



UvA-DARE (Digital Academic Repository)

On the diminishing returns of higher-order terms in asymptotic expansions of bias

Bun, M.J.G.; Kiviet, J.F.

[Link to publication](#)

Citation for published version (APA):

Bun, M. J. G., & Kiviet, J. F. (2002). On the diminishing returns of higher-order terms in asymptotic expansions of bias. (UvA-econometrics discussion paper; No. 2002/06). Amsterdam: Universiteit van Amsterdam.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <http://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Discussion Paper: 2002/06

On the diminishing returns of higher-order terms in asymptotic expansions of bias

Maurice J.G. Bun and Jan F. Kiviet

www.fee.uva.nl/ke/UvA-Econometrics

Department of Quantitative Economics
Faculty of Economics and Econometrics
Universiteit van Amsterdam
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA  UNIVERSITEIT VAN AMSTERDAM



On the diminishing returns of higher-order terms in asymptotic expansions of bias*

MAURICE J.G. BUN and JAN F. KIVJET
(University of Amsterdam & Tinbergen Institute)

18 September 2002

JEL-code: C13; C23

Keywords: asymptotic expansion, bias approximation, dynamic panel data model, finite sample bias, LSDV estimator

Abstract

The relative magnitudes are compared of successive terms in a higher-order asymptotic expansion of the bias of the LSDV estimator in dynamic panels. We find that the leading term accounts for the major part of the actual bias in small samples. This implies that bias correction procedures can be based on relatively simple bias approximation formulas.

1. Introduction

For dynamic panel data models the inconsistency of the LSDV (least-squares with dummy variables or within groups) estimator for finite T (number of time-series observations) and large N (number of cross-section units) is a well established fact. Nickell (1981) derived for the strongly stationary panel AR(1) model an analytical expression for the inconsistency for $N \rightarrow \infty$ of LSDV which is an expression involving terms in powers of T^{-1} . He also presents a similar but less explicit result for the first-order dynamic panel data model with exogenous regressors. However, Nickell did not consider any terms that contribute to the LSDV bias which are of order N^{-1} , which may be substantial in samples where N is small. Some of these are considered in Kiviet (1995), where an explicit asymptotic approximation formula has been obtained for the finite sample bias of the LSDV estimator in the first-order dynamic panel data model with additional exogenous regressors. That study also examined by simulation its accuracy and effectiveness with respect to improving estimator efficiency by bias correction. In Kiviet (1999) the approximation formula has been slightly extended by taking a particular further higher-order term into account, but the actual performance of this refinement has not been examined yet. Here, we will analyze and compare by simulation the accuracy of these various bias approximations in small samples for the first-order dynamic panel data model with no, or just one, additional exogenous regressor.

2. Approximations of the LSDV bias

We consider the standard linear first-order¹ dynamic panel data model

$$y_{it} = \gamma y_{i,t-1} + \beta' x_{it} + \eta_i + \varepsilon_{it}, \quad i = 1, \dots, N; \quad t = 1, \dots, T. \quad (2.1)$$

*Tinbergen Institute & Faculty of Economics and Econometrics, Universiteit van Amsterdam, Roeterstraat 11, 1018 WB Amsterdam, The Netherlands (M.J.G.Bun@UvA.NL and J.F.Kiviet@UvA.NL); phone +31.20.5254217, fax +31.20.5254349.

¹For ease of exposition we restrict ourselves to the first-order dynamic panel data model, but all results can be extended to higher-order dynamic models.

In this model the dependent variable y_{it} is regressed on a vector of K exogenous explanatory variables x_{it} and the one period lagged value of the dependent variable. The composite disturbance term $\eta_i + \varepsilon_{it}$ consists of an unobserved individual specific effect η_i , which is either fixed or random, and a white noise error term ε_{it} with variance σ_ε^2 . Stacking the observations over time and across individuals one gets

$$y = W\delta + (I_N \otimes \iota_T)\eta + \varepsilon, \quad (2.2)$$

where $\delta = (\gamma, \beta')'$, y and $W = (y_{-1}, X)$ are $NT \times 1$ and $NT \times (K + 1)$ matrices of stacked observations, ε is the $NT \times 1$ vector of disturbances and $\iota_T = (1, \dots, 1)'$ a $T \times 1$ vector of ones.

Estimation of the $N + K + 1$ coefficients $(\delta', \eta')'$ of (2.2) by ordinary least squares yields² regarding δ

$$\hat{\delta}_{LSDV} = (W'AW)^{-1}W'Ay = \delta + (W'AW)^{-1} \begin{pmatrix} y'_{-1}A\varepsilon \\ X'A\varepsilon \end{pmatrix}, \quad (2.3)$$

where the $NT \times NT$ matrix $A = I_N \otimes (I_T - \frac{1}{T}\iota_T\iota_T')$ is the within transformation which wipes out the individual effects. This procedure is valid for both fixed and random effects. In case of random effects it is especially useful in case of non-zero correlation between the effects and the time varying regressors. Although strict exogeneity of the regressors X implies $E(X'A\varepsilon) = 0$, we have $\text{plim}_{N \rightarrow \infty} \frac{1}{N}W'A\varepsilon \neq 0$, since it can be shown that $\text{plim}_{N \rightarrow \infty} \frac{1}{N}y'_{-1}A\varepsilon \neq 0$ for T finite. Therefore, $\hat{\delta}_{LSDV}$ is inconsistent for $N \rightarrow \infty$ and T finite.

Nickell (1981) examines the bias of the LSDV estimator for $N \rightarrow \infty$, which is of order $O(T^{-1})$. However, in finite samples where both N and T are moderate or small additional bias terms may play a major role as well³. Kiviet (1995) derives an approximation formula for the bias of the LSDV estimator in the normal stationary dynamic panel data model with a strictly exogenous regressor matrix X while conditioning on start-up values (v_{10}, \dots, v_{N0}) from which the accumulated individual effects have been removed, i.e. $v_{i0} = y_{i0} - \eta_i/(1 - \gamma)$. The resulting expression includes all $O(N^{-1}T^{-1})$ contributions to the bias and has an approximation error which has a magnitude of order $O(N^{-1})$ and of order $O(T^{-2})$ at the same time. In Kiviet (1999) this analysis has been extended and an approximation to the bias is produced which is not only accurate to order $O(T^{-1})$ again, but now it is accurate to order $O(N^{-1})$ as well, yielding

$$E(\hat{\delta}_{LSDV} - \delta) = c_1(T^{-1}) + c_2(N^{-1}T^{-1}) + c_3(N^{-1}T^{-2}) + O(N^{-2}T^{-2}), \quad (2.4)$$

where

$$\begin{aligned} c_1(T^{-1}) &= QE(W'A\varepsilon) = O(T^{-1}) \\ c_2(N^{-1}T^{-1}) &= -QE(W'AW - Q^{-1})Q[W'A\varepsilon - E(W'A\varepsilon)] = O(N^{-1}T^{-1}) \\ c_3(N^{-1}T^{-2}) &= -QE\{[(W'AW - Q^{-1})Q]^2\}E(W'A\varepsilon) = O(N^{-1}T^{-2}), \end{aligned}$$

with

$$\begin{aligned} Q &= [E(W'AW)]^{-1} = [\bar{W}'A\bar{W} + \sigma_\varepsilon^2 \text{tr}(\Pi'\Pi)e_1e_1']^{-1} \\ \bar{W} &= E(W) \\ e_1 &= (1, 0, \dots, 0)' \text{ has } K + 1 \text{ elements} \\ \Pi &= AL\Gamma, \Gamma = I_N \otimes \Gamma_T, \Gamma_T = (I_T - \gamma L_T)^{-1}, L = I_N \otimes L_T, \end{aligned}$$

²It is assumed that all the explanatory variables are time variant so that $W'AW$ is invertible.

³Some general background to this type of bias approximation in the context of dynamic panel data models can be found in Beggs and Nerlove (1988).

where the $T \times T$ matrix L_T has ones on its first lower subdiagonal and all other elements equal to zero. Using the shorthand notation $q_1 = Qe_1$ and $q_{11} = e_1'q_1$ and following the derivations as presented in the papers referred to above, one obtains

$$c_1(T^{-1}) = \sigma_\varepsilon^2 \text{tr}(\Pi)q_1 \quad (2.5)$$

$$c_2(N^{-1}T^{-1}) = -QE(W'AWQW'A\varepsilon) + QE(W'A\varepsilon) \quad (2.6)$$

$$\begin{aligned} &= -\sigma_\varepsilon^2 [Q\bar{W}'\Pi A\bar{W} + \text{tr}(Q\bar{W}'\Pi A\bar{W})I_{K+1} + 2\sigma_\varepsilon^2 q_{11} \text{tr}(\Pi'\Pi\Pi)I_{K+1}]q_1 \\ c_3(N^{-1}T^{-2}) &= QE(W'AWQW'AW)QE(W'A\varepsilon) - QE(W'A\varepsilon) \quad (2.7) \\ &= \sigma_\varepsilon^4 \text{tr}(\Pi) \{ 2q_{11}Q\bar{W}'\Pi\Pi'\bar{W}q_1 \\ &\quad + [(q_1'\bar{W}'\Pi\Pi'\bar{W}q_1) + q_{11}\text{tr}(Q\bar{W}'\Pi\Pi'\bar{W}) + 2\text{tr}(\Pi'\Pi\Pi')q_{11}^2]q_1 \}. \end{aligned}$$

In Kiviet (1995) only the first two terms (2.5) and (2.6) have been taken into account, so the approximation error is still $O(N^{-1}T^{-2})$. The more general result with approximation error of order $O(N^{-2}T^{-2})$ can already be found in Kiviet (1999) but here the expressions are simpler due to removing factors $\bar{W}'A\bar{W}$ by exploiting $\bar{W}'A\bar{W} = Q^{-1} - \sigma_\varepsilon^2 \text{tr}(\Pi'\Pi)e_1e_1'$.

Below we shall examine the three bias approximations

$$B_1 = c_1(T^{-1}), \quad B_2 = B_1 + c_2(N^{-1}T^{-1}), \quad B_3 = B_2 + c_3(N^{-1}T^{-2}), \quad (2.8)$$

by comparing them with the true bias as obtained from Monte Carlo simulations. The simulation design is basically the same as in Kiviet (1995), which has also been used by Judson and Owen (1999). Data for y have been generated according to equation (2.1) with $K = 1$. The generating equation for the explanatory variable x is

$$x_{it} = \rho x_{i,t-1} + \xi_{it}, \quad i = 1, \dots, N; \quad t = 1, \dots, T, \quad (2.9)$$

where $\xi_{it} \sim \text{iIN}(0, \sigma_\xi^2)$. The long-run effect $\beta/(1 - \gamma)$ of x on y has been set equal to unity in all experiments. This implies that the impact multiplier β varies with the chosen values for γ . The variance of the disturbance term σ_ε^2 is set at the value of one. The parameter σ_ξ has been determined by controlling the signal-to-noise ratio σ_s^2 of the model, see Kiviet (1995). The following three sample sizes have been analyzed: $T = N = 20$; $T = 10, N = 40$ and $T = 40, N = 10$. Two values for γ and ρ , viz. 0.2 and 0.8, are considered together with $\sigma_s^2 = \{2, 9\}$, which corresponds with a population R^2 equal to $\frac{2}{3}$ and $\frac{9}{10}$ respectively. Hence, for the three chosen sample sizes we consider 8 different designs⁴.

In Table 2.1 we compare the bias approximations (2.8), calculated at the true parameter values, with the actual LSDV bias as estimated from 10,000 Monte Carlo replications. Columns 1 to 5 indicate the various parametrizations, while columns 6 and 10 show the actual LSDV bias in estimating γ and β respectively. Regarding finite sample bias we observe the following patterns in the simulation results. The absolute value of the bias in estimating γ increases with both γ and ρ , hence more persistence in the processes for y and x leads to larger finite sample bias. For all the chosen parametrizations we find a negative finite sample bias in estimating γ , which is serious in relative terms especially when γ is small. The bias in estimating β increases with ρ , but not with γ . Hence, more persistence in the process for x complicates accurate estimation of β . On the whole, however, the bias in estimating β seems less serious than in estimating γ . We find that the magnitude of the finite sample bias depends heavily on the dimensions T and N . In general, bias decreases as T increases, which is not surprising as the leading term of the LSDV bias is of order $O(T^{-1})$. Seriousness of bias is inversely related with the signal-to-noise ratio.

Columns 7 to 9 and columns 11 to 13 of Table 2.1 show the numerical values of the bias approximations (2.8), labelled $B_{j,\gamma}$ and $B_{j,\beta}$ ($j = 1, 2, 3$) respectively. We find that adding

⁴The reported simulation results are a selection of a wider range of parametrizations examined.

Table 2.1: Actual LSDV bias and bias approximations

σ_s^2	T	N	γ	ρ	$Bias$	γ	$B_{1,\gamma}$	$B_{2,\gamma}$	$B_{3,\gamma}$	$Bias$	β	$B_{1,\beta}$	$B_{2,\beta}$	$B_{3,\beta}$
2	10	40	0.2	0.2	-0.045	-0.044	-0.045	-0.045	-0.045	0.002	0.002	0.002	0.002	0.002
				0.8	-0.077	-0.076	-0.077	-0.077	-0.077	0.039	0.038	0.039	0.039	0.039
			0.8	0.2	-0.207	-0.204	-0.204	-0.207	-0.207	-0.008	-0.008	-0.008	-0.008	-0.008
				0.8	-0.216	-0.212	-0.213	-0.216	-0.216	0.024	0.024	0.024	0.024	0.024
	20	20	0.2	0.2	-0.023	-0.022	-0.022	-0.022	-0.022	0.002	0.002	0.002	0.002	0.002
				0.8	-0.039	-0.036	-0.038	-0.038	-0.038	0.024	0.023	0.024	0.024	0.024
			0.8	0.2	-0.101	-0.097	-0.099	-0.101	-0.101	-0.001	-0.001	-0.001	-0.001	-0.001
				0.8	-0.106	-0.101	-0.104	-0.106	-0.106	0.023	0.022	0.023	0.023	0.024
	40	10	0.2	0.2	-0.011	-0.011	-0.011	-0.011	-0.011	0.002	0.001	0.002	0.002	0.002
				0.8	-0.020	-0.018	-0.020	-0.020	-0.020	0.014	0.013	0.014	0.014	0.014
			0.8	0.2	-0.050	-0.046	-0.049	-0.050	-0.050	0.001	0.001	0.001	0.001	0.001
				0.8	-0.053	-0.047	-0.052	-0.053	-0.053	0.017	0.015	0.017	0.017	0.017
9	10	40	0.2	0.2	-0.014	-0.014	-0.014	-0.014	-0.014	0.001	0.001	0.001	0.001	0.001
				0.8	-0.033	-0.032	-0.032	-0.033	-0.033	0.017	0.016	0.017	0.017	0.017
			0.8	0.2	-0.067	-0.067	-0.067	-0.067	-0.067	-0.003	-0.003	-0.003	-0.003	-0.003
				0.8	-0.112	-0.110	-0.110	-0.111	-0.111	0.012	0.012	0.012	0.012	0.012
	20	20	0.2	0.2	-0.007	-0.007	-0.007	-0.007	-0.007	0.001	0.001	0.001	0.001	0.001
				0.8	-0.016	-0.015	-0.016	-0.016	-0.016	0.010	0.009	0.010	0.010	0.010
			0.8	0.2	-0.032	-0.030	-0.031	-0.031	-0.031	-0.000	-0.000	-0.000	-0.000	-0.000
				0.8	-0.048	-0.045	-0.047	-0.048	-0.048	0.011	0.010	0.011	0.011	0.011
	40	10	0.2	0.2	-0.003	-0.003	-0.003	-0.003	-0.003	0.000	0.000	0.000	0.000	0.000
				0.8	-0.008	-0.007	-0.008	-0.008	-0.008	0.006	0.005	0.006	0.006	0.006
			0.8	0.2	-0.015	-0.014	-0.015	-0.015	-0.015	0.000	0.000	0.000	0.000	0.000
				0.8	-0.022	-0.020	-0.022	-0.022	-0.022	0.007	0.006	0.007	0.007	0.007

$B_{j,\gamma}$ and $B_{j,\beta}$ ($j=1,2,3$) are based on (2.8)

higher-order terms improves the approximations, with B_3 almost corresponding to the true bias. However, the successive higher-order terms lead to improvements of diminishing significance. In fact, the leading term B_1 comprises already around 90% of the true bias or more. Further contributions to the bias, which are $O(N^{-1}T^{-1})$ and $O(N^{-1}T^{-2})$, are generally found to lead to improvements, but these are minor.

3. Special results for the panel AR(1) model

For a special case of model (2.1), viz. the AR(1) panel data model with $K = 0$, a further decomposition of the leading term $\sigma_\varepsilon^2 tr(\Pi)q_1$ of the expansion is possible. Here we have $\bar{W} = \bar{y}_{-1}$ and thus $Q = q_1 = q_{11} = [\bar{y}'_{-1}A\bar{y}_{-1} + \sigma_\varepsilon^2 tr(\Pi'\Pi)]^{-1}$, where

$$\bar{y}'_{-1}A\bar{y}_{-1} = \sum_{i=1}^N v_{i0}^2 \left(\frac{1 - \gamma^{2T}}{1 - \gamma^2} - \frac{1}{T} \frac{(1 - \gamma^T)^2}{(1 - \gamma)^2} \right) = O(NT^0)$$

and

$$tr(\Pi'\Pi) = \frac{NT}{1 - \gamma^2} + O(NT^0).$$

So we have $q_{11}^{-1} = \sigma_\varepsilon^2 \frac{NT}{1 - \gamma^2} + O(NT^0)$. With

$$tr(\Pi) = -\frac{N}{1 - \gamma} + \frac{N(1 - \gamma^T)}{T(1 - \gamma)^2}$$

this gives

$$\sigma_\varepsilon^2 tr(\Pi)q_1 = -\frac{1+\gamma}{T} \left(1 - \frac{1}{T} \frac{1-\gamma^T}{1-\gamma} \right) + O(T^{-3}). \quad (3.1)$$

This corresponds – to the order of the approximation – with Nickell’s (1981) formula (18), which has been derived for the strongly stationary model with fully random start-up values. Although approximation (3.1) is of order $O(T^{-1})$, it contains contributions of order $o(T^{-1})$ too. Its first term is the ”pure” $O(T^{-1})$ bias. The approximation to the LSDV bias

$$c_{11}(T^{-1}) = -\frac{1}{T}(1+\gamma) \quad (3.2)$$

is an extremely simple one, which may work well for moderately large N and T , because the ”pure” $O(T^{-1})$ contribution to the bias has been separated here from $O(T^{-2})$ and $O(N^{-1}T^{-1})$ contributions, together constituting a hybrid remainder term. Note that also in this simple panel AR(1) model the bias terms c_h ($h = 1, 2, 3$) depend not only on γ and T , but also on N and the conditioning start-up values $v_{i,0}$.

Figure 3.1 shows the actual bias in estimating γ for values of γ ranging from -0.99 to 0.99 for the selected sample dimensions T and N . Finite sample bias in estimating γ is generally negative and in absolute value increasing in γ . Figure 3.2 shows the difference of the actual finite sample bias with three different bias approximations, viz. $B_{11} = c_{11}(T^{-1})$, B_1 and B_3 as defined in (2.8). The decomposition of the bias approximation in terms of different orders in T and N shows that the pure $O(T^{-1})$ bias approximation accounts already for the major part of the finite sample bias, especially for low values of γ . However, when T is small or moderate B_{11} is inaccurate especially for large positive values of γ . Keeping into the approximation the contributions which are $o(T^{-1})$, but which are still $O(N^0)$ at the same time, becomes profitable in this part of the parameter space. The contributions to the bias which are $O(N^{-1}T^{-1})$ and $O(N^{-1}T^{-2})$ are found to be of very limited actual magnitude for all values of γ , T and N . Hence, B_1 and B_3 are almost similar for the panel AR(1) model.

Figure 3.1: Actual bias as a function of γ and (T, N)

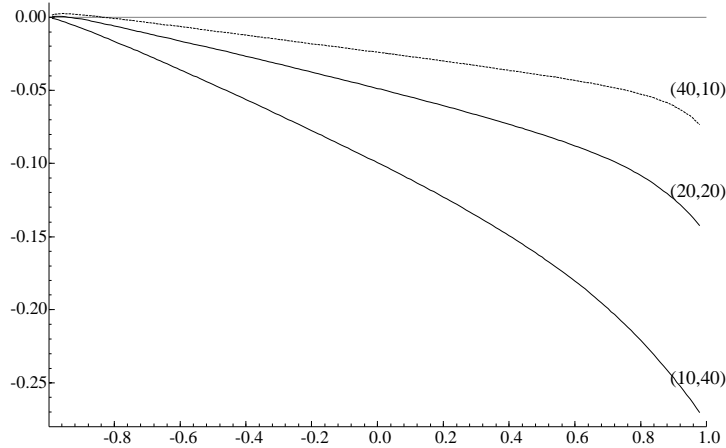
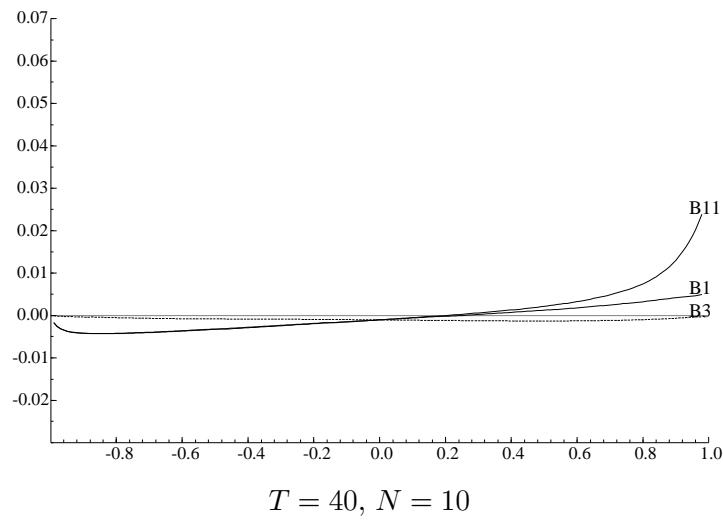
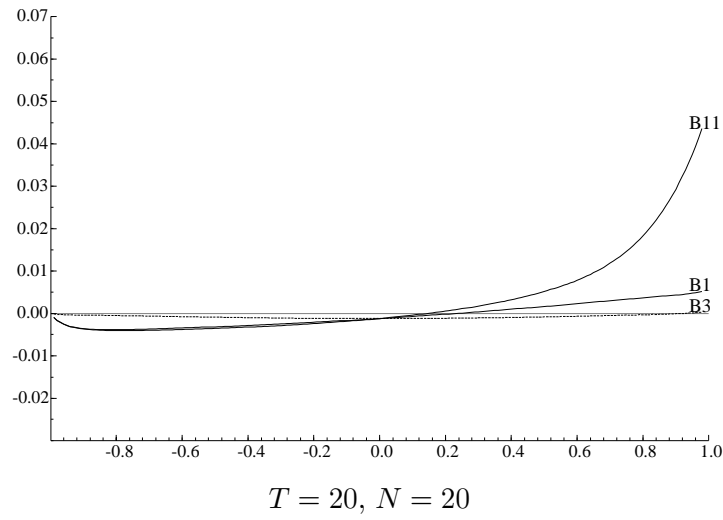
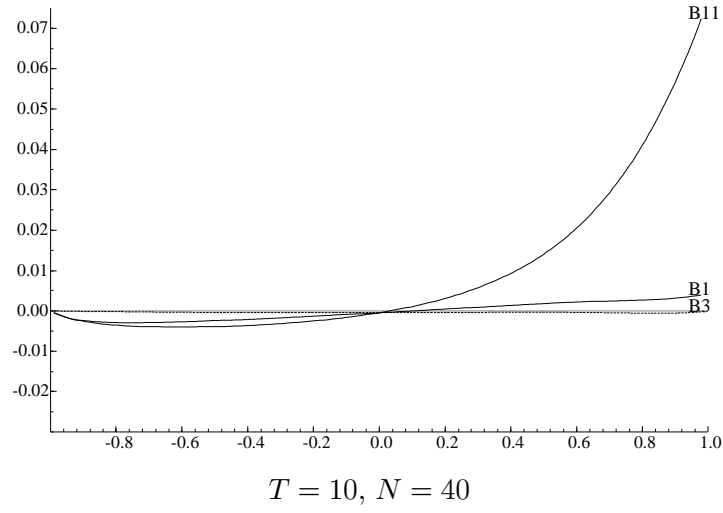


Figure 3.2: Actual bias minus bias approximation as a function of γ



4. Concluding remarks

In this study the accuracy of various analytical approximations for the finite sample bias of the LSDV coefficient estimator for dynamic panel data models have been analyzed through Monte Carlo experiments. The simulation results show that such approximations can be very accurate for a wide range of parametrizations. Adding an extra term to the approximation formula for the bias of the LSDV estimator, as derived in Kiviet (1999), does not improve the bias approximation substantially. In fact, the present results indicate that including $O(T^{-1})$ and $o(T^{-1})$ terms, i.e. all the $O(N^0)$ terms, and omitting the $O(N^{-1}T^{-1})$ contributions accounts for most of the bias in the LSDV estimator when $NT \geq 400$ and $N \geq 10$. These findings are useful for developing corrected estimators, which exploit bias approximations to mitigate bias and to increase efficiency. When both N and T are two digit numbers it seems appropriate to base these bias corrections simply on an estimate of B_1 , i.e. the $O(N^0T^{-1})$ terms, especially in the panel AR(1) model. Invoking higher-order terms yields some minor improvements in models with exogenous regressors.

References:

- Beggs, J.J., Nerlove, M., 1988. Biases in dynamic models with fixed effects. *Economics Letters* 26, 29-31.
- Judson, R.A., Owen, A.L., 1999. Estimating dynamic panel data models: a guide for macroeconomists. *Economics Letters* 65, 9-15.
- Kiviet, J.F., 1995. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics* 68, 53-78.
- Kiviet, J.F., 1999. Expectations of expansions for estimators in a dynamic panel data model; some results for weakly exogenous regressors. In: Hsiao, C., Lahiri, K., Lee, L-F, Pesaran, M.H. (eds.). *Analysis of Panels and Limited Dependent Variables*. Cambridge University Press, Cambridge.
- Nickell, S., 1981. Biases in dynamic models with fixed effects. *Econometrica* 49, 1417-1426.