Bias-Corrected Estimation in Dynamic Panel Data Models

Maurice J. G. BUN
Faculty of Economics and Econometrics, University of Amsterdam, The Netherlands (*m.j.g.bun@uva.nl*)

Martin A. CARREE
Faculty of Economics and Business Administration, University of Maastricht, The Netherlands (*m.carree@os.unimaas.nl*)

This study develops a new bias-corrected estimator for the fixed-effects dynamic panel data model and derives its limiting distribution for finite number of time periods, T, and large number of cross-section units, N. The bias-corrected estimator is derived as a bias correction of the least squares dummy variable (within) estimator. It does not share some of the drawbacks of recently developed instrumental variables and generalized method-of-moments estimators and is relatively easy to compute. Monte Carlo experiments provide evidence that the bias-corrected estimator performs well even in small samples. The proposed technique is applied in an empirical analysis of unemployment dynamics at the U.S. state level for the 1991–2000 period.

KEY WORDS: Bias correction; Dynamic panel data model; Unemployment dynamics.

1. INTRODUCTION

The estimation of fixed-effects dynamic panel data models has been one of the main challenges in econometrics during the last two decades. Various instrumental variables (IV) estimators and generalized method-of-moments (GMM) estimators have been proposed and compared (see, e.g., Anderson and Hsiao 1981, 1982; Arellano and Bond 1991; Arellano and Bover 1995; Ahn and Schmidt 1995; Kiviet 1995; Wansbeek and Bekker 1996; Ziliak 1997; Blundell and Bond 1998; Hahn 1999; Judson and Owen 1999). The development and comparison of such new estimators was necessary because the traditional least squares dummy variable (LSDV) estimator is inconsistent for fixed T. Despite the increasing sophistication of the IV and GMM estimators, they have two important drawbacks. First, the complexity of the new estimators is a barrier for applied researchers (see, e.g., Baltagi, Griffin, and Xiong 2000). This should be only a temporary drawback, however, as the new estimators are incorporated into the statistical packages. But the newly developed estimators may require additional decisions on, for example, which and how many instruments to use. For example, by evaluating the expectation of asymptotic expansions of estimation errors, Bun and Kiviet (2002b) showed that finite-sample bias of GMM estimators increases with the number of moment conditions used. This makes application less straightforward. In addition, the new estimators introduce problems of their own. For example, the performance of some GMM estimators depends strongly on the ratio of variance of the individual-specific effects and the variance of the general error term (see, e.g., Kitazawa 2001; Bun and Kiviet 2002b).

This article introduces a new and simple estimator for dynamic panel data models with or without additional exogenous explanatory variables. An important advantage of this estimator is that it does not depend on the ratio of the variance of the individual-specific effects and the variance of the general error term. It is computed as a bias correction to the LSDV estimator (also referred to as the within estimator) and as such is related to estimators developed by Kiviet (1995), Hansen (2001), and Hahn and Kuersteiner (2002). MacKinnon and Smith (1998) already indicated that bias of parameter estimates may be virtually eliminated in some common cases, albeit at the expense of increased variance of the estimators. The present article confirms this for the case of dynamic panel data models. Regarding dynamic panel data models, Kiviet (1995) and Judson and Owen (1999) presented Monte Carlo evidence indicating that the bias-corrected estimator proposed by Kiviet (1995) may outperform IV and GMM estimators.

This article provides evidence of the usefulness of bias correction, but the resulting estimator does not share some limitations of existing bias-corrected procedures. First, Kiviet (1995) proposed consistently estimating the extent of the bias by using a preliminary consistent estimator. This allows for a consistent corrected estimator based on additive bias correction. An obvious disadvantage of such a procedure is that its finite-sample accuracy depends on the preliminary estimator chosen. Bias adjustment of the newly developed estimator is done without resorting to outside initial consistent estimates and appears to perform well in comparison. Second, Hansen (2001) proposed a somewhat similar bias-corrected estimator as in this study, but did not derive its limiting distribution. Also, the bias-correction procedure proposed by Hansen does not take into account the inconsistency of the LSDV estimator of the variance of the error term. Finally, Hahn and Kuersteiner (2002) recently introduced a bias-corrected estimator related to that developed by Kiviet (1995); however, their estimator is not designed for samples with small T.

The rest of the article is organized as follows. In Section 2 we explain the principle of bias correction in dynamic panel data models. In Section 3 we derive the limiting distribution of the bias-corrected estimator for finite T and large N. In Section 4 we discuss the special case of the AR(1) model in which no additional exogenous variables are included. We compare the bias-corrected estimator with other possible corrections on the

Journal of Business & Economic Statistics
April 2005, Vol. 23, No. 2
DOI 10.1198/073500104000000532

LSDV estimator. In Section 5 we present results from Monte Carlo experiments for the model with an additional exogenous regressor. In Section 6 we apply the estimators to a simple model of intertemporal dynamics of the unemployment rate in U.S. states in the 1991–2000 period. Finally, in Section 7 we discuss extensions and limitations of the proposed estimator in more general models and provide concluding remarks.

2. BIAS–CORRECTED ESTIMATION IN DYNAMIC PANEL DATA MODELS

In this section we illustrate the principle of bias-corrected estimation in the first-order dynamic panel data model. For ease of exposition, we assume only one additional time-varying regressor (next to the lagged dependent variable regressor) and the panel to be balanced. Consider the following first-order dynamic panel data model

$$y_{it} = \gamma y_{i,t-1} + \beta x_{it} + \eta_i + \varepsilon_{it}, \qquad i = 1, \ldots, N; t = 1, \ldots, T. \tag{1}$$

In this model the dependent variable y_{it} is determined by the one-period lagged value of the dependent variable $y_{i,t-1}$, the additional regressor x_{it}, the unobserved individual-specific effect η_i, and a general disturbance term ε_{it}. The regressor x_{it} may be correlated with the individual-specific effect η_i, but we assume that it is strictly exogenous with respect to the general error term ε_{it}. Regarding the latter, we assume that it has mean 0, constant variance $\sigma_{\boldsymbol{\varepsilon}}^2$, and finite fourth moment, not correlated either over time or across individuals and not correlated with η_i. Considering the startup observations y_{i0}, we assume that they are uncorrelated with subsequent error terms ε_{it}. Finally, there are no assumptions about the value of γ; that is, it is not necessary to assume that model (1) is dynamically stable.

The unknown individual effects in (1) can be eliminated by expressing each variable in deviation of its individual-specific mean. We introduce $\tilde{y}_{it} = y_{it} - \bar{y}_i$, $\tilde{y}_{i,t-1} = y_{i,t-1} - \bar{y}_{i,-1}$, $\tilde{x}_{it} = x_{it} - \bar{x}_i$, and $\tilde{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_i$ and rewrite model (1) as

$$\tilde{y}_{it} = \gamma \tilde{y}_{i,t-1} + \beta \tilde{x}_{it} + \tilde{\varepsilon}_{it}, \qquad i = 1, \ldots, N; t = 1, \ldots, T. \tag{2}$$

We compute the LSDV estimators by applying ordinary least squares (OLS) to this equation to give

$$\hat{\gamma}_{\text{lsdv}} = \frac{\sum\sum \tilde{x}_{it}^2 \sum\sum \tilde{y}_{i,t-1}\tilde{y}_{it} - \sum\sum \tilde{x}_{it}\tilde{y}_{i,t-1} \sum\sum \tilde{x}_{it}\tilde{y}_{it}}{\sum\sum \tilde{x}_{it}^2 \sum\sum \tilde{y}_{i,t-1}^2 - (\sum\sum \tilde{x}_{it}\tilde{y}_{i,t-1})^2} \tag{3}$$

and

$$\hat{\beta}_{\text{lsdv}} = \frac{-\sum\sum \tilde{x}_{it}\tilde{y}_{i,t-1} \sum\sum \tilde{y}_{i,t-1}\tilde{y}_{it} + \sum\sum \tilde{y}_{i,t-1}^2 \sum\sum \tilde{x}_{it}\tilde{y}_{it}}{\sum\sum \tilde{x}_{it}^2 \sum\sum \tilde{y}_{i,t-1}^2 - (\sum\sum \tilde{x}_{it}\tilde{y}_{i,t-1})^2}, \tag{4}$$

where the double summations are for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.

The LSDV estimators of γ and β are biased and inconsistent for fixed T because of the correlation between $\tilde{y}_{i,t-1}$ and $\tilde{\varepsilon}_{it}$. The extent of the inconsistency can be computed as follows. We rewrite (3) and (4) as

$$\hat{\gamma}_{\text{lsdv}} = \gamma + \frac{\sum\sum \tilde{x}_{it}^2 \sum\sum \tilde{y}_{i,t-1}\tilde{\varepsilon}_{it} - \sum\sum \tilde{x}_{it}\tilde{y}_{i,t-1} \sum\sum \tilde{x}_{it}\tilde{\varepsilon}_{it}}{\sum\sum \tilde{x}_{it}^2 \sum\sum \tilde{y}_{i,t-1}^2 - (\sum\sum \tilde{x}_{it}\tilde{y}_{i,t-1})^2} \tag{5}$$

and

$$\hat{\beta}_{\text{lsdv}} = \beta - \frac{\sum\sum \tilde{x}_{it}\tilde{y}_{i,t-1} \sum\sum \tilde{y}_{i,t-1}\tilde{\varepsilon}_{it} - \sum\sum \tilde{y}_{i,t-1}^2 \sum\sum \tilde{x}_{it}\tilde{\varepsilon}_{it}}{\sum\sum \tilde{x}_{it}^2 \sum\sum \tilde{y}_{i,t-1}^2 - (\sum\sum \tilde{x}_{it}\tilde{y}_{i,t-1})^2}. \tag{6}$$

From (1), we use continuous substitution to obtain

$$y_{it} = \gamma^t y_{i0} + \beta(x_{it} + \gamma x_{i,t-1} + \cdots + \gamma^{t-1} x_{i1}) + \frac{1-\gamma^t}{1-\gamma}\eta_i + \varepsilon_{it} + \gamma\varepsilon_{i,t-1} + \cdots + \gamma^{t-1}\varepsilon_{i1}. \tag{7}$$

Note that this also holds for the specific case of $\gamma = 1$, because we have $\lim_{\gamma\to 1}(1-\gamma^t)/(1-\gamma) = t$. To obtain an expression for $\tilde{y}_{i,t-1}$, we require the mean $\bar{y}_{i,-1}$. The sum of y_{i0} through $y_{i,T-1}$ equals

$$\sum_{t=1}^{T} y_{i,t-1} = \frac{1-\gamma^T}{1-\gamma} y_{i0} + \beta\left(x_{i,T-1} + \cdots + \frac{1-\gamma^{T-1}}{1-\gamma} x_{i1}\right) + \frac{(T-1) - T\gamma + \gamma^T}{(1-\gamma)^2}\eta_i + \varepsilon_{i,T-1} + \cdots + \frac{1-\gamma^{T-1}}{1-\gamma}\varepsilon_{i1}. \tag{8}$$

From this, it can be derived that when y_{i0} is uncorrelated with subsequent error terms ε_{it},

$$\operatorname*{plim}_{N\to\infty} \frac{\sum\sum \tilde{y}_{i,t-1}\tilde{\varepsilon}_{it}}{N(T-1)} = -\sigma_{\boldsymbol{\varepsilon}}^2 \frac{(T-1) - T\gamma + \gamma^T}{T(T-1)(1-\gamma)^2} = -\sigma_{\boldsymbol{\varepsilon}}^2 h(\gamma, T). \tag{9}$$

This expression is always negative (for $\gamma \geq -1$), because the function $h(\gamma, T)$ is positive. Having N tending to infinity and using $\operatorname{plim}_{N\to\infty} \frac{1}{N(T-1)} \sum\sum \tilde{x}_{it}\tilde{\varepsilon}_{it} = 0$ (because the error term $\tilde{\varepsilon}_{it}$ is assumed to be uncorrelated with $\tilde{x}_{it}$), we find that the inconsistency of the LSDV coefficient estimators equals (see also Nickell 1981, p. 1424; Kiviet 1995, p. 61)

$$\gamma^* = \operatorname*{plim}_{N\to\infty} (\hat{\gamma}_{\text{lsdv}} - \gamma) = \operatorname*{plim}_{N\to\infty} \frac{\sum\sum \tilde{y}_{i,t-1}\tilde{\varepsilon}_{it} / \sum\sum \tilde{y}_{i,t-1}^2}{1 - (\sum\sum \tilde{x}_{it}\tilde{y}_{i,t-1})^2 / (\sum\sum \tilde{x}_{it}^2 \sum\sum \tilde{y}_{i,t-1}^2)} \tag{10}$$

and

$$\beta^* = \operatorname*{plim}_{N\to\infty} (\hat{\beta}_{\text{lsdv}} - \beta) = -\operatorname*{plim}_{N\to\infty} \frac{\sum\sum \tilde{x}_{it}\tilde{y}_{i,t-1}}{\sum\sum \tilde{x}_{it}^2} (\hat{\gamma}_{\text{lsdv}} - \gamma). \tag{11}$$

We introduce the following expressions of the (asymptotic) variances of $\tilde{y}_{i,t-1}$ and $\tilde{x}_{it}$ and their (limiting) covariance: $\sigma_{\mathbf{y}_{-1}}^2 = \operatorname{plim}_{N\to\infty} \frac{1}{N(T-1)} \sum\sum \tilde{y}_{i,t-1}^2$, $\sigma_{\mathbf{x}}^2 = \operatorname{plim}_{N\to\infty} \frac{1}{N(T-1)} \times \sum\sum \tilde{x}_{it}^2$, and $\sigma_{\mathbf{xy}_{-1}} = \operatorname{plim}_{N\to\infty} \frac{1}{N(T-1)} \sum\sum \tilde{x}_{it}\tilde{y}_{i,t-1}$. The inconsistency of the LSDV coefficient estimators is now conveniently expressed as

$$\gamma^* = \frac{-\sigma_{\boldsymbol{\varepsilon}}^2 h(\gamma, T)}{(1-\rho_{\mathbf{xy}_{-1}}^2)\sigma_{y_{-1}}^2}, \qquad \beta^* = -\zeta\gamma^*, \tag{12}$$

where $\rho_{\mathbf{xy}_{-1}} = \sigma_{\mathbf{xy}_{-1}}/\sigma_{\mathbf{x}}\sigma_{\mathbf{y}_{-1}}$ and $\zeta = \sigma_{\mathbf{xy}_{-1}}/\sigma^2_{\mathbf{x}}$ are the (asymptotic) correlation coefficient between $\tilde{y}_{i,t-1}$ and $\tilde{x}_{it}$ and the (asymptotic) regression coefficient of $\tilde{y}_{i,t-1}$ on $\tilde{x}_{it}$. Note that the denominator $(1-\rho^2_{\mathbf{xy}_{-1}})\sigma^2_{\mathbf{y}_{-1}}$ in the first expression of (12) is the conditional variance of $\tilde{y}_{i,t-1}$ given $\tilde{x}_{it}$.

From the first expression in (12), it is clear that the LSDV estimator $\hat{\gamma}_{\text{lsdv}}$ is downward-biased. The extent of the (asymptotic) bias depends on five parameters: γ, T, $\sigma^2_{\boldsymbol{\varepsilon}}$, $\sigma^2_{\mathbf{y}_{-1}}$, and $\rho^2_{\mathbf{xy}_{-1}}$. The bias of the LSDV estimator will be especially severe when (a) the value of γ is close to 1 or even exceeds 1; (b) the number of time periods, T, is low; (c) the ratio of variances, $\sigma^2_{\boldsymbol{\varepsilon}}/\sigma^2_{\mathbf{y}_{-1}}$, is high; or (d) the lagged endogenous variable and the exogenous variable are highly correlated, either positively or negatively. The second expression in (12) shows that the inconsistency of $\hat{\beta}_{\text{lsdv}}$ is proportional to that of $\hat{\gamma}_{\text{lsdv}}$. The bias of the LSDV estimator $\hat{\beta}_{\text{lsdv}}$ can be either positive or negative, depending on the sign of the (asymptotic) covariance between $\tilde{y}_{i,t-1}$ and $\tilde{x}_{it}$.

The principle of bias correction can be explained straightforwardly using (12). First, assume that we would know the values for $\sigma^2_{\boldsymbol{\varepsilon}}$, $\rho_{\mathbf{xy}_{-1}}$, $\sigma^2_{\mathbf{y}_{-1}}$, and ζ. Then we may use as a bias-corrected estimator, $\hat{\gamma}_{\text{bc}}$ (where the subscript *bc* means "bias-corrected"; the fact that *bc* also are the initials of the authors' surnames is purely coincidental), that value of γ for which

$$\hat{\gamma}_{\text{lsdv}} = \gamma - \frac{\sigma^2_{\boldsymbol{\varepsilon}} h(\gamma, T)}{(1-\rho^2_{\mathbf{xy}_{-1}})\sigma^2_{\mathbf{y}_{-1}}}. \tag{13}$$

This estimator can then be inserted into the second expression in (12) to find the bias-corrected estimator $\hat{\beta}_{\text{bc}} = \hat{\beta}_{\text{lsdv}} + \zeta(\hat{\gamma}_{\text{lsdv}} - \hat{\gamma}_{\text{bc}})$. The function $h(\gamma, T)$ as defined in (9) plays an important role in this nonlinear bias-correction procedure. This function is always positive and monotonically increasing for $\gamma \ge -1$, a condition that usually can be safely assumed to hold in applications. For $\gamma = 1$, the function $h(\gamma, T)$ has a value of $h(1, T) = 1/2$ (using l'Hôpital's rule) irrespective of the length of time period T. For $T = 2$, the function $h(\gamma, 2)$ is equal to $1/2$, and for $T = 3$, the function $h(\gamma, 3)$ is equal to $(2+\gamma)/6$. Hence for $T = 2$, the bias-corrected estimator can be expressed explicitly as

$$\hat{\gamma}_{\text{bc}} = \hat{\gamma}_{\text{lsdv}} + \frac{\sigma^2_{\boldsymbol{\varepsilon}}}{2(1-\rho^2_{\mathbf{xy}_{-1}})\sigma^2_{\mathbf{y}_{-1}}} \quad \text{for } T = 2. \tag{14}$$

For $T = 3$, it can be expressed explicitly as

$$\hat{\gamma}_{\text{bc}} = \frac{6\hat{\gamma}_{\text{lsdv}} + 2\sigma^2_{\boldsymbol{\varepsilon}}/(1-\rho^2_{\mathbf{xy}_{-1}})\sigma^2_{\mathbf{y}_{-1}}}{6 - \sigma^2_{\boldsymbol{\varepsilon}}/(1-\rho^2_{\mathbf{xy}_{-1}})\sigma^2_{\mathbf{y}_{-1}}} \quad \text{for } T = 3. \tag{15}$$

For $T > 3$, (13) must be solved numerically. Equation (13) can for example be solved numerically as follows. Define $C = \sigma^2_{\boldsymbol{\varepsilon}}/(1-\rho^2_{\mathbf{xy}_{-1}})\sigma^2_{\mathbf{y}_{-1}}$ and take $\hat{\gamma}_{(0)} = \hat{\gamma}_{\text{lsdv}}$. An iterative procedure to converge toward the bias-corrected estimate (from below) is $\hat{\gamma}_{(j+1)} = \hat{\gamma}_{\text{lsdv}} + Ch(\hat{\gamma}_{(j)}, T)$.

In practice, we do not know the values for $\sigma^2_{\boldsymbol{\varepsilon}}$, $\rho_{\mathbf{xy}_{-1}}$, $\sigma^2_{\mathbf{y}_{-1}}$, and ζ. The values of the latter three variables can be estimated consistently using their sample analogs $\hat{\rho}_{\mathbf{xy}_{-1}} = \hat{\sigma}_{\mathbf{xy}_{-1}}/\hat{\sigma}_{\mathbf{x}}\hat{\sigma}_{\mathbf{y}_{-1}}$, $\hat{\sigma}^2_{\mathbf{y}_{-1}}$, and $\hat{\zeta} = \hat{\sigma}_{\mathbf{xy}_{-1}}/\hat{\sigma}^2_{\mathbf{x}}$. However, the LSDV estimator of $\sigma^2_{\boldsymbol{\varepsilon}}$ is inconsistent, and the variance of the error term can be consistently estimated only when the LSDV estimators for γ and β have been bias-corrected. We discuss three solutions to this problem that lead to the same bias-corrected estimates. First, we can use an iterative procedure for (13). We then substitute the LSDV estimate for $\sigma^2_{\boldsymbol{\varepsilon}}$ in (13) to achieve one-step estimates for γ and β. These estimates are used to compute the one-step estimate for $\sigma^2_{\boldsymbol{\varepsilon}}$. This one-step estimate is again substituted in (13) to achieve two-step estimates for γ and β and so on until convergence. Second, an alternative procedure is to use the expression for the inconsistency of the LSDV estimate for $\sigma^2_{\boldsymbol{\varepsilon}}$, that is,

$$\begin{aligned}
&\operatorname*{plim}_{N\to\infty} \hat{\sigma}^2_{\text{lsdv}} \\
&\quad = \operatorname*{plim}_{N\to\infty} \frac{\sum\sum(\tilde{y}_{it} - \hat{\gamma}_{\text{lsdv}}\tilde{y}_{i,t-1} - \hat{\beta}_{\text{lsdv}}\tilde{x}_{it})^2}{N(T-1)} \\
&\quad = \operatorname*{plim}_{N\to\infty} \frac{\sum\sum((\gamma - \hat{\gamma}_{\text{lsdv}})\tilde{y}_{i,t-1} + (\beta - \hat{\beta}_{\text{lsdv}})\tilde{x}_{it} + \tilde{\varepsilon}_{it})^2}{N(T-1)} \\
&\quad = \sigma^2_{\boldsymbol{\varepsilon}} - (1-\rho^2_{\mathbf{xy}_{-1}})\sigma^2_{\mathbf{y}_{-1}}\gamma^{*2}.
\end{aligned} \tag{16}$$

The expression for $\sigma^2_{\boldsymbol{\varepsilon}} = \hat{\sigma}^2_{\text{lsdv}} + (1-\rho^2_{\mathbf{xy}_{-1}})\sigma^2_{\mathbf{y}_{-1}}(\hat{\gamma}_{\text{lsdv}} - \gamma)^2$ is then substituted into (13) to arrive at an expression from which $\hat{\gamma}_{\text{bc}}$ can be derived in one step, that is,

$$\hat{\gamma}_{\text{lsdv}} = \gamma - \frac{\hat{\sigma}^2_{\text{lsdv}} h(\gamma, T)}{(1-\rho^2_{\mathbf{xy}_{-1}})\sigma^2_{\mathbf{y}_{-1}}} - (\hat{\gamma}_{\text{lsdv}} - \gamma)^2 h(\gamma, T). \tag{17}$$

Equation (17) can, for example, be solved numerically as follows. Define $C = \hat{\sigma}^2_{\text{lsdv}}/(1-\rho^2_{\mathbf{xy}_{-1}})\sigma^2_{\mathbf{y}_{-1}}$ and take $\hat{\gamma}_{(0)} = \hat{\gamma}_{\text{lsdv}}$. An iterative procedure to converge toward the bias-corrected estimate is $\hat{\gamma}_{(j+1)} = \hat{\gamma}_{\text{lsdv}} + \frac{1}{2h(\hat{\gamma}_{(j)}, T)}(1 - \sqrt{1 - 4Ch(\hat{\gamma}_{(j)}, T)^2})$. For $T = 2$, an analytic expression for the bias-corrected estimate can be derived as

$$\hat{\gamma}_{\text{bc}} = \hat{\gamma}_{\text{lsdv}} + 1 - \sqrt{1 - \hat{\sigma}^2_{\text{lsdv}}/(1-\rho^2_{\mathbf{xy}_{-1}})\sigma^2_{\mathbf{y}_{-1}}} \quad \text{for } T = 2. \tag{18}$$

Finally, to achieve bias-corrected estimates, $\sigma^2_{\boldsymbol{\varepsilon}}$ in equation (13) can be replaced by the infeasible estimate (we are grateful to a referee for suggesting this procedure)

$$\tilde{\sigma}^2_{\boldsymbol{\varepsilon}} = \frac{\sum\sum(\tilde{y}_{it} - \gamma\tilde{y}_{i,t-1} - \beta\tilde{x}_{it})^2}{N(T-1)}, \tag{19}$$

and we solve the following equivalent of (12) for γ and β simultaneously:

$$\begin{aligned}
\hat{\gamma}_{\text{lsdv}} &= \gamma - \frac{\tilde{\sigma}^2_{\boldsymbol{\varepsilon}}(\gamma, \beta) h(\gamma, T)}{(1-\rho^2_{\mathbf{xy}_{-1}})\sigma^2_{\mathbf{y}_{-1}}}, \\
\hat{\beta}_{\text{lsdv}} &= \beta - \zeta(\hat{\gamma}_{\text{lsdv}} - \gamma).
\end{aligned} \tag{20}$$

At first sight, this procedure would appear more cumbersome because there is an optimization with two arguments (γ and β) instead of one argument (γ). However, there is an advantage to deriving the expression for (asymptotic) standard errors, because the (asymptotic) distribution of $\hat{\sigma}^2_{\text{lsdv}}$ is not necessary. Using any one of the iterative procedures (13), (17), or (20) results in the same bias-corrected estimate, $\hat{\gamma}_{\text{bc}}$, $\hat{\beta}_{\text{bc}}$, or $\hat{\sigma}^2_{\text{bc}}$.

3. ASYMPTOTIC PROPERTIES OF BIAS–CORRECTED ESTIMATORS

In this section we discuss the asymptotic properties of the proposed bias-corrected estimators. We derive consistency and asymptotic normality for the corrected estimators for finite T and N large. We generalize the discussion to the case with K additional exogenous variables, x_{1it} through x_{Kit}, and use matrix notation. Stacking the observations over time, that is, $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$, $\mathbf{y}_{i,-1} = (y_{i0}, \ldots, y_{i,T-1})'$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)'$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$, and $\mathbf{X}_i$ a matrix with the (t, k) element equal to x_{itk}, we extend (1) to

$$\mathbf{y}_i = \gamma \mathbf{y}_{i,-1} + \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\iota}_T \eta_i + \boldsymbol{\varepsilon}_i, \qquad i = 1, \ldots, N, \tag{21}$$

where $\boldsymbol{\iota}_T = (1, \ldots, 1)'$ is a $T \times 1$ vector of 1's. Stacking once more over individuals, that is, $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_N')'$, $\mathbf{y}_{-1} = (\mathbf{y}_{1,-1}', \ldots, \mathbf{y}_{N,-1}')'$, $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_N)'$, $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1', \ldots, \boldsymbol{\varepsilon}_N')'$, and $\mathbf{X} = (\mathbf{X}_1', \ldots, \mathbf{X}_N')'$, we have the following model:

$$\begin{aligned} \mathbf{y} &= \gamma \mathbf{y}_{-1} + \mathbf{X}\boldsymbol{\beta} + (\mathbf{I}_N \otimes \boldsymbol{\iota}_T)\boldsymbol{\eta} + \boldsymbol{\varepsilon} \\ &= \mathbf{W}\boldsymbol{\delta} + (\mathbf{I}_N \otimes \boldsymbol{\iota}_T)\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \end{aligned} \tag{22}$$

where we have defined the $NT \times (K+1)$ matrix $\mathbf{W} = [\mathbf{y}_{-1} \vdots \mathbf{X}]$ and the $(K+1)$ parameter vector $\boldsymbol{\delta} = (\gamma, \boldsymbol{\beta}')'$. The LSDV estimator for model (22) is equal to

$$\begin{aligned} \hat{\boldsymbol{\delta}}_{\text{lsdv}} &= (\mathbf{W}'\mathbf{A}\mathbf{W})^{-1}\mathbf{W}'\mathbf{A}\mathbf{y} \\ &= \begin{pmatrix} \hat{\sigma}^2_{\mathbf{y}_{-1}} & \hat{\boldsymbol{\Sigma}}'_{\mathbf{x}\mathbf{y}_{-1}} \\ \hat{\boldsymbol{\Sigma}}_{\mathbf{x}\mathbf{y}_{-1}} & \hat{\boldsymbol{\Sigma}}_{\mathbf{xx}} \end{pmatrix}^{-1} \begin{pmatrix} \hat{\sigma}_{\mathbf{y}_{-1}\mathbf{y}} \\ \hat{\boldsymbol{\Sigma}}_{\mathbf{xy}} \end{pmatrix}, \end{aligned} \tag{23}$$

where the $NT \times NT$ idempotent matrix $\mathbf{A} = \mathbf{I}_N \otimes (\mathbf{I}_T - \frac{1}{T}\boldsymbol{\iota}_T\boldsymbol{\iota}_T')$ is the within-transformation matrix that eliminates the individual effects and $\hat{\boldsymbol{\Sigma}}_{\mathbf{xy}}$, $\hat{\boldsymbol{\Sigma}}_{\mathbf{xx}}$, $\hat{\sigma}_{\mathbf{y}_{-1}\mathbf{y}}$, and $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}\mathbf{y}_{-1}}$ are sample analogs of $\boldsymbol{\Sigma}_{\mathbf{xy}} = \text{plim}_{N\to\infty} \frac{1}{N(T-1)}\mathbf{X}'\mathbf{A}\mathbf{y}$, $\boldsymbol{\Sigma}_{\mathbf{xx}} = \text{plim}_{N\to\infty} \frac{1}{N(T-1)} \times \mathbf{X}'\mathbf{A}\mathbf{X}$, $\sigma_{\mathbf{y}_{-1}\mathbf{y}} = \text{plim}_{N\to\infty} \frac{1}{N(T-1)}\mathbf{y}_{-1}'\mathbf{A}\mathbf{y}$, and $\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{y}_{-1}} = \text{plim}_{N\to\infty} \frac{1}{N(T-1)}\mathbf{X}'\mathbf{A}\mathbf{y}_{-1}$.

Define the inconsistency of the LSDV estimator as $\boldsymbol{\delta}^* = \text{plim}_{N\to\infty}(\hat{\boldsymbol{\delta}}_{\text{lsdv}} - \boldsymbol{\delta})$. We now introduce $\rho^2_{\mathbf{X}\mathbf{y}_{-1}} = \boldsymbol{\Sigma}'_{\mathbf{x}\mathbf{y}_{-1}} \times \boldsymbol{\Sigma}^{-1}_{\mathbf{xx}}\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{y}_{-1}}/\sigma^2_{\mathbf{y}_{-1}}$ as the (asymptotic) squared multiple correlation coefficient of the regression of $\tilde{y}_{i,t-1}$ on $\tilde{x}_{1it}$ through $\tilde{x}_{Kit}$ and $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_K) = \boldsymbol{\Sigma}^{-1}_{\mathbf{xx}}\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{y}_{-1}}$ as the corresponding vector of regression coefficients. This allows us to generalize (12) and express the inconsistency $\boldsymbol{\delta}^* = (\gamma^*, \boldsymbol{\beta}^{*\prime})'$ as

$$\gamma^* = \frac{-\sigma^2_{\boldsymbol{\varepsilon}} h(\gamma, T)}{(1 - \rho^2_{\mathbf{X}\mathbf{y}_{-1}})\sigma^2_{\mathbf{y}_{-1}}}, \quad \beta^*_k = -\zeta_k \gamma^*, \qquad k = 1, \ldots, K. \tag{24}$$

Although inconsistent, the LSDV estimator has a limiting distribution for $N \to \infty$ and fixed T. Bun and Kiviet (2001) derived the limiting distribution as

$$\sqrt{N}(\hat{\boldsymbol{\delta}}_{\text{lsdv}} - \boldsymbol{\delta}^* - \boldsymbol{\delta}) \xrightarrow[N\to\infty]{d} \mathcal{N}[\mathbf{0}, \mathbf{V}_{\mathbf{X}}], \tag{25}$$

where

$$\mathbf{V}_{\mathbf{X}} = \sigma^2_{\boldsymbol{\varepsilon}} \boldsymbol{\Sigma}^{-1}_{\mathbf{WAW}} + \sigma^4_{\boldsymbol{\varepsilon}} z(\gamma, T)\boldsymbol{\Sigma}^{-1}_{\mathbf{WAW}}\mathbf{e}_{K+1}\mathbf{e}'_{K+1}\boldsymbol{\Sigma}^{-1}_{\mathbf{WAW}}, \tag{26}$$

with $\boldsymbol{\Sigma}_{\mathbf{WAW}} = \text{plim}_{N\to\infty} \frac{1}{N}\mathbf{W}'\mathbf{A}\mathbf{W}$, $\mathbf{e}_{K+1}$ the $(K+1)$ vector with the first element equal to 1 and the other elements equal to 0, and $z(\gamma, T)$ equal to

$$z(\gamma, T) = -\frac{1 + 2\gamma^{T-1}}{(1-\gamma)^2} + \frac{2(1-\gamma^T)}{T(1-\gamma)^3} + \frac{(1-\gamma^T)^2}{T^2(1-\gamma)^4}. \tag{27}$$

The function $z(\gamma, T)$ is equal to $\text{tr}(\boldsymbol{\Pi}^2_T)$, where $\boldsymbol{\Pi}_T = \mathbf{A}_T\mathbf{L}_T\boldsymbol{\Gamma}_T$, with $\mathbf{A}_T = \mathbf{I}_T - \frac{1}{T}\boldsymbol{\iota}_T\boldsymbol{\iota}_T'$ the within-transformation matrix,

$$\mathbf{L}_T = \begin{bmatrix} 0 & 0 & \cdot & \cdot & 0 & 0 \\ 1 & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & 1 & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 & 0 \\ 0 & 0 & \cdot & 0 & 1 & 0 \end{bmatrix} \quad \text{and}$$
$$\boldsymbol{\Gamma}_T = \begin{bmatrix} 1 & 0 & \cdot & \cdot & 0 & 0 \\ \gamma & 1 & 0 & \cdot & \cdot & \cdot \\ \gamma^2 & \gamma & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & 0 \\ \gamma^{T-1} & \gamma^{T-2} & \cdot & \gamma^2 & \gamma & 1 \end{bmatrix}, \tag{28}$$

(see Bun and Kiviet 2001). We have that $\lim_{\gamma\to 1} z(\gamma, T) = -(T-1)(T-5)/12$. For $T = 2$, the value of $z(\gamma, T)$ is equal to $1/4$; for $T = 3$, it is equal to $(\gamma^2 + 4\gamma - 2)/9$. The expression for the inconsistency holds irrespective of the distribution of the error term ε_{it}. However, the specific expression for the matrix $\mathbf{V}_{\mathbf{X}}$ holds under normality of the error term only. Using notation introduced earlier, the variance–covariance matrix $\mathbf{V}_{\mathbf{X}}$ of the limiting distribution (25) of the LSDV estimator can be expressed as

$$\begin{aligned} \mathbf{V}_{\mathbf{X}} &= \begin{pmatrix} V^{11}_{\mathbf{X}} & V^{12}_{\mathbf{X}} \\ V^{21}_{\mathbf{X}} & V^{22}_{\mathbf{X}} \end{pmatrix} \\ &= \frac{\sigma^2_{\boldsymbol{\varepsilon}}}{T-1} \begin{pmatrix} \sigma^2_{\mathbf{y}_{-1}} & \boldsymbol{\Sigma}'_{\mathbf{x}\mathbf{y}_{-1}} \\ \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{y}_{-1}} & \boldsymbol{\Sigma}_{\mathbf{xx}} \end{pmatrix}^{-1} \\ &\quad + \frac{\sigma^4_{\boldsymbol{\varepsilon}} z(\gamma, T)}{(T-1)^2(1-\rho^2_{\mathbf{X}\mathbf{y}_{-1}})^2\sigma^4_{\mathbf{y}_{-1}}} \begin{pmatrix} 1 & -\boldsymbol{\zeta}' \\ -\boldsymbol{\zeta} & \boldsymbol{\zeta}\boldsymbol{\zeta}' \end{pmatrix}. \end{aligned} \tag{29}$$

The result (25) of Bun and Kiviet (2001) showed that the LSDV estimator has a limiting normal distribution for finite T and $N \to \infty$, but it is not centered at $\boldsymbol{\delta}$, and it has a nonstandard variance–covariance matrix.

We now turn to bias-corrected estimation of $\boldsymbol{\delta} = (\gamma, \boldsymbol{\beta}')'$. We first assume $\sigma^2_{\boldsymbol{\varepsilon}}$ to be given. Generalizing the results of Section 2 [see (13)], using (24), the bias-corrected estimator for γ is that γ which solves

$$\hat{\gamma}_{\text{lsdv}} = \gamma - \frac{\sigma^2_{\boldsymbol{\varepsilon}} h(\gamma, T)}{(1-\rho^2_{\mathbf{X}\mathbf{y}_{-1}})\sigma^2_{\mathbf{y}_{-1}}}. \tag{30}$$

The resulting estimator can then be inserted into the second expression in (24) to find the bias-corrected estimator for $\boldsymbol{\beta}$. In short, we solve $\hat{\boldsymbol{\delta}}_{\text{lsdv}} = \mathbf{g}(\boldsymbol{\delta})$ for $\boldsymbol{\delta}$ with

$$\mathbf{g}(\boldsymbol{\delta}) = \boldsymbol{\delta} + \boldsymbol{\delta}^* = \begin{pmatrix} \gamma - \sigma^2_{\boldsymbol{\varepsilon}} h(\gamma)/\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}} \\ \boldsymbol{\beta} + \sigma^2_{\boldsymbol{\varepsilon}} \boldsymbol{\zeta} h(\gamma)/\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}} \end{pmatrix}, \tag{31}$$

where $\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}} = (1-\rho^2_{\mathbf{X}\mathbf{y}_{-1}})\sigma^2_{\mathbf{y}_{-1}}$ is the conditional variance of $\tilde{\mathbf{y}}_{-1}$. Defining $\mathbf{f}(\boldsymbol{\delta}) = \mathbf{g}^{-1}(\boldsymbol{\delta})$, the expression for the bias-corrected estimator is

$$\hat{\boldsymbol{\delta}}_{\text{bc}} = \mathbf{f}(\hat{\boldsymbol{\delta}}_{\text{lsdv}}). \tag{32}$$

The function $\mathbf{f}$ is unknown but can be evaluated numerically using only a few lines of computer code; for details, see Section 2.

From (32), we see that

$$\operatorname*{plim}_{N\to\infty} \hat{\boldsymbol{\delta}}_{\text{bc}} = \operatorname*{plim}_{N\to\infty} \mathbf{f}(\hat{\boldsymbol{\delta}}_{\text{lsdv}}) = \mathbf{g}^{-1}\Big(\operatorname*{plim}_{N\to\infty} \hat{\boldsymbol{\delta}}_{\text{lsdv}}\Big) = \boldsymbol{\delta},$$

and hence the bias-corrected estimator is a consistent estimator of $\boldsymbol{\delta}$ for finite T and $N\to\infty$. Furthermore, exploiting (25) and using the delta method, we have

$$\sqrt{N}(\hat{\boldsymbol{\delta}}_{\text{bc}} - \boldsymbol{\delta}) \xrightarrow[N\to\infty]{d} \mathcal{N}[\mathbf{0}, \mathbf{F}\mathbf{V}_{\mathbf{X}}\mathbf{F}'], \tag{33}$$

where $\mathbf{F}$ is the $(K+1)\times(K+1)$ matrix of first partial derivatives of the vector function $\mathbf{f}$. Hence the bias-corrected estimator (32) has a limiting normal distribution centered at $\boldsymbol{\delta}$. Its asymptotic variance depends on $\mathbf{V}_{\mathbf{X}}$ and $\mathbf{F}$. The latter matrix is simply $\mathbf{F} = \mathbf{G}^{-1}$ with

$$\mathbf{G} = \begin{pmatrix} 1 - \sigma^2_{\boldsymbol{\varepsilon}} h'(\gamma)/\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}} & 0 \\ \sigma^2_{\boldsymbol{\varepsilon}} \boldsymbol{\zeta} h'(\gamma)/\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}} & \mathbf{I} \end{pmatrix}$$

as the Jacobian matrix of $\mathbf{g}(\boldsymbol{\delta})$ and

$$h'(\gamma) = \frac{(T-2)(1-\gamma^T) - T\gamma(1-\gamma^{T-2})}{T(T-1)(1-\gamma)^3}.$$

Using results on partitioned matrix inversion, the matrix $\mathbf{F} = \mathbf{G}^{-1}$ can be written as

$$\mathbf{F} = \begin{pmatrix} 1/(1 - \sigma^2_{\varepsilon} h'(\gamma)/\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}}) & 0 \\ -\sigma^2_{\boldsymbol{\varepsilon}} \boldsymbol{\zeta} h'(\gamma)/(\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}} - \sigma^2_{\boldsymbol{\varepsilon}} h'(\gamma)) & \mathbf{I} \end{pmatrix}. \tag{34}$$

This implies that the first diagonal element of the matrix $\mathbf{F}\mathbf{V}_{\mathbf{X}}\mathbf{F}'$, or $N * \text{var}(\hat{\gamma}_{\text{bc}})$, is simply equal to $V^{11}_{\mathbf{X}}/(1 - \sigma^2_{\boldsymbol{\varepsilon}} h'(\gamma)/\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}})^2$ and that $N * \text{var}(\hat{\boldsymbol{\beta}}_{\text{bc}})$ is equal to $V^{22}_{\mathbf{X}}$. For $T = 2$, the matrix $\mathbf{F}$ equals the unity matrix $\mathbf{I}$, because then $h'(\gamma) = 0$.

In general, $\sigma^2_{\boldsymbol{\varepsilon}}$ is unknown and also must be estimated. There are at least three equivalent approaches leading to the same bias-corrected estimator; see Section 2 for details. First, we can use an iterative procedure. Second, we can extend $\boldsymbol{\delta}$ to $\boldsymbol{\delta} = (\gamma, \boldsymbol{\beta}', \sigma^2_{\boldsymbol{\varepsilon}})'$. Exploiting (16), we now solve $\hat{\boldsymbol{\delta}}_{\text{lsdv}} = \mathbf{g}(\boldsymbol{\delta})$ for $\boldsymbol{\delta}$ with

$$\mathbf{g}(\boldsymbol{\delta}) = \begin{pmatrix} \gamma - \sigma^2_{\boldsymbol{\varepsilon}} h(\gamma)/\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}} \\ \boldsymbol{\beta} + \sigma^2_{\boldsymbol{\varepsilon}} \boldsymbol{\zeta} h(\gamma)/\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}} \\ \sigma^2_{\boldsymbol{\varepsilon}} + \sigma^4_{\boldsymbol{\varepsilon}} h^2(\gamma)/\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}} \end{pmatrix}. \tag{35}$$

Finally, we can use the original $\boldsymbol{\delta} = (\gamma, \boldsymbol{\beta}')'$ and the (infeasible) estimate $\tilde{\sigma}^2_{\boldsymbol{\varepsilon}}(\gamma, \boldsymbol{\beta})$, where

$$\tilde{\sigma}^2_{\boldsymbol{\varepsilon}}(\gamma, \boldsymbol{\beta}) = \frac{(\mathbf{y} - \gamma\mathbf{y}_{-1} - \mathbf{X}\boldsymbol{\beta})'\mathbf{A}(\mathbf{y} - \gamma\mathbf{y}_{-1} - \mathbf{X}\boldsymbol{\beta})}{N(T-1)}. \tag{36}$$

We then have that (31) is replaced by

$$\mathbf{g}(\boldsymbol{\delta}) = \begin{pmatrix} \gamma - \tilde{\sigma}^2_{\boldsymbol{\varepsilon}}(\gamma, \boldsymbol{\beta}) h(\gamma)/\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}} \\ \boldsymbol{\beta} + \tilde{\sigma}^2_{\boldsymbol{\varepsilon}}(\gamma, \boldsymbol{\beta}) \boldsymbol{\zeta} h(\gamma)/\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}} \end{pmatrix}. \tag{37}$$

The corresponding matrix $\mathbf{G}$ is then derived as

$$\mathbf{G} = \begin{pmatrix} 1 - (\tilde{\sigma}^2_{\gamma} h(\gamma) + \tilde{\sigma}^2_{\boldsymbol{\varepsilon}} h'(\gamma))/\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}} & -\tilde{\sigma}^{2\prime}_{\boldsymbol{\beta}} h(\gamma)/\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}} \\ \boldsymbol{\zeta}(\tilde{\sigma}^2_{\gamma} h(\gamma) + \tilde{\sigma}^2_{\boldsymbol{\varepsilon}} h'(\gamma))/\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}} & \mathbf{I} + \boldsymbol{\zeta}\tilde{\sigma}^{2\prime}_{\boldsymbol{\beta}} h(\gamma)/\sigma^2_{\mathbf{y}_{-1}|\mathbf{X}} \end{pmatrix},$$

where $\tilde{\sigma}^2_{\gamma} = 2\gamma\hat{\sigma}^2_{\mathbf{y}_{-1}} - 2\hat{\sigma}_{\mathbf{y}_{-1}\mathbf{y}} + 2\hat{\boldsymbol{\Sigma}}'_{\mathbf{x}\mathbf{y}_{-1}}\boldsymbol{\beta}$ and $\tilde{\sigma}^2_{\boldsymbol{\beta}} = 2\hat{\boldsymbol{\Sigma}}_{\mathbf{x}\mathbf{x}}\boldsymbol{\beta} - 2\hat{\boldsymbol{\Sigma}}_{\mathbf{x}\mathbf{y}} + 2\gamma\hat{\boldsymbol{\Sigma}}_{\mathbf{x}\mathbf{y}_{-1}}$ are the derivates of $\tilde{\sigma}^2_{\boldsymbol{\varepsilon}}(\gamma, \boldsymbol{\beta})$ with respect to γ and the $\boldsymbol{\beta}$-vector. The latter ($\tilde{\sigma}^2_{\boldsymbol{\beta}}$) can be shown to equal 0 when evaluated for the bias-corrected estimators, somewhat simplifying calculation of $\mathbf{G}$. We use this last approach to compute asymptotic standard errors in the simulation and empirical exercises. They are estimated consistently by $\frac{1}{N}\hat{\mathbf{F}}\hat{\mathbf{V}}_{\mathbf{X}}\hat{\mathbf{F}}'$ using the bias-corrected estimators and $\hat{\mathbf{F}} = \hat{\mathbf{G}}^{-1}$. We now turn to the specific case of having no additional exogenous variables.

4. BIAS–CORRECTION IN THE PANEL AR(1) MODEL

In this section we apply the limiting distribution theory of the previous section to a special case, the first-order dynamic panel data model without additional exogenous variables. We analyze the model

$$y_{it} = \gamma y_{i,t-1} + \eta_i + \varepsilon_{it}, \qquad i = 1,\dots,N;\ t = 1,\dots,T. \tag{38}$$

This model is a special case of (21) where $\beta = 0$. An important difference from the preceding sections is that here we make explicit assumptions about the admissible values for γ and about the distribution of the initial observations y_{i0}. Regarding γ, we now assume that $|\gamma| < 1$, and for the initial observations, we assume that the process (38) has been going on for a long time, that is,

$$y_{i0} = \frac{1}{1-\gamma}\eta_i + \frac{\varepsilon_{i0}}{\sqrt{1-\gamma^2}}, \qquad i = 1,\dots,N, \tag{39}$$

with the same assumptions about ε_{i0} as for the other disturbance terms ε_{it}, $t = 1,\dots,T$ (see Sec. 2). Note that this specific assumption about y_{i0} matches our earlier assumption about the initial observations made in Section 2; that is, all N startup observations y_{i0} are uncorrelated with all ε_{it} for $t > 0$. However, the additional assumptions about γ and y_{i0} enable us to derive explicit expressions for the inconsistency of the LSDV estimator and its asymptotic variance as a function of γ and T, as we discuss later. This makes it possible to analytically compute and compare the asymptotic efficiency of original and bias-corrected LSDV estimators.

Stacking the observations over time and across individuals, we get

$$\mathbf{y} = \gamma\mathbf{y}_{-1} + (\mathbf{I}_N \otimes \boldsymbol{\iota}_T)\eta + \boldsymbol{\varepsilon}. \tag{40}$$

Focusing on the autoregressive parameter γ, estimation of model (40) by OLS yields

$$\hat{\gamma}_{\text{lsdv}} = (\mathbf{y}'_{-1}\mathbf{A}\mathbf{y}_{-1})^{-1}\mathbf{y}'_{-1}\mathbf{A}\mathbf{y} = \gamma + (\mathbf{y}'_{-1}\mathbf{A}\mathbf{y}_{-1})^{-1}\mathbf{y}'_{-1}\mathbf{A}\boldsymbol{\varepsilon}. \tag{41}$$

The inconsistency of the LSDV estimator for γ when N tends to infinity can be expressed as (Nickell 1981; Hsiao 1986)

$$\begin{aligned}\gamma^* &= \operatorname*{plim}_{N\to\infty}(\hat{\gamma}_{\text{lsdv}} - \gamma) \\ &= \left(\operatorname*{plim}_{N\to\infty}\frac{1}{N}(\mathbf{y}_{-1}'\mathbf{A}\mathbf{y}_{-1})\right)^{-1}\operatorname*{plim}_{N\to\infty}\frac{1}{N}(\mathbf{y}_{-1}'\mathbf{A}\boldsymbol{\varepsilon}) \\ &= -\frac{1+\gamma}{T-1}\left(1-\frac{1-\gamma^T}{T(1-\gamma)}\right) \\ &\quad\times\left(1-\frac{2\gamma}{(1-\gamma)(T-1)}\left(1-\frac{1-\gamma^T}{T(1-\gamma)}\right)\right)^{-1}. \qquad (42)\end{aligned}$$

Note that the inconsistency of the LSDV estimator is a function of γ for fixed T and does not depend on $\sigma_{\boldsymbol{\varepsilon}}^2$; that is, we have $\text{plim}_{N\to\infty}(\hat{\gamma}_{\text{lsdv}}) = \gamma^* + \gamma = g(\gamma)$ for given T. In the interval $[-1, 1)$, the function g is a monotonically increasing function of γ with minimum value $g(-1) = -1$ and maximum value $g(1) = 1 - 3/(T+1)$, the latter of which is computed using l'Hôpital's rule. Hence it is possible to invert the function g and express γ as a function of $\text{plim}_{N\to\infty}(\hat{\gamma}_{\text{lsdv}})$, that is, $\gamma = f(\text{plim}_{N\to\infty}(\hat{\gamma}_{\text{lsdv}}))$ with $f = g^{-1}$. Analogous to previous sections, a consistent bias-corrected estimator thus can be constructed as

$$\hat{\gamma}_{\text{bc}} = f(\hat{\gamma}_{\text{lsdv}}). \qquad (43)$$

For example, when $T = 2$, we find from (42) that $\text{plim}_{N\to\infty}(\hat{\gamma}_{\text{lsdv}}) = (\gamma - 1)/2$. Hence we use $2\hat{\gamma}_{\text{lsdv}} + 1$ as a bias-corrected estimator for γ. However, for higher values of T, the function f is unknown but can be evaluated numerically or approximated by a known function. However, in the latter case consistency is lost, due to the approximation. Carree (2002) proposed approximating the function f by a linear specification. His estimate is easy to calculate but requires using a table to obtain values for the intercept and slope. Furthermore, the estimator is inconsistent for $N \to \infty$, due to the approximation. These properties make this estimator less appealing.

We now turn to limiting distributions of the LSDV estimator and the proposed bias-corrected estimator. Exploiting (25), the limiting distribution for $\hat{\gamma}_{\text{lsdv}}$ for finite T and large N is

$$\sqrt{N}(\hat{\gamma}_{\text{lsdv}} - \gamma^* - \gamma) \xrightarrow[N\to\infty]{d} \mathcal{N}[0, V], \qquad (44)$$

where $V = \sigma_{\boldsymbol{\varepsilon}}^2/(T-1)\sigma_{\mathbf{y}_{-1}}^2 + \sigma_{\boldsymbol{\varepsilon}}^4 z(\gamma, T)/(T-1)^2\sigma_{\mathbf{y}_{-1}}^4$ and $z(\gamma, T)$ as in (27). For given T, this limiting distribution depends only on γ as the factor $\sigma_{\boldsymbol{\varepsilon}}^2$ in V cancels out, because $\sigma_{\mathbf{y}_{-1}}^2$ is proportional to $\sigma_{\boldsymbol{\varepsilon}}^2$. Using equation (14) of Nickell (1981), we have that

$$\begin{aligned}V = &\frac{T(1-\gamma^2)(1-\gamma)^2}{T(T-1)(1-\gamma)^2 - 2T\gamma(1-\gamma) + 2\gamma(1-\gamma^T)} \\ &+ z(\gamma, T) \\ &\times\left(\frac{T(1-\gamma^2)(1-\gamma)^2}{T(T-1)(1-\gamma)^2 - 2T\gamma(1-\gamma) + 2\gamma(1-\gamma^T)}\right)^2. \end{aligned} \qquad (45)$$

Regarding the bias-corrected estimator (43), we find, using (33), that

$$\sqrt{N}(\hat{\gamma}_{\text{bc}} - \gamma) \xrightarrow[N\to\infty]{d} \mathcal{N}\left(0, \frac{V}{(g'(\gamma))^2}\right). \qquad (46)$$

The asymptotic variance depends on V and the first derivative of the function g. Evaluating the latter factor analytically is cumbersome, but it can be approximated numerically. In fact, to compute the variance of $\hat{\gamma}_{\text{bc}}$, we insert this estimate into (45) to find $\hat{V}$. We then approximate the first derivative of g using the expression for $\gamma^*(\gamma)$ as given in (42) by $g'(\hat{\gamma}_{\text{bc}}) = 1 + [\gamma^*(\hat{\gamma}_{\text{bc}}) - \gamma^*(\hat{\gamma}_{\text{bc}} - \mu)]/\mu$, with μ a small number, say .001. We could also actually derive the analytic first derivative from (42), but this is not an elegant expression.

For the dynamic panel data model without additional exogenous regressors (38), other estimators can be used that are not consistent for fixed T but are simple to compute, being linear functions of the LSDV estimator. It is interesting to compare their asymptotic efficiency with that of the $\hat{\gamma}_{\text{bc}}$ estimator. A first estimator emerges from taking a linear approximation to (42). When we insert in (42) values for γ equal to 0 and 1 (using l'Hôpital's rule), we find that for $\gamma = 0$, $\text{plim}_{N\to\infty}(\hat{\gamma}_{\text{lsdv}}) = -1/T$ and that for $\gamma \to 1$, $\text{plim}_{N\to\infty}(\hat{\gamma}_{\text{lsdv}}) = 1 - 3/(T+1)$. A linear approximation for the function f in the $\gamma \in [0, 1)$ interval is found by connecting the points $(-1/T, 0)$ and $(1 - 3/(T+1), 1)$. Hence the proposed estimator is

$$\hat{\gamma}_c = \frac{T^2+T}{T^2-T+1}\hat{\gamma}_{\text{lsdv}} + \frac{T+1}{T^2-T+1}. \qquad (47)$$

The estimator in (47) strongly resembles an estimator proposed by Hahn and Kuersteiner (2002),

$$\hat{\gamma}_{\text{hk}} = \frac{T+1}{T}\hat{\gamma}_{\text{lsdv}} + \frac{1}{T}. \qquad (48)$$

Although the estimators (47) and (48) are also inconsistent for finite T, the leading bias term of order $O(T^{-1})$ has been accounted for. Hence these estimators may perform reasonably well for moderate T.

Each of the three estimators (43), (47), and (48) are functions of $\hat{\gamma}_{\text{lsdv}}$, for which we know the limiting distribution (44), which is dependent on γ and T. This makes it possible to analytically compute asymptotic bias and variance of the estimators. These are presented in Table 1 for values of T equal to 3, 6, and 10 and values of γ equal to 0, .4, and .8. The bias-corrected estimator $\hat{\gamma}_{\text{bc}}$ has (by definition) the lowest bias, whereas the Hahn and Kuersteiner estimator has considerable bias for small T. The latter estimator has the lowest asymptotic variance of the three estimators, however. In terms of mean squared error (MSE), $\hat{\gamma}_{\text{bc}}$ would be preferable if we had small T and N large, because the extent of bias would dominate this measure for such dimensions.

5. MONTE CARLO EXPERIMENTS

In this section we compare the performance of the bias-corrected estimator (32), denoted by *bc*, with some alternative estimators in a first-order dynamic panel model with an additional exogenous regressor. We compare *bc* with the original LSDV estimator (*lsdv*), an additive bias-corrected estimator (*ac*), and the GMM estimator (*gmm*) of Arellano and Bond (1991). For *ac*, we use a slightly different version of Kiviet's (1995) estimator in which there is bias correction of the first-order term only. Bun and Kiviet (2002a) showed

Table 1. Asymptotic Bias and Variance for the Panel AR(1) Model

T	γ	γ^*	g'	V	N*var($\hat{\gamma}_{bc}$)	bias($\hat{\gamma}_c$)	N*var($\hat{\gamma}_c$)	bias($\hat{\gamma}_{hk}$)	N*var($\hat{\gamma}_{hk}$)
3	0	−.333	.611	.444	1.191	0	1.306	−.111	.790
3	.4	−.494	.587	.607	1.763	.010	1.785	−.192	1.080
3	.8	−.663	.569	.814	2.513	.006	2.391	−.284	1.447
6	0	−.167	.811	.174	.265	0	.320	−.028	.237
6	.4	−.251	.762	.161	.278	.028	.296	−.059	.219
6	.8	−.361	.684	.146	.312	.020	.268	−.121	.198
10	0	−.100	.891	.101	.128	0	.148	−.010	.123
10	.4	−.148	.864	.086	.116	.026	.126	−.023	.104
10	.8	−.218	.768	.051	.087	.024	.075	−.060	.062

NOTE: The asymptotic bias($\hat{\gamma}_{bc}$) is always equal to 0. The value for V is N*var($\hat{\gamma}_{lsdv}$).

that this first-order term is responsible for most of the finite-sample bias in the LSDV estimator. We use the GMM estimator as the first-step–consistent estimate. Assuming strict exogeneity of x_{it}, we have $T(T-1)/2 + T(T-1)$ moment conditions for *gmm*, that is, $E[y_{i,t-s}\Delta\varepsilon_{it}] = 0$ $(t = 2, \dots, T;$ $s = 2, \dots, t)$ and $E[x_{is}\Delta\varepsilon_{it}] = 0$ $(t = 2, \dots, T; s = 1, \dots, T)$. We do not exploit additional moment conditions due to imposing homoscedasticity, because Ahn and Schmidt (1995) noted that efficiency gains are small. Regarding the strict exogeneity of x_{it} to economize on the number of moment conditions, we also experimented with a GMM estimator using $E[y_{i,t-s}\Delta\varepsilon_{it}] = 0$ $(t = 2, \dots, T; s = 2, \dots, t)$ and $E[\mathbf{x}_i'\Delta\varepsilon_i] = \mathbf{0}$, and hence $T(T-1)/2 + 1$ moment conditions. However, this resulted in lower efficiency, that is, higher root mean squared error (RMSE). Under the assumptions made in Section 2, the GMM estimator is consistent for finite T and large N, and hence it is a reasonable benchmark for evaluating the corrected LSDV variants.

We generated data for y according to (1) with $\eta_i \sim \mathcal{IIN}[0, \sigma_\eta^2]$ and $\varepsilon_{it} \sim \mathcal{IIN}[0, \sigma_{\boldsymbol{\varepsilon}}^2]$. The generating equation for the explanatory variable x is

$$x_{it} = \rho x_{i,t-1} + \xi_{it}, \qquad i = 1, \dots, N; t = 1, \dots, T, \qquad (49)$$

where $\xi_{it} \sim \mathcal{IIN}[0, \sigma_\xi^2]$. We used three different research designs. In the first design we chose $\beta = 1$, $\rho = .8$, and $\sigma_\eta = \sigma_{\boldsymbol{\varepsilon}} = \sigma_\xi = 1$. We used two different values for γ, .4 and .8. We assumed that the panel dataset comprises 600 observations and conducted experiments for several combinations of T and N for which $NT = 600$. The second design was equal to the first design, except that we allowed for time series heteroscedasticity in the general error term ε_{it}. (We also experimented with cross-sectional heteroscedasticity in the general error term ε_{it}, but the results were qualitatively not very different from those obtained in the first design.) In this design, $\sigma_{\boldsymbol{\varepsilon}}^2$ is varying over time; that is, we specify $\sigma_{\boldsymbol{\varepsilon},t}^2 = .95 - .05T + .1t$. This specification ensures that $\frac{1}{T}\sum_{t=1}^{T}\sigma_{\boldsymbol{\varepsilon},t}^2 = 1$; hence a proper comparison can be made with the simulation results in case of homoscedasticity. The third research design has identical parameter settings to the design used by Kiviet (1995, table 1). In all of Kiviet's experiments the long-run effect $\beta/(1-\gamma)$ of $\mathbf{x}$ on $\mathbf{y}$ is set equal to unity, and hence the impact multiplier β varies with the chosen values for γ. Homoscedasticity is assumed, and the value of $\sigma_{\boldsymbol{\varepsilon}}^2$ is set equal to 1, but the values of the variances σ_η^2 and σ_ξ^2 differ across experiments. By varying σ_η^2, the relative impact on $\mathbf{y}$ of the two error components η and $\boldsymbol{\varepsilon}$ is changed, whereas the parameter σ_ξ^2 determines the signal-to-noise ratio of the model (for details, see Kiviet 1995). For each experiment, we performed 10,000 Monte Carlo replications.

Selected simulation results for the first, second, and third designs are presented in Tables 2–4. Regarding coefficient estimators, these tables present in the bias in estimating γ and β together with the RMSE. In calculating the RMSE of coefficient estimators, we use the variance as estimated from the Monte Carlo as a measure of true variance. Next to bias and RMSE, we report actual size of (two-sided) simple t-tests of the parameters γ or β to be equal to the values chosen in the respective designs. The nominal size is 5% for each research design. Actual size is calculated as the percentage of replications for which the ratio of coefficient estimator and its standard deviation estimator is larger than 1.96 in absolute value. Regarding the variance estimators used to calculate t ratios, for *lsdv* and *ac* we use the standard variance expression, $\hat{\sigma}_{\boldsymbol{\varepsilon}}^2(\mathbf{W}'\mathbf{AW})^{-1}$. For *bc*, we use the expression in (33) (unreported simulation

Table 2. Homoscedasticity, $\gamma = \rho = .8$ and $\beta = 1$

(N, T)	(300, 2)	(200, 3)	(150, 4)	(100, 6)	(60, 10)	(40, 15)
bias γ						
lsdv	−.365	−.214	−.143	−.079	−.038	−.021
ac	.157	.089	.059	.031	.013	.007
bc	.005	.000	.001	.000	−.001	−.000
gmm	−.002	−.008	−.009	−.011	−.014	−.016
RMSE γ						
lsdv	.368	.217	.146	.082	.041	.025
ac	.172	.100	.068	.039	.021	.015
bc	.073	.044	.033	.023	.016	.013
gmm	.072	.046	.036	.027	.022	.021
bias β						
lsdv	−.101	−.030	−.003	.014	.021	.019
ac	.044	.015	.003	−.007	−.007	−.006
bc	.002	.002	.001	−.001	.001	.000
gmm	−.000	.000	.001	.001	.008	.013
RMSE β						
lsdv	.124	.066	.052	.046	.044	.039
ac	.099	.065	.053	.045	.039	.035
bc	.082	.060	.052	.044	.038	.034
gmm	.081	.060	.052	.044	.039	.037
% actual size γ (nominal is 5%)						
lsdv	100.0	100.0	100.0	97.5	71.3	40.0
ac	78.5	66.9	54.7	34.6	16.7	10.1
bc	2.3	3.0	4.1	4.4	4.8	5.3
gmm	5.7	6.6	7.3	8.6	14.0	23.3
% actual size β (nominal is 5%)						
lsdv	29.9	9.2	5.9	6.9	9.3	8.6
ac	8.6	5.9	5.1	5.3	5.5	5.4
bc	5.0	5.3	5.2	5.2	5.3	5.1
gmm	5.6	5.6	5.7	5.6	6.5	7.9

NOTE: For the variances, we assume that $\sigma_{\boldsymbol{\varepsilon}}^2 = \sigma_\eta^2 = \sigma_\xi^2 = 1$.

Table 3. Time Series Heteroscedasticity, $\gamma = \rho = .8$ and $\beta = 1$

(N, T)	(300, 2)	(200, 3)	(150, 4)	(100, 6)	(60, 10)	(40, 15)
bias γ						
lsdv	−.353	−.203	−.133	−.072	−.033	−.018
ac	.178	.104	.071	.040	.019	.011
bc	.035	.020	.015	.010	.005	.003
gmm	−.002	−.008	−.009	−.010	−.013	−.014
RMSE γ						
lsdv	.356	.206	.136	.075	.036	.022
ac	.192	.114	.079	.046	.025	.017
bc	.084	.050	.037	.025	.017	.013
gmm	.072	.046	.036	.026	.021	.020
bias β						
lsdv	−.098	−.029	−.003	.013	.018	.015
ac	.050	.017	.003	−.008	−.010	−.009
bc	.010	.005	.001	−.003	−.002	−.003
gmm	−.000	.001	.001	.001	.007	.012
RMSE β						
lsdv	.121	.066	.052	.046	.042	.038
ac	.102	.066	.054	.045	.039	.036
bc	.084	.061	.052	.044	.038	.034
gmm	.081	.060	.052	.044	.039	.037
% actual size γ (nominal is 5%)						
lsdv	100.0	100.0	99.9	94.2	59.5	29.2
ac	86.1	77.9	68.0	48.5	25.4	15.4
bc	1.1	3.0	4.7	5.8	5.7	6.0
gmm	5.8	6.4	7.0	8.1	12.8	21.1
% actual size β (nominal is 5%)						
lsdv	28.0	8.8	5.9	6.7	8.5	7.5
ac	9.4	6.0	5.1	5.5	6.0	5.9
bc	4.9	5.3	5.2	5.1	5.3	5.0
gmm	5.6	5.6	5.7	5.5	6.7	7.7

NOTE: We assume that $\sigma^2_{\varepsilon,t} = .95 - .05T + .1t$ and $\sigma^2_\eta = \sigma^2_\xi = 1$.

results show that the accuracy of t-tests based on the *bc* estimator depends on the normality assumption needed to derive asymptotic standard errors), whereas for *gmm*, we exploit the so-called one-step estimates.

Regarding the first design, Table 2 presents the results for $\gamma = .8$. The results for $\gamma = .4$ are similar and hence are deleted to save space. We observe the following patterns in the simulation results for the coefficient estimators. First, bias in estimating the autoregressive parameter γ is always negative for *lsdv* and *gmm*, whereas positive bias has been found for *ac*. Second, for (bias-corrected) LSDV, the bias in estimating both γ and β decreases for larger T (and smaller N), but not so for *gmm*. This is to be expected because *gmm* should perform well, especially for T small and N large. Third, especially for γ, bias in *gmm* carries over to bias in *ac*, demonstrating the dependence of additive bias correction on preliminary consistent estimators. Fourth, in estimating both γ and β, *bc* is virtually unbiased. Finally, based on a MSE criterion, *bc* is almost never beaten by the other coefficient estimators. Regarding simple t-tests for *bc*, we observe that the actual size is close to the nominal size in most cases (except for γ in the case of small T, when the actual size is somewhat low), indicating the accuracy of the asymptotic approximation in this design.

Table 3 presents simulation results for the second design with time series heteroscedasticity. Again, we show the results for $\gamma = .8$ only. In general, results for bias-corrected estimators (*ac* and *bc*) are worse here than in the case of homoscedasticity. This is not surprising, because bias-corrected estimators are not consistent in cases of time series heteroscedasticity. The additive bias-corrected estimator *ac* is especially vulnerable to the presence of heteroscedasticity, but the detrimental effects on *bc* seem modest. Based on an MSE criterion, *bc* is now sometimes beaten by *gmm*, especially for smaller T. The actual size for *bc* is still quite close to the nominal size of 5% (except for testing γ when $T = 2$).

Finally, we turn to simulation results using the third design, that is, the parameterizations used by Kiviet (1995). Table 4 presents the simulation results for a selection of parameterizations. The first part of the table gives the parameterizations used. We need to make several points before discussing the simulation results. First, the relative impact on y_{it} of the two error components η_i and ε_{it} is 1 in experiments I–VIII, but increases to 5 in experiments IX and X. Hence the individual-specific effect is relatively strong in experiments IX and X. Second, the signal-to-noise ratio corresponds to an expected R^2 of 2/3 in all experiments except VIII and X, where it increases to 8/9.

Regarding the third design, we observe the following patterns for the coefficient estimators. First, except for *ac*, bias in estimating the autoregressive parameter γ increases with γ for all estimators. Especially for larger values of γ, substantial coefficient bias is found for *gmm*. Second, there is no one estimation method with the lowest RMSE across all parameterizations. The bias-corrected estimator performs well in all of the designs except designs III and VI, in which there is a relatively high value for γ combined with a relatively low signal-to-noise ratio. The bias-corrected estimator fails to converge in about 40% of the replications in these two designs. We decided to skip such replications completely for each of the estimators. In all other experiments, we found very limited or no convergence problems. We also simulated designs III and VI with $N = 1{,}000$ and found much less convergence problems (around 5%), indicating that this is a small sample issue. Regarding simple t-tests in the third design, we observe that for *bc*, again actual size is quite close to nominal size, whereas coefficient bias of other estimators clearly carries over to the accuracy of t-tests.

Summarizing, regarding coefficient estimators and simple t-tests, we find large bias for *lsdv*, moderate bias for *ac* and *gmm*, and little bias for *bc*. In addition, based on an RMSE criterion, the bias-corrected estimator performs comparatively well for a range of parameter combinations. However, the Monte Carlo results do not suggest that one estimation technique is superior for all parameter combinations. Hence in empirical applications, it may be advisable to compare results using different (consistent) estimation techniques.

6. EMPIRICAL APPLICATION: UNEMPLOYMENT DYNAMICS AT THE U.S. STATE LEVEL

In this section we apply the bias-corrected estimation procedure (denoted by *bc*) to a model of unemployment dynamics at the U.S. state level. We compare the coefficient estimates and estimated standard errors with those of the LSDV, additive bias-corrected LSDV, and GMM estimators (denoted by *lsdv*, *ac*, and *gmm*); see the previous section for more details. As a benchmark, we included also the pooled OLS estimator (*ols*) in the empirical analysis.

The model relates the current unemployment rate (U_{it}) to the unemployment rate and economic growth rate (G_{it}) in the previous year. The model includes individual-specific and time effects (η_i and λ_t),

$$U_{it} = \gamma U_{i,t-1} + \beta G_{i,t-1} + \eta_i + \lambda_t + \varepsilon_{it}. \qquad (50)$$

Table 4. Simulation Results for Selected Designs From Kiviet (1995)

Design nr.	*I*	*II*	*III*	*IV*	*V*	*VI*	*VII*	*VIII*	*IX*	*X*
T	6	6	6	6	6	6	3	3	3	3
γ	0	.4	.8	0	.4	.8	.4	.4	.4	.4
ρ	.8	.8	.8	.99	.99	.99	.8	.8	.8	.8
σ_η	1	.6	.2	1	.6	.2	.6	.6	3	3
σ_ξ	.85	.88	.4	.2	.19	.07	.88	1.84	.88	1.84
bias γ										
lsdv	−.104	−.175	−.366	−.163	−.247	−.375	−.381	−.215	−.381	−.215
ac	.040	.060	.007	.058	.069	−.005	.141	.088	.133	.087
bc	−.001	−.001	−.037	−.001	−.001	−.042	.005	.000	.005	.000
gmm	−.017	−.035	−.179	−.034	−.069	−.197	−.047	−.017	−.067	−.019
RMSE γ										
lsdv	.109	.179	.368	.167	.251	.377	.386	.221	.386	.221
ac	.055	.076	.066	.077	.091	.067	.173	.110	.169	.109
bc	.037	.045	.077	.049	.060	.080	.109	.064	.109	.064
gmm	.044	.061	.200	.067	.098	.217	.116	.066	.145	.073
bias β										
lsdv	.044	.045	.015	.085	.095	.049	.013	.008	.013	.008
ac	−.019	−.017	−.001	−.037	−.033	−.007	−.005	−.003	−.004	−.003
bc	−.001	−.001	.001	−.004	−.005	−.001	−.000	.000	−.000	.000
gmm	.006	.008	.007	.014	.023	.022	.002	.001	.003	.001
RMSE β										
lsdv	.069	.068	.116	.238	.257	.678	.092	.046	.092	.046
ac	.057	.053	.109	.211	.221	.622	.102	.048	.101	.048
bc	.053	.050	.109	.211	.221	.619	.096	.046	.096	.046
gmm	.054	.051	.110	.215	.227	.635	.095	.046	.095	.046
% actual size γ (nominal is 5%)										
lsdv	85.3	99.8	100.0	97.1	100.0	100.0	100.0	99.0	100.0	99.0
ac	21.9	36.7	20.1	27.6	35.8	20.5	48.3	38.4	45.1	37.9
bc	4.8	4.0	5.6	4.3	3.1	6.1	2.8	3.3	2.8	3.3
gmm	7.6	10.8	51.1	9.7	17.1	54.8	9.5	7.2	10.7	7.5
% actual size β (nominal is 5%)										
lsdv	13.7	16.7	9.1	9.0	11.0	10.2	6.6	6.4	6.6	6.4
ac	6.4	6.2	5.2	4.7	4.6	5.4	5.0	5.2	5.0	5.2
bc	5.1	5.1	5.2	5.0	5.0	5.2	5.1	5.3	5.1	5.3
gmm	5.8	5.8	6.0	5.7	5.8	6.3	5.8	5.9	5.8	5.9

NOTE: We assume that $\sigma_\varepsilon^2 = 1$, $N = 100$, and $\beta = 1 - \gamma$ in all experiments.

Equation (50) can be rewritten in an easier-to-interpret, from,

$$\Delta U_{it} = (\gamma - 1)(U_{i,t-1} - \alpha_i) + \beta(G_{i,t-1} - \delta) + \lambda_t + \varepsilon_{it}, \qquad (51)$$

where $(1 - \gamma)\alpha_i - \beta\delta = \eta_i$. Equation (51) indicates that the change in the unemployment rate is determined by an adjustment of the unemployment rate toward a "natural" or "equilibrium" rate of unemployment, α_i, which may differ across the states, and by the previous economic growth rate. The speed of adjustment of the unemployment rate toward the "natural" or "equilibrium" rate is equal to $1 - \gamma$. Partial adjustment, $0 < \gamma < 1$, is expected. A state that has relatively high economic growth is more likely to have reduced unemployment rates compared with states in which the economy is growing more slowly. This would imply that $\beta < 0$.

The data for the unemployment rate for the 1991–2000 period are obtained from the U.S. Bureau of Labor Statistics, and data for the (current dollar) gross state product are obtained from the U.S. Bureau of Economic Analysis. The economic growth rate is taken to be the relative growth of the gross state product. Data are available for all U.S. states and Washington, DC ($N = 51$). The number of time periods in estimation is $T = 9$, because the year 1991 is taken as the starting observation.

Table 5 presents the various coefficient estimates and their estimated standard deviations. The value of the LSDV estimate of γ is .484, which would imply an adjustment rate of around 50% per year. In contrast, the bias-corrected estimate (*bc*) is equal to .615, which implies an adjustment rate of less than 40%. Hence the speed of adjustment toward a "natural rate of unemployment" is not as large as the original LSDV estimator would suggest. The value of the LSDV estimate of β equals −.064, whereas the value of the bias-corrected estimate is −.057. This implies a somewhat smaller effect of economic growth on the change in unemployment than indicated by the traditional within estimate. The results for the additive bias-corrected estimator (*ac*) are somewhat different from those of the bias-corrected estimator introduced in this article. However, the results for the GMM estimator (*gmm*) are more or less equal to that of *bc*.

A restrictive assumption of bias-corrected LSDV estimators is that consistency depends on strict exogeneity of the lagged growth rate, $G_{i,t-1}$. Because we have assumed strict exogeneity

Table 5. Empirical Results for the Unemployment-Growth Model

	ols	*lsdv*	*ac*	*bc*	*gmm*
$\hat{\gamma}$	.840	.484	.763	.615	.600
sd($\hat{\gamma}$)	.022	.037	.040	.047	.048
$\hat{\beta}$	−.041	−.064	−.049	−.057	−.057
sd($\hat{\beta}$)	.008	.012	.013	.012	.013

NOTE: $T = 9$ and $N = 51$; time dummies are included.

of $G_{i,t-1}$ in GMM estimation, we can test against exogeneity using the Sargan test. To increase power, we do not use all moment conditions, only $E[U_{i,t-s}\Delta\varepsilon_{it}] = 0$ $(t = 2, \ldots, T; s = 2, 3)$ and $E[G_{i,t-1}\Delta\varepsilon_{it}] = 0$ $(t = 2, \ldots, T)$. The test has a value of 25.09 (p value .29), and hence the validity of the moment conditions is not rejected. We conclude that the problem of $G_{i,t-1}$ being only weakly exogenous is not an issue in this particular application.

7. EXTENSIONS AND CONCLUDING REMARKS

In this article we have developed a new bias-corrected estimator for dynamic panel data models. The proposed estimator has desirable asymptotic properties for finite T and large N, but these have been derived under certain restrictive assumptions, including strict exogeneity of regressors in x_{it}, homoscedasticity of the disturbances, and balanced panels. In this final section we discuss the limitations and possible extensions of our approach with respect to each of these three assumptions.

First, regarding the exogeneity assumption, some regressors in x_{it} could be predetermined as well. Inconsistencies originating from this source are not accounted for in the current bias corrections. It can be shown that the order of magnitude of such inconsistency terms equals that of lagged dependent variable regressors, that is, of order $O(T^{-1})$. But addressing the importance of this source of bias requires full specification of the marginal process of the regressors x_{it}, which is a major complication in practice. Simulation evidence for the dynamic panel data model with predetermined or endogenous regressors x_{it} has been given by Bun and Kiviet (2002b) and Blundell, Bond, and Windmeijer (2000). In general, these simulation results show that lack of strict exogeneity of x_{it} does influence the finite-sample properties of estimators, and hence it is expected that in practice estimators will be affected as well. Note, however, that in the current application on unemployment dynamics, strict exogeneity of the additional regressor (lagged growth rate) is not rejected. Second, regarding homoscedasticity of the disturbances, we have provided some simulation results allowing for either cross-section or time series heteroscedasticity. From the simulation results, we see that in the latter case, bias-corrected estimators behave somewhat worse, as expected.

Finally, the proposed method in this study can be extended to unbalanced panels. In this case not all time observations are available for each individual i. That is, the data may be observed for certain individuals i only from a certain date, or the data may be observed for other individuals only up until a certain date. This implies that the starting date and ending date of the data are individual-specific. Denoting the beginning of the data period by B_i and the final time period of observation by T_i, we have $1 \leq B_i \leq T_i \leq T$. We then order the individuals in terms of the length of the time period, $T_i - B_i + 1$. The largest value for this length of time period is T, and the smallest value is 2. Denote by φ_t the fraction of observations with period of time length $t = 2, \ldots, T$; that is, $\sum_{t=2}^{T} \varphi_t = 1$. Then we replace the function $h(\gamma)$ in Sections 2 and 3 with

$$h_u(\gamma) = \sum_{t=2}^{T} \varphi_t \frac{(t-1) - t\gamma + \gamma^t}{t(t-1)(1-\gamma)^2}, \tag{52}$$

and likewise derive expressions for the limiting distribution of the estimator. Note that we do not take possible sample selection issues into account in this way. Research into problems of sample selection in dynamic panel data models has started only recently, with Kyriazidou (2001) providing a first contribution in this area.

Given the assumptions, the bias-corrected estimator performs well when T is small and N is large. Simulation results on various designs show that based on an RMSE criterion, bias-corrected LSDV estimators perform well against GMM estimators. In cases where both T and N are small, the limiting distributions for the estimators may have little to say about the actual distribution (especially when γ is close to unity). However, given the strong (relative) performance of the bias-corrected estimator in the Monte Carlo exercises in cases where T is as small as 2 or 3, this estimator appears suitable for research efforts with samples with large numbers of individuals/firms and a (very) small number of time periods. Many datasets, especially those in which data are collected yearly, have these panel dimensions.

New estimators for the dynamic panel data model have recently been introduced. Each of these estimators has advantages and disadvantages, and it is not clear that any of them would uniformly outperform the bias-corrected estimator. Hahn and Kuersteiner (2002) introduced an estimator that requires that the number of time periods be at least moderate. They also paid the most attention to the case of no exogenous variables. Hsiao, Pesaran, and Tahmiscioglu (2002) introduced a maximum likelihood estimator based on first differencing the dynamic panel data model to get rid of the unobserved individual effects. Methods based on first differencing are conceptually different from methods based on removing unobserved effects by subtracting the individual-specific means. One potential source of distinction between methods based on either first difference or within transformations is the influence of measurement errors. Mairesse, Hall, and Mulkay (1999), for example, argued that biases from random measurement errors are more severe in cases of first-differenced estimates than in cases of within estimates. Alvarez and Arellano (2003) introduced a random-effects maximum likelihood estimator, but did not consider the case of exogenous variables included, and assumed in deriving the asymptotic distribution that both N and T tend to infinity. Finally, Lancaster (2002) took a Bayesian approach to dynamic panel data models, finding a relatively simple set of first-order conditions for the maximum of the posterior. However, Lancaster's work still has some unresolved issues concerning priors, and its inference may not be completely comparable to the classic inference used in the present article. Nevertheless, research into these and other newly developed estimators for dynamic panel data models remains a very vivid and important area for both theoreticians and practitioners.

ACKNOWLEDGMENTS

The authors thank the Royal Academy of Arts and Sciences (KNAW) for financial support. The article was strongly improved by the comments of the associate editor and referees; the authors thank them for their careful reading. The authors are also grateful to Boris Lokshin for reporting some inaccuracies in an earlier version of the manuscript.

[Received November 2001. Revised May 2004.]

REFERENCES

Ahn, S. C., and Schmidt, P. (1995), "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, 68, 29–52.

Alvarez, J., and Arellano, M. (2003), "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," *Econometrica*, 71, 1121–1159.

Anderson, T. W., and Hsiao, C. (1981), "Estimation of Dynamic Models With Error Components," *Journal of the American Statistical Association*, 76, 598–606.

——— (1982), "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47–82.

Arellano, M., and Bond, S. R. (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277–297.

Arellano, M., and Bover, O. (1995), "Another Look at the Instrumental Variable Estimator of Error-Components Models," *Journal of Econometrics*, 68, 29–52.

Baltagi, B. H., Griffin, J. M., and Xiong, W. (2000), "To Pool or Not to Pool: Homogeneous versus Heterogeneous Estimators Applied to Cigarette Demand," *Review of Economics and Statistics*, 82, 117–126.

Blundell, R., and Bond, S. (1998), "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115–143.

Blundell, R., Bond, S., and Windmeijer, F. (2000), "Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimators," in *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, ed. B. H. Baltagi, Amsterdam: Elsevier Science, pp. 53–91.

Bun, M. J. G., and Kiviet, J. F. (2001), "The Accuracy of Inference in Small Samples of Dynamic Panel Data Models," Discussion Paper 2001-006/4, Tinbergen Institute, Amsterdam.

——— (2002a), "On the Diminishing Returns of Higher-Order Terms in Asymptotic Expansions of Bias," *Economics Letters*, 79, 145–152.

——— (2002b), "The Effects of Dynamic Feedbacks on LS and MM Estimator Accuracy in Panel Data Models," UvA-Econometrics Discussion Paper 2002/05, University of Amsterdam.

Carree, M. A. (2002), "Nearly Unbiased Estimation in Dynamic Panel Data Models," Discussion Paper 2002-008/2, Tinbergen Institute, Rotterdam.

Hahn, J. (1999), "How Informative Is the Initial Condition in the Dynamic Panel Model With Fixed Effects?" *Journal of Econometrics*, 93, 309–326.

Hahn, J., and Kuersteiner, G. (2002), "Asymptotically Unbiased Inference for a Dynamic Panel Model With Fixed Effects When Both n and T Are Large," *Econometrica*, 70, 1639–1657.

Hansen, G. (2001), "A Bias-Corrected Least Squares Estimator of Dynamic Panel Models," *Allgemeines Statistisches Archiv*, 85, 127–140.

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

Hsiao, C., Pesaran, M. H., and Tahmiscioglu, A. K. (2002), "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods," *Journal of Econometrics*, 109, 107–150.

Judson, R. A., and Owen, A. L. (1999), "Estimating Dynamic Panel Data Models: A Guide for Macroeconomists," *Economics Letters*, 65, 9–15.

Kitazawa, Y. (2001), "Exponential Regression of Dynamic Panel Data Models," *Economics Letters*, 73, 7–13.

Kyriazidou, E. (2001), "Estimation of Dynamic Panel Data Sample Selection Models," *Review of Economic Studies*, 68, 543–572.

Kiviet, J. F. (1995), "On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models," *Journal of Econometrics*, 68, 53–78.

Lancaster, T. (2002), "Orthogonal Parameters and Panel Data," *Review of Economic Studies*, 69, 647–666.

MacKinnon, J. G., and Smith, A. A., Jr. (1998), "Approximate Bias Correction in Econometrics," *Journal of Econometrics*, 85, 205–230.

Mairesse, J., Hall, B. H., and Mulkay, B. (1999), "Firm-Level Investment in France and the United States: An Exploration of What We Have Learned in Twenty Years," *Annales d'Économie et de Statistique*, 55/56, 27–67.

Nickell, S. (1981), "Biases in Dynamic Models With Fixed Effects," *Econometrica*, 49, 1417–1426.

Wansbeek, T., and Bekker, P. (1996), "On IV, GMM and ML in a Dynamic Panel Data Model," *Economics Letters*, 51, 145–152.

Ziliak, J. P. (1997), "Efficient Estimation With Panel Data When Instruments Are Predetermined: An Empirical Comparison of Moment-Condition Estimators," *Journal of Business & Economic Statistics*, 15, 419–431.