working paper
department
of economics

MULTIPLICATIVE PANEL DATA MODELS WITHOUT THE STRICT EXOGENEITY ASSUMPTION

Jeffrey M. Wooldridge

Number 574 March 1991

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

MULTIPLICATIVE PANEL DATA MODELS WITHOUT THE STRICT EXOGENEITY ASSUMPTION

Jeffrey M. Wooldridge

Number 574 March 1991

MULTIPLICATIVE PANEL DATA MODELS WITHOUT THE STRICT EXOGENEITY ASSUMPTION

Jeffrey M. Wooldridge
Department of Economics
Massachusetts Institute of Technology, E52-262C
Cambridge, MA 02139
(617) 253-3488
JMWOOLDR@ATHENA.MIT.EDU

December 1990
Revised: March 1991

I am grateful to Whitney Newey, Danny Quah, and Paul Ruud for helpful comments and suggestions.

Abstract

This paper studies estimation of multiplicative, unobserved components panel data models without imposing the strict exogeneity assumption on the explanatory variables. The method of moments estimators proposed have significant robustness properties; they require only a conditional mean assumption, and apply to models with lagged dependent variables, finite distributed lag models that allow arbitrary feedback from the explained to the explanatory variables, and models with contemporaneous endogeneity. The model can be applied to any nonnegative explained variable, including count variables, binary variables, and continuously distributed nonnegative variables. An extension of the basic model applies to certain Euler equation applications with individual data.

1. Introduction

Textbook treatments of the standard linear unobserved effects model assume that the explanatory variables are nonrandom. While the assumption of fixed regressors is mostly harmless in analyzing pure cross section problems, in models with a time dimension it is operationally the same as assuming strict exogeneity of the explanatory variables. Importantly, as shown, for example, in Chamberlain (1984), the usual fixed effects (within) estimator is inconsistent when the explanatory variables are not strictly exogenous. For certain applications the strict exogeneity assumption can be quite restrictive. One case where strict exogeneity cannot hold is when a lagged dependent variable is included among the regressors, as in the models studied by Balestra and Nerlove (1966) and Anderson and Hsiao (1982). But the strict exogeneity assumption can also be restrictive in static and finite distributed lag models, as it rules out certain kinds of feedback from the explained variable to future explanatory variables. Standard fixed effects estimators in nonlinear panel data models, such as the count models considered by Hausman, Hall, and Griliches (1984) (HHG), are also inconsistent if the explanatory variables are not strictly exogenous; see Chamberlain (1984) and Wooldridge (1990).

To formalize the discussion, let y_t, $t=1,\ldots,T$, denote the variable to be explained at time t and let z_t, $t=1,\ldots,T$, denote a vector of contemporaneous conditioning variables at time t. Let ϕ denote the latent, time-constant effect. Interest lies in $E(y_t|\phi,x_t)$, where the vector of explanatory variables x_t consists of elements from $(z_t,y_{t-1},z_{t-1},\ldots,z_1,y_1)$, with the restriction that the lag lengths appearing in x_t do not depend on t. Examples are $x_t \equiv y_{t-1}$, $x_t \equiv z_t$, $x_t \equiv (y_{t-1},z_{t-1},y_{t-2},z_{t-2})$, and $x_t \equiv$

$(z_t, \ldots, z_{t-Q})$. When interest centers on the conditional mean, the strict exogeneity assumption (conditional on the latent effect) is stated as

$$E(y_t | \phi, x_1, \ldots, x_T) = E(y_t | \phi, x_t). \tag{1.1}$$

Equation (1.1) is necessarily false if x_t contains y_{t-1} or other lags of y, the result mentioned above. Further, (1.1) can easily fail even if x_t contains only current and lagged z_t, i.e. $x_t \equiv (z_t, \ldots, z_{t-Q})$. Provided enough lags of z_t can be included in x_t to account for all of the distributed lag dynamics, (1.1) is equivalent to

$$E(y_t | \phi, z_1, \ldots, z_T) = E(y_t | \phi, z_1, \ldots z_t). \tag{1.2}$$

When all conditional expectations are linear, (1.2) is equivalent to the Granger noncausality condition

$$E(z_t | \phi, z_{t-1}, y_{t-1}, \ldots, z_1, y_1) = E(z_t | \phi, z_{t-1}, \ldots, z_1) \tag{1.3}$$

(Chamberlain (1982)), so that past y does not help to predict z_t once the latent effect and past z's have been controlled for. When the conditional expectations are nonlinear, neither condition implies the other (Chamberlain (1982)), but if there is reason to believe that (1.3) fails then one should also be concerned about the strict exogeneity assumption (1.2).

It is easy to think of cases where (1.3) might fail, especially when z_t contains variables that are directly or indirectly affected by the economic unit under analysis. An example is the Hausman, Hall, and Griliches (1984) (HHG) patents-R&D application. The vector z_t contains expenditures on R&D by a particular firm and y_t is patents applied for or awarded to the firm in year t. Condition (1.3) requires that firms do not adjust future R&D expenditures based on the number of patents awarded in previous years. This condition could easily fail, depending on the underlying model of firm behavior.

A second example appears in Rose (1990), who uses panel data to examine the effects of financial variables -- in particular, measures of profitability -- on airline accident counts. For this application, condition (1.3) requires that the accident rate today does not affect current or subsequent profitability, an assumption that could be false. It is important to see that, when fixed effects techniques are used, lagging variables that are suspected of being contemporaneously endogenous, such as measures of profitability, does not solve the endogeneity problem. This is because (1.1) requires that all future values of explanatory variable be uncorrelated with the implicit error at time t. For example, if $\mathbf{x}_t$ contains profit variables dated at time t-1, then $\mathbf{x}_{t+1}$ contains these variables dated at time t, and these must be uncorrelated with the error at time t (and the errors at all other time periods) for standard fixed effects estimators to be consistent.

One important consequence of this discussion is that fixed-effects type estimators, which allow arbitrary dependence between the unobserved effect ϕ and the explanatory variables $(\mathbf{x}_1,\ldots,\mathbf{x}_T)$, are <u>not</u> necessarily more robust than their non fixed-effects counterparts. For example, in an additive, linear context, consistency of the usual OLS estimator using the pooled data relies on the assumption

$$E(\phi|\mathbf{x}_t) = E(\phi),\ t=1,\ldots,T, \tag{1.4}$$

but does not require (1.1). As a result, a significant difference between the pooled OLS estimator and the fixed effects estimator can be due to violations of (1.4) <u>or</u> (1.1). The fixed effects estimator is not more robust than pooled OLS, unless one takes (1.1) as a maintained assumption. Similar statements apply in nonlinear contexts.[1]

For the case of linear models, Holtz-Eakin, Newey, and Rosen (1988) (HNR) have recently offered a general approach to consistently estimating linear unobserved effects models without the strict exogeneity assumption. HNR focus on vector autoregressions, but their approach can be used whenever x_t is not strictly exogeneous, provided that the errors are unpredictable given at least some lagged variables of observable variables. HNR's instrumental variables method consists of first differencing to remove the unobserved effect, and then selecting instruments that can be used in the first differenced equations.

This paper offers estimation methods for nonlinear, multiplicative unobserved effects models without the strict exogeneity assumption. In an earlier paper (Wooldridge (1990)) I showed how to obtain consistent and asymptotically normal estimators in multiplicative models when the strict exogeneity assumption is maintained but without distributional assumptions (see Chamberlain (1990) for another approach to this problem). Here, I show how to relax the strict exogeneity assumption, thereby allowing for arbitrary feedback from the explained to future explanatory variables. In addition, the model is expanded to allow nonlinear transformations of the endogenous variables which depend on unknown parameters, so that these methods can be applied to estimating Euler equations with unobserved effects using individual data.

Section 2 introduces the basic model and the assumptions, and offers a differencing-like transformation that can be used to construct orthogonality conditions in the observable data. Generalized method of moments estimation of the model is covered in section 3. Several examples of the basic framework are provided in section 4. The model is extended in section 5 to allow for nonlinear transformations of the explained variables that depend on unknown parameters. Section 6 contains concluding remarks.

2. A Multiplicative Unobserved Effects Model without Strict Exogeneity

Let $\{(\mathbf{y}_i, \mathbf{x}_i, \phi_i): i=1,2,\ldots\}$ be a sequence of i.i.d. random variables, where $\mathbf{y}_i \equiv (y_{i1},\ldots,y_{iT})'$ is an observable Tx1 vector of nonnegative variables, $\mathbf{x}_i \equiv (\mathbf{x}'_{i1},\mathbf{x}'_{i2},\ldots,\mathbf{x}'_{iT})'$ is a TxK matrix of observable explanatory variables ($\mathbf{x}_{it}$ is 1xK, t=1,...,T), and ϕ_i is a nonnegative unobservable random scalar. The explanatory variables $\mathbf{x}_{it}$ can contain lags of y_{it} or current and lagged values of some conditioning variables, say z_{it}. Thus, a finite distributed lag (DL) model would take $\mathbf{x}_{it} \equiv (z_{it},\ldots,z_{i,t-Q})$, while a simple first-order dynamic model model would take $\mathbf{x}_{it} \equiv (z_{it}, y_{i,t-1})$. For simplicitly, the number of lags showing up in $\mathbf{x}_{it}$ is assumed not to depend on t; hence the assumption that $\mathbf{x}_{it}$ is 1xK for all t. Note that $\mathbf{x}_{it}$ can contain a set of time dummies that are constant across i. For notational simplicity, the time index appearing in the model starts at t = 1, which should be interpreted as the first time period for which a full set of explanatory variables is available. No restrictions are imposed on the distribution of ϕ_i given the observed explanatory variables $\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}$.

The basic model studied in this paper can be expressed as

$$y_{it} = \phi_i \mu(\mathbf{x}_{it},\beta_o) u_{it}, \quad t=1,\ldots,T, \tag{2.1}$$

where β_o is a Px1 vector of unknown parameters, $\mu(\mathbf{x}_{it},\beta) > 0$ for all x_{it} and all β is a known function, and u_{it} is an unobserved multiplicative error. Note that y_{it} can be a binary variable, a count variable, or a nonnegative continuous variable. The most popular form for μ is $\mu(x_{it},\beta) \equiv \exp(x_{it}\beta)$, but there is no need to use this particular functional form. A more flexible form is $\mu(\mathbf{x}_{it},\beta) = \exp[x_{it}(\rho)\delta]$, where $\mathbf{x}_{it}(\rho)$ denotes a vector of transformed explanatory variables, such as Box-Cox transformed variables. Depending on the form of μ, not all of the elements of $\mathbf{x}_{it}$ need be time-varying; this is discussed further in section 3. Section 5 extends (2.1) to the case where y_{it} is replaced by the scalar transformation $\tau(y_{it},\lambda_o)$, where y_{it} can now be a vector and λ_o is a vector of unknown parameters.

In Wooldridge (1990) I analyzed (2.1) under the strict exogeneity assumption

$$E(u_{it}|\phi_i,\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}) = 1,\ t=1,\ldots,T, \tag{2.2}$$

which is equivalent to assuming $E(y_{it}|\phi_i,\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}) = E(y_{it}|\phi_i,\mathbf{x}_{it}) = \phi_i\mu(\mathbf{x}_{it},\beta_o)$. (Note that, because ϕ_i has an unrestricted mean, the assumption that $E(u_{it}) = 1$ is without loss of generality.) As stated in the introduction, (2.2) rules out lagged dependent variables in $\mathbf{x}_{it}$ and, more generally, certain forms of feedback from y_{it} to $\mathbf{x}_{ir}$, $r > t$. Under (2.1) and (2.2), it is possible to consistently estimate β_o, as shown by Chamberlain (1990) and Wooldridge (1990). One simple consistent estimator, regardless of the actual distribution of y_{it} conditional on ϕ_i and $\mathbf{x}_i$, is the fixed effects Poisson or multinomial quasi-conditional maximum likelihood estimator (QCMLE) (Wooldridge (1990)). Unfortunately, this estimator is no longer guaranteed to be consistent if (2.2) fails.

Instead of (2.2), consider the weaker condition

$$E(u_{it}|\phi_i,\mathbf{x}_{i1},\ldots,\mathbf{x}_{it}) = 1,\quad t=1,\ldots,T, \tag{2.3}$$

which is equivalent to

$$E(y_{it}|\phi_i,\mathbf{x}_{i1},\ldots,\mathbf{x}_{it}) = \phi_i\mu(\mathbf{x}_{it},\beta_o).[2] \tag{2.4}$$

Condition (2.3) is perhaps not as weak as one would like; in particular, it presumes a kind of correct dynamic specification. This is not very restrictive for autoregressive, distributed lag, and other dynamic models because one usually assumes that enough lags are included in $\mathbf{x}_{it}$ so that (2.3) holds (although the truncation effect in finite distributed lag models can be problematical). Condition (2.3) is most restrictive for static models where $\mathbf{x}_{it} \equiv \mathbf{z}_{it}$ and one does not care about distributed lag dynamics. Some of the discussion focuses on (2.3), but below I offer some weaker assumptions. At a minimum, (2.3) allows feedback from y_{it} to $\mathbf{x}_{ir}$, $r > t$, so it is much more applicable than (2.2). It essentially corresponds to the linear counterpart used by HNR, except that here $\mathbf{x}_{it}$ might not include lagged dependent variables.

The primary problem addressed in this paper is estimation of β_o under (2.1) and (2.3). As mentioned above, the methods of Chamberlain (1990) and Wooldridge (1990) no longer produce consistent estimators. The differencing method proposed by HNR does not work in the multiplicative case. Nevertheless, there is a transformation that leads to orthogonality conditions that can be exploited in estimation. To state these, define $\mu_{it}(\beta) \equiv \mu(\mathbf{x}_{it},\beta)$ and

$$r_{it}(\beta) \equiv y_{it}/\mu_{it}(\beta) - y_{i,t-1}/\mu_{i,t-1}(\beta), \quad t=2,\ldots,T. \qquad (2.5)$$

Note that $r_{it}(\beta)$ depends only on the observed data and the parameter vector β through the known function $\mu(\mathbf{x}_{it},\beta)$. From (2.1), $r_{it}(\beta_o)$ can be expressed as

$$r_{it}(\beta_o) = \phi_i u_{it} - \phi_i u_{i,t-1} = \phi_i(u_{it} - u_{i,t-1}). \qquad (2.6)$$

This representation leads to a simple but useful lemma.

LEMMA 2.1: Under (2.1) and (2.3),

$$E[r_{it}(\beta_o)|\phi_i,\mathbf{x}_{i1},\ldots,\mathbf{x}_{i,t-1}] = 0, \quad t=2,\ldots,T. \qquad (2.7)$$

PROOF: From (2.6), it is enough to show that

$$E[\phi_i(u_{it} - u_{i,t-1})|\phi_i, x_{i1}, \ldots, x_{i,t-1}] = 0.$$

But this expectation equals

$$\phi_i\{E(u_{it}|\phi_i, x_{i1}, \ldots, x_{i,t-1}) - E(u_{i,t-1}|\phi_i, x_{i1}, \ldots, x_{i,t-1})\},$$

$$= \phi_i\{1 - 1\}$$

$$= 0$$

by (2.3) and the law of iterated expectations. ■

Lemma 2.1 immediately provides orthogonality conditions that can be used to estimate β_o under standard identification assumptions. Let w_{it} be a $1xL_t$ vector of functions of $x_{i1}, \ldots, x_{i,t-1}$, $t=2, \ldots, T$. Then, from Lemma 2.1 and finite moment assumptions,

$$E[w'_{it} r_{it}(\beta_o)] = 0, \quad t=2, \ldots, T; \tag{2.8}$$

for $\beta \neq \beta_o$, $E[w'_{it} r_{it}(\beta)] \neq 0$ in general. Note that the list of available instruments generally grows with t under (2.3).

From Lemma 2.1 it is clear that (2.3) can be relaxed to some degree. More generally, it is useful to view w_{it} as a function of some <u>subset</u> of $(x_{i1}, \ldots, x_{i,t-1})$, say v_{it}, such that

$$E(u_{it}|\phi_i, v_{it}) = E(u_{i,t-1}|\phi_i, v_{it}) = 1. \tag{2.9}$$

For example, if in (2.1) x_{it} contains some endogenous elements that are correlated with u_{it}, then v_{it} would exclude the endogenous elements of $x_{i,t-1}$. This causes no problems provided v_{it} is rich enough to identify β_o. As is usual in nonlinear contexts, the requirements for formally identifying β_o are modest. Some additional examples are given in section 4.

Sometimes there are elements of x_{it} that can be maintained as being strictly exogenous. In this case, all leads and lags of strictly exogenous variables can be included in v_{it}.

3. Method of Moments Estimation

The orthogonality conditions (2.8) can be used to obtain a variety of method of moments estimators. In what follows, assume that the $1xL_t$ vector of instruments $\hat{w}_{it}$ is obtained as

$$\hat{w}_{it} = g_t(v_{it}, \hat{\gamma}_N), \quad t=2,\ldots,T,$$

where g_t is a known, continuously differentiable function of γ, v_{it} satisfies (2.9), and $\hat{\gamma}_N$ is a Gx1 vector estimator of some nuisance parameters. A regularity condition used in the subsequent analysis is

$$\sqrt{N}(\hat{\gamma}_N - \gamma^*) = O_p(1) \tag{3.1}$$

for <u>some</u> $\gamma^* \in \mathbb{R}^G$. This allows for the case where $\hat{\gamma}_N$ is an initial $\sqrt{N}$-consistent estimator of β_o, as well as when $\hat{\gamma}_N$ estimates parameters outside of the model (2.1). It is also useful to note that $\hat{\gamma}_N$ can be an initial <u>inconsistent</u> estimator of β_o, such as the multinomial QCMLE when the strict exogeneity assumption fails.

Define the (T-1)x1 vector $r_i(\beta) \equiv (r_{i2}(\beta),\ldots,r_{iT}(\beta))'$. Let $L \equiv L_2 + L_3 + \ldots + L_T$, and define a (T-1)xL matrix of instruments as

$$\hat{w}_i \equiv w_i(\hat{\gamma}_N) \equiv \begin{bmatrix} \hat{w}_{i2} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \hat{w}_{i3} & 0 & \cdots & 0 & 0 \\ & & & \cdot & & \\ & & & \cdot & & 0 \\ 0 & 0 & 0 & \cdot & 0 & \hat{w}_{iT} \end{bmatrix}. \tag{3.2}$$

Under (2.1) and (2.9),

$$E[w_i(\gamma^*)'r_i(\beta_o)] = 0. \tag{3.3}$$

Thus, as usual, β_o can be estimated by setting the Lx1 vector

$$N^{-1}\sum_{i=1}^{N}\hat{w}_i' r_i(\beta)$$

as close as possible to zero. Given the choice of instruments, the minimum chi square estimator $\hat{\beta}_N$ is obtained by solving

$$\min_{\beta}\left[\sum_{i=1}^{N}\hat{w}_i' r_i(\beta)\right]'\hat{\Omega}_N^{-1}\left[\sum_{i=1}^{N}\hat{w}_i' r_i(\beta)\right], \tag{3.4}$$

where the LxL symmetric, positive definite matrix $\hat{\Omega}_N$ is a consistent estimator of

$$\Omega^* \equiv E[w_i(\gamma^*)' r_i(\beta_o) r_i(\beta_o)' w_i(\gamma^*)]. \tag{3.5}$$

In many cases $\gamma^* = \beta_o$, so if $\hat{\gamma}_N$ is a $\surd$N-consistent estimator of β_o, then $\hat{\Omega}_N$ can be taken to be

$$\hat{\Omega}_N \equiv N^{-1}\sum_{i=1}^{N} w_i(\hat{\gamma}_N)' r_i(\hat{\gamma}_N) r_i(\hat{\gamma}_N)' w_i(\hat{\gamma}_N). \tag{3.6}$$

The asymptotic variance of the minimum chi-square estimator is obtained from Hansen (1982). Let $\nabla_\beta r_i(\beta)$ denote the (T-1)xP derivative of $r_i(\beta)$ with respect to β, with rows given by the 1xP vectors

$$\nabla_\beta r_{it}(\beta) = -\{y_{it}/[\mu_{it}(\beta)]^2\}\nabla_\beta \mu_{it}(\beta) \tag{3.7}$$

$$+ \{y_{i,t-1}/[\mu_{i,t-1}(\beta)]^2\}\nabla_\beta \mu_{i,t-1}(\beta), \quad t=2,\ldots,T.$$

Here, $\nabla_\beta \mu_{it}(\beta)$ is the 1xP derivative of $\mu_{it}(\beta)$. The asymptotic variance of $\surd N(\hat{\beta}_N - \beta_o)$ is $(R^{*\prime}\Omega^{*-1}R^*)^{-1}$, where R^* is the LxP matrix

$$R^* \equiv E[w_i(\gamma^*)'\nabla_\beta r_i(\beta_o)]. \tag{3.8}$$

The asymptotic variance of $\hat{\beta}_N$ is estimated by $(\hat{R}_N'\hat{\Omega}_N^{-1}\hat{R}_N)^{-1}/N$, where

$$\hat{R}_N \equiv N^{-1}\sum_{i=1}^{N} w_i(\hat{\gamma}_N)'\nabla_\beta r_i(\hat{\beta}_N). \tag{3.9}$$

When $\gamma^* = \beta_o$, so that $\hat{\gamma}_N$ is an initial $\surd$N-consistent estimator of β_o, it is straightforward to obtain a one-step estimator which is first-order equivalent to the corresponding minimum chi-square estimator. Generally, given a $\surd N$-consistent estimator $\ddot{\beta}_N$ of β_o, the one-step estimator $\hat{\beta}_N$ is given by

$$\hat{\beta}_N = \ddot{\beta}_N + (\ddot{R}_N' \ddot{\Omega}_N^{-1} \ddot{R}_N)^{-1} N^{-1} \sum_{i=1}^{N} \ddot{s}_i, \tag{3.10}$$

where $\ddot{s}_i \equiv \ddot{R}_N' \ddot{\Omega}_N^{-1} \ddot{w}_i' \ddot{r}_i$ is a Px1 vector and all quantities on the right hand side are evaluated at $\ddot{\beta}_N$.

For the general minimum chi-square estimator that uses L > P nonredundant orthogonality conditions, a test of (2.1) and (2.9) is easily obtained from the GMM overidentification statistic, given by

$$N^{-1}\left[\sum_{i=1}^{N} \hat{w}_i' r_i(\hat{\beta}_N)\right]' \hat{\Omega}_N^{-1} \left[\sum_{i=1}^{N} \hat{w}_i' r_i(\hat{\beta}_N)\right]; \tag{3.11}$$

under (2.1), (2.9), and standard regularity conditions, (3.11) has a limiting χ^2_{L-P} distribution. This statistic can detect misspecified functional form as well as instruments that are not appropriately orthogonal to u_{it}.

For models where x_{it} contains no lags of y_{it} or no contemporaneous endogenous elements, one might want to estimate the model under the strict exogeneity assumption. As mentioned above, a simple, consistent estimator is given by the multinomial quasi-conditional MLE. Given this estimator, it is straightforward to test the strict exogeneity assumption by constructing additional orthogonality conditions and computing a specification test; see Wooldridge (1990) for details. Alternatively, one can include functions of $x_{it},\ldots,x_{iT}$ in w_{it} above and compute the minimum chi-square test statistic. Of course, this test cannot distinguish between a misspecified functional form and violation of the strict exogeneity assumption.

Before turning to some examples, I should briefly discuss identification of β_o. Although identification in these contexts is frequently taken on faith, it is sometimes easy to show that β_o is not identified. One important case where identification fails is when $\mathbf{x}_{it}$ contains some time-constant variables and $\mu(\mathbf{x}_{it},\beta) \equiv \exp(\mathbf{x}_{it}\beta)$. Partition $\mathbf{x}_{it}$ as $(\mathbf{x}_{it1},\mathbf{x}_{i2})$, where $\mathbf{x}_{i2}$ is a $1xK_2$ vector of time constant variables. Then, for any β_2,

$$r_{it}(\beta_{o1},\beta_2) = \phi_i \exp[x_{i2}(\beta_{o2} - \beta_2)](u_{it} - u_{i,t-1}),$$

so that

$$E[r_{it}(\beta_{o1},\beta_2)|\phi_i,v_{it}] = 0,$$

whenever v_{it} contains $\mathbf{x}_{i2}$, which is usually the case. Thus, as in the case of strictly exogenous explanatory variables, coefficients on the time-constant regressors cannot be identified. The term $\exp(\mathbf{x}_{i2}\beta_{o2})$ is simply absorbed in ϕ_i. This does not mean that time constant variables can never be included in the analysis; it depends on the functional form for $\mu(\mathbf{x}_{it},\beta)$. For example, interaction terms between time-varying and time constant variables can be included in an exponential model.

4. Examples

This section covers some specific choices of instruments in models that might arise in practice. The question of efficient estimation is difficult enough under the strict exogeneity assumption (see Chamberlain (1990)), and relaxing the strict exogeneity assumption introduces further difficulties. Thus, we restrict ourselves to estimators that cannot be expected to achieve efficiency bounds. Recall, however, that the minimum chi-square estimator is efficient for the given choice of instruments.

EXAMPLE 4.1: Let $\mu(\mathbf{x}_{it},\beta) \equiv \exp(\mathbf{x}_{it}\beta)$ (or, more generally, $\exp[\psi(\mathbf{x}_{it})\beta]$ for a known 1xP function $\psi(\mathbf{x}_{it})$). Assume that (2.3) holds. Then a consistent estimator of β_o is obtained simply by choosing $\hat{\mathbf{w}}_{it} \equiv \mathbf{w}_{it} \equiv (1,\mathbf{x}_{i1},\ldots,\mathbf{x}_{i,t-1})$, t=2,..,T, and $\hat{\Omega}_N \equiv \mathbf{I}_L$. Given this consistent estimator, say $\hat{\gamma}_N$, a more efficient estimator is obtained by setting

$$\hat{\mathbf{w}}_{it} \equiv (1,\mathbf{x}_{i1},\ldots,\mathbf{x}_{i,t-1},\exp(\mathbf{x}_{i1}\hat{\gamma}_N)\mathbf{x}_{i1},\ldots,\exp(\mathbf{x}_{i,t-1}\hat{\gamma}_N)), \quad t=2,\ldots,T \qquad (4.1)$$

and

$$\hat{\Omega}_N \equiv N^{-1}\sum_{i=1}^{N}\hat{\mathbf{w}}_i'\mathbf{r}_i(\hat{\gamma}_N)\mathbf{r}_i(\hat{\gamma}_N)'\hat{\mathbf{w}}_i. \qquad (4.2)$$

(Or, $\mathbf{x}_{it}$ can be replaced by $\psi(\mathbf{x}_{it})$.) After obtaining the minimum chi-square estimator $\hat{\beta}_N$, the test statistic (3.11) can be used to test the validity of (2.1) and (2.3).

EXAMPLE 4.2: A special case of Example 4.1 is a simple distributed lag model for count data, as in HHG: $\mathbf{x}_{it} \equiv (\mathbf{z}_{it},\ldots,\mathbf{z}_{i,t-Q})$. The estimator obtained requires only $E(y_{it}|\phi_i,\mathbf{z}_{it},\ldots,\mathbf{z}_{i1}) = E(y_{it}|\phi_i,\mathbf{z}_{it},\ldots,\mathbf{z}_{i,t-Q}) = \phi_i\exp(x_{it}\beta_o)$. Neither distributional assumptions nor the strict exogeneity assumption (1.2) are needed.

EXAMPLE 4.3: A nonlinear, first order model is also a special case of Example 4.1. Let $\mu(\mathbf{x}_{it},\beta) \equiv \exp[\psi(\mathbf{x}_{it})\beta] = \exp[\alpha_o\tau(y_{i,t-1}) + \mathbf{z}_{i,t-1}\delta_o]$, where $\tau(y_{i,t-1})$ is a known function of lagged y_{it}. This is analogous to the linear vector autoregressions covered by HNR, under the assumption

$$E(y_{it}|\phi_i,y_{i,t-1},\mathbf{z}_{i,t-1},y_{i,t-2},\mathbf{z}_{i,t-2},\ldots,y_{i1},\mathbf{z}_{i1})$$
$$= E(y_{it}|\phi_i,y_{i,t-1},\mathbf{z}_{i,t-1}).$$

Note that $\mathbf{v}_{it} = (y_{i1},\mathbf{z}_{i1},\ldots,y_{i,t-2},\mathbf{z}_{i,t-2})$, t=2,...,T.

EXAMPLE 4.4: Models with contemporaneous endogeneity can also be estimated in this framework. Consider a static model where $\mathbf{x}_{it} \equiv (z_{it1}, z_{it2})$:

$$y_{it} = \phi_i \exp(z_{it1}\beta_{o1} + z_{it2}\beta_{o2}) u_{it},$$

where $E(u_{it}|z_{it2})$ is not necessarily constant. If we assume that

$$E(u_{it}|z_{it1}, z_{i,t-1,1}, z_{i,t-1,2}, \ldots) = 1$$

then this falls under (2.9) with

$$\mathbf{v}_{it} \equiv (z_{i11}, z_{i12}, \ldots, z_{i,t-2,1}, z_{i,t-2,2}, z_{i,t-1,1}).$$

After obtaining an intial consistent estimator $\hat{\gamma}_N$ of β_o, the instruments $\hat{\mathbf{w}}_{it}$ can be any function of $\mathbf{v}_{it}$ and $\hat{\gamma}_N$ and the weighting matrix $\hat{\Omega}_N$ can be chosen as in (4.2).

EXAMPLE 4.5: Consider an exponential model containing expectations of future variables:

$$y_{it} = \phi_i \exp[z_{it1}\beta_{o1} + E_t(z_{i,t+1,2})\beta_{o2}] e_{it},$$

where $E_t(\cdot)$ denotes the rational expectation given information at time t and

$$E(e_{it}|\phi_i, z_{it}, \ldots, z_{i1}) = 1.$$

Writing

$$z_{t,t+1,2} = E_t(z_{i,t+1,2}) + c_{i,t+1},$$

assume in addition that $c_{i,t+1}$ is <u>independent</u> of $\phi_i, e_{it}, z_{it}, z_{i,t-1}, \ldots, z_{i1}$. Then, defining $u_{it} \equiv \exp(-c_{i,t+1}\beta_{o2}) e_{it}$,

$$E(u_{it}|\phi_i, e_{it}, z_{it}, z_{i,t-1}, \ldots, z_{i1}) = E[\exp(-c_{i,t+1}\beta_{o2})] e_{it},$$

so

$$E(u_{it}|\phi_i, z_{it}, z_{i,t-1}, \ldots, z_{i1}) = \delta_o.$$

Setting $\delta_o = 1$ is without loss of generality, so β_{o1} and β_{o2} can be estimated by writing the model as

$$y_{it} = \phi_i \exp(z_{it1}\beta_{o1} + z_{i,t+1,2}\beta_{o2})u_{it};$$

thus, this model fits into (2.1) and (2.9), with $\mathbf{x}_{it} \equiv (z_{it1}, z_{i,t+1,2})$ and $\mathbf{v}_{it} \equiv (z_{i1}, \ldots, z_{i,t-1})$.

EXAMPLE 4.6: Finally, consider a particular binary choice model for panel data. Let y^*_{it} and a^*_{it} be latent variables and let y_{it} be the observed binary variable. For each individual, let θ_i be an unobserved, time-constant effect, and let $\mathbf{x}_{it}$ denote the observable variables. Assume that y_{it} is determined by

$$\begin{aligned} y_{it} &= 1 \quad \text{if} \quad y^*_{it} \geq 0 \ \underline{\text{and}} \ a^*_{it} \geq 0 \\ &= 0 \quad \text{otherwise,} \end{aligned}$$

where

$$y^*_{it} = \mathbf{x}_{it}\beta_o + \eta_{it} \tag{4.3}$$

$$a^*_{it} = \theta_i + \nu_{it}. \tag{4.4}$$

The errors η_{it} and ν_{it} are assumed to be independent of $(\theta_i, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{it})$ with distribution functions $F(\cdot)$ and $G(\cdot)$, respectively. In addition, η_{it} and ν_{it} are independent of each other conditional on $(\theta_i, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{it})$. Under these assumptions,

$$\begin{aligned} E(y_{it}|\theta_i, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{it}) &= P(y_{it}=1|\theta_i, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{it}) \\ &= P(y^*_{it} \geq 0,\ a^*_{it} \geq 0|\theta_i, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{it}) \\ &= P(y^*_{it} \geq 0|\theta_i, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{it})P(a^*_{it} \geq 0|\theta_i, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{it}) \\ &= P(\eta_{it} \geq -\mathbf{x}_{it}\beta_o)P(\nu_{it} \geq -\theta_i) \\ &= [1 - F(-\mathbf{x}_{it}\beta_o)][1 - G(-\theta_i)] \\ &\equiv \phi_i \mu(\mathbf{x}_{it}, \beta_o), \end{aligned} \tag{4.5}$$

where $\phi_i \equiv 1 - G(-\theta_i)$ and $\mu(\mathbf{x}_{it}, \beta) \equiv 1 - F(-\mathbf{x}_{it}\beta)$. Thus, this model fits into the framework of section two, namely equations (2.1) and (2.3). If the strict exogeneity assumption (2.2) is imposed, then the multinomial QCMLE consistently estimates β_o. Without the strict exogeneity assumption the GMM procedures offered in section 3 and in the previous examples can be used. Typically, $F(\cdot)$ would be the logistic function or the standard normal cdf, in which case $\mu(\mathbf{x}_{it},\beta) = F(\mathbf{x}_{it}\beta)$.

Note that this example differs from the usual method for introducing latent effects into binary choice models. Typically, the model is

$$y_{it}^{*} = \theta_i + \mathbf{x}_{it}\beta_o + \eta_{it}, \tag{4.6}$$

where η_{it} is independent of $(\theta_i,\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT})$. Then $y_{it} = 1$ if $y_{it}^{*} \geq 0$, so

$$P(y_{it}=1|\theta_i,\mathbf{x}_i) = F(\theta_i + \mathbf{x}_{it}\beta_o),$$

where $F(\cdot)$ is the symmetric distribution function of η_{it}. If $F(\cdot)$ is the logistic function, then conditional ML techniques can be used if y_{it} and y_{ir}, $t \neq r$, are independent conditional on $(\theta_i,\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT})$. If $F(\cdot)$ is the normal cdf then conditional MLE cannot be used to eliminate θ_i. Instead, one must specify the distribution of θ_i given $(\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT})$.

In (4.4) and (4.5), under the assumptions imposed, we do not need strict exogeneity, $F(\cdot)$ can be any known distribution function, $G(\cdot)$ need not be specified, and the distribution of θ_i conditional on $\mathbf{x}_i$ is unrestricted and unknown. Thus, (4.4)-(4.5) might be a useful alternative to (4.6), but one needs to to investigate whether (4.4)-(4.5) describes interesting economic behavior. One possible example is in modelling the choice of renting or owning a home. Equation (4.4) could describe financial considerations, while the unobserved effect in (4.5) might capture the idea that some are consistently more opposed to owning than others, regardless of the financial incentives.

5. Extension to Nonlinear Transformations of y_{it}

The model analyzed in sections 2, 3, and 4 is easily extended to allow Euler equation-type applications and other models where a nonlinear transformation of y_{it}, possibly a vector, is of interest. Let the model now be

$$\tau(y_{it},\lambda_o) = \phi_i\mu(x_{it},\beta_o)u_{it}, \quad t=1,\ldots,T, \tag{5.1}$$

where $\tau(y_{it},\lambda)$ is a known scalar transformation and λ_o is a Qx1 vector of parameters.[3] The remaining quantities are as in previous sections. The assumption on the errors is as in (2.9):

$$E(u_{it}|\phi_i,v_{it}) = E(u_{i,t-1}|\phi_i,v_{it}) = 1, \tag{5.2}$$

where v_{it} is some subset of $(\mathbf{x}_{i1},\ldots,x_{i,t-1})$. Models of the form (5.1) and (5.2) often arise as Euler equations from dynamic models of individual behavior, where an expected discounted future stream of consumption, cash flow, etc. is maximized. To the best of my knowledge, to date these models have been estimated first by linearizing them and then removing fixed effects by differencing or subtracting off time averages. Then, OLS or IV procedures are applied; see, for example, Shapiro (1984), Zeldes (1989), and Morduch (1990). It is easy to see that such a procedure is inconsistent (even if one ignores the approximation due to linearization) if the strict exogeneity assumption is violated. This is potentially important, as the rational expectations hypothesis does not imply the strict exogeneity assumption.

Estimating β_o and λ_o follows as in sections 2 and 3. Let $\theta_o \equiv (\beta_o',\lambda_o')'$ be the (P+Q)x1 vector of parameters to estimate, and define

$$r_{it}(\theta) \equiv \tau(y_{it},\lambda)/\mu_{it}(\beta) - \tau(y_{i,t-1},\lambda)/\mu_{i,t-1}(\beta), \quad t=2,\ldots,T. \tag{5.3}$$

As with Lemma 2.1, the following lemma is simple to verify under (5.1) and (5.2):

LEMMA 5.1: Under (5.1) and (5.2),

$$E[r_{it}(\theta_o)|\phi_i,v_{it}] = 0, \quad t=2,\dots,T. \quad \blacksquare \tag{5.4}$$

Lemma 5.1 provides orthogonality conditions that can be used to estimate β_o and λ_o, as well as to test any overidentification restrictions imposed in the estimation. Let $\mathbf{w}_{it}(\gamma^*)$ be a $1 \times L_t$ vector of functions of $\mathbf{v}_{it}$, so that

$$E[\mathbf{w}_{it}(\gamma^*)' r_{it}(\theta_o)] = 0, \quad t=2,\dots,T.$$

The GMM analysis in section 3 carries over with only some slight changes. First, note that

$$\nabla_\theta r_{it}(\theta) = [\nabla_\beta r_{it}(\theta), \nabla_\lambda r_{it}(\theta)] \tag{5.6}$$

is now a $1 \times (P+Q)$ vector, where

$$\nabla_\beta r_{it}(\theta) = -\{\tau(y_{it},\lambda)/[\mu_{it}(\beta)]^2\}\nabla_\beta \mu_{it}(\beta) \tag{5.7}$$

$$+ \{\tau(y_{i,t-1},\lambda)/[\mu_{i,t-1}(\beta)]^2\}\nabla_\beta \mu_{i,t-1}(\beta), \quad t=2,\dots,T$$

and

$$\nabla_\lambda r_{it}(\theta) = \nabla_\lambda \tau(y_{it},\lambda)/\mu_{it}(\beta) - \nabla_\lambda \tau(y_{i,t-1},\lambda)/\mu_{i,t-1}(\beta). \tag{5.8}$$

Here, $\nabla_\lambda \tau(y_{it},\lambda)$ is the $1 \times Q$ derivative of $\tau(y_{it},\lambda)$ with respect to λ. Let $\hat{\theta}_N$ denote a minimum chi-square estimator, based on the weighting matrix $\hat{\Omega}_N$ given by (3.6), except that $\hat{\gamma}_N$ now denotes a preliminary consistent estimator of θ_o. The asymptotic variance of $\sqrt{N}(\hat{\theta}_N - \theta_o)$ is $(R^{*\prime}\Omega^{*-1}R^*)^{-1}$, where $\Omega^* \equiv \text{plim } \hat{\Omega}_N$, R^* is the $L \times (P+Q)$ matrix

$$R^* \equiv E[\mathbf{w}_i(\gamma^*)' \nabla_\theta r_i(\theta_o)], \tag{5.9}$$

and $w_i(\gamma^*)$ and $r_i(\theta)$ are defined as in section 3. The asymptotic variance of the minimum chi-square estimator $\hat{\theta}_N$ is still estimated by $(\hat{R}_N'\hat{\Omega}_N^{-1}\hat{R}_N)^{-1}/N$, where

$$\hat{R}_N \equiv N^{-1}\sum_{i=1}^{N} w_i(\hat{\gamma}_N)'\nabla_\theta r_i(\hat{\theta}_N). \tag{5.10}$$

The usual overidentification test statistic can be used to test model specification.

EXAMPLE 5.1: Shapiro (1984), Zeldes (1989), and Morduch (1990) analyze life-cycle models of consumption using panel data, the latter two authors under liquidity constraints. For simplicity, I consider only one version of the model without constraints; the same issues arise in the analysis of more general models. Let c_{it} denote consumption of family i at time t, let θ_{it} denote taste shifters, let δ_i denote the rate of time preference for family i, and let $r^j_{i,t+1}$ denote the return from holding asset j from period t to t+1. Under the assumption that utility is given by

$$u(c_{it},\theta_{it}) = \exp(\theta_{it})c_{it}^{1-\alpha_o}/(1 - \alpha_o), \tag{5.11}$$

$$\theta_{it} = z_{it}\beta_o + \eta_i, \tag{5.12}$$

where η_i is an unobserved effect, the Euler equation is easily seen to be

$$E\left[(c_{i,t+1}/c_{it})^{-\alpha_o}(1 + r^j_{i,t+1})\middle|\psi_{it}\right] = (1 + \delta_i)^{-1}\exp(x_{it}\beta_o), \tag{5.13}$$

where ψ_{it} is family i's information set at time t and $x_{it} \equiv z_{it} - z_{i,t+1}$; note that (5.13) assumes that $z_{i,t+1} - z_{it} \in \psi_{it}$. This could be relaxed, but in these studies it reduces to assuming that family size at time t+1 is perfectly predictable at time t, not an unreasonable assumption. In fact, x_{it} might satisfy the strict exogeneity assumption, in which case α_o and β_o can be estimated by GMM by taking $y_{it} \equiv (c_{i,t+1}/c_{it}, r^j_{i,t+1})$ and v_{it} to be a subset of $(\mathbf{x}_i, c_{i,t-1}, \ldots, c_{i1}, r^j_{i,t-1}, \ldots, r^j_{i1})$, $t=2,\ldots,T$. If the $\mathbf{x}_{it}$ are not assumed to be strictly exogenous, $\mathbf{v}_{it}$ simply excludes leads of $\mathbf{x}_{it}$.

Instead, Shapiro (1984), Zeldes (1989), and Morduch (1990) linearize (5.13) and remove time averages to account for the fixed effects. Depending on the context, either OLS or IV procedures are applied to the demeaned data. For example, Shapiro (1984) and Zeldes (1989) take $\log(1 + r^j_{i,t+1})$ to the right hand side and use income as an instrument for (demeaned) $\log(1 + r^j_{i,t+1})$ in IV estimation. This procedure, even ignoring the linearization, is not consistent unless the innovations implicit in (5.13) do not help to predict future income. This strict exogeneity assumption on income is not implied by the theory; whether it is empirically important or not is another issue. But GMM estimators are available that do not require linearization or the strict exogeneity assumption.

Estimating and testing Euler equations for multiple assets is carried out by the straightforward extension of (5.1) and (5.2) suggested in endnote three.

6. Concluding Remarks

The estimators proposed here can be used to estimate dynamic or static multiplicative unobserved effects models, including Euler equations, under fairly weak assumptions. The main contributions of the paper are relaxing the strict exogeneity assumption and allowing for nonlinear transformations of y_{it} that can depend on unknown parameters. No additional distributional assumptions are imposed. Strict exogeneity is implicit in all of the count data applications that use fixed or random effects methods, and in all of the Euler equation examples on individual data, that I am aware of. There are several examples in the literature where the strict exogeneity assumption can at least be questioned. Whether failure of this assumption is empirically important remains to be seen. At a minimum, the methods in this paper allow for estimation of nonlinear autoregressive models with unobserved effects. As far as I know, no other approaches to relaxing the strict exogeneity assumption have been offered for multiplicative, nonlinear models.

Multiplicative models can be used for count variables, nonnegative continuously distributed variables, and sometimes binary variables. One weakness of the GMM estimators discussed in sections 3, 4, and 5 is that they cannot be expected to be the efficient instrumental variables estimator. Deriving the semiparametric efficiency bound in such models is an interesting open question; in fact, it is still an open question in additive, linear models without the strict exogeneity assumption.

Endnotes

1. Consistency of the usual random effects estimator relies on

$$E(\phi | \mathbf{x}_1, \ldots, \mathbf{x}_T) = E(\phi)$$

in addition to the strict exogeneity assumption (1.1). Thus, it is the least robust of the three estimators. The same ranking holds for the estimators of the count models studied by HHG (pooled Poisson, fixed effects Poisson, and random effects Poisson).

2. The models covered in this paper can be written with an additive error rather than a multiplicative error, under the assumption

$$E(u_{it} | \phi_i, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{it}) = 0.$$

Without further restrictions on the errors, the multiplicative and additive models are observationally equivalent, so I analyze only the more natural multiplicative form.

3. Extending the model to a vector of transformations is straightforward:

$$\tau_j(y_{it}, \lambda_o) = \phi_i \mu(\mathbf{x}_{it}, \beta_o) u_{itj},$$

$$E(u_{itj} | \phi_i, v_{it}) = E(u_{i,t-1,j} | \phi_i, v_{it}) = 1, \quad j=1, \ldots, M.$$

This allows multiple asset treatment in consumption CAPM-type models. The details are omitted for brevity.

References

Anderson, T.W. and C. Hsiao (1982), "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics* 18, 67-82.

Balestra, P. and M. Nerlove (1966), "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas," *Econometrica* 34, 585-612.

Chamberlain, G. (1982), "The General Equivalence of Granger and Sims Causality," *Econometrica* 50, 569-581.

Chamberlain, G. (1984), "Panel Data," in: Z. Griliches and M. Intriligator, eds., *Handbook of Econometrics*, Volume I, Amsterdam: North Holland.

Chamberlain, G. (1990), "Efficiency Bounds for Semiparametric Regression," Harvard Institute of Economic Research Discussion Paper No. 1494.

Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50, 1029-1054.

Hausman, J.A., B.H. Hall, and Z. Griliches (1984), "Econometric Models for Count Data with an Application to the Patents-R&D Relationship, *Econometrica* 52, 909-938.

Holtz-Eakin, D., W.K. Newey, and H.S. Rosen (1988), "Estimating Vector Autoregressions with Panel Data," *Econometrica* 56, 1371-1395.

Morduch, J. (1990), "Risk, Production, and Saving: Theory and Evidence from Indian Households," mimeo, Harvard University Department of Economics.

Rose, N.L. (1990), "Profitability and Product Quality: Economic Determinants of Airline Safety Performance," *Journal of Political Economy* 98, 944-961.

Shapiro, M.D. (1984), "The Permanent Income Hypothesis and the Real Interest Rate: Some Evidence from Panel Data," *Economics Letters* 14, 93-100.

Wooldridge, J.M. (1990), "Distibution-Free Estimation of Some Nonlinear Panel Data Models," MIT Department of Economics Working Paper No. 564.

Zeldes, S.P. (1989), "Consumption and Liquidity Constraints: An Empirical Investigation," *Journal of Political Economy* 97, 305-346.