INDIRECT INFERENCE FOR DYNAMIC PANEL MODELS

By

Christian Gouriéroux, Peter C. B. Phillips and Jun Yu

Indirect Inference for Dynamic Panel Models[*]

Christian Gouriéroux[†], Peter C. B. Phillips[‡], Jun Yu[§]

First version: June 2005 *This version*: December 2005

Abstract

It is well-known that 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 (T) and large cross section sample size (N) asymptotics. The estimation bias is particularly relevant in practical applications when T is small and the autoregressive parameter is close to unity. The present paper proposes a general, computationally inexpensive method of bias reduction that is based on indirect inference (Gouriéroux et al., 1993), shows unbiasedness and analyzes efficiency. The method is implemented in a simple linear dynamic panel model, but has wider applicability and can, for instance, be easily extended to more complicated frameworks such as nonlinear models. Monte Carlo studies show that the proposed 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 bias-corrected ML estimators previously proposed in the literature and is shown to have superior finite sample properties to GMM and the bias-corrected ML of Hahn and Kuersteiner (2002). Finite sample performance is compared with that of a recent estimator proposed by Han and Phillips (2005).

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 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, 2004) 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, Newey and Rosen (1988), Arellano and Bond (1991), Ahn and Schmidt (1992), Hahn (1997), Blundell and Bond (1998), and Alvares 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 achieved by 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, 2005a) and nonparametric approaches (Lee, 2005b).

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

1

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, and 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.

The present paper seeks to address the problem of bias reduction in dynamic panel modeling by using the technique of indirect inference (Gouriéroux et al., 1993). While indirect inference methods have been extensively used in various time series models for bias correction (e.g., Gouriéroux et al., 2000), 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 (2005b) 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 (2004a) 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 performs 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.

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 (2005), respectively.

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 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 $\epsilon_{it} \sim iidN(0, \sigma^2)$, $i = 1, \cdots, N$, $t = 1, \cdots, T$, 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 $\{\epsilon_{it}, \ i = 1, \cdots, N, \ t = 1, \cdots, T\}$, 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 $y = (y_1, \cdots, y_N)'$ with $y_i = (y_{i1}, \cdots, y_{iT})'$, $A = I_N \otimes A_T$ with $A_T = I_T - \frac{1}{T}\iota'_T \iota_T$, $y_- = (y_{1-}, \cdots, y_{N-})'$ with $y_{i-} = (y_{i0}, \cdots, y_{iT-1})'$.

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 $y_{i\bullet} = \sum_{i=1}^{T} y_{it}/T$, $y_{i\bullet-1} = \sum_{i=0}^{T-1} y_{it}/T$, $\epsilon_{i\bullet} = \sum_{i=1}^{T} \epsilon_{it}/T$. 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

$$\text{plim}_{N\to\infty}(\hat{\phi}_{NT}^{ML} - \phi_0) = -\frac{(1-\phi_0^2)f_T(\phi_0)}{T-1}\left(1 - \frac{2\phi_0 f_T(\phi_0)}{T-1}\right)^{-1} = G_T(\phi_0), \qquad (3)$$

where $f_T(\phi) = \frac{1}{1-\phi}(1 - \frac{1-\phi^T}{T(1-\phi)})$. The bias disappears as $T \to \infty$, but may be considerable for small values of T and the smaller is T, the larger the bias. If $\phi_0 > 0$, the bias is always negative, and the larger is ϕ_0, 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

$$\Delta y_{it} = \phi \Delta y_{it-1} + \Delta \epsilon_{it}, \qquad (4)$$

which gives rise to the following moment conditions,

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

Equation (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, Newey and Rosen (1988) and Arellano and Bond (1991), and the resulting estimator is consistent as long as $N \to \infty$ regardless of T. More sophisticated GMM/IV procedures have been proposed in recent years by, among others, Arellano and Bover (1995) and Blundell and Bond (1998).

Despite the consistency property of GMM/IV, it is known that its finite sample properties can be poor. A particular handicap of this approach is that as the autoregressive parameter moves close to unity the instruments become weak and attendant problems of weak instrumentation arise. In such cases, the GMM/IV estimator of the autoregressive parameter can suffer from substantial bias and large variation. In other circumstances, when the number of moment conditions becomes large, the GMM/IV estimator also suffers from large finite sample bias (Bun and Kiviet, 2002, Ziliak, 1997). Finally, the GMM/IV estimator is designed for linear dynamic systems and is not readily applicable to nonlinear models.

To overcome the weak instrumentation problem that arises in near unit root panels, Han and Phillips (2005) replaced the weak moment conditions (5) by a set of new moment conditions, i.e.,

$$E(\Delta y_{it-1} \times [(2\Delta y_{it} + \Delta y_{it-1}) - \phi \Delta y_{it-1}]) = 0. \qquad (6)$$

4

This approach leads to a new estimator of the form

$$\hat{\phi}_{NT}^{HP} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\Delta y_{it-1}(2\Delta y_{it} + \Delta y_{it-1})}{\sum_{i=1}^{N}\sum_{t=1}^{T}(\Delta y_{it-1})^2}. \tag{7}$$

Han and Phillips (2005) established the following large N and large T asymptotics for this estimator

$$\sqrt{NT}(\hat{\phi}_{NT}^{HP} - \phi) \Rightarrow N(0, 2(1 + \phi)). \tag{8}$$

and fournd that these asymptotics work very well when ϕ is close to 1, even for T as small as 3.

Hahn and Kuersteiner (2002) also resorted to large N and T asymptotics. In particular, they showed that when both N and T approach infinity and $0 < \lim \frac{N}{T} = c < \infty$,

$$\sqrt{NT}(\hat{\phi}_{NT}^{ML} - (\phi - \frac{1}{T}(1 + \phi))) \Rightarrow N(0, (1 - \phi^2)). \tag{9}$$

As T passes to infinity, the ML estimator becomes consistent. However, the asymptotic distribution is not centered at the origin and there is an asymptotic bias in the limiting distribution. Accordingly, Hahn and Kuersteiner (2002) introduced a bias-corrected ML estimator centered at the origin, which is a feasible version of the bias-corrected ML estimator of Kiviet (1995) because it does not require that the true value ϕ_0 be known. If the the bias-corrected MLE is denoted by $\hat{\phi}_{NT}^{HK}$, Hahn and Kuersteiner (2002) showed that

$$\sqrt{NT}(\hat{\phi}_{NT}^{HK} - \phi) \Rightarrow N(0, (1 - \phi^2)). \tag{10}$$

Since $1 - \phi^2 < 2(1 + \phi)$, $\hat{\phi}_{NT}^{HK}$ always has a smaller asymptotic variance than $\hat{\phi}_{NT}^{HP}$. Bun and Carree (2005) proposed alternative bias-corrected ML estimators under the assumption that T may be small.

Using simulations, Kiviet (1995) showed that in many practically relevant cases, the bias-corrected ML estimator has smaller RMSE than various GMM estimators. Hahn and Kuersteiner (2002) also examined the finite sample properties of the bias-correct ML estimator and made comparisons with GMM. From these studies, the superiority of the bias-corrected ML estimator over GMM is now documented in many empirically relevant circumstances. The improvement is particularly substantial when ϕ is close to unity.

3 Estimating Dynamic Panel Models via Indirect Inference

3.1 Estimating AR(1) models via indirect inference

The indirect inference procedure, first introduced by Gouriéroux et al (1993) and independently proposed by Smith (1993) and Gallant and Tauchen (1996), can be understood as a generalization of the simulated method of moments approach of Duffie and Singleton (1993). It has been found to be a highly useful procedure when the moments and the likelihood function of the true model are difficult to deal with, but the true model is amenable to data simulation. Gouriéroux et al (1993) provided conditions under which the indirect inference estimator has desirable large sample properties, such as consistency and asymptotical normality.

A carefully designed indirect inference estimator can have good small sample properties, too, as shown by Gouriéroux, et al (2000) in the time series context. Because our procedure is closely related to one given in Gouriéroux, et al (2000), we first review that method in the context of a simple AR(1) model.

Suppose we need to estimate the parameter ϕ in the AR(1) model

$$y_t = \phi y_{t-1} + \epsilon_t,$$

from observations $y = \{y_0, y_1, \cdots, y_T\}$, where the true value of ϕ is ϕ_0 which lies in a compact set Φ of the stable region and $|\phi_0| < 1$. It is well known that standard procedures such as ML and least squares (LS) produce downward biased coefficient estimators of ϕ in finite samples. Using analytic techniques in a simple case, Hurwicz (1950) demonstrated this AR bias effect, showed that the bias does not go to zero as the AR coefficient goes to zero and that the bias increases as the AR coefficient moves towards unity. It is now well-known that this bias is accentuated in models with fitted intercept and trends (Orcutt, 1969).

Various techniques have been proposed to correct the bias in the ML estimator of ϕ in the AR(1). Examples include Kendall (1954), Quenouille (1956), Efron (1979), and Andrews (1993). Some of these methods, such as Kendall's procedure, requires explicit knowledge of the first term of the asymptotic expansion of the bias in powers of $\frac{1}{T}$.

The indirect inference method proposed by Gouriéroux et al (2000) makes use of simulations to calibrate the bias function and requires neither the explicit form of the bias, nor the bias expansion. This advantage seems important when the computation of the bias expression is analytically involved, and it becomes vital when the bias and the first term of the bias asymptotic expansions are too difficult to compute explicitly.

The idea of indirect inference here is as follows. Given a parameter choice ϕ, let $\tilde{y}^h(\phi) = \{\tilde{y}_0^h, \tilde{y}_1^h, \cdots, \tilde{y}_T^h\}$ be data simulated from the true model, where $h = 1, \cdots, H$

with H being the number of simulated paths. It should be emphasized that it is important to choose the number of observations in $\tilde{y}^h(\phi)$ to be the same as the number of observations in the observed sequence y for the purpose of the bias calibration.

The central idea is then to match various functions of the simulated data with those of observed data in order to estimate parameters. Suppose Q_T is the objective function of a certain estimation method applied to an auxiliary model which is indexed by the parameter θ. Define the corresponding estimator based on the observed data by

$$\hat{\theta}_T = \mathrm{argmax}_{\theta \in \Theta} Q_T(y), \tag{11}$$

and the corresponding estimator based on the hth simulated path by

$$\tilde{\theta}_T^h(\phi) = \mathrm{argmax}_{\theta \in \Theta} Q_T(\tilde{y}^h(\phi)), \tag{12}$$

where Θ is a compact set.

The indirect inference estimator is defined by

$$\hat{\phi}_{T,H}^{II} = \mathrm{argmin}_{\phi \in \Phi} \parallel \hat{\theta}_T - \frac{1}{H} \sum_{h=1}^{H} \tilde{\theta}_T^h(\phi) \parallel, \tag{13}$$

where $\parallel \cdot \parallel$ is some finite dimensional distance metric. In the case where H tends to infinity, the indirect inference estimator becomes

$$\hat{\phi}_T^{II} = \mathrm{argmin}_{\phi \in \Phi} \parallel \hat{\theta}_T - E(\tilde{\theta}_T^h(\phi)) \parallel. \tag{14}$$

It is useful to define the so-called binding function as

$$b_T(\phi) = E(\tilde{\theta}_T^h(\phi)).$$

In the case where the number of parameters in the auxiliary model is the same as that in the true model (this is always the case when the auxiliary model is chosen to be the true model), and b_T is invertible, the indirect inference estimator is given by

$$\hat{\phi}_T^{II} = b_T^{-1}(\hat{\theta}_T).$$

The procedure essentially builds in a small-sample bias correction to parameter estimation, with the bias being computed directly by simulation. To see this, suppose the true value of $\phi = 0.9$, and the given estimator (like OLS in the present case) has downward bias. For example, suppose $\hat{\phi}_T = 0.85$ is the realized value of the estimate. We do not use the value 0.85 to estimate ϕ, but instead use the value of ϕ that yields

the averaged estimated ϕ of 0.85 from simulated data. Since the bias occurs in $\hat{\phi}_T$, it should also occur in the binding function $b_T(\phi)$. Hence, with the bias correction that is built into the inversion $\hat{\phi}_T^{II} = b_T^{-1}(\hat{\phi}_T)$, the estimator is exactly "b_T-mean-unbiased" for ϕ. That is, $E(b_T(\hat{\phi}_T^{II})) = b_T(\phi_0)$. Gouriéroux et al (2000) established conditions under which the indirect inference estimator is "b_T-mean-unbiased", and related the indirect inference estimator to the median unbiased estimator of Andrews (1993) and the bootstrap estimator of Efron (1979).

In practice, three choices have to be made: H, the estimation criterion Q_T, and the distribution of the data used in the simulation. Of course, H cannot be infinite and the choice of H has to be made to ensure that $E(\tilde{\theta}_T^h(\phi))$ is well approximated by $\frac{1}{H}\sum_{h=1}^{H}\tilde{\theta}_T^h(\phi)$, which will be guaranteed by the use of large H. When the true model is easy to estimate (although the resulting estimator may be severely biased) – for example, when the likelihood function has a closed-form expression – the estimation criterion can be maximum likelihood applied to the true model itself. The simulation results reported in Gouriéroux et al (2000) suggest that the indirect inference method, when $H = 15,000$ and estimation criterion is maximum likelihood, works as well as the median unbiased estimator of Andrews (1993). Both these methods are, of course, dependent on the validity of the assumed data distribution for the validity of the finite sample binding formula.

It is necessary to apply the Common Random Numbers (CRNs) technique during the numerical optimization to enforce a smooth surface for the objective function. That is, the H simulated paths are always obtained from a fixed set of canonical random numbers, which are typically uniform variates or standardized normals.

3.2   Estimating panel models via indirect inference

In the context of dynamic panel models, the bias correction methods proposed by Kiviet (1995), Bun and Carree (2005), and Hahn and Kuersteiner (2002) all work under linear specifications and for a given first order lag structure. When the panel model becomes more complicated, analytic derivations of the bias function become much more involved, if not possible. Some recent generalizations for higher order dynamic structures and nonlinear models have been developed by Lee (2005a, 2005b). Use of approaches to bias elimination that require knowledge of the bias function imposes further challenges as the model complexity increases. Moreover, when the model is nonlinear, GMM is not readily available as the classical moment conditions are no longer valid.

As a general principle to correct for bias, indirect inference has the advantage that it can be applied to many models and estimators. The present paper proposes indirect inference in conjunction with the MLE as the baseline estimator and chooses

the auxiliary model to be the true model.

When applying ML to estimate the linear panel model (1) with the observed data, we obtain $\hat{\phi}_{NT}^{ML}$ defined by (2). Let the ML estimator of (1) with the hth simulated path, given ϕ, be denoted by $\tilde{\phi}_{NT}^{ML,h}(\phi)$, that is,

$$\tilde{\phi}_{NT}^{h,ML}(\phi) = ((\tilde{y}_-^h)'A\tilde{y}_-^h)^{-1}(\tilde{y}_-^h)'A\tilde{y}^h, \tag{15}$$

where $\tilde{y}^h = (\tilde{y}_1^h, \cdots, \tilde{y}_N^h)'$ with $\tilde{y}_i^h = (\tilde{y}_{i1}^h, \cdots, \tilde{y}_{iT}^h)'$, $\tilde{y}_-^h = (\tilde{y}_{1-}^h, \cdots, \tilde{y}_{N-}^h)'$ with $\tilde{y}_{i-}^h = (\tilde{y}_{i0}^h, \cdots, \tilde{y}_{iT-1}^h)'$. Note that $\tilde{y}^h$ depends on (ϕ). For the sake of presentation we simply write $\tilde{y}^h = \tilde{y}^h(\phi)$.

The indirect inference estimator is defined by

$$\hat{\phi}_{NT}^{II} = \mathrm{argmin}_{\phi \in \Phi} \parallel \hat{\phi}_{NT}^{ML} - b_{NT}(\phi) \parallel, \tag{16}$$

where $\parallel \cdot \parallel$ is a distant metric and $b_{NT}(\phi)$ is the binding function defined by

$$b_{NT}(\phi) = E(\tilde{\phi}_{NT}^{h,ML}(\phi)).$$

In practice, of course, we replace $b_{NT}(\phi)$ in *(16)* by $\frac{1}{H} \sum_{h=1}^{H} \tilde{\phi}_{NT}^{h,ML}(\phi)$. Since ML generally has small variance in dynamic panel models because N is large, even small values of H appear to be sufficient to ensure good finite sample performance of the estimator, as shown in the simulation study below.

To discuss the "unbiasedness" property, we impose the following condition.

Assumption 1: *The binding function $b_{NT}(\cdot)$, mapping from Φ to $b_{NT}(\Phi)$, is uniformly continuous and one-to-one.*

By construction when $H = \infty$, we have

$$E(b_{NT}(\hat{\phi}_{NT}^{II})) = E(\hat{\phi}_{NT}^{ML}) = E(\tilde{\phi}_{NT}^{h,ML}(\phi_0)) = b_{NT}(\phi_0).$$

By Assumption 1, b_{NT} is invertible and hence $b_{NT}^{-1}(E(b_{NT}(\hat{\phi}_{NT}^{II}))) = \phi_0$, from which we deduce that $\hat{\phi}_{NT}^{II}$ is "b_{NT}-mean-unbiased". Formally stated, we have the following result.

THEOREM 1 *If Assumption 1 holds, the indirect inference estimator defined in (16) is "b_{NT}-mean-unbiased", that is,*

$$b_{NT}^{-1}(E(b_{NT}(\hat{\phi}_{NT}^{II})) = \phi_0.$$

Remark 1: The property of "b_{NT}-unbiasedness" derived above does not impose any restriction on N or T. This is in contrast with the existing bias-corrected estimators which require either large N, or large T, or both. But the procedure does make use of explicit distributional assumptions on data generation - here the normality of the inputs ϵ_{it} in (1). However, as $N \rightarrow \infty$, the binding function $b_{NT}(\phi)$ will depend on T and certain moments of the data that will be consistently estimated in the simulations as $N \rightarrow \infty$, so that some robustness to the distribution of the input variables can be expected in this case. Similarly, while limiting normality is obtained for the Han and Phillips (2005) estimator when $N \rightarrow \infty$ for fixed T, the expression for the variance in that limit distribution depends on T and on certain moments of the data, which again may be estimated consistently using the cross section observations. So, the methods may be regarded as having similar forms of distributional dependence, at least for large N.

Remark 2: The "b_{NT}-unbiasedness" in general does not imply "mean-unbiasedness" or vice versa. In the case where $b_{NT}(\phi)$ is a linear function in ϕ, however, these two concepts are equivalent. When $b_{NT}(\phi)$ is close to a linear function, which is perhaps a practically relevant case, we may expect $\hat{\phi}_{NT}^{II}$ to be close to "mean-unbiasedness".

We now discuss the issue of efficiency and impose a smoothness assumption on the function b_{NT}, which seems mild under the given distributional assumption and the restriction to the stable region Φ.

Assumption 2: *The binding function $b_{NT}(\cdot)$ and its inverse $b_{NT}^{-1}(\cdot)$ are continuously differentiable on Φ.*

There is no formal extension of the asymptotically minimax limit theory in the present framework of the panel model (1). Even for double index asymptotics as both $N \rightarrow \infty$ and $T \rightarrow \infty$, where consistency and asymptotic normality for the autoregressive coefficient ML estimator applies, the model still involves an infinite number of nuisance parameters and we do not know of an extension of the Hájek-LeCam asymptotically locally minimax theory (e.g. van de Vaart, 2000) to such cases.

Notwithstanding the absence of this asymptotically minimax limit theory, we may mechanically apply the standard Cramér-Rao bound theory in this framework as follows. Under Assumption 2, the variance of an unbiased estimator of $b(\phi)$ is no less than

$$Bound(b) = \left(\frac{\partial b(\phi_0)}{\partial \phi} \right)^2 I^{\phi\phi},$$

where $I^{\phi\phi}$ is the element of the inverse of the information matrix corresponding to parameter ϕ. Note that the information matrix in this case concerns all the parameters

in the model, viz., $\phi, \alpha_1, \cdots, \alpha_N, \sigma^2$.

Any biased estimator of the parameter ϕ can be considered as an unbiased estimator of its mean. Let us assume that this estimator $\hat{\phi}$ has mean $b_{NT}(\phi)$ which is dependent only on ϕ and the sample sizes.[1] Then, according to the above we have

$$Var(\hat{\phi}) \geq \left(\frac{\partial b_{NT}(\phi_0)}{\partial \phi}\right)^2 I^{\phi\phi},$$

and its "lack of efficiency" can be measured by

$$\left(\frac{\partial b_{NT}(\phi_0)}{\partial \phi}\right)^{-2} \frac{Var(\hat{\phi})}{I^{\phi\phi}}.$$

Similarly the lack of efficiency of any unbiased estimator $\hat{\phi}$ of ϕ may be measured by $\frac{Var(\hat{\phi})}{I^{\phi\phi}}$.

Consider the indirect inference estimator $\hat{\phi}_{NT}^{II}$ associated with $\hat{\phi}_{NT}^{ML}$. We have,

$$\hat{\phi}_{NT}^{II} = b_{NT}^{-1}(\hat{\phi}_{NT}^{ML}).$$

Let us now assume that N is large, in which case the estimator $\hat{\phi}_{NT}^{ML}$ converges to a limit $b_T(\phi_0)$ as $N \to \infty$. By virtue of the delta method appled to

$$\hat{\phi}_{NT}^{II} = b_{NT}^{-1}(\hat{\phi}_{NT}^{ML}) = b_{NT}^{-1}(b_{NT}(\phi_0) + \hat{\phi}_{NT}^{ML} - b_{NT}(\phi_0)),$$

we have

$$Var(\hat{\phi}_{NT}^{II}) \approx \left(\frac{\partial b_T(\phi_0)}{\partial \phi}\right)^{-2} Var(\hat{\phi}_{NT}^{ML}), \tag{17}$$

and hence

$$\frac{Var(\hat{\phi}_{NT}^{II})}{I^{\phi\phi}} \approx \left(\frac{\partial b_T(\phi_0)}{\partial \phi}\right)^{-2} \frac{Var(\hat{\phi}_{NT}^{ML})}{I^{\phi\phi}}. \tag{18}$$

The asymptotic approximation (18) suggests that the indirect inference estimator should inherit some the "efficiency" properties of the initial estimator treated as an estimator of its mean.

Remark 3: The change in the mean squared error (MSE) of $\hat{\phi}_{NT}^{II}$ over that of $\hat{\phi}_{NT}^{ML}$ is due to the reduction (often substantial) that takes place in the bias of the estimator and to the fact that the change in variance is often minor. In fact, the change in the

---

[1]This is justified by formula (15) and holds for the present simple panel dynamic model.

variance depends largely on $\frac{\partial b_T(\phi_0)}{\partial \phi}$, as seen above. For $\left| \frac{\partial b_T(\phi_0)}{\partial \phi} \right| > 1$, $\hat{\phi}_{NT}^{\hat{II}}$ has a smaller variance than the initial estimator, and for $\left| \frac{\partial b_T(\phi_0)}{\partial \phi} \right| < 1$, $\hat{\phi}_{NT}^{\hat{II}}$ has a larger variance than the initial estimator. For the present model, the following expression for $b_T(\phi)$ follows from the Nickell bias formula (6) and the asymptotic expansion of this bias for large T given in Phillips and Sul (2004):

$$b_T(\phi) = \phi + G_T(\phi) = \begin{cases} \phi - \frac{1+\phi}{T-1} + O\left(T^{-2}\right) & \text{for } |\phi| < 1 \\ \phi - \frac{3}{T+1} & \text{for } |\phi| = 1 \end{cases} . \tag{19}$$

Note that although $b_T(\phi)$ is continuous in ϕ as ϕ passes through unity, its asymptotic expansion as $T \to \infty$ is not, and the bias expression given in (19) for the case $\phi = 1$ is exact. The derivative $\frac{\partial b_T(\phi_0)}{\partial \phi} = 1 + O\left(T^{-1}\right)$ is well behaved and for large T has a magnitude that is less than unity. Hence, according to this asymptotic expression, the variance of $\hat{\phi}_{NT}^{\hat{II}}$ should be greater than that of the initial ML estimator, an outcome confirmed in the simulations below.

4   Monte Carlo Results

This section reports the results of some simulation experiments examining the relative performance of the proposed procedure against certain alternative methods. Following Hahn and Kuersteiner (2002), the data are generated from the following linear dynamic panel model,

$$y_{it} = \alpha_i + \phi_0 y_{it-1} + \epsilon_{it},$$

where $\epsilon_{it} \sim iidN(0,1)$, $\alpha_i \sim$ iid $N(0,1)$, and $\phi_0 = 0, 0.3, 0.6, 0.9$. α_i and ϵ_{it} are assumed to be independently distributed. The initial condition is

$$y_{i0}|\alpha_i \sim N(\frac{\alpha_i}{1 - \phi_0}, \frac{1}{\sqrt{1 - \phi_0^2}}).$$

We choose $N = 100, 200$ and $T = 5, 10, 20$. For each combination of N and T, we employ five methods to estimate ϕ: ML, GMM, the method proposed by Han and Phillips (2005), the bias-corrected ML method of Hahn and Kuersteiner (2002), and the indirect inference method developed here. The design of the experiment is identical to that in Hahn and Kuersteiner (2002) to aid comparisons. Although a linear model is considered in these experiments so that GMM, the Han-Phillips method and the Hahn-Kuersteiner method can be compared, it is worth pointing out that the indirect inference approach can be applied to more complicated models. For GMM and the bias-corrected ML, we simply report the results of Hahn and Kuersteiner (2002). For the

indirect inference method, we first choose $H = 10$ and later investigate the performance of our estimator for larger values of H.

Table 1 reports the biases and RMSEs of all five estimates obtained from 5,000 replications. The following general results emerge. First, ML has serious bias problems in all cases. In general, the ML bias becomes larger as ϕ moves closer to unity or N gets larger, but becomes smaller as T gets larger, all of which corroborates the asymptotic theory.

Second, although GMM alleviates the bias problems in all cases, the biases remain substantial when ϕ is close to unity. Compared with ML, GMM generally has smaller RMSEs. However, some exceptions to this occur when T is small and ϕ is close to unity. The large values of the variance and bias in cases where ϕ is close to unity are evidence of the weak instrumentation of GMM in these cases. It is interesting that these effects are strongly manifested even at $\phi = 0.9$ which is some distance from unity.

Third, the bias-corrected ML substantially alleviates the bias problems in all cases, as it is designed to do, at least when T is modestly large. Like ML and GMM, the bias in the bias-corrected ML becomes larger when ϕ gets larger, but becomes smaller when T is larger. Interestingly, the bias is still substantial in this bias corrected version for $\phi = 0.9$. However, the bias-corrected ML has smaller RMSE than ML in all cases and has smaller RMSE than GMM in almost all cases.

Fourth, the Han-Phillips estimator provides very good bias correction in all cases, including those cases where ϕ is close to unity. This is not surprising as the problem of weak instrumentation is avoided in this approach. Like the three methods discussed above, the RMSE becomes larger as ϕ gets larger. Unlike these other methods, however, the bias does not seem to depend on ϕ. Moreover, the method dominates ML in terms of RMSEs in all cases due to its ability to remove the bias. It also dominates GMM and bias-corrected ML in terms of RMSE when ϕ is close to unity except when T is large. This result is interesting and somewhat surprising as the bias-corrected ML estimator is asymptotically more efficient than the Han-Phillips estimator.

Finally, the most important comparisons are between the indirect inference estimates with the other four estimates. With $H = 10$, the indirect inference procedure removes the bias more successfully than GMM, ML and the bias-corrected ML, but less successfully than the Han-Phillips method. As shown later, however, with increased values for H, the indirect inference method is much more effective in removing bias and has performance that is comparable with the Han-Phillips method in terms of bias correction. Like Han-Phillips, the bias does not seem to depend on ϕ. In terms of RMSE, indirect inference estimates clearly dominate ML, GMM and the Han-Phillips estimates in all cases and dominate the bias-corrected ML estimates in almost all cases. The larger is ϕ, the more substantial is the improvement of the indirect inference method over the existing methods. For example, when $T = 5$, $N = 100$,

13

$\phi = 0.9$, the RMSE of the indirect inference estimates is 85.5%, 57.2%, 82.9%, and 28% smaller than that of GMM, the bias-corrected ML, ML, and Han and Phillips's estimates, respectively. When $T = 10$, $N = 200$, $\phi = 0.9$, the RMSE of the indirect inference estimates is 84.7%, 66.2%, 88.7%, and 41.8% smaller than that of the other four estimates, respectively.

To investigate the sensitivity of the performance of the indirect inference method to the choice of H, Table 2 reports the biases and the RMSEs when $H = 10, 50, 250$. With large values of H, we expect indirect inference to have better finite sample properties. This is confirmed in Table 2. When $H = 250$, the biases almost completely disappear. However, the improvement in terms of RMSE is marginal, especially from $H = 50$ to $H = 250$. This finding suggests that the initial estimator (ML) indeed has a small variance and hence a small value of H delivers satisfactory approximation of the binding function by $H^{-1} \sum_{h=1}^{H} \tilde{\phi}_{NT}^{h,ML}(\phi)$. Consequently, despite being a simulation-based estimation procedure, the indirect inference method is not computationally expensive in the context of panel models.

5   Conclusions

Bias in the estimation of the parameters of dynamic panel models by standard methods such as ML is generally not negligible in short (time span) panels and conventional GMM approaches encounter difficulties of bias and variance when the autoregressive coefficient is close to unity, as it commonly is in practical work. The procedure we propose here for reducing the bias involves the use of indirect inference to calibrate the bias function and operates with only small increases in variance. Simulations show the procedure to be highly effective in the simple linear dynamic panel model. Although the present paper implements the approach only in the context of a simple first order linear dynamic panel model estimated by ML as the base estimation method, the technique itself is quite general and can be applied in many other panel models and other base estimation methods with little modification.

Being a simulation-based estimation method, the indirect inference procedure is computationally more involved than other methods. However, since the base estimator employed here has a small variance, only a small number of simulated paths are needed for the indirect inference estimator to have good finite sample properties. Therefore, the computational cost of the indirect inference procedure seems low and its finite sample gains substantial. The procedure has potential for even greater gains in cases where the bias is known to be larger, as for instance in the case of dynamic panel with incidental trends (Phillips and Sul, 2004).

14

References

[1] Ahn, S. and P. Schmidt, 2003, "Efficient estimation of models for dynamic panel data", *Journal of Econometrics,* 68, 5-28.

[2] Alvarez, J. and M. Arellano, 2003, "The time series and cross-section asymptotics of dynamic panel data estimators", *Econometrica*, 71, 1121-1159.

[3] Andersen, T.W. and C. Hsiao, 1981, "Estimation of dynamic models with error components", *Journal of the American Statistical Association*, 76, 598-606.

[4] Andersen, T.W. and C. Hsiao, 1982, "Formulation and estimation of dynamic models using panel data", *Journal of Econometrics*, 18, 47-82.

[5] Andrews, D. W. K., 1993, "Exactly median-unbiased estimation of first-order autoregressive/unit root models," *Econometrica*, 61, 139–166.

[6] Arellano, M. and O. Bover, 2003, "Another look at the instrumental-variable estimation of error-components models", *Journal of Econometrics*, 68, 29-51.

[7] Arellano, M. and S. Bond, 1991, "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", *Review of Economic Studies*, 58, 277-297.

[8] Bun, M.J. and M.A. Carree, 2005, "Bias-corrected estimation in dynamic panel data models", *Journal of Business and Economic Statistics*, 23, 200-210.

[9] Duffie, D., and K. J. Singleton, 1993, "Simulated moments estimation of Markov models of asset prices," *Econometrica*, 61, 929–952.

[10] Efron, B., 1993, "Bootstrap methods: Another look at the jackknife," *Annals of Statistics*, 7, 1–26.

[11] Gallant, A. R., and G. Tauchen, 1996, "Which moments to match?" *Econometric Theory*, 12, 657–681.

[12] Gouriéroux, C., A. Monfort, and E. Renault, 1993, "Indirect inference," *Journal of Applied Econometrics*, 8, S85–S118.

15

[13] Gouriéroux, C., E. Renault, and N. Touzi, 2000, "Calibration by simulation for small sample bias correction," in R.S. Mariano, T. Schuermann and M. Weeks (eds.), *Simulation-Based Inference in Econometrics: Methods and Applications*, Cambridge University Press, p328-358.

[14] Hahn, J., 1997, "Efficient estimation of panel data models with sequential moment restrictions", *Journal of Econometrics*, 79, 1–21.

[15] Hahn, J. and G. Kuersteiner, 2002, "Asymptotically unbiased inference for a dynamic model with fixed effects when both n and T are large", *Econometrica*, 70, 1639–1657.

[16] Han, C. and P.C.B. Phillips, 2005, "GMM Estimation for Dynamic Panels with Fixed Effects and Strong Instruments at Unity", Yale University, mimeographed.

[17] Holtz-Eakin, D., W. Newey, and H. S. Rosen, 1988, "Estimating vector autoregressions with panel data", *Econometrica*, 56, 1371-1395.

[18] Hurvicz, L., 1950, "Least square bias in time series", in T. Koopmans (ed.), *Statistical Inference in Dynamic Economic Models*, New York, Wiley, 365-383.

[19] Kendall, M. G., 1954, "Notes on bias in the estimation of autocorrelation", *Biometrika*, 41, 403-404.

[20] Kiviet, I., 1995, "On bias, inconsistency and efficiency of various estimators in dynamic panel data models", *Journal of Econometrics,* 68, 53-78.

[21] Lee, Y. (2005a) "A General Approach to Bias Correction in Dynamic Panels under Time Series Misspecification". Yale University, mimeographed.

[22] Lee, Y. (2005b) "Nonparametric Estimation for Dynamic Panel Models". Yale University, mimeographed.

[23] Nickell. S., 1981, "Biases in dynamic models with fixed effects", *Econometrica*, 49, 1417–1426.

[24] Phillips, P. C. B. and D. Sul (2004). "Bias in Dynamic Panel Estimation with Fixed Effects, Incidental Trends and Cross Section Dependence" Cowles Foundation Discussion Paper, No. 1438.

[25] Quenouille, M. H., 1956, "Notes on bias in estimation," *Biometrika*, 43, 353-360.

[26] Smith, A.A., 1993, "Estimating nonlinear time-series models using simulated vector autoregressions," *Journal of Applied Econometrics*, 8, S63–S84.

[27] van de Vaart (2000). *Asymptotic Statistics.* Cambridge University Press: New York.

[28] Ziliak, J.P., 1997, "Efficient estimation with panel data when instruments are predetermined: An empirical comparison of moment-condition estimators", *Journal of Business and Economic Statistics*, 15, 419-431.

Table 1. Monte Carlo comparison of the bias and RMSE of the GMM estimator of Arellano and Bover, the corrected ML estimator of Hahn and Kuersteiner (HK), ML, the new GMM estimator of Han and Phillips (HP), and the indirect inference estimator of ϕ for the dynamic panel model. The number of simulated paths is set to be 10 for indirect inference.

Case			Bias in $\hat{\phi}$					RMSE of $\hat{\phi}$				
T	N	ϕ	GMM	HK	ML	HP	II	GMM	HK	ML	HP	II
5	100	0	-.011	-.039	-.1993	.0038	-.0297	.074	.065	.2041	.0877	.0635
5	100	0.3	-.027	-.069	-.2741	.0041	-.0384	.099	.089	.2779	.0948	.0868
5	100	0.6	-.074	-.115	-.3619	.0044	-.0291	.160	.129	.3650	.1021	.0761
5	100	0.9	-.452	-.178	-.4642	.0039	-.0282	.552	.187	.4667	.1111	.0799
5	200	0	-.006	-.041	-.2002	-.0003	.0117	.053	.055	.2026	.0624	.0433
5	200	0.3	-.014	-.071	-.2751	-.0004	.0068	.070	.081	.2771	.0676	.0770
5	200	0.6	-.038	-.116	-.3631	-.0008	.0233	.111	.124	.3647	.0729	.0564
5	200	0.9	-.337	-.178	-.4654	-.0021	.0273	.443	.183	.4668	.0792	.0616
10	100	0	-.011	-.010	-.0996	.0023	-.0198	.044	.036	.1044	.0512	.0407
10	100	0.3	-.021	-.019	-.1350	.0022	-.0147	.053	.040	.1387	.0572	.0404
10	100	0.6	-.045	-.038	-.1791	.0021	-.0046	.075	.051	.1818	.0626	.0392
10	100	0.9	-.218	-.079	-.2448	.0016	.0052	.248	.085	.2465	.0682	.0408
10	200	0	-.006	-.011	-.1001	.0004	.0038	.031	.027	.1025	.0366	.0253
10	200	0.3	-.011	-.019	-.1352	.0001	.0054	.038	.032	.1371	.0406	.0263
10	200	0.6	-.025	-.037	-.1789	-.0003	.0034	.051	.045	.1802	.0441	.0262
10	200	0.9	-.152	-.079	-.2439	-.0006	.0041	.181	.082	.2447	.0476	.0277
20	100	0	-.011	-.003	-.0497	.0012	-.0050	.029	.024	.0545	.0338	.0239
20	100	0.3	-.017	-.005	-.0663	.0012	-.0025	.033	.024	.0699	.0383	.0240
20	100	0.6	-.029	-.011	-.0859	.0011	.0018	.042	.024	.0883	.0423	.0224
20	100	0.9	-.100	-.032	-.1203	.0010	-.0044	.109	.037	.1215	.0460	.0209
20	200	0	-.006	-.003	-.1001	.0002	.0038	.020	.017	.0525	.0240	.0175
20	200	0.3	-.009	-.005	-.1352	.0000	.0054	.022	.017	.0683	.0270	.0174
20	200	0.6	-.016	-.010	-.1789	-.0002	.0034	.027	.018	.0869	.0296	.0163
20	200	0.9	-.065	-.031	-.2439	-.0002	.0041	.074	.034	.1204	.0317	.0152

Table 2. Monte Carlo comparison of the bias and RMSE of the indirect inference estimator of ϕ for the dynamic panel model with different numbers (H) of simulated paths. The number of simulated paths is set to be $H = 10, 50,$ and 250.

Case			Bias in $\hat{\phi}$			RMSE of $\hat{\phi}$		
T	N	ϕ	H=10	H=50	H=250	H=10	H=50	H=250
5	100	0	-0.0297	-0.0082	0.0007	0.0635	0.0571	0.0570
5	100	0.3	-0.0384	-0.0181	-0.0074	0.0868	0.0817	0.0814
5	100	0.6	-0.0291	-0.0112	0.0005	0.0761	0.0706	0.0696
5	100	0.9	-0.0282	-0.0088	0.0000	0.0799	0.0777	0.0760
5	200	0	0.0117	-0.0043	-0.0002	0.0433	0.0413	0.0408
5	200	0.3	0.0068	-0.0126	-0.0102	0.0770	0.0742	0.0741
5	200	0.6	0.0233	-0.0023	-0.0003	0.0564	0.0503	0.0503
5	200	0.9	0.0273	-0.0058	-0.0030	0.0616	0.0540	0.0542
10	100	0	-0.0198	-0.0068	0.0009	0.0407	0.0357	0.0353
10	100	0.3	-0.0147	-0.0087	0.0004	0.0404	0.0376	0.0366
10	100	0.6	-0.0046	-0.0088	0.0001	0.0392	0.0386	0.0375
10	100	0.9	0.0052	-0.0066	-0.0011	0.0408	0.0412	0.0395
10	200	0	0.0038	-0.0030	0.0001	0.0253	0.0251	0.0248
10	200	0.3	0.0054	-0.0014	-0.0001	0.0263	0.0259	0.0258
10	200	0.6	0.0034	0.0003	0.0002	0.0262	0.0265	0.0264
10	200	0.9	0.0041	0.0027	0.0021	0.0277	0.0285	0.0286
20	100	0	-0.0050	0.0025	0.0010	0.0239	0.0233	0.0235
20	100	0.3	-0.0025	0.0028	0.0007	0.0240	0.0235	0.0237
20	100	0.6	0.0044	0.0025	0.0004	0.0224	0.0220	0.0221
20	100	0.9	0.0052	0.0032	0.0001	0.0209	0.0205	0.0209
20	200	0	0.0038	-0.0011	0.0003	0.0175	0.0166	0.0165
20	200	0.3	0.0054	-0.0003	0.0001	0.0174	0.0165	0.0164
20	200	0.6	0.0034	0.0010	0.0000	0.0163	0.0154	0.0153
20	200	0.9	0.0041	0.0022	0.0002	0.0152	0.0144	0.0142