entER
for
nomic Research

Discussion paper

[redacted]

TILBURG UNIVERSITY

Center
for
Economic Research

No. 9455

TWO-STEP ESTIMATION OF SIMULTANEOUS EQUATION PANEL DATA MODELS WITH CENSORED ENDOGENOUS VARIABLES

by Francis Vella and
Marno Verbeek

August 1994

ISSN 0924-7815

Two-Step Estimation of Simultaneous Equation Panel Data Models with Censored Endogenous Variables *

Francis VELLA
Department of Economics
Rice University

Marno VERBEEK
Department of Econometrics and CentER
Tilburg University

This version June 1994

Abstract

This paper presents a two-step approach to estimating simultaneous equation panel data models with censored endogenous variables and sample selection. The procedure employs the residuals from the reduced form estimation of the endogenous variable to adjust for the heterogeneity in the primary equation. The panel nature of the data allows adjustment, and testing, for three forms of endogeneity.

*This is a substantially shortened and revised version of an earlier paper, circulated under the title "Estimating and Testing Simultaneous Equation Panel Data Models with Censored Endogenous Variables". 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, and while Vella was visiting the CentER for Economic Research at Tilburg University. Helpful comments by Bertrand Melenberg, Robin Sickles, Jeffrey Wooldridge and detailed suggestions by two anonymous referees are gratefully acknowledged. We alone are responsible for any remaining errors.

1 Introduction

This paper proposes a two step estimator for panel data models with censored endogenous variables and/or sample selection bias. The model comprises a primary equation with an endogenous explanatory variable and the reduced form for the endogenous explanator. We derive estimates of the heterogeneity generating the endogeneity to include as additional explanatory variables in the primary equation. These are obtained through a decomposition of the reduced form residuals. The procedure is applicable in cases where the primary equation has an uncensored dependent variable and the endogenous explanator is either uncensored or censored, as well as cases where the primary equation has a censored dependent variable. In addition, it is applicable to estimation over non-randomly chosen sub-samples.

Previously proposed procedures examining this family of models commonly assume that the endogeneity is due to time-invariant individual effects (see, for example, Hausman and Taylor [1981], Amemiya and MaCurdy [1986] and Honoré [1992, 1993]). We, however, allow the endogeneity to operate through an individual component, a time component and an individual specific/time component. This allows a richer econometric, and economic, structure. Moreover, it extends several well known cross-sectional estimators to panel data (see, for example, Heckman [1979], Smith and Blundell [1986], Rivers and Vuong [1988] and Vella [1993]) and encompasses existing panel data procedures for sample selection and attrition bias (see, for example, Ridder [1990], Nijman and Verbeek [1992] and Wooldridge [1993]).

Two-step estimators are generally inefficient in comparison to limited information maximum likelihood (LIML) (see, for example, Newey [1987]). The efficiency loss is model specific, while under restrictive conditions the two-step estimator attains the Cramér-Rao lower bound. We are confident that our procedure is sufficiently simple that it will be often preferable to LIML. If efficiency is a primary objective our method provides initial consistent estimators for a LIML approach[1]. It is possible to estimate a subset of the models we examine under more general conditions, for example by instrumental variables (IV). We, however, do not consider this a shortcoming as there is a trade-off between the economic information extracted via estimation and the distributional assumptions employed. Furthermore, the distributional assumptions can be tested by generalizing the approach

[1] A computationally attractive alternative is simulated maximum likelihood, in which the integrals in the loglikelihood function are replaced by simulators (cf. Gourieroux and Monfort [1993]).

in Pagan and Vella [1989] or, for special cases, through comparisons with semi-parametric alternatives, such as Honoré [1992], via the methodology in Peters and Smith [1991]. Moreover, we show that IV estimators, including those in Hausman and Taylor [1981], Amemiya and MaCurdy [1986] and Breusch, Mizon and Schmidt [1989], can be replicated by our "residual" approach.

The following section presents the model and outlines the estimation procedure. Sections 3 and 4 consider two important cases in detail, while Section 5 discusses the relationship between our procedure and available IV estimators. Concluding comments are presented in Section 6.

2 The general framework

Consider a general model where equations (1)–(2) are of primary focus and equations (3)–(4) constitute the reduced form for the endogenous explanatory variable.

$$
\begin{aligned}
y_{it}^* &= g_1(x_{it}, z_{it}, z_{it}^*; \theta_1) + \mu_i + \varepsilon_t + \eta_{it} && (1)\\
y_{it} &= g_2(y_{it}^*; \theta_2) \quad \text{if} \;\; g_3(z_{i1}, ..., z_{iT}; t) \neq 0 && (2)\\
& \quad \text{unobserved elsewhere} && \\
z_{it}^* &= g_4(x_{it}, z_{i,t-1}; \theta_4) + \alpha_i + \rho_t + v_{it} && (3)\\
z_{it} &= g_5(z_{it}^*; \theta_5), && (4)
\end{aligned}
$$

where i indexes individuals ($i = 1, ..., N$) and t indexes time ($t = 1, ..., T$). y_{it}^* and z_{it}^* are latent endogenous variables; y_{it} and z_{it} are observed variables produced by the censoring functions g_2 and g_5 noting that these functions may be characterized by the unknown parameters θ_2 and θ_5. The functions g_1 and g_4 are assumed to be continuously differentiable with respect to the parameter vectors θ_1, θ_4, respectively. While we make no assumptions about the function g_3, determining sample selection, we require that the applicable regime is observed. The main parameters of interest are θ_1 and, when applicable, θ_2. It is assumed that the parameters are identified up to some normalization. In nonlinear cases identification is in principle guaranteed through distributional assumptions only, but, preferably, functional form and exclusion restrictions are imposed. Also, standard restrictions on the functions g_j, $j = 1, 2, 4, 5$ apply[2]. Finally, the variables in x_{it} are assumed to

[2] For example, when g_5 is the identity mapping, we cannot have both z_{it} and z_{it}^* entering g_1 in equation (1).

be independent of all error components such that, potentially, each function g_j can depend on x_{is} for any s.

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 v_{it}. These are assumed to be i.i.d. jointly normal with zero mean and variances $\sigma_\ell^2, \ell = \alpha, \mu, \varepsilon, \rho, \eta, v$. Each effect is potentially correlated with its counterpart of the same dimension in the other equation, with covariances $\sigma_{\mu\alpha}$, $\sigma_{\varepsilon\rho}$ and $\sigma_{\eta v}$.[3] The endogeneity of z_{it}^* and z_{it} thus operates through the three factors common across the two equations.

Consider some models this framework encompasses. The conventional sample selection model (see Nijman and Verbeek [1992]) is obtained when z_{it} and z_{it}^* do not appear in g_1, g_5 is an index function and g_3 equals z_{it}. In this case z_{it} is a zero-one variable indicating selection into the sample. If g_3 equals $\prod_t z_{it}$, attention is restricted to a balanced sub-panel. A dummy endogenous variable (see Heckman [1978]) appears in (1) when z_{it} appears in g_1 and g_5 is an index function. More general g_5 functions allow for categorical and censored endogenous variables with reduced forms corresponding to ordered probit or tobit specifications. This allows the inclusion of a range of dummy-variables in (1) corresponding to different values for z_{it}. Another feature is the inclusion of either z_{it} or z_{it}^*, as discussed in Blundell and Smith [1993] and Vella [1993], in the primary equation, which allows the relationship of interest to be between y_{it} and z_{it}^* rather than y_{it} and z_{it}. Finally, the general specification permits non-linear transformations of the endogenous variables which may depend on unknown parameters.

Our general strategy is the following. We estimate (3)-(4) by maximum likelihood to obtain consistent estimators for θ_4, θ_5 and the variances of the error components. This requires the usual regularity conditions and z_{i0} to be strictly exogenous (see Heckman [1981]). We then condition (1) on the observed outcomes of the endogenous explanatory variable and employ the relevant conditional moment restrictions, or the conditional likelihood function, to estimate the parameters in (1) and (2). For cross-sectional estimation the conditioning set includes z_{it} for the relevant time period (see Smith and Blundell [1986] and Vella [1993]). To exploit the correlation structure of the panel, it is natural to condition upon the vector of *all* outcomes. This is required to seperately identify the three sources of endogeneity and, moreover, to guarantee consistency for general sample se-

[3]Consistent estimation of $\sigma_{\mu\alpha}$ requires N to be large while $\sigma_{\varepsilon\rho}$ requires large T. For $\sigma_{\eta v}$ we require either N and/or T to be large.

lection functions g_3. This allows estimation over subsamples corresponding to non-randomly chosen values of the endogenous explanator. In addition, it increases the estimator's efficiency as it incorporates additional information into estimation[4]. It is often useful to include the time effects ρ_t from (3) in the conditioning set. Accordingly, the conditioning set reduces to $\{z_i, \rho\}$ where $z_i = (z_{i1}, ..., z_{iT})'$.

Our approach has two primary advantages over LIML. First, it is computationally attractive as it generally does not require higher order integrals. Second, as the specifications of g_1 and g_2 are only relevant for the second stage estimation it is relatively easy to conduct specification searches for the primary equation. These advantages, however, cannot be realized for all specifications of g_j. Two general classes of models that satisfy the requirements can be characterized as:

1. Case I: g_2 is the identity mapping; g_1 is linear in z^*_{it}.

2. Case II: $g_5(z^*_{it}) = z^*_{it}\,\mathrm{I}[z^*_{it} \in A]$, for some $A \subseteq \mathbb{R}$, where I is an indicator function; $g_3(z_{i1}, ..., z_{iT}; t) = \prod_s z_{is} g^*_3(z_{i1}, ..., z_{iT}; t)$ for some function g^*_3.

The conditions for case II imply that the distribution of z_i is continuous if y_{it} is observed (for any t). The next two sections focus on these cases which we refer to as conditional moment estimation and conditional maximum likelihood (CML) estimation.

3 Conditional moment estimation

Conditioning[5] (1) on the NT vector of outcomes z_{it}, denoted Z, produces

$$\begin{aligned} E\{y^*_{it} \mid Z\} &= E\{g_1(x_{it}, z_{it}, z^*_{it}, \theta_1) \mid Z\} \\ &\quad + E\{\mu_i \mid Z\} + E\{\varepsilon_t \mid Z\} + E\{\eta_{it} \mid Z\}. \end{aligned} \quad (5)$$

When g_2 is the identity mapping, it is straightforward to estimate the parameters in (5), from the appropriate conditional moment restrictions, given expressions for the conditional expectations in (5) and consistent estimates

[4] If $g_3(.;t)$ depends on z_{it} only and g_4 does not involve $z_{i,t-1}$, a consistent estimator for θ_1 can be obtained by conditioning upon z_{it} only and pooling the cross-sections (see Wooldridge (1993) for an example).

[5] All conditional expectations that follow are also conditional upon the exogenous variables in x_{it} (for all t).

for the parameters in (3) and (4). When g_2 is not the identity mapping we employ the conditional distribution of y_{it}. We focus on this latter case in Section 4. Our initial task is to find expression for the conditional expectations in (5). Note that the conditional expectation of g_1 is taken over z^*_{it} only. If g_1 is linear in z^*_{it} this is a straightforward function of g_4 and the conditional expectations of the error components in (3).

We proceed by deriving $E\{\mu_i \mid U\}$, $E\{\varepsilon_t \mid U\}$ and $E\{\eta_{it} \mid U\}$ where $u_{it} = \alpha_i + \rho_t + v_{it}$ and U is the NT vector of u_{it}'s. We subsequently take expectations with respect to U given Z, noting this second iteration of the expectations is influenced by the censoring function g_5. Joint normality and straightforward matrix manipulations (using Hsiao [1986, eq. (3.6.20)]), produce

$$E\{\mu_i \mid U\} = \sigma_{\mu\alpha}\left[\frac{T}{\sigma_v^2 + T\sigma_\alpha^2}\bar{u}_{i.} - \frac{\sigma_\rho^2 NT}{(\sigma_v^2 + T\sigma_\alpha^2)(\sigma_v^2 + T\sigma_\alpha^2 + N\sigma_\rho^2)}\bar{u}_{..}\right] \quad (6)$$

$$E\{\varepsilon_t \mid U\} = \sigma_{\varepsilon\rho}\left[\frac{N}{\sigma_v^2 + N\sigma_\rho^2}\bar{u}_{.t} - \frac{\sigma_\alpha^2 NT}{(\sigma_v^2 + N\sigma_\rho^2)(\sigma_v^2 + T\sigma_\alpha^2 + N\sigma_\rho^2)}\bar{u}_{..}\right] \quad (7)$$

$$E\{\eta_{it} \mid U\} = \sigma_{\eta v}\left[\frac{1}{\sigma_v^2}u_{it} - \frac{T\sigma_\alpha^2}{\sigma_v^2(\sigma_v^2 + T\sigma_\alpha^2)}\bar{u}_{i.} - \frac{N\sigma_\rho^2}{\sigma_v^2(\sigma_v^2 + N\sigma_\rho^2)}\bar{u}_{.t}\right.$$
$$\left. + \frac{T\sigma_\alpha^2}{\sigma_v^2 + T\sigma_\alpha^2}\frac{N\sigma_\rho^2}{\sigma_v^2 + N\sigma_\rho^2}\frac{2\sigma_v^2 + T\sigma_\alpha^2 + N\sigma_\rho^2}{\sigma_v^2(\sigma_v^2 + T\sigma_\alpha^2 + N\sigma_\rho^2)}\bar{u}_{..}\right] \quad (8)$$

where $\bar{u}_{..} = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} u_{it}$; $\bar{u}_{.t} = \frac{1}{N}\sum_{i=1}^{N} u_{it}$ and $\bar{u}_{i.} = \frac{1}{T}\sum_{t=1}^{T} u_{it}$.

To compute the conditional expectations given Z requires an expression for $E\{u_{it} \mid Z\}$. When g_5 is a one-to-one function this is equal to u_{it}. In general, the conditional expectation is more complicated. As T is small for most panel data studies a first step towards a general solution is to condition upon the time effects in u_{it}. This corresponds to treating the time effects in (3) as fixed unknown parameters[6]. The conditional expectations are then given by (6) and (8) with the terms involving σ_ρ^2 set to zero, while (7) reduces to

[6] It may also be more appropriate to treat the time effects in (1) as fixed. This decreases the difficulty in estimation as the fixed effects can be captured through time specific dummies. It also ensures that the approach is robust to incorrect specification of the distribution of the time effects, and relaxes the requirement for T to be large. However, the estimated fixed time effects in (1) will now comprise the direct effect of time and the indirect effect of time through the endogeneity of z_{it} and z^*_{it}. Naturally, this approach does not allow one to identify the correlation between the time effects.

$$E\{\varepsilon_t \mid U, \rho\} = \frac{\sigma_{\varepsilon\rho}}{\sigma_\rho^2}\rho_t. \tag{9}$$

Conditional on ρ, we have $E\{u_{it} \mid Z, \rho\} = E\{u_{it} \mid z_i, \rho\}$ and

$$E\{u_{it} \mid z_i, \rho\} = \rho_t + \int [\alpha_i + E\{v_{it} \mid z_{it}, \rho, \alpha_i\}]\, f(\alpha_i \mid z_i, \rho) d\alpha_i, \tag{10}$$

where $f(\alpha_i \mid z_i, \rho)$ denotes the conditional density of α_i. The conditional expectation of v_{it} given z_{it}, ρ and α_i is the generalized residual from (3) as, conditional on ρ and α_i, the errors from (3) are independent across observations. The form of the generalized residual depends on g_5 (see, for example, Gourieroux et al. [1987], Pagan and Vella [1989] or Vella [1993]).

The conditional distribution of α_i given z_i can be derived by using the result that[7]

$$f(\alpha_i \mid z_i, \rho) = \frac{f(z_i \mid \alpha_i, \rho) f(\alpha_i \mid \rho)}{f(z_i \mid \rho)} \tag{11}$$

where $f(z_i \mid \rho) = \int f(z_i \mid \alpha_i, \rho) f(\alpha_i \mid \rho) d\alpha_i$ is the likelihood contribution of individual i in (3)–(4), conditional on ρ. Furthermore, $f(\alpha_i \mid \rho) = f(\alpha_i)$ is a normal density and $f(z_i \mid \alpha_i, \rho)$ is the conditional likelihood contribution given α_i and ρ. Finally, $f(z_i \mid \alpha_i, \rho) = \prod_t f(z_{it} \mid \alpha_i, \rho)$, where $f(z_{it} \mid \alpha_i, \rho)$ has the form of the likelihood contribution in the cross sectional case.

Computation of the conditional expectations in (6)–(8) when g_5 is not a one-to-one mapping requires an expression for the likelihood contribution in an i.i.d. context, the corresponding generalized residual and the numerical evaluation of two one-dimensional integrals. Given consistent estimates of the parameters in (3) and (4), including the variances of the error components, the only unknowns in the expectations are the covariances. Due to the linearity these are easily estimated consistently in the second step from the conditional moment restrictions. The null hypothesis of no endogeneity is a linear restriction on these covariances and is thus easily tested.

The simplest way to obtain the second stage estimates is by standard GMM or nonlinear least squares based on (5).[8] Standard errors can be obtained in the usual way after accounting for the generation of the expectations (see appendix).

---

[7] We use $f(.)$ as generic notation for any density/mass function.

[8] It is possible to exploit the covariance matrix of the errors in an additional round of estimation. This requires analytical expressions for the conditional (co)variances of μ_i, ε_t and η_{it} and is therefore, in general, computationally unattractive.

4 Conditional Maximum Likelihood Estimation

When g_2 is not the identity mapping the conditional distribution of y_{it} is needed for estimation. This section discusses case II where, *inter alia*, $g_5(z_{it}^*) = z_{it}^* I[z_{it}^* \in A]$, with the identity mapping as a special case. Under our assumptions the conditional distribution of the error terms in (1) is still normal and the error components structure is preserved. Consequently, we can estimate (1)-(2) conditional on the estimated parameters from (3) using the random effects likelihood function, after making appropriate adjustments for the mean and noting that the variances now reflect the conditional variances. Moreover, the unconditional variances and the cross-equation covariances can be recovered from this procedure, while endogeneity tests are easily performed.

For computational purposes it is unattractive, due to the integration, to allow for random time effects in the error terms in nonlinear models. Accordingly, we treat the time effects in both equations as fixed and include time dummies in both mean functions. Write the joint density of $y_i = (y_{i1}, ..., y_{iT})'$ and z_i given $X_i = [x_{i1} ... x_{iT}]'$ (including the time effects) as

$$f_1(y_i \mid z_i, X_i; \theta^{(1)}, \theta^{(2)}) f_2(z_i \mid X_i; \theta^{(2)}), \tag{12}$$

where $\theta^{(1)}$ denotes $(\theta_1, \theta_2, \sigma_\mu^2, \sigma_\eta^2, \sigma_{\mu\alpha}, \sigma_{\eta v})$ and $\theta^{(2)}$ denotes $(\theta_4, \theta_5, \sigma_\varepsilon^2, \sigma_v^2)$. The conditional maximum likelihood approach involves first estimating $\theta^{(2)}$ by maximizing the marginal likelihood function of the z_i's. Subsequently, the conditional likelihood function

$$\prod_i f_1(y_i \mid z_i; \theta^{(1)}, \hat{\theta}^{(2)}) \tag{13}$$

is maximized with respect to $\theta^{(1)}$. Given our restrictions on g_3 and g_5 the latter step has the same computational complexity as when z_{it} is strictly exogenous. This is due to the fact that the conditional distribution of y_{it}^* given z_i and sample selection ($g_3 \neq 0$) is normal with expectations given in (5), and a covariance structure corresponding 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

$$\sigma_1^2 := V\{\nu_{1i}\} = \sigma_\eta^2 - \sigma_{\eta v}^2 \sigma_v^{-2}, \tag{14}$$

$$\sigma_2^2 := V\{\nu_{2,it}\} = \sigma_\mu^2 - \frac{T\sigma_{\mu\alpha}^2 \sigma_v^2 + 2\sigma_{\mu\alpha}\sigma_{\eta v}\sigma_v^2 - \sigma_{\eta v}^2 \sigma_\alpha^2}{\sigma_v^2(\sigma_v^2 + T\sigma_\alpha^2)}, \tag{15}$$

which follows from the usual matrix manipulations needed to derive a conditional covariance matrix. The results in (14) and (15) show that the error components structure is preserved and the conditional likelihood function of (1)–(2) has the same form as the marginal likelihood function without endogenous regressors or sample selection. The asymptotic variance of the conditional ML estimator, and an expression for the efficiency loss, is provided in the appendix. Under the null hypothesis of exogeneity, no loss of efficiency is incurred such that routinely computed t-statistics for the covariances provide efficient tests of endogeneity (compare Smith and Blundell [1986]).

The algebraic manipulations simplify if σ_1^2 and σ_2^2 replace the unconditional variances σ_η^2 and σ_μ^2 in $\theta^{(1)}$. 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_v^2 and σ_α^2, and the estimated covariances from the mean function.

5 Relationship with IV estimators

For several cases of our model alternative estimators are available which are more robust to the distributional assumptions. We now analyse the relationship between our two-step approach and the appropriate IV estimators. To do this, we concentrate on the special case where g_1 is linear in x_{it}, z_{it} and does not depend upon z_{it}^*, both g_2 and g_5 are identity mappings and $g_3 \neq 0$ (no sample selection). The model of interest is thus given by

$$y_{it} = x_{it}'\beta + z_{it}\psi + \mu_i + \eta_{it}, \tag{16}$$

where there is only one endogenous explanatory variable, z_{it}, and the time effects, if present, are included in x_{it}. To simplify notation we do not transform (16) to obtain a scalar error covariance matrix. This does not affect the primary result.

Let w_{it} denote a set of appropriate instruments for z_{it} (including x_{it}). It is well known that the linear IV estimator for $\theta_1 = (\beta', \psi)'$ in (16) can be obtained through least squares on

$$y_{it} = x_{it}'\beta + z_{it}\psi + \hat{\xi}_{it}\phi + e_{it}, \tag{17}$$

where e_{it} denotes a zero mean error term orthogonal to the regressors in this equation, and $\hat{\xi}_{it}$ is the residual from projecting z_{it} on the instruments w_{it}. In our two-step approach a reduced form for z_{it} is specified as

$$z_{it} = m_{it}'\gamma + u_{it}, \tag{18}$$

where $u_{it} = \alpha_i + v_{it}$ and m_{it} is a vector of exogenous variables, including x_{it}. The appropriate residuals are derived from (6) and (8) and are linear functions of $\hat{u}_{it}$ and $\sum_s \hat{u}_{is}$, where $\hat{u}_{it}$ is the least squares residual from (18).

Following Hausman and Taylor [1981], Amemiya and MaCurdy [1986] and Breusch, Mizon and Schmidt [1989], first assume that the endogeneity operates only through a nonzero covariance between μ_i and α_i. The appropriate residual in our two-step approach is proportional to $\sum_s u_{is}$. IV thus produces algebraically the same estimator if m_{it} and w_{it} are such that this residual is proportional to $\hat{\xi}_{it}$. 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 proportionality is attained if m_{it} in (18) is chosen as $[x_{it}, \bar{x}_i]$. Amemiya and MaCurdy [1986] extend the Hausman-Taylor approach by including $[x_{i1}, ..., x_{iT}]$ in the instrument set. The two-step approach is identical to this IV estimator when we include in the reduced form in (18) the exogenous variables from all periods, as well as x_{it} in deviation from its individual mean, i.e. if we choose $m_{it} = [x_{i1}, ..., x_{iT}, x_{it} - \bar{x}_i]$. If we extend m_{it} to include $z_{i1} - \bar{z}_i, ..., z_{iT} - \bar{z}_i$, our two-step procedure produces identical results to the IV estimator suggested by Breusch, Mizon and Schmidt [1989].

Next, assume that the endogeneity potentially operates through both μ_i and η_{it}. As no transformation of z_i will provide a valid instrument in this case, a possible instrument set is given by $w_{it} = [x_{it}, \bar{x}_i]$. If $m_{it} = w_{it}$, it can be shown (see appendix) that the linear IV estimator and the two-step estimator are, again, algebraically the same. This general equivalence of the two-step approach and IV indicates that normality is not required in the linear case and imposing normality does not increase efficiency. However, normality is needed to interpret the coefficients on the residuals as covariance-variance ratios.

6 Concluding remarks

This paper presents a two-step approach to estimating simultaneous equation panel data models with censored endogenous variables and sample selection. Compared to LIML, this approach is computationally simpler as only one-dimensional numerical integration is required. The costs are in a loss of efficiency, which depends, *inter alia*, upon the magnitude of the covariances responsible for the endogeneity.

In contrast to IV methods, our method can handle a larger variety of models. In particular it allows for truncation, censoring, sample selection and the inclusion of a latent explanatory variable[9]. Moreover, the two-step method identifies an additional number of parameters that may have economic appeal, including two or three sources of endogeneity. Direct test for endogeneity are also provided.

As our procedure requires distributional assumptions this may seem restrictive. However, the general functions (g_2 and g_5) allow for smooth monotonic transformations of the endogenous variables, while the assumption of normality is testable. Normality of the error components in (3) can be tested extending results from, for example, Ruud [1984]. In the second stage normality is needed to guarantee linearity of the conditional expectations and, in the conditional ML case, normality of the conditional distribution. A test for normality in this second stage can be based upon work by Lee [1984] and Gallant and Nychka [1987], by including powered up values of the conditional expectations of the errors to capture departures from normality.

Appendix

Covariance matrix estimation

First, consider case I. With fixed time effects the resulting estimator is $\sqrt{N}$ consistent and asymptotically normal under weak regularity conditions. The asymptotic covariance matrix can be obtained using the results in Newey [1984]. Let $\theta^{(2)}$ denote the parameter vector from (3)–(4) and let the $\sqrt{N}$ consistent maximum likelihood estimator for $\theta^{(2)}$ be given by $\hat{\theta}^{(2)}$ with asymptotic covariance matrix V_2. Write

$$g_1(x_{it}, z_{it}, z_{it}^*; \theta_1) = g_{11}(x_{it}, z_{it}; \theta_1^{(1)}) + \lambda z_{it}^*$$

and let

$$\theta^{(1)\prime} = (\theta_1^{(1)\prime}, \theta_2^{(1)\prime}),$$

with $\theta_2^{(1)} = (\lambda, \sigma_{\mu\alpha}, \sigma_{\eta v})'$. In vector notation, write

$$y_i = g_{11}(x_i, z_i; \theta_1^{(1)}) + Z_i(\theta^{(2)})\theta_2^{(1)} + e_i,$$

where $Z_i(\theta^{(2)})$ denotes the generated regressors and where e_i is an error term corresponding to the conditional distribution of $\iota_T \mu_i + \eta_i$, ι_T being a

[9] It is also a straightforward extension of the current framework to incorporate switching regression models.

T-dimensional vector of ones. Let Ω_i denote the variance of e_i for individual i and let

$$G_i = \frac{\partial g_{11}(x_i, z_i; \theta_1^{(1)})}{\partial \theta_1^{(1)\prime}}.$$

Defining

$$M_N = \frac{1}{N}\sum_{i=1}^{N} E\left\{[G_i \;\; Z_i(\theta^{(2)})]'[G_i \;\; Z_i(\theta^{(2)})]\right\},$$

$$V_N = \frac{1}{N}\sum_{i=1}^{N} E\left\{[G_i \;\; Z_i(\theta^{(2)})]'\Omega_i[G_i \;\; Z_i(\theta^{(2)})]\right\},$$

$$D_N = \frac{1}{N}\sum_{i=1}^{N} E\left\{[G_i \;\; Z_i(\theta^{(2)})]'\frac{\partial Z_i(\theta^{(2)})\theta_2^{(1)}}{\partial \theta^{(2)\prime}}\right\},$$

the asymptotic covariance matrix of the second step estimator for $\theta^{(1)}$ 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 be consistently estimated by replacing expectations with sample moments and unknown parameters by their estimators. The second part within brackets for V_1 is due to the generated regressors problem and equals zero if $\theta_2^{(1)} = 0$. An estimator for V_N, without specifiying the exact form of heteroskedasticity, is given by

$$\hat{V}_N = \frac{1}{N}\sum_{i=1}^{N}[\hat{G}_i \;\; Z_i(\hat{\theta}^{(2)})]'\hat{e}_i\hat{e}_i'[\hat{G}_i \;\; Z_i(\hat{\theta}^{(2)})],$$

where $\hat{e}_i$ is the T dimensional vector of residuals. While this estimator does not exploit the error components structure of the original errors it 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 should be adjusted to include only those observations used in estimation.

Next, consider case II. Let $\hat{\theta}^{(2)}$ denote the (marginal) ML estimator for $\theta^{(2)}$ with asymptotic covariance matrix V_2 and $\hat{\theta}^{(1)}$ the conditional ML estimator obtained from maximizing (13). Define

$$F_{11} = E\left\{-\frac{\partial^2 \ln f_1}{\partial \theta^{(1)}\theta^{(1)\prime}}\right\}, \qquad F_{12} = E\left\{\frac{\partial^2 \ln f_1}{\partial \theta^{(1)}\theta^{(2)\prime}}\right\}.$$

Then, it is easily verified from Taylor expansions that

$$\sqrt{N}(\hat{\theta}^{(1)} - \theta^{(1)}) = F_{11}^{-1} \left[\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{\partial \ln f_1}{\partial \theta^{(1)}} + F_{12}\sqrt{N}(\hat{\theta}^{(2)} - \theta^{(2)}) \right] + o_p(1).$$

Using the result that the two terms in square brackets are asymptotically independent (see Pierce [1982], Parke [1986]) it follows that the conditional ML estimator is asymptotically normal with covariance matrix

$$V_1 = F_{11}^{-1} \left[F_{11} + F_{12} V_2 F_{12}' \right] F_{11}^{-1}.$$

F_{11}^{-1} is estimated by standard ML programs. In two cases $\hat{\theta}^{(1)}$ attains the Cramér-Rao lower bound. The first corresponds to the null hypothesis of exogeneity ($\sigma_{\mu\alpha} = \sigma_{\eta v} = 0$) and is characterized by $F_{12} = 0$. The second case is given by

$$J = F_{22} - F_{12}' F_{11}^{-1} F_{12} = 0,$$

where

$$F_{22} = E\left\{ -\frac{\partial^2 \ln f_1}{\partial \theta^{(2)} \theta^{(2)\prime}} \right\}.$$

Note that J summarizes all information on $\theta^{(2)}$ contained in the conditional distribution f_1. We refer to $J = 0$ as the exactly identified case (see Rivers and Vuong [1988] for an example). In general, the efficiency loss due to the two-step nature is given by

$$F_{11}^{-1} F_{12} \left(V_2 - [V_2^{-1} + J]^{-1} \right) F_{12}' F_{11}^{-1}.$$

The relationship between IV and residual-type estimators

Let M denote the matrix of observations for the variables included in the reduced form (18), and let W denote the matrix of instruments for the IV estimator. P_v is projection matrix producing individual means and $Q_v = I - P_v$ (the "within" transformation). Let P_A denote the projection matrix upon the space spanned by the columns of A, while $Q_A = I - P_A$. Finally, denote G = [X,Z]. Each of the estimators for θ_1 can be written as $(G'PG)^{-1}G'Py$, where $P = P_{IV} = P_W$ for the IV estimator, $P = P_1 = Q_{P_v Q_M Z}$ for the two-step estimator with endogeneity operating through α_i only, and $P = P_2 = Q_{[P_v Q_M Z : Q_v Q_M Z]}$ for the general two-step estimator. Now the following results hold.

Theorem
Provided M has the same column space as $[Q_vM : P_vM]$,

$$(a)\quad P_{IV}G = P_1G \quad \text{if} \quad W = [M : Q_vZ],$$

$$(b)\quad P_{IV}G = P_2G \quad \text{if} \quad W = M.$$

Proof
Given that X is contained in M it is easily verified that $PX = 0$ for each of the choices for P. Furthermore, $P_2Z = P_MZ$, which proves (b). To prove (a), we use that $P_1Z = Z - P_vQ_MZ$, where

$$P_vQ_MZ = P_v(z - P_MZ) = P_v(z - P_{Q_vM}Q_vZ - P_{P_vM}P_vZ)$$

$$= P_v(I - P_{P_vM})Z.$$

Furthermore, for $W = [M : P_vZ]$ we have

$$Z - P_WZ = Q_W(Q_vZ + P_vZ) = P_vZ - P_WP_vZ =$$

$$P_vZ - P_{P_vW}P_vZ = P_vZ - P_{P_vM}P_vZ,$$

which proves (a).

Corollary 1 $(\sigma_{\eta v} = 0)$
The Hausman-Taylor estimator is identical to the two-step estimator for θ_1 for $M = [Q_vX : P_vX]$ or, equivalently, $M = [X : P_vX]$.
Corollary 2 $(\sigma_{\eta v} = 0)$
The Amemiya-MaCurdy estimator is identical to the two-step estimator for θ_1 for $M = [Q_vX : X^*]$.
Corollary 3 $(\sigma_{\eta v} = 0)$
The Breusch-Mizon-Schmidt estimator is identical to the two-step estimator for θ_1 for $M = [Q_vX : X^* : (Q_vZ)^*]$.
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 $P_vX^* = X^*$.
Corollary 4
The IV estimator with instruments $W = [X : P_vX]$ is identical to the two-step estimator for θ_1 for $M = W$.

References

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

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

Blundell, R. and R.J. Smith (1993): "Coherency and Estimation in Simultaneous Models with Censored or Qualitative Dependent Variables," working paper, Dept. of Applied Economics, University of Cambridge.

Gallant, A. and D. Nychka (1987): "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica*, 55, 363-390.

Gourieroux, C. and A. Monfort (1990): "Simulation-Based Inference. A Survey with Special Reference to Panel Data Models," *Journal of Econometrics*, 59, 5-33.

———, ———, E. Renault and A.Trognon (1987): "Generalized Residuals," *Journal of Econometrics*, 34, 5-32.

Hausman, J. and W. Taylor (1981): "Panel Data and Unobservable Individual Effects," *Econometrica*, 49, 1377-1398.

Heckman, J.J. (1978): "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 931-959.

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

——— (1981): "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete-Time-Discrete-Data Stochastic Process," in: C.F. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, MA.

Hsiao, C. (1986): *Analysis of Panel Data*, Cambridge University Press.

Honoré, B. (1992): "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica*, 60, 533-565.

——— (1993): "Orthogonality Conditions for Tobit Models with Fixed Effects and Lagged Dependent Variables", *Journal of Econometrics*, 59, 35-61.

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

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

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

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

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

Parke, W.R. (1986): "Pseudo Maximum Likelihood Estimation: The Asymptotic Distribution", *Annals of Statistics*, 14, 355-357.

Pierce, D.A. (1982): "The Asymptotic Effect of Substituting Estimators for Parameters in Certain Types of Statistics," *Annals of Statistics*, 10, 475-478.

Peters, S. and R.J. Smith (1991): "Distributional Specification Tests against Semi-Parametric Alternatives", *Journal of Econometrics*, 47, 175-194.

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

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

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

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

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

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

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

(For previous papers please consult previous discussion papers.)

No.	Author(s)	Title
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
9360	W.B. van den Hout and J.P.C. Blanc	The Power-Series Algorithm Extended to the *BMAP/PH*/1 Queue
9361	R. Heuts and J. de Klein	An (s,q) Inventory Model with Stochastic and Interrelated Lead Times
9362	K.-E. Wärneryd	A Closer Look at Economic Psychology
9363	P.J.-J. Herings	On the Connectedness of the Set of Constrained Equilibria
9364	P.J.-J. Herings	A Note on "Macroeconomic Policy in a Two-Party System as a Repeated Game"
9365	F. van der Ploeg and A. L. Bovenberg	Direct Crowding Out, Optimal Taxation and Pollution Abatement
9366	M. Pradhan	Sector Participation in Labour Supply Models: Preferences or Rationing?
9367	H.G. Bloemen and A. Kapteyn	The Estimation of Utility Consistent Labor Supply Models by Means of Simulated Scores
9368	M.R. Baye, D. Kovenock and C.G. de Vries	The Solution to the Tullock Rent-Seeking Game When R > 2: Mixed-Strategy Equilibria and Mean Dissipation Rates
9369	T. van de Klundert and S. Smulders	The Welfare Consequences of Different Regimes of Oligopolistic Competition in a Growing Economy with Firm-Specific Knowledge
9370	G. van der Laan and D. Talman	Intersection Theorems on the Simplotope
9371	S. Muto	Alternating-Move Preplays and $vN - M$ Stable Sets in Two Person Strategic Form Games
9372	S. Muto	Voters' Power in Indirect Voting Systems with Political Parties: the Square Root Effect

No.	Author(s)	Title
9373	S. Smulders and R. Gradus	Pollution Abatement and Long-term Growth
9374	C. Fernandez, J. Osiewalski and M.F.J. Steel	Marginal Equivalence in ν-Spherical Models
9375	E. van Damme	Evolutionary Game Theory
9376	P.M. Kort	Pollution Control and the Dynamics of the Firm: the Effects of Market Based Instruments on Optimal Firm Investments
9377	A. L. Bovenberg and F. van der Ploeg	Optimal Taxation, Public Goods and Environmental Policy with Involuntary Unemployment
9378	F. Thuijsman, B. Peleg, M. Amitai & A. Shmida	Automata, Matching and Foraging Behavior of Bees
9379	A. Lejour and H. Verbon	Capital Mobility and Social Insurance in an Integrated Market
9380	C. Fernandez, J. Osiewalski and M. Steel	The Continuous Multivariate Location-Scale Model Revisited: A Tale of Robustness
9381	F. de Jong	Specification, Solution and Estimation of a Discrete Time Target Zone Model of EMS Exchange Rates
9401	J.P.C. Kleijnen and R.Y. Rubinstein	Monte Carlo Sampling and Variance Reduction Techniques
9402	F.C. Drost and B.J.M. Werker	Closing the Garch Gap: Continuous Time Garch Modeling
9403	A. Kapteyn	The Measurement of Household Cost Functions: Revealed Preference Versus Subjective Measures
9404	H.G. Bloemen	Job Search, Search Intensity and Labour Market Transitions: An Empirical Exercise
9405	P.W.J. De Bijl	Moral Hazard and Noisy Information Disclosure
9406	A. De Waegenaere	Redistribution of Risk Through Incomplete Markets with Trading Constraints
9407	A. van den Nouweland, P. Borm, W. van Golstein Brouwers, R. Groot Bruinderink, and S. Tijs	A Game Theoretic Approach to Problems in Telecommunication
9408	A.L. Bovenberg and F. van der Ploeg	Consequences of Environmental Tax Reform for Involuntary Unemployment and Welfare

No.	Author(s)	Title
9409	P. Smit	Arnoldi Type Methods for Eigenvalue Calculation: Theory and Experiments
9410	J. Eichberger and D. Kelsey	Non-additive Beliefs and Game Theory
9411	N. Dagan, R. Serrano and O. Volij	A Non-cooperative View of Consistent Bankruptcy Rules
9412	H. Bester and E. Petrakis	Coupons and Oligopolistic Price Discrimination
9413	G. Koop, J. Osiewalski and M.F.J. Steel	Bayesian Efficiency Analysis with a Flexible Form: The AIM Cost Function
9414	C. Kilby	World Bank-Borrower Relations and Project Supervision
9415	H. Bester	A Bargaining Model of Financial Intermediation
9416	J.J.G. Lemmen and S.C.W. Eijffinger	The Price Approach to Financial Integration: Decomposing European Money Market Interest Rate Differentials
9417	J. de la Horra and C. Fernandez	Sensitivity to Prior Independence via Farlie-Gumbel -Morgenstern Model
9418	D. Talman and Z. Yang	A Simplicial Algorithm for Computing Proper Nash Equilibria of Finite Games
9419	H.J. Bierens	Nonparametric Cointegration Tests
9420	G. van der Laan, D. Talman and Z. Yang	Intersection Theorems on Polytopes
9421	R. van den Brink and R.P. Gilles	Ranking the Nodes in Directed and Weighted Directed Graphs
9422	A. van Soest	Youth Minimum Wage Rates: The Dutch Experience
9423	N. Dagan and O. Volij	Bilateral Comparisons and Consistent Fair Division Rules in the Context of Bankruptcy Problems
9424	R. van den Brink and P. Borm	Digraph Competitions and Cooperative Games
9425	P.H.M. Ruys and R.P. Gilles	The Interdependence between Production and Allocation
9426	T. Callan and A. van Soest	Family Labour Supply and Taxes in Ireland
9427	R.M.W.J. Beetsma and F. van der Ploeg	Macroeconomic Stabilisation and Intervention Policy under an Exchange Rate Band

No.	Author(s)	Title
9428	J.P.C. Kleijnen and W. van Groenendaal	Two-stage versus Sequential Sample-size Determination in Regression Analysis of Simulation Experiments
9429	M. Pradhan and A. van Soest	Household Labour Supply in Urban Areas of a Developing Country
9430	P.J.J. Herings	Endogenously Determined Price Rigidities
9431	H.A. Keuzenkamp and J.R. Magnus	On Tests and Significance in Econometrics
9432	C. Dang, D. Talman and Z. Wang	A Homotopy Approach to the Computation of Economic Equilibria on the Unit Simplex
9433	R. van den Brink	An Axiomatization of the Disjunctive Permission Value for Games with a Permission Structure
9434	C. Veld	Warrant Pricing: A Review of Empirical Research
9435	V. Feltkamp, S. Tijs and S. Muto	Bird's Tree Allocations Revisited
9436	G.-J. Otten, P. Borm, B. Peleg and S. Tijs	The MC-value for Monotonic NTU-Games
9437	S. Hurkens	Learning by Forgetful Players: From Primitive Formations to Persistent Retracts
9438	J.-J. Herings, D. Talman, and Z. Yang	The Computation of a Continuum of Constrained Equilibria
9439	E. Schaling and D. Smyth	The Effects of Inflation on Growth and Fluctuations in Dynamic Macroeconomic Models
9440	J. Arin and V. Feltkamp	The Nucleolus and Kernel of Veto-rich Transferable Utility Games
9441	P.-J. Jost	On the Role of Commitment in a Class of Signalling Problems
9442	J. Bendor, D. Mookherjee, and D. Ray	Aspirations, Adaptive Learning and Cooperation in Repeated Games
9443	G. van der Laan, D. Talman and Zaifu Yang	Modelling Cooperative Games in Permutational Structure
9444	G.J. Almekinders and S.C.W. Eijffinger	Accounting for Daily Bundesbank and Federal Reserve Intervention: A Friction Model with a GARCH Application
9445	A. De Waegenaere	Equilibria in Incomplete Financial Markets with Portfolio Constraints and Transaction Costs
9446	E. Schaling and D. Smyth	The Effects of Inflation on Growth and Fluctuations in Dynamic Macroeconomic Models

No.	Author(s)	Title
9447	G. Koop, J. Osiewalski and M.F.J. Steel	Hospital Efficiency Analysis Through Individual Effects: A Bayesian Approach
9448	H. Hamers, J. Suijs, S. Tijs and P. Borm	The Split Core for Sequencing Games
9449	G.-J. Otten, H. Peters, and O. Volij	Two Characterizations of the Uniform Rule for Division Problems with Single-Peaked Preferences
9450	A.L. Bovenberg and S.A. Smulders	Transitional Impacts of Environmental Policy in an Endogenous Growth Model
9451	F. Verboven	International Price Discrimination in the European Car Market: An Econometric Model of Oligopoly Behavior with Product Differentiation
9452	P.J.-J. Herings	A Globally and Universally Stable Price Adjustment Process
9453	D. Diamantaras, R.P. Gilles and S. Scotchmer	A Note on the Decentralization of Pareto Optima in Economies with Public Projects and Nonessential Private Goods
9454	F. de Jong, T. Nijman and A. Röell	Price Effects of Trading and Components of the Bid-ask Spread on the Paris Bourse
9455	F. Vella and M. Verbeek	Two-Step Estimation of Simultaneous Equation Panel Data Models with Censored Endogenous Variables

PO BOX 90153 5000 LE TILBURG THE NETHERLANDS