Estimating and testing simultaneous equation panel data models with censored endogenous variables

Vella, F.; Verbeek, M.J.C.M.

Publication date:
1993

ntER
for
omic Research

Discussion paper

[redacted]

TILBURG UNIVERSITY

Center
for
Economic Research

No. 9359

ESTIMATING AND TESTING SIMULTANEOUS EQUATION PANEL DATA MODELS WITH CENSORED ENDOGENOUS VARIABLES

by Francis Vella and Marno Verbeek

October 1993

ISSN 0924-7815

Estimating and Testing Simultaneous Equation Panel Data Models with Censored Endogenous Variables

Francis Vella
Department of Economics
Rice University
Houston, TX

and

Marno Verbeek
CentER and Department of Econometrics
Tilburg University
The Netherlands

September 1992
This Version September 1993

Abstract

This paper presents a two-step approach to estimating simultaneous equation panel data models with censored endogenous variables. Our procedure employs the residuals from the reduced form estimation of the endogenous variables to adjust for the unobserved heterogeneity in the primary equation. The panel nature of the data allows the identification of three forms of endogeneity. Although our approach is applicable to a much wider family of models, it is closely related to the existing instrumental variables procedures. Accordingly, we explore the relationship of our estimator with the appropriate IV estimators. Furthermore, as our approach employs distributional assumptions, we introduce appropriate tests of specification. Finally, we present an empirical example featuring the estimation of the wage/hours profile for young women. This example illustrates several valuable aspects of our procedure.

This paper was partially written while the authors were visitors in the Department of Economics, Research School of Social Sciences and the Department of Statistics, The Faculties at the Australian National University, Canberra. Helpful comments by Bertrand Melenberg, Robin Sickles and Jeffrey Wooldridge are gratefully acknowledged. We alone are responsible for any remaining errors.

1. Introduction

Economists are increasingly employing panel data to estimate economic relationships. Empirical studies include, for example, labor supply (MaCurdy (1981)), the cyclicality of real wages (Keane, Moffitt and Runkle (1988)), consumption patterns (Nijman and Verbeek (1992)), and retirement behavior (Sickles and Taubman (1986)). However, despite the popularity of panel data investigations the available estimators for simultaneous panel systems containing censored endogenous explanators is limited. Moreover, while there is a substantial literature on the treatment of endogenous regressors in panel data (see for example Hausman and Taylor (1981), Amemiya and MaCurdy (1986) and Breusch, Mizon and Schmidt (1989)) relatively few papers discuss sample selection and attrition bias (see Hausman and Wise (1979), Ridder (1990), Verbeek and Nijman (1992) and Wooldridge (1993)) or the inclusion of censored endogenous regressors (Vella and Verbeek (1993a)). This is a major shortcoming given that these are commonly encountered problems in this area of estimation.

We confront this shortcoming by proposing an estimation procedure for simultaneous equation panel data models with censored endogenous variables. Our general structure comprises a two equation system where the first is a primary equation of major focus, possessing an endogenous explanatory variable, while the second is the reduced form for the explanatory variable. Our approach is to derive explicit estimates of the heterogeneity responsible for the endogeneity and include these constructed variables as additional explanatory variables. We obtain these estimates through a decomposition of the reduced form residuals. Our procedure is applicable to models where the primary equation has an uncensored dependent variable and the endogenous explanator is either uncensored or censored. It is also applicable to primary equations with censored dependent variables and uncensored explanatory variables. We also examine cases where both equations have censored dependent variables. The approach generalizes a number of well known cross sectional two-step estimators to panel data. These include the methods discussed in Heckman (1976, 1979), Amemiya (1984), Smith and Blundell (1986), Rivers and Vuong (1988) and Vella (1993).

We allow three forms of heterogeneity to operate through the two equations. We decompose the error terms into individual, time, and individual/time specific random effects and allow for cross equation correlation between errors of the same dimension. For generality we initially treat all the errors as random effects. However, for many models, particularly when the number of periods is small, it may be more realistic to treat the time effects as fixed. We discuss how to adapt our procedure accordingly.

In estimating models with censored or limited dependent variables, we employ the normality assumption to construct the likelihood function. Normality is also used in deriving the conditional expectations of the random effects. As the consistency of the estimates relies on correct distributional assumptions we propose tests of our assumption of normality. As heteroskedasticity can result in inconsistent estimates in censored and limited dependent variable models we also test the assumption of homoskedasticity. In testing for normality we propose two tests. The first is based on the lagrange multiplier principle. The second borrows an idea employed elsewhere in the diagnostic testing literature (see, e.g., Lee (1984)) and captures departures from normality by suitably powering up the conditional expectations of the random effects from the reduced form equation and including these additional terms in the primary equation. Our test of heteroskedasticity is derived in the lagrange multiplier framework.

Our approach is closely related to the existing instrumental variable estimators available for models with uncensored endogenous variables. Vella and Verbeek (1993b) explore the relationship between the instrumental variables estimates and control function estimates of "treatment" effects in cross sectional models. We show that the estimators proposed in Hausman and Taylor (1981), Amemiya and MaCurdy (1986) and Breusch, Mizon and Schmidt (1989) can be replicated by our "residual" approach.

An issue not confronted in this paper is the relative efficiency loss from estimating the models in a two step manner. Two step estimators are known to be relatively inefficient in comparison to maximum likelihood (see, e.g., Newey (1987)) and the efficiency loss is model specific. We are confident that our procedure is so simple that it will be often preferable to full maximum likelihood methods. If efficiency is considered an important issue, our approach provides initial consistent estimators for a full maximum likelihood approach, so that one iteration in the optimization process is sufficient to attain efficiency[1].

Consistency of the estimators may require either, or both, the number of cross sectional units, denoted by N, and time periods, (denoted T) to be large. Clearly in the treatment of random individual and time effects it is necessary that both N and T are large. In case of fixed time effects, consistency requires large N only.

The following section presents the general model and the estimation procedure. In section three we discuss estimation when the primary equation has a continuous uncensored dependent variable and the regressors are potentially endogenous and censored. Section four discusses estimation of a primary equation with a censored dependent variable and an uncensored endogenous regressor. This extends the conditional maximum likelihood methods of Smith and Blundell (1986) and Rivers and

[1] An alternative method that may be computationally attractive is the use of simulated maximum likelihood, in which the integrals in the log likelihood function are replaced by simulators (cf. Gourieroux and Monfort (1990)).

Vuong (1988) to panel data. In section five we discuss some extensions such as testing for, and estimation in the presence of, selection bias and the introduction of unknown mappings between the dependent variable and the endogenous explanatory variable. Specification test are presented in section six. In section seven we discuss the relationship between our procedure and available instrumental variable estimators. Section eight features an empirical example focusing on the wage/hours relationship and concluding comments are presented in section nine.

2. The general framework

2.1. The model

Consider a general model where equation (1) is our primary focus and equation (2) is the reduced form for the endogenous explanatory variable:

(1) $y^*_{it} = x_{it}'\beta + \psi_1(z_{it}) + \psi_2(z^*_{it}) + D_{it}'\delta + \mu_i + \varepsilon_t + \eta_{it}$ $\quad i=1..N: t=1..T$

(2) $z^*_{it} = m_{it}'\gamma + \alpha_i + \rho_t + \upsilon_{it}$ $\quad i=1..N: t=1..T$

(3) $y_{it} = h(y^*_{it})$

(4) $z_{it} = k(z^*_{it})$

(5) $D_{jit} = I_j(z^*_{it})$ where I_j is an indicator function denoting the occurrence of event j;

where y^*_{it} and z^*_{it} are latent endogenous variables; y_{it} and z_{it} are observed variables produced by the censoring functions h and k; x and m are vectors of exogenous variables; D_{it} is a J dimensional vector of dummy variables indicating that z^*_{it} is in a specified range; ψ_1 and ψ_2 are unspecified functions mapping z_{it} and z^*_{it}, respectively, into y^*_{it}; and β, δ, and γ are parameters to be estimated. We assume that the parameters of the model are identified up to some normalization. To ensure consistent estimation of the covariances between the random effects we assume that both N and T are large.

Each equation's error can be decomposed into individual effects μ_i and α_i; time effects ε_t and ρ_t; and individual specific time effects η_{it} and υ_{it}. These are assumed to be i.i.d. jointly normal with zero mean and variances σ^2_l, $l=\alpha,\mu,\varepsilon,\rho,\nu,\eta$, respectively. Each effect is potentially correlated with its counterpart, of the same dimension, in the other equation. It is this correlation which induces the endogeneity of z_{it}, z^*_{it} and D_{it} in equation (1).[2]

Consider some of the models this framework encompasses. First, it covers the conventional case of an uncensored dependent variable and an uncensored endogenous regressor. Second, it allows the inclusion of a censored dependent variable in either equation or, in special cases, both equations. Thus it includes models with censored endogenous regressors and continuous dependent variables. It also allows a censored dependent variable in the primary equation with an endogenous continuous regressor thus representing an extension to panel data of the conditional maximum likelihood models. Finally, it covers the sample selection model where y_{it} is only observed if $z^*_{it} > 0$ (cf. Nijman and Verbeek (1992)).

A key aspect is the structure of the error terms. The endogeneity of z^*_{it}, z_{it} and D_{it} operates through the three factors common across the two equations. Another feature is the inclusion of either z_{it} or z^*_{it}, as discussed in Blundell and Smith (1992) and Vella (1993), in the primary equation. This is a potentially important issue as the appropriate economic trade off may be between y_{it} and z^*_{it} rather than y_{it} and z_{it}. Also we do not constrain ψ_1 and ψ_2 to be linear parameters thereby allowing a non-linear mapping from z_{it}, or z^*_{it}, to y^*_{it}.

2.2. The estimator

To derive an appropriate estimator equation (1) is first conditioned on the observed outcomes. For cross sectional models this is done with respect to the observed value for z_{it} in the relevant time period. However, given the presence of both individual and time effects, it is necessary for their identification to condition on the vector of *all* outcomes. This also increases the efficiency of the estimator as it incorporates additional information into the procedure. Also, by conditioning on the whole vector of z_{it}'s we are able to consider any subsample based on values of z_{it} without affecting the consistency of our estimator. This is useful when estimating over subsamples or considering selectivity bias issues. Define the NT vector of outcomes of z_{it} as **Z**. Conditioning[3] on **Z** produces the conditional expectation of (1)

$$(6)\quad E\{y^*_{it}|\mathbf{Z}\} = x_{it}'\beta + E\{\psi_1(z_{it})|\mathbf{Z}\} + E\{\psi_2(z^*_{it})|\mathbf{Z}\} + E\{D_{it}|\mathbf{Z}\}'\delta + E\{\mu_i|\mathbf{Z}\} + E\{\varepsilon_t|\mathbf{Z}\} + E\{\eta_{it}|\mathbf{Z}\}.$$

[2] It is straightforward to show that consistent estimation of $\sigma_{\mu\alpha}$ requires N to be large while $\sigma_{\rho\varepsilon}$ requires large T. Consistency of $\sigma_{\eta\upsilon}$ requires either N and/or T to be large.

[3] All conditional expectations that follow are also conditional upon the exogenous variables in x_{it} and m_{it}.

As the conditional expectations of the z_{it} and the indicator functions are themselves we rewrite (6) as

$$(7)\quad y^*_{it} = x_{it}'\beta + \psi_1(z_{it}) + E\{\psi_2(z^*_{it})|\mathbf{Z}\} + D_{it}'\delta + E\{\mu_i|\mathbf{Z}\} + E\{\varepsilon_t|\mathbf{Z}\} + E\{\eta_{it}|\mathbf{Z}\} + e_{it}$$

where e_{it} is a zero mean error term orthogonal to the explanatory variables by construction included to present the estimable form of the relationship. If h is the identity mapping it is straightforward to estimate the parameters from (7) given consistent estimates of the conditional expectations. Where h is not the identity mapping, the conditional distribution of e_{it} will usually be required for drawing inferences (see Section 4 below).

We proceed by obtaining $E\{\mu_i|\mathbf{U}\}$, $E\{\varepsilon_t|\mathbf{U}\}$ and $E\{\eta_{it}|\mathbf{U}\}$ where $u_{it}=\alpha_i+\rho_t+\upsilon_{it}$ and $\mathbf{U}$ is the NT vector of u_{it}'s. Subsequently we take expectations with respect to $\mathbf{U}$ given $\mathbf{Z}$ noting that only the second iteration of the expectations is influenced by the censoring function k. It is reasonably straightforward, due to our assumption of joint normality, to show that the first round of expectations results in (8), (9) and (10)

$$(8)\quad E\{\mu_i|\mathbf{U}\} = \sigma_{\mu\alpha}\left\{\frac{T}{\sigma^2_\upsilon + T\sigma^2_\alpha}\bar{u}_{i.} - \frac{\sigma^2_\rho NT}{(\sigma^2_\upsilon + T\sigma^2_\alpha)(\sigma^2_\upsilon + T\sigma^2_\alpha + N\sigma^2_\rho)}\bar{u}_{..}\right\}$$

$$(9)\quad E\{\varepsilon_t|\mathbf{U}\} = \sigma_{\varepsilon\rho}\left\{\frac{N}{\sigma^2_\upsilon + N\sigma^2_\rho}\bar{u}_{.t} - \frac{\sigma^2_\alpha NT}{(\sigma^2_\upsilon + T\sigma^2_\alpha)(\sigma^2_\upsilon + T\sigma^2_\alpha + N\sigma^2_\rho)}\bar{u}_{..}\right\}$$

$$(10)\quad E\{\eta_{it}|\mathbf{U}\} = \sigma_{\eta\upsilon}\left\{\frac{1}{\sigma^2_\upsilon}u_{it} - \frac{T\sigma^2_\alpha}{\sigma^2_\upsilon(\sigma^2_\upsilon + T\sigma^2_\alpha)}\bar{u}_{i.} - \frac{N\sigma^2_\rho}{\sigma^2_\upsilon(\sigma^2_\upsilon + N\sigma^2_\rho)}\bar{u}_{.t} + \frac{T\sigma^2_\alpha}{\sigma^2_\upsilon + T\sigma^2_\alpha}\,\frac{N\sigma^2_\rho}{\sigma^2_\upsilon + N\sigma^2_\rho}\,\frac{2\sigma^2_\upsilon+T\sigma^2_\alpha+N\sigma^2_\rho}{\sigma^2_\upsilon(\sigma^2_\upsilon + T\sigma^2_\alpha + N\sigma^2_\rho)}\bar{u}_{..}\right\}$$

where $\bar{u}_{..} = (1/NT)\,\Sigma^T_{t=1}\Sigma^N_{i=1}u_{it}$; $\bar{u}_{.t} = (1/N)\,\Sigma^N_{i=1}u_{it}$ and $\bar{u}_{.t} = (1/T)\,\Sigma^T_{t=1}u_{it}$.

To compute the conditional expectations given $\mathbf{Z}$ requires the conditional expectation of the error term from (2). This is given by $E\{u_{it}|\mathbf{Z}\}$ and depends on the censoring function k. When k is the identity mapping $E\{u_{it}|\mathbf{Z}\}$ is equal to u_{it}. However, in general the conditional expectation is more complicated. As T is small for most panel data studies, it is an attractive solution to condition upon the time effects in u_{it}, i.e. to treat the time effects in (2) as fixed unknown parameters. The appropriate expressions for the conditional expectations are then given by (8) and (10), with the terms involving σ_ρ^2 set to zero, while (9) becomes

$$(9') \quad E\{\varepsilon_t|\mathbf{U}, \rho\} = \sigma_{\varepsilon\rho}/\sigma_\rho^2\, \rho_t.$$

It may also be appropriate from an economic viewpoint to treat the time effects in (1) as fixed. Moreover, this decreases the difficulty in estimation as the fixed effects can be captured through time specific dummies. This makes the approach robust to incorrect specification of the distribution of the time effects and also relaxes the requirement for T to be large. However, the estimated fixed time effects in equation (1) will now comprise the direct effect of time and the indirect effect of time through the endogeneity of z_{it}, z^*_{it}, or D_{it}. Naturally this treatment does not allow us to identify the correlation between the time effects. Given the appeal of this alternative specification for the time effects we occasionally focus on this approach below.

Conditioning on the time effects ρ_t allows us to express the conditional expectations of u_{it} as

$$(11) \quad E\{u_{it}|\mathbf{Z}, \rho\} = E\{u_{it}|Z_i, \rho\}$$

where Z_i is the T vector of outcomes for individual i. Express (11) as

$$(12) \quad E\{u_{it}|Z_i,\rho\} = \rho_t + \int [\alpha_i + E\{\upsilon_{it}|z_{it},\rho,\alpha_i\}]\, f(\alpha_i|Z_i,\rho)d\alpha_i$$

where $f(\alpha_i|Z_i,\rho)$ denotes the conditional density of α_i. The conditional expectation of υ_{it} given z_{it}, ρ and α_i is the usual generalized residual from (2) as, conditional on ρ and α, the errors from (2) are independent across observations. Depending on the form of k we replace this expectation with the appropriate generalized residual (see for example Gourieroux et. al (1987), Pagan and Vella (1989) and Vella (1993)).

The more difficult part of (12) is the conditional distribution of α_i given Z_i and ρ. We derive this from the joint distribution of Z_i and α_i, conditional on a given ρ using the result that

$$(13)\quad f(\alpha_i|Z_i,\rho) = \frac{f(Z_i|\alpha_i,\rho)\ f(\alpha_i|\rho)}{f(Z_i|\rho)}$$

where $f(Z_i|\rho) = \int f(Z_i|\alpha_i,\rho)\ f(\alpha_i|\rho)\ d\alpha_i$, which is the likelihood contribution of individual i in (2), conditional on ρ. Furthermore, $f(\alpha_i|\rho) = f(\alpha_i)$ and $f(Z_i|\alpha_i,\rho)$ are the conditional likelihood contributions of individual i given α_i and ρ. Since the errors in (2) are independent across i and t, conditional on individual and time effects, $f(Z_i|\alpha_i,\rho)$ is the product of T individual contributions. This can be written

$$(14)\quad f(Z_i|\alpha_i,\rho) = \prod_{t=1}^{T} f(z_{it}|\alpha_i,\rho)$$

where $f(z_{it}|\alpha_i,\rho)$ has the form of the likelihood contribution in the cross sectional case. Thus the general expressions necessary to compute the conditional expectations in equation (7) are now available. The form of the expressions, and the method of estimation, are functions of h and k as they determine the likelihood functions and generalized residuals. The following two sections consider some interesting cases for h and k.

3. Estimation with a continuous dependent variable and a censored endogenous regressor

This section discusses cases where the dependent variable y_{it} in the primary equation is uncensored. First, consider where the explanatory variable appearing from equation (2) is endogenous but uncensored. This is characterized by the following assumptions:

Case 1:a) $y=y^*$; b) $z=z^*$; c) ψ_1 is known; d) ψ_2 and δ set equal to zero

This is the simplest case due to the observability of y^*_{it} and z^*_{it} and the assumption that ψ_1 is known up to a finite number of unknown parameters. Assumptions 1c and 1d are imposed to avoid unnecessary complications.[4] Equation (2) is estimated by random effects maximum likelihood or alternative methods producing consistent estimates. With the estimates of γ, σ_α, σ_ρ and σ_υ expressions (8), (9) and (10) can be computed. These terms are included as additional explanatory variables in the primary equation which is estimated by ordinary or generalized least squares. This generates

[4] These assumptions are dropped in the treatment of later cases and in Section 5.2.

consistent estimates of the parameter vectors β and ψ and the covariances (i.e $\sigma_{\alpha\mu}$, $\sigma_{\rho\varepsilon}$ and $\sigma_{\nu\eta}$).

The distribution of e_{it} in (7) corresponds with the *conditional* distribution of $\mu_i + \varepsilon_t + \eta_{it}$ in (1) *given* **Z**. If z_i^* is observed, e_{it} will be normal and have the same error components structure, but with the variances reflecting conditional variances (see Section 4 below). Unlike the original specification these components will be uncorrelated with the regressors. Incorporating this error structure in the second step estimation will increase efficiency. After replacing the expectations in (7) with their computed values, the error term for estimation purposes is $\nu_{it} = e_{it} + \sigma_{\mu\alpha}(\mu_i - \hat{\mu}_i) + \sigma_{\varepsilon\rho}(\varepsilon_t - \hat{\varepsilon}_t) + \sigma_{\eta\upsilon}(\eta_{it} - \hat{\eta}_{it})$ where the ^'s denote the values using the estimated values from the first step random effects model. Under the null hypothesis that the three covariances are zero, routinely computed standard errors are correct, and consequently the individual t-tests on these terms are tests of weak exogeneity. In general, however, standard errors should be adjusted for heteroskedasticity and for the additional terms being generated regressors (cf. Newey (1984) and Pagan (1986)). This is discussed in the appendix.

Note that the function ψ_1 need not be linear. Irrespective of the functional form employed the conditional expectations needed to correct for endogeneity are unchanged. This allows one to specify and estimate a flexible relationship between y_{it} and z_{it} and makes the estimation procedure less vulnerable to the normality assumption of the errors in the reduced form equation. Instead of z_{it} any one-to-one transformation of this variable may be employed in equation (2).

If the endogeneity of z_{it} only operates through the individual specific effects the existing instrumental variables techniques of Hausman and Taylor (1981), Amemiya and MaCurdy (1986) and Breusch, Mizon and Schmidt (1989) can be employed. In this case, and where z_{it} is uncensored, our estimation method will produce identical results to these instrumental variable methods. We feel that the equivalence of the instrumental variable and "residual" approaches is of interest and pursue this issue in section seven. Having noted the equivalence in this simpler model we feel that our approach has substantial attraction. The primary advantage of our approach over the instrumental variables is that it allows any form of censoring of z_{it}. Moreover, it can directly identify different sources of endogeneity[5].

Case 2: a) $y=y^*$; c) ψ_1 is known; d) $\psi_2=0$, $\delta=0$

Now consider where k is a mapping generating a censored z_{it}. While models of this form have appeared for cross sectional data (see, for example, Heckman (1976,

[5] The instrumental variables procedures can identify the form of endogeneity through a series of Hausman tests.

1978), and Vella (1993)) they are much less common in a panel data setting[6]. For illustrative purposes, consider a form of k such that

$$(15)\quad \begin{array}{l} z_{it}=1 \text{ if } z^*_{it}>0 \\ z_{it}=0 \text{ otherwise.} \end{array}$$

Equation (2) can now be estimated by random effects probit maximum likelihood. The generalized residual is given by

$$(16)\quad E\{\upsilon_{it}|Z_i,\rho,\alpha_i\}= (2z_{it}-1)\sigma_\upsilon\{\phi(b_{it})/\Phi(b_{it})\}$$

where $b_{it}=(2z_{it}-1)(m_{it}'\gamma+\alpha_i+\rho_t)/\sigma_\upsilon$ and ϕ and Φ denote the probability density and cumulative density functions of the standard normal distribution. Some normalization regarding the variances in the probit model is required. An obvious choice is to set the total variance of the error for any observation to one, while conditioning on the time effects (i.e. including time dummies), that is, set $\sigma^2_\upsilon+\sigma^2_\alpha=1$. To compute the expectations of the random effects we use

$$(17)\quad f(z_{it}|\alpha_i,\rho)= \Phi(b_{it}).$$

The form of censoring shown in (15) is simple. However, the method of estimation is easily adapted for alternative censoring functions. Suppose the censoring is given by (18)

$$(18)\quad \begin{array}{l} z_{it}=z^*_{it} \text{ if } z^*_{it}>0 \\ z_{it}=0 \text{ otherwise.} \end{array}$$

The first step estimates are obtained by random effects tobit maximum likelihood, and the generalized residual is given by

$$(19)\quad E\{\upsilon_{it}|Z_i,\rho,\alpha_i\} = \begin{cases} -\sigma_\upsilon\{\phi(b_{it})/\Phi(b_{it})\} & \text{if } z_{it}=0 \\ z_{it}-m_{it}'\gamma-\alpha_i-\rho_t & \text{if } z_{it}>0. \end{cases}$$

The term corresponding to (17) is

---

[6] Two exceptions are Keane, Moffitt and Runkle (1988) and Vella and Verbeek (1993a))

$$(20)\quad f(z_{it}|\alpha_i,\rho) = \Phi(b_{it})^{I(z^*_{it}<0)}[\sigma_\upsilon^{-1}\phi\{(z_{it}-m_{it}'\gamma-\alpha_i-\rho_t)/\sigma_\upsilon\}]^{I(z^*_{it}\geq 0)},$$

which represents the individual contribution to the standard tobit likelihood. Thus estimation is relatively straightforward although numerical integration is required (see, for example, Butler and Moffitt (1982)), as in the probit case, for the estimation of the first step parameters and the computation of the conditional expectations. The second step is again estimated by generalized least squares.

Our approach is sufficiently general to carry over to other forms of censoring. The ordered probit is a particularly interesting case on which we focus in case 3.

Case 3: a) $y=y^*$; b) ψ_1 and ψ_2 =0

Consider the following form of censoring

$$(21)\quad z_{it}= 0 \text{ if } \lambda_0 \leq z^*_{it} < \lambda_1;\ z_{it}=1 \text{ if } \lambda_1 \leq z^*_{it} < \lambda_2;\ z_{it}=2 \text{ if } \lambda_2 \leq z^*_{it} < \lambda_3;\ldots$$
$$z_{it}= m\text{-}1 \text{ if } \lambda_{m-1} \leq z_{it} < \lambda_m$$

where the λ's represent separation points and it is assumed λ_0 and λ_m are equal to $-\infty$ and $+\infty$, respectively, and $\lambda_1=0$. Thus, z_{it} is an ordinal variable retaining the ordering in z^*_{it}. The dummy variables in (5) are generated in the following manner

$$(22)\quad D_{jit}=1 \text{ if } \lambda_{j-1} < z^*_{it} \leq \lambda_j.$$

Estimation of equation (2) with censoring rule (21) can be performed by random effects ordered probit. The corresponding conditional generalized residual has the form

$$(23)\quad E\{\upsilon_{it}|Z_i,\rho,\alpha_i\} = \sigma_\upsilon\{[\phi(\lambda_{j-1}-c_{it})-\phi(\lambda_j-c_{it})]/[\Phi(\lambda_j-c_{it})-\Phi(\lambda_{j-1}-c_{it})]\}$$

if $D_{jit}=1$ and where $c_{it}=(m_{it}'\gamma+\alpha_i+\rho_t)/\sigma_\upsilon$. The corresponding conditional probability mass function is given by

$$(24)\quad f(z_{it}|\alpha_i,\rho)= \Phi(\lambda_j-c_{it})-\Phi(\lambda_{j-1}-c_{it}) \text{ for } \lambda_{j-1} < z^*_{it} \leq \lambda_j.$$

This is a useful extension as explanatory variables from micro data sets are often recorded in the manner given in (21). An example of this is a time varying vector of dummy variables, denoting the highest level of education acquired, included as explanators in an earnings equation (see Vella and Gregory (1993)).

Case 4: a) $y=y^*$; b) ψ_2 is linear; c) ψ_1 and $\delta=0$

Cases 1 to 3 consider the relationship between the censored explanator and the continuous dependent variable. However, the appropriate relationship may be between y_{it} and z^*_{it}. Equation (6) indicates that the conditional expectation of the latent variable, term $E\{(z^*_{it}|Z\})$, also needs to be computed. This will generally be relatively straightforward and its exact form will depend on k. The steps required are very similar to the previous cases although we now compute an additional conditional expectation.

4. Conditional maximum likelihood panel data estimation

Now consider models with censored dependent variables and uncensored endogenous explanatory variables. For cross sectional data a computationally simple approach is conditional maximum likelihood (see Smith and Blundell (1986) and Rivers and Vuong (1988)). In this method equation (2) is estimated by maximum likelihood and, subsequently, the conditional likelihood function of (1), given z_i and the estimated parameters from (2), is maximized. Provided the errors in the conditional distribution are normal this procedure is straightforward. We now consider the panel data extensions of these models. For simplicity we shall concentrate on the most relevant case where the time effects in both equations are treated as fixed. Consequently, time dummies are included in both mean functions.

Write the joint likelihood of $y_i = (y_{i1},\ldots,y_{iT})'$ and $z_i = (z_{i1},\ldots,z_{iT})'$ given X_i, and where $X_i = (x_{i1},\ldots,x_{it})'$, as

$$f(y_i \mid z_i, X_i, \vartheta_1, \vartheta_2)\, f(z_i \mid X_i, \vartheta_2)$$

where ϑ_2 denotes $(\gamma,\sigma_\alpha,\sigma_\upsilon)$ and ϑ_1 denotes $(\beta,\varphi_1,\varphi_2,\delta,\sigma_\mu,\sigma_\eta,\sigma_{\alpha\mu},\sigma_{\upsilon\eta})$ and the time effects are included in γ and β, respectively. The conditional maximum likelihood approach involves first estimating ϑ_2 by maximizing the marginal likelihood function of the z_i's. This is straightforward given the reduced form in (2). Subsequently, the conditional likelihood function

$$\prod_{i=1}^{N} f(y_i \mid z_i, X_i, \vartheta_1, \hat{\vartheta}_2)$$

is maximized with respect to θ_1. This step depends upon the form of the conditional distribution of y_i given z_i and X_i. When z_i is continuously and completely observed the latter step has the same computational complexity as when z_i is strictly exogenous. We shall illustrate this below.

Case 5: a) $z_{it}=z^*_{it}$; b) ψ_1 is known; c) ψ_2 and $\delta=0$

The primary equation has a censored dependent variable and an endogenous continuous variable appears as the dependent variable in the reduced form equation. As the endogenous regressor is observed $E\{u_{it}|Z\} = E\{u_{it}|U\} = u_{it}$ and the conditional expectations are provided in equations (8) and (10)[7]. Due to the normality assumption the error term e_{it}, given by $y^*_{it} - E\{y^*_{it}|Z\}$, is also normal. Also, given the form of (8) and (10) its distribution retains the same error components structure, although the variances now reflect the conditional variances.

The first step parameters are obtained through random effects estimation and the second step is estimated by maximum likelihood depending on the form of h. Consistent estimates of the unconditional variances σ^2_η and σ^2_μ can be obtained from the estimated covariances and conditional variances. To illustrate this model, consider censoring of the form

$$y_{it} = 1 \text{ if } y^*_{it} > 0$$
$$y_{it} = 0 \text{ otherwise.}$$

The model to be estimated is

$$(25) \quad y^*_{it} = x_{it}'\beta + \psi_1 z_{it} + \mu_i + \eta_{it}$$

$$(26) \quad z_{it} = m_{it}'\gamma + \alpha_i + \upsilon_{it}$$

First, (26) can be estimated by random effects maximum likelihood and with the appropriate conditional expectations one can estimate the following probit model

$$(27) \quad y^*_{it} = x_{it}'\beta + \psi_1 z_{it} + \sigma_{\mu\alpha}\hat{V}_{1i} + \sigma_{\eta\upsilon}\hat{V}_{2it} + \nu_{it}$$

where

[7] With the terms involving σ^2_ρ set to zero.

$$V_{1i} = \frac{1}{\sigma^2_\upsilon + T\sigma^2_\alpha} \sum_{t=1}^{T} u_{it}$$

and

$$V_{2it} = \frac{1}{\sigma^2_\upsilon} u_{it} - \frac{1}{\sigma^2_\upsilon(\sigma^2_\upsilon + T\sigma^2_\alpha)} \sum_{t=1}^{T} u_{it},$$

which are obtained from (8) and (10) after eliminating the time effects in both equations. It can be shown that ν_{it} is a zero mean normally distributed error term with the following covariance properties

$$\mathrm{Var}\{\nu_{it}\} = \mathrm{Var}\{\mu_i + \eta_{it} | u_{i1}, .. u_{iT}\} =$$

$$(\sigma^2_\eta - \sigma^2_{\eta\upsilon}\sigma^{-2}_\upsilon) + (\sigma^2_\upsilon - \frac{T\sigma^2_{\mu\alpha}\sigma^2_\upsilon + 2\sigma_{\mu\alpha}\sigma_{\eta\mu}\sigma^2_\upsilon + \sigma^2_{\eta\upsilon}\sigma^2_\alpha}{\sigma^2_\upsilon(\sigma^2_\upsilon + T\sigma^2_\alpha)}) = \sigma^2_1 + \sigma^2_2,$$

which defines σ^2_1 and σ^2_2, and

$$\mathrm{Cov}\ \{\nu_{it}, \nu_{is}\} = \sigma^2_2, \quad s \neq t.$$

Thus, the error components structure is preserved and the conditional likelihood function of (25) has the same form as the marginal likelihood function without endogenous regressors. The appropriate conditional log likelihood is given by log L= Σ_i log L_i where

$$(28) \quad L_i = \int \prod_{t=1}^{T} \Phi\left([2y_{it}-1] \frac{x_{it}'\beta + \psi\ z_{it} + \sigma_{\mu\alpha}\hat{V}_{1i} + \sigma_{\eta\upsilon}\hat{V}_{2it} + \xi}{\sigma_1}\right) \frac{1}{\sigma_2}\ \phi(\xi/\sigma_2)d\xi$$

and where some normalization should be made with respect to the variances, as in the usual probit model. By maximizing (28) consistent estimates of the parameters from (27) are obtained. While a random effects probit model is estimated here, it is straightforward to accommodate alternative forms of censoring for $\hat{y}_{it}$. As with the previous models, the t-statistics for the coefficients for $\hat{V}_{1i}$ and $\hat{V}_{2it}$ provide tests of endogeneity. In general, however, routinely computed standard errors should be corrected for the two-step character of the estimation procedure.

5. Other applications and extensions

5.1. Selection bias

Several additional models can be estimated in this framework. The first is an extension to panel data of the cross section selection bias procedure of Heckman (1976,1979). Tests for selection bias in panel data models have been suggested by Verbeek and Nijman (1992) and Wooldridge (1993). However, our method can provide consistent estimates in the presence of selection bias or attrition bias (cf. Ridder (1990)).

Selection bias arises when values of y_{it} are only observed when z^*_{it} passes some threshold. Assume that the censoring rule shown in (15) can be again employed but note that the dependent variable in (1) is only observed if $z_{it}=1$. The parameters in equation (2) are estimated over all observations. Having computed the necessary correction terms, one can estimate the second step over the observations for which z_{it} is equal to one. If the censoring generates unbalanced panels the data can be "trimmed" to balance the panels as the exclusion of additional observations does not induce any selection bias. The loss from trimming is a decrease in efficiency. An application of this procedure is given in Nijman and Verbeek (1992).

Selection bias is usually raised in the context of a continuous dependent variable in the primary equation. Recently Vella (1992) has developed a test for selection bias and endogeneity for a cross section in models when z_i and y_i are both censored. Verbeek and Nijman (1992) also suggest a number of tests of selection bias in panel data models which are applicable to models with censored or limited dependent variables. The extension of the approach of Vella (1992) to panel data is relatively straightforward. The first step is the estimation over all NT observations by random effects probit. Then we compute the necessary correction terms and include them in the primary equation. We then estimate this primary equation over the available observations. A joint test on the coefficients of the included variables is a test of endogeneity or selectivity bias.

5.2. Flexible estimation of unknown functions

While we did not specify particular forms for ψ_1 and ψ_2 we assumed above that they were known operators. This assumption can be relaxed through the methodology discussed in Newey, Powell and Vella (1993) where it is shown that various approximations to unknown functions can be easily employed once the endogeneity is accounted for through the inclusion of the reduced form residuals. In the case of normality only the three correction terms discussed above are necessary to account for the endogeneity irrespective of the number of terms employed in the approximation of the unknown functions. Generally, it is only sensible to concentrate on the estimation of ψ_1 in this context. For example, estimating higher order terms for censored variables will usually have no intuitive appeal although it may be possible to compute the conditional higher moments.

5.3. Dynamic specifications

The recent panel data literature has featured an increasing emphasis on correct dynamic specification. This often implies including one or more lagged endogenous variables. In our general approach, particular dynamic specifications can easily be allowed. First, as the conditional expectations in (7) are conditional on the *vector* Z_i (or the NT vector Z), one can easily include a lagged observed variable $z_{i,t-1}$ in (2). Estimation of the first step follows from above and this results in consistent estimates provided the problem of initial conditions can be solved. Unless the initial values z_{i0} are truly exogenous, some ad hoc solution will be required (cf. Heckman (1981)). If k is the identity mapping, consistent estimates for the reduced form can be obtained by instrumental variables techniques.

Second, a lagged dependent variable in the primary equation can be handled in the conditional maximum likelihood framework. Although the conditional distribution of the error given Z_i and the exogenous variables is not changed, deriving the corresponding conditional distrubtion of y_i requires some assumption of the first observation y_{i0}. Unless this first observation can be assumed to be exogenous, the standard conditional maximum likelihood method will not result in consistent estimates and alternative estimation methods (with alternative assumptions on y_{i0}) are to be employed. This is a topic left for future research[8].

6. Diagnostic tests

The previous sections outlined tests for endogeneity and sample selection bias which were byproducts from consistent estimation of the model. We now focus on misspecification which may result in inconsistency. Initially we concentrate on testing the crucial assumptions needed when estimating the censored or discrete choice random effects models. The assumptions on which we focus are those of normality and homoskedasticity. We then discuss tests of normality when the primary equation is estimated by least squares.

---

[8] Some results on the estimation of autoregressive models for panel data with sample selectivity are available in Arellano, Bover and Labeaga (1992).

6.1. Specification tests for discrete and censored dependent variable models

To illustrate the tests we focus on the random effects probit model and note that tests for alternative forms of k will differ in their treatment of the likelihood function. We assume for simplicity that the time effects are fixed and are absorbed into the vectors x and m in equations (1) and (2), respectively. The model is shown in equations (1a) and (2a) and is an example of case 2:

$$\text{(1a)}\quad y_{it} = x_{it}'\beta + \psi_1(z_{it}) + \mu_i + \eta_{it}$$

$$\text{(2a)}\quad z^*_{it} = m_{it}'\gamma + \alpha_i + \upsilon_{it}$$

with censoring given by equation (15). The likelihood function for (2a) under the null hypothesis has the form

$$\text{(29)}\quad \sum_i \log L_i = \sum_i \log \int \prod_{t=1}^{T} \Phi\left(\frac{m_{it}'\gamma + \alpha_i}{\sigma_\upsilon}\right)^{z_{it}} \left(1 - \Phi\left(\frac{m_{it}'\gamma + \alpha_i}{\sigma_\upsilon}\right)\right)^{1-z_{it}} \sigma_\alpha^{-1}\phi\left(\frac{\alpha_i}{\sigma_\alpha}\right) d\alpha_i .$$

Denote the unknown k-dimensional vector of parameters from (29) as θ and their maximum likelihood estimator as $\hat{\theta}$. Define the extended parameter vector, where the additional parameters characterize the departures from the null of correct specification, as ζ. The lagrange multiplier test statistics can be written as

$$\text{(30)}\quad \xi_{LM} = \sum \frac{\partial \log \tilde{L}_i}{\partial \zeta'} \left[\sum \frac{\partial \log \tilde{L}_i}{\partial \zeta} \frac{\partial \log \tilde{L}_i}{\partial \zeta'}\right]^{-1} \sum \frac{\partial \log \tilde{L}_i}{\partial \zeta} \Bigg|_{\zeta=\hat{\zeta}_0}$$

where $\hat{\zeta}_0$ is the ML estimate for ζ under the null hypothesis and $\tilde{L}_i$ is the contribution based on the likelihood function involving the alternative. (30) can be computed as N times the non-centered R^2 from a regression of a vector of ones against the vectors $\partial \log \tilde{L}_i/\partial\zeta_j$, $j=1,\ldots,k+l$, where l is the number of restrictions. Under the null hypothesis ξ_{LM} is asymptotically Chi-squared distributed with l degrees of freedom.

The score contributions with respect to the parameters θ that are unrestricted under the null do not depend upon the alternative involved, i.e. $\partial \log \tilde{L}_i/\partial\theta = \partial \log L_i/\partial\theta$ when evaluated at $\theta = \hat{\theta}$. To derive these scores, denote $f(\alpha_i) = \phi(\alpha_i/\sigma_\alpha)/\sigma_\alpha$ and write the log likelihood contributions as

$$(31) \quad \log L_i = \log \int \prod_t \Phi(b_{it}) f(\alpha_i) d\alpha_i \ .$$

Since[9]

$$(32) \quad \frac{\partial L_i}{\partial \zeta} = \int \sum_{s=1}^{T} \prod_{t=1,\, t\neq s}^{T} \Phi(b_{it}) \frac{\partial \Phi(b_{is})}{\partial \zeta} f(\alpha_i) d\alpha_i$$

we require expressions for $\frac{\partial \Phi(b_{it})}{\partial \zeta}$. Since

$$(33) \quad \frac{\partial \Phi(b_{it})}{\partial \zeta} = \phi(b_{it}) \frac{\partial b_{it}}{\partial \zeta} = \Phi(b_{it}) \frac{\phi(b_{it})}{\Phi(b_{it})} \frac{\partial b_{it}}{\partial \zeta}$$

we can write

$$(34) \quad \frac{\partial L_i}{\partial \zeta} = \int \prod_{t=1}^{T} \Phi(b_{it}) \left(\sum_{s=1}^{T} \lambda(b_{is}) \frac{\partial (m_{it}'\gamma/\sigma_\upsilon)}{\partial \zeta} \right) f(\alpha_i) d\alpha_i$$

where $\lambda(.)$ corresponds to the generalized residual for the standard probit model. As the integral in (34) corresponds to a conditional expectation over α_i given z_i we can write

$$(35) \quad \frac{\partial L_i}{\partial \zeta} = \int \sum_{s=1}^{T} E\{\upsilon_{is} | \alpha_i, z_i\} \frac{\partial (m_{it}'\gamma/\sigma_\upsilon)}{\partial \zeta} f(\alpha_i | z_i) d\alpha_i.$$

The score contribution with respect to the intercept term, for example, is obtained from

$$(36) \quad \frac{\partial L_i}{\partial \gamma_0} = \sigma_\upsilon^{-1} \sum_{s=1}^{T} E\{\upsilon_{is} | z_i\} = \sigma_\upsilon^{-1} \int \sum_{s=1}^{T} E\{\upsilon_{is} | \alpha_i, z_i\} f(\alpha_i | z_i) d\alpha_i.$$

To construct a test of non-normality of υ_{it} we employ a parameterization suggested by Ruud (1984). Assume the distribution function of υ_{it} is

[9] Many of the equalities that follow are valid under the null hypothesis only.

$F(t)=\Phi(\omega_0+t+\omega_1 t^2+\omega_2 t^3)$ where ω_1 and ω_2 satisfy $1+2\omega_1 t+3\omega_2 t^2>0$ for all t. We assume that ω_0 is equal to zero and test for nonnormality through the hypothesis $\omega_1=\omega_2=0$. By replacing (33) with

$$\partial F(b_{it})/\partial\omega_{j-1} = \phi(b_{it})(2z_{it}-1)\left[(m_{it}'\gamma+\alpha_i)/\sigma_\upsilon\right]^j, \ j = 2, 3$$

the scores with respect to ω_1 and ω_2 are easily derived. Note that it is necessary to integrate out α_i.

A second form of mis-specification of interest is that of heteroskedasticity of υ_{it}. For simplicity, assume that the heteroskedasticity is related to the observed and included m_{it}'s only. The general likelihood function now involves the function $g(m_{it}'d)$ instead of σ_υ, where d is a vector of parameters, possibly with some elements set to zero. The score required to test that $g(.)$ is non-constant requires

$$(37)\quad \partial\Phi\left[(m_{it}'\gamma+\alpha_i)/g(m_{it}'d)\right]/\partial d = \kappa\ \phi(b_{it})(2z_{it}-1)(m_{it}'\gamma+\alpha_i)m_{it}$$

where κ is some constant. As in the linear regression case (see Breusch and Pagan (1979)), the LM test does not depend on the form of the function $g(.)$ because the constant κ is irrelevant.

A general feature of these tests, apparent from (35), is that unlike the standard probit model the scores are no longer the product of the generalized residual and a set of exogenous variables. This is due to the presence of the individual effects, α_i.

Our tests of heteroskedasticity and non-normality of υ_{it} are based on the assumption that the distribution of α_i is correctly specified. This assumption can be tested by generalizing the form of $f(.)$ in (31). This changes the form of (32) to

$$(38)\quad \frac{\partial L_i}{\partial\zeta} = \int \prod_{t=1}^{T} \Phi(b_{it}) \frac{\partial f(\alpha_i)}{\partial\zeta}\, d\alpha_i$$

where it is assumed that the order of integration and differentiation can be changed. An interesting alternative is non-normality of α_i. The density function corresponding to the distribution function F(t) above is given by

$$(39)\quad \tau(t) = (1+2\omega_1 t+3\omega_2 t^2)\phi(t+\omega_1 t^2+\omega_2 t^3).$$

Consequently $f(.)$ in (32) is replaced by $\tau(\alpha_i/\sigma_\alpha)/\sigma_\alpha$. The derivatives with respect to

ω_j are given by

$$(40) \quad \frac{\partial f(\alpha_i/\sigma_\alpha)}{\partial \omega_1} = \sigma_\alpha^{-1}\phi(\alpha_i/\sigma_\alpha)(\alpha_i/\sigma_\alpha)(2-(\alpha_i/\sigma_\alpha)^2)$$

and

$$(41) \quad \frac{\partial f(\alpha_i/\sigma_\alpha)}{\partial \omega_2} = \sigma_\alpha^{-1}\phi(\alpha_i/\sigma_\alpha)(\alpha_i/\sigma_\alpha)^2(3-(\alpha_i/\sigma_\alpha)^2).$$

To complete the required scores the derivative with respect to σ_α is given by

$$(42) \quad \frac{\partial f(\alpha_i/\sigma_\alpha)}{\partial \sigma_\alpha} = -\sigma_\alpha^{-2}\left[1-\left(\alpha_i^2/\sigma_\alpha^2\right)\right]\phi(\alpha_i/\sigma_\alpha).$$

Mis-specification of the types discussed above in censored models often represents an insuperable problem. Incorrect distributional assumptions imply that inconsistent parameter estimates will be obtained. Given the relatively complicated structure of the model it is very difficult to relax the distributional assumptions and still proceed although it may be possible to adopt alternative distributions. However, note that any one-to-one function of z_{it} may be explained by the reduced form, because of the flexibility of the functional form ψ_1.

6.2. Testing linearity of the conditional expectations

Now assume that the effects in (2) remain normally distributed but rather than assume joint normality we express the conditional expectations of the components in equation (1) as a function of those in (2). We thus replace the joint normality assumption with

$$(43) \quad E\{\mu_i|\alpha_i+\upsilon_{i1},\ldots.\alpha_i+\upsilon_{iT}\} = r[\sum_{t=1}^{T}(\alpha_i+\upsilon_{it})\ ]$$

and

$$(44) \quad E\{\eta_{it}|\alpha_i+\upsilon_{i1},\ldots.\alpha_i+\upsilon_{iT}\} = q_1(\alpha_i+\upsilon_{it}) + q_2[\sum_{s=1}^{T}(\alpha_i+\upsilon_{is})]$$

where r and q are functions mapping the reduced form errors into the primary equation error. Under joint normality r, q_1 and q_2 are linear mappings with a restriction that identifies r and q_2. To capture departures from normality we follow Lee (1984) and Gallant and Nychka (1987) by multiplying the normal distribution by some suitably chosen polynomial. We capture non-normality in the primary equation effects by including powered up values of the random effects as shown in (44)

$$(45)\quad E\{\mu_i|\alpha_i+\upsilon_{i1},\ldots\alpha_i+\upsilon_{iT}\}=\sum_{j=1}^{J}\lambda_{\mu j}\,[\sum_{t=1}^{T}(\alpha_i+\upsilon_{it})]^j$$

and

$$(46)\quad E\{\eta_{it}|\alpha_i+\upsilon_{i1},\ldots\alpha_i+\upsilon_{iT}\}=\sum_{j=1}^{J}\lambda_{\eta j}\,[\,(\alpha_i+\upsilon_{it})-\frac{\sigma_\alpha^2}{\sigma_\upsilon^2(\sigma_\upsilon^2+T\sigma_\alpha^2)}\sum_{s=1}^{T}(\alpha_i+\upsilon_{is})]^j$$

where the length J can be chosen by cross validation. Thus we estimate the first step by random effects maximum likelihood and then estimate

$$(47)\quad y_{it}=x_{it}'\beta+\psi_1 z_{it}+\lambda_{\mu1}\hat{\mu}_i+\lambda_{\eta1}\hat{\eta}_{it}+\lambda_{\mu2}\hat{\mu}_i^2+\lambda_{\eta2}\hat{\eta}_{it}^2+\ldots\lambda_{\mu J}\hat{\mu}_i^J+\lambda_{\eta J}\hat{\eta}_{it}^J$$

by least squares, where we denote the higher order terms with the subscripts from 2 to J.[10] Note that in this model the higher order terms capture non-normality so t-tests indicating statistical significance represent a rejection of normality[11]

7. Relationship with instrumental variable estimation

Vella and Verbeek (1993b) discuss the relationship between instrumental variables[12] and control function estimation from cross sectional data. It is useful to extend this to panel data for the following reasons. First, it illustrates that the instrumental variables estimates can be obtained from this residual approach. Second, our approach is easily extended to models with censored endogenous variables while the instrumental

[10] Newey et. al (1993) discusses the rate at which J should grow with the sample size.

[11] Strictly speaking, the standard t-tests are only valid in the case where J is kept fixed.

[12] We use the term "instrumental variables" estimator for the linear IV estimator, cf. Amemiya (1985, p. 239). Following Newey (1987) we could refer to the estimators of Section 3 as two-stage instrumental variables estimators.

variables methods are not. Third, our modeling approach produces tests of endogeneity as a by-product and does not require any additional testing procedures. Also, we allow simultaneous identification of the sources of the endogeneity. Finally, our approach is immediately extended to sample selection type models.

While we feel our procedure is more appealing than the instrumental variable approach it is important to establish they produce identical results for models where the instrumental variables approach is applicable. We discuss the relationship of our estimators with the instrumental variables estimators proposed by Hausman and Taylor (1981) and Amemiya and MaCurdy (1986). These papers consider linear models with uncensored endogenous regressors. Furthermore, the endogeneity operates only through the individual effects. While we now discuss the relationship between these competing estimators the exact nature of the relationship is given in a theorem in the appendix. The model is given by

$$(48) \quad y_{it} = x_{it}'\beta + z_{it}\psi_1 + \mu_i + \eta_{it}$$

where there is only one endogenous variable, z_{it}. Furthermore, z_{it} is assumed to be time-varying although the vector x_{it} of exogenous variables may contain time-invariant regressors. Let w_{it} denote the set of instruments. To simplify notation we do not transform (48) to obtain a scalar error covariance matrix. This does not affect the primary result.

It is well known that the instrumental variables estimator can be obtained from estimating

$$(49) \quad y_{it} = x_{it}'\beta + z_{it}\psi_1 + \hat{v}_{it}\phi + e_{it}$$

by least squares, where e_{it} denotes a zero mean error term orthogonal to the regressors in this equation, and $\hat{v}_{it}$ is the residual from projecting z_{it} on the set of instruments.

Now consider our approach. First, write the reduced form equation for z_{it} as

$$(50) \quad z_{it} = m_{it}'\gamma + \alpha_i + \upsilon_{it}.$$

Assuming the endogeneity operates only through a nonzero covariance between μ_i and α_i the appropriate correction term is given by (8) and reduces to

$$(51) \quad \kappa \sum_{t=1}^{T} (z_{it} - m_{it}'\gamma)$$

which equals a constant κ times the time average of the residual from (50). Consequently, the instrumental variables procedure and the residual two-step method produce identical results if $\hat{v}_{it}$ in (49) is equal to (51) multiplied by some scaling factor.

Consider the instrument set suggested by Hausman and Taylor, i.e. $w_{it} = [x_{it}, \bar{x}_i, z_{it} - \bar{z}_i]$. It can be shown (see appendix) that $\hat{v}_{it} = \bar{z}_i - P_{\bar{x}_i}\bar{z}_i$, where the latter element denotes the projection of $\bar{z}_i$ upon $\bar{x}_i$. If the vector m_{it} in (50) is chosen as $[x_{it}, \bar{x}_i]$, it can also be shown that (51) corresponds to $\bar{z}_i - P_{\bar{X}_i}\bar{z}_i$. Thus, in the linear case our two-step procedure (with $m_{it} = [x_{it}, \bar{x}_i]$), results in algebraically the same estimator as the Hausman-Taylor approach. It also indicates that the normality assumption is not required in the linear case and imposing normality does not increase the estimator's efficiency.

Amemiya and MaCurdy (1986) extend the Hausman-Taylor approach by including $[x_{i1},\ldots,x_{iT}]$ in the instrument set. Our two-step estimator is identical to this instrumental variables estimator when we include in the reduced form in (50) the exogenous variables from all periods as well as x_{it} in deviation from its individual mean, i.e. if we choose $m_{it} = [x_{i1},\ldots, x_{iT}, x_{it}-\bar{x}_i]$. If we extend m_{it} to include $z_{i1} - \bar{z}_i, \ldots z_{iT} - \bar{z}_i$ our two-step procedure produces identical results to the instrumental variables estimator suggested by Breusch, Mizon and Schmidt (1989).

8. Empirical example

To illustrate our estimation strategy we examine an empirical example in which we estimate the impact of weekly hours worked on the offered hourly wage rate while accounting for the endogeneity of hours. This issue has attracted attention in the labor economics literature (see, for example, Moffitt (1984) Biddle and Zarkin (1989)). The model has the form

$$(52)\quad w_{it} = X_{it}\beta + Z_{it}\gamma + f(H_{it}) + \mu_i + \eta_{it}$$

$$(53)\quad H^*_{it} = X_{it}\theta + \alpha_i + \upsilon_{it}$$

$$(54)\quad \begin{array}{ll} H_{it} = H^*_{it} & \text{if } H^*_{it} > 0 \\ H_{it} = 0,\ w_{it} \text{ not observed} & \text{if } H^*_{it} \le 0 \end{array}$$

where w_{it} represents the log of the hourly wage of individual i in time period t; X are variables representing individual characteristics; Z are characteristics of the individuals work place; H^*_{it} represents the desired number of hours worked by individual i in time t and H_{it} denotes the observed number of hours; μ_i and α_i are normally distributed individual effects with covariance $\sigma_{\mu\alpha}$; η_{it} and υ_{it} are normally distributed error terms with covariance $\sigma_{\eta\upsilon}$; f denotes an unknown function mapping changes in weekly hours worked into wage changes; β, γ and θ are parameters to be estimated. This is a useful example as it highlights several aspects of our approach. First, as the hours variable is censored, the existing instrumental variables methods are inappropriate. Second, as the wage equation is estimated over the subsample reporting positive hours it illustrates our ability to control for selection bias. Third, the inclusion of the correction terms enables the identification of the form of the endogeneity. Finally, the potential non-linear relationship between wages and hours can be captured.

To estimate the model we employ data for young females taken from the National Longitudinal Survey (Youth Sample) for the period 1980-1987. For the period examined there were 2,300 observations for each of the eight years. From these 18,400 total observations there were 12,039 observations reporting positive hours of work in a given period.[13] Given the nature of the censoring we estimate equation (53) by random effects tobit. We assume that the individual effects are random and capture the the time effects by including dummy variables in equations (52) and (53).

The summary statistics are reported in Table 1 and the random effects tobit model results are presented in Table 2. Several points are worth noting. First, the parameter estimates are generally in keeping with expectations and most variables have a statistically significant impact on hours worked. Second, the coefficients on the time dummies indicate an increasing trend in hours worked over the period examined. Finally, the estimate of σ^2_α indicates that 39 percent of the total variance in the hours equation is explained by individual effects. Unfortunately, the tests outlined in section 6.1, adapted for the tobit type of censoring, reject the normality of both the individual specific component and the idiosyncratic component[14]. This is not surprising given the large number of observations (eight periods of 2300 individuals). Skewness of the empirical error distribution appears to be the main reason for the rejection of normality.

As this example is primarily illustrative, we continue despite this rejection and estimate the wage equation for the subsample of working women. To do this we need to

---

[13] Four hundred and eighty one women report positive hours of work over all eight years.

[14] The LM test statistics for testing normality of α_i and η_{it} are 297.6 and 814.5, respectively.

specify a form for the unknown function f and specify the length of J which captures departures from normality, as given in equation (47). We first estimate the wage equation without any correction for the potential endogeneity of hours. On the basis of experimentation we conclude that the f function can be approximated by a third order polynomial in hours. The results reported in column 1 of Table 3 indicate that there is a significant relationship between the number of weekly hours worked and the weekly wage rate. Although the linear effect of hours is statistically insignificant the higher order effects have a significant impact. The remaining coefficients in this column are reasonable in sign and magnitude although the coefficient reflecting the returns to schooling is rather high.

Column 2 of Table 3 reports the results from estimating the wage equation following the inclusion of the expected value of the individual effects and the individual/time specific effects. While the former is statistically insignificant the latter appears to have a strong negative impact on wages. The inclusion of these correction terms induces a number of important changes in the results. First, the time effects appear to be much stronger in this adjusted equation in comparison to the unadjusted equation. However, this increased time effect is partially due to the reduced experience effects. Second, the inclusion of the correction terms substantially reduces the impact of schooling from almost 10 percent to a more reasonable 6.3 percent. Finally, the linear hours effect is now statistically significant and the hours/wage profile appears to have noticeably changed from that implied in the previous column. Under the assumption of joint normality, the coefficients on the two correction terms correspond with the covariances between the individual effects and between the idiosyncratic terms in both equations, respectively. Both are estimated to be negative, the latter one being significant. The implied correlation coefficients[15] are high with estimates of -0.74 for the idiosyncratic effects and -0.93 for the individual effects. However, as the assumption of normality is rejected these estimates no longer are interpreted as correlation coefficients. Furthermore, the variances associated with these estimates are large.

Before comparing the implied wage/hours profiles we focus on the potentially non-normality of the log wage equation. These results are reported in the final column of Table 3. While there appears to be only very minor changes on the coefficients of primary interest the higher order terms are all statistically significant. This suggests that our assumption of normality in this wage equation is rejected.

To examine the effect of weekly hours worked on the wage rate we employ the

[15] The variances of μ_i and η_{it} are estimated as 0.075 and 0.114, respectively, implying a correlation coefficient across time of 0.40.

results from the estimation of the wage equation and plot the implied relationship. We do this by plotting the hours effect against hours worked for the last year of the data. The three plots corresponding to the results from columns 1, 2 and 3 are shown in Figure 1. As the hours effects do not operate through any of the other explanatory variables we plot the percentage change due to hours changes rather than the actual wage changes.

Two important points are worth noting from Figure 1. First, the unadjusted figure clearly lays below the two adjusted profiles. This suggests that the impact of endogeneity is reducing the estimated hours effect on wages. Second, the adjusted profiles both peak to the right of the peak for the unadjusted equation. While the unadjusted results suggest that hours peak in the 35-40 range the adjusted results indicate that the overtime effects are much stronger and lead to wages peaking at approximately 45 hours per week. Finally, despite the statistical significance of the higher order correction terms the implied profile from the two adjusted equations are almost identical. This indicates that the incorrect imposition of the normality assumption in the second step, in this case, is relatively harmless. Experimenting with the inclusion of higher order residuals appeared to have a surprisingly small impact on the estimated wage/hours profile.

9. Conclusions

This paper presents a new approach to estimating simultaneous equation panel data models with censored endogenous variables. The procedure we propose employs the residuals from the reduced form estimation of the endogenous variables to adjust for the heterogeneity in the primary equation. The panel nature of the data allows adjustment, and testing, for three forms of endogeneity. As the procedure we employ requires correct specification of the models we also propose some specification tests for limited dependent panel data models.

The cross section counterpart of the procedure we propose has long been considered as an alternative treatment of endogeneity to instrumental variables. We show that in special cases the method we propose is identical to instrumental variables procedures. This is a useful result as our procedure is easily extended to cover models not easily handled by instrumental variables. Finally we present an empirical example examining the relationship between the hourly wage rate and weekly hours worked. The results suggest that the estimation procedure works well.

Table 1: Variable Definitions and Means

Name	*definition*	*mean*	*st.dev.*	*mean*
SCHOOL	Years of schooling	11.91	1.55	11.67
HISP	Dummy, 1 if hispanic	0.126	0.331	0.132
BLACK	Dummy, 1 if black	0.189	0.189	0.208
AGE	Age in years	24.08	2.73	23.89
MAR	Dummy, 1 if married	0.386	0.487	0.415
WAGE	Log of real hourly wage	1.285	0.640	-
EXPER	Actual exp. in yrs.	5.22	2.45	4.95
EX2	Actual exp. squared	33.27	29.09	30.33
HLTH	Dummy, 1 if health disability	0.038	0.191	0.047
HOURS	Hours of work per week	33.98	11.82	-
UNION	Dummy, 1 if wage set by collective bargaining	0.168	0.374	-
RUR	Dummy, 1 if rural area	0.181	0.385	0.192
S	Dummy, 1 if lives in South	0.370	0.483	0.370
NC	Dummy, 1 if Northern Central	0.240	0.427	0.248
NE	Dummy, 1 if North East	0.204	0.403	0.187
Industry dummies				
PUB	Public Sector	0.048	0.215	-
AG	Agricultural	0.008	0.087	-
MIN	Mining	0.003	0.057	-
CON	Construction	0.008	0.087	-
MAN	Manufacturing	0.159	0.366	-
TRA	Transportation	0.040	0.195	-
TRAD	Trade	0.237	0.425	-
FIN	Finance	0.097	0.296	-
BUS	Business and Repair Service	0.045	0.206	-
PER	Personal Service	0.061	0.240	-
ENT	Entertainment	0.013	0.115	-
PRO	Professional and Related Services	0.281	0.450	-

Note: Averages in column 3 are based on observations over 2300 women over all eight years (1980-1987); averages and standard deviations in columns 1 and 2 are for the subsample of observations with positive hours only (12039 cases).

Table 2: Hours Equation Results

Constant	-186.903 (10.16)	**Dummy 1981**	1.816 (0.777)
Age	13.368 (0.796)	**Dummy 1982**	1.697 (0.854)
Age-squared	-0.291 (0.015)	**Dummy 1983**	3.807 (1.017)
Married	-5.682 (0.384)	**Dummy 1984**	6.561 (1.241)
Black	-6.419 (0.940)	**Dummy 1985**	8.592 (1.475)
Hispanic	-0.510 (1.113)	**Dummy 1986**	11.849 (1.679)
Rural	-1.893 (0.534)	**Dummy 1987**	14.313 (1.922)
Health	-3.870 (0.285)		
School	4.172 (0.285)		
		σ_α^2	259.09 (12.36)
Observations	18400	σ_ε^2	402.87 (4.93)

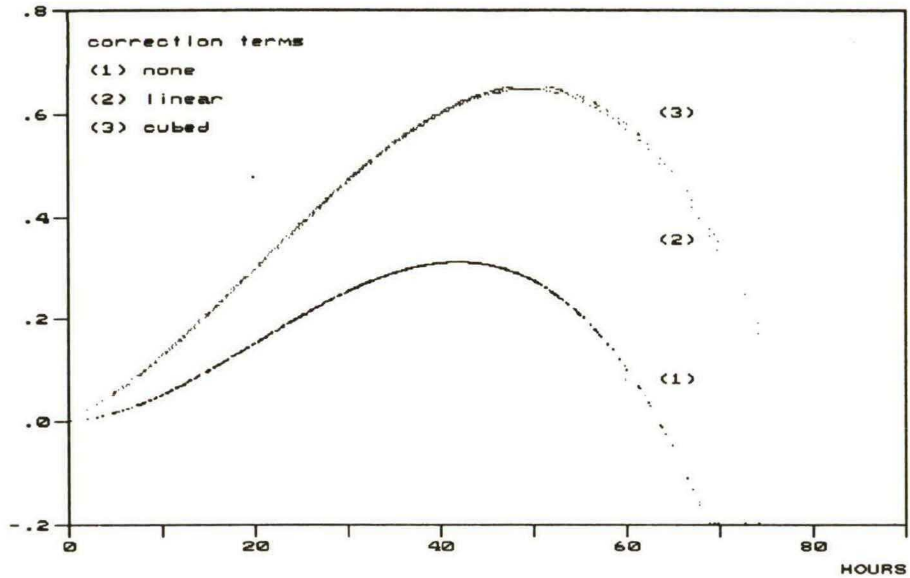
Note: model also includes regional dummies

Table 3: Wage equation results

	wage	wage	wage
Married	-0.015 (0.011)	0.047 (0.016)	0.048 (0.017)
Black	-0.144 (0.014)	-0.098 (0.021)	-0.098 (0.021)
Hispanic	-0.028 (0.017)	-0.022 (0.016)	-0.020 (0.016)
Rural	-0.115 (0.014)	-0.101 (0.014)	-0.101 (0.014)
School	0.096 (0.003)	0.063 (0.009)	0.063 (0.009)
Union	0.116 (0.014)	0.120 (0.014)	0.120 (0.014)
Exper	0.038 (0.009)	0.009 (0.009)	0.008 (0.009)
Exper2	-0.0009 (0.0007)	-0.0005 (0.0007)	-0.0003 (0.007)
Hours	0.001 (0.004)	0.009 (0.004)	0.009 (0.004)
Hours squared	0.0005 (0.0001)	0.0004 (0.0001)	0.0004 (0.0001)
Hours cubed	-0.000008 (0.000001)	-0.000007 (0.000001)	-0.000007 (0.000001)
Correction terms:			
α	-	-4.097 (3.267)	-4.845 (3.371)
ε	-	-5.018 (0.911)	-5.048 (0.931)
α^2	-	-	-117.24 (48.25)
ε^2	-	-	15.33 (6.73)
α^3	-	-	5729.1 (2877.4)
ε^3	-	-	-140.42 (53.30)
$\bar{R}^2$	0.246	0.267	0.268
Observations	12039	12039	12039

Note: all specifications also include regional, industry and time dummies

Figure 1: Estimated Effect of Hours on Wages (ceteris paribus)



Appendix

1. The relationship between instrumental variable and residual-type estimators.

Theorem:
Let **M** denote the set of variables included in the reduced form given by (2), and let **W** denote the set of instruments for the IV estimator. Let P_v denote the projection matrix producing individual means and Q_v the projection matrix giving deviations from individual means ("within" transformation). Finally, let $P_{\mathbf{A}}$ denote the projection matrix upon the space spanned by the columns of **A**, while $M_{\mathbf{A}} = I - P_{\mathbf{A}}$. Apart from scaling, the correction term from the two-step method obtained from (8) is given by v = $P_v M_{\mathbf{M}} z$, while the residual from the IV projection upon **W** is given by $M_{\mathbf{W}} z$. Now the following result holds:

$$P_v M_{\mathbf{M}} z = M_{\mathbf{W}} z \quad \text{if} \quad \mathbf{W} = [\mathbf{M} : Q_v z],$$

provided **M** has the same column space as $[Q_v \mathbf{M} : P_v \mathbf{M}]$.

Proof
Since

$$M_{\mathbf{W}} z = M_{\mathbf{W}}(Q_v z + P_v z) = P_v z - P_{\mathbf{W}} P_v z = P_v z - P_{P_v \mathbf{W}} P_v z = P_v z - P_{P_v \mathbf{M}} P_v z$$

we have that $M_{\mathbf{W}} z = P_v(I - P_{P_v \mathbf{M}})z$. On the other hand, it holds that

$$P_v M_{\mathbf{M}} z = P_v(z - P_{\mathbf{M}} z) = P_v(z - P_{Q_v \mathbf{M}} Q_v z - P_{P_v \mathbf{M}} P_v z) = P_v(I - P_{P_v \mathbf{M}})z,$$

which proves the required results.

Corollary 1
The Hausman-Taylor estimator is identical to the two-step estimator for $\mathbf{M} = [Q_v X : P_v X]$ or, equivalently, $\mathbf{M} = [X : P_v X]$.
Corollary 2
The Amemiya-MaCurdy estimator is identical to the two-step estimator for $\mathbf{M} = [Q_v X : X^*]$.

Corollary 3
The Breusch-Mizon-Schmidt estimator is identical to the two-step estimator for $\mathbf{M} = [Q_v X : X^* : (Q_v z)^*]$.

The proofs of these corollaries are straightforwardly obtained from the general theorem. The notation X^* is taken from Breusch, Mizon and Schmidt (1989). Each column of X^* contains values of $X_{k,it}$ for one t only. Note that $P_v X^* = X^*$.

2. Computing correct standard errors

The estimator for the parameters in (1) is a two-step estimator and can be viewed as a member of a class of method of moments estimators (cf. Newey (1984)). Considering the case with fixed time effects as the most relevant case, this estimator is root N consistent and asymptotically normal under weak regularity conditions. The asymptotic covariance matrix can be obtained using the results in Newey (1984) or Pagan (1986) taking into account the fact that in most cases the error terms exhibit heteroskedasticity.

Let θ_2 denote the parameter vector in (2) and let the (root N) consistent maximum likelihood estimator for θ_2 be given by $\hat{\theta}_2$ and its asymptotic covariance matrix by V_2 (with estimate $\hat{V}_2$). Assume, for simplicity, that k is the identity mapping and that (1) is linear in parameters, denoted $\theta_1 = (\theta_{11}, \theta_{12})$ where θ_{11} corresponds to the observed regressors X and θ_{12} to the generated regressors denoted $Z(\hat{\theta}_2)$. In vector notation, the estimated version of (1) can be written as

$$(1') \quad y_i = X_i\theta_{11} + Z_i(\hat{\theta}_2)\theta_{12} + e_i$$

where e_i is an error term vector corresponding to the conditional distribution of[16] $\iota_T\mu_i + \eta_i$ given Z_i, ι_T being a T-dimensional vector of ones. Let Ω_i denote the variance of e_i for individual i. Introducing the following notation:

$$M_N = N^{-1}\sum_{i=1}^{N} E\,[X_i\ Z_i(\theta_2)]'\,[X_i\ Z_i(\theta_2)]$$

$$V_N = N^{-1}\sum_{i=1}^{N} E\,[X_i\ Z_i(\theta_2)]'\,\Omega_i\,[X_i\ Z_i(\theta_2)]$$

[16] Recall that the time effects are excluded from the error term.

$$D_N = N^{-1} \sum_{i=1}^{N} E\,[X_i\ Z_i(\theta_2)]'\ [\frac{\partial Z_i(\theta_2)\theta_{12}}{\partial \theta_2'}]$$

the asymptotic covariance matrix of the second step (least squares) estimator for θ_1 is given by

$$V_1 = \lim_{N\to\infty} M_N^{-1}\ (V_N + D_N V_2 D_N')\ M_N^{-1}$$

which can easily be consistently estimated replacing expectations by sample moments and unknown parameters by their estimates. The second part within brackets in the expression for V_1 is due to the generated regressors problem; the first part gives the standard covariance matrix in case of heteroskedasticity. The simplest way to get an estimate for V_N without specifying the exact form of heteroskedasticity follows from generalizing the results of White (1980) to the panel data case. The appropriate estimator for V_N is given by

$$\hat{V}_N = N^{-1} \sum_{i=1}^{N}\ [X_i\ Z_i(\hat{\theta}_2)]'\ \hat{e}_i\hat{e}_i'\ [X_i\ Z_i(\hat{\theta}_2)],$$

where $\hat{e}_i$ is the T-dimensional vector of least squares residuals from (1). Note that this estimator does not make use of the error components structure of the original errors, but is attractive because of its simplicity.

Finally, if conditioning upon Z_i implies that certain observations are excluded from the second stage estimation, the dimensions of all vectors and matrices above should be adjusted to include only those observations that are actually used in estimation.

References

Amemiya,T., "Tobit Models: A Survey," Journal of Econometrics 24 (1984), 3-61.

——, Advanced Econometrics, Basil Blackwell, Oxford (1985).

—— and T. MaCurdy, "Instrumental-Variable Estimation of an Error-Components Model," Econometrica 54 (1986), 869-881.

Arellano, M., O. Bover, and J.M. Labeaga, "Autoregressive Models with Sample Selectivity for Panel Data, working paper, CEMFI, Madrid.

Biddle, J. and G. Zarkin, "Choice among Wage-Hours Packages: An Empirical Investigation of Male Labor Supply", Journal of Labor Economics, 7 (1989), 415-437.

Breusch, T.S. and A.R. Pagan, "A Simple Test for Heteroskedasticity and Random Coefficient Variation," Econometrica 47 (1979), 1287-1294.

——, G.E. Mizon and P. Schmidt, "Efficient Estimation Using Panel Data," Econometrica 57 (1989), 695-700.

Blundell, R. & R. Smith, "Coherency and Estimation in Simultaneous Models with Censored or Qualitative Dependent Variables" *forthcoming* (1992).

Butler, J. and R. Moffitt, "A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model" Econometrica, 50 (1982), 761-764.

Gallant, A. and D. Nychka, "Semi-nonparametric maximum likelihood estimation," Econometrica 55 (1987), 363-390.

Gourieroux, C. and A. Monfort, "Simulation Based Econometric Methods", working paper, ENSAE-CREST, Paris (1990).

Gourieroux, C., A. Monfort, E. Renault and A.Trognon, "Generalized Residuals," Journal of Econometrics 34 (1987), 5-32.

Hausman, J., "Specification Tests in Econometrics," Econometrica 46 (1978), 1251-1271.

—— and D. Wise, "Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment," Econometrica 46 (1979), 455-473.

—— and W. Taylor "Panel Data and Unobservable Individual Effects," Econometrica 49 (1981), 1377-1398.

Heckman, J., "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variable Models and a Simple Estimator for Such Models," Annals of Economic and Social Measurement 5 (1976), 475-492.

——, "Dummy Endogenous Variables in a Simultaneous Equation System," Econometrica 46 (1978), 931-959.

——, "Sample Selection Bias as a Specification Error," Econometrica 47 (1979), 153-161.

——, "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete-Time-Discrete-Data Stochastic Process", in: Cf. Manski and D.

McFadden, eds., Structural Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, MA (1981).

Keane, M., R. Moffitt and D. Runkle, "Real Wages over the Business Cycle: Estimating the Impact of Heterogeneity with Micro Data," Journal of Political Economy 96 (1988), 1232-1265.

Lee, L., "Tests for the Bivariate Normal Distribution in Econometric Models with Selectivity," Econometrica 52 (1984), 843-863.

MaCurdy, T., "An Empirical Model of Labor Supply in a Life-Cycle Setting" Journal of Political Economy 89 (1981), 1059-1085.

Moffitt, R., "The Estimation of a Joint Wage-Hours Labor Supply Model", Journal of Labor Economics 2 (1984), 550-566.

Newey, W., "A Method of Moments Interpretation of Sequential Estimators," Economics Letters 14 (1984), 201-206.

——, "Maximum Likelihood Specification Testing and Conditional Moment Tests," Econometrica 53 (1985), 1047-1070.

——, "Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables", Journal of Econometrics 36 (1987), 231-250.

——, J. Powell and F. Vella, "Non-parametric Instrumental Variables Estimation via Additive Models" (in progress)

Nijman, T., and M. Verbeek, "Non response in Panel Data: The Impact on Estimates of a Life Cycle Consumption Function," Journal of Applied Econometrics 7 (1992), 243-257.

Pagan, A., "Two Stage and Related Estimators and their Applications," Review of Economic Studies 53 (1986), 517-538.

—— and F. Vella, "Diagnostic Tests for Models Based on Individual Data: A Survey," Journal of Applied Econometrics 4 (1989), S29-S60.

Ridder, G., "Attrition in Multi-Wave Panel Data" in J. Hartog, G. Ridder and J. Theeuwes, eds., Panel Data and Labor Market Studies, Elsevier, North Holland (1990)

Rivers, D. and Q. Vuong, "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models, " Journal of Econometrics 39 (1988), 347-366.

Ruud, P.A., "Tests for Specification in Econometrics," Econometric Reviews, 3 (1984) 221-242.

Sickles, R., and P. Taubman, "An Analysis of the Health and Retirement Status of the Elderly" Econometrica 54 (1986), 1339 - 1356.

Smith, R. and R. Blundell, "An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply," Econometrica 54 (1986), 679-685.

Vella, F., "Simple Tests for Sample Selection Bias in Censored and Discrete Choice Models," Journal of Applied Econometrics 7 (1992), 413-421..

——, "A Simple Estimator for Simultaneous Models with Censored Endogenous Regressors, International Economic Review 34 (1993), 441-457.

——, and R. Gregory, "Selection Bias and Human Capital Investment: Estimating the Rates of Return to Education for Young Males," discussion paper, Rice University (1993).

—— and M. Verbeek, "Estimating the Impact of Endogenous Union Choice on Wages using Panel Data," discussion paper, Tilburg University (1993a).

—— and ——, "Estimating and Interpreting Models with Endogenous Treatment Effects: The Relationship between Competing Estimators of the Union Impact on Wages", CentER Discussion Paper 9351, Tilburg University (1993b)

Verbeek, M., & T. Nijman, "Testing for Selectivity Bias in Panel Data Models," International Economic Review 33 (1992) 681-703.

White, H., "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, Econometrica 48 (1980) 817-838.

Wooldridge, J., "Selection Corrections for Panel Data Models under Conditional Mean Independence Assumptions", working paper, Michigan State University (1993).

Discussion Paper Series, CentER, Tilburg University, The Netherlands:

(For previous papers please consult previous discussion papers.)

No.	Author(s)	Title
9221	S. Smulders and Th. van de Klundert	Monopolistic Competition, Product Variety and Growth: Chamberlin vs. Schumpeter
9222	H. Bester and E. Petrakis	Price Competition and Advertising in Oligopoly
9223	A. van den Nouweland, M. Maschler and S. Tijs	Monotonic Games are Spanning Network Games
9224	H. Suehiro	A "Mistaken Theories" Refinement
9225	H. Suehiro	Robust Selection of Equilibria
9226	D. Friedman	Economically Applicable Evolutionary Games
9227	E. Bomhoff	Four Econometric Fashions and the Kalman Filter Alternative - A Simulation Study
9228	P. Borm, G.-J. Otten and H. Peters	Core Implementation in Modified Strong and Coalition Proof Nash Equilibria
9229	H.G. Bloemen and A. Kapteyn	The Joint Estimation of a Non-Linear Labour Supply Function and a Wage Equation Using Simulated Response Probabilities
9230	R. Beetsma and F. van der Ploeg	Does Inequality Cause Inflation? - The Political Economy of Inflation, Taxation and Government Debt
9231	G. Almekinders and S. Eijffinger	Daily Bundesbank and Federal Reserve Interventions - Do they Affect the Level and Unexpected Volatility of the DM/$-Rate?
9232	F. Vella and M. Verbeek	Estimating the Impact of Endogenous Union Choice on Wages Using Panel Data
9233	P. de Bijl and S. Goyal	Technological Change in Markets with Network Externalities
9234	J. Angrist and G. Imbens	Average Causal Response with Variable Treatment Intensity
9235	L. Meijdam, M. van de Ven and H. Verbon	Strategic Decision Making and the Dynamics of Government Debt
9236	H. Houba and A. de Zeeuw	Strategic Bargaining for the Control of a Dynamic System in State-Space Form
9237	A. Cameron and P. Trivedi	Tests of Independence in Parametric Models: With Applications and Illustrations

No.	Author(s)	Title
9238	J.-S. Pischke	Individual Income, Incomplete Information, and Aggregate Consumption
9239	H. Bloemen	A Model of Labour Supply with Job Offer Restrictions
9240	F. Drost and Th. Nijman	Temporal Aggregation of GARCH Processes
9241	R. Gilles, P. Ruys and J. Shou	Coalition Formation in Large Network Economies
9242	P. Kort	The Effects of Marketable Pollution Permits on the Firm's Optimal Investment Policies
9243	A.L. Bovenberg and F. van der Ploeg	Environmental Policy, Public Finance and the Labour Market in a Second-Best World
9244	W.G. Gale and J.K. Scholz	IRAs and Household Saving
9245	A. Bera and P. Ng	Robust Tests for Heteroskedasticity and Autocorrelation Using Score Function
9246	R.T. Baillie, C.F. Chung and M.A. Tieslau	The Long Memory and Variability of Inflation: A Reappraisal of the Friedman Hypothesis
9247	M.A. Tieslau, P. Schmidt and R.T. Baillie	A Generalized Method of Moments Estimator for Long-Memory Processes
9248	K. Wärneryd	Partisanship as Information
9249	H. Huizinga	The Welfare Effects of Individual Retirement Accounts
9250	H.G. Bloemen	Job Search Theory, Labour Supply and Unemployment Duration
9251	S. Eijffinger and E. Schaling	Central Bank Independence: Searching for the Philosophers' Stone
9252	A.L. Bovenberg and R.A. de Mooij	Environmental Taxation and Labor-Market Distortions
9253	A. Lusardi	Permanent Income, Current Income and Consumption: Evidence from Panel Data
9254	R. Beetsma	Imperfect Credibility of the Band and Risk Premia in the European Monetary System
9301	N. Kahana and S. Nitzan	Credibility and Duration of Political Contests and the Extent of Rent Dissipation
9302	W. Güth and S. Nitzan	Are Moral Objections to Free Riding Evolutionarily Stable?

No.	Author(s)	Title
9303	D. Karotkin and S. Nitzan	Some Peculiarities of Group Decision Making in Teams
9304	A. Lusardi	Euler Equations in Micro Data: Merging Data from Two Samples
9305	W. Güth	A Simple Justification of Quantity Competition and the Cournot-Oligopoly Solution
9306	B. Peleg and S. Tijs	The Consistency Principle For Games in Strategic Form
9307	G. Imbens and A. Lancaster	Case Control Studies with Contaminated Controls
9308	T. Ellingsen and K. Wärneryd	Foreign Direct Investment and the Political Economy of Protection
9309	H. Bester	Price Commitment in Search Markets
9310	T. Callan and A. van Soest	Female Labour Supply in Farm Households: Farm and Off-Farm Participation
9311	M. Pradhan and A. van Soest	Formal and Informal Sector Employment in Urban Areas of Bolivia
9312	Th. Nijman and E. Sentana	Marginalization and Contemporaneous Aggregation in Multivariate GARCH Processes
9313	K. Wärneryd	Communication, Complexity, and Evolutionary Stability
9314	O.P.Attanasio and M. Browning	Consumption over the Life Cycle and over the Business Cycle
9315	F. C. Drost and B. J. M. Werker	A Note on Robinson's Test of Independence
9316	H. Hamers, P. Borm and S. Tijs	On Games Corresponding to Sequencing Situations with Ready Times
9317	W. Güth	On Ultimatum Bargaining Experiments - A Personal Review
9318	M.J.G. van Eijs	On the Determination of the Control Parameters of the Optimal Can-order Policy
9319	S. Hurkens	Multi-sided Pre-play Communication by Burning Money
9320	J.J.G. Lemmen and S.C.W. Eijffinger	The Quantity Approach to Financial Integration: The Feldstein-Horioka Criterion Revisited
9321	A.L. Bovenberg and S. Smulders	Environmental Quality and Pollution-saving Technological Change in a Two-sector Endogenous Growth Model

No.	Author(s)	Title
9322	K.-E. Wärneryd	The Will to Save Money: an Essay on Economic Psychology
9323	D. Talman, Y. Yamamoto and Z. Yang	The $(2^{n+m+1} - 2)$-Ray Algorithm: A New Variable Dimension Simplicial Algorithm For Computing Economic Equilibria on $S^n \times R^m_+$
9324	H. Huizinga	The Financing and Taxation of U.S. Direct Investment Abroad
9325	S.C.W. Eijffinger and E. Schaling	Central Bank Independence: Theory and Evidence
9326	T.C. To	Infant Industry Protection with Learning-by-Doing
9327	J.P.J.F. Scheepens	Bankruptcy Litigation and Optimal Debt Contracts
9328	T.C. To	Tariffs, Rent Extraction and Manipulation of Competition
9329	F. de Jong, T. Nijman and A. Röell	A Comparison of the Cost of Trading French Shares on the Paris Bourse and on SEAQ International
9330	H. Huizinga	The Welfare Effects of Individual Retirement Accounts
9331	H. Huizinga	Time Preference and International Tax Competition
9332	V. Feltkamp, A. Koster, A. van den Nouweland, P. Borm and S. Tijs	Linear Production with Transport of Products, Resources and Technology
9333	B. Lauterbach and U. Ben-Zion	Panic Behavior and the Performance of Circuit Breakers: Empirical Evidence
9334	B. Melenberg and A. van Soest	Semi-parametric Estimation of the Sample Selection Model
9335	A.L. Bovenberg and F. van der Ploeg	Green Policies and Public Finance in a Small Open Economy
9336	E. Schaling	On the Economic Independence of the Central Bank and the Persistence of Inflation
9337	G.-J. Otten	Characterizations of a Game Theoretical Cost Allocation Method
9338	M. Gradstein	Provision of Public Goods With Incomplete Information: Decentralization vs. Central Planning
9339	W. Güth and H. Kliemt	Competition or Co-operation
9340	T.C. To	Export Subsidies and Oligopoly with Switching Costs
9341	A. Demirgüç-Kunt and H. Huizinga	Barriers to Portfolio Investments in Emerging Stock Markets

No.	Author(s)	Title
9342	G.J. Almekinders	Theories on the Scope for Foreign Exchange Market Intervention
9343	E.R. van Dam and W.H. Haemers	Eigenvalues and the Diameter of Graphs
9344	H. Carlsson and S. Dasgupta	Noise-Proof Equilibria in Signaling Games
9345	F. van der Ploeg and A.L. Bovenberg	Environmental Policy, Public Goods and the Marginal Cost of Public Funds
9346	J.P.C. Blanc and R.D. van der Mei	The Power-series Algorithm Applied to Polling Systems with a Dormant Server
9347	J.P.C. Blanc	Performance Analysis and Optimization with the Power-series Algorithm
9348	R.M.W.J. Beetsma and F. van der Ploeg	Intramarginal Interventions, Bands and the Pattern of EMS Exchange Rate Distributions
9349	A. Simonovits	Intercohort Heterogeneity and Optimal Social Insurance Systems
9350	R.C. Douven and J.C. Engwerda	Is There Room for Convergence in the E.C.?
9351	F. Vella and M. Verbeek	Estimating and Interpreting Models with Endogenous Treatment Effects: The Relationship Between Competing Estimators of the Union Impact on Wages
9352	C. Meghir and G. Weber	Intertemporal Non-separability or Borrowing Restrictions? A Disaggregate Analysis Using the US CEX Panel
9353	V. Feltkamp	Alternative Axiomatic Characterizations of the Shapley and Banzhaf Values
9354	R.J. de Groof and M.A. van Tuijl	Aspects of Goods Market Integration. A Two-Country-Two -Sector Analysis
9355	Z. Yang	A Simplicial Algorithm for Computing Robust Stationary Points of a Continuous Function on the Unit Simplex
9356	E. van Damme and S. Hurkens	Commitment Robust Equilibria and Endogenous Timing
9357	W. Güth and B. Peleg	On Ring Formation In Auctions
9358	V. Bhaskar	Neutral Stability In Asymmetric Evolutionary Games
9359	F. Vella and M. Verbeek	Estimating and Testing Simultaneous Equation Panel Data Models with Censored Endogenous Variables

P.O. BOX 90153, 5000 LE TILBURG, THE NETHERLAND