Journal of Econometrics 157 (2010) 68–77

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Indirect inference for dynamic panel models

Christian Gouriéroux [a,b], Peter C.B. Phillips [c,d,e,f], Jun Yu [g,*]

[a] CREST-INSEE, 92245 Malakoff, France
[b] Department of Economics, University of Toronto, Canada
[c] Yale University, United States
[d] University of Auckland, New Zealand
[e] University of York, United Kingdom
[f] Singapore Management University, Singapore
[g] School of Economics and Sim Kee Boon Institute for Financial Economics, Singapore Management University, 90 Stamford Road, Singapore 178903, Singapore

ARTICLE INFO

Article history:
Available online 30 October 2009

JEL classification:
C33

Keywords:
Autoregression
Bias reduction
Dynamic panel
Fixed effects
Indirect inference

ABSTRACT

Maximum likelihood (ML) estimation of the autoregressive parameter of a dynamic panel data model with fixed effects is inconsistent under fixed time series sample size and large cross section sample size asymptotics. This paper proposes a general, computationally inexpensive method of bias reduction that is based on indirect inference, shows unbiasedness and analyzes efficiency. Monte Carlo studies show that our procedure achieves substantial bias reductions with only mild increases in variance, thereby substantially reducing root mean square errors. The method is compared with certain consistent estimators and is shown to have superior finite sample properties to the generalized method of moment (GMM) and the bias-corrected ML estimator.

1. Introduction

It is well known to econometricians that in dynamic panel models with fixed effects conventional estimation procedures such as (Gaussian) maximum likelihood (ML) or least-squares dummy-variable (LSDV) are asymptotically justified only when the number of time series observations (T) is large. For instance, when T is small and fixed (a single digit number, say, as occurs in many practical short time span panels), the ML estimator (MLE) is inconsistent under large N asymptotics. Nickell (1981) derived analytic formulae for the asymptotic bias under such fixed T, large N asymptotics. Using this formula and related formulae for cases with incidental trends (Phillips and Sul, 2007), it is easy to see that in many practically relevant cases the magnitude of the bias is considerable, and sometimes substantial enough to change the sign of the autoregressive coefficient estimate. At a more general level, the problem of estimation bias is of great importance in the practical use of econometric estimates, for instance, in testing theories and evaluating policies.

In the search for consistent estimators, much of the literature in the past two decades has focused on generalized method of moment (GMM) procedures and estimation methods based on instrumental variable (IV) methods, often involving lagged variables as instruments. Important contributions include Holtz-Eakin et al. (1988), Arellano and Bond (1991), Ahn and Schmidt (1995), Hahn (1997), Blundell and Bond (1998), and Alvarez and Arellano (2003). Although GMM/IV estimators are consistent when designed properly to take into account the number of lags in the given model, consistency comes at a cost. In particular, the reduction in asymptotic bias in various GMM/IV estimators is the cost at an increase, which can be substantial, in the variance. Moreover, most of the consistent GMM estimates proposed in the literature are highly model specific. For example, the methods fail when the dynamic lag order is misspecified, and it is difficult to use the standard panel GMM estimators in more complicated frameworks, for instance, when there is nonlinearity in the dynamics (Hahn and Kuersteiner, 2002). Some new developments addressing these particular issues involve generalized model choice (Lee, 2008a) and nonparametric approaches (Lee, 2008b).

In the recent literature also, several improved estimation methods have been proposed, some of them motivated by the following idea. If a bias-corrected ML estimator can be found, such an estimator may outperform the consistent GMM/IV estimator on root mean squared error (RMSE) criteria (Bun and Carree, 2005; Kiviet, 1995; Hahn and Kuersteiner, 2002). Consequently, some attempts have been made to pursue this approach and correct for bias in the ML estimator under various circumstances.

* Corresponding author.
E-mail addresses: gouriero@ensae.fr (C. Gouriéroux), peter.phillips@yale.edu (P.C.B. Phillips), yujun@smu.edu.sg (J. Yu).

0304-4076/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2009.10.024

The present paper seeks to address the problem of bias reduction in dynamic panel modeling by using the technique of indirect inference. The indirect inference methodology was first introduced by Smith (1993) and Gouriéroux et al. (1993). It has proven to be a useful method for simulation-based estimation and inference in intractable structural models. Effective applications on indirect inference include Monfort (1996) to continuous time models, Dridi and Renault (2000) to semi-parametric models, Keane and Smith (2003) to discrete choice models, Garcia et al. (2004) to stable distributions, and Monfardini (1998) to stochastic volatility models. In the paper that is closest to the present contribution, Gouriéroux et al. (2000) demonstrate that indirect inference methods can be used in various time series models for bias correction. However, we know of no earlier implementation in the context of dynamic panel models.

Indirect inference has several advantages in dynamic panels. Its primary advantage is its generality. Unlike other bias reduction methods, such as those based on explicit analytic expressions for the bias function or the leading terms in an asymptotic expansion of the bias, the indirect inference technique calibrates the bias function via simulation and hence does not require a given explicit form for the bias function or its expansion. Consequently, the method is applicable in a broad range of model specifications including nonlinear models (but note also the recent work of Lee (2008b) on alternative nonparametric estimation methods). Since panel models are two-dimensional in the sample size, the bias term is often of a complicated form and may in some cases be infeasible to obtain, although Lee (2008a) provides some general expressions for higher order dynamic specifications. Even the asymptotic bias expansions can be complicated, especially as the model itself becomes more complex and includes other incidental effects such as trends. In all these cases, the versatility of indirect inference is a significant advantage and makes the method well suited for empirical implementation.

A second advantage of indirect inference is that the approach to bias reduction can be used with many different estimation methods, including general methods like ML or LSDV, and in doing so may inherit some of the nice properties of the initial estimators. For instance, it is well known that MLE has very small dispersion relative to many consistent estimators and indirect inference applied to the MLE should preserve its good dispersion characteristic while at the same time achieving substantial bias reductions. Accordingly, indirect inference can perform very well on RMSE comparisons, as our own simulations later confirm. Unlike some other bias correction techniques, which are designed specifically for particular cases (such as when T is either small or large), the method developed here is generic and works extremely well for any values of N and T. Finally, although indirect inference is a simulation-based method, which can in some cases be computationally involved, it is computationally inexpensive in the context of dynamic panel models. This is because we propose to use the MLE as the base estimator, and since the MLE has small variance only a small number of simulated paths is sufficient to ensure an accurate calibration of the bias function that is needed for the implementation of indirect inference. This is in sharp contrast to time series models.

Our findings indicate that indirect inference provides a very substantial improvement over existing methods. For example, when $T = 5$, and $N = 100$ in a simple dynamic panel model with autoregressive coefficient $\phi = 0.9$, the RMSE of the indirect inference estimator is 85.5%, 57.2%, 82.9%, and 28% smaller than that of a GMM estimator, the bias-corrected ML estimator of Hahn and Kuersteiner (2002), the ML estimator, and the new estimator of Han and Phillips (forthcoming), respectively.

Recently, an alternative simulation-based bias correction method via the bootstrap has been proposed by Everaert and Pozzi (2007). Gouriéroux et al. (2000) compared these two simulation-based methods in the context of time series models and found no theoretical evidence for the dominance of one of them.

The paper is organized as follows. Section 2 briefly reviews various estimation methods in the context of a simple linear dynamic panel model. Section 3 introduces a generic version of the indirect inference procedure and gives some statistical properties of the resulting estimator related to unbiasedness and efficiency. In Section 4, the finite sample performance of the indirect inference estimate is compared with that of some existing approaches. Section 5 extends the method to more general specifications and Section 6 concludes.

2. Some existing estimation methods in dynamic panel models

We start the discussion with a brief review of the well-known bias result for the following simple dynamic panel model with fixed effects:

$$y_{it} = \alpha_i + \phi y_{it-1} + \epsilon_{it}, \tag{1}$$

where $\epsilon_{it} \sim iid\, N(0, \sigma^2), i = 1, \ldots, N, t = 1, \ldots, T$, the true value of ϕ is $\phi_0 \in \Phi$ with Φ being a compact set in the stable region and $|\phi_0| < 1$. The initial condition is set to be

$$y_{i0} = \frac{\alpha_i}{1-\phi} + \frac{\epsilon_{i0}}{\sqrt{1-\phi^2}},$$

where $\epsilon_{i0} \sim N(0, \sigma^2)$, independent of $\{\epsilon_{it}, i = 1, \ldots, N, t = 1, \ldots, T\}$, so that the distribution of y_{i0} follows the stationary distribution of the AR(1) process (1).

The ML (fixed effects or within-group or LSDV) estimator of ϕ is given by

$$\hat{\phi}_{NT}^{ML} = (y_-' A y_-)^{-1} y_-' A y, \tag{2}$$

where $y = (y_1, \ldots, y_N)'$ with $y_i = (y_{i1}, \ldots, y_{iT})'$, $A = I_N \otimes A_T$ with $A_T = I_T - \frac{1}{T}\iota_T'\iota_T$, $y_- = (y_{1-}, \ldots, y_{N-})'$ with $y_{i-} = (y_{i0}, \ldots, y_{iT-1})'$.

Nickell (1981) showed that the ML estimator is inconsistent when $N \to \infty$ and T is fixed. The reason for the inconsistency comes from the endogeneity of the regressor in the de-meaned regression,

$$y_{it} - y_{i\bullet} = \phi(y_{it-1} - y_{i\bullet-1}) + (\epsilon_{it} - \epsilon_{i\bullet}),$$

where $y_{i\bullet} = \sum_{t=1}^{T} y_{it}/T$, $y_{i\bullet-1} = \sum_{t=0}^{T-1} y_{it}/T$, $\epsilon_{i\bullet} = \sum_{t=1}^{T} \epsilon_{it}/T$. Since the regressor and the disturbance term are correlated in this regression and this correlation does not disappear as $N \to \infty$ when T is finite, the ML estimator (2) is asymptotically biased. Nickell (1981)'s expression for the asymptotic bias is

$$\begin{aligned} \mathrm{plim}_{N\to\infty}(\hat{\phi}_{NT}^{ML} - \phi_0) &= -\frac{(1-\phi_0^2) f_T(\phi_0)}{T-1}\left(1 - \frac{2\phi_0 f_T(\phi_0)}{T-1}\right)^{-1} \\ &= G_T(\phi_0), \end{aligned} \tag{3}$$

where $f_T(\phi) = \frac{1}{1-\phi}\left(1 - \frac{1-\phi^T}{T(1-\phi)}\right)$. The bias disappears as $T \to \infty$, but may be considerable for small values of T, and the smaller T is, the larger the bias. If $\phi_0 > 0$, the bias is always negative, and the larger ϕ_0 is, the larger the bias. But the bias does not disappear as ϕ_0 goes to zero.

Applying the first difference transformation to (1), we have

$$\Delta y_{it} = \phi \Delta y_{it-1} + \Delta\epsilon_{it}, \tag{4}$$

which gives rise to the following moment conditions:

$$E(\Delta y_{it-1} \times y_{it-s}) = 0, \quad \text{for } s = 2, 3, \ldots, t-1. \tag{5}$$

Eq. (5) suggests a GMM/IV approach to estimation for the equation in first difference form. This GMM/IV procedure was introduced and developed by Andersen and Hsiao (1981, 1982), Holtz-Eakin