15. Multinomial Outcomes

A. Colin Cameron Pravin K. Trivedi

These slides were prepared in 1999.
They cover material similar to Sections 15.3-15.6 of our subsequent book *Microeconometrics: Methods and Applications*, Cambridge University Press, 2005.

INTRODUCTION

- Consider data on several discrete outcomes, usually mutually exclusive.
- Examples:
  - Transportation: several ways to commute to work
  - Labor: employment status is be full-time, part-time or no work.

OUTLINE

- General results for MLE of all multinomial/multivariate models.
- Specific multinomial models
  - multinomial logit
  - random parameters multinomial logit
  - nested logit
  - multinomial probit
  - ordered logit and probit
  - sequential models

- Multivariate models such as bivariate probit
- Simultaneous equations

GNERAL RESULTS

- There are $m + 1$ mutually-exclusive alternatives.
- The dependent variable y takes value j if the j^{th} alternative is taken, $j = 0, ..., m + 1$.
- Define the probability that choose alternative j

$$p_j = \Pr[y = j], \qquad j = 0, ..., m.$$

- Introduce $(m+1)$ binary variables for each observed y

$$y_j = \begin{cases} 1 \text{ if } y = j \\ 0 \text{ if } y \neq j. \end{cases}.$$

- Thus y_j equals 1 if alternative j is chosen and equals 0 for all other non-chosen alternatives, so for an individual exactly one of $y_0, y_1, ..., y_m$ will be non-zero.
- The density for one observation can then be conveniently written as

$$f(y) = p_0^{y_0} \times p_1^{y_1} \times ... \times p_m^{y_m} = \prod_{j=0}^{m} p_j^{y}.$$

- The likelihood function for a sample of size n is then $\mathcal{L} = \prod_{i=1}^{n} \prod_{j=0}^{m} p_{ij}^{y_{ij}}$
- The log-likelihood function is

$$\mathcal{L} = \ln \mathsf{L} = \sum_{i=1}^{n} \sum_{j=0}^{m} y_{ij} \ln p_{ij}.$$

- All that is needed is parameterization of p_{ij} in terms of observed data $\mathbf{x}_{ij}$ and a finite number of parameters $\boldsymbol{\beta}_j$, that is

$$p_{ij} = \Pr[y_i = j] = F_j(\mathbf{x}_{ij}, \boldsymbol{\beta}_j), \qquad j = 0, ..., m.$$

- These probabilities should lie between 0 and 1 and sum over j to one.

- Then the MLE for $\boldsymbol{\beta} = (\boldsymbol{\beta}_0', ..., \boldsymbol{\beta}_m')'$ maximizes

$$\mathcal{L} = \sum_{i=1}^{n} \sum_{j=0}^{m} y_{ij} \ln F_j(\mathbf{x}_{ij}, \boldsymbol{\beta}_j).$$

The first-order conditions are

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}_j} = \sum_{i=1}^{n} \sum_{j=0}^{m} \frac{y_{ij}}{F_j(\mathbf{x}_{ij}, \boldsymbol{\beta}_j)} \frac{\partial F_j(\mathbf{x}_{ij}, \boldsymbol{\beta}_j)}{\partial \boldsymbol{\beta}_j}, \qquad j = 0, ..., m,$$

- By the usual asymptotic theory

$$\widehat{\boldsymbol{\beta}} \overset{a}{\sim} \mathsf{N}\left[\boldsymbol{\beta}_0, -\left(\mathsf{E}\left[\left.\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}_0}\right]\right)^{-1}\right]$$

  if the dgp is correctly specified.
- The distribution is necessarily multinomial so correct specification of the dgp, as for binary outcome models, means correct specification of the functional forms $F_j(\mathbf{x}_{ij}, \boldsymbol{\beta}_j)$ for the probabilities.
- There are a number of ways to parameterize F_j. These different ways correspond to specific models.

- An important distinction should be made between
  - 1. Alternative-specific regressors, such as travel costs in a model of transportation mode choice. Model identification requires that the parameters be constant across alternatives, i.e. $\boldsymbol{\beta}_j = \boldsymbol{\beta}$.
  - 2. Alternative-invariant regressors, such as individual socio-economic characteristics in a model of transportation mode choice. These parameters $\boldsymbol{\beta}_j$ vary across alternatives. Common in economics.

MULTINOMIAL LOGIT: CASE 1

- When regressors do not vary over choices the *multinomial logit* (MNL) model specifies

  $$p_{ij} = \frac{e^{\mathbf{x}_i'\boldsymbol{\beta}_j}}{\sum_{k=0}^{m} e^{\mathbf{x}_i'\boldsymbol{\beta}_k}}, \qquad j = 0, ..., m,$$

  where $\beta_0 = 0$ is the usual restriction made to ensure model identification.

- Clearly these probabilities lie between 0 and 1 and sum over j to one.

- The parameters $\beta_1, ..., \beta_m$ are estimated by MLE which maximizes above with $F_j = p_{ij}$.

- If probabilities are correctly specified the MLE has the asymptotic distribution where the information matrix has jk^{th} block

$$\mathsf{E}\left[\frac{\partial^2\mathcal{L}}{\partial\boldsymbol{\beta}_j\partial\boldsymbol{\beta}_k'}\right] = \sum_{i=1}^{n} p_{ij}(\delta_{jk} - \mathbf{p}_{ik})\mathbf{x}_i\mathbf{x}_i{}',$$

$$j = 0,...,m, \ k = 0,...,m,$$

where

$\delta_{jk} = 1$ if $j = k$ and $\delta_{jk} = 0$ if $j \neq k$.

MULTINOMIAL LOGIT: CASE 2

- When instead regressors do vary over choices, the MNL probabilities are

$$p_{ij} = \frac{e^{\mathbf{x}_{ij}'\boldsymbol{\beta}}}{\sum_{k=0}^{m} e^{\mathbf{x}_{ij}'\boldsymbol{\beta}}}, \qquad j = 0, ..., m,$$

- It can be shown that

$$\widehat{\boldsymbol{\beta}}_{ML} \overset{a}{\sim} \mathsf{N}\left[\boldsymbol{\beta}, \left(\sum_{i=1}^{n}\sum_{j=0}^{m} p_{ij}(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)'\right)^{-1}\right],$$

  where $\bar{\mathbf{x}}_i = \sum_{k=0}^{m} p_{ik}\mathbf{x}_{ik}$ is a weighted average of the regressors over alternatives.

MULTINOMIAL LOGIT: COMPARISON

- The first formulation is often used in labor economics. For example, for choice of occupation all individual specific regressors, such as education, age and gender, are invariant across alternatives.
- The second formulation is more commonly used in transportation mode choice.

  Then data is available on mode attributes such as price and time which vary over both individuals and alternatives. This formulation is sometimes called the conditional logit model.

- Such studies will also include individual characteristics that do not vary across alternatives. These can also be incorporated, leading to what some authors call a mixed model.

MULTINOMIAL LOGIT: COMPARISON

- The first mdel can in fact can be re-expressed as the second.
- Suppose $\mathbf{x}_i$ is $k \times 1$ and define $\mathbf{x}_{ij}$ to be $k(m+1) \times 1$ vector with zeros everywhere except that the $(j+1)^{th}$ block is $\mathbf{x}_i$, that is,

  $$\mathbf{x}_{ij} = [\mathbf{0}' \cdots \mathbf{0}' \quad \mathbf{x}_i' \quad \mathbf{0}' \cdots \mathbf{0}']',$$

  and define $\boldsymbol{\beta} = [\boldsymbol{\beta}_0' \quad \cdots \quad \boldsymbol{\beta}_m']'$.
- Then $\mathbf{x}_{ij}'\boldsymbol{\beta} = \mathbf{x}_i'\boldsymbol{\beta}_j$ or $\mathbf{x}_i'\boldsymbol{\beta}_j = \mathbf{x}_{ij}'\boldsymbol{\beta}$.
- This result can be used, for example, to rewrite a mixed

model as a conditional logit model.

- An obvious generalization of the multinomial logit model is

$$p_{ij} = \frac{\mu_{ij}}{\sum_{k=0}^{m} \mu_{ij}}, \qquad j = 0, ..., m,$$

where $\mu_{ij} > 0$ can be quite general functions of regressors $\mathbf{x}_i$ and parameters $\boldsymbol{\beta}$.

INDEPENDENCE OF IRRELEVANT ALTERNATIVES

- A limitation of the multinomial logit model is the assumption of independence of irrelevant alternatives (IIA).
- The multinomial logit probabilities imply that the conditional probability of observing alternative j given that either alternative j or alternative k is observed is

  $$\Pr[y = j | y = j \text{ or } k] = \frac{p_j}{p_j + p_k} = \frac{e^{\mathbf{x}_j' \boldsymbol{\beta}}}{e^{\mathbf{x}_j' \boldsymbol{\beta}} + e^{\mathbf{x}_k' \boldsymbol{\beta}}},$$

  upon some simplification.

- This equals $\exp((\mathbf{x}_j - \mathbf{x}_k)'\boldsymbol{\beta})/[1 + \exp((\mathbf{x}_j - \mathbf{x}_k)'\boldsymbol{\beta})]$, a binary logit model.
- The conditional probability does not depend on other alternatives, a major limitiation.
- As an extreme example, the conditional probability of commute by car given commute by car or red bus is assumed in a MNL model to be independent of whether commuting by blue bus is an option.
- But in practice we expect introduction of a blue bus, same as red bus in every aspect except color to

  - have little impact on car use
  - halve use of blue bus
  - leading to an increase in the conditional probability of car use given car or blue bus.
- This weakness of MNL has led to extensions which are obtained by using the random utility approach.

RANDOM UTILITY MODEL

- The multinomial logit model can be motivated by the following random utility formulation.
- Consider a 3-choice model

$$\begin{aligned} U_0 &= \mu_0 + \varepsilon_0 \\ U_1 &= \mu_1 + \varepsilon_1 \\ U_2 &= \mu_2 + \varepsilon_2, \end{aligned}$$

- where μ_j is deterministic, e.g. $\mu_j = \mathbf{x}'\boldsymbol{\beta}_j$, and the errors ε_j are iid log Weibull (or type I extreme value) distributed with density

$$f(\varepsilon_j) = e^{-\varepsilon_j} \exp(e^{-\varepsilon_j}), \qquad j = 0, 1, 2.$$

- Then

$$\begin{aligned} \Pr[y=2] &= \Pr[U_2 > U_1, \quad U_2 > U_0] \\ &= \Pr[\mu_2 + \varepsilon_2 > \mu_1 + \varepsilon_1, \quad \mu_2 + \varepsilon_2 > \mu_0 + \varepsilon_0] \\ &= \Pr[\varepsilon_1 < \varepsilon_2 + \mu_2 - \mu_1, \quad \varepsilon_0 < \varepsilon_2 + \mu_2 - \mu_0] \\ &= \int_{-\infty}^{\infty} f(\varepsilon_2) \left\{ \int_{-\infty}^{\varepsilon_2+\mu_2-\mu_1} f(\varepsilon_1) \times \left\{ \int_{-\infty}^{\varepsilon_2+\mu_2-\mu_0} f(\varepsilon_0) d\varepsilon_0 \right\} d\varepsilon_1 \right. \end{aligned}$$

- After much manipulation, similar to that in the binary case, this simplifies to

$$\Pr[y=2] = \frac{e^{\mu_2}}{e^{\mu_0} + e^{\mu_1} + e^{\mu_2}},$$

which is the multinomial logit when $\mu_j = \mathbf{x}'\boldsymbol{\beta}_j$ or

$$\mu_j = \mathbf{x}_j' \boldsymbol{\beta}.$$

- A weakness of the multinomial logit model is that the errors ε_j are assumed to be independent across j.
- This is certain to be violated if two alternatives are similar.
- For example, suppose alternatives 1 and 2 are similar.
- A low draw of ε_1 leads to overprediction of the utility of alternative 1.
- We then also expect to overpredict the utility of alternative 1, i.e. ε_2 is low.
- Since low values of ε_1 and ε_2 tend to go together, and

similarly for high values, the errors must be correlated.

- The "*red bus - blue bus*" problem is an extreme case.
- The models in the remainder of this section and in the next two sections are models that overcome this weakness of the multinomial logit, at the expense of increased computational burden which in some cases is very large and complex.

NESTED LOGIT

- Consider the 3-choice random utility model.
- Suppose alternatives 1 and 2 are similar, and 0 dissimilar.
- For example, 1 is commute by bus (public transit), 2 is commute by train (public transit) and 0 is commute by car.
- Assume

  ε_1 and ε_2 are correlated, with joint distribution Gumbel's Type B bivariate extreme value

ε_0 is independent of the other errors with Type I extreme value distribution.

- Then the cdf's of the errors are

$$\begin{aligned} F(\varepsilon_1, \varepsilon_2) &= \exp(-[e^{-\varepsilon_1/\rho} + e^{-\varepsilon_2/\rho}]^{\rho}) \\ F(\varepsilon_0) &= \exp(-e^{-\varepsilon_0}). \end{aligned}$$

- The parameter ρ should lie between 0 and 1.
- It can be shown after much algebra that for the nested logit model

$$p_0 = \Pr[y = 0] = \frac{e^{\mu_0}}{e^{\mu_0} + (e^{\mu_1/\rho} + e^{\mu_2/\rho})^{\rho}},$$

and

$$\frac{p_1}{p_1 + p_2} = \Pr[y = 1 | y \neq 0] = \frac{e^{\mu_1/\rho}}{(e^{\mu_1/\rho} + e^{\mu_2/\rho})}.$$

- The model for $\Pr[y = 0]$ is logit-like, except we take a weighted average of the similar alternatives. This weighted average

$$v = e^{\mu_1/\rho} + e^{\mu_2/\rho},$$

is called the *inclusive value*, and then

$$p_0 = \frac{e^{\mu_0}}{e^{\mu_0} + e^{\rho \ln v}}.$$

- For choosing between the similar alternatives, the model is a logit model, except for the scale factor ρ.

- From the above, and $p_0 + p_1 + p_2 = 1$ we can solve for p_1 and p_2 to get the density for an observation

$$\begin{aligned}
\ln f(y) &= p_0^{y_0} p_1^{y_1} p_2^{y_2} \\
&= p_0^{y} (p_1 + p_2)^{y_1 + y_2} \left(\frac{p_1}{p_1 + p_2} \right)^{y_1} \left(\frac{p_2}{p_1 + p_2} \right)^{y_2} \\
&= p_0^{y_0} (1 - p_0)^{1 - y_0} \left(\frac{p_1}{p_1 + p_2} \right)^{y_1} \left(\frac{p_2}{p_1 + p_2} \right)^{y_2} \\
&= \left(\frac{e^{\mu_0}}{e^{\mu_0} + e^{\rho \ln v}} \right)^{y_0} \left(\frac{e^{\rho \ln v}}{e^{\mu_0} + e^{\rho \ln v}} \right)^{1 - y_0} \\
&\quad \times \left(\frac{e^{\mu_1/\rho}}{(e^{\mu_1/\rho} + e^{\mu_2/\rho})} \right)^{y_1} \left(\frac{e^{\mu_2/\rho}}{(e^{\mu_1/\rho} + e^{\mu_2/\rho})} \right)^{y_2}
\end{aligned}$$

- Thus

$$\mathsf{L}(\beta, \rho) = \prod_{i=1}^{n} \left(\frac{e^{\mu_{i0}}}{e^{\mu_{i0}} + e^{\rho \ln v_i}} \right)^{y_{i0}} \left(\frac{e^{\rho \ln v_i}}{e^{\mu_{i0}} + e^{\rho \ln v_i}} \right)^{1-y_{i0}} \times \left(\frac{e^{\mu_{i1}/\rho}}{(e^{\mu_{i1}/\rho} + e^{\mu_{i2}/\rho})} \right)^{y_{i1}} \left(\frac{e^{\mu_{i2}/\rho}}{(e^{\mu_{i1}/\rho} + e^{\mu_{i2}/\rho})} \right)^{y_{i2}}$$

COMPUTATION

- Two possible estimation methods.
- The first is the MLE which maximizes $\ln \mathsf{L}$ and is fully efficient.
- The second is a two-step procedure
  - Estimate logit for alternative 1 versus 2
    i.e. maximize over the last two terms in the log-likelihood function.
    This yields estimates $\widehat{\mu_1/\rho}$ and $\widehat{\mu_2/\rho}$ then used to form an estimate $\widehat{v}$ of the inclusive value.

  - Estimate a logit model of alternative 0 versus alternatives 1 and 2 where $\ln \widehat{\upsilon}$ is an additional regressor. i.e. maximize over the first two terms in the log-likelihood function, to give estimates $\widehat{\mu}_1$ and $\widehat{\rho}$.

- The two-step procedure yields consistent but inefficient estimates.
- It is useful for obtaining starting values for the MLE.
- It is not so useful on its own as getting the correct standard errors is difficult.

- The main problem with nested l;ogit can be estimated values of ρ that lie outside $[0, 1]$.
  It can be useful to do a grid search over ρ to constrain ρ to the unit interval and to enumerate the reduction in log-likelihood, if any, due to doing so.
- The nested logit model can be extended to
  - more alternatives
  - higher levels of alternatives (or nesting)
  - generalized extreme value distribution.

MULTINOMIAL PROBIT

- The three-choice example of the multinomial probit model is similar to earlier only now the errors are assumed to be multivariate normal distributed and correlated over the three choices

$$\begin{bmatrix} \varepsilon_{i0} \\ \varepsilon_{i1} \\ \varepsilon_{i2} \end{bmatrix} \sim \mathsf{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{02} \\ \sigma_{10} & \sigma_1^2 & \sigma_{12} \\ \sigma_{20} & \sigma_{21} & \sigma_2^2 \end{bmatrix} \right).$$

- Not all the variance components are identified.
  Here only 3 parameters are identified.

- Then

$$\begin{aligned}\Pr[y = 2] &= \Pr[U_2 > U_1, \quad U_2 > U_0] \\ &= \Pr[\mu_2 + \varepsilon_2 > \mu_1 + \varepsilon_1, \quad \mu_2 + \varepsilon_2 > \mu_0 + \varepsilon_0] \\ &= \Pr[\varepsilon_1 < \varepsilon_2 + \mu_2 - \mu_1, \quad \varepsilon_0 < \varepsilon_2 + \mu_2 - \mu_0] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\varepsilon_2+\mu_2-\mu_1} \int_{-\infty}^{\varepsilon_2+\mu_2-\mu_0} f(\varepsilon_0, \varepsilon_1, \varepsilon_2) d\varepsilon_0 d\varepsilon_1 d\varepsilon_2\end{aligned}$$

  and similarly for p_1 and p_0.
- Estimation is by MLE, though for identification some restrictions will have to be placed on parameters such as those in the error covariance matrix. The model can easily be extended to permit random coefficients β_i

which are normally distributed. All that matters is that the utilities U_{ij} be normally distributed. And there is no need to assume independence of irrelevant alternatives.

IDENTIFICATION

- Bunch (1991) demonstrates that identification of the MNP model can be achieved by considering the difference $U_j - U_0$ between utility of alternative j and that of a benchmark alternative, say 0.
- Then all but one of the parameters of the covariance matrix of the errors $\varepsilon_1 - \varepsilon_0$, can be estimated. One way to achieve this is to normalize $\varepsilon_0 = 0$ and then restrict one covariance element.
- Keane (JBES, 1992) demonstrated that even if just-identification is technically achieved, in practice it can

be practically difficult to estimate with any precision the parameters of the MNP model, in models with regressors that do not vary with the alternative. Further restrictions are needed.

- Keane finds that exclusion restrictions on the regressors (one exclusion for each utility index) work well. Others consider placing further restrictions on the covariance parameters.

COMPUTATION

- The problem is in implementation.
- For the three-choice model above computation of the probabilities can be only reduced to a bivariate normal integral, and an $(m + 1)$ choice model will require an m-variate integral.
- This is computationally burdensome, as the integral needs to be evaluated for every individual in the sample at every iteration of the iterative method used to compute the MLE.

- Until recently at most 3-choice multinomial probit models have been used. Solving this problem is an active area of research.
- The estimation methods are variants of the *method of simulated moments* proposed by McFadden (1989). A recent survey is the book by Gourieroux and Monfort (1996).

ORDERED PROBIT

- Begin with the single latent variable

$$y^* = \mathbf{x}'\boldsymbol{\beta} + u.$$

- Suppose the outcome depends on how large y^* is, with

$$y_i = \begin{cases} 0 & \text{if } y^* \le \alpha_1 \\ 1 & \text{if } \alpha_1 < y^* \le \alpha_2 \\ 2 & \text{if } y^* > \alpha_2. \end{cases}$$

- An example is y^* is a person's propensity to work and we observe whether the person does not work ($y = 0$), works part-time ($y = 1$) or works full-time ($y_i = 2$).

- The ordered probit model specifies $u \sim \mathsf{N}[0,1]$.
- Then

$$\begin{aligned} p_0 &= \Pr[\mathbf{x}'\boldsymbol{\beta} + u \leq \alpha_1] = \Phi(\alpha_1 - \mathbf{x}_i'\boldsymbol{\beta}) \\ p_1 &= \Pr[\alpha_1 < \mathbf{x}'\boldsymbol{\beta} + u \leq \alpha_2] = \Phi(\alpha_2 - \mathbf{x}'\boldsymbol{\beta}) - \Phi(\alpha_1 - \mathbf{x}_i'\boldsymbol{\beta}) \\ p_2 &= 1 - p_0 - p_1. \end{aligned}$$

- The likelihood function is then easily obtained and estimation is by maximum likelihood.
- The ordered logit replaces $\Phi(\cdot)$ in the above by $\Lambda(\cdot)$, the logistic cdf.

SEQUENTIAL MODELS

- An example of a sequential model is sequential probit with three alternatives.
- First choose whether $y = 2$ or $y \neq 2$.
- Second, if $y \neq 2$ choose whether $y = 0$ or $y = 1$.
- Assume a probit model at each stage, with regressors $\mathbf{x}_2$ at the first stage and regressors $\mathbf{x}_1$ at the second stage. Then clearly

$$\begin{aligned} p_2 &= \Pr[y = 2] = \Phi(\mathbf{x}_2'\boldsymbol{\beta}_2), \\ \frac{p_1}{p_0 + p_1} &= \Pr[y_i = 1 | y_i \neq 2] = \Phi(\mathbf{x}_1'\boldsymbol{\beta}_1), \end{aligned}$$

This implies after some algebra

$p_1 = \Pr[y \neq 2] \times \Pr[y = 1|y \neq 2] = (1 - \Phi(\mathbf{x}_2'\boldsymbol{\beta}_2)) \times \Phi(\mathbf{x}_1'\boldsymbol{\beta}_1).$

Finally

$$p_0 = 1 - p_1 - p_2.$$

- The likelihood function is then easily obtained and estimation is by maximum likelihood.

MULTIVARIATE MODELS

- To date we have considered only one discrete dependent variable.
- Now consider more than one.
- For example, jointly model labor supply and fertility

$$y_1 = \begin{cases} 0 & \text{if do not work} \\ 1 & \text{if work} \end{cases}$$

$$y_2 = \begin{cases} 0 & \text{if no children} \\ 1 & \text{if children} \end{cases}$$

- There are four probabilities

$$\begin{aligned} p_{00} &= \Pr[y_1 = 0, y_2 = 0] \\ p_{01} &= \Pr[y_1 = 0, y_2 = 1] \\ p_{10} &= \Pr[y_1 = 1, y_2 = 0] \\ p_{11} &= \Pr[y_1 = 1, y_2 = 1]. \end{aligned}$$

- These are mutually exclusive and exhaust all possibilities, so that $p_{00} + p_{00} + p_{00} + p_{00} = 1$.
- From these probabilities one can form the log-likelihood, and estimate by ML.
- This is essentially the same as a four-choice multino-

mial model.

- All that differs is the story told to derive the functional forms for the probabilities.
- A leading example is the bivariate probit model.

OTHER TOPICS

- Ranked Data: With stated preference data know the second-preferred choice, not just the most-preferred choice.
- Simultaneous Equations: Two binary variables that are simultaneous.
  Easiest if simultaneity is in the latent variables.

- **APPLICATION: LABOR SUPPLY**
- Use data of Mroz (1987) on 753 married women from the 1976 Panel Survey of Income Dynamics (PSID).
- Dependent variable DWORK is a discrete indicator variable that equals
  - 0 if no work in the previous year,
  - 1 if work part-time ($<$ 1000 hours per year) and
  - 2 if work full-time ($>$ 1000 hours)

- The regressors are a constant term and

1. KL6: Number of children less than six
2. K618: Number of children more than six
3. AGE: Age
4. ED: Education (years of schooling completed)
5. NLINCOME: annual nonlabor income of wife measured in $10,000’s.

Variable	Coeff			t-stat	
	1 vs. 0	2 vs. 0	2 vs. 1	1 vs. 0	2 vs. 0
ONE	-1.58	1.20	2.78	-1.6	1.4
$KL6$	$-.93$	-2.01	-1.04	-4.1	-7.7
$K618$	$.10$	$-.19$	$-.09$	1.2	-2.4
AGE	$-.04$	$-.07$	$-.03$	-2.5	-5.0
ED	$.27$	$.25$	$-.02$	5.2	5.6
$NLINCOME$	$-.29$	$-.39$	$-.10$	-2.9	-4.2

- MNL estimates, with coefficients $\boldsymbol{\beta}_0 = 0$ for $\Pr[y = 0]$ normalized to zero, are presented in the table.
- The first column gives $\boldsymbol{\beta}_1$, which gives $\Pr[y = 1]$ vs. $\Pr[y = 0]$.
- The second column gives $\boldsymbol{\beta}_2$, which gives $\Pr[y = 2]$ vs. $\Pr[y = 0]$.
- The third column gives the implied $\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1$, which gives $\Pr[y = 2]$ vs. $\Pr[y = 1]$.
- The t-statistics are also given.