TI 2002-007/2
Tinbergen Institute Discussion Paper

Nearly Unbiased Estimation in Dynamic Panel Data Models with Exogenous Variables

Martin A. Carree

Faculty of Economics, Erasmus University Rotterdam,Tinbergen Institute, and
Faculty of Economics and Business Administration, Maastricht University

Nearly Unbiased Estimation in Dynamic Panel Data Models with Exogenous Variables

Martin A. Carree

Faculty of Economics
Erasmus University Rotterdam
and
Faculty of Economics and Business Administration
Maastricht University

Abstract

This paper introduces a new estimator for the fixed effects dynamic panel data model with exogenous variables. This estimator does not share some of the drawbacks of recently developed IV and GMM estimators and has a good performance even in small samples. The nearly unbiased estimator is derived as a bias correction of the within estimator (least squares dummy variable estimator). The estimator is applied to a model of unemployment dynamics at the U.S. state level for the 1991-2000 period.

Version October 2001

Keywords: panel data, fixed effects, nearly unbiased estimation

JEL-code: C23

email: carree@few.eur.nl or m.carree@mw.unimaas.nl

Correspondence to:

Martin A. Carree
MW-ORG
FdEWB
Maastricht University
P.O. Box 616
6200 MD Maastricht
The Netherlands

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 estimators or generalized method of moments 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, Wansbeek and Bekker, 1996, Ziliak, 1997, Blundell and Bond, 1998, Hahn, 1999 and Judson and Owen, 1999). The development of such new estimators was necessary because the traditional within estimator (least-squares dummy variable estimator) is inconsistent for fixed T. Despite the increasing sophistication of the IV and GMM estimators, two important drawbacks remain. First, the complexity of the new estimators is a barrier for applied researchers (see e.g. Baltagi et al., 2000). This should partly be a temporary drawback as the new estimators will be incorporated in the statistical packages. However, the newly developed estimators may require additional decisions on, for example, which instruments to use. This makes application less straightforward. Second, the new estimators introduce problems of their own. For example, the GMM-estimators suffer from an important upward bias in case the autoregressive parameter becomes close to one (see Blundell and Bond, 1998, Kitazawa, 2001). Furthermore, the performance of these estimators depends strongly upon the ratio of variance of the fixed effects and the variance of the error term (see e.g. Kitazawa, 2001).

This paper introduces a new and simple estimator for dynamic panel data models. It is computed as a bias correction to the within estimator (least squares dummy variable estimator) and is, as such, related to the bias-corrected estimator developed by Kiviet (1995).[1] The newly developed estimator is computationally simpler than Kiviet's bias-corrected estimator and, in addition, appears to perform well in comparison. MacKinnon and Smith (1998) already indicate that bias of parameter estimates may be virtually eliminated in some common cases, though at the expense of increased variance of the estimators. This paper shows that this is also the case for fixed effects panel data models and that for a wide range of parameter values the mean squared error of the nearly unbiased estimator is far less than that of the traditional within estimator. An applied researcher can compute the estimates using a very basic statistical package (with regression analysis). As an example the estimator is applied to the intertemporal dynamics of the unemployment rate in U.S. states in the 1991-2000 period. Another advantage of the estimator is that it provides an intuitive decomposition of the bias of the within estimator in terms of (i) the correlation between the lagged endogenous variable and the error term and (ii) the (multiple) correlation between the lagged endogenous variable and the exogenous variable(s).

[1] Judson and Owen (1999) present Monte Carlo simulation results indicating that Kiviet's bias-corrected estimator outperforms the Anderson-Hsiao IV estimator and the Arellano-Bond GMM estimators. The current paper provides further support for the usefulness of Kiviet's central idea of bias-correcting the least squares dummy variable estimator.

The rest of this paper is as follows. In section 2 the nearly unbiased esimators are derived. In section 3 Monte Carlo exercises are performed. In section 4 the estimators are applied to a simple model of intertemporal dynamics of the unemployment rate. Section 5 concludes.

2. Nearly unbiased estimation in dynamic panel data models

Consider the dynamic panel data model with an exogenous variable x_{it} :

(1) $y_{it} = g y_{i,t-1} + b x_{it} + h_i + u_{it}$ for $i = 1,\ldots,N$ and $t = 1,\ldots,T$

In this equation the h_i are the fixed effects. We assume that $|g| < 1$ and that the u_{it} are i.d.d. with mean zero and variance s_u^2. In order to derive the least squares dummy variable (LSDV) estimator 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 $\tilde{u}_{it} = u_{it} - \bar{u}_i$. Equation (1) can then be rewritten as

(2) $\tilde{y}_{it} = g\tilde{y}_{i,t-1} + b\tilde{x}_{it} + \tilde{u}_{it}$

The LSDV-estimator is computed by applying ordinary least squares to equation (2):

(3) $$\hat{g} = \frac{\sum\sum \tilde{x}_{it}^2 \sum\sum \tilde{y}_{i,t-1}\tilde{y}_{it} - \sum\sum \tilde{x}_{it}\tilde{y}_{i,t-1} \sum\sum \tilde{x}_{it}\tilde{y}_{it}}{\sum\sum \tilde{x}_{it}^2 \sum\sum \tilde{y}_{i,t-1}^2 - \left(\sum\sum \tilde{x}_{it}\tilde{y}_{i,t-1}\right)^2}$$

This estimator is biased because there is correlation between $\tilde{y}_{i,t-1}$ and $\tilde{u}_{it}$ (see e.g. Hsiao, 1986, p.74). The bias does not disappear by having *N* go to infinity and the LSDV estimator is therefore inconsistent for fixed *T*. The extent of the bias can be computed as follows. We rewrite equation (3) as:

(4) $$\hat{g} = g + \frac{\sum\sum \tilde{x}_{it}^2 \sum\sum \tilde{y}_{i,t-1}\tilde{u}_{it} - \sum\sum \tilde{x}_{it}\tilde{y}_{i,t-1} \sum\sum \tilde{x}_{it}\tilde{u}_{it}}{\sum\sum \tilde{x}_{it}^2 \sum\sum \tilde{y}_{i,t-1}^2 - \left(\sum\sum \tilde{x}_{it}\tilde{y}_{i,t-1}\right)^2}$$

From equation (1) we use continuous substitution to obtain:

(5) $y_{it} = g^{t} y_{i0} + b\left(x_{it} + g x_{i,t-1} + \ldots + g^{t-1} x_{i1}\right) + \frac{1-g^{t}}{1-g} h_i + u_{it} + g u_{i,t-1} + \ldots + g^{t-1} u_{i1}$

Summing $y_{i,t-1}$ over time gives:

(6) $\sum_{t=1}^{T} y_{i,t-1} = \frac{1-g^{T}}{1-g} y_{i0} + b\left(x_{i,T-1} + \ldots + \frac{1-g^{T-1}}{1-g} x_{i1} \right) + \frac{(T-1) - Tg + g^{T}}{(1-g)^2} h_i +$

$$+ u_{i,T-1} + \ldots + \frac{1-g^{T-1}}{1-g} u_{i1}$$

From this it can be derived that

(7a) $\underset{N\to\infty}{p\lim} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{y}_{i,t-1} \tilde{u}_{it} / NT = \underset{N\to\infty}{p\lim} \sum_{i=1}^{N} \bar{y}_{i,-1} \bar{u}_i / N = -s_u^2 \frac{(T-1) - Tg + g^{T}}{T^2 (1-g)^2} = -s_u^2 f(g, T).$

The problem of subtracting the cross-section specific mean does not extend to the exogenous variables, and

(7b) $\underset{N\to\infty}{p\lim} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}_{it} \tilde{u}_{it} / NT = 0$

Having *N* tending to infinity and using equation (7b) we have that equation (4) becomes:

(8) $\underset{N\to\infty}{p\lim}(\hat{g} - g) = \underset{N\to\infty}{p\lim} \frac{\sum\sum \tilde{y}_{i,t-1} \tilde{u}_{it} / \sum\sum \tilde{y}_{i,t-1}^2}{1 - \left(\sum\sum \tilde{x}_{it} \tilde{y}_{i,t-1}\right)^2 / \left(\sum\sum \tilde{x}_{it}^2 \sum\sum \tilde{y}_{i,t-1}^2\right)}$

Let us now introduce the overall variance of $\tilde{y}_{i,t-1}$, $s_{y_{-1}}^2$, and the correlation of $\tilde{y}_{i,t-1}$ and $\tilde{x}_{it}$, $r_{xy_{-1}}$.

(9a) $s_{y_{-1}}^2 = \underset{N\to\infty}{p\lim} \sum\sum \tilde{y}_{i,t-1}^2 / NT$

(9b) $r^2_{xy_{-1}} = \underset{N\to\infty}{p\,lim}\left(\sum\sum \tilde{x}_{it}\tilde{y}_{i,t-1}\right)^2 / \left(\sum\sum \tilde{x}^2_{it}\sum\sum \tilde{y}^2_{i,t-1}\right)$

Inserting these variables into equation (8) gives our final expression for the bias of the LSDV estimator:

(10) $\underset{N\to\infty}{p\,lim}(\hat{g} - g) = \dfrac{-s^2_u f(g,T)}{(1-r^2_{xy_{-1}})s^2_{y_{-1}}} = -g(s^2_u, s^2_{y_{-1}}, r_{xy_{-1}})f(g,T)$

In this expression both the functions f and g are positive. It implies that the LSDV estimator is downward biased. The extent of the bias depends upon five parameters, viz. g, T, s^2_u, $s^2_{y_{-1}}$ and $r_{xy_{-1}}$. The bias of the LSDV estimator is especially severe in case (i) the value of g is high; (ii) the number of time periods is low; (iii) the ratio of variances $s^2_u / s^2_{y_{-1}}$ is high; (iv) the lagged endogenous variable and the exogenous variable are highly correlated. Equation (10) will be used to construct a bias-corrected estimator by assuming that N in an actual panel is large enough for the bias to be close to $-g(s^2_u, s^2_{y_{-1}}, r_{xy_{-1}})f(g,T)$. The estimator is constructed by solving equation (10) for g as a function of the LSDV-estimator $\hat{g}$ and the function *g*.

The function $f(g,T)$ is equal to $1/4$ for $T = 2$, to $(2+g)/9$ for $T = 3$ and $(3+2g+g^2)/16$ for $T = 4$. For values of $T > 4$ equation (10) cannot be directly used to obtain a simple bias-corrected estimator because the function f is non-linear. However, the function f can be very well approximated by the following function:

(11) $f(g,T) = \dfrac{(T-1) - Tg + g^T}{T^2(1-g)^2} \approx a_T + b_T g + \dfrac{c_T}{d_T - g}$ for $T > 3$

Values for a_T, b_T, c_T and d_T are given in Table 1. These values are computed by taking values for g from 0.000 to 0.999 with step size 0.001 and to compute the value of the function *f*.[2] The value of this function is then taken as dependent variable in a non-linear least squares regression to obtain estimates for the four parameters. The fit of this ‘regression equation’ is very good (the percentage of

[2] We assume that $0 \le g < 1$. This is the common case in practice. A negative value of g would imply that, in terms of error-correction models, the one period adjustment towards the ‘equilibrium’ is more than 100%. It is of course possible to extend the approximation function (11) to negative values of g, although somewhat at the expense of fit for the positive interval of g.

variance explained is 99.98% or higher). We use the approximation for T = 4 as well as not to introduce an additional special case. The two special cases of solving g as a function of $\hat{g}$ and $\hat{g}$ from equation (10) are for T = 2 and T = 3 (the subscript BC stands for bias-corrected):

$$\text{(12a)}\quad \hat{g}_{BC} = \hat{g} + \hat{g}\,/\,4 \qquad \text{for } T = 2$$

$$\text{(12b)}\quad \hat{g}_{BC} = \frac{9\hat{g} + 2\hat{g}}{9 - \hat{g}} \qquad \text{for } T = 3$$

with $\hat{g}$ as an estimate for the function g. For $T > 3$ we solve equation (10) for g after inserting equation (11) and get the following quadratic equation:

$$\text{(12c)}\quad \hat{g}_{BC}^{2}\left(1 - b_T\,\hat{g}\right) + \hat{g}_{BC}\left(-d_T - \hat{g} - a_T\,\hat{g} + b_T d_T\,\hat{g}\right) + \left(d_T\hat{g} + a_T d_T\,\hat{g} + c_T\,\hat{g}\right) = 0 \qquad \text{for } T > 3$$

Equation (12c) will have either no, one unique or two solutions. In case of two solutions, the solution to choose is the smallest one in value. Let us define

$$\text{(13)}\quad D_T(\hat{g},\hat{g}) = (\,d_T + \hat{g} + (\,a_T - b_T d_T\,)\hat{g}\,)^2 - (\,4 - 4b_T\,\hat{g}\,)(\,d_T\hat{g} + (\,a_T d_T + c_T\,)\hat{g}\,)$$

Using this variable the estimator can be written as follows:

$$\text{(12c')}\quad \hat{g}_{BC} = \frac{d_T + \hat{g} + (\,a_T - b_T d_T\,)\hat{g} - \sqrt{D_T}}{2 - 2b_T\,\hat{g}} \qquad \text{for } T > 3$$

In case of $\hat{g} = 0$, or the absence of any residual in equation (1), we have that in each of the cases (12a), (12b) and (12c'), $\hat{g}_{BC} = \hat{g}$. This is a desirable property of the approximation function (11) which results in the estimation in equation (12c'). That is, in case the LSDV estimates result in a perfect fit, there is no problem of correlation between the error term and the lagged endogenous variables, as relationship (1) is estimated to be without error!

The estimators developed in equations (12a), (12b) and (12c') can be addressed to as *nearly* unbiased estimators. The estimators are not unbiased as in practice N is less than infinity and the function f is not perfectly approximated by equation (11). To achieve nearly unbiased estimates of the parameters of equation (2), the following steps are to be taken:

STEP 1: Compute the LSDV-estimators $\hat{g}$ and $\hat{b}$.

STEP 2: Compute

$$\hat{s}_u^2 = \sum\sum\left(\tilde{y}_{it} - \hat{g}\,\tilde{y}_{i,t-1} - \hat{b}\,\tilde{x}_{it}\right)^2 / N(T-1)$$

$$\hat{s}_{y_{-1}}^2 = \sum\sum\tilde{y}_{i,t-1}^2 / NT$$

$$R_{xy_{-1}}^2 = \left(\sum\sum\tilde{x}_{it}\tilde{y}_{i,t-1}\right)^2 / \left(\sum\sum\tilde{x}_{it}^2\sum\sum\tilde{y}_{i,t-1}^2\right)$$

$$\hat{g} = \frac{\hat{s}_u^2}{(1-R_{xy_{-1}}^2)\hat{s}_{y_{-1}}^2}$$

STEP 3: Use equation (12) to determine the new $\hat{g}_{BC}$, which is the 1-step estimator $\hat{g}_{BC}^{(i)}$, substitute this estimate into equation (2) and determine the least squares estimate $\hat{b}_{BC}$ for b .

STEP 4: Compute

$$\hat{s}_u^2 = \sum\sum\left(\tilde{y}_{it} - \hat{g}_{BC}\,\tilde{y}_{i,t-1} - \hat{b}_{BC}\,\tilde{x}_{it}\right)^2 / N(T-1)$$

$$\hat{g} = \frac{\hat{s}_u^2}{(1-R_{xy_{-1}}^2)\hat{s}_{y_{-1}}^2}$$

STEP 5: Use equation (12) to determine the new $\hat{g}_{BC}$, which is the 2-step estimator $\hat{g}_{BC}^{(ii)}$, substitute this estimate into equation (2) and determine the least squares estimate $\hat{b}_{BC}$ for b .

STEP 6: In case of a sizeable difference between the 1-step and 2-step estimators, repeat steps 4 and 5 until the $\hat{g}_{BC}$ converges.

The reason for the estimation procedure to be iterative is that the intial estimate(s) for s_u^2 are based upon biased estimates of the parameters b and g . In the Monte Carlo exercises we will find that in many cases the 2-step estimator or even the 1-step estimator is already close to being unbiased. The procedure as described will have the (asymptotic) bias of the estimator becoming smaller with each step, or:

(14) $\underset{N\to\infty}{plim}\,\hat{g} \le \underset{N\to\infty}{plim}\,\hat{g}_{BC}^{(i)} \le \underset{N\to\infty}{plim}\,\hat{g}_{BC}^{(ii)} \le \ldots \le \underset{N\to\infty}{plim}\,\hat{g}_{BC}^{lim} \approx g$.

where $\hat{g}_{BC}^{lim}$ is the converged estimate. In case of T = 2 or T = 3 we may replace the "approximately equal to" by an "equal to", because we do not use approximation function (11). In practice N is less than infinite and there are two potential problems with the estimation procedure. First, the estimation procedure may not converge. This may occur in case $\hat{g}$ has a very high value. Second, for values of T > 3 the value of D_T may turn out to be negative (usually in later stages of the convergence procedure). This implies that the estimator in equation (12c') cannot be computed. In the Monte Carlo experiments we will find that these problems occur on a limited scale in samples where both T and N are small and/or the exogenous variable(s) have very little explanatory power (e.g. when b =0).

Lack of convergence of the nearly unbiased estimator would render comparison with other estimators like the LSDV and Kiviet's (1995) bias-corrected estimator difficult. It would of course be unfair to eliminate the cases for which the nearly unbiased estimator was found not to converge. For this reason we introdue a "combined" estimator, $\hat{g}_{BC}^{com}$, which is equal to $\hat{g}_{BC}^{lim}$ when there is convergence and equal to $\hat{g}_{BC}^{(i)}$ when there is no convergence. In none of the simulation experiments a negative value of D_T already in step 3 was found.[3] Hence, it was always possible to calculate a value for $\hat{g}_{BC}^{com}$.

The estimation procedure is simply extended to the case of more than one exogenous variable. It implies that the simple correlation between the lagged endogenous variable and the exogenous variable has to be replaced by the multiple correlation between the endogenous variable and all the exogenous variables combined. It means that the R-squared of the following regression equation can be taken (K is the number of exogenous variables):

$$(15)\quad \tilde{y}_{i,t-1} = l + m_1\tilde{x}_{1it} + ... + m_K\tilde{x}_{Kit} + n_{it}$$

Standard errors for the nearly unbiased estimates can be computed using a bootstrapping procedure. The standard errors will exceed that of the LSDV estimates as is most easily shown for the case of T = 3. Equation (12b) shows two sources of increased standard error. First, the derivate of $\hat{g}_{BC}$ with respect to $\hat{g}$ exceeds one. Second, an additional parameter g has to be estimated causing additional variability. A third source is that the percentage of explained variance is less ($\hat{s}_u^2$ is higher) using the nearly unbiased estimates when compared to the within estimator.

[3] A range of simulation experiments showed that the case of a negative D_T for the 1-step estimator was sometimes found only for cases with very small N and little or no explanatory power of the exogenous variable.

3. Monte Carlo experiments

The Monte Carlo experiments have three main objectives. The first objective is to study the finite (small) sample properties of the estimator. Equation (14) indicates that the estimator is (nearly) consistent, but this property will have little value in case finite (small) sample properties are nevertheless poor. The second objective is to consider the scale at which the problems of non-convergence and/or a negative value of D_T occur. The third oblective is to compare the performance of the estimator with Kiviet's (1995) bias-corrected estimator. The design of the Monte Carlo experiments closely corresponds to that of Arellano and Bond (1991), Kiviet (1995) and Judson and Owen (1999). In the Monte Carlo experiments we assume that the (one) exogenous variable has the following dynamic process:

(16) $x_{it} = r x_{i,t-1} + e_{it}$

The value of r is assumed to be less than unity in absolute value to guarantee stationarity. In the experiments we assume that $e_{it} \sim N(0, s_e^2)$, $u_{it} \sim N(0, s_u^2)$ and $h_i \sim N(0, s_h^2)$. These three random variables are assumed to be independently distributed. In the experiments we will use a range of values for *N*, *T*, r, b and g.

In the *first* experiment we will have assume that there are 600 observations in total. However, the dimensions of the panel range from *T* = 2, and hence *N* = 300, to *T* = 30, and hence *N* = 20. We keep the value of b fixed at 1 and the value of r fixed at 0.8. We consider three values for g : 0.3, 0.7 and 0.9. In the *second* experiment we assume that there are only 60 observations in total, which can be considered a small sample. In the *third* experiment we keep *T* fixed at 6, *N* fixed at 100 and g fixed at 0.7. We consider three different values for b : 0, 1 and 4. We also consider three different values for r : 0.8, 0.99 and –0.99. In the *fourth* experiment the exact same parameter combinations as used by Kiviet (1995, p.67) are replicated. The number of time periods, *T*, is equal to either 3 or 6 and *N* is fixed at 100 in these experiments. Three different values of g are considered: 0.0, 0.4 and 0.8. In addition, two values of r are considered: 0.8 and 0.99. In each of the experiments we first have forty periods of the dynamic processes (1) and (16) to eliminate starting-up problems. The number of replications is 500 for each case.

The results of the first experiment are displayed in Table 2. In the first experiment with 600 observations we find that none of the potential problems of non-convergence or of a negative D_T occur. The simulations indicate that convergence is fast for values of *T* equal to five or higher. In these cases the mean of the estimates across the replications of even the 1-step estimate is within 0.01

distance of the true value. For values of T equal to three or four one additional step is to be recommended: the 2-step estimate has a 0.01 distance or less of the true value. For the case of T equal to two convergence is somewhat slower, but even then the 3-step estimate is within 0.01 distance of the true value. The nearly unbiased estimator strongly outperforms the within estimator in terms of root mean squared error (RMSE) for small values of T. For $T = 2$ the RMSE is less than *one fifth* of that of the within estimator. For $T = 6$ the RMSE is still less than *one third* of that of the within estimator.

The results of the second experiment are given in Table 3. In the second experiment we consider a much smaller sample of 60 observations. This is likely to pose more cases of no convergence or a negative D_T. The largest problems occur for the case of $T = 2$, and hence $N = 30$. But even in this extreme case the large majority of estimates (90% or more) shows convergence. For $T = 3$ the percentage of convergence increases already to 96% or higher. For values of T higher than three problems of non convergence are quite unlikely to occur. It should of course be noted that in other experimental settings the percentage of non-convergence may be higher. Given the small sample size, both the 1-step estimate and the "combined" estimator perform satisfactorily, on average.

The results of the third experiment are presented in Table 4. Nine different combinations of values of b and r are used to examine the sensitivity of the estimator. In case of b equal to 4 the within estimator is already close to the true value and convergence is fast. In case of an absence of the influence of the exogenous variable (b = 0) convergence is slower. This simulation experiment has a large range of values of correlation between the lagged endogenous variable and the exogenous variable (between 0.001 and 0.965). The nearly unbiased estimator appears to perform well for either low or high values of correlation. In the case of b equal to zero there is a substantial percentage of non-convergence of the estimate (up to 11.4%). However, even in this case the "combined" estimator, $\hat{g}_{BC}^{com}$, performs very well. The simulation results in Table 4 show that for this simulation design the presence of an exogenous variable *limits* the bias of the LSDV estimator instead of aggravating it (as was originally claimed by Nickell, 1981).

The last experiment has identical parameter combinations as chosen by Kiviet (1995) in his Table 1 (p.67). The results are presented in Table 5. The mean bias and the root mean squared errors for the LSDV estimator in the different cases were very close to those reported by Kiviet in his Table 2-4. This is a confirmation that the exact same research design is assumed. In each of the ten different cases both the 1-step estimator and the "combined" estimator appear to outperform Kiviet's bias-corrected estimator in terms of root mean squared error. In five out of ten cases there were cases of non-convergence (there was no case of a negative D_6 for the 1-step estimate). These five cases were exactly those (out of ten cases) for which Kiviet's estimator had the highest root mean squared errors. In these cases also the speed of convergence (when it did converge) of the nearly unbiased estimate was low. For example, for case III (g =0.8) the 1-step, 2-step and 3-step estimate were 0.700, 0.739 and 0.753, on average, in case of converging estimates. For the non-converging estimates the 1-step estimate was 0.797, on average. This is very close to the actual value of 0.8. Also for the cases VI, XI and XII the average bias of the 1-step estimate of *non*-converging estimates was *less* than that of converging estimates. However, already the 2-step estimates of the non-converging estimates start to show poor performance.

The Monte Carlo experiments show that the nearly unbiased estimator performs satisfactorily for a wide range of parameter combinations and values of *N* and *T*. In case the estimate converges it is indeed nearly unbiased even for small values of *N*. In a number of cases in which either the sample is very small or the exogenous variable has (almost) no explanatory power the estimator does not converge. Application of the estimator in empirical settings will indicate the importance of this non-convergence problem. However, even in the case of non-convergence the 1-step estimator strongly outperforms the LSDV estimator.

4. Empirical application: unemployment dynamics at the US state level

In this section we apply the nearly unbiased estimator to a model of unemployment dynamics at the US state level. 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 has state fixed effects (h_i) included and is as follows:

(17) $U_{it} = h_i + bG_{i,t-1} + gU_{i,t-1} + e_{it}$

This equation can be rewritten in a form which is more easy to interpret:

(18) $\Delta U_{it} = (g-1)(U_{i,t-1} - a_i) + b(G_{i,t-1} - d) + e_{it}$

where $(1-g)a_i - bd = h_i$. Equation (18) indicates that the change in the unemployment rate is determined by an adjustment of the unemployment rate towards a “natural” or “equilibrium” rate of unemployment a_i which may be different across the states and by the previous economic growth rate. The speed of adjustment of the unemployment rate towards the “natural” or “equilibrium” rate is equal to $1-g$. It is to be expected that there is partial adjustment: $1 > g \geq 0$. A state which shows relatively high economic growth is more likely to show reduced unemployment rates when compared to states in which the economy is growing relatively slowly. This would imply that $b > 0$.

The data for the unemployment rate for the 1991-2000 period are from the U.S. Bureau of Labor Statistics and data for the (current dollar) gross state product are 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 50 U.S. states and Washington D.C. (*N* = 51). The number of time periods *T* is 9 because the year 1991 is taken as year 0. The variables $U_{i,t-1} - \overline{U}_{i,-1}$ and $G_{i,t-1} - \overline{G}_{i,-1}$ have an estimated correlation coefficient of 0.029 ($R^2_{xy_{-1}}$ = 0.029). Because *T* = 9 we use the following values from Table 1: a_9 = -0.144, b_9 = -0.081, c_9 = 0.383 and d_9 = 1.570.

The within (LSDV) estimates and the 1-step, 2-step and 3-step estimates are given in Table 6. The nearly unbiased estimates converge in three steps. The value of the LSDV estimate of g equals 0.805 which would imply an adjustment rate of almost 20%. In contrast, the nearly unbiased estimate is equal to 0.942 which implies an adjustment rate of less than 6 %. Hence, the speed of adjustment towards a "natural rate of unemployment" is not nearly as large of the within estimator would suggest. The value of the LSDV estimate of b equals –0.083 while the value of the nearly unbiased estimate is –0.071. It implies a somewhat smaller effect of economic growth on the change in unemployment than indicated by the traditional estimates.

The average value of the estimated fixed effects is higher for the LSDV estimation procedure (0.00847) when compared to that of the 3-step estimator (0.00833). This is in accordance with the upward bias of the LSDV estimator of the fixed effects. The standard errors of the LSDV and 3-step estimates are computed by using a bootstrap procedure with 100 replications. For the LSDV estimate of g the standard deviation of the 100 estimates is equal to 0.0298. This is clearly less than the standard deviation of the 3-step estimate which is 0.0446. For the LSDV estimate of b the standard deviation is equal to 0.0155. This is about equal to that of the 3-step estimate which is 0.0158.

5. Conclusion

This paper introduces a new estimator for the fixed effects dynamic panel data model. This estimator does not share some of the drawbacks of recently developed IV and GMM estimators and has a good performance even in small samples. The nearly unbiased estimator is derived as a bias correction of the within estimator (least squares dummy variable estimator). Hence, it is related to Kiviet's (1995) bias corrected estimator. However, it is computationally simpler and does not require a first-stage consistent estimate of the parameter vector as needed in the case of Kiviet's correction.

The within estimator is popular amongst applied researchers using panel data sets. This popularity is not confined to the case of exogenous variables only, but extends to some extent to the case of a lagged endogenous variable despite the inconsistency of the estimator. The current paper suggests a simple procedure to consider the extent of bias which has likely occured by applying the within estimator. In general it can be said that the bias will be worse (i) the smaller the number of time periods in the panel; (ii) the stronger the (multiple) correlation between the lagged endogenous variable and the exogenous variables over time, and (iii) the less the explanatory power of the exogenous variables in the model.

The nearly unbiased estimator performs well in the Monte Carlo experiments. Across a wide range of simulation designs it either converges always or in the large majority of cases. Convergence has been found to guarantee the estimator to have desirable properties in terms of low bias and low mean squared error. However, in a minority of (extreme) cases the estimator may run into the problem of non-convergence. Not surprisingly, this problem occurs most frequently in cases in which Kiviet's bias-corrected estimator has relatively poor performance. Even in such cases the large majority of simulations showed convergence of the nearly unbiased estimates. A first solution to the problem of non-converging estimates is to take a 1-step estimate. Although this appears to work satisfactorily in the simulation exercises, additional research into this issue of non-convergence is necessary. It may be the case that non-convergence suggests that either (i) the sample dimensions (*NxT*) are too small to successfully apply fixed effects dynamic panel data models or (ii) model (1) is misspecified in that the exogenous variable(s) incorporated in the model lack any explanatory power.

The expression for the nearly unbiased estimator provides some additional insight into whether the presence of exogenous regressors aggravates the "Nickell" bias or not. Nickell (1981), for example, argued that this presence would aggravate the bias. This paper suggests that it depends upon the exogenous variables introduced. An exogenous variable which is very highly correlated with the lagged endogenous variable and which provides little additional explanatory power will lead to worse bias. However, an exogenous variable that is not highly correlated with the lagged endogenous variable and which has additonal explanatory power will limit the bias.

References

Ahn, S.C. and P. Schmidt (1995), Efficient estimation of models for dynamic panel data, *Journal of Econometrics* 68, 29-52.

Anderson, T.W. and C. Hsiao (1981), Estimation of dynamic models with error components, *Journal of the American Statistical Association* 76, 598-606.

Anderson, T.W. and C. Hsiao (1982), Formulation and estimation of dynamic models using panel data, *Journal of Econometrics* 18, 47-82.

Arellano, M. and S.R. 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.

Arellano, M. and O. Bover (1995), Another look at the instrumental variable estimator of error-components models, *Journal of Econometrics* 68, 29-52.

Baltagi, B.H., J.M. Griffin and W. Xiong (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 S. Bond (1998), Initial conditions and moment restrictions in dynamic panel data models, *Journal of Econometrics* 87, 115-143.

Hahn, J. (1999), How informative is the initial condition in the dynamic panel model with fixed effects?, *Journal of Econometrics* 93, 309-326.

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

Judson, R.A. and A.L. Owen (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.

Kiviet, J.F. (1995), On bias, inconsistency, and inefficiency of various extimators in dynamic panel data models, *Journal of Econometrics* 68, 53-78.

MacKinnon, J.G. and A.A. Smith, Jr. (1998), Approximate bias correction in econometrics, *Journal of Econometrics* 85, 205-230.

Nickell, S. (1981), Biases in dynamic models with fixed effects, *Econometrica* 49, 1417-1426.

Wansbeek, T. and P. Bekker (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 and Economic Statistics* 15, 419-431.

Table 1: Approximation of the $f(g, T)$-function

T	a_T	b_T	c_T	d_T	$R^2_{approximation}$
4	-9.164	-0.592	121.436	12.986	1.00000
5	-1.362	-0.259	6.167	4.052	1.00000
6	-0.505	-0.154	1.607	2.494	1.00000
7	-0.289	-0.115	0.816	1.978	0.99999
8	-0.195	-0.094	0.526	1.722	0.99999
9	-0.144	-0.081	0.383	1.570	0.99998
10	-0.112	-0.071	0.298	1.470	0.99997
11	-0.090	-0.064	0.244	1.398	0.99996
12	-0.075	-0.058	0.205	1.345	0.99995
13	-0.063	-0.054	0.177	1.304	0.99993
14	-0.054	-0.050	0.155	1.272	0.99992
15	-0.047	-0.046	0.139	1.245	0.99991
16	-0.042	-0.043	0.125	1.223	0.99990
17	-0.037	-0.041	0.113	1.205	0.99989
18	-0.033	-0.039	0.104	1.189	0.99988
19	-0.030	-0.037	0.096	1.176	0.99987
20	-0.027	-0.035	0.089	1.164	0.99986
21	-0.025	-0.034	0.083	1.153	0.99984
22	-0.023	-0.032	0.078	1.144	0.99983
23	-0.021	-0.031	0.073	1.136	0.99982
24	-0.019	-0.030	0.069	1.129	0.99981
25	-0.018	-0.029	0.065	1.122	0.99980
26	-0.017	-0.028	0.062	1.116	0.99979
27	-0.016	-0.027	0.059	1.111	0.99978
28	-0.015	-0.026	0.056	1.106	0.99978
29	-0.014	-0.025	0.054	1.101	0.99977
30	-0.013	-0.024	0.051	1.097	0.99976

Note: The values of a_T, b_T, c_T and d_T have been calculated as the best least squares approximation of the $f(g, T)$-function by the function $a_T + b_T g + c_T / (d_T - g)$. The values of g for which the approximation has been made range from 0.000 to 0.999 with step size 0.001. The R^2 of the approximation is given in the last column of the table.

Table 2: Monte Carlo simulation exercises with 600 observations

T N	2 300	3 200	4 150	5 120	6 100	8 75	10 60	12 50	15 40	20 30	30 20
$g = 0.3$											
$\hat{g}$	-0.078	0.069	0.136	0.177	0.200	0.230	0.245	0.255	0.265	0.275	0.283
$\hat{g}_{BC}^{(i)}$	0.228	0.280	0.290	0.296	0.297	0.299	0.298	0.297	0.299	0.299	0.298
$\hat{g}_{BC}^{(ii)}$	0.275	0.299	0.298	0.300	0.299	0.300	0.298	0.297	0.299	0.299	0.298
$\hat{g}_{BC}^{(iii)}$	0.291	0.303	0.299	0.300	0.299	0.300	0.298	0.297	0.299	0.299	0.298
$\hat{g}_{BC}^{lim}$	0.300	0.303									
RMSE $\hat{g}$	0.381	0.234	0.167	0.127	0.104	0.075	0.062	0.052	0.043	0.035	0.030
RMSE $\hat{g}_{BC}^{(iii)}$	0.068	0.048	0.038	0.036	0.032	0.030	0.029	0.028	0.027	0.024	0.025
$g = 0.7$											
$\hat{g}$	0.313	0.471	0.544	0.586	0.612	0.641	0.657	0.666	0.675	0.683	0.688
$\hat{g}_{BC}^{(i)}$	0.625	0.673	0.687	0.693	0.696	0.698	0.700	0.698	0.700	0.700	0.698
$\hat{g}_{BC}^{(ii)}$	0.674	0.694	0.697	0.697	0.699	0.699	0.700	0.698	0.700	0.700	0.698
$\hat{g}_{BC}^{(iii)}$	0.691	0.698	0.698	0.698	0.699	0.699	0.700	0.698	0.700	0.700	0.698
$\hat{g}_{BC}^{lim}$	0.700	0.699									
RMSE $\hat{g}$	0.389	0.232	0.159	0.117	0.091	0.062	0.046	0.038	0.030	0.022	0.018
RMSE $\hat{g}_{BC}^{(iii)}$	0.076	0.046	0.034	0.029	0.024	0.021	0.019	0.016	0.016	0.015	0.014
$g = 0.9$											
$\hat{g}$	0.565	0.707	0.773	0.808	0.830	0.855	0.868	0.875	0.882	0.888	0.893
$\hat{g}_{BC}^{(i)}$	0.845	0.881	0.892	0.895	0.897	0.899	0.899	0.899	0.899	0.900	0.900
$\hat{g}_{BC}^{(ii)}$	0.885	0.897	0.900	0.899	0.899	0.900	0.900	0.899	0.900	0.900	0.900
$\hat{g}_{BC}^{(iii)}$	0.897	0.900	0.901	0.899	0.900	0.900	0.900	0.899	0.900	0.900	0.900
$\hat{g}_{BC}^{lim}$	0.902	0.901									
RMSE $\hat{g}$	0.338	0.195	0.130	0.095	0.072	0.047	0.034	0.028	0.021	0.015	0.010
RMSE $\hat{g}_{BC}^{(iii)}$	0.062	0.038	0.030	0.025	0.020	0.016	0.013	0.012	0.010	0.009	0.007

Note: The simulation exercises have $b = 1$, $r = 0.8$ and $s_u = s_h = s_e = 1$. There were no cases of lack of convergence of the estimate. There were no cases for $T > 3$ in which D_T was found to be negative. The limiting value of the estimate is different from the 3-step estimate only for $T = 2$ and $T = 3$. The reported values are means of the estimates over 500 replications and the root mean squared errors for the within estimator and the 3-step estimator.

Table 3: Monte Carlo simulation exercises with 60 observations

T	2	3	4	5	6	10
N	30	20	15	12	10	6
$g = 0.3$						
% no estimate	8.2	2.4	0.4	0.0	0.0	0.0
$\hat{g}$	-0.070	0.068	0.126			
$\hat{g}_{BC}^{(i)}$	0.252	0.289	0.284			
$\hat{g}_{BC}^{com}$	0.329	0.319	0.296			
$g = 0.7$						
% no estimate	10.2	3.8	1.2	0.2	0.2	0.0
$\hat{g}$	0.306	0.467	0.533	0.570	0.603	
$\hat{g}_{BC}^{(i)}$	0.633	0.686	0.685	0.686	0.694	
$\hat{g}_{BC}^{com}$	0.708	0.721	0.701	0.694	0.699	
$g = 0.9$						
% no estimate	6.2	2.2	0.8	1.2	0.0	0.4
$\hat{g}$	0.568	0.695	0.769	0.797		0.862
$\hat{g}_{BC}^{(i)}$	0.862	0.884	0.898	0.896		0.902
$\hat{g}_{BC}^{com}$	0.924	0.913	0.911	0.904		0.903

Note: see note to Table 2. The percentages of “no estimate” are based upon the relative number of replications in which there was no convergence. There was no case of $D_T < 0$ for the 1-step estimate.

Table 4: Monte Carlo simulation exercises with $g = 0.7$

r	0.8			0.99			-0.99		
b	4	1	0	4	1	0	4	1	0
$\hat{g}$	0.693	0.612	0.366	0.695	0.631	0.366	0.648	0.454	0.367
$\hat{g}_{BC}^{(i)}$	0.700	0.696	0.641	0.700	0.698	0.640	0.698	0.672	0.642
$\hat{g}_{BC}^{com}$	0.700	0.699	0.690	0.700	0.700	0.689	0.699	0.702	0.691
% no estimate	0.0	0.0	10.6	0.0	0.0	10.8	0.0	0.2	11.4
RMSE $\hat{g}$	0.009	0.091	0.337	0.007	0.072	0.337	0.054	0.249	0.335
RMSE $\hat{g}_{BC}^{(i)}$	0.006	0.025	0.085	0.006	0.022	0.085	0.019	0.056	0.084
RMSE $\hat{g}_{BC}^{com}$	0.006	0.025	0.072	0.006	0.023	0.072	0.019	0.059	0.070
$R^2_{xy_{-1}}$	0.039	0.031	0.003	0.150	0.125	0.004	0.965	0.893	0.001

Note: In all cases T = 6, N = 100, $g = 0.7$ and $s_u = s_h = s_e = 1$. There were no cases of $D_T < 0$ for the 1-step estimator. The reported values are means of the estimates over 500 replications and the root mean squared errors for the within estimator, the 1-step estimator and the combined estimator.

Table 5: Monte Carlo results for the cases used by Kiviet (1995)

	T	g	r	s_e	$\hat{g}$	$\hat{g}_{BC}^{(i)}$	$\hat{g}_{BC}^{com}$	rmse	rmse	rmse	%n.e.
I	6	0.0	0.8	0.85	-0.103	-0.001	0.001	0.108	0.037	0.038	0.0
II	6	0.4	0.8	0.88	0.222	0.388	0.397	0.182	0.046	0.046	0.0
III	6	0.8	0.8	0.40	0.443	0.722	0.771	0.359	0.100	0.073	22.6
IV	6	0.0	0.99	0.20	-0.162	-0.005	0.000	0.167	0.048	0.049	0.0
V	6	0.4	0.99	0.19	0.151	0.377	0.397	0.253	0.060	0.061	0.0
VI	6	0.8	0.99	0.07	0.435	0.718	0.764	0.367	0.104	0.073	26.8
VII	3	0.4	0.8	0.88	0.019	0.333	0.403	0.386	0.108	0.108	1.6
VIII	3	0.4	0.8	1.84	0.187	0.383	0.405	0.218	0.062	0.065	0.0
XI	3	0.4	0.99	0.19	-0.087	0.287	0.387	0.492	0.152	0.129	10.0
XII	3	0.4	0.99	0.40	-0.064	0.303	0.404	0.469	0.137	0.129	7.6

Note: The Roman numbers refer to the cases in Kiviet (1995, Table 1). In each of the simulation experiments $s_h = b = 1 - g$ and $N = 100$. The last column (%n.e.) contains the percentage of no convergence of estimates. Kiviet (1995) has four more cases (IX, X, XIII and XIV), but the results of these cases are very similar to those of four cases which are already incorporated in the table (VII, VIII, XI and XII). The three columns of "rmse" are in the same order as the three estimates in the previous three columns.

Table 6: Estimates of the intemporal relation (17) between unemployment and growth

	$\hat{g}$	$\hat{b}$	$\hat{s}_u^2 / \hat{s}_{y_{-1}}^2$	$\hat{s}_u$
LSDV	0.805	-0.083	0.321	0.00625
1-step	0.931	-0.072	0.339	0.00642
2-step	0.941	-0.071	0.342	0.00645
3-step	0.942	-0.071	0.342	0.00645