Microeconometrics

Limited Dependent Variable Models

by Vanessa Berenguer-Rico

University of Oxford

Michaelmas Term 2016

Limited Dependent Variable Models

"A limited dependent variable is broadly defined as a dependent variable whose range of values is substantively restricted," Wooldrige p. 583

Limited Dependent Variable Models

Examples

- Binary Response: y takes only two values: 1 or 0, which indicate whether or not a certain even has occurred
- Multinomial Response: y takes on values $0, 1, 2, ..., J$ for J a positive integer, which denote responses or choices from multiple alternatives without an intrinsic ordering (no natural numerical vales)
- Ordered Response: y takes on values $0, 1, 2, ..., J$ for J a positive integer, which denote responses or choices from multiple alternatives with an intrinsic ordering (no natural numerical vales)
- Count Data: y takes on values $0, 1, 2, ..$ which denote a count of the number of occurrences (natural numerical vales)

Limited Dependent Variable Models

Examples: Which is which?

- A mortgage application is accepted or denied
- Mode of transport chosen by a commuter: subway, bus, drive, walk/bike
- Mode of transport: public or private
- Obtaining a high school degree, some college education (but not graduating), graduating from college, MPhil, DPhil
- Number of restaurant meals eaten by a consumer in a week
- Decision on whether to participate in the labor market or not
- Number of crimes committed

Limited Dependent Variable Models

Examples: Which is which?

- Survey questions about strength of feelings about a particular commodity, such as a movie
- Number of visits to a recreation site
- Number of defects per unit of time in a production process
- Variable indicating whether a student's grade in an intermediate macro course was higher than that in the principles course
- Scale of occupations: unskilled employees, machine operators, skilled manual employees, clerical and sales workers and technicians, etc..
- Choice of automobile models from a varied menu of features

Limited Dependent Variable Models

Examples

- Binary Response: y takes only two values: 1 or 0, which indicate whether or not a certain even has occurred
- Multinomial Response: y takes on values $0, 1, 2, ..., J$ for J a positive integer, which denote responses or choices from multiple alternatives without an intrinsic ordering (no natural numerical vales)
- Ordered Response: y takes on values $0, 1, 2, ..., J$ for J a positive integer, which denote responses or choices from multiple alternatives with an intrinsic ordering (no natural numerical vales)
- Count Data: y takes on values $0, 1, 2, ..$ which denote a count of the number of occurrences (natural numerical vales)

Limited Dependent Variable Models

"In these and any number of other cases, the dependent variable is not a quantitative measure of some economic outcome, but rather an indicator of whether or not some outcome occurred. It follows that the regression methods we have used up to this point are largely inappropriate. We turn, instead to modeling probabilities and using econometric tools to make probabilistic statements about the occurrence of these events. We will also examine models for counts of occurrences," Greene p. 721

Binary Response Models

Binary Response Models

- Dependent variable

$$y_i = \begin{cases} 1 & \text{if event } A \text{ is observed for } i \\ 0 & \text{otherwise} \end{cases}$$

- Observable characteristics: x_i

Cross-sectional Data: Labor Force Participation

- **Labor Force Participation**: Data: Wooldrige (p. 239)
  - $inlf_i$: 1 if woman i reports working for a wage outside the home, 0 otherwise
  - $nwifeinc_i$: husband's earnings
  - $educ_i$: years of education
  - $\exp er_i$: past years of labor market experience
  - $kidslt6_i$: number of children less than six years old
  - $kidsge6_i$: number of kids between 6 and 18 years of age

Binary Response Models

- Dependent variable: $y_i = 1(A)$
- Explanatory variables: x_i
- Objective: Modelling the probability that $y_i = 1$
- We specify

$$
\begin{aligned}
P(y = 1|x) &= G(x, \beta) \\
P(y = 0|x) &= 1 - G(x, \beta)
\end{aligned}
$$

- β: measures the impact of changes in x on the probability
- Notice

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

Binary Response Models

- Linear Probability Model: $G(x, \beta)$ linear
- Probit: $G(x, \beta)$ Normal
- Logit: $G(x, \beta)$ Logistic

Linear Probability Model

Binary Response Models

Linear Probability Model

- Objective: Modelling the probability that $y_i = 1$:

$$\begin{aligned} P\left(y = 1|x\right) &= G\left(x, \beta\right) \\ P\left(y = 0|x\right) &= 1 - G\left(x, \beta\right) \end{aligned}$$

- The linear probability model specifies:

$$P\left(y = 1|x\right) = G\left(x, \beta\right) = x'\beta$$

Binary Response Models

Linear Probability Model

- Objective: Modelling the response probability (that $y = 1$):

$$P(y = 1|x) = G(x, \beta)$$

- The linear probability model considers

$$P(y = 1|x) = G(x, \beta) = x'\beta = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k$$

and specifies

$$y_i = x_i'\beta + u_i$$

Cross-sectional Data: Labor Force Participation

- **Labor Force Participation**: Data: Wooldrige (p. 239)
  - $inlf_i$: 1 if woman i reports working for a wage outside the home, 0 otherwise
  - $nwifeinc_i$: husband's earnings
  - $educ_i$: years of education
  - $\exp er_i$: past years of labor market experience
  - $kidslt6_i$: number of children less than six years old
  - $kidsge6_i$: number of kids between 6 and 18 years of age

Linear Probability Model: Labor Force Participation
Wooldrige, p. 250

$$\begin{aligned}\widehat{inlf} &= \underset{(0.154)}{0.586} - \underset{(0.0014)}{0.0034} nwifeinc + \underset{(0.007)}{0.038} educ + \underset{(0.006)}{0.039} exper \\ &\quad - \underset{(0.00018)}{0.00060} exper^2 - \underset{(0.002)}{0.016} age - \underset{(0.034)}{0.262} kidslt6 + \underset{(0.013)}{0.013} kidsge6\end{aligned}$$

- Coefficients interpretation:
  - *educ*: (ceteris paribus) another year of education increases the probability of labor force participation by 0.038

Binary Response Models

Linear Probability Model: Labor Force Participation
Wooldrige, p. 250

$$
\begin{aligned}
\widehat{inlf} \quad = \quad & \underset{(0.154)}{0.586} - \underset{(0.0014)}{0.0034}nwifeinc + \underset{(0.007)}{0.038}educ + \underset{(0.006)}{0.039}exper \\
& -\underset{(0.00018)}{0.00060}exper^2 - \underset{(0.002)}{0.016}age - \underset{(0.034)}{0.262}kidslt6 + \underset{(0.013)}{0.013}kidsge6
\end{aligned}
$$

- Coefficients interpretation:
  - *nwifeinc*: (ceteris paribus) if $\Delta nwifeinc = 10$ (i.e. an increase of \$10000), then the probability that a woman is in the labor market falls by 0.034

Binary Response Models

Linear Probability Model: Labor Force Participation
Wooldrige, p. 250

$$
\begin{aligned}
\widehat{inlf} &= \underset{(0.154)}{0.586} - \underset{(0.0014)}{0.0034} nwifeinc + \underset{(0.007)}{0.038} educ + \underset{(0.006)}{0.039} exper \\
&\quad - \underset{(0.00018)}{0.00060} exper^2 - \underset{(0.002)}{0.016} age - \underset{(0.034)}{0.262} kidslt6 + \underset{(0.013)}{0.013} kidsge6
\end{aligned}
$$

- Coefficients interpretation:
  - *exper*: Quadratic: past experience has a diminishing effect on the labor force participation. In particular, (ceteris paribus) the estimated change in the probability is approximated as $0.039 - 2\,(0.0006) = 0.039 - 0.0012 exper$

Linear Probability Model: Labor Force Participation
Wooldrige, p. 250

$$\begin{aligned}\widehat{inlf} &= \underset{(0.154)}{0.586} - \underset{(0.0014)}{0.0034} nwifeinc + \underset{(0.007)}{0.038} educ + \underset{(0.006)}{0.039} exper \\ &\quad - \underset{(0.00018)}{0.00060} exper^2 - \underset{(0.002)}{0.016} age - \underset{(0.034)}{0.262} kidslt6 + \underset{(0.013)}{0.013} kidsge6\end{aligned}$$

- Coefficients interpretation:
  - *kidslt6* and *kidsge6*: very different effect of the number of younger and older children. (ceteris paribus) Having one additional child less than six years old reduces the probability of participation by 0.262.

Linear Probability Model

- The linear probability model considers

$$P\left(y=1|x\right)=G\left(x,\beta\right)=x'\beta=\beta_0+\beta_1x_1+...+\beta_kx_k$$

and specifies

$$y_i=x_i'\beta+u_i$$

- Easy to estimate and interpret

Linear Probability Model: Shortcomings

- Heteroscedasticity (by construction):
  - Notice that $x_i'\beta + u_i$ must equal zero or one
  - Therefore, u_i equals either $-x_i'\beta$ or $1 - x_i'\beta$ with probabilities $1 - G$ and G, respectively. (Not normally distributed in finite samples)
  - Hence,
$$Var\left[u|x\right] = x'\beta\left(1 - x'\beta\right)$$
  - Gauss-Markov does not apply
  - We could use HAC standard errors

Binary Response Models

Linear Probability Model: Labor Force Participation
HAC standard errors
Wooldrige 2, p. 250

$$
\begin{aligned}
\widehat{inlf} \quad = \quad & \underset{(0.151)}{0.586} - \underset{(0.0015)}{0.0034} nwifeinc + \underset{(0.007)}{0.038} educ + \underset{(0.006)}{0.039} exper \\
& - \underset{(0.00019)}{0.00060} exper^2 - \underset{(0.002)}{0.016} age - \underset{(0.032)}{0.262} kidslt6 + \underset{(0.013)}{0.013} kidsge6
\end{aligned}
$$

- Similar standard errors!

Binary Response Models

Linear Probability Model: Shortcomings
Wooldrige , p. 251

- Predicted probabilities may not belong to $[0, 1]$:

$$\hat{y}_i = x'\hat{\beta} \notin [0, 1]$$

Binary Response Models

Linear Probability Model: Shortcomings
Wooldrige , p. 251

$$\begin{aligned}\widehat{inlf} &= \underset{(0.151)}{0.586} - \underset{(0.0015)}{0.0034} nwifeinc + \underset{(0.007)}{0.038} educ + \underset{(0.006)}{0.039} exper \\ &\quad - \underset{(0.00019)}{0.00060} exper^2 - \underset{(0.002)}{0.016} age - \underset{(0.032)}{0.262} kidslt6 + \underset{(0.013)}{0.013} kidsge6\end{aligned}$$

- A probability cannot be linearly related to the independent variables for all their positive values
- Labor force participation example: from 0 to 1 young child reduces the probability of working by 0.262, the same reduction as going from 1 to 2 children
- Extreme: going from 0 to four young children reduces the probability of working by $\Delta\widehat{inlf} = 0.262\,(\Delta kidslt6) = 0.262 \times 4 = 1.048$

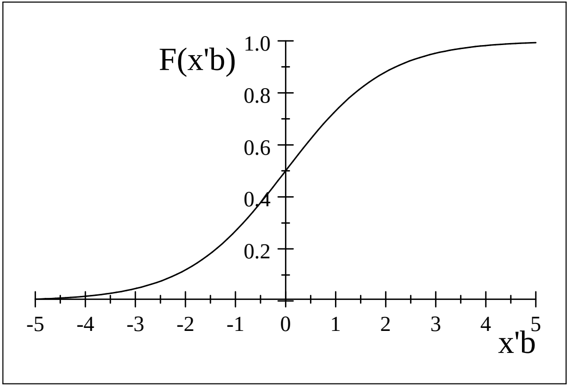
Binary Response Models

- How can these shortcomings be overcome?
- We need

$$\begin{aligned}\lim_{x'\beta \to +\infty} P(y = 1|x) &= 1 \\ \lim_{x'\beta \to -\infty} P(y = 1|x) &= 0\end{aligned}$$

- Cumulative Distribution Functions

Binary Response Models



Probit and Logit

Binary Response Models

- Normal distribution: **Probit**

$$P\left(y=1|x\right)=\int_{-\infty}^{x'\beta}\phi\left(t\right)dt=\Phi\left(x'\beta\right)$$

where $\Phi\left(x'\beta\right)$ denotes the standard normal distribution function and ϕ the standard normal density

- Logistic distribution: **Logit**

$$P\left(y=1|x\right)=\frac{\exp\left(x'\beta\right)}{1+\exp\left(x'\beta\right)}=\Lambda\left(x'\beta\right)$$

Binary Response Models

- In general,

$$P\left(y=1|x\right)=E\left[y|x\right]=G\left(x'\beta\right)=G\left(\beta_0+\beta_1x_1+...+\beta_kx_k\right)$$

- G is **nonlinear** in Probit and Logit models!!!

Marginal Effects

Binary Response Models

Marginal Effects: Binary explanatory variable

- Model

$$P(y=1|x) = E[y|x] = G(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki})$$

  where x_{1i} is a binary explanatory variable (dummy)
- Partial effect from changing x_{1i} from zero to one is

$$G(\beta_0 + \beta_1 + \beta_2 x_{2i} + ... + \beta_k x_{ki}) - G(\beta_0 + \beta_2 x_{2i} + ... + \beta_k x_{ki})$$

- This depends on all the values of the other x_j. We can evaluate the partial effect at some particular value of the x_j's

Marginal Effects: Continuous explanatory variable

$$P\left(y=1|x\right)=G\left(x'\beta\right)=G\left(\beta_0+\beta_1 x_1+...+\beta_k x_k\right)$$

- **Marginal effects**
$$\frac{\partial E\left[y|x\right]}{\partial x_j}$$
- What are they in the probit and logit models?

Probit: Normal Distribution

- **Recall**: for $X \sim N(0,1)$
- Density function

$$g(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

- Cumulative distribution function

$$G(x) = \Phi(x) = \int_{-\infty}^{x} \phi(x)\, dx$$

Binary Response Models

Cumulative distribution (cdf) vs Density (pdf)

$$G'(x) = \frac{dG(x)}{dx} = g(x)$$

Binary Response Models

Probit Marginal Effects

$$P\left(y=1|x\right)=\int_{-\infty}^{x'\beta}\phi\left(t\right)dt=\Phi\left(x'\beta\right)$$

- Marginal effect:

$$\begin{aligned}\frac{\partial E\left[y|x\right]}{\partial x_j} &= \frac{d\Phi\left(x'\beta\right)}{d\left(x'\beta\right)}\frac{\partial\left(x'\beta\right)}{\partial x_j}\\ &= \frac{d\Phi\left(x'\beta\right)}{d\left(x'\beta\right)}\beta_j\\ &= \phi\left(x'\beta\right)\beta_j\end{aligned}$$

where $\phi\left(x'\beta\right)$ is known as the scale factor

Probit Marginal Effects

- How do we evaluate the marginal effect:

$$\frac{\partial E\left[y|x\right]}{\partial x_j} = \phi\left(x'\beta\right)\beta_j$$

- We could evaluate marginal effects for mean values of the x:

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

Binary Response Models

Logit Marginal Effects

$$P\left(y=1|x\right)=\frac{\exp\left(x'\beta\right)}{1+\exp\left(x'\beta\right)}=\Lambda\left(x'\beta\right)$$

- Marginal effect:

$$\begin{aligned}
\frac{\partial E\left[y|x\right]}{\partial x_j} &= \frac{d\Lambda\left(x'\beta\right)}{d\left(x'\beta\right)}\frac{\partial\left(x'\beta\right)}{\partial x_j} \\
&= \frac{d\Lambda\left(x'\beta\right)}{d\left(x'\beta\right)}\beta_j \\
&= \frac{\exp\left(x'\beta\right)\left(1+\exp\left(x'\beta\right)\right)-\exp\left(x'\beta\right)\exp\left(x'\beta\right)}{\left(1+\exp\left(x'\beta\right)\right)^2}\beta_j \\
&= \Lambda\left(x'\beta\right)\left(1-\Lambda\left(x'\beta\right)\right)\beta_j
\end{aligned}$$

where $\Lambda\left(x'\beta\right)\left(1-\Lambda\left(x'\beta\right)\right)$ is known as the scale factor

Logit Marginal Effects

- How do we evaluate the marginal effect:

$$\frac{\partial E\left[y|x\right]}{\partial x_j} = \Lambda\left(x'\beta\right)\left(1 - \Lambda\left(x'\beta\right)\right)\beta_j$$

- We could evaluate marginal effects for mean values of the x:

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

Marginal Effects Summary

- **Probit**: $P\left(y=1|x\right)=\Phi\left(x'\beta\right)$

$$\frac{\partial E\left[y|x\right]}{\partial x_j}=\phi\left(x'\beta\right)\beta_j$$

- **Logit**: $P\left(y=1|x\right)=\Lambda\left(x'\beta\right)$

$$\frac{\partial E\left[y|x\right]}{\partial x_j}=\Lambda\left(x'\beta\right)\left(1-\Lambda\left(x'\beta\right)\right)\beta_j$$

Probit Marginal Effects: Example

- Estimated model

$$\hat{y}_i = \Phi\left(\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}\right)$$

- Mean values

$$\bar{x}'\hat{\beta} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2$$

- Estimated marginal effect at $\bar{x}$

$$\left.\frac{\widehat{\partial E\left[y|x\right]}}{\partial x_2}\right|_{\bar{x}} = \phi\left(\hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2\right)\hat{\beta}_2$$

Probit Marginal Effects: Example

- Estimated model

$$\hat{y}_i = \Lambda\left(\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}\right)$$

- Mean values

$$\bar{x}'\hat{\beta} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2$$

- Estimated marginal effect at $\bar{x}$

$$\left.\frac{\widehat{\partial E\left[y|x\right]}}{\partial x_2}\right|_{\bar{x}} = \Lambda\left(\hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2\right)\left(1 - \Lambda\left(\hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2\right)\right)\hat{\beta}_2$$

Estimation

Estimation

- Nonlinear Least Squares: not efficient
- ML: *"For estimating limited dependent variable models, maximum likelihood methods are indispensable. Because maximum likelihood estimation is based on the distribution of y given x, the heteroskedasticity in $Var(y|x)$ is automatically accounted for"* Wooldrige, p. 587

Binary Response Models

Estimation: Maximum Likelihood

- Let (y_i, x_i) be a random sample of size n
- Let the density of y_i given x_i be

$$P\left(y_i|x_i,\beta\right) = \left[G\left(x_i'\beta\right)\right]^{y_i}\left[1-G\left(x_i'\beta\right)\right]^{1-y_i} \quad \text{where } y_i = 0, 1$$

- The joint density of $y_1, ..., y_n$ given x_i is then

$$P\left(y_1, ..., y_n|x_i,\beta\right) = \prod_{i=1}^{n}\left[G\left(x_i'\beta\right)\right]^{y_i}\left[1-G\left(x_i'\beta\right)\right]^{1-y_i}$$

Estimation: Maximum Likelihood

- The joint density of $y_1, ..., y_n$ given x_i is then

$$P\left(y_1, ..., y_n | x_i, \beta\right) = \prod_{i=1}^{n} \left[G\left(x_i'\beta\right)\right]^{y_i} \left[1 - G\left(x_i'\beta\right)\right]^{1-y_i}$$

- The log-likelihood is then

$$\begin{aligned} \mathcal{L}_n\left(\beta\right) &= \ln\left(P\left(y_1, ..., y_n | x_i, \beta\right)\right) \\ &= \ln\left(\prod_{i=1}^{n} \left[G\left(x_i'\beta\right)\right]^{y_i} \left[1 - G\left(x_i'\beta\right)\right]^{1-y_i}\right) \\ &= \sum_{i=1}^{n} y_i \ln\left[G\left(x_i'\beta\right)\right] + \sum_{i=1}^{n} \left(1 - y_i\right) \ln\left[1 - G\left(x_i'\beta\right)\right] \end{aligned}$$

Estimation: Maximum Likelihood

- The log-likelihood is then

$$\mathcal{L}_n(\beta) = \sum_{i=1}^{n} y_i \ln\left[G\left(x_i'\beta\right)\right] + \sum_{i=1}^{n} (1-y_i) \ln\left[1 - G\left(x_i'\beta\right)\right]$$

- Objective: Maximize the log-likelihood function. FOC:

$$\begin{aligned}\frac{\partial \mathcal{L}_n(\beta)}{\partial \beta} &= \sum_{i=1}^{n} y_i \frac{g\left(x_i'\beta\right) x_i}{G\left(x_i'\beta\right)} - \sum_{i=1}^{n} (1-y_i) \frac{g\left(x_i'\beta\right) x_i}{1 - G\left(x_i'\beta\right)} \\ &= \sum_{i=1}^{n} \left[y_i \frac{g\left(x_i'\beta\right)}{G\left(x_i'\beta\right)} - (1-y_i) \frac{g\left(x_i'\beta\right)}{1 - G\left(x_i'\beta\right)} \right] x_i = 0\end{aligned}$$

Binary Response Models

Logit: Maximum Likelihood

- Logit

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

- Objective: Maximize the log-likelihood function. FOC:

$$\frac{\partial \mathcal{L}_n\left(\beta\right)}{\partial \beta} = \sum_{i=1}^{n}\left[y_i \frac{g\left(x_i'\beta\right)}{G\left(x_i'\beta\right)} - \left(1-y_i\right)\frac{g\left(x_i'\beta\right)}{1-G\left(x_i'\beta\right)}\right] x_i = 0$$

- Logit model:

$$\frac{\partial \mathcal{L}_n\left(\beta\right)}{\partial \beta} = \sum_{i=1}^{n}\left[y_i \frac{\frac{e^{x_i'\beta}}{\left(1+e^{x_i'\beta}\right)^2}}{\frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}} - \left(1-y_i\right)\frac{\frac{e^{x_i'\beta}}{\left(1+e^{x_i'\beta}\right)^2}}{1-\frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}}\right] x_i = 0$$

Binary Response Models

Logit: Maximum Likelihood

- Logit

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

- Logit model:

$$\frac{\partial \mathcal{L}_n(\beta)}{\partial \beta} = \sum_{i=1}^{n}\left[y_i \frac{\frac{e^{x_i'\beta}}{\left(1+e^{x_i'\beta}\right)^2}}{\frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}} - (1-y_i)\frac{\frac{e^{x_i'\beta}}{\left(1+e^{x_i'\beta}\right)^2}}{1-\frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}}\right] x_i = 0$$

- Simplifying

$$\frac{\partial \mathcal{L}_n(\beta)}{\partial \beta} = \sum_{i=1}^{n}\left[y_i - \Lambda\left(x_i'\beta\right)\right] x_i = 0$$

Binary Response Models

Probit: Maximum Likelihood

- Probit

$$G\left(x_i'\beta\right) = \Phi\left(x_i'\beta\right)$$

- FOC

$$\frac{\partial \mathcal{L}_n\left(\beta\right)}{\partial \beta} = \sum_{i=1}^{n}\left[y_i \frac{g\left(x_i'\beta\right)}{G\left(x_i'\beta\right)} - \left(1 - y_i\right)\frac{g\left(x_i'\beta\right)}{1 - G\left(x_i'\beta\right)}\right] x_i = 0$$

- Probit model:

$$\frac{\partial \mathcal{L}_n\left(\beta\right)}{\partial \beta} = \sum_{i=1}^{n}\left[y_i \frac{\phi\left(x_i'\beta\right)}{\Phi\left(x_i'\beta\right)} - \left(1 - y_i\right)\frac{\phi\left(x_i'\beta\right)}{1 - \Phi\left(x_i'\beta\right)}\right] x_i = 0$$

Logit: Maximum Likelihood

- Probit

$$G\left(x_i'\beta\right) = \Phi\left(x_i'\beta\right)$$

- Probit model. FOC:

$$\frac{\partial \mathcal{L}_n\left(\beta\right)}{\partial \beta} = \sum_{i=1}^{n}\left[y_i \frac{\phi\left(x_i'\beta\right)}{\Phi\left(x_i'\beta\right)} - \left(1-y_i\right)\frac{\phi\left(x_i'\beta\right)}{1-\Phi\left(x_i'\beta\right)}\right] x_i = 0$$

- Simplifying

$$\frac{\partial \mathcal{L}_n\left(\beta\right)}{\partial \beta} = \sum_{i=1}^{n} w_i\left[y_i - \Phi\left(x_i'\beta\right)\right] x_i = 0$$

where $w_i = \phi\left(x_i'\beta\right) / \left[\Phi\left(x_i'\beta\right)\left(1-\Phi\left(x_i'\beta\right)\right)\right]$

Probit or Logit in practice?
Cameron and Trivedi, p. 472

- Empirically, either logit and probit can be used
- Often little difference between predicted probabilities
- The difference is greater in the tail (probabilities close to 0 or 1)
- Less difference if interest is averaged marginal effects

Binary Response Models

Probit or Logit in practice?
Cameron and Trivedi, p. 472

- Natural metric to compare models: Fitted log-likelihood

$$\mathcal{L}_n\left(\hat{\beta}\right) = \sum_{i=1}^{n} y_i \ln\left[G\left(x_i'\hat{\beta}\right)\right] + \sum_{i=1}^{n}\left(1-y_i\right)\ln\left[1-G\left(x_i'\hat{\beta}\right)\right]$$

- Often log-likelihoods are similar

Binary Response Models

Probit or Logit in practice?

Cameron and Trivedi, p. 473

- Different models yield different $\hat{\beta}$ (artifact of using different models). What needs to be compared are marginal effects across models
- Rule of thumb:

$$
\begin{aligned}
\hat{\beta}_{Logit} &\simeq 4\hat{\beta}_{OLS} \\
\hat{\beta}_{Probit} &\simeq 2.5\hat{\beta}_{OLS} \\
\hat{\beta}_{Logit} &\simeq 1.6\hat{\beta}_{Probit}
\end{aligned}
$$

- Amemiya (1981, p. 1488) shows that this rule of thumb works well if $0.1 \leq p \leq 0.9$
- Greater departures across models across models occur in the tails

Determining Model Adequacy
Cameron and Trivedi, p. 473

- Pseudo-R^2
- Predicted Probabilities

Binary Response Models

Determining Model Adequacy
Cameron and Trivedi, p. 473

- **Pseudo-R^2**: (proposed by McFadden, 1974)

$$
\begin{aligned}
R^2_{Binary} &= 1 - \frac{\mathcal{L}_n\left(\hat{\beta}\right)}{\mathcal{L}_n\left(\bar{y}\right)} \\
&= 1 - \frac{\sum_{i=1}^{n}\left(y_i \ln \hat{p}_i + (1 - y_i)\ln\left(1 - \hat{p}_i\right)\right)}{n\left[\bar{y}\ln\bar{y} + (1 - \bar{y})\ln(1 - \bar{y})\right]}
\end{aligned}
$$

where $\hat{p}_i = G\left(x_i'\hat{\beta}\right)$ and $\bar{y} = n^{-1}\sum_{i=1}^{n} y_i$

- Compares the likelihood function with all regressors to the likelihood with none
- Interpretation is as with the usual R^2

Determining Model Adequacy
Cameron and Trivedi, p. 473

- **Predicted Probabilities**: Based on the correctly predicted percentage
- Consider the following rule: Predict $Y = 1$ if model estimates that $P(y = 1) > 0.5$; predict $Y = 0$ otherwise
- Construct the following table

Frequencies	*predicted* $y = 0$	*predicted* $y = 1$
Observed $y = 0$	✓	$\times$
Observed $y = 1$	$\times$	✓

Testing

Asymptotics
Wooldridge, p. 588

- The general theory of MLE for random samples applies
- Under very general conditions, the MLE is consistent, asymptotically normal, and asymptotically efficient
- Hence, each $\hat{\beta}$ comes with an (asymptotic) standard error

$$\widehat{Avar}\left(\hat{\beta}\right) = \left(\sum_{i=1}^{n} \frac{\left[g\left(x_i'\hat{\beta}\right)\right]^2 x_i x_i'}{G\left(x_i'\hat{\beta}\right)\left[1 - G\left(x_i'\hat{\beta}\right)\right]} \right)^{-1}$$

which is a $k \times k$ matrix (see Wooldrige, p. 631)

Significance Test
Wooldridge, p. 588

- As in previous topics, once we have the standard errors, we can construct (asymptotic) t tests (as with OLS)
- Hypothesis:

$$\begin{aligned} H_o &: \quad \beta_j = 0 \\ H_a &: \quad \beta_j \neq 0 \end{aligned}$$

- Test Statistic:

$$t = \frac{\hat{\beta}_j}{s.e.\left(\hat{\beta}_j\right)}$$

- Decision Rule: Reject the null if $|t| > 1.96$

Testing General Hypothesis
Greene, p. 564

- The Trinity:
  - Wald Test
  - Likelihood Ratio Test
  - Lagrange Multiplier (Scores) Test

Testing General Hypothesis

- **Wald Test**: If the restriction is valid, then $c\left(\hat{\beta}\right)$ should be close to q. (Based on estimates of the unrestricted model)
- Hypothesis

$$\begin{aligned} H_o &: c\left(\beta\right) = q \\ H_a &: c\left(\beta\right) \neq q \end{aligned}$$

- Test Statistic

$$W = \left(c\left(\hat{\beta}\right) - q\right)' \left[\left(\widehat{AsyVar}\left(c\left(\hat{\beta}\right) - q\right)\right)\right]^{-1} \left(c\left(\hat{\beta}\right) - q\right) \sim \chi^2_q$$

- Decision Rule: Reject the null if $W > \chi^2_{\alpha/2;q}$

Testing General Hypothesis

- **Likelihood Ratio Test**: If the restriction is valid, then imposing it shoud not lead to a large reduction in the log-likelihood. (Based on estimates of both the unrestricted and restricted models)
- Hypothesis

$$
\begin{aligned}
H_o &: c(\beta) = q \\
H_a &: c(\beta) \neq q
\end{aligned}
$$

- Test Statistic

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

- Decision Rule: Reject the null if $LR > \chi^2_{\alpha/2;q}$

Testing General Hypothesis

- **Lagrange Test**: If the restriction is valid, then the restricted estimator should be near the point that maximizes the log-likelihood. (Based on estimates of the restricted models)
- Hypothesis

$$\begin{aligned} H_o &: c(\beta) = q \\ H_a &: c(\beta) \neq q \end{aligned}$$

- Test Statistic

$$LM = \left(\frac{\partial \ln L_n \left(\hat{\beta}_R \right)}{\partial \hat{\beta}_R} \right)' \left[I \left(\hat{\beta}_R \right) \right]^{-1} \left(\frac{\partial \ln L_n \left(\hat{\beta}_R \right)}{\partial \hat{\beta}_R} \right) \sim \chi^2_q$$

  where $I\left(\hat{\beta}_R\right)$ is the information matrix, that is, minus the expected Hessian matrix (second derivatives)
- Decision Rule: Reject the null if $LM > \chi^2_{\alpha/2;q}$

Example:

Labor Force Participation

Linear Probability Model: Labor Force Participation

Wooldrige, p. 250

$$\begin{aligned}
\widehat{inlf} \quad = \quad & \underset{(0.154)}{0.586} - \underset{(0.0014)}{0.0034} nwifeinc + \underset{(0.007)}{0.038} educ + \underset{(0.006)}{0.039} exper \\
& - \underset{(0.00018)}{0.00060} exper^2 - \underset{(0.002)}{0.016} age - \underset{(0.034)}{0.262} kidslt6 + \underset{(0.013)}{0.013} kidsge6 \\
& \textit{percentage correctly predicted} \quad 73.4 \\
& \log - \textit{likelihood} \quad - \\
& \textit{Pseudo } R^2 \quad 0.264
\end{aligned}$$

Binary Response Models

Logit (MLE): Labor Force Participation
Wooldrige, p. 594

$$\begin{aligned}
\widehat{inlf} \quad = \quad & \Lambda(\underset{(0.860)}{0.425} - \underset{(0.008)}{0.021} nwifeinc + \underset{(0.043)}{0.221} educ + \underset{(0.032)}{0.206} exper \\
& - \underset{(0.0010)}{0.0032} exper^2 - \underset{(0.015)}{0.088} age - \underset{(0.204)}{1.443} kidslt6 + \underset{(0.075)}{0.060} kidsge6) \\
& \textit{percentage correctly predicted} \quad 73.6 \\
& \log - \textit{likelihood} \quad -401.77 \\
& \textit{Pseudo } R^2 \quad 0.220
\end{aligned}$$

Binary Response Models

Probit (MLE): Labor Force Participation

Wooldrige, p. 594

$$
\begin{aligned}
\widehat{inlf} \quad = \quad & \Phi(\underset{(0.509)}{0.270} - \underset{(0.005)}{0.012}nwifeinc + \underset{(0.025)}{0.131}educ + \underset{(0.019)}{0.123}exper \\
& -\underset{(0.0006)}{0.0019}exper^2 - \underset{(0.008)}{0.053}age - \underset{(0.119)}{0.868}kidslt6 + \underset{(0.043)}{0.036}kidsge6) \\
& \textit{percentage correctly predicted} \quad 73.4 \\
& \log - likelihood \quad -401.30 \\
& Pseudo\ R^2 \quad 0.221
\end{aligned}
$$

Binary Response Models

Labor Force Participation
Wooldrige, p. 594

- Consistent story from the 3 models
- Sign of coefficients the same across models
- Same variables are statistically significant in each model
- The pseudo R^2 for the LPM is the usual R^2. For logit and probit, it is the measure based on the log-likelihoods

Binary Response Models

Labor Force Participation
Wooldrige, p. 594

- Magnitudes of $\hat{\beta}_j$ across models not directly comparable. Instead, we compare marginal effects (or scale factors)
- Recall:
  - Probit Marginal effect:

$$\frac{\partial E\left[y|x\right]}{\partial x_j} = \phi\left(x'\beta\right)\beta_j$$

    where $\phi\left(x'\beta\right)$ is known as the scale factor
  - Logit Marginal effect:

$$\frac{\partial E\left[y|x\right]}{\partial x_j} = \Lambda\left(x'\beta\right)\left(1 - \Lambda\left(x'\beta\right)\right)\beta_j$$

    where $\Lambda\left(x'\beta\right)\left(1 - \Lambda\left(x'\beta\right)\right)$ is known as the scale factor

Binary Response Models

Labor Force Participation
Wooldrige, p. 594

- Estimated Marginal Effects:
  - Probit Marginal effect:

$$\frac{\partial E\left[y|x\right]}{\partial x_j} = \phi\left(x'\hat{\beta}\right)\hat{\beta}_j$$

    where $\phi\left(x'\hat{\beta}\right)$ is known as the scale factor
  - Logit Marginal effect:

$$\frac{\partial E\left[y|x\right]}{\partial x_j} = \Lambda\left(x'\hat{\beta}\right)\left(1-\Lambda\left(x'\hat{\beta}\right)\right)\hat{\beta}_j$$

    where $\Lambda\left(x'\hat{\beta}\right)\left(1-\Lambda\left(x'\hat{\beta}\right)\right)$ is known as the scale factor

Labor Force Participation
Wooldrige, p. 594

- Probit Scale Factor evaluated at the mean values:

$$\phi\left(\bar{x}'\hat{\beta}\right) = 0.391$$

- Logit Scale Factor evaluated at the mean values:

$$\Lambda\left(\bar{x}'\hat{\beta}\right)\left(1 - \Lambda\left(\bar{x}'\hat{\beta}\right)\right) = 0.243$$

- Comparing:

$$0.391/0.243 \approx 1.61$$

which is close to the rule of thumb

Labor Force Participation
Wooldrige, p. 594

- Estimated Marginal Effects of Educ:
  - Probit Marginal effect:

$$\frac{\partial E\left[y|x\right]}{\partial Educ_i} = \phi\left(\bar{x}'\hat{\beta}\right)\hat{\beta}_{Educ} = 0.391 \times 0.221 = 0.086$$

  - Logit Marginal effect:

$$\frac{\partial E\left[y|x\right]}{\partial Educ_i} = \Lambda\left(\bar{x}'\hat{\beta}\right)\left(1 - \Lambda\left(\bar{x}'\hat{\beta}\right)\right)\hat{\beta}_{Educ} = 0.243 \times 0.221 = 0.053$$

The Poisson Regression Model

Count Data

- Count Data: y takes on values $0, 1, 2, ..$ which denote a count of the number of occurrences (natural numerical vales)
- Examples:
  - Number of children ever born to a woman
  - Number of times someone is arrested
  - Number of patents applied for by a firm in a year
  - Number of visits to a recreation site
  - Number of defects per unit of time in a production process
  - Number of people in a community who survive to age 100
  - Number of customers entering a store on a given day
  - etc... etc...

The Poisson Regression Model

- Count Data: y takes on values $0, 1, 2, ..$ which denote a count of the number of occurrences (natural numerical vales)
- Examples:
  - Number of children ever born to a woman
  - Number of times someone is arrested
  - Number of patents applied for by a firm in a year
  - Number of visits to a recreation site
  - Number of defects per unit of time in a production process
  - Number of people in a community who survive to age 100
  - Number of customers entering a store on a given day
  - etc... etc...

The Poisson Regression Model
Wooldrige, p. 604

- As with binary response models, a linear model for $E\left[y|x_1, ..., x_k\right]$ might not provide the best fit over all values of the x's
- Poisson Regression Model

$$E\left[y|x_1, ..., x_k\right] = \exp\left(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k\right)$$

- Because $\exp\left(.\right)$ is always positive, the Poisson regression model ensures that predicted values for y will also be positive

The Poisson Regression Model: Marginal Effects
Wooldrige2, p. 726

- Poisson Regression Model

$$E\left[y|x_1,...,x_k\right] = \exp\left(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k\right)$$

- Marginal effect:

$$\frac{\partial E\left[y|x\right]}{\partial x_j} = \exp\left(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k\right)\beta_j$$

- For a dummy regressor, say x_1:

$$\begin{aligned} & E\left[y|x_1 = 1,...,x_k\right] - E\left[y|x_1 = 0,...,x_k\right] \\ = \quad & \exp\left(\beta_0 + \beta_1 + ... + \beta_k x_k\right) - \exp\left(\beta_0 + ... + \beta_k x_k\right) \end{aligned}$$

- Evaluate the marginal effects for some representative individual

The Poisson Regression Model: Coefficients Interpretation
Wooldrige, p. 581

- Poisson Regression Model

$$E\left[y|x_1,\ldots,x_k\right] = \exp\left(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k\right)$$

- Change in $E\left[y|x_1,\ldots,x_k\right]$

$$\begin{aligned} & \{E[y|x_1,\ldots,x_k^{(1)}] - E[y|x_1,\ldots,x_k^{(0)}]\}/E[y|x_1,\ldots,x_k^{(0)}] \\ = & \frac{\exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k^{(1)}) - \exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k^{(0)})}{\exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k^{(0)})} \\ = & \exp(\beta_k \Delta x_k) - 1 \end{aligned}$$

where $\Delta x_k = x_k^{(1)} - x_k^{(0)}$

- $100\exp(\hat{\beta}_k)$ is the percentage change

The Poisson Regression Model: Coefficients Interpretation
Wooldrige, p. 581

- Poisson Regression Model

$$E\left[y|x_1,...,x_k\right] = \exp\left(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k\right)$$

- Take logs:

$$\ln\left(E\left[y|x_1,...,x_k\right]\right) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k$$

so that the log of the expected value is linear

- Using the approximation properties of the log

$$\%\Delta E\left(y|x\right) \approx \left(100\beta_j\right)\Delta x_j$$

- $\left(100\beta_j\right)$ is roughly the percentage change in $E\left(y|x\right)$ given one-unit increase in x_j

Estimation

The Poisson Regression Model

Maximum Likelihood Estimation

(Cameron and Trivedi p. 117 or Greene p. 843)

- Primary equation of the model

$$P\left(Y = y_i | \lambda_i\left(x_i\right)\right) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, \quad y = 0, 1, 2, \ldots$$

- The most common formulation for λ_i is the loglinear model:

$$\ln \lambda_i = x_i'\beta$$

- In this case:

$$E\left[y_i | x_i\right] = V\left[y_i | x_i\right] = \lambda_i = e^{x_i'\beta}$$

The Poisson Regression Model

Maximum Likelihood

(Cameron and Trivedi p. 117 or Greene p. 843)

$$y_i = E\left[y_i | x_i\right] + u_i$$

in this case is

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

and our objective is to estimate β by ML

The Poisson Regression Model

Maximum Likelihood

- Density: By iid

$$f\left(y_1, ..., y_n | x_i, \beta\right) = \prod_{i=1}^{n} f\left(y_i | x_i, \beta\right) = \prod_{i=1}^{n} \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}$$

- Conditional Likelihood function:

$$L_n\left(y_1, ..., y_n | x_1, ...x_n, \beta\right) = \prod_{i=1}^{n} L\left(y_i | x_i, \beta\right) = \prod_{i=1}^{n} \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}$$

Maximum Likelihood

- Conditional Likelihood function:

$$L_n\left(y_1,...,y_n|x_1,...x_n,\beta\right)=\prod_{i=1}^{n}L\left(y_i|x_i,\beta\right)=\prod_{i=1}^{n}\frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}$$

- Conditional Log-Likelihood function:

$$\mathcal{L}_n\left(\beta\right)=\ln\left(L_n\left(y_1,...,y_n|x_1,...x_n,\beta\right)\right)=\ln\left(\prod_{i=1}^{n}\frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}\right)$$

The Poisson Regression Model

Maximum Likelihood

Log-Likelihood function: $\ln(\lambda_i) = x_i'\beta$ or equivalently $\lambda_i = e^{x_i'\beta}$

$$
\begin{aligned}
\mathcal{L}_n(\beta) &= \ln\left(\prod_{i=1}^{n} \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}\right) \\
&= \sum_{i=1}^{n} \ln\left(\frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}\right) \\
&= \sum_{i=1}^{n} \left(-\lambda_i \ln(e) + y_i \ln(\lambda_i) - \ln(y_i!)\right) \\
&= \sum_{i=1}^{n} \left(-e^{x_i'\beta} + y_i x_i'\beta - \ln(y_i!)\right)
\end{aligned}
$$

The Poisson Regression Model

Extremum Estimators Examples: Maximum Likelihood

The Maximum Likelihood (ML) estimator is defined as

$$\hat{\beta} = \arg\max_{\beta} Q_n(\beta)$$

where

$$Q_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} \ln f(y_i|x_i, \beta)$$

and where $f(y_i|x_i, \beta)$ is the conditional likelihood for observation i.

Maximum Likelihood Example

Example: **The Poisson regression model**: The Maximum Likelihood (ML) estimator is defined as

$$\hat{\beta} = \arg\max_{\beta} Q_n(\beta)$$

where

$$\begin{aligned} Q_n(\beta) &= \frac{1}{n} \sum_{i=1}^{n} \ln f(y_i | x_i, \beta) \\ &= \frac{1}{n} \sum_{i=1}^{n} \left(-e^{x_i' \beta} + y_i x_i' \beta - \ln(y_i!) \right) \end{aligned}$$

Maximum Likelihood Example

Example: **The Poisson regression model**:

- MLE

$$\hat{\beta} = \arg\max_{\beta} Q_n(\beta) = \arg\max_{\beta} \left(\frac{1}{n} \sum_{i=1}^{n} \left(-e^{x_i'\beta} + y_i x_i'\beta - \ln(y_i!) \right) \right)$$

- FOC:

$$\frac{\partial Q_n(\beta)}{\partial \beta} = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - e^{x_i'\beta} \right) x_i = 0$$

- **System of Nonlinear equations!!! Numerical Methods**

Testing

The Poisson Regression Model

Asymptotics
Wooldridge, p. 606

- The general theory of MLE for random samples applies
- Under very general conditions, the MLE is consistent, asymptotically normal, and asymptotically efficient
- Hence, each $\hat{\beta}$ comes with an (asymptotic) standard error

$$\widehat{Avar}\left(\hat{\beta}\right) = \hat{\sigma}^2 \left(\sum_{i=1}^{n} \exp\left(x_i' \beta\right) x_i x_i' \right)^{-1}$$

which is a $k \times k$ matrix (see Wooldrige, p. 631)

Example:

Number of Arrests

Cross-sectional Data: Crime Data

- **Crime**: Data: Wooldrige (p. 4, 78,172, 295, 583)
  - $crime_i$: some measure of the frequency of criminal activity
  - Ex: $narr86_i$: number of times a man was arrested during 1986
  - $pcnv_i$: proportion of prior arrests leading to conviction
  - $tottime_i$: total time the man has spent in prison prior to 1986 since reaching the age of 18
  - $ptime86_i$: months spent in prison in 1986
  - $qemp86_i$: number of quarters in 1986 during which the man was legally employed

The Poisson Regression Model

Linear Model: Number of Arrests

Wooldrige, p. 608

$$
\begin{aligned}
\widehat{narr86} \quad = \quad & \underset{(0.038)}{0.577} - \underset{(0.040)}{0.132}pcnv - \underset{(0.012)}{0.011}avgsen \\
& + \underset{(0.009)}{0.012}tottime - \underset{(0.009)}{0.041}ptime86 \\
& - \underset{(0.014)}{0.051}qemp86 - \underset{(0.0003)}{0.0015}inc86 \\
& + \underset{(0.045)}{0.327}black + \underset{(0.040)}{0.194}hispan - \underset{(0.033)}{0.022}born60
\end{aligned}
$$

$$
\begin{array}{ll}
\log -likelihood & \\
R^2 & 0.073
\end{array}
$$

The Poisson Regression Model

Poisson Model: Number of Arrests

Wooldrige, p. 608

$$\begin{aligned}
\widehat{narr86} \quad = \quad & \exp(\underset{(0.067)}{-0.600} - \underset{(0.085)}{0.402}pcnv - \underset{(0.020)}{0.024}avgsen \\
& + \underset{(0.015)}{0.024}tottime - \underset{(0.021)}{0.099}ptime86 \\
& - \underset{(0.029)}{0.038}qemp86 - \underset{(0.0010)}{0.0081}inc86 \\
& + \underset{(0.074)}{0.661}black + \underset{(0.074)}{0.500}hispan - \underset{(0.064)}{0.051}born60)
\end{aligned}$$

$$\log -likelihood \quad -2248.76$$

$$R^2 \quad 0.077$$

The Poisson Regression Model

Interpreting the results
Wooldrige, p. 608

- OLS and Poisson coefficients: not directly comparable, very different meanings. Example: coefficient on *pcnv*:
- Linear Model: -0.132: if $\Delta pcnv = 0.1$, expected number of arrests falls by 0.013
- Poisson Model: -0.402: if $\Delta pcnv = 0.1$, expected number of arrests falls by 4% ($0.402(0.1) = 0.0402$ and multiply by 100 to get the percentage effect)
- Poisson coefficient on *black*: 0.661: expected number of arrests for a black man is $100\left(\exp\left(0.661\right) - \exp\left(0\right)\right) \approx 93.7\%$ higher than for a white man with the same values for the other explanatory variables