected heterogeneity
- Omitted variables
- Endogeneity
  - These problems are relevant for all index models
  -Since the normal distribution allows us to obtain concrete results, the focus is on Probit models.

• In linear models:
- Heterogeneity causes OLS to be inefficient, though it is still consistent and unbiased.

- Omitted variables can lead to inconsistent estimates, unless...
  - The omitted variable does not affect y
  - The omitted variable is uncorrelated with x

Hetersocedasticity (Greene)

• Scaling each individual by its variance.

Steps:

(1) Parameterize: Var[ε_n] = exp($\boldsymbol{z}_n{}'\gamma$)

(2) Reformulate probabilities

Binary Probit or Logit: $P_n[y_n = 1 | \boldsymbol{x}_n] = P(\boldsymbol{x}_n{}'\boldsymbol{\beta} / \exp(\boldsymbol{z}_n{}'\gamma))$

• Marginal effects are more complicated. If $\mathbf{x}_n = \mathbf{z}_n$, signs and magnitudes of marginal effects tend to be ambiguous.

Example: For the univariate case:

$E[y_n | \boldsymbol{x}_n, \boldsymbol{z}_n] = \Phi[\boldsymbol{x}_n{}'\boldsymbol{\beta}/\exp(\boldsymbol{z}_n{}'\gamma)]$

$\partial E[y_n | \boldsymbol{x}_n, \boldsymbol{z}_n]/\partial \boldsymbol{x}_n = \varphi[\boldsymbol{x}_n{}'\boldsymbol{\beta}/\exp(\boldsymbol{z}_n{}'\gamma)] * \boldsymbol{\beta}$

$\partial E[y_n | \boldsymbol{x}_n, z_n]/\partial \boldsymbol{z}_n = \varphi[\boldsymbol{x}_n{}'\boldsymbol{\beta}/\exp(\boldsymbol{z}_n{}'\gamma)] * [-\boldsymbol{x}_n{}'\boldsymbol{\beta}/\exp(\boldsymbol{z}_n{}'\gamma)] * \boldsymbol{\gamma}$

Hetersocedasticity (Greene)

• Scaling with a dummy variable. For example,

Var$[\varepsilon_n]$ = exp$(\boldsymbol{z}_n'\gamma)$

$$\text{Prob(Doctor=1)} = F\left(\frac{\beta'\mathbf{x}_i}{\exp(\gamma\text{Female}_i)}\right) \text{ is equivalent to}$$

$$\text{Prob(Doctor=1)} = F\left(\beta'\mathbf{x}_i\right) \text{ for men}$$

$$\text{Prob(Doctor=1)} = F\left(\lambda\beta'\mathbf{x}_i\right) \text{ for women where } \lambda = e^{-\gamma}$$

Heteroscedasticity of this type is equivalent to an implicit scaling of the preference structure for the two (or G) groups.

Hetersocedasticity – Application I

• Determinants of Doctor's visits (Logit Model).

Model for Variance: Var$[\varepsilon_n]$ = exp(Female´γ)

```
----------------------------------------------------------------------
Binary Logit Model for Binary Choice
Dependent variable               DOCTOR
Log likelihood function     -2096.42765
Restricted log likelihood   -2169.26982
Chi squared [   4 d.f.]       145.68433
Significance level               .00000
McFadden Pseudo R-squared      .0335791
Estimation based on N =    3377, K =   6
Heteroscedastic Logit Model for Binary Data
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
        |Characteristics in numerator of Prob[Y = 1]
Constant|    1.31369***       .43268        3.036   .0024
     AGE|    -.05602***       .01905       -2.941   .0033       42.6266
   AGESQ|     .00082***       .00021        3.838   .0001       1951.22
  INCOME|     .11564          .47799         .242   .8088        .44476
 AGE_INC|    -.00704          .01086        -.648   .5172       19.0288
        |Disturbance Variance Terms
  FEMALE|    -.81675***       .12143       -6.726   .0000        .46343
--------+-------------------------------------------------------------
```

Hetersocedasticity – Partial Effects Application I

• Determinants of Doctor's visits (Logit Model).

Model for Variance: Var[ε_n] = exp(Female´γ)

Partial derivatives of probabilities with respect to the vector of characteristics. They are computed at the means of the Xs. Effects are the sum of the mean and variance term for variables which appear in both parts of the function.

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Elasticity
AGE	-.02121***	.00637	-3.331	.0009	-1.32701
AGESQ	.00032***	.717036D-04	4.527	.0000	.92966
INCOME	.13342	.15190	.878	.3797	.08709
AGE_INC	-.00439	.00344	-1.276	.2020	-.12264
FEMALE	.19362***	.04043	4.790	.0000	.13169
Disturbance Variance Terms					
FEMALE	-.05339	.05604	-.953	.3407	-.03632
Sum of terms for variables in both parts					
FEMALE	.14023***	.02509	5.588	.0000	.09538
Marginal effect for variable in probability – Homoscedastic Model					
AGE	-.02266***	.00677	-3.347	.0008	-1.44664
AGESQ	.00034***	.747582D-04	4.572	.0000	.99890
INCOME	.11363	.16552	.687	.4924	.07571
AGE_INC	-.00409	.00375	-1.091	.2754	-.11660
Marginal effect for dummy variable is P|1 - P|0.					
FEMALE	.14306***	.01619	8.837	.0000	.09931

Hetersocedasticity Test (Greene)

• Determinants of Doctor's visits (Probit Model).

To test for heteroscedasticity, we do a LR test: 1) Fit restricted model (H_0: No heteroscedasticity), and 2) Fit unrestricted model (H_1: Var[ε_n] = exp($\boldsymbol{z}_n{}'\gamma$)). Then, 3) Do a LR test. LR ~ $\chi^2_{df=4}$

	Heteroscedastic				Homoscedastic				
	LogL = -2888.328				LogL = -2890.288				
	LogLR = -2890.288				LogL0 = -3010.421				
	Chisq = 3.920				Chisq = 240.266				
	Wald = 3.728								
	LM = 3.858								
Variable	Coef.	S.E.	t	P	Coef.	S.E.	t	P	Mean of X
Constant	.7595	.2349	3.233	.0012	.4816	.1423	3.383	.0007	1.0000
AGE	-.0329	.0032	-10.266	.0000	-.0203	.0020	-10.386	.0000	43.4401
EDUC	.0860	.0148	5.805	.0000	.0520	.0089	5.872	.0000	11.4181
INCOME	.3454	.2083	1.658	.0972	.2180	.1265	1.724	.0847	.34874
MARRIED	-.0483	.0828	-.584	.5592	-.0311	.0508	-.612	.5403	.75217
KIDS	.1278	.0756	1.692	.0907	.0800	.0463	1.727	.0841	.37943
	Variance Function								
INCOME	.0141	.5193	.027	.9784					.34874
KIDS	-.1608	.1975	-.814	.4158					.37943
FEMALE	.0291	.1073	.271	.7864					.48405
WORKING	-.1831	.1350	-1.356	.1750					.67232
	Partial Effects				Partial Effects				
AGE	-.0080	.0008	-9.469	.0000	-.0078	.0008	-10.392	.0000	43.4401
EDUC	.0190	.0035	5.443	.0000	.0200	.0034	5.875	.0000	11.4181
INCOME	.0859	.1539	.558	.5769	.0838	.0486	1.724	.0847	.34874
MARRIED	-.0171	.0217	-.789	.4301	-.0119	.0194	-.614	.5394	.75217
KIDS	.0314	.0478	.657	.5113	.0307	.0177	1.733	.0831	.37943
FEMALE	-.0029	.0104	-.282	.7779					.48405
WORKING	.0184	.0186	.989	.3227					.67232

Binary Data – Robust Covariance Matrix

• Determinants of Doctor's visits (Probit Model).
We calculate the "Robust" Covariance Matrix: Var[**b**] = $\boldsymbol{A}^{-1}\boldsymbol{B}\,\boldsymbol{A}^{-1}$

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
	Robust Standard Errors				
Constant	1.86428***	.68442	2.724	.0065	
AGE	-.10209***	.03115	-3.278	.0010	42.6266
AGESQ	.00154***	.00035	4.446	.0000	1951.22
INCOME	.51206	.75103	.682	.4954	.44476
AGE_INC	-.01843	.01703	-1.082	.2792	19.0288
FEMALE	.65366***	.07585	8.618	.0000	.46343
	Conventional Standard Errors Based on Second Derivatives				
Constant	1.86428***	.67793	2.750	.0060	
AGE	-.10209***	.03056	-3.341	.0008	42.6266
AGESQ	.00154***	.00034	4.556	.0000	1951.22
INCOME	.51206	.74600	.686	.4925	.44476
AGE_INC	-.01843	.01691	-1.090	.2756	19.0288
FEMALE	.65366***	.07588	8.615	.0000	.46343

• Not a big difference. Harmless to use the robust estimator.

Odds Ratio

• A popular descriptive statistics is the odds ratio. The odds ratio is just the ratio of two choice probabilities:

$$\text{Odds Ratio} = \frac{P(y_n = 1 \mid x_n)}{P(y_n = 0 \mid x_n)}$$

• For the Logit Model the odds ratio is very simple to calculate:

$$\frac{P(y_n = 1 \mid x_n)}{P(y_n = 0 \mid x_n)} = \frac{e^{X_n\beta}/(1+e^{X_n\beta})}{1/(1+e^{X_n\beta})} = e^{X_n\beta}$$

• We may be interested in measuring the effect of a unit change in the odds ratio. Simple to do for a dummy variable (D=1 to D=0). For the Logit Model, this ratio simplifies to exp(coeff. of dummy):

$$\text{Ratio of Odds Ratio} = \frac{e^{X_n\beta+\gamma D}/(1+e^{X_n\beta+\gamma D})}{1/(1+e^{X_n\beta+\gamma D})} \Big/ \frac{e^{X_n\beta}/(1+e^{X_n\beta})}{1/(1+e^{X_n\beta})} = e^{\gamma}$$

Odds Ratio – Application (Greene)

• We are interested in estimating the change in odds of buying public insurance for a female headed household (D=1) compared to a male headed household (D=0). For the Logit Model:

Odds ratio: exp(.23427) = 1.26399 (odds up by 26%)

Binomial Probit Model
Dependent variable PUBLIC

PUBLIC	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Index function for probability					
Constant	3.46197***	.07059	49.04	.0000	3.32361	3.60033
AGE	.00169*	.00100	1.70	.0894	-.00026	.00365
EDUC	-.16763***	.00408	-41.09	.0000	-.17563	-.15963
HHNINC	-.99851***	.05497	-18.16	.0000	-1.10625	-.89077
FEMALE	.11847***	.02209	5.36	.0000	.07518	.16176

Binary Logit Model for Binary Choice

PUBLIC	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
	Characteristics in numerator of Prob[PUBLIC=1]					
Constant	6.13526***	.13143	46.68	.0000	5.87767	6.39285
AGE	.00274	.00190	1.44	.1494	-.00098	.00646
EDUC	-.29933***	.00711	-42.09	.0000	-.31327	-.28539
HHNINC	-1.78997***	.10129	-17.67	.0000	-1.98849	-1.59145
FEMALE	.23427***	.04239	5.53	.0000	.15118	.31737

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

Endogeneity

• In the doctor's visits problem, we want to study the effect of public health insurance, h, on doctor's visits, y_n. But, individuals also make a decision to take public health insurance or not.

$\Rightarrow$ endogeneity problem!

• Two cases: (1) h is continuous (complicated); (2) h is binary (easier).

• There are many approaches to estimate this problem: ML. GMM, ad-hoc solutions, especially for case (2).

• We focus on MLE. It requires full specification of the model, including the assumption that underlies the endogeneity of h_n.

• We present an example for the Probit Model.

Endogeneity – ML – Continuous Variable

- **CASE 1** – h continuous
- Full specification:

- RUM: $U_n = \boldsymbol{x}_n{}'\boldsymbol{\beta} + h_n{}'\boldsymbol{\theta} + \varepsilon_n$

- Revealed preference: $y_n = 1$ if $U_n > 0$

- Endogenous variable(s): $h_n = \boldsymbol{z}_n{}'\boldsymbol{\alpha} + u_n$,

with E$[\varepsilon_n \mid h] \neq 0 \Leftrightarrow$ Cov$[u_n, \varepsilon_n] \neq 0$ (ρ=Corr$[u_n, \varepsilon_n]$)

- Additional Assumptions:

1) $u_n, \varepsilon_n \sim \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_u \\ \rho\sigma_u & \sigma_u^2 \end{pmatrix}\right]$

2) $\boldsymbol{z}$ = IV (exogenous variables), uncorrelated with (u_n, ε_n)

- ML becomes a simultaneous equations model.

Endogeneity - ML

- FIML estimation. Steps:
  - Write down the joint density: $f(y_n \mid \boldsymbol{x}_n, \boldsymbol{z}_n) * f(\boldsymbol{z}_n)$
  - Assume a Probit Model $\Rightarrow$ Normal for $f(y_n \mid \boldsymbol{x}_n, \boldsymbol{z}_n)$:
    P$[y_n=1 \mid \mathbf{x}_n, \mathbf{z}_n] = \Phi(\boldsymbol{x}_n{}'\boldsymbol{\beta} + h_n{}'\boldsymbol{\theta} + \varepsilon_n)$.
  - Assume marginal for $f(\boldsymbol{z}_n)$, a normal distribution.
  - Use the projection:
    $\varepsilon_n \mid u_n = [(\rho\sigma_u)/\sigma_u^2]\, u_n + v_n$, with $\sigma_v^2 = (1 - \rho^2)$.
  - Insert projection in
    P$[y_n=1 \mid \boldsymbol{x}_n, \boldsymbol{z}_n] = \Phi(\boldsymbol{x}_n{}'\boldsymbol{\beta} + h_n{}'\boldsymbol{\theta} + [(\rho\sigma_u)/\sigma_u^2]\, u_n)$
  - Replace $u_n = h_n - \boldsymbol{z}_n{}'\boldsymbol{\alpha}$ in P(y_n).
  - Maximize Log L(.) w.r.t. $(\boldsymbol{\beta}, \boldsymbol{\alpha}, \theta, \rho, \sigma_u)$

Endogeneity – ML: Probit (Greene)

Probit fit of y to $\mathbf{x}$ and h will not consistently estimate $(\boldsymbol{\beta},\theta)$ because of the correlation between h and ε induced by the correlation of u and ε. Using the bivariate normality,

$$\text{Prob}(y=1\,|\,\mathbf{x},h)=\Phi\left[\frac{\boldsymbol{\beta}'\mathbf{x}+\theta h+(\rho/\sigma_u)u}{\sqrt{1-\rho^2}}\right]$$

Insert $u_i = (h_i - \boldsymbol{\alpha}'\mathbf{z})/\sigma_u$ and include f(h|$\mathbf{z}$) to form logL

$$\text{logL}=\sum_{i=1}^{N}\left\{\begin{array}{l}\log\Phi\left[(2y_i-1)\left(\dfrac{\boldsymbol{\beta}'\mathbf{x}_i+\theta h_i+\rho\left(\dfrac{h_i-\boldsymbol{\alpha}'\mathbf{z}_i}{\sigma_u}\right)}{\sqrt{1-\rho^2}}\right)\right]+\\ \log\dfrac{1}{\sigma_u}\phi\left[\left(\dfrac{h_i-\boldsymbol{\alpha}'\mathbf{z}_i}{\sigma_u}\right)\right]\end{array}\right\}$$

Endogeneity - ML

• Two step limited information ML (Control Function) is also possible:

- Use OLS to estimate $\boldsymbol{\alpha}$, σ_u ⇒ get estimates a and s.
- Compute the residual u_n. Standardize them: $\hat{u}/s$
- Plug residuals $\hat{u}/s$ into Φ.
- Fit the Probit Model.
- Transform the estimated coefficients into the structural ones.
- Use delta method to calculate standard errors.

Endogeneity – ML - Application

• Health Satisfaction example, with income as endogenous variable.

```
--------------------------------------------------------------------
Probit with Endogenous RHS Variable
Dependent variable              HEALTHY
Log likelihood function      -6464.60772
--------+-----------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-----------------------------------------------------------
        |Coefficients in Probit Equation for HEALTHY
Constant|    1.21760***        .06359       19.149   .0000
     AGE|    -.02426***        .00081      -29.864   .0000      43.5257
 MARRIED|    -.02599           .02329       -1.116   .2644       .75862
  HHKIDS|     .06932***        .01890        3.668   .0002       .40273
  FEMALE|    -.14180***        .01583       -8.959   .0000       .47877
  INCOME|     .53778***        .14473        3.716   .0002       .35208
        |Coefficients in Linear Regression for INCOME
Constant|    -.36099***        .01704      -21.180   .0000
     AGE|     .02159***        .00083       26.062   .0000      43.5257
   AGESQ|    -.00025***     .944134D-05    -26.569   .0000      2022.86
    EDUC|     .02064***        .00039       52.729   .0000      11.3206
 MARRIED|     .07783***        .00259       30.080   .0000       .75862
  HHKIDS|    -.03564***        .00232      -15.332   .0000       .40273
  FEMALE|     .00413**         .00203        2.033   .0420       .47877
        |Standard Deviation of Regression Disturbances
Sigma(w)|     .16445***        .00026      644.874   .0000
        |Correlation Between Probit and Regression Disturbances
Rho(e,w)|    -.02630           .02499       -1.052   .2926
--------+-----------------------------------------------------------
```

Endogeneity – Partial Effects (Greene)

• Partial effects have to be re-scaled.

Conditional Mean

$$E[y\,|\,\mathbf{x},h] = \Phi(\beta'\mathbf{x} + \theta h)$$

$$h = \alpha'\mathbf{z} + u = \alpha'\mathbf{z} + \sigma_u v \text{ where } v \sim N[0,1]$$

$$E[y|x,\mathbf{z},v] = \Phi[\beta'\mathbf{x} + \theta(\alpha'\mathbf{z} + \sigma_u v)]$$

Partial Effects. Assume z = x (just for convenience)

$$\frac{\partial E[y|\mathbf{x},\mathbf{z},v]}{\partial \mathbf{x}} = \phi[\beta'\mathbf{x} + \theta(\alpha'\mathbf{z} + \sigma_u v)](\beta + \theta\alpha)$$

$$\frac{\partial E[y|\mathbf{x},\mathbf{z}]}{\partial \mathbf{x}} = E_v\left[\frac{\partial E[y|\mathbf{x},\mathbf{z},v]}{\partial \mathbf{x}}\right] = (\beta + \theta\alpha)\int_{-\infty}^{\infty} \phi[\beta'\mathbf{x} + \theta(\alpha'\mathbf{z} + \sigma_u v)]\phi(v)dv$$

The integral does not have a closed form, but it can easily be simulated :

$$Est.\frac{\partial E[y|\mathbf{x},\mathbf{z}]}{\partial \mathbf{x}} = (\beta + \theta\alpha)\frac{1}{R}\sum_{r=1}^{R} \phi[\beta'\mathbf{x} + \theta(\alpha'\mathbf{z} + \sigma_u v_r)]$$

For variables only in x, omit $\theta\alpha_k$. For variables only in z, omit β_k.

Endogeneity – Partial Effects (Greene)

• Health Satisfaction example, with income as endogenous variable.

```
+--------------------------------------+
| Listed Calculator Results            |
+--------------------------------------+
SB      =       .380919
```

Matrix - BMODEL ([7, 1] Cell: 1.2176) | Matrix - ALPHA ([7, 1] Cell: -0.36099) | Matrix - SCALEDB ([7, 1] Cell: 0.299023)

	BMODEL 1	ALPHA 1	SCALEDB 1
1	1.2176	-0.36099	0.299023
2	-0.02426	0.02159	-0.00369571
3	0	-0.00025	-3.92804e-005
4	0	0.02064	0.00324299
5	-0.02599	0.07783	0.00463535
6	0.06932	-0.03564	0.0146532
7	-0.1418	0.00413	-0.0407803

The scale factor is computed using the model coefficients, means of the variables and 35,000 draws from the standard normal population.

Endogeneity – Binary Variable - Application

• **CASE 2** – h is binary. (From Greene).

Doctor's visit example. Public insurance option = endogen. variable.

```
----------------------------------------------------------------------
FIML Estimates of Bivariate Probit Model
Dependent variable               DOCPUB
Log likelihood function      -25671.43905
Estimation based on N =  27326, K =  14
--------+-------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er.  P[|Z|>z]   Mean of X
--------+-------------------------------------------------------------
        |Index    equation  for DOCTOR
Constant|     .59049***      .14473        4.080   .0000
     AGE|    -.05740***      .00601       -9.559   .0000      43.5257
   AGESQ|     .00082***    .681660D-04    12.100   .0000      2022.86
  INCOME|     .08883*        .05094        1.744   .0812       .35208
  FEMALE|     .34583***      .01629       21.225   .0000       .47877
  PUBLIC|     .43533***      .07357        5.917   .0000       .88571
        |Index    equation  for PUBLIC
Constant|    3.55054***      .07446       47.681   .0000
     AGE|     .00067         .00115         .581   .5612      43.5257
    EDUC|    -.16839***      .00416      -40.499   .0000      11.3206
  INCOME|    -.98656***      .05171      -19.077   .0000       .35208
 MARRIED|    -.00985         .02922        -.337   .7361       .75862
  HHKIDS|    -.08095***      .02510       -3.225   .0013       .40273
  FEMALE|     .12139***      .02231        5.442   .0000       .47877
        |Disturbance correlation
RHO(1,2)|    -.17280***      .04074       -4.241   .0000
--------+-------------------------------------------------------------
```

Endogeneity – Binary Variable - Application

• Doctor's visit example, with public insurance option as the endogenous binary variable. (From Greene.)

- MODEL PREDICTIONS

```
+--------------------------------------------------------+
| Bivariate Probit Predictions for DOCTOR   and PUBLIC   |
| Predicted cell (i,j) is cell with largest probability  |
| Neither DOCTOR   nor PUBLIC   predicted correctly      |
|                         1599 of   27326 observations   |
| Only    DOCTOR   correctly predicted                   |
|         DOCTOR   = 0:    1062 of   10135 observations  |
|         DOCTOR   = 1:     632 of   17191 observations  |
| Only    PUBLIC   correctly predicted                   |
|         PUBLIC   = 0:     140 of    3123 observations  |
|         PUBLIC   = 1:     632 of   24203 observations  |
| Both    DOCTOR   and PUBLIC   correctly predicted      |
|         DOCTOR   = 0 PUBLIC   = 0:      69 of    1403  |
|         DOCTOR   = 1 PUBLIC   = 0:      92 of    1720  |
|         DOCTOR   = 0 PUBLIC   = 1:     252 of    8732  |
|         DOCTOR   = 1 PUBLIC   = 1:   15008 of   15471  |
+--------------------------------------------------------+
```

Endogeneity – Binary Variable – Partial Effects

Conditional Mean

$$\mathrm{E}[y\,|\,\mathbf{x},h] = \Phi(\boldsymbol{\beta}'\mathbf{x} + \theta h)$$

$$\begin{aligned}\mathrm{E}[y\,|\,\mathbf{x},\mathbf{z}] &= E_h E[y\,|\,\mathbf{x},h] \\ &= \mathrm{Prob}(h=0\,|\,\mathbf{z})\mathrm{E}[y\,|\,\mathbf{x},h=0] + \mathrm{Prob}(h=1\,|\,\mathbf{z})\mathrm{E}[y\,|\,\mathbf{x},h=1] \\ &= \Phi(-\boldsymbol{\alpha}'\mathbf{z})\Phi(\boldsymbol{\beta}'\mathbf{x}) + \Phi(\boldsymbol{\alpha}'\mathbf{z})\Phi(\boldsymbol{\beta}'\mathbf{x}+\theta)\end{aligned}$$

Partial Effects

Direct Effects

$$\frac{\partial \mathrm{E}[y\,|\,\mathbf{x},\mathbf{z}]}{\partial \mathbf{x}} = \left[\Phi(-\boldsymbol{\alpha}'\mathbf{z})\phi(\boldsymbol{\beta}'\mathbf{x}) + \Phi(\boldsymbol{\alpha}'\mathbf{z})\phi(\boldsymbol{\beta}'\mathbf{x}+\theta)\right]\boldsymbol{\beta}$$

Indirect Effects

$$\begin{aligned}\frac{\partial \mathrm{E}[y\,|\,\mathbf{x},\mathbf{z}]}{\partial \mathbf{z}} &= \left[-\phi(-\boldsymbol{\alpha}'\mathbf{z})\Phi(\boldsymbol{\beta}'\mathbf{x}) + \phi(\boldsymbol{\alpha}'\mathbf{z})\Phi(\boldsymbol{\beta}'\mathbf{x}+\theta)\right]\boldsymbol{\alpha} \\ &= \phi(\boldsymbol{\alpha}'\mathbf{z})\left[\Phi(\boldsymbol{\beta}'\mathbf{x}+\theta) - \Phi(\boldsymbol{\beta}'\mathbf{x})\right]\boldsymbol{\alpha}\end{aligned}$$

Endogeneity – Application: Selection Model

• Sample selection problem. We only observe data if a condition is met; for example, an individual decides to invest in stocks.

• Adapted framework, to this problem:

- RUM: $U_n = \boldsymbol{x}_n'\boldsymbol{\beta} + h_n'\boldsymbol{\theta} + \varepsilon_n$
- Revealed preference: $y_n = 1$ if $U_n > 0$
- Endogenous variable(s): $h_n = \boldsymbol{z}_n'\boldsymbol{\alpha} + u_n$,

with E[ε_n | h] ≠ 0 ⇔ Cov[u_n, ε_n] ≠ 0 (ρ=Corr[u_n, ε_n])

- Sample selection: (y_n, $\boldsymbol{x}_n$) are observed only when h = 1
- Additional Assumptions:

1) $u_n, \varepsilon_n \sim \left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\sigma_u\\ \rho\sigma_u & \sigma_u^2\end{pmatrix}\right]$

2) $\boldsymbol{z}$ = IV (exogenous variables), uncorrelated with (u_n, ε_n)

Endogeneity – Application: Selection Model

• Doctor's visits, with public insurance option.

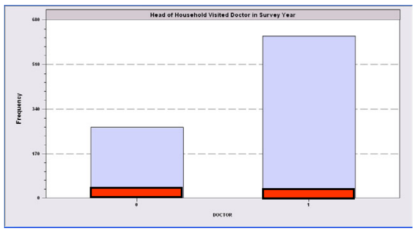
DATA: 3 Groups of observations:

(Public=0), (Doctor=0|Public=1), (Doctor=1|Public=1)

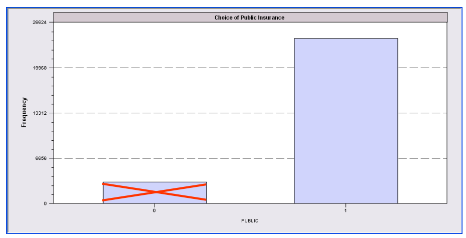
```
+-----------------------------------------------------------------+
|Cross Tabulation                                                 |
|Row variable is DOCTOR   (Out of range 0-49:      0)             |
|Number of Rows =  2      (DOCTOR   =  0 to  1)                   |
|Col variable is PUBLIC   (Out of range 0-49:      0)             |
|Number of Cols =  2      (PUBLIC   =  0 to  1)                   |
|Chi-squared independence tests:                                  |
|Chi-squared[   1] =   92.77760   Prob value =  .00000            |
|G-squared  [   1] =   90.86999   Prob value =  .00000            |
+-----------------------------------------------------------------+
|                  PUBLIC                                         |
+--------+-------------+------+                                   |
| DOCTOR|      0      1| Total|                                   |
+--------+-------------+------+                                   |
|       0|   1403   8732|  10135|                                 |
|       1|   1720  15471|  17191|                                 |
+--------+-------------+------+                                   |
|   Total|   3123  24203|  27326|                                 |
+-----------------------------------------------------------------+
```

Endogeneity – Application: Selection Model

Doctor = F(age,age^2,income,female,**Public=1**)



Public = F(age,educ,income,married,kids,female)



Endogeneity – Application: Selection Model

- Selected Sample

Joint Frequency Table for Bivariate Probit Model
Predicted cell is the one with highest probability

	PUBLIC		
DOCTOR	0	1	Total
0 Fitted	0 (0)	8732 (511)	8732 (511)
1 Fitted	0 (477)	15471 (23215)	15471 (23692)
Total Fitted	0 (477)	24203 (23726)	24203 (24203)

Counts based on 24203 selected of 27326 in sample

Endogeneity – Application: Selection Model

• ML Estimates, with Probit Model

```
---------------------------------------------------------------------
FIML Estimates of Bivariate Probit Model
Dependent variable               DOCPUB
Log likelihood function    -23581.80697
Estimation based on N =  27326, K =  13
Selection model based on PUBLIC
Means for vars. 1-  5 are after selection.
--------+------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+------------------------------------------------------------
        |Index    equation for DOCTOR
Constant|    1.09027***       .13112        8.315   .0000
     AGE|    -.06030***       .00633       -9.532   .0000       43.6996
   AGESQ|     .00086***    .718153D-04     11.967   .0000       2041.87
  INCOME|     .07820          .05779        1.353   .1760        .33976
  FEMALE|     .34357***       .01756       19.561   .0000        .49329
        |Index    equation for PUBLIC
Constant|    3.54736***       .07456       47.580   .0000
     AGE|     .00080          .00116         .690   .4899       43.5257
    EDUC|    -.16832***       .00416      -40.490   .0000       11.3206
  INCOME|    -.98747***       .05162      -19.128   .0000        .35208
 MARRIED|    -.01508          .02934        -.514   .6072        .75862
  HHKIDS|    -.07777***       .02514       -3.093   .0020        .40273
  FEMALE|     .12154***       .02231        5.447   .0000        .47877
        |Disturbance correlation
RHO(1,2)|    -.19303***       .06763       -2.854   .0043
--------+------------------------------------------------------------
```

Endogeneity – Application: Selection Model

• Partial Effects in the Selection Model

Conditional Mean : Case 1, Given Selection

$$E[y|\mathbf{x},\text{Selection}] = \text{Prob}(y=1|\mathbf{x},h=1)$$

$$= \frac{\text{Prob}(y=1,h=1|\mathbf{x},\mathbf{z})}{\text{Prob}(h=1|\mathbf{z})}$$

$$= \frac{\Phi(\beta'\mathbf{x},\alpha'\mathbf{z},\rho)}{\Phi(\alpha'\mathbf{z})}$$

Partial Effects

$$\frac{\partial E[y|\mathbf{x},\mathbf{z},\text{Selection}]}{\partial \mathbf{x}} = \frac{\left(\partial\Phi(\beta'\mathbf{x},\alpha'\mathbf{z},\rho)/\partial\beta'\mathbf{x}\right)}{\Phi(\alpha'\mathbf{z})}\beta$$

$$\frac{\partial E[y|\mathbf{x},\mathbf{z},\text{Selection}]}{\partial \mathbf{z}} = \left\{\frac{\left(\partial\Phi(\beta'\mathbf{x},\alpha'\mathbf{z},\rho)/\partial\alpha'\mathbf{z}\right)}{\Phi(\alpha'\mathbf{z})} - \frac{\phi(\alpha'\mathbf{z})\Phi(\beta'\mathbf{x},\alpha'\mathbf{z},\rho)}{[\Phi(\alpha'\mathbf{z})]^2}\right\}\alpha$$

$$\partial\Phi_2(a,b,\rho)/\partial a = \phi(a)\Phi\left(\frac{b-\rho a}{\sqrt{1-\rho^2}}\right)$$

For variables that appear in both $\mathbf{x}$ and $\mathbf{z}$, the effects are added.