Panel Data Econometrics (PhD)
Non-linear Static & Dynamic Models

Yonas Alem (PhD)

Department of Economics
University of Gothenburg
Environment for Development Initiative (EfD)

July 4, 2012

Maximum Likelihood Estimation Recap

Introduction

- ML method assumes knowledge of the entire distribution, not just of a number of its moments as in GMM method
- If these distributional assumptions are correct, the ML estimator, is under weak regularity conditions, consistent and asymptotically normal.
- It is also asymptotically efficient since it fully exploits the assumptions about the distribution.
- Starting assumptions of ML:
  - The conditional distribution of an observed phenomenon is known, except for a finite number of unknown parameters.
  - These parameters will be estimated by taking those values for them that give the observed values the highest probability, the highest likelihood.
  - It provides an approach of estimating a set of parameters characterizing a distribution, if we know, or assume we know, the form of this distribution

Maximum Likelihood Estimation Recap

Introduction

Example -1

- Consider a large pool of balls filled with red and yellow balls
- One could be interested in the fraction p of red balls in this pool
- Take random sample of N balls only
- Let $y_i = 1$ if ball i is red and $y_i = 0$ otherwise
- Thus, $P\{y_i = 1\} = p$
- Suppose the pool of ball contains $N_1 = \sum_{i=1}^{N} y_i$ red and $N - N_1$ yellow balls

Maximum Likelihood Estimation Recap

Introduction Cont.

Example -1 Cont.

- The likelihood (probability) of obtaining such a sample is given by:

$$P\{N_1 \quad red \quad balls, \quad N - N_1 \quad yellow \quad balls\} = p^{N_1}(1-p)^{N-N_1} \tag{1}$$

- Equation 1 is what is called the **Likelihood Function** and it is a function of the unknown parameter p.
- In ML estimation, we choose a value for p such that the likelihood function is maximal, and obtain $\hat{p}$.
- The conventional practice is to maximize the log-likelihood, which is a simple monotonic transformation of equation [1] (for computational convenience)

$$logL(p) = N_1 log(p) + (N - N_1) log(1-p) \tag{2}$$

Maximum Likelihood Estimation Recap

Introduction Cont.

Example -1 Cont.

- FOC to maximize [1]:

$$\frac{dlogL(p)}{dp} = \frac{N_1}{p} - \frac{N-N_1}{1-p} = 0 \tag{3}$$

- Solving [3] for p gives the ML estimator $\hat{p} = N_1/N$
- It corresponds to the sample proportion of red balls, and most likely to your best guess for p based on the sample drawn
- SOC:

$$\frac{d^2logL(p)}{dp^2} = \frac{N_1}{p^2} - \frac{N-N_1}{(1-p)^2} < 0 \tag{4}$$

- Indicating that we indeed have a maximum
- Another example:

Maximum Likelihood Estimation Recap

Introduction Cont.

Example 2.

- Consider the simple regression model

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i \tag{5}$$

- Keep assumptions [A1][A4]
- The assumptions imply that $E\{y_i|x_i\} = \beta_1 + \beta_s x_i$ & $V\{y_i|x_i\} = \sigma^2$
- To estimate the above model, we need to impose distributional assumption on ε, the most common being assumption [A5] (normal dist.)

Maximum Likelihood Estimation Recap

Introduction Cont.

Example 2.

- The contribution of the i^{th} observation to the likelihood function is the value of the density function at the observed point y_i. Which, for a normal distribution yields,

$$f(y_i|x_i;\beta,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}exp\{-\frac{1}{2}\frac{(y_i-\beta_1-\beta_2 x_i)^2}{\sigma^2}\} \quad (6)$$

- **Note**: y_i has a continues distribution, hence the likelihood of observing a particular outcome y for y_i is zero for any y
- Where $\beta = (\beta_1, \beta_2)$
- The joint density of $(y_1, ..., y_N)$ conditional on $X = (x_1, ..., x_N)'$ is stated as

$$f(y_1,...,y_N|X;\beta,\sigma^2) = \prod_{i=1}^{N} f(y_i|x_i;\beta,\sigma^2) = (\frac{1}{\sqrt{2\pi\sigma^2}})^N \prod_{i=1}^{N} exp\{-\frac{1}{2}\frac{(\ldots$$

Maximum Likelihood Estimation Recap

Introduction Cont.

Example 2. Cont.

- The likelihood function and the joint density function of $y_1, ..., y_N$ are similar except the fact that the former is considered as a function of the unknown parameters β, σ^2
- The LL function is given by

$$logL(\beta, \sigma^2) = -\frac{N}{2}log(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{N}\frac{(y_i - \beta_1 - \beta_2 x_i)^2}{\sigma^2} \quad (8)$$

- Maximizing (8) w.r.t $\beta_1 \& \beta_2$ corresponds to minimizing the residual sum of squares $S(\beta)$, as shown in OLS. Do you see why?

Maximum Likelihood Estimation Recap

Introduction Cont.

Example 2. Cont.

- Meaning that the ML estimators of $\beta_1 \& \beta_2$ are identical to the OLS estimators!
- Denote these estimators by $\hat{\beta}_1$ and $\hat{\beta}_2$, and define the residuals $e_i = y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i$ and maximize (8) w.r.t σ^2. FOC:

$$-\frac{N}{2}\frac{2\pi}{2\pi\sigma^2} + \frac{1}{2}\sum_{i=1}^{N}\frac{e_i^2}{\sigma^4} = 0 \tag{9}$$

- Solve(9) for σ^2 to get the ML estimator for σ^2 given by

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} e_i^2 \tag{10}$$

Maximum Likelihood Estimation Recap

Introduction Cont.

Example 2. Cont.

- Note that however this estimator is consistent but not unbiased (a small sample problem) as the estimator in OLS which was given by

$$s^2 = \frac{1}{N-K}\sum_{i=1}^{N} e_i^2 \tag{11}$$

- In many cases, the ML estimator cannot be shown to be unbiased (unknown small sample properties)
- Its use generally is defended based on asymptotic grounds
- Analytical solution of the ML estimator is also difficult in many cases except in some general cases as shown above

Maximum Likelihood Estimation Recap

Specification Tests

- Three types of tests
  1. The Wald test: pretty much in line with t and F tests
  2. The likelihood ratio test: used to compare two alternative nested models
  3. The lagrange multiplier test: used to test restrictions imposed in estimation

Cross-sectional Binary Choice Models

Recap

- Used to model phenomenal that are of discrete nature
  - Do married women participate in the labor force?
  - Which sections of society are poor?
  - What are the determinants of an agricultural technology adoption?
- For such kinds of models, OLS is generally inappropriate - we rather use binary choice models
- Mostly (although not exclusively) the problems analyzed are micro-economic nature

Cross-sectional Binary Choice Models

Recap Cont.

- Suppose we want to study the impact of income (assumed as the only variable here) on the probability of owning a car:

$$y_i = \beta_1 + \beta_2 x_{i2} + \varepsilon_i = x_i' \beta + \varepsilon_i \tag{12}$$

- Where, $y_i = 1$ if family i owns a car, 0 if family i does not own a car
- $x_i = (x_{i1}, x_{i2})'$
- The standard assumptions:

$$E\{\varepsilon_i | x_i\} = 0 \quad such \;\; that \quad E\{y_i | x_i\} = x_i' \beta \Longrightarrow \tag{13}$$

$$E\{y_i | x_i\} = 1.P\{y_i = 1 | x_i\} + 0.P\{y_i = 0 | x_i\} \tag{14}$$

$$= 1.P\{y_i = 1 | x_i\} = x_i' \beta \tag{15}$$

Cross-sectional Binary Choice Models

Recap Cont.

- Thus, the linear model implies that $x_i'\beta$ is a probability and should therefore lie between 0& 1.
- This is only possible if the x_i values are bounded and if certain restrictions on β are satisfied.
  - Hard to achieve this in practice
- Another fundamental problem:
  - ε_i in equation [1] has a highly non-normal distribution and suffers from heteroscedasticity
  - Because y_i has only two possible outcomes, so does the error term for a given value of x_i
- The distribution of ε_i can be summarized as:

Cross-sectional Binary Choice Models

Recap Cont.

- $$P\{\varepsilon_i = -x_i'\beta | x_i\} = P\{y_i = 0 | x_i\} = 1 - x_i'\beta \qquad (16)$$

$$P\{\varepsilon_i = 1 - x_i'\beta | x_i\} = P\{y_i = 1 | x_i\} = x_i'\beta \qquad (17)$$

- This implies that the variance of the error term is not constant but dependent upon the explanatory variables

$$V\{\varepsilon_i | x_i\} = x_i'\beta(1 - x_i'\beta) \qquad (18)$$

Cross-sectional Binary Choice Models

Recap Cont.

- We therefore use binary choice models (or univariate dichotomous models)
- Describe the probability $y_i = 1$ directly (but derived from an underlying latent variable model (see next pages)
- The general formulation is:

$$P\{y_i = 1|x_i\} = G(x_i, \beta) \quad (19)$$

for some function $G(.)$

- Equation [8] says that the probability of having $y_i = 1$ depends on x_i
- But, clearly, $G(.)$ should take on values in the interval $[0, 1]$ only
- Usually, we assume:

$$G(x_i, \beta) = F(x_i'\beta) \quad (20)$$

Cross-sectional Binary Choice Models

Recap Cont.

- Common choices of F are the standard normal distribution

$$F(x'\beta) = \Phi(x'\beta) = \int_{-\infty}^{x'\beta} \Phi(z)dz, \tag{21}$$

giving rise to the so-called **Probit Model**, and the standard logistic function given by:

$$L(x'\beta) = \frac{e^{x'\beta}}{(1 + e^{x'\beta})} \tag{22}$$

leading to the **Logit Model**

Cross-sectional Binary Choice Models

Recap Cont.

- A third option is a uniform distribution over the interval [0,1] with distribution function:

$$F(x'\beta) = 0, x'\beta < 0; \tag{23}$$

$$F(x'\beta) = x'\beta, 0 \leq x'\beta \leq 1; \tag{24}$$

$$F(x'\beta) = 1, x'\beta > 1. \tag{25}$$

- Leading to what is called the **Linear Probability Model** pretty similar with [12], except that the probabilities are set to 0 or 1 if $x_i'\beta$ exceeds the lower upper limit respectively.

Cross-sectional Binary Choice Models

Recap Cont.

- Probit and logit are more common on applied work.
- Both the standard normal and the standard logistic random variable have an expectation of zero, while the latter has a variance of $\pi^2/3$ instead of 1.
- Correcting for the scaling difference would give similar results
- Apart from their signs, the coefficients in these binary choice models are not easy to interpret directly
- One needs to compute the marginal effects of changes in the explanatory variables

Cross-sectional Binary Choice Models

Recap Cont.

- For a continuous explanatory variable, x_{ik}, say the marginal effect is defined as the partial derivative of the probability that y_i equals one.
- For the three models above, we obtain

$$\frac{\partial \Phi(x_i'\beta)}{\partial x_{ik}} = \phi(x_i'\beta)\beta_k; \tag{26}$$

$$\frac{\partial L(x_i'\beta)}{\partial x_{ik}} = \frac{e^{x_i'\beta}}{(1+e^{x_i'\beta})}\beta_k \tag{27}$$

$$\frac{\partial (x_i'\beta)}{\partial x_{ik}} = \beta_k; (or \;\; 0) \tag{28}$$

Cross-sectional Binary Choice Models

Recap Cont.

- MEs are typically computed for the "average" observation, replacing x_i in the previous expressions with the sample averages.
- For the logit model, re-write [19] as

$$log \frac{p_i}{1-p_i} = x_i'\beta, \tag{29}$$

where $p_i = P\{y_i = 1|x_i\}$ is the probability of observing outcome 1.

- The lhs expression is known as the "log odds ratio"
- For example, an odds ratio of 3 mean that the odds of $y_i = 1$ are 3 times those of $y_i = 0$
- The β coefficients therefore can be interpreted as describing the effect upon the odds ratio
- If $\beta_k = 0.1$, a one-unit increase of x_{ik} increases the odds ratio by about 10% (ceteris paribus)

Cross-sectional Binary Choice Models

Recap Cont.

Estimation

- Very often binary choice modes are derived from underlying behavioral model - following the latent model approach.

$$y^* = x_i'\beta + \epsilon_i \quad (30)$$

- y^* is referred to as the latent variable because it is unobserved
- Assume a probability model of working where a person chooses to work if the utility difference exceeds a certain threshold level
- Thus, one observes $y_i = 1$ (working) if and only if $y_i^* > 0$, and $y_i = 0$ (not working) otherwise.
- Hence,

$$P\{y_i = 1\} = P\{y_i^* > 0\} = P\{x_i'\beta + \epsilon_i > 0\} = P\{-\epsilon_i \leq x_i'\beta\} = F(x \quad (31)$$

Where F denotes the distribution function of $-\epsilon_i$

Cross-sectional Binary Choice Models

Recap Cont.

Estimation

- The likelihood contribution of observation i with $y_i = 1$ is given by $P\{y_i = 1|x_i\}$ as a function of β. We do the same for $y_i = 0$
- We can write the likelihood function to be maximized for the entire sample as

$$L(\beta) = \prod_{i=1}^{N} P\{y_i = 1|x_i;\beta\}^{y_i} P\{y_i = 0|x_i;\beta\}^{1-y_i} \tag{32}$$

and the corresponding loglikelihood function (which is convenient to work with) will be given by

$$logL(\beta) = \sum_{i=1}^{N} y_i logF(x_i'\beta) + \sum_{i=1}^{N} (1-y_i) log(1-F(x_i'\beta)). \tag{33}$$

Non-linear Panel Data Models

Panel Binary Choice Models

- A binary choice model is formulated in terms of an underlying latent model

$$y_{it}^* = x_{it}'\beta + \alpha_i + u_{it} \quad (34)$$

- Where:

$$y_{it} = \begin{cases} 1 & \text{if } y_{it}^* > 0; \\ 0, & \text{otherwise.} \end{cases} \quad (35)$$

- Assume u_{it} has a symmetric distribution with distribution function $F(.), i.i.d$ across $i \& t$ and independent of all x_{is}
- The presence of α_i complicates estimation even when one treats them as fixed unknown parameters, and as random error terms.
- Treating as fixed unknown = including N dummy variables. Hence:

$$logL(\beta, \alpha_1, ..., \alpha_N) = \sum_{i,t} y_{it} logF(\alpha_i + x_{it}'\beta) + \sum_{i,t} (1 - y_{it}) log[1 - F(\alpha_i +$$

Non-linear Models

Panel Binary Choice Models Cont.

- Max w.r.t β & $\alpha_i(i = 1, ..., N)$ will result in consistent estimates only when $T \sim \infty$
- For fixed T & $N \sim \infty$, the estimators are inconsistent!
- Why? Due to what is called "incidental parameter" problem
  - For fixed T, the number of parameters grow with N
  - α_i can be estimated consistently only when $T \sim \infty$ (i.e the number of observations for individual i grows
  - The inconsistency of $\hat{\alpha}_i$ for fixed T will carry over to the estimator for β
- Why was the incidental parameter problem not an issue in the linear FE model?

Non-linear Models

Panel Binary Choice Models Cont.

- Alternative strategy: conditional maximum likelihood estimator
- Considers the likelihood function conditional upon as set of statistics t_i that are sufficient for α_i
  - Conditional upon t_i an individual's likelihood contribution no longer depends upon α_i ut still depends upon the other parameters β
- In the panel data binary choice model, the existence of a sufficient statistics depends upon the functional form of F, i.e., depends upon the distribution of u_{it}
- Let the joint density or probability mass function of $y_{i1}, ..., y_{iT} = f(y_{i1}, ..., y_{iT} | \alpha_i, \beta)$
- If a sufficient statistics t_i exists $\Rightarrow$ $f(y_{i1}, ..., y_{iT} | t_i, \alpha_i, \beta) = f(y_{i1}, ..., y_{iT} | t_i, \beta)$
- Maximize the conditional likelihood function based up on $f(y_{i1}, ..., y_{iT} | t_i, \beta)$ to get a consistent estimator of β

Non-linear Models

Panel Binary Choice Models Cont.

- Takes $t_i = \overline{y}_i$ as a sufficient statistic for α_i and applies CML
- Note: The conditional distribution of $y_{i1}, ..., y_{iT}$ is degenerate if $t_i = 0$ or $t_i = 1$
- Such individuals do not contribute to the CL and shoudl be discarded in estimation
- Only individuals that change status at least once are relevant for estimating β
- Consider the case where $T = 2$
- By conditioning upon $t_i = 1/2$, we restrict the sample to the observation for which y_{it} changes. $(0,1)\&(1,0)$ will be the two possible outcomes

Non-linear Models

Panel Binary Choice Models - FE Logit

- The conditional probability of the first outcome is

$$P\{(0,1)|t_i = 1/2, \alpha_i, \beta\} = \frac{P\{(0,1)|\alpha_i, \beta\}}{P\{(0,1)|\alpha_i, \beta\} + P\{(1,0)|\alpha_i, \beta\}} \tag{37}$$

Making use of

$$P\{(0,1)|\alpha_i, \beta\} = P\{y_{i1} = 0|\alpha_i, \beta\}P\{y_{i2} = 1|\alpha_i, \beta\} \tag{38}$$

and

$$P\{y_{i2} = 1|\alpha_i, \beta\} = \frac{exp\{\alpha_i + x'_{i2}\beta\}}{1 + exp\{\alpha_i + x'_{i2}\beta\}} \tag{39}$$

- Thus the conditional probability is given by

$$P\{(0,1)|t_i = 1/2, \alpha_i, \beta\} = \frac{exp\{(x_{i2} - x_{i1})'\beta\}}{1 + exp\{(x_{i2} - x_{i1})'\beta\}} \tag{40}$$

- And it does not depend on α_i any more

Non-linear Models

Panel Binary Choice Models FE Logit Cont.

- In a similar fashion:

$$P\{(1,0)|t_i = 1/2, \alpha_i, \beta\} = \frac{1}{1 + exp\{(x_{i2} - x_{i1})'\beta\}} \qquad (41)$$

- The conditional distribution of $(y_{i1}, y_{i2}$, given $t_i \& \alpha_i$ is independent of the individual specific effects
- Thus, for $T = 2$, estimation is possible using a standard logit with $x_{i2} - xi1$ as explanatory variables and the change in y_{it} as the endogenous event (1 for positive change, 0 for a negative one)
- Conditioning upon $t_i = 1/2 \Rightarrow$ first differencing (or within transforming)as in the case of linear panel data models
- The principle is straightforward in the case of larger T.
- The CML approach is also extendable to the multinomial logit model

Non-linear Models

Panel Binary Choice Models - RE Probit

- Follow the latent variable specification

$$y_{it}^* = x_{it}'\beta + \varepsilon_{it} \tag{42}$$

- Where:

$$y_{it} = \begin{cases} 1 & \text{if } y_{it}^* > 0; \\ 0, & \text{otherwise.} \end{cases} \tag{43}$$

$$\varepsilon_{it} \sim (0,1) \tag{44}$$

$$E\{\varepsilon_{it}, x_{it}\} = 0 \tag{45}$$

Non-linear Models

Panel Binary Choice Models - RE Probit Cont.

- To estimate β by ML, one needs an assumption about the joint distribution of $(\varepsilon_{i1}, ..., \varepsilon_{iT}$
- If we assume that the ε_{it} are independent, $\Longrightarrow$ $f(y_{i1}, ..., y_{iT}, \beta) = \Pi_t f(y_{it}, \beta)$, which involves T one-dimensional integrals only (as in the cross-sectional case)
- Assume:

$$\varepsilon = \alpha_i + u_{it} \tag{46}$$

$$E\{u_{it}, u_{is}\} = 0 \tag{47}$$

- The joint probability will be:

$$f(y_{i1}, ..., y_{iT} | x_{i1}, ..., x_{iT}, \beta) = \int_{-\infty}^{\infty} f(y_{i1}, ..., y_{iT} | x_{i1}, ..., x_{iT}, \alpha_i, \beta) f(\alpha_i) d\alpha_i \tag{48}$$

Non-linear Models

Panel Binary Choice Models - RE Probit Cont.

- Requires numerical integration over one dimension $\Longrightarrow$ feasible to allow for correlation of the composite error term ε_{it} (not the u_{it})

$$\lambda = Corr(\varepsilon_{it}, \varepsilon_{is}) = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_u^2} \tag{49}$$

- In principle, arbitrary assumptions can be made about distributions of α_i & u_{it}, but in practice, not attractive!
- The most common assumption is the multivariate normal distribution which gives rise to the **Random Effects Probit Model**.
- Let's see how it works:

Non-linear Models

Panel Binary Choice Models - RE Probit Cont.

- Assumptions

$$\varepsilon_{it} \sim N(0,1) \tag{50}$$

- As in the cross-sectional case, the error variance is normalized to 1

$$cov\{\varepsilon_{it}, \varepsilon_{is}\} = \sigma_{\alpha}^2, s \neq t \tag{51}$$

- $\Longrightarrow$

$$\alpha_i \sim NID(0, \sigma_{\alpha}^2) \tag{52}$$

$$u_{it} \sim NID(0, 1 - \sigma_{\alpha}^2) \tag{53}$$

Non-linear Models

Panel Binary Choice Models - RE Probit Cont.

- The Likelihood function is given by:

$$f(y_{it}|x_{it},\alpha_i,\beta) = \Phi(\frac{x'_{it}\beta + \alpha_i}{\sqrt{1-\sigma_\alpha^2}}) \quad if \quad y_{it} = 1 \tag{54}$$

$$= 1 - \Phi(\frac{x'_{it}\beta + \alpha_i}{\sqrt{1-\sigma_\alpha^2}}) \quad if \quad y_{it} = 0 \tag{55}$$

- Where Φ denotes the CDF of the standard normal distribution
- The density of α_i is given by

$$f(\alpha_i) = \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} exp\{-\frac{1}{2}\frac{\alpha_i 2}{\sigma_\alpha^2}\} \tag{56}$$

- The integral in equation [48] can be computed numerically using the Gauss-Hermite quadrature (Butler & Moffitt, 1982)
- The RE Probit is estimable in a standard software such as Stata

Non-linear Models

Panel Binary Choice Models - RE Tobit.

- The RE Tobit is pretty similar to the RE Probit

$$y_{it}^{*} = x_{it}'\beta + \alpha_i + u_{it} \tag{57}$$

- Where:

$$y_{it} = \begin{cases} y_{it}^{*} & \text{if } y_{it}^{*} > 0; \\ 0, & \text{otherwise.} \end{cases} \tag{58}$$

- The standard RE assumptions

$$\alpha_i \sim N(0, \sigma_\alpha) \tag{59}$$

$$u_{it} \sim N(0, \sigma_u) \tag{60}$$

Non-linear Models

Panel Binary Choice Models - RE Tobit Cont.

- $$E\{x_{it}, \alpha_i\} = 0 \quad (61)$$

$$E\{x_{it}, u_{it}\} = 0 \quad (62)$$

- The likelihood function is given as:

$$f(y_{i1}, ..., y_{iT} | x_{i1}, ..., x_{it}, \beta) = \int_{-} \infty^{\infty} \prod_t f(y_{it} | x_{it}, \alpha_i, \beta) f(\alpha_i) d\alpha_i, \quad (63)$$

$$f(\alpha_i) = \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} exp\{-\frac{1}{2}\frac{\alpha_i 2}{\sigma_\alpha^2}\} \quad (64)$$

Non-linear Models

Panel Binary Choice Models - RE Tobit Cont.

- &

(65)

- NOte: the later two expressions are similar to the likelihood contributions in the cross-sectional case (the only difference is the inclusion of α_i in the conditional mean.

Dynamic Probit Models & The Initial Conditions Problem

Example - Poverty Persistence

- A poverty probability model is estimated using a dynamic probit model due to state dependence in poverty
- State dependence in poverty: An individual (a household) experiencing a poverty spell today is much more likely to experience it again in the future
- Five possible reasons for true state dependence in poverty
    1. Unwillingness to take-up a new job or continue working when wage is too low
    2. Deterioration of human capital during a spell of unemployment
    3. Social exclusion due to poverty - addiction to drug and alcohol
    4. Accepting social welfare support as a way of living
    5. Inability to engage in marriage or cohabitation - reduce the opportunity of economies of scale

Dynamic Probit Models & The Initial Conditions Problem

Example - Poverty Persistence Contd...

- Let the probability of being poor be specified as

$$p_{it}^* = \gamma p_{it-1} + x_{it}'\beta + \alpha_i + u_{it} \quad (66)$$

- $(i = 1, ..., N; t = 2, ..., T)$,
- where p_{it}^* is a latent dependent variable;
- x_{it} is a vector of exogenous explanatory variables,
- α_i are (unobserved) individual-specific random effects, and
- $u_{it} \sim N(0, \sigma_u^2)$
- It is assumed that N is large but T is small, which implies that asymptotics depend on N alone.

Dynamic Probit Models & The Initial Conditions Problem

Example - Poverty Persistence Contd...

- p_{it} is the observed binary outcome variable defined as
- Even when u_{it} is serially independent, the composite error term $v_{it} = \alpha_i + u_{it}$ will be correlated overtime because of α_i
- The RE specification adopted implies equicorrelation between the v_{it} in any two periods:

$$\rho = Corr(v_{it}, v_{is}) = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_u^2} \qquad t, s = 2, ..., T; t \neq s. \quad (67)$$

The Initial Conditions Problem

Example - Poverty Persistence Contd...

- The standard RE model also assumes α_i uncorrelated with x_{it}, but correlation can be allowed following Mundlak (1978) & Chamberlain (1984).
  - Correlation between α_i and the observed characteristics can be allowed by assuming a relationship between α_i and either the time means of the x variables or a combination of their lags and leads.
  - E.g $\alpha_i = \overline{x}_i' a + \zeta_i$ where $\zeta_i \sim IID(0, \sigma_\zeta^2)$ independent of x_{it} and u_{it} for all i, t.
- p is a binary variable $\longrightarrow$ we require normalization
- A convenient one is that $\sigma_u^2 = 1$

The Initial Conditions Problem

Example - Poverty Persistence Contd...

- Under the assumption of normality, the transition probability for individual i at time t given α_i, is then given by

$$P[p_{it}|x_{it}, p_{it-1}, \alpha_i] = \Phi[(\gamma p_{it-1} + x'_{it}\beta + \alpha_i)(2p_{it} - 1)], \quad (68)$$

- Φ is the cumulative distribution function of the standard normal

The Initial Conditions Problem

Example - Poverty Persistence Contd...

RE Probit

- Estimation of the model requires an assumption about the initial observations, p_{i1} and its relationship with α_i.
- One option: assume p_{i1} to be exogenous $\Longrightarrow p_{it-1}$ can be incorporated in x_{it} and estimated using the standard RE probit model.
- The assumption that p_{it-1} is exogenous is however a strong assumption because of what is called *the initial conditions problem*
- The start of the observation period does not coincide with the start of the stochastic process generating households' poverty persistence

The Initial Conditions Problem

Example - Poverty Persistence Contd...

RE Probit

- Consequently, estimation of (1) by RE will tend to overstate the degree of state dependence, γ
- Thus, α should be integrated out
- Three available methods
  - Heckman's Estimator
  - Orme Two-Stage Estimator
  - Wooldridge's Conditional Maximum Likelihood Estimator

Dynamic Probit Models

Heckman's Estimator

- Heckman's (1981) approach starts by specifying a linearized reduced-form equation for the initial value of the latent variable:

$$p_{i1}^{*} = z_{i1}'\pi + \eta_i \tag{69}$$

- Where z_{i1} is a vector of exogenous instruments (for example pre-sample variables) and includes x_{i1}, and η_i is correlated with α_i, but uncorrelated with u_{it} for $t \geq 2$.

$$\eta_i = \theta\alpha_i + u_{i1} \tag{70}$$

where $\theta > 0$, and α_i & u_{i1} are independent of one another.

Dynamic Probit Models

Heckman's Estimator

- We also assume that u_{i1} satisfies the same distributional assumptions as u_{it} for $t = 2, ..., T$.
- One can therefore write the linearized reduced form for the latent variable for the initial period as

$$p_{i1}^{*} = z_{i1}^{\prime}\pi + \theta\alpha_i + u_{i1} \tag{71}$$

- The joint probability of the observed binary sequence for individual i given α_i, is thus:

$$\Phi[(z_{i1}^{\prime}\pi + \theta\alpha_i)(2p_{i1} - 1)] \prod_{t=2}^{T} \Phi[(\gamma p_{it-1} + x_{it}^{\prime}\beta + \alpha_i)(2p_{it} - 1)] \tag{72}$$

Dynamic Probit Models

Heckman's Estimator

- For a random sample of households, the likelihood to be maximized is given by

$$\prod_i \int \{\Phi[(z_i'\pi + \theta\sigma_\alpha\alpha*)(2p_{it}-1)] \times \prod_{t=2}^{T} \Phi[(\gamma p_{it-1} + x_{it}'\beta + \sigma_\alpha\alpha^*)(2p_{it}-1)]\} dF(\alpha^*) \quad (73)$$

- Where F is the distribution function of $\alpha^* = \alpha/\sigma_\alpha$
- Under the normalization used,

$$\sigma_\alpha = \sqrt{\rho/1-\rho} \quad (74)$$

- with α assumed to be normally distributed, the integral over α^* can be evaluated using Gaussian-Hermite quadrature

Dynamic Probit Models

Orme's Two-Step Estimator

- Another approach to address the initial conditions problem suggested by Orme (1997, 2001)
- Write the initial period latent equation as:

$$p_{i1}^* = z_i' \lambda^* + v_{i1}. \tag{75}$$

- The cause of the initial conditions problem is the correlation between the regressor p_{it-1} and the unobservable α_i,
- Thus, Orme uses an approximation to substitute α_i with another unobservable component that is uncorrelated with the initial observation
- Assume $(v_{i1}, \alpha_i) \sim BVN(0, 0, \sigma_v^2, \sigma_\alpha^2, r)$

$$\alpha_i | v_{i1} \sim N[r \frac{\sigma_\alpha}{\sigma_v}, \sigma_\alpha^2 (1 - r^2)] \tag{76}$$

Dynamic Probit Models

Orme's Two-Step Estimator Cont

- One can rewrite:

$$\alpha_i = r\frac{\sigma_\alpha}{\sigma_v}v_{i1} + \sigma_\alpha\sqrt{(1-r^2)w_i} \tag{77}$$

- Where w_i is orthogonal to v_{i1} by construction and distributed as $N(0,1)$.
- Substituting (16) for α_i in equation (1) yields,

$$p^*_{it} = \gamma p_{it-1} + x'_{it}\beta + [r\frac{\sigma_\alpha}{\sigma_v}v_{i1} + \sigma_\alpha\sqrt{(1-r^2)w_i}] + u_{it} \tag{78}$$

- Equation (17) has two time invariant unobserved components, v_{i1} and w_i.
- Since $E(w_i|p_{i1}) = 0$, allowing for the correlation of v_{i1}, and p_{i1} addresses the initial conditions problem.

Dynamic Probit Models

Orme's Two-Step Estimator Cont

- Orme shows that Eq(17) and the assumption of BVN imply that,

$$e_i \equiv E(v_{i1}|p_{i1} = (2p_{i1}-1)\sigma_v \phi(\lambda^{*'} z_i/\sigma_v)/\Phi((2p_{i1}-1)\lambda^{*'} z_i/\sigma_v), \quad (79)$$

- Where ϕ and Φ are the Normal density and distribution functions, respectively.
- This is the generalized error from a first period probit equation, analogous to that used in Heckman's sample selection model estimator.
- Thus, one can estimate Eq(17) as a random-effects probit model with an estimate of e_i (obtained after the estimation of (17) using a simple probit for the initial period) used in place of v_{i1}.

Dynamic Probit Models

Wooldridge's Conditional Maximum Likelihood (CML) Estimator

- Proposes an alternative CML estimator that considers the distribution of $p_2, p_3, ..., p_T$ conditional on the initial period value p_1 (and exogenous variables)
- Let the joint density for the observed sequence of the dependent variable $(p_2, p_3, ..., p_T | p_1)$ be written as $(p_T, p_{T-1}, ..., p_2 | p_1, x, \alpha)$.
- To integrate out α_i, specify an approximation for its density conditional on p_{i1}
- In the case of RE probit, the following specification is assumed

$$\alpha_i | p_{i1}, z_i \sim N(\zeta_0 + \zeta_1 p_{i1} + z_i' \zeta, \sigma_a^2) \tag{80}$$

where

$$\alpha_i = \zeta_0 + \zeta_1 p_{i1} + z_i' \zeta + a_i \tag{81}$$

Dynamic Probit Models

Wooldridge's Conditional Maximum Likelihood (CML) Estimator Cont

- in which z_i includes variables that are correlated with α_i.
- z here differs from that in the Heckman specification
- The trick here is that the correlation between p_{i1} and α is handled by the use of Eq(20) which gives rise to a_i - a new unobservable individual-specific heterogeneity term uncorrelated with p_{i1}
- Substituting Eq(17) into Eq(1) gives

$$Pr(p_{it} = 1|a_i, p_{i1}) = \Phi[(x'_{it}\beta + \gamma p_{it-1} + \zeta_1 p_{i1} + z'_i\zeta + a_i] \qquad t = 2, .. \tag{82}$$

Dynamic Probit Models

Wooldridge's Conditional Maximum Likelihood (CML) Estimator Cont

- Consequently, the likelihood function for household i is given by

$$L_i = \int \{\prod_{t=2}^{T} \Phi[(x_{it}'\beta + \gamma p_{it-1} + \zeta_1 p_{i1} + z_i'\zeta + a)(2p_{it} - 1)]\} g^*(a)da, \quad (83)$$

where $g^*(a)$ is the normal probability density function of the new unobservable term a_i.

Non-linear Static and Dynamic Models

End of Lecture!

- **Thank You!**