1 Censoring, Sample Selection and Attrition

- The general structure of the models considered here is represented by the following system of simultaneous equations:

$$
\begin{aligned}
y_{it}^{*} &= m_1(x_{it}, z_{it}, y_{i,t-1}; \theta_1) + u_{it}, && (1)\\
z_{it}^{*} &= m_2(x_{it}, x_{1it}, z_{i,t-1}; \theta_2) + v_{it}, && (2)\\
z_{it} &= h(z_{it}^{*}; \theta_3), && (3)\\
y_{it} &= y_{it}^{*} \text{ if } g_t(z_{i1}, ..., z_{iT}) = 1, && (4)\\
&= 0 \text{ (or unobserved) otherwise,}
\end{aligned}
$$

where i indexes individuals $(i = 1, ..., N)$ and t indexes time $(t = 1, ..., T)$; y_{it}^{*} and z_{it}^{*} are latent endogenous variables with observed counterparts y_{it} and z_{it}; x_{it} and x_{1it} are vectors of exogenous variables; m_1 and m_2 denote general functions characterized by the unknown parameters in θ_1 and θ_2, respectively.

- While we will generally focus on the case where we impose index restrictions on the conditional means, we write the model in the more general form by employing the unknown functions m_1 and m_2 to capture possible non-linearities.
- The mapping from the latent variable to its observed counterpart occurs through the censoring functions h and g_t noting that the former may depend on the unknown parameter vector θ_3.
- We will generally focus on the case where $h(.)$ is an indicator function producing the value 1 if $z_{it}^* > 0$, in which case there are no unknown parameters in the censoring process.

- When we consider the available two-step estimators we will also consider some popular alternative selection rules and these may involve the estimation of additional parameters.
- The function g_t indicates that y_{it}^* may only be observed for certain values of $z_{i1}, ..., z_{iT}$. This includes sample selection where y_{it} is only observed if, for example, $z_{it} = 1$ or, alternatively in the balanced subsample case, if $z_{i1} = ... = z_{iT} = 1$.

- Alternatively, we will consider a special case of interest in which we replace the censoring mechanism in (4) with

$$y_{it} = y_{it}^* \cdot I(y_{it}^* > 0), \tag{5}$$

  where $I(.)$ is an indicator function operator which produces the value 1 if event (.) occurs and zero otherwise.

- The model which incorporates (4) as the censoring or selection rule corresponds with the sample selection model. The model with (5) as the censoring mechanism corresponds to the censored regression model.

- The above model nests many models of interest. For example, it encompasses the static sample selection and censored regression models in which we only observe the dependent variable of primary interest for some subset of the data depending on the operation of a specific selection rule.
- The model also incorporates a potential role for dynamics in both the y equation and the censoring process.
- That is, while panel data are frequently seen as a mechanism for eliminating unobservables which create difficulties in estimation, an important feature and major attraction of panel data is that it provides the ability to estimate the dynamics of various economic relationships based on individual behavior.

- An important feature of these models is related to identification. In many of the models that we consider it is possible to obtain identification of the parameters of interest by simply relying on non-linearities which arise from the distributional assumptions.

- In general, this is not an attractive, nor frequently accepted, means of identification. As these issues are frequently quite complicated we avoid such a discussion by assuming that the elements in the vector x_{1it} appear as explanatory variables in the selection equation (2) but are validly excluded from the primary equation (1).

- In this way the models are generally identified.

- A key aspect of any panel data model is the specification and treatment of its disturbances. We write the respective equations' errors as

$$u_{it} = \alpha_i + \varepsilon_{it} \tag{6}$$
$$v_{it} = \xi_i + \eta_{it} \tag{7}$$

  which indicates that they comprise individual effects, α_i and ξ_i, and individual specific time effects, ε_{it} and η_{it}, which are assumed to be independent across individuals.

- This corresponds to the typical one-way error components model.

- Moreover, we allow the errors of the same dimension to be correlated across equations. In some instances we will assume that both the individual effects and the idiosyncratic disturbances can be treated as random variables, distributed independently of the explanatory variables.
- In such cases, we will often assume that the error components are drawn from known distributions.

- For many empirical applications, however, these assumption are not appropriate. For example, one may expect that some subset of the explanatory variables are potentially correlated with the one or both of the different forms of disturbances.
- Accordingly, it is common to treat the individual effects as fixed effects, which are potentially correlated with the independent variable, and we will consider the available procedures for estimating under such conditions.
- Second, while distributional assumptions are frequently useful from the sake of implementation, for many applications they may not be appropriate.

- As many of the procedures we examine are likelihood based any misspecification of the parametric component may lead to the resulting estimators being inconsistent.
- Thus, while we begin the analysis of each submodel by making distributional assumptions regarding the disturbances we will also examine some semi-parametric estimators which do not rely on distributional assumptions.
- Finally, note that for the majority of models the parameters of primary interest are those contained in the vector θ_1, the variance σ_ε^2 and, when appropriate, σ_α^2. In some instances, however, there may be interest in the θ_2 vector.

2 Sample selection bias and robustness of standard estimators

- One can easily illustrate the problems generated by the presence of attrition or selection bias by examining the properties of standard estimators for the primary equation where we estimate only over the sample of uncensored observations.
- Consider the simplest case of (1) where the dependent variable is written as a linear function of only the exogenous explanatory variables:

$$y_{it} = x_{it}'\beta + \alpha_i + \varepsilon_{it}, \tag{8}$$

where we consider that each of the selection rules captured in (4) and (5) can be written as $z_{it} = 1$.

- To illustrate the problems with such pooled estimation of (8) we can take expectations of (8) conditional upon y_{it} being observed, which gives

$$E(y_{it}|x_{it}, z_{it} = 1) = x'_{it}\beta + E(\alpha_i|x_{it}, z_{it} = 1) + E(\varepsilon_{it}|x_{it}, z_{it} = 1), \quad (9)$$

  noting that the last two terms will in general have non-zero values, which are potentially correlated with the $x's,$ due to the dependence between α_i and ξ_i, and ε_{it} and ξ_{it}.

- These terms will, in general, be non-zero whenever $Pr\{z_{it} = 1|y_{it}, x_{it}\}$ is not independent of y_{it}. Accordingly, least squares estimation of (8) will lead to biased estimates of β due to this misspecification of the mean.

- This above result is well known in the cross-sectional case and is a restatement of the results of Heckman (1979).
- However, given that in the panel data setting we have repeated observations on the individual one might think that the availability of panel data estimators which exploit the nature of the error structure might provide some scope to eliminate this bias without the use of such a variable.
- Accordingly, it is useful to discuss the properties of the standard fixed effects and random effects estimators in the linear model when the selection mechanism is endogenous. Thus we first consider estimation of (8) by the standard linear fixed effects or random effects procedures.

- To consider these estimators we first introduce some additional notation.
- Observations on y_{it} are treated as available if $z_{it} = 1$ and missing if $z_{it} = 0$.
- We define $c_i = \prod_{t=1}^{T} z_{it}$, so that $c_i = 1$ if and only if y_{it} is observed for all t.
- The first estimator for β that we consider are the standard random effects estimators.

- Defining

$$\lambda_i = 1 - \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T_i \sigma_\alpha^2}}$$

where $T_i = \sum_{t=1}^{T} z_{it}$ denotes the number of time periods y_{it} is observed, the random effects estimator based on the unbalanced panel (using all available cases) can be written as

$$\hat{\beta}_{RE}^{U} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} z_{it}(x_{it} - \lambda_i \bar{x}_i)(x_{it} - \lambda_i \bar{x}_i)' \right)^{-1} \times \left(\sum_{i=1}^{N} \sum_{t=1}^{T} z_{it}(x_{it} - \lambda_i \bar{x}_i)(y_{it} - \lambda_i \bar{y}_i) \right) \quad (10)$$

where $\bar{x}_i = T_i^{-1} \sum_{t=1}^{T} z_i x_{it}$ and $\bar{y}_i = T_i^{-1} \sum_{t=1}^{T} z_i y_{it}$ denote averages over the available observations.

- In some cases attention may be restricted to the balanced sub-panel comprising only those individuals that have completely observed records.
- The resulting random effects estimator is given by

$$\hat{\beta}^B_{RE} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} c_i (x_{it} - \lambda_i \bar{x}_i)(x_{it} - \lambda_i \bar{x}_i)' \right)^{-1} \times \left(\sum_{i=1}^{N} \sum_{t=1}^{T} c_i (x_{it} - \lambda_i \bar{x}_i)(y_{it} - \lambda_i \bar{y}_i) \right). \quad (11)$$

- Note that all units for which $c_i = 1$ will have the same value for λ_i.

- Under appropriate regularity conditions, these two estimators are consistent for $N \to \infty$ if

$$E(\alpha_i + \varepsilon_{it}|z_i) = 0, \tag{12}$$

where $z_i = (z_{i1}, ..., z_{iT})'$.

- This condition states that the two components of the error term in the model are mean independent of the sample selection indicators in z_i (conditional upon the exogenous variables).
- This appears to be a very strong condition and essentially implies that the selection process is independent of both of the unobservables in the model.
- One would suspect that for a large range of empirical cases this is unlikely to be true and this does not appear to be an attractive assumption to impose.

- Given that the random effects estimator does not appear to be useful in the presence of selection bias it is worth focussing on the suitability of the fixed effects estimators of β.
- For the unbalanced panel the estimator can be written as

$$\hat{\beta}_{FE}^{U} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} z_{it}(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' \right)^{-1} \times \left(\sum_{i=1}^{N} \sum_{t=1}^{T} z_{it}(x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) \right), \tag{13}$$

- The corresponding estimator for the balanced sub-panel is given by

$$\hat{\beta}_{FE}^{B} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} c_i (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' \right)^{-1} \times \left(\sum_{i=1}^{N} \sum_{t=1}^{T} c_i (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) \right). \tag{14}$$

- Under appropriate regularity conditions, consistency of these two estimators requires that

$$E(\varepsilon_{it} - \bar{\varepsilon}_i | z_i) = 0, \tag{15}$$

where $\bar{\varepsilon}_i = T_i^{-1} \sum_{t=1}^{T} z_{it}\varepsilon_{it}$.

- This indicates that the estimation over the subsample for which $z_{it} = 1$ will produce consistent estimates if the random component determining whether $z_{it} = 1$ is eliminated in the fixed effects transformation.
- That is, the unobservable component determining selection for each individual is time-invariant.
- While this may be true in certain instances it is likely that in many empirical examples such an assumption would not be reasonable as it imposes that the selection process is independent of the idiosyncratic errors.

- This discussion illustrates that the conventional linear panel data estimators are inappropriate for the linear model with selection.
- The random effects estimator essentially requires that selection is determined outside the model while the fixed effects estimator imposes that, conditional on the individual effects, the selection process is determined outside the model.
- While the fixed effects estimator is more robust, it still is unsatisfactory for most empirical examples of panel data models with selectivity.
- Accordingly, we now begin to examine a range of estimators which handle the situation for which (12) and (15) are not satisfied.

3 Tobit and Censored Regression Models

- The first model considered can be fully described by a subset of the equations capturing the general model outlined above.
- The model has the form

$$
\begin{aligned}
y_{it}^* &= m_1(x_{it}, y_{i,t-1}; \theta_1) + u_{it}, & (16)\\
y_{it} &= y_{it}^* \text{ if } y_{it}^* > 0, & (17)\\
&= 0 \text{ (or unobserved) otherwise.}
\end{aligned}
$$

- This considers a latent variable y_{it}^*, decomposed into a conditional mean depending upon x_{it} and possibly a lagged observed outcome $y_{i,t-1}$, and an unobserved mean zero error term u_{it}.
- The observed outcome equals the latent value if the latter is positive and zero otherwise.
- This model is the panel data extension of the tobit type I (under certain distributional assumptions) or censored regression model which is commonly considered in cross-sectional analyses.

- We now consider estimation of this standard censored regression model in (16)-(17) under different sets of assumptions.
- The simplest case arises when the lagged dependent variable is excluded from (16), and when ε_{it} is assumed to be drawn from a normal distribution, independent of the explanatory variables.
- We then consider the model where we allow for a lagged dependent variable.
- As we will see the estimation is somewhat more difficult because one has to incorporate the additional complications arising from the initial conditions.
- We then proceed to a consideration of the model where we relax the distributional assumptions that we impose on the error terms.

3.1 Random Effects Tobit

- First, we consider the static tobit model, given by
-
$$y_{it}^* = m_1(x_{it}; \theta_1) + u_{it},$$
where the censoring rule is stated in (5)
$$y_{it} = y_{it}^* \text{ if } y_{it}^* > 0, \qquad (18)$$
$$y_{it} = 0 \text{ if otherwise.}$$
- We also assume that u_{it} has mean zero and constant variance, independent of $(x_{i1}, ..., x_{iT})$.
- In order to estimate θ_1 by maximum likelihood we add an additional assumption regarding the joint distribution of $u_{i1}, ..., u_{iT}$.

- The likelihood contribution of individual i is the (joint) probability/density of observing the T outcomes $y_{i1}, \ldots, y_{iT}$, which is determined from the joint distribution of the latent variables $y_{i1}^*, \ldots, y_{iT}^*$ by integrating over the appropriate intervals.
- In general, this will imply T integrals, which in estimation are typically to be computed numerically.
- When $T = 4$ or more, this makes maximum likelihood estimation infeasible.

- If the u_{it} are assumed to be independent, we have that

$$f(y_{i1}, ..., y_{iT}|x_{i1}, ..., x_{iT}; \vartheta_1) = \prod_t f(y_{it}|x_{it}; \vartheta_1),$$

  where ϑ_1 contains all relevant parameters (including θ_1), which involves T one-dimensional integrals only (as in the cross-sectional case).

- This, however is highly restrictive.

- If, instead, we impose the error components assumption that $u_{it} = \alpha_i + \varepsilon_{it}$, where ε_{it} is i.i.d. over individuals and time, we can write the joint probability/density as

$$
\begin{aligned}
f(y_{i1},...,y_{iT}|x_{i1},...,x_{iT};\vartheta_1) &= \int_{-\infty}^{\infty} f(y_{i1},...,y_{iT}|x_{i1},...,x_{iT},\alpha_i;\vartheta_1)f(\alpha_i)d\alpha_i \\
&= \int_{-\infty}^{\infty} \left[\prod_t f(y_{it}|x_{it},\alpha_i;\vartheta_1)\right] f(\alpha_i)d\alpha_i, \quad (19)
\end{aligned}
$$

  where f is generic notation for a density or probability mass function.

- This is a feasible specification that allows the error terms to be correlated across different periods, albeit in a restrictive way. The crucial step in (19) is that conditional upon α_i the errors from different periods are independent.
- In principle arbitrary assumptions can be made about the distributions of α_i and ε_{it}. For example, one could assume that ε_{it} is i.i.d. normal while α_i has a Student t-distribution.
- However, this may lead to distributions for $\alpha_i + \varepsilon_{it}$ that are nonstandard and this is unattractive.
- Accordingly, it is more common to start from the joint distribution of $u_{i1}, ..., u_{iT}$.

- We assume that the joint distribution of $u_{i1}, ..., u_{iT}$ is normal with zero means and variances equal to $\sigma_\alpha^2 + \sigma_\varepsilon^2$ and $\text{cov}\{u_{it}, u_{is}\} = \sigma_\alpha^2,\ s \neq t$.
- This is the same as assuming that α_i is $NID(0, \sigma_\alpha^2)$ and ε_{it} is $NID(0, \sigma_\varepsilon^2)$.
- The likelihood function can then be written as in (19), where

$$f(\alpha_i) = \frac{1}{\sqrt{2\pi}\sigma_\alpha} \exp\left\{-\frac{1}{2}\left(\frac{\alpha_i}{\sigma_\alpha}\right)^2\right\}. \tag{20}$$

- and

$$f(y_{it}|x_{it},\alpha_i;\vartheta_1) = \frac{1}{\sqrt{2\pi}\sigma_\varepsilon}\exp\left\{-\frac{1}{2}\left(\frac{y_{it}-m_1(x_{it};\theta_1)-\alpha_i}{\sigma_\varepsilon}\right)^2\right\} \text{ if } y_{it}>0$$
$$= 1-\Phi\left(\frac{m_1(x_{it};\theta_1)+\alpha_i}{\sigma_\varepsilon}\right) \text{ if } y_{it}=0, \qquad (21)$$

  where Φ denotes the standard normal cumulative density function.

- The latter two expressions are similar to the likelihood contributions in the cross-sectional case, with the exception of the inclusion of α_i in the conditional mean.

- The estimation of this model is identical to estimation of the tobit model in the cross-sectional setting except that we now have to account for the inclusion of the individual specific effect.
- As this individual effect is treated as a random variable, and the disturbances in the model are normally distributed, the above procedure is known as random effects tobit.
- Note that while we do not do so here, it would be possible to estimate many of the models considered in the survey of cross-sectional tobit models by Amemiya (1978) by allowing for an individual random effect.

3.2 Random Effects Tobit with Endogenous Explanatory Variables

- The discussion of the random effects tobit model in the previous section assumed that the disturbances are independent of the explanatory variables
- One useful extension of the model would be instances where some of the explanatory variables were treated as endogenous.
- This is similar to the cross-sectional model of Smith and Blundell (1986) who present a conditional ML estimator to account for the endogeneity of the explanatory variables.

- The estimator simply requires estimating the residuals from the model for the endogenous explanatory and including them as an additional explanatory variable in the cross-sectional tobit likelihood function.

- Vella and Verbeek (1999) extend this to the panel case by exploiting the error components structure of the model. We now present this case where we assume the endogenous explanatory variable is fully observed.

- The model has the following form:

$$y^*_{it} = m_1(x_{it}, z_{it}; \theta_1) + \alpha_i + \varepsilon_{it} \quad (22)$$

$$z_{it} = m_2(x_{it}, x_{1it}, z_{i,t-1}; \theta_2) + \xi_i + \eta_{it} \quad (23)$$

$$y_{it} = y^*_{it} \cdot (y^*_{it} > 0) \quad (24)$$

  The model's disturbances are assumed to be generated by the following distribution:

- $$\left(\begin{array}{c}\alpha_i\iota+\varepsilon_i\\ \xi_i\iota+\eta_i\end{array}\right)|_{X_i}\sim NID\left(\left(\begin{array}{c}0\\0\end{array}\right),\left(\begin{array}{cc}\sigma_\alpha^2\iota\iota'+\sigma_\varepsilon^2 I & \sigma_{\alpha\xi}\iota\iota'+\sigma_{\varepsilon\eta}I\\ & \sigma_\xi^2\iota\iota'+\sigma_\eta^2 I\end{array}\right)\right) \tag{25}$$

  where ι is a T-vector of ones.

- Exploiting this joint normality assumption allows us to write

  $$E(u_{it}|X_i,v_i)=\tau_1 v_{it}+\tau_2\bar{v}_i, \tag{26}$$

  where $\tau_1=\sigma_{\varepsilon\eta}/\sigma_\varepsilon^2$, $\tau_2=T(\sigma_{\alpha\xi}-\sigma_{\varepsilon\eta}\sigma_\xi^2/\sigma_\varepsilon^2)/(\sigma_\eta^2+T\sigma_\xi^2)$ and $\bar{v}_i=T^{-1}\sum_{t=1}^T v_{it}$.

- As the endogenous explanatory variable is uncensored the conditional distribution of the error terms in (22) given z_i remains normal with an error components structure.
- Thus one can estimate the model in (22) and (24) conditional on the estimated parameters from (23) using the random effects likelihood function, after making appropriate adjustments for the mean and noting that the variances now reflect the conditional variances.

- Write the joint density of $y_i = (y_{i1}, ..., y_{iT})'$ and z_i given X_i as:[1]

$$f(y_i|z_i, X_i; \vartheta_1, \vartheta_2) f(z_i|X_i; \vartheta_2), \tag{27}$$

  where ϑ_1 denotes $(\theta_1, \sigma_\alpha^2, \sigma_\varepsilon^2, \sigma_{\alpha\xi}, \sigma_{\varepsilon\eta})$ and ϑ_2 denotes $(\theta_2, \sigma_\xi^2, \sigma_\eta^2)$.

- We first estimate ϑ_2 by maximizing the marginal likelihood function of the z_i's.

- Subsequently, the conditional likelihood function

$$\prod_i f(y_i|z_i, X_i; \vartheta_1, \hat{\vartheta}_2) \tag{28}$$

  is maximized with respect to ϑ_1 where $\hat{\vartheta}_2$ denotes a consistent estimate of ϑ_2.

[1] When (23) is dynamic with an exogenous initial value z_{io}, (27) is valid if z_{io} is included in X_i. When the initial value is endogenous, we need to include z_{i0} in z_i.

- The conditional distribution of y_i given z_i is multivariate normal with an error components structure.
- The conditional expectation can be derived directly from (26), substituting $v_{it} = z_{it} - m_2(x_{it}, x_{1it}, z_{i,t-1}; \theta_2)$, while the covariance structure corresponds to that of $\nu_{1i} + \nu_{2,it}$, where ν_{1i} and $\nu_{2,it}$ are zero mean normal variables with zero covariance and variances

$$\sigma_1^2 = V\{\nu_{1i}\} = \sigma_\varepsilon^2 - \sigma_{\varepsilon\eta}^2 \sigma_\eta^{-2}, \tag{29}$$

$$\sigma_2^2 = V\{\nu_{2,it}\} = \sigma_\alpha^2 - \frac{T\sigma_{\alpha\xi}^2 \sigma_\eta^2 + 2\sigma_{\alpha\xi}\sigma_{\varepsilon\eta}\sigma_\eta^2 - \sigma_{\varepsilon\eta}^2 \sigma_\xi^2}{\sigma_\eta^2(\sigma_\eta^2 + T\sigma_\xi^2)}. \tag{30}$$

- These follow from straightforward matrix manipulations and show that the error components structure is preserved and the conditional likelihood function of (22) and (24) has the same form as the marginal likelihood function without endogenous explanatory variables.[2]
- The conditional maximum likelihood estimator can be extended to account for multiple endogenous variables as the appropriate conditional expectation is easily obtained as all endogenous regressors are continuously observed.
- Even if the reduced form errors of the endogenous regressors are correlated, provided they are characterized by an error components structure it can be shown that the conditional distribution of $\alpha_i + \varepsilon_{it}$ also has an error components structure.

[2] The algebraic manipulations are simplified if σ_1^2 and σ_2^2 replace the unconditional variances σ_ε^2 and σ_α^2 in ϑ_1. In this case, estimates for the latter two variances are easily obtained in a third step from the estimates from the first stage for σ_ξ^2 and σ_η^2, and the estimated covariances from the mean function, using the equalities in (29) and (30).

- In general the conditional maximum likelihood estimator cannot be employed when $z_{it} \neq z_{it}^*$.
- Thus the family of sample selection models considered below cannot be estimated by conditional maximum likelihood. One interesting exception, however, is when the primary equation is estimated over the subsample of individuals that have $z_{is} = z_{is}^*$, for all $s = 1, ..., T$.

3.3 Dynamic Random Effects Tobit

- The ability to estimate dynamic relationships from individual level data is an important attraction of panel data.
- Accordingly, an extension to the above model which involves the inclusion of a lagged dependent variable is of economic interest.
- Let us now reconsider the random effects tobit model, and generalize the latent variable specification to

$$y^*_{it} = m_1(x_{it}, y_{i,t-1}; \theta_1) + \alpha_i + \varepsilon_{it}, \tag{31}$$

with $y_{it} = y^*_{it}$ if $y^*_{it} > 0$ and 0 otherwise.

- Now consider maximum likelihood estimation of this dynamic random effects tobit model, making the same distributional assumptions as above.
- In general terms, the likelihood contribution of individual i is given by (compare (19))

$$f(y_{i1}, ..., y_{iT}|x_{i1}, ..., x_{iT}; \vartheta_1) = \int_{-\infty}^{\infty} f(y_{i1}, ..., y_{iT}|x_{i1}, ..., x_{iT}, \alpha_i; \vartheta_1) f(\alpha_i) d\alpha_i$$

$$= \int_{-\infty}^{\infty} \left[\prod_{t=2}^{T} f(y_{it}|y_{i,t-1}, x_{it}, \alpha_i; \vartheta_1)\right] f(y_{i1}|x_{i1}, \alpha_i; \vartheta_1) f(\alpha_i) d\alpha_i, \quad (32)$$

- where

$$f(y_{it}|y_{i,t-1}, x_{it}, \alpha_i; \vartheta_1)$$
$$= \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} \exp\left\{-\frac{1}{2}\left(\frac{y_{it} - m_1(x_{it}, y_{i,t-1}; \theta_1) - \alpha_i}{\sigma_\eta}\right)^2\right\} \text{ if } y_{it} > 0,$$
$$= 1 - \Phi\left(\frac{m_1(x_{it}, y_{i,t-1}; \theta_1) + \alpha_i}{\sigma_\varepsilon}\right) \text{ if } y_{it} = 0.$$

- This is analogous to the static case and $y_{i,t-1}$ is simply included as an additional explanatory variable.
- However, the term $f(y_{i1}|x_{i1}, \alpha_i; \theta_1)$ in the likelihood function may cause problems. It gives the distribution of y_{i1} without knowing its previous value but conditional upon the unobserved heterogeneity term α_i.

- If the initial value is exogenous in the sense that its distribution does not depend upon α_i, we can place the term $f(y_{i1}|x_{i1}, \alpha_i; \vartheta_1) = f(y_{i1}|x_{i1}; \vartheta_1)$ outside the integral.
- In this case, we can simply consider the likelihood function conditional upon y_{i1} and ignore the term $f(y_{i1}|x_{i1}; \vartheta_1)$ in estimation.
- The only consequence may be a loss of efficiency if $f(y_{i1}|x_{i1}; \vartheta_1)$ provides information about ϑ_1.
- This approach would be appropriate if the starting value is necessarily the same for all individuals or if it is randomly assigned to individuals.

- However, it may be hard to argue in many applications that the initial value y_{i1} is exogenous and does not depend upon a person's unobserved heterogeneity.
- In that case we would need an expression for $f(y_{i1}|x_{i1}, \alpha_i; \vartheta_1)$ and this is problematic. If the process that we are estimating has been going on for a number of periods before the current sample period, $f(y_{i1}|x_{i1}, \alpha_i; \vartheta_1)$ is a complicated function that depends upon person i's unobserved history.
- This means that it is typically impossible to derive an expression for the marginal probability $f(y_{i1}|x_{i1}, \alpha_i; \vartheta_1)$ that is consistent with the rest of the model.

- Heckman (1981) suggests an approximate solution to this initial conditions problem that seems to work reasonably well in practice.
- It requires an approximation for the marginal distribution of the initial value by a tobit function using as much pre-sample information as available, without imposing restrictions between its coefficients and the structural parameters in θ_1.

3.4 Fixed Effects Tobit Estimation

- The fully parametric estimation of the tobit model assumes that both error components have a normal distribution, independent of the explanatory variables.
- Clearly, this is restrictive and a first relaxation arises if we treat the individual-specific effects α_i as parameters to be estimated, as is done in the linear fixed effects model.
- However, such an approach is generally not feasible in non-linear models.

- The loglikelihood function for the fixed effects tobit model has the general form

$$\log L = \sum_{i=1}^{N} \left[\sum_{t=1}^{T} \log \ f(y_{it}|x_{it}, \alpha_i; \vartheta_1) \right], \tag{33}$$

where

-

$$
\begin{aligned}
f(y_{it}|x_{it}, \alpha_i; \vartheta_1) &= \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} \exp\left\{-\frac{1}{2}\left(\frac{y_{it} - m_1(x_{it}; \theta_1) - \alpha_i}{\sigma_\varepsilon}\right)^2\right\} \text{ if } y_{it} > 0 \\
&= 1 - \Phi\left(\frac{m_1(x_{it}; \theta_1) + \alpha_i}{\sigma_\varepsilon}\right) \text{ if } y_{it} = 0.
\end{aligned}
$$

- Maximization of (33) can proceed through the inclusion of N dummy variables to capture the fixed effects or using an alternative strategy, described in Greene (2004), which bypasses the large computation demands of including so many additional variables.

- This fixed effects tobit estimators is subject to the incidental parameter problem (Neyman and Scott, 1948, Lancaster, 2000), and results in inconsistent estimators of the parameters of interest if the number of individuals goes to infinity while the number of time periods is fixed.
- It was generally believed that the bias resulting from fixed effects tobit was large although more recent evidence provided by Greene suggests this may not be the case.
- On the basis of Monte Carlo evidence, Greene (2004) concludes that there is essentially no bias in the estimates of θ_1. However, the estimate of σ_ε is biased and this generates bias in the estimates of the marginal effects.
- Greene also concludes that the bias is small if T is 5 or greater.

- Hahn and Newey (2004) suggest two approaches to bias reduction in fixed effects estimation of non-linear models such as the fixed effects tobit model.
- The first procedure is based on the use of jackknife methods and exploits the variation in the fixed effects estimator when each of the observations are, in turn, separately deleted.
- By doing so one is able to form a bias-corrected estimator using the Quenouille (1956) and Tukey (1958) jackknife formula.

- For simplicity, let $m_1(x_{it};\theta_1) = x_{it}'\beta$ and let $\widehat{\beta}_{(t)}$ denote the fixed effects estimator based on the subsample excluding the observations for the t^{th} wave.
- The jackknife estimator ($\widehat{\beta}_{JK}$) is defined to be

$$\widehat{\beta}_{JK} = T\widehat{\beta} - (T-1)\sum_{t=1}^{T}\widehat{\beta}_{(t)}/T,$$

  where $\widehat{\beta}$ is the fixed effects estimator based on the entire panel.
- Hahn and Newey note that the panel jackknife is not particularly complicated. While it does require $(T+1)$ fixed effects estimations of the model one can employ the algorithm proposed by Greene, discussed above, and the estimates of $\widehat{\beta}$ and $\widehat{\alpha}_i$ can be used as starting values.

- The second procedure is an analytic bias correction using the bias formula obtained from an asymptotic expansion as the number of periods grows.
- This is based on an approach suggested by Waterman et al. (2000) and is also related to the approach adopted by Woutersen (2002).
- Note that while none of these authors examine the fixed effect tobit model, preferring to focus mainly on discrete choice models, the approaches are applicable.
- Hahn and Newey (2004) provide some simulation evidence supporting the use of their procedures in the fixed effects probit model.

3.5 Semi-Parametric Estimation

- As shown in Honoré (1992) is also possible to estimate the parameters of panel data tobit models like (16)–(17) with no assumptions on the distribution of the individual specific effects and with much weaker assumptions on the transitory errors.
- To fix ideas, consider a model with a linear index restriction, that is

-
$$y^*_{it} = x'_{it}\beta + \alpha_i + \varepsilon_{it},$$

and

$$y_{it} = y^*_{it} \text{ if } y^*_{it} > 0,$$
$$y_{it} = 0 \text{ otherwise.}$$

- The method proposed in Honoré (1992) is based on a comparison of any two time periods, t and s.
- The key insight behind the estimation strategy is that if ε_{it} and ε_{is} are identically distributed conditional on (x_{it}, x_{is}) then

- $$\begin{aligned} v_{ist}(\beta) &= \max\{y_{is}, (x_{is} - x_{it})' \beta\} - \max\{0, (x_{is} - x_{it})' \beta\} \\ &= \max\{\alpha_i + \varepsilon_{is}, -x'_{is}\beta, -x'_{it}\beta\} - \max\{-x'_{is}\beta, -x'_{it}\beta\} \end{aligned}$$

  and

  $$\begin{aligned} v_{its}(\beta) &= \max\{y_{it}, (x_{it} - x_{is})' \beta\} - \max\{0, (x_{it} - x_{is})' \beta\} \\ &= \max\{\alpha_i + \varepsilon_{it}, -x'_{it}\beta, -x'_{is}\beta\} - \max\{-x'_{it}\beta, -x'_{is}\beta\} \end{aligned}$$

  are also identically distributed conditional on (x_{it}, x_{is}).

- This can be used to construct numerous moment conditions of the form

$$E\left[\left(g\left(v_{ist}\left(\beta\right)\right)-g\left(v_{its}\left(\beta\right)\right)\right)h\left(x_{it},x_{is}\right)\right]=0 \tag{34}$$

- If g is increasing and $h\left(x_{it},x_{is}\right)=x_{is}-x_{it}$, these moment conditions can be turned into a minimization problem which identifies β subject to weak regularity conditions.

- For example, with $g\left(d\right)=d$, (34) corresponds to the first-order conditions of the minimization problem

$$\underset{b}{\text{minimize}} E\left[\left(\max\{y_{is}, (x_{is} - x_{it})' b\} - \max\{y_{it}, -(x_{is} - x_{it})' b\} - (x_{is} - x_{it})' b\right)^2 + 2 \cdot 1\{y_{is} < (x_{is} - x_{it})' b\}((x_{is} - x_{it})' b - y_{is})y_{it} + 2 \cdot 1\{y_{it} < -(x_{is} - x_{it})' b\}(-(x_{is} - x_{it})' b - y_{it})y_{is}\right]$$

which suggests estimating β by minimizing

$$
\begin{aligned}
\sum_{i=1}^{n} \sum_{s<t} \Big(& \max\{y_{is}, (x_{is} - x_{it})' \, b\} \\
& - \max\{y_{it}, -(x_{is} - x_{it})' \, b\} - (x_{is} - x_{it})' \, b \Big)^2 \\
& + 2 \cdot 1\{y_{is} < (x_{is} - x_{it})' \, b\}((x_{is} - x_{it})' \, b - y_{is}) y_{it} \\
& + 2 \cdot 1\{y_{it} < -(x_{is} - x_{it})' \, b\}(-(x_{is} - x_{it})' \, b - y_{it}) y_{is}
\end{aligned}
\tag{35}
$$

- The objective function in (35) is convex in b, as are other objective functions based on (34).
- Honoré and Kyriazidou (2000) discuss estimators defined by a general $g(d)$ as well as estimators based on moment conditions that are derived under the stronger assumption that the distribution of $(\varepsilon_{it}, \varepsilon_{is})$ is exchangeable conditional on (x_{it}, x_{is}).

3.6 Semi-Parametric Estimation in the Presence of Lagged Dependent Variables.

- Honoré (1993), Hu (2002) and Honoré and Hu (2004) show how one can modify the moment conditions in (34) in such a way that one can allow for lagged dependent variables as explanatory variables.
- The specifics for this differs depending on whether the lagged latent or the lagged censored variable is used, and the main difficulty in this literature is that it is not easy to show that the moment conditions are uniquely satisfied at the true parameter values.

4 Models of Sample Selection and Attrition

- As discussed above the tobit model has the somewhat unattractive feature that the index that explains the censoring also is required to explain the variation in the dependent variable of primary interest.
- We now turn our attention to the estimation of the model where the selection process is driven by a different index to that generating the dependent variable of primary interest.

- We introduce the following form of the model

$$y_{it}^* = x_{it}'\beta + \alpha_i + \varepsilon_{it}, \tag{36}$$
$$z_{it}^* = x_{it}'\theta_{21} + x_{1it}'\theta_{22} + \xi_i + v_{it}, \tag{37}$$
$$z_{it} = I(z_{it}^* > 0), \tag{38}$$
$$y_{it} = y_{it}^* \cdot z_{it}$$

where we again highlight that the vector x_{1it} is nonempty (and not collinear with x_{it}).

4.1 Maximum likelihood estimators

- Given that we can make distributional assumptions regarding the error components it is natural to construct a maximum likelihood estimator for all the parameters in (36)-(38).

- Consider the case where the individual effect is treated as a random effect and the disturbances are all normally distributed.

- To derive the likelihood function of the vectors z_i and y_i, we first write

$$\log f(z_i, y_i) = \log f(z_i|y_i) + \log f(y_i) \tag{39}$$

- where $f(z_i|y_i)$ is the likelihood function of a conditional T-variate probit model and $f(y_i)$ is the likelihood function of a T_i-dimensional error components regression model, where $T_i = \sum_t z_{it}$.

- The second term can be written as

$$\log f(y_i) = \frac{-T_i}{2}\log 2\pi - \frac{T_i - 1}{2}\log \sigma_\varepsilon^2 - \frac{1}{2}(\sigma_\varepsilon^2 + T_i\sigma_\alpha^2) \qquad (40)$$

$$- \frac{1}{2\sigma_\varepsilon^2}\sum_{t=1}^{T} z_{it}(y_{it} - x_{it}'\beta)^2 - \frac{T_i}{2(\sigma_\varepsilon^2 + T_i\sigma_\alpha^2)}(\bar{y}_i - \bar{x}_i'\beta)^2.$$

- The first term in (39) requires the derivation of the conditional distribution of the error term in the probit model.
- From the assumption of joint normality and defining $\pi_{it} = z_{it}(\alpha_i + \varepsilon_{it})$, the conditional expectation of $v_{it} = \xi_i + \eta_{it}$ is given by

$$E(\xi_i + \eta_{it}|\pi_{i1}, ..., \pi_{iT}) = z_{it}\frac{\sigma_{\varepsilon\eta}}{\sigma_\varepsilon^2}\left[\pi_{it} - \frac{\sigma_\alpha^2}{\sigma_\varepsilon^2 + T_i\sigma_\alpha^2}\sum_{t=1}^{T}\pi_{it}\right] + \frac{\sigma_{\alpha\xi}}{\sigma_\varepsilon^2 + T_i\sigma_\alpha^2}\sum_{t=1}^{T}\pi_{it} \qquad (41)$$

- Using our distributional assumptions the conditional distribution of $\xi_i + \eta_{it}$ given $\pi_{i1}, ..., \pi_{iT}$ corresponds to the unconditional distribution of the sum of three normal variables $e_{it} + \omega_{1i} + z_{it}\omega_{2i}$ whose distribution is characterized by

$$
\begin{aligned}
E(\omega_{1i}) &= E(\omega_{2i}) = 0,\ E(e_{it}) = c_{it} \\
V(e_{it}) &= \sigma_\eta^2 - z_{it}\sigma_{\varepsilon\eta}^2/\sigma_\varepsilon^2 = s_t^2 \\
V(\omega_{1i}) &= \sigma_\xi^2 - T_i\sigma_{\alpha\xi}^2(\sigma_\varepsilon^2 + T_i\sigma_\alpha^2)^{-1} = k_1 \\
V(\omega_{2i}) &= \sigma_{\varepsilon\eta}^2\sigma_\alpha^2\sigma_\varepsilon^{-2}(\sigma_\varepsilon^2 + T_i\sigma_\alpha^2)^{-1} = k_2 \\
\mathrm{cov}(\omega_{1i}, \omega_{2i}) &= -\sigma_{\alpha\xi}\sigma_{\varepsilon\eta}(\sigma_\varepsilon^2 + T_i\sigma_\alpha^2)^{-1} = k_{12},
\end{aligned}
$$

- where the other covariances are all zero and note that we do not explicitly add an index i to the variances s_t^2, k_1 and k_2.

- Similar to the unconditional error components probit model the likelihood contribution can be written as

$$f(z_i|y_i) = \int \int \prod_{t=1}^{T} \Phi \left(d_{it} \frac{x_{it}'\theta_{21} + x_{1it}'\theta_{22} + c_{it} + \omega_{1i} + z_{it}\omega_{2i}}{s_t} \right) \\ \times f(\omega_{1i}, \omega_{2i}) d\omega_{1i} d\omega_{2i}$$

where $d_{it} = 2z_{it} - 1$ and $f(.,.)$ is the density of ω_{1i} and ω_{2i}.

- Using these various expressions it is now possible to construct the complete likelihood function.
- Computation of the maximum likelihood estimator requires numerical integration over two dimensions for all individuals which are not observed in each period.
- Thus the computational demands are reasonably high and as a result this approach has not been proven to be popular in empirical work.

4.2 Two-step estimators

- To present the two-step estimators in the panel setting we follow the approach of Vella and Verbeek (1999).
- We again start with the model presented in (36), (37) and (38).
- Note that although we focus on estimating the above model we retain some degree of generality.
- This allows us to more easily talk about extensions of the above model to alternative forms of censoring.
- The approach that we adopt is a generalization of the Heckman (1979) cross-sectional estimator to the panel data model.

- To motivate a two-step estimator in this setting we begin by conditioning (36) on the vector z_i (and the matrix of exogenous variables X_i) to get

$$E(y_{it}|X_i, z_{i0}, z_i) = x'_{it}\beta + E(u_{it}|X_i, z_{i0}, z_i). \tag{42}$$

- If the mean function of (37) does not depend upon $z_{i,t-1}$ and sample selection only depends on the current value of z_{it}, it is possible to condition only on z_{it} and not $z_i = (z_{i1} \ldots . z_{iT})$ and base estimators on the corresponding conditional moments (see Wooldridge, 1995).

- In this case z_{i0} drops from the conditioning set.
- We assume, as before, that the error terms in the selection equation $v_{it} = \xi_i + \eta_{it}$ exhibit the usual one-way error components structure, with normally distributed components.
- That is

$$v_i|X_i \sim NID(\sigma_\xi^2 u' + \sigma_\eta^2 I).$$

- Note that while we do make explicit distributional assumptions about the disturbances in the main equation we assume

$$E(u_{it}|X_i, v_i) = \tau_1 v_{it} + \tau_2 \overline{v}_i. \tag{43}$$

- Equation (43) implies that the conditional expectation $E(\varepsilon_{it}|X_i, z_{i0}, z_i)$ is a linear function of the conditional expectation of v_{it} and its individual specific mean noting that the $\tau's$ are parameters to be estimated.

- To derive the conditional expectation of the terms on the right hand side of (43) we use

$$E(u_{it}|X_i, z_{i0}, z_i) = \int [\xi_i + E(\eta_{it}|X_i, z_{i0}, z_i, \xi_i)] f(\xi_i|X_i, z_{i0}, z_i) \, d\xi_i, \quad (44)$$

  where $f(\xi_i|X_i, z_{i0}, z_i)$ is the conditional density of ξ_i.

- The conditional expectation $E(\eta_{it}|X_i, z_{i0}, z_i, \xi_i)$ is the usual cross-sectional generalized residual (see Gourieroux et al., 1987, Vella, 1993) from (37) and (38), since, conditional on ξ_i, the errors from this equation are independent across observations.

- The conditional distribution of ξ_i can be derived using the result

$$f(\xi_i|X_i, z_{i0}, z_i) = \frac{f(z_i, z_{i0}|X_i, \xi_i)f(\xi_i)}{f(z_i, z_{i0}|X_i)}, \tag{45}$$

where we have used that ξ_i is independent of X_i and

$$f(z_i, z_{i0}|X_i) = \int f(z_i, z_{i0}|X_i, \xi_i)f(\xi_i)d\xi_i \tag{46}$$

is the likelihood contribution of individual i in (37) and (38).

- Finally

$$f(z_i, z_{i0}|X_i, \xi_i) = \left[\prod_{t=1}^{T} f(z_{it}|X_i, z_{i,t-1}, \xi_i)\right] f(z_{i0}|X_i, \xi_i), \qquad (47)$$

  where $f(z_{it}|X_i, z_{i,t-1}, \xi_i)$ has the form of the likelihood function in the cross-sectional case.

- If we assume that $f(z_{i0}|X_i, \xi_i)$ does not depend on ξ_i, or any of the other error components, then z_{i0} is exogenous and $f(z_{i0}|X_i, \xi_i) = f(z_{i0}|X_i)$.

- Thus we can condition on z_{i0} in (46) and (47) and obtain valid inferences neglecting its distribution. In general, however, we require an expression for the distribution of the initial value conditional on the exogenous variables and the ξ_i.
- The typical manner in which this is done is to follow Heckman (1981) in which the reduced form for z_{i0} is approximated using all presample information on the exogenous variables.

- Thus the two-step procedure takes the following form.
- The unknown parameters in (37) and (38) are estimated by maximum likelihood while exploiting the random effects structure.
- Equation (44) is then evaluated at these ML estimates by employing the expression for the likelihood function in an i.i.d. context, the corresponding generalized residual, and the numerical evaluation of two one dimensional integrals.
- This estimate, and its average over time for each individual provide two additional terms to be included in the primary equation.

- As we noted above, the correction terms have been written to allow greater flexibility with respect to the censoring process.

- We address this issue in the following section. However, as the model in (36), (37) and (38) is perhaps the most commonly encountered for panel data models with selectivity it is useful to see the form of the correction terms.

- The first step is to estimate the model by random effects probit to obtain estimates of the $\theta_2' s$ and the variances σ_ξ^2 and σ_η^2.
- We then compute (44) and its individual specific average after inserting the following terms

$$E(\eta_{it}|X_i, z_{i0}, z_i, \xi_i) = d_{it}\sigma_\eta \frac{\phi\left(\frac{x_{it}'\theta_{21}+x_{1it}'\theta_{22}+\xi_i}{\sigma_\eta}\right)}{\Phi\left(d_{it}\frac{x_{it}'\theta_{21}+x_{1it}'\theta_{22}+\xi_i}{\sigma_\eta}\right)}, \qquad (48)$$

where ϕ denotes the standard normal density function, and

$$f\left(\xi_i|X_i, z_{i0}, z_i\right) = \frac{\prod_{t=1}^{T} \Phi\left(d_{it}\frac{x_{it}'\theta_{21}+x_{1it}'\theta_{22}+\xi_i}{\sigma_\eta}\right)\frac{1}{\sigma_\xi}\phi\left(\frac{\xi_i}{\sigma_\xi}\right)}{\int_{-\infty}^{\infty}\prod_{t=1}^{T} \Phi\left(d_{it}\frac{x_{it}'\theta_{21}+x_{1it}'\theta_{22}+\xi_i}{\sigma_\eta}\right)\frac{1}{\sigma_\xi}\phi\left(\frac{\xi_i}{\sigma_\xi}\right)d\xi}, \quad (49)$$

where $d_{it} = 2z_{it} - 1$.

- The model can be estimated by maximum likelihood if we make some additional distributional assumptions regarding the primary equation errors.
- If all error components are assumed to be homoskedastic and jointly normal, excluding autocorrelation in the time-varying components, it follows that (43) holds with $\tau_1 = \sigma_{\varepsilon\eta}/\sigma_\varepsilon^2$ and $\tau_2 = T(\sigma_{\alpha\xi} - \sigma_{\varepsilon\eta}\sigma_\xi^2/\sigma_\varepsilon^2)/(\sigma_\eta^2 + T\sigma_\xi^2)$.
- This shows that τ_2 is nonzero even when the individual effects α_i and ξ_i are uncorrelated. In contrast, the two-step approach readily allows for heteroskedasticity and autocorrelation in the primary equation.

- Moreover, the assumption in (43) can easily be relaxed to, for example:
- $$E(u_{it}|X_i, v_i) = \lambda_{1t}v_{i1} + \lambda_{2t}v_{i2} + ... + \lambda_{Tt}v_{iT}. \tag{50}$$

  By altering equation (43) this approach can be extended to multiple sample selection rules. With two selection rules, $z_{1,it}$ and $z_{2,it}$, say, with reduced form errors $v_{1,it}$ and $v_{2,it}$, respectively, (43) is replaced by

$$E(u_{it}|X_i, v_{1,i}, v_{2,i}) = \tau_{11} v_{1,it} + \tau_{12} \bar{v}_{1,i} + \tau_{21} v_{2,it} + \tau_{22} \bar{v}_{2,i}. \qquad (51)$$

- Computation of the generalized residuals, however, now requires the evaluation of $E\{v_{j,it}|X_i, z_{1,i}, z_{2,i}\}$ for $j = 1, 2$.

- Unless $z_{1,i}$ and $z_{2,i}$ are independent, conditional upon X_i, the required expressions are different from those obtained from (44) and (45) and generally involve multi-dimensional numerical integration

- Alternative Selection Rules
- The first is the extension to panel data of the Tobit type 3 model given by

$$
\begin{aligned}
y_{it}^* &= x_{it}'\beta_1 + \beta_2 z_{it} + u_{it}, \\
z_{it}^* &= x_{it}'\theta_{21} + x_{1it}'\theta_{22} + v_{it}, \\
z_{it} &= z_{it}^* \cdot I(z_{it}^* > 0), \\
y_{it} &= y_{it}^* \cdot I(z_{it}^* > 0).
\end{aligned}
$$

- In this case one sees that the primary equation may or may not have the censoring variable as an endogenous explanatory variable and the censoring equation is censored at zero but observed for positive values.
- In our wage example discussed above, the extension implies that we observe not only whether the individual works but also the number of hours.
- We also allow the number of hours to affect the wage rate.
- For this model we would first estimate the censoring equation by random effects tobit.

- We would then use these estimates, along with the appropriate likelihood contribution and tobit generalized residual, to compute (44) which are to be included in the main equation.
- Note that due to the structure of the model the inclusion of the correction terms accounts for the endogeneity of z_{it} in the main equation.

- A second model of interest is where the z_{it} is observed as an ordinal variable, taking values j for $j = 1, ..., J,$ and where the values of y_{it} are only observed for certain values of $j.$

- In this case, where the dummies denoting the value of z_{it} do not appear in the model, we would conduct estimation in the following way.

- Estimate the censoring equation by random effects ordered probit and then compute the corrections based on (44) accordingly.

- Then estimate the main equation over the subsample for z_{it} corresponding to a specific value and including the correction terms.

- When one wishes to include the dummies denoting the value of z_{it} as additional explanatory variable it is necessary to pool the sample for the different values of z_{it} and include the appropriate corrections.

4.3 Two-Step Estimators with Fixed Effects

- A feature of the two-step estimator discussed above is their reliance on the assumption that the individual effect is random variable and independent of the explanatory variables.
- While the approach proposed by Vella and Verbeek (1999) is somewhat able to relax the latter assumption it is generally difficult to overcome.
- For this reason, as we noted above in the discussion of the censored regression model, it is generally more appealing to treat the individual fixed component of the error term as a fixed effect which may be correlated with the explanatory variables.
- We noted above that the results of Hahn and Newey (2004) would allow one to estimate a fixed effects tobit model and then perform the appropriate bias correction.

- Accordingly, it would be useful to adopt the same approach in the sample selection model and this has been studied by Fernandez-Val and Vella (2005).
- The basic model they study has the form

$$y_{it}^* = x_{it}'\beta + \alpha_i + \varepsilon_{it}, \tag{52}$$

$$z_{it}^* = x_{it}'\theta_{21} + x_{1it}'\theta_{22} + \xi_i + \eta_{it}, \tag{53}$$

$$z_{it} = I(z_{it}^* > 0), \tag{54}$$

$$y_{it} = y_{it}^* \cdot z_{it}, \tag{55}$$

- where the α_i and ξ_i are individual specific fixed effects, potentially correlated with each other and the explanatory variables, and the ε_{it} and ξ_{it} are random disturbances which are jointly normally distributed and independent of the explanatory variables.
- While Fernandez-Val and Vella (2005) consider various forms of the censoring function, such as described in the previous section, we focus here on the standard case where the selection rule is a binary censoring rule.
- The estimators proposed by Fernandez-Val and Vella (2005) are based on the following approach.

- One first estimates the reduced form censoring rule by the appropriate fixed effects procedure.
- Once these estimates are obtained one uses the bias correction approaches outlined in Hahn and Newey (2004) to adjust the estimates. With these bias corrected estimates one then computes the appropriate correction terms which generally correspond to the cross-sectional generalized residuals. One then estimates the main equation, (52), by a linear fixed effects procedure and bias correct the estimates.
- Fernandez-Val and Vella (2005) study the performance of this procedure to a range of models for alternative forms of censoring. These include the static and dynamic binary selection rule, and the static and dynamic tobit selection rule.

4.4 Semi-Parametric Sample Selection Models

- Kyriazidou (1997) also studied the model in (52)–(55).
- Her approach is semi-parametric in the sense that no assumptions are placed on the individual specific effects α_i and ξ_i and the distributional assumptions on the transitory errors ε_{it} and η_{it} are weak.

- It is clear that $(\theta_{21}, \theta_{22})$ can be estimated by one of the methods for estimation of discrete choice models with individual specific effects, such as Rasch's (1960, 1961) conditional maximum likelihood estimator, Manski's (1987) maximum score estimator or the smoothed versions of the conditional maximum score estimator.
- Kyriazidou's insight into estimation of β combines insights from the literature on the estimation of semi-parametric sample selection models (see Powell, 1987) with the idea of eliminating the individual specific effects by differencing the data.
- Specifically, to difference out the individual specific effects α_i, one must restrict attention to time periods s and t for which y is observed.

- With this "sample selection", the mean of the error term in period t is

$$\lambda_{it} = E(\varepsilon_{it} | \eta_{it} > -x'_{it}\theta_{21} - x'_{1it}\theta_{22} - \xi_i, \eta_{is} > -x'_{is}\theta_{21} - x'_{1is}\theta_{22} - \xi_i, \zeta_i)$$

where $\zeta_i = (x_{is}, x_{1is}, x_{it}, x_{1it}, \alpha_i, \xi_i)$.

- The key observation in Kyriazidou (1997) is that if $(\varepsilon_{it}, \eta_{it})$ and $(\varepsilon_{is}, \eta_{is})$ are independent and identically distributed (conditional on $(x_{is}, x_{1is}, x_{it}, x_{1it}, \alpha_i, \xi_i)$), then for an individual i, who has $x'_{it}\theta_{21} + x'_{1it}\theta_{22} = x'_{is}\theta_{21} + x'_{1is}\theta_{22}$,

$$\begin{aligned}\lambda_{it} &= E(\varepsilon_{it}|\,\eta_{it} > -x_{it}'\theta_{21} - x_{1it}'\theta_{22} - \xi_i, \zeta_i) \\ &= E(\varepsilon_{is}|\,\eta_{is} > -x_{is}'\theta_{21} - x_{1is}'\theta_{22} - \xi_i, \zeta_i) \\ &= \lambda_{is}.\end{aligned} \tag{56}$$

- This implies that for individuals with $x_{it}'\theta_{21} + x_{1it}'\theta_{22} = x_{is}'\theta_{21} + x_{1is}'\theta_{22}$, the same differencing that will eliminate the fixed effect will also eliminate the effect of sample selection.

- This suggests a two–step estimation procedure similar to Heckman's (1976, 1979) two–step estimator of sample selection models: first estimate $(\theta_{21}, \theta_{22})$ by one of the methods mentioned earlier, and then estimate β by applying OLS to the first differences, but giving more weight to observations for which $(x_{it} - x_{is})' \widehat{\theta}_{21} + (x_{1it} - x_{1is}) \widehat{\theta}_{22}$ is close to zero:

$$\hat{\beta}_2 =$$

$$\left[\sum_{i=1}^{n}\sum_{s<t}\left(x_{it}-x_{is}\right)'\left(x_{it}-x_{is}\right)K\left(\frac{\left(x_{it}-x_{is}\right)'\widehat{\theta}_{21}+\left(x_{1it}-x_{1is}\right)\widehat{\theta}_{22}}{h_n}\right)y_{it}y_{is}\right]^{-1}$$

$$\times\left[\sum_{i=1}^{n}\sum_{s<t}\left(x_{it}-x_{is}\right)'\left(x_{it}-x_{is}\right)K\left(\frac{\left(x_{it}-x_{is}\right)'\widehat{\theta}_{21}+\left(x_{1it}-x_{1is}\right)\widehat{\theta}_{22}}{h_n}\right)y_{it}y_{is}\right]$$

where K is a kernel and h_n is a bandwidth which shrinks to zero as the sample size increases.

- Kyriazidou (1997) showed that the resulting estimator is $\sqrt{n}$–consistent and asymptotically normal. Kyriazidou (2001) shows how the same approach can be used to estimate models when lagged dependent variables are included as explanatory variables in (52) or (53).
- As pointed out in Honoré and Kyriazidou (2000), the estimators proposed in Honoré (1992) and Kyriazidou (1997) can be modified fairly trivially to cover static panel data versions of the other tobit–type models discussed in Amemiya (1985).

4.5 Semi-parametric Estimation of a Type-3 Tobit Model

- One paper which explores the semi-parametric estimation of panel data models with a tobit type censoring rule is Lee and Vella (2005). To present this idea first consider the cross-sectional estimator they propose.

- They consider the following model:

$$y_i = x_i'\beta + u_i, \tag{57}$$

$$z_i^* = x_{it}'\theta_{21} + x_{1it}'\theta_{22} + v_i \tag{58}$$

$$z_i = \max(0, z_i^*), \quad s_i = I(z_i > 0), \tag{59}$$

$$(x_i',\ z_i,\ s_i y_i)' \text{ is observed, i.i.d. across } i. \tag{60}$$

  and impose the following mean independence assumption $E(u_i|v_i, x_i, s_i) = E(u_i|v_i, s_i)$.

- The approach to obtain consistent estimates of β is to purge the (57) equation of the component related to the selection equation (58) error.
- To do this they suggest a Robinson (1988) type procedure in which they regress $y_i - E(y_i|v_i, s_i = 1)$ on $x_i - E(x_i|v_i, s_i = 1)$ noting the inclusion of v_i in the conditioning set eliminates the source of the selection problem.
- The model is semi-parametric in that one does not make distributional assumptions about the disturbances.
- Rather, one estimates the selection model (58)-(59) parameters by some appropriate semi-parametric estimator and the estimates $\widehat{v}_i$ as $z_i - x_{1i}\widehat{\theta}_{21} - x'_{2i}\widehat{\theta}_{22}$ (if $s_i = 1$), where the $\widehat{\theta}_{21}$ and $\widehat{\theta}_{22}$ denote the first step semi-parametric estimates.

- The expectations $E(y_i|v_i, s_i = 1)$ and $E(x_i|v_i, s_i = 1)$ can be estimated non-parametrically.
- Lee and Vella (2005) argue that this approach can be extended to additional forms of endogeneity and selectivity by simply including the appropriate reduced form residual(s) in the conditioning set.
- This type of estimator is useful in the two wave panel context and Lee and Vella consider two models which adopt alternative strategies for dealing with dynamics in the model.
- The first is where the lagged dependent variable appears in the conditional mean and the model has the following form:

$$
\begin{aligned}
&y_{it} = y_{i,t-1}\beta_y + x_{it}'\beta + u_{it}, \\
&z_{it}^* = x_{it}'\theta_{21} + x_{1it}'\theta_{22} + v_{it} \\
&z_{it} = \max(0, z_{it}^*), \quad s_{it} = I(z_{it} > 0), \quad t = 1, 2, \qquad (61) \\
&\quad (x_{i1}', x_{i2}',\ z_{i1}, z_{i2},\ s_{i1}y_{i1}, s_{i2}y_{i2})' \text{ is observed, i.i.d. across } i.
\end{aligned}
$$

- The outcome equation can only be estimated over the subpopulation $s_{i1} = s_{i2} = 1$, which poses a double selection problem.

- Thus one estimates over this subsample after subtracting off the component of the outcome equation related to the two selection residuals.
- The mean independence condition assumption required is $E(u_{i2}|v_{i1}, v_{i2}, x_{i2}, y_{i1}, s_i) = E(u_{i2}|v_{i1}, v_{i2}, s_i)$ and one estimates

$$y_{i2} - E(y_{i2}|v_{i1}, v_{i2}) = [y_{i1} - E(y_{i1}|v_{i1}, v_{i2})]\beta_y + [x_{i2} - E(x_{i2}|v_{i1}, v_{i2})]'\beta + \epsilon$$

over the subsample corresponding to $s_{i1} = s_{i2} = 1$.

- Lee and Vella also consider the treatment of dynamics through the inclusion of a time invariant individual fixed effect α_i. The main equation is static and is of the form:

$$y_{it} = x_{it}'\beta + \alpha_i + \varepsilon_{it}.$$

- The double selection problem arises if the first-differenced outcome equation is estimated to eliminate a time-constant error which is potentially related to x_{it}'s:

  $$\Delta y_i = \Delta x_i' \beta + \Delta \varepsilon_i, \quad \Delta y_i \equiv y_{i2} - y_{i1}, \ \Delta x_i \equiv x_{i2} - x_{i1}, \ \Delta \varepsilon_i \equiv \varepsilon_{i2} - \varepsilon_{i1}.$$

  The mean independence assumption required is is $E(\Delta \varepsilon_i | v_{i1}, v_{i2}, \Delta x_i, s_i) = E(\Delta \varepsilon_i | v_{i1}, v_{i2}, s_i)$ and one estimates

- $$\Delta y_i - E(\Delta y_i | v_{i1}, v_{i2}) = [\Delta x_i - E(\Delta x_i | v_{i1}, v_{i2})]' \beta + \epsilon$$

- over the subsample corresponding to $s_{i1} = s_{i2} = 1$.