ECON 613: Applied Econometrics

Methods for Cross-sectional Data

March 4, 2019

Introduction

Binary response models are models where the variable to be explained y is a random variable taking on the values zero and one which indicate whether or not a certain event has occured.

- $y = 1$ if a person is employed
- $y = 1$ if a family contributes to a charity during a particular year
- $y = 1$ if a firm has a particular type of pension plan
- $y = 1$ if a worker goes to college
- Regardless of what y stands for, we refer to $y = 1$ as a success and $y = 0$ as a failure.

An OLS regression of y on dependent variables denoted x ignores the discreteness of the dependent variable and does not constrain predicted probabilities to be between zero and one.

Linear Probability Model (1)

$$P(y=1) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k. \tag{1}$$

- If x_1 is continuous, β_1 is the change in the probability of success given one unit increase in x_1
- If x_1 is discrete, β_1 is the difference in the probability of success when $x_1 = 1$ and $x_1 = 0$, holding other x_j fixed.

Linear Probability Model (2)

Given that y is a random variable (Bernouilli), we also have

$$\begin{aligned} E(y \mid x) &= \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k. && (2) \\ Var(y \mid x) &= x\beta(1 - x\beta) && (3) \end{aligned}$$

Implications:

- OLS regression of y on $x_1, x_2, \ldots, x_k$ produces consistent and unbiaised estimators of the β_j.
- Heteroskedacticity, which can be dealt with using standard heteroskedasticity-robust standard errors.
- Problem: OLS fitted values may not be between zero and one.

Maximum Likelihood Estimation: Logit and Probit Models

For binary outcome data, the dependent variable y takes one of two values. We let

$$y = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } 1 - p \end{cases} \tag{4}$$

Parametrize conditional probabilities:

$$p_i = F_\epsilon(X\beta) \tag{5}$$

And, Marginal Effects

$$\frac{\partial Pr(y_i = 1 \mid x_i)}{\partial x_{ij}} = F'_\epsilon(X\beta)\beta_j \tag{6}$$

Probit Model (1)

The probit model corresponds to the case where $F(x)$ is the cumulative standard normal distribution function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(\tfrac{1}{2}X^2)dX \tag{7}$$

Where $F(X\beta) = \Phi(X\beta)$.

Probit Model (2)

- Consider the latent approach

$$y^\star = X\beta + \epsilon \tag{8}$$

where $\epsilon \sim N(0,1)$. Think of $y^\star$ as the net utility associated with some action. If the action yields positive net utility, it is undertaken otherwise it is not.

- We would care only about the sign of $y^\star$

$$y = \begin{cases} 1, & \text{if } p \\ 0, & \text{with probability } 1-p \end{cases} \tag{9}$$

- Probabilities

$$\begin{aligned} Pr(y=1) &= Pr(y^\star > 0) = Pr(X\beta + \epsilon \geq 0) \\ &= Pr(\epsilon \geq -X\beta) = Pr(\epsilon \leq X\beta) = \Phi(X\beta) \end{aligned}$$

Logit Model

The logit model specifies the cdf function $F(x)$ is now the logistic function

$$\Lambda(x) = \frac{1}{1 + \exp(-x)} = \frac{exp(x)}{1 + \exp(x)} \tag{10}$$

The logit model is most easily derived by assuming that

$$\log\left(\frac{P}{1 - P}\right) = X\beta \tag{11}$$

The logarithm of the odds (ratio of two probabilities) is equal to $X\beta$

Maximum Likelihood Estimation

- Likelihood can not be defined as a joint density function.
- Outcome of a Bernouilli trial

$$f(y_i \mid x_i) = p_i^{y_i}(1 - p_i)^{1-y_i} \tag{12}$$

- Given the independence of individuals, the likelihood can be written

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} F(x_i\beta)^{y_i}(1 - F(x_i\beta))^{1-y_i} \tag{13}$$

Log Likelihood

- The log likelihood

$$\log \mathcal{L}(\beta) = \sum_{i=1}^{n} y_i lnF(x_i\beta) + (1 - y_i) ln(1 - F(x_i\beta)) \tag{14}$$

- First order conditions

$$\sum_{i=1}^{n} \frac{y_i - F(x_i\beta)}{F(x_i\beta)(1 - F(x_i\beta))} F^{'}(x_i\beta)x_i = 0 \tag{15}$$

Empirical considerations

- Probit and logit yield same outcomes. Only difference is how parameters are scaled.
- The natural metric to compare models is the fitted log-likelihood provided that the models have the same number of parameters.
- Although estimated parameters are different, marginal effects are quite similar.

Pseudo R2

$$R^2_{\text{Binary}} = 1 - \frac{\mathcal{L}(\hat{\beta})}{N[\bar{y} ln \bar{y} + (1 - \bar{y}) ln(1 - \bar{y})]} \tag{16}$$

Predicted Outcomes

- The criterion $\sum_i (y_i - \hat{y}_i)^2$ gives the number of wrong predictions.
  - average rule: let $\hat{y} = 1$ when $\hat{p} = F(X\beta) > 0.5$
  - Receiver Operating Characteristics (ROC) curve plots the fractions of $y = 1$ correctly classified against the fractions of $y = 0$ incorrectly specified as the cutoffs $\hat{p} = F(X\beta) > c$ varies.

Example: Describing the data

		No Affair	Affair
Gender	female	0.54	0.48
	male	0.46	0.52
Age	17.5	0.01	0.02
	22	0.22	0.11
	27	0.26	0.24
	32	0.17	0.25
	37	0.14	0.15
	42	0.08	0.12
	47	0.04	0.05
	52	0.03	0.04
	57	0.04	0.02
Years Married	0.125	0.02	0.01
	0.417	0.02	0.01
	0.75	0.06	0.02
	1.5	0.17	0.08
	4	0.17	0.18
	7	0.13	0.15
	10	0.11	0.14
	15	0.31	0.41
Share		0.75	0.25

Example: Affairs

		No Affair	Affair
Children	no	0.319	0.18
	yes	0.681	0.82
Religiousness	1	0.062	0.133
	2	0.273	0.273
	3	0.191	0.287
	4	0.348	0.22
	5	0.126	0.087
Education	9	0.011	0.013
	12	0.069	0.087
	14	0.264	0.233
	16	0.211	0.133
	17	0.137	0.18
	18	0.175	0.22
	20	0.133	0.133

Probit

	Model 1	Model 2	Model 3	Model 4
(Intercept)	-0.74^{***}	-0.02	-0.18	-0.02
	(0.08)	(0.51)	(0.52)	(0.81)
as.factor(gender)male	0.14	0.13	0.14	0.21
	(0.11)	(0.12)	(0.12)	(0.12)
as.factor(age)22		-1.12^{*}	-1.07^{*}	-1.45^{*}
		(0.53)	(0.54)	(0.61)
as.factor(age)27		-0.76	-0.82	-1.44^{*}
		(0.53)	(0.53)	(0.62)
as.factor(age)32		-0.48	-0.60	-1.36^{*}
		(0.53)	(0.53)	(0.63)
as.factor(age)37		-0.70	-0.84	-1.79^{**}
		(0.53)	(0.54)	(0.66)
as.factor(age)42		-0.50	-0.65	-1.61^{*}
		(0.54)	(0.55)	(0.67)
as.factor(age)47		-0.56	-0.68	-1.72^{*}
		(0.58)	(0.59)	(0.71)
as.factor(age)52		-0.62	-0.76	-1.75^{*}
		(0.59)	(0.60)	(0.71)
as.factor(age)57		-1.17	-1.31^{*}	-2.36^{**}
		(0.62)	(0.62)	(0.74)
as.factor(children)yes			0.31^{*}	0.11
			(0.15)	(0.17)
as.factor(yearsmarried)0.417				0.01
				(0.77)
as.factor(yearsmarried)0.75				-0.37
				(0.67)
as.factor(yearsmarried)1.5				0.22
				(0.56)
as.factor(yearsmarried)4				0.62
				(0.57)
Log Likelihood	-336.91	-327.84	-325.74	-319.69

Logit

	Model 1	Model 2	Model 3	Model 4
(Intercept)	-1.22^{***}	-0.03	-0.31	-0.07
	(0.13)	(0.82)	(0.84)	(1.48)
as.factor(gender)male	0.24	0.21	0.23	0.34
	(0.19)	(0.20)	(0.20)	(0.21)
as.factor(age)22		-1.88^{*}	-1.78^{*}	-2.44^{*}
		(0.86)	(0.87)	(1.05)
as.factor(age)27		-1.24	-1.35	-2.41^{*}
		(0.84)	(0.85)	(1.06)
as.factor(age)32		-0.77	-0.99	-2.29^{*}
		(0.84)	(0.86)	(1.08)
as.factor(age)37		-1.14	-1.38	-2.99^{**}
		(0.86)	(0.88)	(1.13)
as.factor(age)42		-0.82	-1.07	-2.70^{*}
		(0.87)	(0.89)	(1.14)
as.factor(age)47		-0.90	-1.11	-2.87^{*}
		(0.94)	(0.95)	(1.21)
as.factor(age)52		-1.00	-1.26	-2.95^{*}
		(0.95)	(0.97)	(1.21)
as.factor(age)57		-1.97	-2.22^{*}	-3.99^{**}
		(1.03)	(1.05)	(1.29)
as.factor(children)yes			0.54^{*}	0.16
			(0.27)	(0.30)
as.factor(yearsmarried)0.417				0.09
				(1.49)
as.factor(yearsmarried)0.75				-0.50
				(1.31)
as.factor(yearsmarried)1.5				0.42
				(1.10)
as.factor(yearsmarried)4				1.11
				(1.10)
Log Likelihood	-336.91	-327.90	-325.85	-320.24

Probit VS Logit

	Probit	Logit
(Intercept)	-0.02	-0.07
	(0.81)	(1.48)
as.factor(gender)male	0.21	0.34
	(0.12)	(0.21)
as.factor(age)22	-1.45^{*}	-2.44^{*}
	(0.61)	(1.05)
as.factor(age)27	-1.44^{*}	-2.41^{*}
	(0.62)	(1.06)
as.factor(age)32	-1.36^{*}	-2.29^{*}
	(0.63)	(1.08)
as.factor(age)37	-1.79^{**}	-2.99^{**}
	(0.66)	(1.13)
as.factor(age)42	-1.61^{*}	-2.70^{*}
	(0.67)	(1.14)
as.factor(age)47	-1.72^{*}	-2.87^{*}
	(0.71)	(1.21)
as.factor(age)52	-1.75^{*}	-2.95^{*}
	(0.71)	(1.21)
as.factor(age)57	-2.36^{**}	-3.99^{**}
	(0.74)	(1.29)
as.factor(children)yes	0.11	0.16
	(0.17)	(0.30)
as.factor(yearsmarried)0.417	0.01	0.09
	(0.77)	(1.49)
AIC	675.38	676.48
BIC	754.55	755.66
Log Likelihood	-319.69	-320.24

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Identification considerations

- β is identified up to a scale.
- We observe only whether $X\beta + \epsilon > 0$.
- Implication for the interpretation of the coefficients.

Model Selection

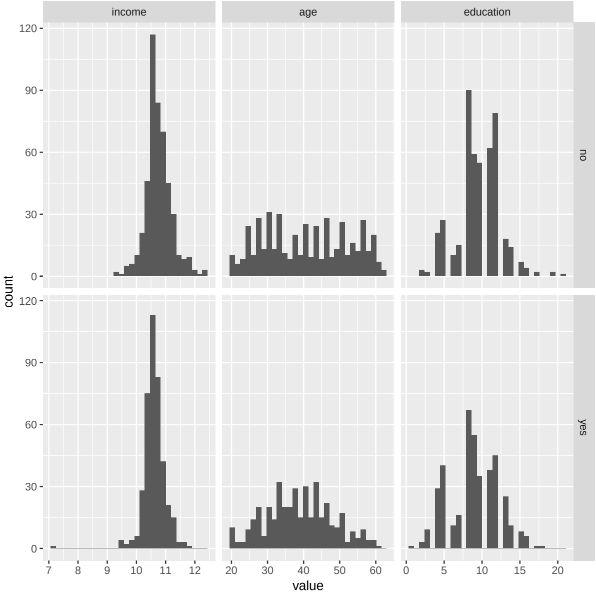
- $AIC = 2k - 2\log\mathcal{L}$
- $BIC = \log(n)k - 2\log\mathcal{L}$

Determinants of Female Labor Supply

Cross-section data originating from the health survey SOMIPOPS for Switzerland in 1981.

- participation Factor: Did the individual participate in the labor force?
- income: Logarithm of nonlabor income.
- age: Age in decades (years divided by 10).
- education: Years of formal education.
- youngkids: Number of young children (under 7 years of age).
- oldkids:Number of older children (over 7 years of age).
- foreign Factor: Is the individual a foreigner (i.e., not Swiss)?

Describing the data (1)



Describing the data (2)

		Participation	
		No	Yes
Number of young kids	0	0.6921	0.8454
	1	0.2123	0.1172
	2	0.0892	0.0324
	3	0.0064	0.005
Number of old kids	0	0.4904	0.404
	1	0.2166	0.2394
	2	0.2166	0.2643
	3	0.0573	0.0698
	4	0.0149	0.0175
	5	0.0042	0
	6	0	0.005

Effect of income

	Model 1	Model 2	Model 3
(Intercept)	5.97^{***}	-16.60	116.89
	(1.19)	(10.14)	(67.96)
income	-0.57^{***}	3.69	-37.41
	(0.11)	(1.91)	(20.52)
$I(income^2)$		-0.20^{*}	3.97
		(0.09)	(2.06)
$I(income^3)$			-0.14^{*}
			(0.07)
Log Likelihood	-587.91	-585.54	-583.06
Num. obs.	872	872	872

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Effect of age

	Model 1	Model 2	Model 3
(Intercept)	0.34^{*}	-3.52^{***}	-1.10
	(0.17)	(0.63)	(2.20)
age	-0.01^{**}	0.19^{***}	-0.00
	(0.00)	(0.03)	(0.18)
$I(age^2)$		-0.00^{***}	0.00
		(0.00)	(0.00)
$I(age^3)$			-0.00
			(0.00)
Log Likelihood	-597.85	-576.39	-575.71
Num. obs.	872	872	872

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Specification

	Model 1	Model 2	Model 3
(Intercept)	7.66^{***}	4.65^{***}	-7.42
	(1.28)	(1.40)	(10.73)
income	-0.56^{***}	-0.74^{***}	1.55
	(0.12)	(0.13)	(2.03)
age	-0.03^{***}	0.23^{***}	0.23^{***}
	(0.01)	(0.04)	(0.04)
education	-0.03	-0.02	-0.02
	(0.02)	(0.02)	(0.02)
youngkids	-0.71^{***}	-0.64^{***}	-0.64^{***}
	(0.10)	(0.10)	(0.10)
oldkids	-0.01	-0.16^{**}	-0.16^{**}
	(0.04)	(0.05)	(0.05)
$I(age^2)$		-0.00^{***}	-0.00^{***}
		(0.00)	(0.00)
$I(income^2)$			-0.11
			(0.10)
Log Likelihood	-550.08	-526.38	-525.86
Num. obs.	872	872	872

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Marginal Effects

- Recall that marginal effects are given by

$$\frac{\partial Pr(y_i = 1 \mid x_i)}{\partial x_{ij}} = F'_{\epsilon}(X\beta)\beta_j \tag{17}$$

- Different definitions
  - Average marginal effects in the sample.
  - Marginal effect evaluated at the mean.

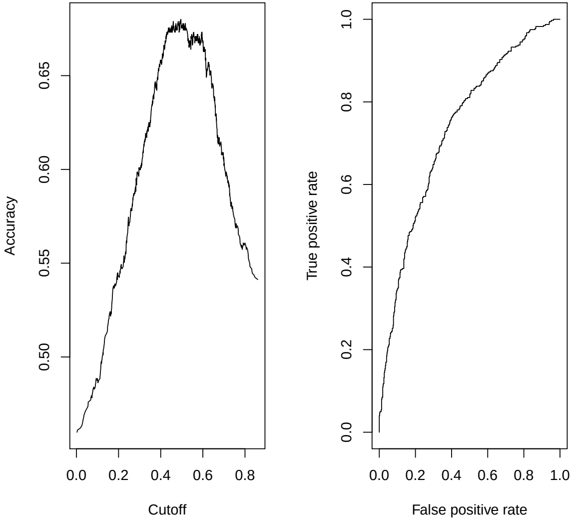
Marginal Effects

income	0.53
age	0.08
education	-0.01
youngkids	-0.22
oldkids	-0.05
I(age^2)	-0.00
I($income^2$)	-0.04

Prediction

	Model	
Data	0	1
0	325	146
1	137	264

ROC



Accuracy
0.50
0.55
0.60
0.65
0.0
0.2
0.4
0.6
0.8
Cutoff
True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate

By citizenship

	Native	Immigrants
(Intercept)	-6.85	-67.57
	(11.24)	(48.18)
income	1.17	13.31
	(2.11)	(9.16)
age	0.24^{***}	0.14
	(0.05)	(0.09)
education	0.05^{*}	-0.03
	(0.02)	(0.04)
youngkids	-0.76^{***}	-0.75^{***}
	(0.13)	(0.18)
oldkids	-0.16^{**}	-0.18
	(0.06)	(0.12)
$\mathrm{I}(age^2)$	-0.00^{***}	-0.00
	(0.00)	(0.00)
$\mathrm{I}(income^2)$	-0.09	-0.66
	(0.10)	(0.43)
Log Likelihood	-384.24	-117.82
Num. obs.	656	216

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Murder Rates

Cross-section data on states in 1950.

- rate: Murder rate per 100,000 (FBI estimate, 1950).
- convictions: Number of convictions divided by number of murders in 1950.
- executions: Average number of executions during 1946–1950 divided by convictions in 1950.
- time: Median time served (in months) of convicted murderers released in 1951.
- income: Median family income in 1949 (in 1,000 USD).
- lfp: Labor force participation rate in 1950 (in percent).
- noncauc: Proportion of population that is non-Caucasian in 1950.
- southern: Factor indicating region.

Warnings

- Binary model of the determinants of having an execution
- This is very bad economics
- An example to illustrate some technical problems

```
glm(formula = I(executions > 0) ~ time + income + noncauc +
    southern, family = binomial, data = MurderRates)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  10.99326   20.77336   0.529   0.5967
time          0.01943    0.01040   1.868   0.0617 .
income       10.61013    5.65409   1.877   0.0606 .
noncauc      70.98785   36.41181   1.950   0.0512 .
lfp          -0.66763    0.47668  -1.401   0.1613
southernyes  17.33126 2872.17069   0.006   0.9952
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
```

- Diagnostic
- Point estimate and Std.

```
table(I(MurderRates$executions > 0), MurderRates$southern)

        no yes
  FALSE  9   0
  TRUE  20  15
```

Quasi Separation

- We have here β^0 such that

$$\begin{aligned} y_i = 0 \quad & \text{whenever} \quad x_i'\beta^0 \leq 0 \\ y_i = 1 \quad & \text{whenever} \quad x_i'\beta^0 \geq 0 \end{aligned}$$

- The maximum likelihood estimate does not exist.

Multinomial Models

Dependent variable has several possible outcomes, that are mutually exclusive

- Commute to work (car, bus, bike, walking)
- Employment status (full time, part time, unemployed)
- Occupation choice, field of study, product choice

- Ordered choices eg. education choices
- Unordered choices eg. fishing mode

Ordered Discrete Response

Suppose that $y^{\star}(= x_i\beta + \epsilon_i)$ is continuously distributed with standard deviation σ but the observed response y_i is an ordered discrete choice (ODR) taking 0,1,...,R-1 determined by fixed threesholds γ_r. Formally, we have

$$y = \begin{cases} 0, & \text{if } x\beta + \epsilon < \gamma_1 \\ 1, & \text{if } \gamma_1 \leq x\beta + \epsilon < \gamma_2 \\ 2, & \text{if } \gamma_2 \leq x\beta + \epsilon < \gamma_3 \\ \dots & \\ R-1, & \text{if } \gamma_{R-1} \leq x\beta + \epsilon \end{cases} \tag{18}$$

Identification

- Not all parameters are identified as in the binary response model.
- Consider $\gamma_r \leq x\beta + \epsilon < \gamma_{r+1}$. Then, we have $\frac{\gamma_r - \gamma_1}{\sigma} \leq \frac{x\beta + \epsilon - \gamma}{\sigma} < \frac{\gamma_{r+1} - \gamma_1}{\sigma}$
- Then, the identified parameters are

$$\alpha = \left(\frac{\beta_1 - \gamma_1}{\sigma}, \frac{\beta_2}{\sigma}, \ldots, \frac{\beta_k}{\sigma} \right) \tag{19}$$

$$\rho_r = \frac{\gamma_r - \gamma_1}{\sigma} \tag{20}$$

for $r = 2, \ldots, R - 1$.

Toward the Likelihood

$y = r$

$$
\begin{aligned}
\gamma_r &\le x\beta + \epsilon &&< \gamma_{r+1} \\
\gamma_r - x\beta &\le \epsilon &&< \gamma_{r+1} - x\beta \\
\frac{\gamma_r - \gamma_1}{\sigma} + \frac{\gamma_1 - x\beta}{\sigma} &\le \frac{\epsilon}{\sigma} &&< \frac{\gamma_{r+1} - \gamma_1}{\sigma} + \frac{\gamma_1 - x\beta}{\sigma} \\
\rho_r - x\alpha &\le \frac{\epsilon}{\sigma} &&< \rho_{r+1} - x\alpha
\end{aligned}
$$

Likelihood

- Choice probabilities

$$\begin{aligned} P(y = r \mid x) &= P(\rho_r - x\alpha \leq \frac{\epsilon}{\sigma} < \rho_{r+1} - x\alpha) \\ &= F(\rho_{r+1} - x\alpha) - F(\rho_r - x\alpha) \end{aligned}$$

- Likelihood

$$\mathcal{L}(a, t) = \prod_{i=1}^{N} \prod_{r=0}^{R-1} [P(y_i = r \mid x)]^{y_{ir}} \tag{21}$$

- Log Likelihood of the probit model

$$\log \mathcal{L}(a, t) = \sum_{i=1}^{N} \sum_{r=0}^{R-1} y_{ir} \log(\Phi(t_{r+1} - xa) - \Phi(t_r - xa)) \tag{22}$$

where a and t are respectively the parameters to be estimated and the cutoffs.

Marginal Effects

- The marginal effects of x on each choice probabilities can be derived as:

$$\frac{\partial P(y = r \mid x)}{\partial x} = -\alpha(\phi(\rho_{r+1} - x\alpha) - \phi(\rho_r - x\alpha)) \tag{23}$$

- Caution

$$\sum_{r=0}^{R} P(y = r \mid x) = 1 \tag{24}$$

Then

$$\sum_{r=0}^{R} \frac{\partial P(y = r \mid x)}{\partial x} = 0 \tag{25}$$

An increase in some choice probability necessarily entails a decrease in some other choice probabilities.

Multinomial Logit

The models differ according to whether or not regressors vary across alternatives.

- Conditional logit model

$$p_{ij} = \frac{\exp(x_{ij}\beta)}{\sum_{l=1}^{m}\exp(x_{il}\beta)} \qquad j = 1, \ldots, m. \tag{26}$$

- Multinomial logit model

$$p_{ij} = \frac{\exp(x_i\beta_j)}{\sum_{l=1}^{m}\exp(x_i\beta_l)} \qquad j = 1, \ldots, m. \tag{27}$$

- Mixed logit model

$$p_{ij} = \frac{\exp(x_{ij}\beta + w_i\gamma_j)}{\sum_{l=1}^{m}\exp(x_{il}\beta + w_i\gamma_l)} \qquad j = 1, \ldots, m. \tag{28}$$

Marginal Effects

- Conditional logit

$$\frac{\partial p_{ij}}{\partial x_{ik}} = p_{ij}(\delta_{ijk} - p_{ik})\beta \tag{29}$$

where δ_{ijk} is an indicator variable equal to 1 if $j = k$ and equal to 0 otherwise.

- Multinomial logit

$$\frac{\partial p_{ij}}{\partial x_i} = p_{ij}(\beta_j - \bar{\beta}_i) \tag{30}$$

where $\bar{\beta}_i = \sum_l p_{il}\beta_l$

Independence of Irrelevant Alternatives

- A property of the conditional logit and multinomial logit is that discrimination among the m alternatives reduces to a series of pairwise comparisons that are unaffected by the characteristics of alternatives other than the pair under consideration.
- The choice probabilities must be unaffected by the removal of one alternative.

That is because

$$Pr(y = j \mid y = k) = \frac{p_j}{p_j + p_k} \tag{31}$$

Testing for IIA

- Estimate the model twice
  - On the full set of alternatives and obtain θ_{full}
  - On a subset of alternatives and obtain θ_{subset}
- Compare $\mathcal{L}_{subset}(\theta_{full})$ and $\mathcal{L}_{subset}(\theta_{subset})$. If there is a significant different, then IIA is violated.

It is very (very) rare that IIA is not violated.

Hausman and McFadden Test

$$HM = (\hat{\beta}^r - \hat{\beta}^f)'[var_{\hat{\beta}^r} - var_{\hat{\beta}^f}]^{-1}(\hat{\beta}^r - \hat{\beta}^f) \tag{32}$$

We have that

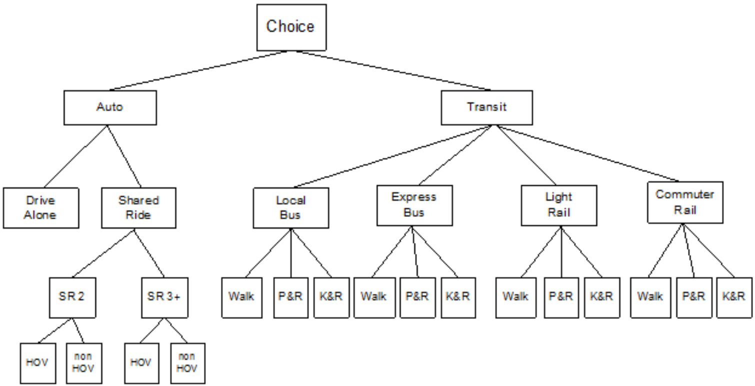
$$HM \sim \chi^2(||\beta^r||) \tag{33}$$

If IIA holds.

Alternatives

- Generalized Extreme Value Model
- Nested Logit Model
- Random Parameters Logit
- Multinomial Probit

Nested Logit



Nested Logit

The nested logit model breaks decision making into groups. The utility for the alternative is given

$$U_{jk} = V_{jk} + \epsilon_{jk} \quad k = 1, 2, \ldots, K_j, \quad j = 1, 2, \ldots, J \tag{34}$$

Utilities are given by:

- $V_{11} + \epsilon_{11}$
- $\ldots$
- $V_{JK_J} + \epsilon_{JK_J}$

Choice Probabilities

- Choice probability
$$p_{jk} = p_j \times p_{k|j}. \tag{35}$$
- This arises from GEV joint distribution
$$F(\epsilon) = exp(-G(e^{-\epsilon_{11}}, \ldots, e^{-\epsilon_{1K_1}}; \ldots; e^{-\epsilon_{J1}}, \ldots, e^{-\epsilon_{JK_J}}))$$
with
$$G(Y) = G(Y_{11}, \ldots, Y_{1K_1}, \ldots, Y_{JK_J}) = \sum_{j=1}^{J} \left(\sum_{k=1}^{K_j} Y_{jk}^{\frac{1}{\rho_j}} \right)^{\rho_j} \tag{36}$$

Model

Consider

$$V_{jk} = z_j\alpha + x_{jk}\beta_j \quad k = 1, \ldots, K_j, \quad j = 1, \ldots, J \tag{37}$$

The probability of the nested logit model

$$p_{jk} = p_j \times p_{k|j} = \frac{\exp(z_j\alpha + \rho_j I_j)}{\sum_{m=1}^{J} \exp(z_m\alpha + \rho_m I_m)} \times \frac{\exp(x_{jk}\beta/\rho_j + \rho_j I_j)}{\sum_{m=1}^{J} \exp(z_m\alpha + \rho_m I_m)}$$

where

$$I_j = ln\left(\sum_{l=1}^{K_j} \exp(x_{jl}\beta_j/\rho_j)\right)$$

is the inclusive value or the log-sum.

Likelihood

For the *ith* observation, we observe $K_1 + \ldots + K_J$ outcomes y_{ijk}, where $y_{ijk} = 1$ if alternative jk is chosen and is zero otherwise. Then the density of one observation y_i can be expressed

$$f(y_i) = \prod_{j=1}^{J} \prod_{k=1}^{K_J} [p_{ij} \times p_{ik|j}]^{y_{ijk}} = \prod_{j=1}^{J} p_{ij}^{y_{ij}} \left(\prod_{k=1}^{K_J} p_{ik|j}^{y_{ijk}} \right)$$

The log likelihood is given by

$$lnL = \sum_{i=1}^{N} \sum_{y=1}^{J} y_{ij} \log(p_{ij}) + \sum_{i=1}^{N} \sum_{y=1}^{J} \sum_{k=1}^{K_J} y_{ijk} \log(p_{ik|j})$$

Discussions and Limitations

- Nested Logit can be estimated in two steps following the construction of the densities
- Not all choices are easy to nest.

Multinomial Probit

- Consider m-choice with

$$U_j = V_j + \epsilon_j \qquad j = 1, 2, \dots, m. \tag{38}$$

  where $\epsilon \sim \mathbb{N}(0, \Sigma)$.
- Σ is the matrix of variance-covariance, which can be left unrestricted to capture the correlation between choices.
- Choice probabilities.. For 4 choices, we have

$$P(Y = 1) = \int_{-\infty}^{-V_{41}} \int_{-\infty}^{-V_{31}} \int_{-\infty}^{-V_{21}} f(x, y, z) dzdydx \tag{39}$$

  where $f(x, y, z)$ is the pdf of the trivariate normal.

Need simulation to compute the choice probabilities and the likelihood.

Applications

- Conditional logit
- Multinomial logit
- Mixed logit

Data: Car

Sample of 4654 individuals stating preferences for cars

- choice: choice of a vehicule amoung 6 propositions,
- college: college education?,
- hsg2: size of household greater than 2?
- coml5: commute lower than 5 miles a day?,
- type: body type, one of regcar (regular car), sportuv (sport utility vehicule), sportcar, stwagon
- (station wagon), truck, van, for each proposition z from 1 to 6,
- fuel: fuel for proposition z, one of gasoline, methanol, cng (compressed natural gas), electric.,
- price: price of vehicule divided by the logarithm of income,
- range: hundreds of miles vehicule can travel between refuelings/rechargings,

Data

- acc: acceleration, tens of seconds required to reach 30 mph from stop,
- speed: highest attainable speed in hundreds of mph,
- pollution: tailpipe emissions as fraction of those for new gas vehicule,
- size: 0 for a mini, 1 for a subcompact, 2 for a compact and 3 for a midsize or large vehicule,
- space: fraction of luggage space in comparable new gas vehicule,
- cost: cost per mile of travel (tens of cents) : home recharging for electric vehicule, station refueling otherwise
- station: fraction of stations that can refuel recharge vehicule

```
summary(car[,6:7])
     type             fuel
regcar  :10930   gasoline:6958
sportuv : 1048   methanol:6952
sportcar:  880   cng     :7016
stwagon : 4446   electric:6998
truck   : 5628
van     : 4992
```

Descriptive Statistics

Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
college	0.429	0.514	0	0	1	1
hsg2	0.571	0.514	0	0	1	1
coml5	0.429	0.514	0	0	1	1
alt	3.214	1.805	1	2	4.8	6
price	4.212	0.620	3.311	3.700	4.717	5.139
range	275.000	79.663	125	250	300	400
acc	4.429	1.530	2	2.9	6	6
speed	115.000	23.534	85	95	140	140
pollution	0.300	0.206	0.000	0.138	0.475	0.600
size	2.571	0.514	2	2	3	3
space	0.914	0.141	0.700	0.775	1.000	1.000
cost	5.714	1.729	4	4	7.5	8
station	0.371	0.421	0.000	0.100	0.825	1.000

Conditional Logit

	Model 1
2:(intercept)	-0.98^{***}
	(0.07)
3:(intercept)	0.46^{***}
	(0.04)
4:(intercept)	-0.70^{***}
	(0.07)
5:(intercept)	0.56^{***}
	(0.04)
6:(intercept)	-0.77^{***}
	(0.08)
typesportuv	-0.05
	(0.15)
typesportcar	-0.18
	(0.16)
typestwagon	-0.57^{***}
	(0.08)
typetruck	-0.49^{***}
	(0.06)
typevan	-0.14^{*}
	(0.07)
price	-0.19^{***}
	(0.03)
cost	-0.07^{***}
	(0.01)
AIC	14473.98
Log Likelihood	-7224.99
Num. obs.	4654

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table: Statistical models

Conditional Logit: Interactions

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
2:(intercept)	-1.19^{***}	-1.19^{***}	-1.19^{***}	-1.19^{***}	-1.19^{***}	-1.19^{***}
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
3:(intercept)	-0.54^{***}	-0.56^{***}	-0.54^{***}	-0.55^{***}	-0.55^{***}	-0.56^{***}
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
4:(intercept)	-1.89^{***}	-1.91^{***}	-1.89^{***}	-1.90^{***}	-1.90^{***}	-1.91^{***}
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
5:(intercept)	-1.02^{***}	-1.05^{***}	-1.02^{***}	-1.03^{***}	-1.03^{***}	-1.05^{***}
	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)	(0.15)
6:(intercept)	-2.61^{***}	-2.64^{***}	-2.61^{***}	-2.62^{***}	-2.62^{***}	-2.64^{***}
	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)
fuelmethanol	-1.18^{***}	-1.30^{***}	-1.28^{***}	-1.26^{***}	-1.30^{***}	-1.30^{***}
	(0.18)	(0.18)	(0.18)	(0.18)	(0.18)	(0.18)
fuelcng	-0.66^{***}	-0.67^{***}	-0.66^{***}	-0.66^{***}	-0.66^{***}	-0.67^{***}
	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)
fuelelectric	-0.13	-0.13	-0.12	-0.13	-0.12	-0.13
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
range	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
price	-0.19^{***}	-0.19^{***}		-0.12^{***}		-0.22^{***}
	(0.03)	(0.03)		(0.03)		(0.03)
cost		-0.08^{***}			-0.05^{**}	-0.11^{***}
		(0.01)			(0.02)	(0.02)
I(price * cost)			-0.02^{***}	-0.01^{***}	-0.01^{*}	0.01
			(0.00)	(0.00)	(0.00)	(0.00)
AIC	14309.01	14196.72	14247.97	14229.76	14240.86	14195.58
Log Likelihood	-7144.50	-7087.36	-7113.98	-7103.88	-7109.43	-7085.79
Num. obs.	4654	4654	4654	4654	4654	4654

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table: Statistical models

Conditional Logit: Effect of size

	Model 1	Model 2	Model 3
2:(intercept)	-0.98^{***}	-0.98^{***}	-0.98^{***}
	(0.07)	(0.07)	(0.07)
3:(intercept)	0.42^{***}	0.41^{***}	0.41^{***}
	(0.04)	(0.04)	(0.04)
4:(intercept)	-0.74^{***}	-0.74^{***}	-0.74^{***}
	(0.07)	(0.07)	(0.07)
5:(intercept)	0.53^{***}	0.53^{***}	0.53^{***}
	(0.04)	(0.04)	(0.04)
6:(intercept)	-0.81^{***}	-0.81^{***}	-0.81^{***}
	(0.07)	(0.07)	(0.07)
typesportuv	-0.04	-0.04	-0.04
	(0.15)	(0.15)	(0.15)
typesportcar	-0.17	-0.17	-0.17
	(0.16)	(0.16)	(0.16)
typestwagon	-0.55^{***}	-0.55^{***}	-0.55^{***}
	(0.08)	(0.08)	(0.08)
typetruck	-0.49^{***}	-0.49^{***}	-0.49^{***}
	(0.06)	(0.06)	(0.06)
typevan	-0.15^{*}	-0.15^{*}	-0.15^{*}
	(0.07)	(0.07)	(0.07)
size	0.05^{*}	0.03	
	(0.02)	(0.10)	
$I(size^2)$		0.01	
		(0.02)	
as.factor(size)1			0.06
			(0.12)
as.factor(size)2			0.09
			(0.12)
as.factor(size)3			0.15
			(0.12)
Log Likelihood	-7296.31	-7296.29	-7296.24
Num. obs.	4654	4654	4654

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Conditional Logit: Effect of size (2)

	Model 1	Model 2	Model 3
2:(intercept)	-1.19^{***}	-1.19^{***}	-1.19^{***}
	(0.07)	(0.07)	(0.07)
3:(intercept)	0.45^{***}	0.45^{***}	0.45^{***}
	(0.04)	(0.04)	(0.04)
4:(intercept)	-0.90^{***}	-0.90^{***}	-0.90^{***}
	(0.06)	(0.06)	(0.06)
5:(intercept)	0.56^{***}	0.56^{***}	0.56^{***}
	(0.04)	(0.04)	(0.04)
6:(intercept)	-1.03^{***}	-1.03^{***}	-1.03^{***}
	(0.07)	(0.07)	(0.07)
price	-0.19^{***}	-0.19^{***}	-0.19^{***}
	(0.03)	(0.03)	(0.03)
cost	-0.07^{***}	-0.07^{***}	-0.07^{***}
	(0.01)	(0.01)	(0.01)
size	0.06^{*}	0.02	
	(0.02)	(0.10)	
$I(size^2)$		0.01	
		(0.02)	
as.factor(size)1			0.05
			(0.12)
as.factor(size)2			0.09
			(0.12)
as.factor(size)3			0.16
			(0.12)
AIC	14544.68	14546.55	14548.50
Log Likelihood	-7264.34	-7264.28	-7264.25
Num. obs.	4654	4654	4654

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Conditional Logit: Effect of size (3)

	Model 1	Model 2	Model 3	Model 4	Model 5
2:(intercept)	-1.19^{***}	-1.19^{***}	-1.19^{***}	-1.19^{***}	-1.19^{***}
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
3:(intercept)	0.45^{***}	0.45^{***}	0.45^{***}	0.45^{***}	0.45^{***}
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
4:(intercept)	-0.90^{***}	-0.90^{***}	-0.90^{***}	-0.89^{***}	-0.90^{***}
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
5:(intercept)	0.56^{***}	0.56^{***}	0.56^{***}	0.56^{***}	0.56^{***}
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
6:(intercept)	-1.03^{***}	-1.03^{***}	-1.03^{***}	-1.03^{***}	-1.03^{***}
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
price	-0.19^{***}	-0.19^{***}	-0.19^{***}	-0.19^{***}	-0.19^{***}
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
cost	-0.07^{***}	-0.07^{***}	-0.07^{***}	-0.07^{***}	-0.07^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
size	0.06^{*}				
	(0.02)				
size1		0.06^{*}			
		(0.02)			
size2			0.06^{*}		
			(0.02)		
log(size1)				0.15^{*}	
				(0.07)	
as.numeric(size $>$ 2)					0.07^{*}
					(0.03)
AIC	14544.68	14544.68	14544.68	14545.34	14545.33
Log Likelihood	-7264.34	-7264.34	-7264.34	-7264.67	-7264.66
Num. obs.	4654	4654	4654	4654	4654

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Conditional Logit: Distance and car type

	Model 1	Model 2	Model 3	Model 4
2:(intercept)	-1.19^{***}	-1.19^{***}	-1.19^{***}	-1.19^{***}
	(0.07)	(0.07)	(0.07)	(0.07)
3:(intercept)	-0.56^{***}	-0.59^{***}	-0.56^{***}	-0.59^{***}
	(0.09)	(0.09)	(0.09)	(0.09)
4:(intercept)	-1.91^{***}	-1.94^{***}	-1.91^{***}	-1.94^{***}
	(0.10)	(0.11)	(0.10)	(0.11)
5:(intercept)	-1.05^{***}	-1.08^{***}	-1.05^{***}	-1.09^{***}
	(0.15)	(0.15)	(0.15)	(0.15)
6:(intercept)	-2.64^{***}	-2.67^{***}	-2.64^{***}	-2.68^{***}
	(0.16)	(0.16)	(0.16)	(0.16)
fuelmethanol	-1.30^{***}	-1.34^{***}	-1.30^{***}	-1.35^{***}
	(0.18)	(0.18)	(0.18)	(0.18)
fuelcng	-0.67^{***}	-0.68^{***}	-0.67^{***}	-0.69^{***}
	(0.13)	(0.13)	(0.13)	(0.13)
fuelelectric	-0.13	-0.13	-0.13	-0.13
	(0.08)	(0.08)	(0.08)	(0.08)
price	-0.19^{***}	-0.19^{***}	-0.19^{***}	-0.19^{***}
	(0.03)	(0.03)	(0.03)	(0.03)
cost	-0.08^{***}	-0.08^{***}	-0.09^{***}	-0.09^{***}
	(0.01)	(0.01)	(0.01)	(0.01)
range	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}
	(0.00)	(0.00)	(0.00)	(0.00)
I(range * coml5)		-0.00		-0.00
		(0.00)		(0.00)
I(cost * coml5)			0.03	0.03^{*}
			(0.02)	(0.02)
Log Likelihood	-7087.36	-7086.12	-7085.96	-7084.18
Num. obs.	4654	4654	4654	4654

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Conditional Logit: Price and College

	Model 1	Model 2	Model 3
2:(intercept)	-0.98^{***}	-0.98^{***}	-0.98^{***}
	(0.07)	(0.07)	(0.07)
3:(intercept)	0.46^{***}	0.46^{***}	0.46^{***}
	(0.04)	(0.04)	(0.04)
4:(intercept)	-0.69^{***}	-0.69^{***}	-0.69^{***}
	(0.07)	(0.07)	(0.07)
5:(intercept)	0.56^{***}	0.56^{***}	0.56^{***}
	(0.04)	(0.04)	(0.04)
6:(intercept)	-0.77^{***}	-0.77^{***}	-0.77^{***}
	(0.08)	(0.08)	(0.08)
typesportuv	-0.05	-0.05	-0.04
	(0.15)	(0.15)	(0.15)
typesportcar	-0.18	-0.18	-0.18
	(0.16)	(0.16)	(0.16)
typestwagon	-0.57^{***}	-0.57^{***}	-0.57^{***}
	(0.08)	(0.08)	(0.08)
typetruck	-0.49^{***}	-0.49^{***}	-0.49^{***}
	(0.06)	(0.06)	(0.06)
typevan	-0.14^{*}	-0.14^{*}	-0.14^{*}
	(0.07)	(0.07)	(0.07)
price	-0.25^{***}	-0.25^{***}	-0.24^{***}
	(0.05)	(0.05)	(0.05)
cost	-0.07^{***}	-0.07^{***}	-0.05^{***}
	(0.01)	(0.01)	(0.01)
I(price * college)	0.08	0.08	0.07
	(0.06)	(0.06)	(0.06)
I(cost * college)			-0.03
			(0.02)
AIC	14474.49	14474.49	14474.06
Log Likelihood	-7224.24	-7224.24	-7223.03
Num. obs.	4654	4654	4654

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Interpretation

- Point estimates
- Non invariant characteristics

Multinomial Logit

	Model 1	Model 2
2:(intercept)	-1.09^{***}	-1.09^{***}
	(0.14)	(0.14)
3:(intercept)	0.41^{***}	0.41^{***}
	(0.09)	(0.09)
4:(intercept)	-0.80^{***}	-0.80^{***}
	(0.13)	(0.13)
5:(intercept)	0.65^{***}	0.65^{***}
	(0.09)	(0.09)
6:(intercept)	-0.75^{***}	-0.75^{***}
	(0.13)	(0.13)
2:college	-0.14	-0.04
	(0.16)	(0.20)
3:college	0.01	0.58^{***}
	(0.11)	(0.13)
4:college	-0.17	0.41^{*}
	(0.15)	(0.20)
5:college	-0.17	0.17
	(0.10)	(0.13)
6:college	-0.43^{**}	-0.17
	(0.15)	(0.21)
2:I(college * cost)		-0.02
		(0.03)
3:I(college * cost)		-0.12^{***}
		(0.02)
4:I(college * cost)		-0.12^{***}
		(0.03)
5:I(college * cost)		-0.07^{***}
		(0.02)
6:I(college * cost)		-0.05
		(0.03)
Num. obs.	4654	4654

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Multinomial Logit

	Model 1	Model 2
2:(intercept)	-1.09^{***}	-1.09^{***}
	(0.08)	(0.08)
3:(intercept)	0.46^{***}	0.46^{***}
	(0.05)	(0.05)
4:(intercept)	-0.87^{***}	-0.87^{***}
	(0.08)	(0.08)
5:(intercept)	0.48^{***}	0.48^{***}
	(0.05)	(0.05)
6:(intercept)	-1.07^{***}	-1.07^{***}
	(0.08)	(0.08)
2:coml5	-0.32^{*}	-0.84^{*}
	(0.15)	(0.43)
3:coml5	-0.11	-0.34
	(0.09)	(0.19)
4:coml5	-0.18	-0.23
	(0.13)	(0.32)
5:coml5	0.11	-0.11
	(0.09)	(0.18)
6:coml5	-0.00	-0.37
	(0.14)	(0.35)
2:I(coml5 * size)		0.21
		(0.16)
3:I(coml5 * size)		0.10
		(0.07)
4:I(coml5 * size)		0.02
		(0.12)
5:I(coml5 * size)		0.09
		(0.07)
6:I(coml5 * size)		0.16
		(0.13)
Num. obs.	4654	4654

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Mixed Logit

	Model 1	Model 2	Model 3	Model 4
price	-0.19^{***}	-0.19^{***}	-0.19^{***}	-0.19^{***}
	(0.03)	(0.03)	(0.03)	(0.03)
cost	-0.07^{***}	-0.07^{***}	-0.07^{***}	-0.07^{***}
	(0.01)	(0.01)	(0.01)	(0.01)
range	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}
	(0.00)	(0.00)	(0.00)	(0.00)
acc	-0.06^{***}	-0.06^{***}	-0.06^{***}	-0.06^{***}
	(0.01)	(0.01)	(0.01)	(0.01)
2:hsg2		0.48^{**}	0.47^{**}	0.46^{**}
		(0.16)	(0.16)	(0.16)
3:hsg2		-0.08	-0.09	-0.09
		(0.11)	(0.11)	(0.11)
4:hsg2		0.59^{***}	0.59^{***}	0.57^{***}
		(0.14)	(0.14)	(0.14)
5:hsg2		-0.02	-0.02	-0.04
		(0.11)	(0.11)	(0.11)
6:hsg2		0.56^{***}	0.56^{***}	0.52^{***}
		(0.15)	(0.15)	(0.15)
speed			0.00^{***}	0.00^{***}
			(0.00)	(0.00)
pollution			-0.03	-0.03
			(0.08)	(0.08)
2:coml5			-0.31^{*}	-0.31^{*}
			(0.15)	(0.15)
3:coml5			-0.16	-0.17
			(0.09)	(0.09)
4:coml5			-0.20	-0.21
			(0.14)	(0.14)
5:coml5			0.04	0.03
			(0.09)	(0.09)
6:coml5			-0.04	-0.08
			(0.14)	(0.14)
2:college				-0.12
				(0.17)

Introduction

Suppose y takes integers 0,1,2,. . . , which are cardinal not just ordinal. Then y is a count response.

- Number of accidents
- Number of live births
- Doctor visits
- Prescription drugs
- Recreational trips

Examples

- Number of children over a specified age interval of the mother, depending on mother's schooling, age, household income
- Visits to a museum to characterize the impact of specific attractions (animals, science room)
- Airline safety (number of accidents) to airline profitability

Motivation

- Linear regression on count data
- Discrete choice methods (ordered, multinomial)

Table 20.1. *Proportion of Zero Counts in Selected Empirical Studies*

Study	Variable	Sample Size	Proportion of Zeros
Cameron et al. (1988)	Doctor visits	5,190	0.798
Pohlmeier and Ulrich (1995)	Specialist visits	5,096	0.678
Grootendorst (1995)	Prescription drugs	5,743	0.224
Deb and Trivedi (1997)	Number of hospital stays	4,406	0.806
Gurmu and Trivedi (1996)	Recreational trips	659	0.632
Geil et al. (1997)	Hospitalizations	30,590	0.899
Greene (1997)	Major derogatory reports	1,319	0.803

Poisson Models

- The natural choice is a Poisson Model

$$Pr(Y) = \frac{\exp(-\mu)\mu^y}{y!}$$

  where μ is the mean and the variance of Y.

- In Poisson MLE, we assume that $y \mid x$ follows a poisson distribution with parameter $\lambda(x) > 0$, and we have

$$P(y = r \mid x) = \frac{\lambda(x)^r}{r!} \exp(-\lambda(x))$$

- We specify

$$\lambda(x) = \exp(x_i'\beta) \tag{40}$$

Likelihood

- The likelihood is given by

$$\log \mathcal{L}(\beta) = \sum_i y_i x_i' \beta - \exp(x_i'\beta) - \log y_i! \tag{41}$$

- FOC

$$\frac{\partial \mathcal{L}(\beta)}{\partial \beta} = \sum_i \{y_i - \exp(x_i'\beta)\} x_i \tag{42}$$

- SOC

$$\frac{\partial^2 \mathcal{L}(\beta)}{\partial \beta \partial \beta'} = \sum_i [-\exp(x_i'\beta)] x_i x_i' \tag{43}$$

Marginal Effects

- Standard solution

$$\frac{\partial E(y \mid x)}{\partial x} = \beta \exp(x\beta) \tag{44}$$

- Better marginal effect: Proportional change in $E(y \mid x)$ as x changes by one unit.

$$\frac{\partial E(y \mid x)/\partial x}{E(y \mid x)} = \beta \tag{45}$$

Over-dispersion problem

- The Poisson Model implies that $E(y \mid x) = Var(y \mid x)$.
- Too restrictive, and leads to a problem of excess zeroes.
- Explanations

Negative Binomial Model

- The negative binomial distribution is a mixture between the Poisson and Gamma distributions.
- Parametrization $\lambda_i \equiv \exp(x_i'\beta)$ and $\psi_i \equiv \frac{1}{\alpha}\lambda_i^{\kappa}$ where β, $\alpha > 0$ and κ are parameters.
- Choice probabilities are given by

$$P(y_i \mid x_i) = \frac{\Gamma(y_i + \psi_i)}{\Gamma(\psi_i)\Gamma(y_i + 1)} \left(\frac{\psi_i}{\lambda_i + \psi_i}\right)^{\psi_i} \left(\frac{\psi_i}{\lambda_i + \psi_i}\right)^{y_i} \tag{46}$$

- The moments are given

$$\begin{aligned} E(y \mid x_i) &= \lambda_i \\ Var(y \mid x_i) &= \lambda_i + \alpha\lambda_i^{2-\kappa} \end{aligned}$$

Zero Inflated Model

- Imagine, two densities $f_1(\cdot)$ and $f_2(\cdot)$.

$$g(y) = \begin{cases} f_1(0) + (1 - f_1(0))f_2(0), & \text{if } y = 0 \\ (1 - f_1(0))f_2(y), & \text{if } y \geq 1 \end{cases} \tag{47}$$

- The likelihood can be derived, and the model is able to cope with many zeros.

Application: Crime and Inequality

- Among industrialized economies, the United States enjoys two unenviable distinctions: high inequality and high rates of crime, particularly of violent crime.
- Individuals allocate time between market and criminal activity by comparing the expected return from each, and taking account of the likelihood and severity of punishment. In these models, inequality leads to crime by placing low-income individuals who have low returns from market activity in proximity to high-income individuals who have things that are worth "taking".

Data

- Data are taken from the 1991 FBI Uniform Crime Reports, which comprise violent crimes and property crimes. Violent crime consists of murder and non-negligent homicide, forcible rape, robbery, and aggravated assault.

Results

TABLE 3.—DETERMINANTS OF CRIME: EXOGENOUS POLICE EXPENDITURE

	Income Inequality				Educational Inequality			
	All Counties		Largest 200		All Counties		Largest 200	
	Violent	Property	Violent	Property	Violent	Property	Violent	Property
Intercept	−11.7507	−0.8078	−12.6106	−0.3304	−11.476	0.1399	−12.3127	0.1865
	(0.7396)	(0.576)	(1.5291)	(1.2082)	(0.7758)	(0.5958)	(1.6045)	(1.2291)
Population	1.2329	0.9568	1.2435	0.9451	1.255	0.9529	1.2659	0.9414
	(0.0137)	(0.0107)	(0.0281)	(0.0225)	(0.0138)	(0.0105)	(0.0283)	(0.0221)
Density	0.0582	0.0362	0.0687	0.0372	0.0245	0.0399	0.0339	0.0404
	(0.0168)	(0.0122)	(0.0344)	(0.0246)	(0.0165)	(0.012)	(0.0338)	(0.0243)
Gini	1.3346	−0.1535	1.3846	−0.1547	1.1906	−0.4168	1.2204	−0.2974
	(0.1509)	(0.1174)	(0.3074)	(0.2378)	(0.1583)	(0.123)	(0.3229)	(0.2459)
Female head	1.5751	0.7923	1.6414	0.7222	1.7032	0.7489	1.7708	0.6867
	(0.0942)	(0.0709)	(0.1926)	(0.1464)	(0.0952)	(0.0709)	(0.1951)	(0.1471)
Nonwhite	0.0568	−0.0305	0.0414	−0.0384	−0.0148	0.0004	−0.0342	−0.015
	(0.0391)	(0.0251)	(0.0808)	(0.0539)	(0.0393)	(0.0268)	(0.0812)	(0.0575)
Unemployed	−0.0489	0.0158	−0.021	0.0258	0.0505	−0.0234	0.0868	0.0049
	(0.0775)	(0.0599)	(0.159)	(0.1251)	(0.084)	(0.0611)	(0.1731)	(0.1269)
Poverty	−0.216	0.3452	−0.2663	0.3888	−0.0776	0.3703	−0.1169	0.3943
	(0.0666)	(0.0483)	(0.1368)	(0.0988)	(0.0616)	(0.0429)	(0.1266)	(0.0875)
Movers	1.6043	1.3384	1.7587	1.2716	1.7753	1.2299	1.9383	1.1985
	(0.1228)	(0.0879)	(0.2564)	(0.1841)	(0.1342)	(0.0922)	(0.2811)	(0.1929)
Young	−0.9055	−1.0359	−0.9524	−1.0251	−1.1988	−0.9752	−1.259	−0.9707
	(0.1017)	(0.0786)	(0.2079)	(0.166)	(0.0997)	(0.073)	(0.2039)	(0.1543)
College	−0.2974	0.1842	−0.3075	0.1857	−0.186	0.2206	−0.1841	0.2048
	(0.0659)	(0.0475)	(0.1356)	(0.0989)	(0.0629)	(0.0467)	(0.1294)	(0.0971)
Police	−0.0209	−0.1046	−0.0267	−0.1128	−0.0709	−0.0808	−0.077	−0.0973
	(0.0387)	(0.0341)	(0.0781)	(0.0692)	(0.0431)	(0.0345)	(0.0875)	(0.0703)
Null dev.	45.511	78.009	39.869	56.525	45.511	78.009	39.870	56.525
Resid. dev.	8.102	22.493	7.365	19.901	8.298	22.228	7.578	19.792

Logistic regression results for all urban counties and largest 200. Standard errors are in parentheses. All explanatory variables are in logs. Deviances multiplied by 10^{-9}.

Affairs Data

	Numbers
0	451
1	34
2	17
3	19
7	42
12	38

Estimation

	Probit	Linear	Poisson	Negative Bin
(Intercept)	-0.00	6.31^{***}	2.39^{***}	1.45
	(0.75)	(1.85)	(0.30)	(1.65)
as.factor(gender)male	0.17	0.01	0.06	0.01
	(0.13)	(0.30)	(0.08)	(0.27)
as.factor(education)12	-0.13	-1.78	-0.80^{***}	-0.11
	(0.57)	(1.36)	(0.24)	(1.23)
as.factor(education)14	-0.29	-2.68^{*}	-1.35^{***}	-0.73
	(0.54)	(1.29)	(0.23)	(1.17)
as.factor(education)16	-0.44	-2.93^{*}	-1.73^{***}	-1.32
	(0.55)	(1.30)	(0.24)	(1.18)
as.factor(age)22	-0.98	-3.33^{*}	-1.89^{***}	-1.57
	(0.54)	(1.37)	(0.25)	(1.21)
as.factor(age)27	-0.77	-2.90^{*}	-1.19^{***}	-0.85
	(0.54)	(1.36)	(0.22)	(1.20)
as.factor(age)32	-0.56	-2.18	-0.72^{***}	-0.35
	(0.54)	(1.37)	(0.22)	(1.21)
as.factor(age)37	-0.80	-2.08	-0.67^{**}	-0.32
	(0.55)	(1.39)	(0.22)	(1.23)
as.factor(age)42	-0.63	-1.93	-0.63^{**}	-0.27
	(0.56)	(1.41)	(0.23)	(1.25)
as.factor(children)yes	0.33^{*}	0.19	0.14	0.15
	(0.16)	(0.35)	(0.10)	(0.33)
Log Likelihood	-323.16		-1558.84	-737.67

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table: Models

Introduction

- Censoring: Data on y is censored if for part of the range of we observe only that y is in that range, rather than observing the exact value of y. e.g. income is top-coded at \$75,000 per year.
- Truncation: Data on y is truncated if for part of the range of y, we do not observe y at all. e.g. people with income above \$75,000 per year are excluded from the sample.

Introduction

- Validity: Meaningful policy analysis requires extrapolation from the restricted sample to the population as a whole.
- regressions on censored or truncated data, without controlling for censoring or truncation, leads to inconsistent parameter estimates.

Examples

- A research project is studying the level of lead in home drinking water as a function of the age of a house and family income. The water testing kit cannot detect lead concentrations below 5 parts per billion (ppb). The EPA considers levels above 15 ppb to be dangerous.
- Consider the situation in which we have a measure of academic aptitude (scaled 200-800) which we want to model using reading and math test scores, as well as, the type of program the student is enrolled in (academic, general, or vocational). The problem here is that students who answer all questions on the academic aptitude test correctly receive a score of 800, even though it is likely that these students are not truly equal in aptitude. The same is true of students who answer all of the questions incorrectly. All such students would have a score of 200, although they may not all be of equal aptitude.

Example

- Marketing program for Hair Volume
- Survey data
- Problems

Introduction

- Focus on the normal distribution
- Limited dependent models

Tobit Model

- Consider a latent variable $y^\star$ given by

$$y^\star = x\beta + \epsilon \tag{48}$$

- $y^\star$ is observed only partially.
- Censoring

$$y = \begin{cases} y^\star, & \text{if } y^\star > 0 \\ 0, & \text{if } y^\star \leq 0 \end{cases} \tag{49}$$

- Truncation

$$y = y^\star \text{ if } y^\star > 0 \tag{50}$$

Assumptions

- The standard estimators require stochastic assumptions about the distribution of y and ϵ.
- Normality:

$$\epsilon \sim N(0, \sigma^2) \rightarrow y^\star \sim N(x\beta, \sigma^2) \tag{51}$$

Simulations

Consider

$$y^{\star} = -2500 + 1000 lnw + \epsilon \quad (52)$$

$$\epsilon \sim \mathbb{N}(0, 10^2) \quad (53)$$

$$lnw \sim \mathbb{N}(2.75, 0.60^2) \quad (54)$$

Results (1)

	Uncensored	Censored	Truncated
(Intercept)	-2507.46^{***}	-1691.52^{***}	-2321.20^{***}
	(24.04)	(29.64)	(39.85)
lw	1002.74^{***}	727.60^{***}	942.02^{***}
	(8.69)	(10.71)	(13.69)
R^2	0.93	0.82	0.86
Adj. R^2	0.93	0.82	0.86
Num. obs.	1000	1000	744
RMSE	99.65	122.87	98.25

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table: Statistical models

Setting 2

Consider

$$y^\star = -2500 + 1000lnw + \epsilon \tag{55}$$

$$\epsilon \sim \mathbb{N}(0,1) \tag{56}$$

$$lnw \sim \mathbb{N}(2.75, 0.60^2) \tag{57}$$

Results (2)

	Uncensored	Censored	Truncated
(Intercept)	-2500.63^{***}	-1811.03^{***}	-2500.58^{***}
	(0.25)	(21.33)	(0.40)
lw	1000.24^{***}	768.07^{***}	1000.23^{***}
	(0.09)	(7.68)	(0.14)
R^2	1.00	0.91	1.00
Adj. R^2	1.00	0.91	1.00
Num. obs.	1000	1000	768
RMSE	1.01	86.63	1.01

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table: Statistical models

Setting 3

Consider

$$
\begin{aligned}
y^\star &= 0.2 + 10lnw + \epsilon && (58) \\
\epsilon &\sim \mathbb{N}(0, 1) && (59) \\
lnw &\sim \mathbb{N}(2.75, 0.60^2) && (60)
\end{aligned}
$$

Results(3)

	Uncensored	Censored	Truncated
(Intercept)	0.36	0.36	0.36
	(0.24)	(0.24)	(0.24)
lw	9.95***	9.95***	9.95***
	(0.09)	(0.09)	(0.09)
R^2	0.93	0.93	0.93
Adj. R^2	0.93	0.93	0.93
Num. obs.	1000	1000	1000
RMSE	1.00	1.00	1.00

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table: Statistical models

Setting 4

$$y^\star = -25 + 10 lnw + \epsilon \tag{61}$$

$$\epsilon \sim \mathbb{N}(0, 10) \tag{62}$$

$$lnw \sim \mathbb{N}(2.75, 0.60^2) \tag{63}$$

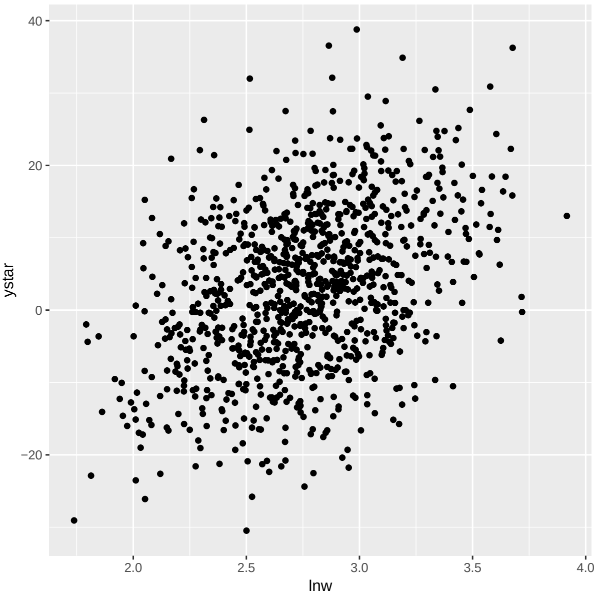
Results (4)

	Uncensored	Censored	Truncated
(Intercept)	-31.32^{***}	-14.74^{***}	-6.67^{**}
	(2.48)	(1.69)	(2.35)
lw	12.45^{***}	7.56^{***}	5.88^{***}
	(0.89)	(0.61)	(0.82)
R^2	0.16	0.13	0.08
Adj. R^2	0.16	0.13	0.08
Num. obs.	1000	1000	604
RMSE	10.06	6.84	6.74

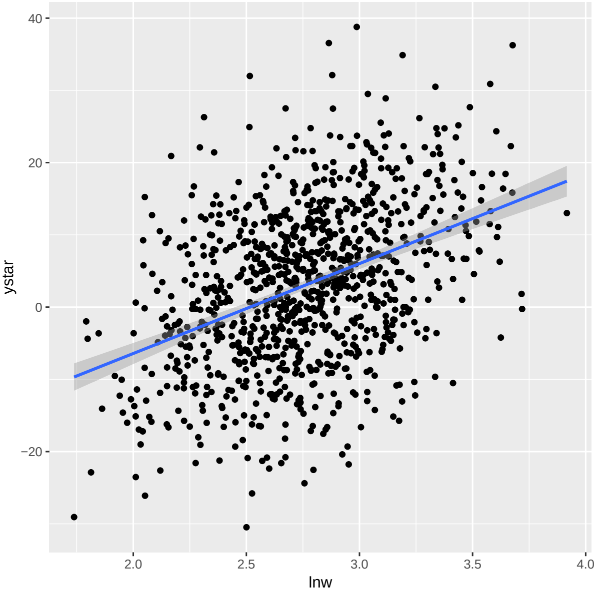
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table: Statistical models

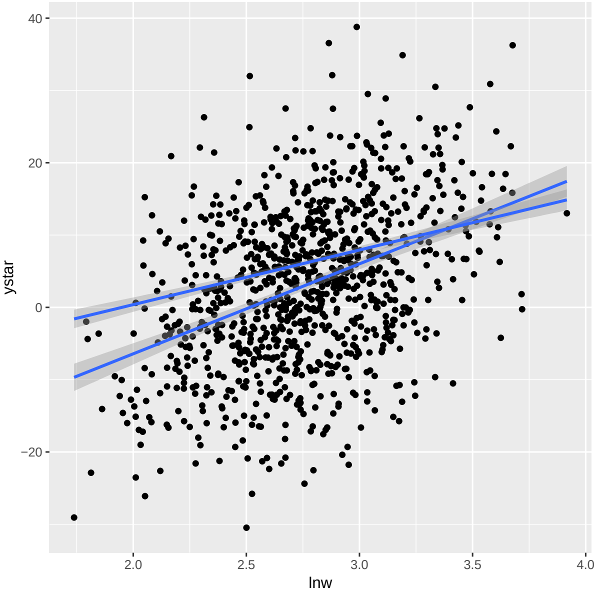
Describing the data (1)



Describing the data (2)- Uncensored



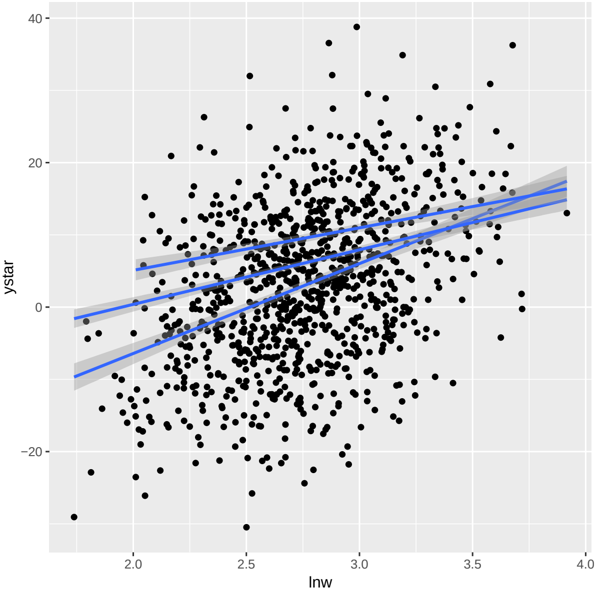
Describing the data (3) - Censored



Findings

- Negative values have been increased to zero.
- Increase the intercept and flattens the regression line

Describing the data (4) - All



Findings

- Observations with negative error draws dropped more than those with positive draws
- The mean of the error is shifted up for low hours.
- Increase the intercept and flattens the regression line

Truncated case: Conditional Mean

- If the data is truncated, we observe y only when $y > 0$
- We need to derive the truncated mean

$$
\begin{aligned}
E(y \mid y > 0) &= E(x\beta + \epsilon \mid x\beta + \epsilon > 0) \\
&= x\beta + E(\epsilon \mid \epsilon > -x\beta) \\
&= x\beta + \sigma E(\tfrac{\epsilon}{\sigma} \mid \tfrac{\epsilon}{\sigma} > -x\tfrac{\beta}{\sigma}) \\
&= x\beta + \sigma \frac{\phi(x\frac{\beta}{\sigma})}{\Phi(x\frac{\beta}{\sigma})}
\end{aligned}
$$

- $\dfrac{\phi(x\frac{\beta}{\sigma})}{\Phi(x\frac{\beta}{\sigma})}$ is the inverse Mills ratio.
- The regression function is nonlinear i.e OLS is inconsistent

Censored Case: Conditional Mean

- For censored data, $y = 0$ if $y^\star < 0$ and $y = y^\star$ otherwise
- The censored sample mean is

$$
\begin{aligned}
E[y] &= E_{y^\star}[E[y \mid y^\star]] \\
&= Pr(y^\star < 0) * 0 + Pr(y^\star > 0) * E(y^\star \mid y^\star > 0) \\
&= \Phi(x\beta) \left\{ x\beta + \sigma \frac{\phi(x\frac{\beta}{\sigma})}{\Phi(x\frac{\beta}{\sigma})} \right\} \\
&= \Phi(x\beta)x\beta + \sigma\phi(x\frac{\beta}{\sigma}))
\end{aligned}
$$

- Which is nonlinear as well.

Maximum Likelihood

- Let $y^\star$ have density $f^\star(y^\star)$ and cdf $F(y^\star)$.
- Consider censored $y = max(y^\star, 0)$
- The density of $y^\star$ is
  - $y > 0$; then $y = y^\star$ so $f(y) = f^\star(y)$
  - $y = 0$; then $y^\star < 0$ so $f(0) = F^\star(0)$
- Define indicator

$$d = \begin{cases} 1, & \text{If } y > 0 \\ 0, & \text{if } y = 0 \end{cases}$$

Likelihood

- The density is given $f(y) = f^\star(y)^d \times F^\star(0)^{1-d}$
- The log likelihood is given

$$\log L = \sum_i d_i \log f^\star(y_i, x_i, \theta) + (1 - d_i) \log F^\star(0, x_i, \theta)$$

which is given by

$$\log L(\beta, \sigma^2) = \sum_i d_i \log \tfrac{1}{\sigma} \phi\left(\frac{y_i - x_i\beta}{\sigma}\right) + (1 - d_i) \log \Phi(-x_i\beta/\sigma)$$

Truncated Case: Conditional Mean

- For truncated data only $y^\star > 0$ is observed.
- The conditional density is

$$f^\star(y^\star \mid y^\star > 0) = \frac{f^\star(y^\star)}{Pr(y^\star \mid y^\star > 0)} \tag{64}$$

- The log likelihood is given

$$\log L = \sum_i \log f^\star(y_i, x_i, \theta) - \log F^\star(0, x_i, \theta)$$

which is given by

$$\log L(\beta, \sigma^2) = \sum_i \log \tfrac{1}{\sigma}\phi\left(\frac{y_i - x_i\beta}{\sigma}\right) - \log \Phi(x_i\beta/\sigma)$$

Estimation

- Standard ML applies...
- NLS on the correct censored or truncated mean (needs to control for Heteroskedacticity).

Heckman Two-Step Estimator: Motivation

- Heckman (1976,1979) proposed estimating the censored model using an ingenious two step estimator.
- Recall for positive y, we have

$$E(y \mid y > 0) = x\beta + \sigma f(-x\beta/\sigma) \tag{65}$$

- Include $f(-x\beta/\sigma)$ as a regressor.

Heckman Two-Step Estimator: Application

- Using censored data, estimate a probit model for whether $y_i > 0$, and calculate $\dfrac{\phi(x\frac{\beta}{\sigma})}{\Phi(x\frac{\beta}{\sigma})}$
- Estimate β and σ from OLS

$$y_i = x_i\beta + \sigma\frac{\phi(x\frac{\beta_0}{\sigma_0})}{\Phi(x\frac{\beta_0}{\sigma_0})} + \varepsilon \tag{66}$$

Pros and Cons

- Consistent estimates
- Allows for weaker assumptions
- Standard Errors are incorrect

Sample Selectivity Model

- Most common generalization of the standard tobit model is the sample selection or self-selection model.
- This is a two-part model
- A latent variable that determines the selection.
- The variable of interest $y_2^\star$ is of interest.
- Major complication comes from unobserved component affecting both processes.

Identifying the causal relation between education and income - From model to estimates

$$\ln w = 7.58 + \underset{(43.8)}{0.070}s \qquad R^2 = 0.067$$

$$\ln w = 6.20 + \underset{(72.3)}{0.107}s + \underset{(75.5)}{0.081}x - \underset{(-55.8)}{0.0012}x^2 \qquad R^2 = 0.285$$

$$\ln w = 4.87 + \underset{(23.4)}{0.255}s - \underset{(-7.1)}{0.0029}s^2 + \underset{(63.7)}{0.148}x - \underset{(-66.2)}{0.0018}x^2 - \underset{(-31.8)}{0.0043}xs \qquad R^2 = 0.309$$

Table: Estimates of wage equations; t designates the duration of schooling, x experience (measured by age minus the duration of schooling minus 6 years), and y the annual earnings of white men working in the non-agricultural sector in the United States in 1959 (t-statistics in parentheses). Source: Mincer (1974)

The theory of human capital (1)

- The aim of estimating earnings functions is to evaluate the returns to education
- This correlation betrays a *causal link* between education and earnings
- Through estimation, it is possible to see the role of education:
  - As a signaling device (enabling employers to select the most efficient workers)
  - As a vector of human capital accumulation
- The main prediction of the theory of human capital is that education is the source of an accumulation of competences, which increases earnings

Identifying the causal relation between education and income - Selection problem

- The correlation between duration of schooling and income does not mean *causal* relation
- The correlation can stem from the fact that the most efficient individuals have higher earnings and stay in school longer *(selection problem)*
- The theory of human capital and signaling theory both predict that the most productive individuals have an interest in studying for the longest period, entailing the possibility of the so called *ability bias*

Identifying the causal relation between education and income - Selection problem

- The relation between the wage and the number of years of education is given by $\ln w(s) = \ln w(0) + \rho s$
- The population regression coefficient ρ minimizes the expected squared errors in the population as a whole, $E(w - \rho s)^2$, where $W = \ln w$. This coefficient is given by:

$$\rho_{OLS} = \frac{Cov(s, W)}{Var(s)}$$

Illustration - Angrist and Krueger (1991)

- Angrist and Krueger (1991) noted that individuals born early in the calendar year have shorter durations of schooling than those born later
- This effect is due to the compulsory duration of schooling where compulsory schooling laws require students to remain in school until their 16th or 17th birthday
- By assuming that the date of one's birth is independent of factors influencing abilities and preferences, this phenomenon can entail an exogenous variation in the duration of schooling, which may be used as an instrument

Illustration - Angrist and Krueger (1991)

- To account for age-related trends in earnings, they estimate the following two-stage least squares model:

$$
\begin{aligned}
S_i &= \boldsymbol{X\pi} + \sum_c \delta_c Y_{ic} + \sum_c \sum_j \theta_{jc} Y_{ic} Q_{ij} + \eta_i \\
w_i &= \boldsymbol{X\beta} + \sum_c \zeta_c Y_{ic} + \rho S_i + \varepsilon_i
\end{aligned}
$$

  - S_i represents the years of education of individual i
  - X_i is a vector of covariates
  - Q_{ij} is a dummy variable for individual i born in quarter j
  - Y_{ic} is a dummy variable for individual i born in year c
  - w_i is weekly log-wage and ρ characterizes the returns to education

Illustration - Angrist and Krueger (1991)

	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS
Years of education	0.0711 (0.0003)	0.0891 (0.0161)	0.0632 (0.0003)	0.0600 (0.0299)
Race (1=black)	–	–	−0.2575 (0.0040)	−0.2676 (0.0458)
SMSA (1=city center)	–	–	0.1763 (0.0029)	0.1797 (0.0305)
Married (1=married)	–	–	0.2479 (0.0032)	0.2486 (0.0073)
8 region of residence dummies	–	–	yes	yes
Age	–	–	−0.0760 (0.0604)	−0.0741 (0.0626)
Age-squared	–	–	0.0008 (0.0007)	0.0007 (0.0007)
Sargan overidentification test: p-value	–	0.6553		0.8798
Number of observations	329,509	329,509	329,509	329,509

Table: OLS and 2SLS estimates of the returns to education for men born 1930-1939: Census 1980. Year of birth dummies are included in all regressions. Standard errors in parentheses.
Source: Angrist and Krueger dataset (1991).

Education and income - Selection problem

- Within Estimators using Siblings or Twins

$$\begin{aligned} Y_{ij} &= \beta S_{ij} + \gamma A_j + \epsilon_{ij} \\ S_{ij} &= \eta A_j + \varepsilon_{ij} \end{aligned}$$

- Eliminate unobserved family ability

$$\Delta Y = \beta \Delta S + \Delta u$$

Education and income (13) - Selection problem

- Behrman, Hrubec and Taubman (1980)
- Ashenfelter and Krueger (1994)
- Behrman and Rosenzweig (1999)
- Assumption: ability is a family specific endowment (same for all children). Twins are identical in their genetic inheritance at conception.
- Unobserved factors that affect the school decision are uncorelated that affect earnings.

Outcomes

TABLE 3—ORDINARY LEAST-SQUARES (OLS), GENERALIZED LEAST-SQUARES (GLS), INSTRUMENTAL-VARIABLES (IV), AND FIXED-EFFECTS ESTIMATES OF LOG WAGE EQUATIONS FOR IDENTICAL TWINS[a]

Variable	OLS (i)	GLS (ii)	GLS (iii)	IV[a] (iv)	First difference (v)	First difference by IV (vi)
Own education	0.084 (0.014)	0.087 (0.015)	0.088 (0.015)	0.116 (0.030)	0.092 (0.024)	0.167 (0.043)
Sibling's education	—	—	−0.007 (0.015)	−0.037 (0.029)	—	—
Age	0.088 (0.019)	0.090 (0.023)	0.090 (0.023)	0.088 (0.019)	—	—
Age squared (÷100)	−0.087 (0.023)	−0.089 (0.028)	−0.090 (0.029)	−0.087 (0.024)	—	—
Male	0.204 (0.063)	0.204 (0.077)	0.206 (0.077)	0.206 (0.064)	—	—
White	−0.410 (0.127)	−0.417 (0.143)	−0.424 (0.144)	−0.428 (0.128)	—	—
Sample size:	298	298	298	298	149	149
R^2:	0.260	0.219	0.219	—	0.092	—

Notes: Each equation also includes an intercept term. Numbers in parentheses are estimated standard errors.

[a]Own education and sibling's education are instrumented for using each sibling's report of the other sibling's education as instruments.

Education and income (14) - Selection problem

- Ashenfelter and Rouse (1998) find that the differences in the returns to education between genetically identical individuals are slightly weaker than those obtained by comparing the duration of schooling and incomes of any two random individuals
- Oreopoulos and Salvanes (2011) have used Norwegian administrative records that supply information on the educational and professional trajectories of all persons born since 1920. They found that siblings with one more year of schooling have more annual income than their less educated siblings, which confirms the results of Ashenfelter and Rouse

Issues Days

- Completely Voluntary - Bonus in grade (+)
- One topic (for example The effects of immigration on natives)
  - Introduction
  - Paper 1 - What
  - Paper 1 - Findings
  - Paper 2 - What
  - Paper 2 - Findings
  - Paper 3 - What
  - Paper 3 - Findings
  - Conclusion
- Class discussion.