TI 2001-120/4
Tinbergen Institute Discussion Paper

Bias Correction in a Stable AD(1,1) Model

Noud P.A. van Giersbergen

Department of Quantitative Economics, Faculty of Economics and Econometrics, Universiteit van Amsterdam

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Keizersgracht 482
1017 EG Amsterdam
The Netherlands
Tel.: +31.(0)20.5513500
Fax: +31.(0)20.5513555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31.(0)10.4088900
Fax: +31.(0)10.4089031

Bias Correction in a Stable AD(1,1) Model: Weak versus Strong Exogeneity

Noud P.A. van Giersbergen*
Department of Quantitative Economics,
Universiteit van Amsterdam
Roetersstraat 11
1018 WB Amsterdam
The Netherlands
E-mail: nvg@fee.uva.nl

December 4, 2001

Abstract

This paper compares the behaviour of a bias-corrected estimator assuming strongly exogenous regressors to the behaviour of a bias-corrected estimator assuming weakly exogenous regressors, when in fact the marginal model contains a feedback mechanism. To this end, the effects of a feedback mechanism on the first-order least-squares coefficient estimation bias is examined through large-sample asymptotics in a stable first-order autoregressive distributed-lag model with weakly exogenous regressors. The derived formulae show explicitly how the bias of the coefficient estimators of the conditional model depends on the parameters that belong to the marginal model. In addition, an explicit approximation in all the system parameters is derived for the first-order bias formula based on strongly exogenous regressors. It is found that the two bias approximations can lead to quite different numerical values. Through a small simulation study, the bias and efficiency of the two bias-corrected estimators is investigated. It appears that the valid bias-corrected estimator based on the whole system is somewhat less biased than the invalid bias-corrected estimator. For a few particular parameter values considered, however, both bias-corrected estimators are inefficient relative to the uncorrected estimator in terms of mean squared error. Somewhat surprisingly, the invalid bias-corrected estimator based on only the conditional model is on average just as efficient as the valid bias-corrected estimator based on the whole system.

Keywords: autoregressive distributed-lag models, estimation bias, large sample asymptotics, Nagar expansions

JEL classification: C13; C22.

*Noud van Giersbergen (1970) graduated from the Erasmus University Rotterdam in 1988 with a degree in econometrics. He then became a PhD student at the Tinbergen Institute and a research assistant at the former Department of Actuarial Science and Econometrics at the Universiteit van Amsterdam. His research cumulated in a PhD-thesis entitled 'Bootstrapping Dynamic Econometric Models', which was published as book no. 184 of the Tinbergen Institute Research Series.

His supervisor was prof. dr. J.F. Kiviet (jfk@fee.uva.nl), who is currently head of the Department of Quantitative Economics at the Universiteit van Amsterdam and a senior Research Fellow of the Tinbergen Institute. Helpful comments from him are gratefully acknowledged.

1 Introduction

The use of asymptotic expansions in approximating the estimation bias in stable autoregressive (AR) models has a relatively long history. The early work focused on the least-squares estimator of the serial correlation coefficient in the first-order autoregressive –AR(1)– model with Gaussian disturbances. In this model, Barlett (1946) and Hurwicz (1950) obtained a first-order approximation of the estimation bias, while White (1961) found higher-order approximations in terms of powers of T^{-1}, where T is the sample size. For the AR(1) model including an intercept, Kendall (1954) and Marriot and Pope (1954) gave an approximation to the bias of the least-squares estimator of the lagged-dependent variable coefficient to the order of T^{-1}. In the stable AR(1) model without a constant, the bias of the AR(1)-coefficient estimator to the order $O(T^{-1})$ is given by $\mathbb{E}[\hat{\lambda} - \lambda] = -2\lambda/T$, where λ denotes the AR(1)-coefficient. When a constant is included, the first-order bias approximation equals $-(1 + 3\lambda)/T$.

More recently, Grubb and Symons (1987) derived the bias to the order T^{-1} for the lagged-dependent variable coefficient estimator in a stable first-order dynamic regression model with exogenous regressors, while Kiviet and Phillips (1993) gave the T^{-1} approximation of the full coefficient vector. Both papers assume strongly exogenous regressors, like a constant or (higher-order) trend(s), which does not allow for any feedback effects between endogenous and explanatory variables. In practice, however, it is unlikely that regressors related to economic variables do not involve feedback mechanisms from past economic outcomes onto current decisions. Hence, one would like to allow for weakly exogenous regressors, which may depend on lagged values of the dependent variable; see Engle *et al.* (1983) for the various concepts of exogeneity. Of course, it is possible to ignore such feedback effects and proceed by assuming that the regressors are strongly exogenous. In doing so, however, a misspecification error is made which may or may not be serious. The aim of this paper is to assess the seriousness of this misspecification error in autoregressive distributed-lag (AD) models by comparing the bias formula that results from ignoring feedback effects to the bias that does take such feedback effects into account.

More specifically, we consider the bias of the ordinary least-squares (OLS) coefficient estimators in the stable AD(1,1) model given by

$$y_t = \lambda y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \theta + \varepsilon_t, \qquad t = 1, ..., T, \tag{1}$$

where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, x_t is stationary and $|\lambda| < 1$. Although the AD(1,1) model seems quite specific, virtually every type of single-equation model in empirical time-series econometrics is a special case of it; see for instance Hendry *et al.* (1984). The marginal model for x_t is assumed to be

$$x_t = \pi y_{t-1} + \gamma x_{t-1} + \eta_t, \tag{2}$$

where $\eta_t \sim N(0, \sigma_\eta^2)$. Since $\{\eta_t\}$ is assumed to be independent of $\{\varepsilon_t\}$, x_t is weakly exogenous for obtaining inference on the parameters in the conditional model (1). Moreover, x_t is strongly exogenous when $\pi = 0$. Weak exogeneity of the regressors implies that asymptotically efficient inference on the model coefficients in (1) can be obtained from analysis of the conditional model (1) in isolation. In deriving the expressions of the estimators, one operates as if x_t were strongly exogenous. However, the finite-sample bias will certainly be affected by the joint stochastic behaviour of y_t and x_t, since it depends on the value of the feedback parameter π. To date, however, no bias expansions have been derived in the AD(1,1) model when x_t is weakly but not strongly exogenous ($\pi \neq 0$). Since the bias expressions depend on the parameters of the marginal model, a bias-corrected inference procedure would also require the analysis of this model, despite the weak exogeneity of x_t. Hence, the pros and cons of each procedure should be weighted and it seems interesting to compare the performance of the bias-corrected coefficient estimator based on the whole system to the bias-corrected estimator obtained in the conditional model (1) only, which does not need to fully specify and analyse the marginal model (2).

The paper is organised as follows. Section 2 introduces some notation and decomposes the regressors into a deterministic and stochastic part. Furthermore, it contains the Nagar-type equation, which is used for deriving the bias. In Section 3, the actual first-order bias formulae for the AD(1,1) model with or without an intercept are presented. The results are specialised to the AD(1,0) model in Section 4. The fifth section analyses the bias formula for the estimator of λ in the AD(1,0) model based on strongly exogenous regressors when in fact the marginal model contains a feedback mechanism. Section 6 contains some graphical illustrations and presents some simulation results. Finally, the main conclusions are summarised in the last section.

2 Notation and Preliminary Analysis

The focus of interest is the bias of the OLS estimator of the regression coefficients in the AD(1,1)-model

$$y = \lambda y_{-1} + \beta_0 x + \beta_1 x_{-1} + \theta + \varepsilon, \tag{3}$$

where $y = (y_1, ..., y_T)'$ is a $T \times 1$ vector of observations on a dependent variable, $y_{-1} = (y_0, ..., y_{T-1})'$ is the vector y lagged one period, x is a $T \times 1$ vector of observations on a weakly but not necessarily strongly exogenous regressor and $\varepsilon \sim N(0, \sigma_\varepsilon^2 I_T)$ is a $T \times 1$ Gaussian vector. The marginal model for x in vector notation is given by

$$x = \pi y_{-1} + \gamma x_{-1} + \eta, \tag{4}$$

where $\eta \sim N(0, \sigma_\eta^2 I_T)$ is a $T \times 1$ Gaussian vector independent of ε. The value of π determines whether x is strongly or only weakly exogenous. To facilitate notation, we shall write model (3) as

$$y = Z\alpha + \varepsilon, \qquad Z = (y_{-1} : x : x_{-1} : \iota_T), \tag{5}$$

where ι_T a $T \times 1$ vector of ones and $\alpha = (\lambda, \beta_0, \beta_1, \theta)'$. The OLS estimator of α is

$$\hat{\alpha} = (Z'Z)^{-1}Z'y = \alpha + (Z'Z)^{-1}Z'\varepsilon, \tag{6}$$

so that the bias of $\hat{\alpha}$ is given by

$$B_\alpha = \mathbb{E}[(\hat{\alpha} - \alpha)] = \mathbb{E}[(Z'Z)^{-1}Z'\varepsilon]. \tag{7}$$

First, we shall derive the conditions for which the whole system is stable. Using the lag-operator L, the data generation process (DGP) can be written as

$$(1-\lambda L)y_t = (\beta_0 + \beta_1 L)x_t + \theta + \varepsilon_t, \tag{8a}$$

$$(1-\gamma L)x_t = \pi L y_t + \eta_t. \tag{8b}$$

Multiplication of (8a) by $(1-\gamma L)$ and substitution of (8b) in (8a) gives

$$(1-\phi L - \psi L^2)y_t = (1-\gamma L)\theta + (1-\gamma L)\varepsilon_t + (\beta_0 + \beta_1 L)\eta_t, \tag{9}$$

where

$$\phi \equiv \pi\beta_0 + \gamma + \lambda \qquad \text{and} \qquad \psi \equiv \pi\beta_1 - \gamma\lambda. \tag{10}$$

To economize on notation, the parameters ϕ and ψ are used as much as possible throughout the paper. For x_t, it can be found that

$$(1-\phi L - \psi L^2)x_t = \pi\theta + \pi L\varepsilon_t + (1-\lambda L)\eta_t. \tag{11}$$

From formulae (9) and (11), it follows that the time-series $\{y_t\}$ and $\{x_t\}$ can be characterised by an ARMA(2,1) process contaminated with an independent MA(1) process. The variables y_t and x_t are stationary if the characteristic polynomial associated with (9) and (11), given by

$$1 - \phi z - \psi z^2 = 0, \tag{12}$$

has all its root outside the unit circle. This leads to three inequalities; see for instance Harvey (1981). For a particular set of values of the parameters in the conditional model, these inequalities are given by

$$\psi < 1-\phi \iff \pi\beta_1 - \gamma\lambda < 1 - \pi\beta_0 - \gamma - \lambda, \tag{13a}$$

$$\psi < 1+\phi \iff \pi\beta_1 - \gamma\lambda < 1 + \pi\beta_0 + \gamma + \lambda, \tag{13b}$$

$$\psi > -1 \iff \pi\beta_1 - \gamma\lambda > -1. \tag{13c}$$

In order to derive the large T-bias approximation, the regressors are decomposed into several (independent) stochastic and deterministic components. To do so, define the $T \times T$ matrices

$$\Gamma = \begin{bmatrix} 1 & 0 & 0 & . & . & . & 0 \\ -\phi & 1 & 0 & & & & . \\ -\psi & -\phi & 1 & 0 & & & . \\ 0 & -\psi & -\phi & 1 & & & . \\ . & & . & . & . & 0 & . \\ . & & & -\psi & -\phi & 1 & 0 \\ 0 & . & . & 0 & -\psi & -\phi & 1 \end{bmatrix} \quad \text{and} \quad L_T = \begin{bmatrix} 0 & 0 & . & . & . & . & 0 \\ 1 & 0 & & & & & . \\ 0 & 1 & 0 & & & & . \\ . & . & . & . & & & . \\ . & & . & . & . & & . \\ . & & & & 1 & 0 & 0 \\ 0 & . & . & . & 0 & 1 & 0 \end{bmatrix}. \tag{14}$$

Let U denote the inverse of Γ, which for $T = 5$ is equal to

$$U = \begin{bmatrix} 1 & 0 & 0 & . & 0 \\ \phi & 1 & 0 & & . \\ \phi^2 + \psi & \phi & 1 & 0 & . \\ \phi^3 + 2\phi\psi & \phi^2 + \psi & \phi & 1 & 0 \\ \phi^4 + 3\phi^2\psi + \psi^2 & \phi^3 + 2\phi\psi & \phi^2 + \psi & \phi & 1 \end{bmatrix}. \tag{15}$$

Note that the non-zero elements of the i^{th} row of this matrix contain the coefficients $\varphi_{i-1}, ..., \varphi_0$, where $(\varphi_0 + \varphi_1 L + \varphi_2 L^2 + ...) = (1 - \phi L - \psi L^2)^{-1}$ with $\varphi_0 = 1$. Hence, the matrix U can be interpreted as the matrix analogue of the scalar lag-polynomial $(1 - \phi L - \psi L^2)^{-1}$. An explicit expression for the elements of U is given by

$$U_{ij} = \begin{cases} (-1)^{(i-j)} \psi^{(i-j+1)} \left(\lambda_1^{(i-j+1)} - \lambda_2^{(i-j+1)} \right) / \sqrt{\phi^2 + 4\psi} & i \geq j, \\ 0 & \text{otherwise,} \end{cases} \tag{16}$$

with

$$\lambda_1 = \frac{-\phi + \sqrt{\phi^2 + 4\psi}}{2\psi} \qquad \text{and} \qquad \lambda_2 = \frac{-\phi - \sqrt{\phi^2 + 4\psi}}{2\psi}. \tag{17}$$

The scalars λ_1 and λ_2 are the roots of the polynomial given by (12). In addition, define the matrices

$$\begin{array}{lcl} V = U L_T & \text{and} & H = \beta_0 U + \beta_1 V, \\ W = V L_T & \text{and} & C = H L_T = \beta_0 V + \beta_1 W, \\ G = U - \gamma V & \text{and} & J = U - \lambda V, \\ P = G L_T = V - \gamma W & \text{and} & S = J L_T = V - \lambda W. \end{array} \tag{18}$$

The matrix G corresponds to the matrix analogue of the scalar lag-polynomial $(1 - \phi L - \psi L^2)^{-1}(1 - \gamma L)$, while H equals the matrix analogue of $(1 - \phi L - \psi L^2)^{-1}(\beta_0 + \beta_1 L)$, *et cetera*. Similarly to (9), the vector y_{-1} may be written as

$$y_{-1} = \theta P \iota_T + P\varepsilon + C\eta + Ge_1 y_0 + (\gamma\beta_0 + \beta_1) V e_1 x_0, \tag{19}$$

where $e_1 = (1, 0, ..., 0)'$ denotes a $T \times 1$ unit vector. Although the second-order bias is likely to involve the starting values y_0 and x_0, see for instance Kiviet and Phillips (1998) for the ARX(1) model with fixed regressors, the first-order bias does not depend on the starting values (provided that the initial values are finite). Hence, we shall ignore them in the remainder of the analysis. Apart from the starting values, the other two regressors can be written as

$$x = \theta\pi V \iota_T + \pi V\varepsilon + J\eta \qquad \text{and} \qquad x_{-1} = \theta\pi W \iota_T + \pi W\varepsilon + S\eta. \tag{20}$$

In order to distinguish the deterministic and stochastic part of the matrix of regressors Z, decompose $Z = \bar{Z} + \tilde{Z}$, where $\bar{Z}$ is defined as the mathematical expectation of Z, so that

$$\begin{aligned} Z &= \mathbb{E}[Z] + (Z - \mathbb{E}[Z]) \\ &= \bar{Z} + \tilde{Z}. \end{aligned} \tag{21}$$

From equations (19) and (20), it follows that the non-stochastic matrix $\bar{Z}$ and stochastic matrix $\tilde{Z}$ are given by

$$\bar{Z} = (\theta P \iota_T : \theta\pi V \iota_T : \theta\pi W \iota_T : \iota_T) \tag{22}$$

and

$$\tilde{Z} = (P\varepsilon + C\eta : \pi V\varepsilon + J\eta : \pi W\varepsilon + S\eta : 0). \tag{23}$$

For notational convenience, we shall denote the inverse of $\mathbb{E}[Z'Z] = \bar{Z}'\bar{Z} + \mathbb{E}[\tilde{Z}'\tilde{Z}]$ by Q, *i.e.* $Q = (\bar{Z}'\bar{Z} + \mathbb{E}[\tilde{Z}'\tilde{Z}])^{-1}$. The Nagar-type expansion, named after Nagar (1959), that is utilised in this paper follows from the identity

$$(Z'Z)^{-1} = Q\left[I + (\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q + (\tilde{Z}'\tilde{Z} - \mathbb{E}[\tilde{Z}'\tilde{Z}])Q\right]^{-1}, \tag{24}$$

where the stochastic terms $(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q$ and $(\tilde{Z}'\tilde{Z} - \mathbb{E}[\tilde{Z}'\tilde{Z}])Q$ both are $O_p(T^{-1/2})$. The inverse of the form $(I + A)^{-1}$ with $A = O_p(T^{-1/2})$ may be expanded in $(I - A + A^2 - A^3 ...)$, with successive terms that are of decreasing order in probability. The two theorems stated in this paper are based on the next lemma, which is proved in, *e.g.*, Kiviet and Phillips (1998, Appendix B).

Lemma 1 *The bias in the AD(1,1) model shown in (5) is given by*

$$B_\alpha(T^{-1}) = \mathbb{E}[Q\tilde{Z}'\varepsilon - Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q\bar{Z}'\varepsilon - Q(\tilde{Z}'\tilde{Z} - \mathbb{E}[\tilde{Z}'\tilde{Z}])Q\bar{Z}'\varepsilon] + o(T^{-1}). \tag{25}$$

Lemma 1 holds quite general and does not depend on the exogeneity status of the regressors. On the other hand, the content of the matrices $\bar{Z}$ and $\tilde{Z}$ does depend on the exogeneity assumption.

3 Bias Approximation in the AD(1,1) Model

The starting point for our analysis can be summarised as follows.

Assumption 1 *In the system consisting of the conditional model (3) and marginal model (4), we have*

- (i) $|\lambda| < 1$,
- (ii) *given* $(\lambda, \beta_0, \beta_0)$, *the parameters* (π, γ) *satisfy the three inequalities* $(13a) - (13c)$,
- (iii) $\varepsilon \sim N(0, \sigma_\varepsilon^2 I_T)$, *with* $0 < \sigma_\varepsilon^2 < \infty$,
- (iv) $\eta \sim N(0, \sigma_\eta^2 I_T)$, *with* $0 < \sigma_\eta^2 < \infty$,
- (v) ε *and* η *are mutually independent,*
- (vi) y_0 *and* x_0 *are finite and fixed.*

The assumptions (*i*) and (*ii*) ensure the stability of the conditional model, while assumption (*v*) formally states the weak exogeneity assumption. It turns out that the bias $B_\alpha(T^{-1})$ can be decomposed into a bias component due to the inclusion of a constant (c), denoted by $B_\alpha^c(T^{-1})$, and another bias component due to the stochastic regressors (r), denoted by $B_\alpha^r(T^{-1})$. Hence, we have

$$B_\alpha(T^{-1}) = B_\alpha^c(T^{-1}) + B_\alpha^r(T^{-1}), \tag{26}$$

and the bias of α to $O(T^{-1})$ in the estimation model without intercept is given by the bias component $B^r_\alpha(T^{-1})$ only. Using Lemma 1 and tedious but straightforward algebra (done by Mathematica 4.0), the following result is derived in Appendix B.

Theorem 1 *Under Assumption 1, the bias B_α of the OLS estimator $\hat{\alpha}$ in model (5) can be approximated to first order as*

$$\begin{aligned}
B^c_\lambda(T^{-1}) &= \omega_1\{-(1+\lambda)(-1+\gamma\lambda)^2\sigma^2_\varepsilon\sigma^2_\eta \\
&\quad +\pi\sigma^2_\varepsilon[-\pi(1-\gamma\lambda+\pi\beta_1)\sigma^2_\varepsilon+(\lambda\beta_0(1-\gamma\lambda+\pi\beta_1) \\
&\qquad +\beta_1((1+2\lambda)(\gamma\lambda-1)-\pi\lambda\beta_1))\sigma^2_\eta]\}\,T^{-1}+o(T^{-1}), \quad (27)\\
B^r_\lambda(T^{-1}) &= \omega_1\{-(\gamma\lambda-1)(\gamma(3\lambda^2-1)-2\lambda)\sigma^2_\varepsilon\sigma^2_\eta \\
&\quad -\pi\sigma^2_\varepsilon[\pi(\gamma+\lambda+\pi\beta_0)\sigma^2_\varepsilon \\
&\qquad +(\lambda\beta_1(4-6\gamma\lambda+3\pi\beta_1)-\pi\lambda\beta_0^2-(2\gamma\lambda-1+\lambda^2)\beta_0)\sigma^2_\eta]\}\,T^{-1}+o(T^{-1}), \quad (28)\\
B^c_{\beta_0}(T^{-1}) &= 0+o(T^{-1}), \quad (29)\\
B^r_{\beta_0}(T^{-1}) &= 0+o(T^{-1}), \quad (30)\\
B^c_{\beta_1}(T^{-1}) &= \omega_1\{-(1+\lambda)(\gamma\lambda-1)(\beta_0+\gamma\beta_1)\sigma^2_\varepsilon\sigma^2_\eta \\
&\quad +\sigma^2_\varepsilon[(1+\lambda)(\gamma\lambda-1)(\beta_0+\gamma\beta_1)\sigma^2_\eta-(\gamma\lambda-1-\pi\beta_1) \\
&\quad \times(-\pi(1+\gamma)\sigma^2_\varepsilon+(\beta_1(\gamma+\gamma\lambda-\pi\beta_1)+\beta_0(1+\lambda+\pi\beta_1))\sigma^2_\eta)]\}\,T^{-1}+o(T^{-1}), \quad (31)
\end{aligned}$$

$$\begin{aligned}
B^r_{\beta_1}(T^{-1}) &= \omega_1\{-(\gamma(3\lambda^2-1)-2\lambda)(\beta_0+\gamma\beta_1)\sigma^2_\varepsilon\sigma^2_\eta \\
&\quad -\pi\sigma^2_\varepsilon[(1+\gamma^2-2\gamma\lambda+\pi\gamma\beta_0+3\pi\beta_1)\sigma^2_\varepsilon \\
&\qquad -(2(\gamma+2\lambda)\beta_0\beta_1+\beta_1^2(6\gamma\lambda-1-3\pi\beta_1)+\beta_0^2(1+\pi\beta_1))\sigma^2_\eta]\}\,T^{-1}+o(T^{-1}), \quad (32)\\
B_\theta(T^{-1}) &= \omega_2\{(\theta\sigma^2_\varepsilon(\pi^2(3+\gamma+\lambda-5\gamma\lambda+\pi\beta_0+5\pi\beta_1)\sigma^2_\varepsilon-((-1+\gamma)(-1+\gamma\lambda) \\
&\quad \times(-1-3\lambda+\gamma(-1+\lambda+4\lambda^2))+\pi(\pi\beta_0^2(1+\lambda-\gamma\lambda+\pi\beta_1) \\
&\quad +\beta_1(-1-6\lambda+\gamma(2+\gamma-2(-5+\gamma)\lambda-4(-2+3\gamma)\lambda^2) \\
&\quad +\pi\beta_1(-2+\gamma-4\lambda+12\gamma\lambda-4\pi\beta_1))+\beta_0(\gamma^2(-2+\lambda)\lambda+\lambda(4+\lambda)+\gamma(2-6\lambda^2) \\
&\quad +\pi\beta_1(2+2\gamma+6\lambda-2\gamma\lambda+\pi\beta_1))))\sigma^2_\eta))\}\,T^{-1}+o(T^{-1}) \quad (33)
\end{aligned}$$

where $\omega_1=\{\pi^2\sigma^4_\varepsilon+((-1+\gamma\lambda)^2-2\pi\lambda\beta_0-2\pi\gamma\lambda\beta_1+\pi^2\beta_1^2)\sigma^2_\varepsilon\sigma^2_\eta+(\lambda\beta_0+\beta_1)^2\sigma^4_\eta\}^{-1}$ *and* $\omega_2=\{(\gamma-1)(\lambda-1)-\pi(\beta_0+\beta_1)\}^{-1}\omega_1$.

To facilitate the reading of the formulae, the bias expressions (except for the constant) are decomposed in expressions not involving the feedback parameter π (first line) and expressions that do involve the feedback parameter π (other than first line). So, in the absence of a feedback effect ($\pi=0$), the first-order bias can be directly read off from the first line of the approximations. Note that the elementary bias formulae given in the introduction are recovered if the appropriate substitutions are made, *i.e.* $B^r_\lambda(T^{-1})=-2\lambda/T+o(T^{-1})$ and

$B_\lambda(T^{-1}) = -(1+3\lambda)/T + o(T^{-1})$ for $\beta_0 = \beta_1 = \pi = \gamma = 0$. Furthermore, it turns out that the expressions for the bias of the stochastic regressors do not depend on the value of the intercept. Hence, they are invariant with respect to parameter θ. Finally, it appears that the OLS estimator $\hat{\beta}_0$ is unbiased to first order, which is in contrast to the other three coefficient estimators.

4 Bias Approximation in the AD(1,0) Model

In this section, we shall consider the slightly more specific AD(1,0) model, *i.e.* without the lagged exogenous variable. To distinguish the AD(1,0) model from the AD(1,1) model, the coefficient β_0 is replaced by β, so that the estimation model becomes

$$y = \lambda y_{-1} + \beta x + \theta + \varepsilon, \tag{34}$$

while the marginal model remains unchanged. In matrix notion, we obtain

$$y = Z\alpha + \varepsilon, \qquad Z = (y_{-1} : x : \iota_T), \tag{35}$$

where $\alpha = (\lambda, \beta, \theta)'$. Although the AD(1,1) model encompasses the AD(1,0) model, the results for the bias do not result directly from Theorem 1 for $(\beta_0, \beta_1) = (\beta, 0)$, since the matrices $(Z'Z)$ and Q for the AD(1,0) model are not submatrices of their counterparts in the AD(1,1) model.

Using the results in Lemma 1 and straightforward algebra, the following result can be derived (see Appendix C for some additional remarks).

Theorem 2 *Under Assumption 1 for $(\beta_0, \beta_1) = (\beta, 0)$, the bias B_α of the OLS estimator $\hat{\alpha}$ in model (35) can be approximated to first order as*

$$\begin{aligned}
B_\lambda^c(T^{-1}) &= \omega_3\{-(1+\lambda)(-1+\gamma\lambda)^2\sigma_\varepsilon^2\sigma_\eta^2 \\
&\quad -\pi(\gamma\lambda-1)\sigma_\varepsilon^2(\pi\gamma\sigma_\varepsilon^2 - \beta\sigma_\eta^2)\}\,T^{-1} + o(T^{-1}), &(36)\\
B_\lambda^r(T^{-1}) &= \omega_3\{-(\gamma\lambda-1)(\gamma(3\lambda^2-1) - 2\lambda)\sigma_\varepsilon^2\sigma_\eta^2 \\
&\quad +\pi\sigma_\varepsilon^2[\pi\gamma(1-3\gamma\lambda)\sigma_\varepsilon^2 + 2\beta(2\gamma\lambda-1)\sigma_\eta^2]\}\,T^{-1} + o(T^{-1}), &(37)\\
B_\beta^c(T^{-1}) &= \omega_3\{-\beta\gamma(1+\lambda)(\gamma\lambda-1)\sigma_\varepsilon^2\sigma_\eta^2 \\
&\quad +\pi\gamma(1+\gamma)(\gamma\lambda-1)\sigma_\varepsilon^4\}\,T^{-1} + o(T^{-1}), &(38)\\
B_\beta^r(T^{-1}) &= \omega_3\{\beta\gamma(\gamma+2\lambda-3\gamma\lambda^2)\sigma_\varepsilon^2\sigma_\eta^2 \\
&\quad -\pi\gamma\sigma_\varepsilon^2((1+\pi\beta\gamma+\gamma^2-2\gamma\lambda)\sigma_\varepsilon^2 - \beta^2\sigma_\eta^2)\}\,T^{-1} + o(T^{-1}), \\
B_\theta(T^{-1}) &= \omega_4\{-(\theta\sigma_\varepsilon^2(\pi^2\gamma^2(3+\pi\beta+\gamma+\lambda-5\gamma\lambda)\sigma_\varepsilon^2 \\
&\quad -(\pi^2\beta^2\gamma + (-1+\gamma)(-1+\gamma\lambda)(-1-3\lambda+\gamma(-1+\lambda+4\lambda^2)) \\
&\quad +\pi\beta(-3+\gamma(4+\gamma+8\lambda-6\gamma\lambda-4\gamma\lambda^2)))\sigma_\eta^2))\}\,T^{-1} + o(T^{-1}). &(39)
\end{aligned}$$

where $\omega_3 = \{\pi^2\gamma^2\sigma_\varepsilon^4 + (-2\pi\beta\gamma + (-1+\gamma\lambda)^2)\sigma_\varepsilon^2\sigma_\eta^2 + \beta^2\sigma_\eta^4\}^{-1}$ and $\omega_4 = \{-1+\pi\beta+\gamma+\lambda-\gamma\lambda\}^{-1}\omega_3$.

The bias expressions in Theorem 2 are considerably more simple than in Theorem 1. Moreover, the β coefficient estimator is biased. Since the results of Theorem 1 continue to hold when the regressor x_{t-1} is redundant, *i.e.* $\beta_1 = 0$, we conclude that β can be estimated unbiasedly up to first order by estimating the overparameterised AD(1,1) model. So, in this particular model, the finite-sample problems with respect to the estimation bias of β can be reduced by including a redundant regressor, *viz.* x_{t-1}.

Grubb and Symons (1987) studied the bias of the AR(1) parameter in model (34) without intercept when there is no feedback effect ($\pi = 0$). For $\pi = 0$, formula (37) reduces to

$$\begin{aligned} B_\lambda^r(T^{-1}) &= \frac{(1-\gamma\lambda)(\gamma(-1+3\lambda^2)-2\lambda)\sigma_\varepsilon^2}{(-1+\lambda\gamma)^2\sigma_\varepsilon^2+\beta^2\sigma_\eta^2}T^{-1}+o(T^{-1}) \\ &= \frac{-(1-\lambda^2)\gamma/(1-\lambda\gamma)-2\lambda}{1+\beta^2(1-\lambda\gamma)^{-2}\sigma_\eta^2\sigma_\varepsilon^{-2}}T^{-1}+o(T^{-1}). \end{aligned} \tag{40}$$

which corresponds to formula (18) of Grubb and Symons (1987) for $(\gamma, \sigma_\varepsilon^2, \sigma_\eta^2) = (\rho, \sigma_u^2, \sigma_\varepsilon^2)$.

5 Approximation to the Strong Exogenous Bias Formula

In this section, we shall investigate the behaviour of the bias-correction formula (40) assuming strongly exogenous regressors, *i.e.* no feedback mechanism is assumed, when in fact the regressors are only weakly but not strongly exogenous, *i.e.* there does exist a feedback mechanism. To this end, the bias formula ignoring feedback mechanisms, which was first derived by Grubb and Symons (1987), is studied under a DGP that does involve a feedback mechanism.

Consider the dynamic regression model

$$y = \lambda y_{-1} + X\beta + \varepsilon, \qquad \varepsilon \sim N(0, \sigma_\varepsilon^2 I_T), \tag{41}$$

where X is a $T \times k$ matrix of fixed regressors, β denotes here a vector of unknown parameters and $|\lambda| < 1$. In model (41), only y_{-1} is a stochastic regressor, which can be decomposed into

$$y_{-1} = y_0 F + CX\beta + C\varepsilon = y_D + C\varepsilon, \tag{42}$$

where F denotes a $T \times 1$ vector equal to the first column of matrix U given in (14) for $(\phi, \psi) = (\lambda, 0)$ and C denotes a $T \times T$ matrix equal to the matrix $V = UL_T$ for $(\phi, \psi) = (\lambda, 0)$. Kiviet and Phillips (1993) show that the bias of the vector $\alpha' = (\lambda, \beta')$ under the assumption that the regressors are strongly exogenous can be approximated by

$$\begin{aligned} B_\alpha^{KPh}(T^{-1}) &= -\sigma_\varepsilon^2 \bar{D}^{-1}\{\bar{Z}'C\bar{Z}\bar{D}^{-1}e_1 + Tr(\bar{Z}'C\bar{Z}\bar{D}^{-1})e_1 \\ &\quad +2\sigma_\varepsilon^2 e_1'\bar{D}^{-1}e_1 Tr(CC'C)e_1\} + o(T^{-1}), \end{aligned} \tag{43}$$

where $\bar{Z} = [y_D : X]$ and $\bar{D} = \bar{Z}'\bar{Z} + \sigma_\varepsilon^2 Tr(C'C)e_1e_1'$.

Since explicit expressions become quickly rather lengthy, we shall focus on the estimation bias in the AD(1,0) model without intercept

$$y = \lambda y_{-1} + \beta x + \varepsilon, \qquad x \text{ strongly exogenous}, \tag{44}$$

where β is now a scalar and $|\lambda| < 1$. Define the scalars $G_m = T^{-1}\sum_{t=1}^{T-m} x_{t+m}x_t$ for $m = 0, ..., T-1$, where x_t denotes the t-th element of x. G_0 is the sample variance $T^{-1}(x'x)$ of the exogenous variable, while G_m $(m \neq 0)$ consist of the m-th order sample autocovariance of x. In addition, let $\Lambda = \sum_{m=1}^{T-1} \lambda^m G_m$ and $\dot{\Lambda} = \sum_{m=1}^{T-1} m\lambda^{m-1} G_m$. Grubb and Symons (1987) show that equation (43) for the OLS estimator $\hat{\lambda}$ in model (44) can be written as

$$B_\lambda^{GS}(T^{-1}) = -\sigma_\varepsilon^2\left(\frac{(1-\lambda^2)\lambda^{-1}Tr(G_0^{-1}\Lambda)+2\lambda}{(\sigma_\varepsilon^2+\beta' H_0\beta)} - \frac{2\lambda^{-3}(1-\lambda^2)\beta' H_1 H_2\beta}{(\sigma_\varepsilon^2+\beta' H_0\beta)^2}\right)T^{-1} + o(T^{-1}), \tag{45}$$

where

$$H_0 = G_0 + 2\Lambda - \lambda^{-2}(1-\lambda^2)2\Lambda/G_0, \tag{46a}$$

$$H_1 = \lambda^2 I - (1-\lambda^2)\Lambda/G_0, \tag{46b}$$

$$H_2 = \Lambda + \Lambda^2/G_0 - \lambda\dot{\Lambda}. \tag{46c}$$

Equation (45) differs slightly from equation (11) of Grubb and Symons (1987), since they omit the effect of σ_ε^2 and the second part of their formula appears to have a wrong sign.

Suppose that, contrary to the condition that x is a fixed regressor, the marginal model for x is given by equation (4), so that the model for x contains a feedback mechanism. Hence, formula (45) is in error since it was derived under the assumption of no feedback effect. However, it seems interesting to compare the behaviour of the erroneous bias formula, *i.e.* $B_\lambda^{GS}(T^{-1})$, with that of the correct bias formula given in Theorem 2, *i.e.* $B_\lambda^r(T^{-1})$. To this end, closed-form approximations in the underlying parameters $(\lambda, \beta, \gamma, \pi, \sigma_\varepsilon^2, \sigma_\eta^2)$ are obtained for the quantities G_0, Λ, $\dot{\Lambda}$, H_0, H_1 and H_0 under the assumption that x is generated by equation (4). The approximations for G_0, Λ and $\dot{\Lambda}$ are derived in Appendix D. Substitution of these expressions in (46a)-(46c), lead to the following approximations

$$H_0 = \frac{\pi^4(-1+\gamma^2\lambda^4)\sigma_\varepsilon^2 - 2\pi^2\kappa(-1+\gamma\lambda^3)\sigma_\varepsilon^2\sigma_\eta^2 + \kappa^2(-1+\lambda^2)\sigma_\eta^4}{\kappa^2(-1+\gamma\lambda)(\pi^2(1+\gamma\lambda)\sigma_\varepsilon^2 - (\kappa+\lambda\pi\beta)\sigma_\eta^2)} + O_p(T^{-1/2}), \tag{47a}$$

$$H_1 = \lambda\frac{\pi^2(\gamma(-1+\lambda^2)^2 + \pi(\beta+\beta\gamma\lambda^3))\sigma_\varepsilon^2 - \kappa(\pi\beta+\gamma-\lambda+(\pi\beta-\gamma)\lambda^2+\lambda^3)\sigma_\eta^4}{\kappa(\pi^2(1+\gamma\lambda)\sigma_\varepsilon^2 - (\kappa+\lambda\pi\beta)\sigma_\eta^2)} + O_p(T^{-1/2}), \tag{47b}$$

$$H_2 = \pi\lambda^3\frac{-\pi^3\gamma\sigma_\varepsilon^4 + \pi((\gamma-\lambda)(-1+\gamma\lambda)+\pi(\beta+\beta\gamma\lambda))\sigma_\varepsilon^2\sigma_\eta^2 - \kappa\beta\sigma_\eta^4}{\kappa(\pi^2(1+\gamma\lambda)\sigma_\varepsilon^2 - (\kappa+\lambda\pi\beta)\sigma_\eta^2)} + O_p(T^{-1/2}), \tag{47c}$$

where $\kappa \equiv -1 + \lambda(\phi + \lambda\psi) = -1 + \lambda(\pi\beta + \gamma + \lambda - \gamma\lambda^2)$. Although it is possible to give a closed-form approximation to the bias formula $B_\lambda^{GS}(T^{-1})$ in the underlying parameters, the formula is not shown here since it is rather lengthy. Of course, this is a result from the fact that the bias formula $B_\lambda^{GS}(T^{-1})$ is evaluated under a probability model that differs from regression equation (44) in which it was derived. When there is no feedback effect, *i.e.* $\pi = 0$, the bias formula $B_\lambda^r(T^{-1})$ shown in Theorem 2 reduces to $B_\lambda^{GS}(T^{-1})$. In general, however, the two bias formulae are different. The numerical discrepancy between the two formulae for particular values of the underlying parameters is illustrated in the next section, which reports some numerical and simulation results.

6 Some Numerical and Simulation Results

In this section, the various bias approximations derived in this paper are illustrated and compared for several values of the population parameters. The focus is on the dependence of the bias formulae on the parameters (π, γ), which appear in the marginal model. In addition, the finite-sample performance of two different bias-corrected estimators is examined through Monte Carlo experiments. Section 6.1 considers the AD(1,0) model, while Section 6.2 considers the more general AD(1,1) model. In both sections, we chose to focus on just a few particular but empirically relevant number of parametrisations to keep the number of figures and tables manageable.

6.1 The AD(1,0) model

The parameters of the AD(1,0) model $y = \lambda y_{-1} + \beta x + \varepsilon$ are chosen as

$$(\lambda, \beta) = (0.8, 0.2), \tag{48}$$

so that the long-run multiplier of x with respect to y is equal to 1. For this particular set of values, the three inequalities given in (13a)-(13c) lead to the triangular admissible region of (π, γ) given by

$$\gamma < 1 - \pi, \tag{49a}$$

$$\gamma > -1 - 1/9\pi, \tag{49b}$$

$$\gamma < 5/4. \tag{49c}$$

Without loss of generality, the variance of the marginal model σ_η^2 is normalized to 1. To make the outcomes more comparable when varying π and γ, we shall control for the signal-to-noise ratio. The population R^2 of the conditional model for y_t, *i.e.*

$$R^2 = 1 - \frac{\sigma_\varepsilon^2}{\mathrm{Var}(y_t)}, \tag{50}$$

is set to 0.8. In addition, σ_ε^2 is set to 0.1 since the bias aggravates as σ_ε^2 increases. All these restrictions lead to the following relationship between π and γ

$$\frac{5 + 4\gamma}{5(4(5 - 4\gamma - 5\gamma^2 + 4\gamma^3) + 8(-5 - \gamma + 5\gamma^2)\pi + (-5 + 4\gamma)\pi^2)} = 0.1. \tag{51}$$

The graph in Figure 1 illustrates this implicit relationship as π is varied over the interval $(-11.3, 0.6)$. Below, we shall examine the various bias formulae along Line 1 and Line 2; these two lines are shown in Figure 1.

Insert Figure 1 about here.

Figure 2 shows the first-order bias which was derived under the assumption of strong exogeneity, *i.e.* $-T\, B_\lambda^{GS}(T^{-1})$ based on the approximations shown in (47a)-(47c), and the bias which was derived under the assumption of only weak exogeneity, *i.e.* $-T\, B_\lambda^{r}(T^{-1})$ based on the approximation shown in Theorem 2 , along Line 1. From this figure, we observe that $-T\ B^{r}_{\lambda}(T^{-1})$ varies more widely than $-T\ B^{GS}_{\lambda}(T^{-1})$. Furthermore, the two functions intersect at several points; one of them being of course $\pi = 0$. Moreover, in the area $\pi \in (-11.0, -6.6)$, $\hat{\lambda}$ is upward biased although λ is positive. This is in contrast to the simple AR(1) model where the OLS estimator is downward biased when λ is positive. Figure 3 shows the two bias formulae along Line 2. Again the discrepancy between the two formulae can be substantial. For $\pi < -0.9$, the incorrect bias formula $-T\ B^{GS}_{\lambda}(T^{-1})$ is smaller than the correct formula $-T\ B^{r}_{\lambda}(T^{-1})$. Hence, the bias formula ignoring a possible feedback effect is expected to undercorrect for the estimation bias when π is sufficiently negative.

Insert Figures 2 and 3 about here.

Next, we shall examine the bias and efficiency of two bias-corrected estimators for the AR(1) coefficient estimator $\hat{\lambda}$. Since these estimators are obtained by replacing population parameters by their estimators, the analysis is based on Monte Carlo experiments. Theorem 2 can be used to define a bias-corrected (BC) least-squares estimator for $\alpha = (\lambda, \beta)'$ in the AD(1,0) model as follows:

$$\breve{\alpha}(T^{-1}) = \hat{\alpha} - \hat{B}_{\alpha}(T^{-1}), \tag{52}$$

where $\hat{B}_{\alpha}(T^{-1})$ is $B_{\alpha}(T^{-1})$ with population values replaced by OLS estimators. Hence, the population parameters of the conditional model $(\lambda, \beta, \sigma^2_{\varepsilon})$ are replaced by the OLS estimators $(\hat{\lambda}, \hat{\beta}, s^2_{\varepsilon})$, the parameters of the marginal model $(\pi, \gamma, \sigma^2_{\eta})$ are replaced by the OLS estimators $(\hat{\pi}, \hat{\gamma}, s^2_{\eta})$ and the approximation term $o(T^{-1})$ is set to zero. Note that the BC estimator based on Theorem 2 requires the analysis of the full system rather than the partial conditional model alone. The need to analyse the full system is a serious drawback of this bias-correction procedure because this makes is more vulnerable to specification errors with respect to the marginal model. If one only considers the conditional model and is willing to ignore the marginal model, a BC estimator can also be obtained by the bias approximation $B^{KPh}_{\alpha}(T^{-1})$ derived by Kiviet and Phillips (1993), which is reproduced in equation (43). This latter procedure is incorrect when the system does contain a feedback mechanism, *i.e.* $\pi \neq 0$. Since both bias formulae are derived under the assumption $|\lambda| < 1$, it seems sensible to truncate $\hat{\lambda}$ to 0.99 when the event $\hat{\lambda} > 0.99$ occurs in a particular sample.

The bias and mean squared error of the OLS and BC estimators $\hat{\alpha}$ and $\breve{\alpha}$ are compared by Monte Carlo experiments. In the simulations, we chose $\pi = \{-10, -8, ..., 0\}$ and γ such that $R^2 = 0.8$ and $\sigma^2_{\varepsilon} = 0.1$, *i.e.* on the two lines shown in Figure 1. The simulation results are based on 10,000 ($= N$) replications. Since the first-order bias does not depend upon the starting values, we choose $(y_0, x_0) = (0, 0)$. In each replication, $\hat{\alpha}$ and $\breve{\alpha}$ are estimated on a sample of size $T = 25$. A Monte Carlo estimator of the bias for $\hat{\alpha}$ is given by

$$\text{Bias}(\hat{\alpha}) = N^{-1} \sum_{i=1}^{N} \hat{\alpha}_i - \alpha, \tag{53}$$

where $\hat{\alpha}_i$ denotes the OLS estimates obtained in the i-th Monte Carlo replication. A Monte Carlo estimator of the mean squared error (MSE) for $\hat{\alpha}$ is given by

$$\text{MSE}(\hat{\alpha}) = N^{-1} Diag \left(\sum_{i=1}^{N} (\hat{\alpha}_i - \alpha)(\hat{\alpha}_i - \alpha)' \right), \tag{54}$$

where the function $Diag(\cdot)$ selects the diagonal elements. The bias and MSE for $\breve{\alpha}$ are defined analogously. All simulations were carried out on a personal computer using Gauss 2.2.

Insert Tables 1 and 2 about here.

Table 1 shows the simulation results with respect to $\hat{\lambda}$ in the AD(1,0) model without intercept. Since $\breve{\lambda}$ based on the correct bias approximation is virtually unbiased, it seems that, on average, the bias approximation derived in Theorem 2 and shown in Figure 2 closely corresponds to the actual bias. As observed before, there is a region for which the bias is positive; see the results for $(\pi, \gamma) = (-10, 0.258)$ and $(\pi, \gamma) = (-8, -0.033)$. In this region, the bias correction ignoring any feedback effects does not seem to pick up the bias correctly but instead aggravates it. For almost every combination of (π, γ), the BC estimator based on the whole system is more unbiased than the BC estimator based on the conditional model only. Judge by the efficiency criterion, however, no clear-cut picture emerges from the reported results. Sometimes the BC estimator based on the whole system is more efficient than the BC estimator based on the partial conditional model, although not always. More importantly, in just two instances, the BC estimator based on Theorem 2 is more efficient than the uncorrected estimator. Surprisingly, the BC estimator based on only the conditional model seems to be more efficient for a larger range of (π, γ)-values.

Table 2 shows the simulation results when a redundant intercept is added to the estimation model ($\theta = 0$). Comparing these numbers with the results shown in Table 1, we conclude that the bias has enlarged. With respect to the (remaining) bias of both BC estimators, the results are qualitatively the same. As before, the reduction in bias can be very substantial especially for the BC estimator based on the whole system, although the bias reduction is less successful than in the model without intercept. With respect to the efficiency, we now observe that there is a larger range for which the BC estimators are more efficient than the uncorrected estimator. Again, there is no global winner since there are combinations of (π, γ) where the BC estimator based on the whole system is more efficient than the BC based on the conditional model and *vice versa*.

Overall, we conclude that the BC estimator based on the whole system exhibits the least bias. Somewhat unexpectedly, the incorrect BC estimator ignoring any feedback effects is comparable to the correct BC estimator taking feedback effects into account in terms of efficiency. These two findings lead us to conclude that the effectiveness of the highly non-linear bias formula assuming weakly exogenous regressors is limited when the population parameters are replaced by their estimates. Furthermore, the region in which an efficiency gain is achieved seems to be limited.

6.2 The AD(1,1) Model

Next, we consider the AD(1,1) model including an intercept with parameters that are chosen to correspond closely to particular estimates obtained in practice. In accordance with an annual consumption function, *cf.* Hendry (1983), we chose

$$(\lambda, \beta_0, \beta_1) = (0.8, 0.5, -0.3). \tag{55}$$

The parametrisation in (55) gives rise to a total multiplier of x with respect to y of $1 = (\beta_0 + \beta_1)/(1 - \lambda)$ and an immediate impact multiplier of $0.5 = \beta_0$. For the marginal model, we took $\sigma_\eta^2 = 1$ without loss of generality. Since the bias of the OLS estimators $\hat{\lambda}$, $\hat{\beta}_0$ and $\hat{\beta}_1$ are invariant to θ (see Theorem 1), we took $\theta = 0$. The parameter σ_ε^2 is chosen in such a way that the population R^2 of the conditional model for y_t given in (3) equals 0.8; *cf.* Hendry (1983, eq. 13). By the decomposition given in (19), it follows that the multipliers of the variance of y_t with respect to σ_ε^2 and σ_η^2 are equal to the leading terms of respectively $Tr(P'P)$ and $Tr(C'C)$, so that

$$\text{Var}(y_t) = \frac{-1 - \gamma^2 + 2\gamma\phi + \psi + \gamma^2\psi}{(1 + \phi - \psi)(1 + \psi)(-1 + \phi + \psi)}\sigma_\varepsilon^2 + \frac{-\beta_0^2 + \psi\beta_0^2 - 2\phi\beta_0\beta_1 - \beta_1^2 + \psi\beta_1^2}{(1 + \phi - \psi)(1 + \psi)(-1 + \phi + \psi)}\sigma_\eta^2. \tag{56}$$

Substitution of (56) into $R^2 = 1 - \sigma_\varepsilon^2/\text{Var}(y_t)$, evaluated and simplified at $(\lambda, \beta_0, \beta_1, \theta, R^2, \sigma_\eta^2) = (0.8, 0.5, -0.3, 0.0, 0.8, 1)$ gives

$$\sigma_\varepsilon^2 = \frac{25 - 12\pi - 7\gamma}{5(12\pi^3 + \pi^2(71\gamma - 25) + 4\pi(-23 - 10\gamma + 29\gamma^2) + 8(5 - 4\gamma - 5\gamma^2 + 4\gamma^3))}. \tag{57}$$

For our particular choice of $(\lambda, \beta_0, \beta_1)$, the system is stable when

$$\gamma < 1 - \pi, \tag{58a}$$

$$\gamma > -1 - \tfrac{4}{9}\pi, \tag{58b}$$

$$\gamma < \tfrac{5}{4} - \tfrac{3}{8}\pi. \tag{58c}$$

This triangular region is shown in Figure 4.

Insert Figure 4 about here.

Figure 5 shows $-T$ times the first-order bias for the AR(1) coefficient as function of the parameters of the marginal model under the additional restriction that $|\pi| < 1$ and $|\gamma| < 1$. From this figure we observe that there are indeed regions where the bias of the AR(1)-coefficient estimator depends heavily on the values of the parameters in the marginal model, *i.e.* (π, γ). The bias seems to be most severe when π and γ are both positive.

Insert Figure 5 about here.

In the AD(1,1) model, Theorem 1 can be utilised to obtain a BC estimator for $\alpha' = (\lambda, \beta_0, \beta_1, \theta)$ which is valid in case the regressors are weakly exogenous. A first-order BC estimator for α based on strongly exogenous regressors can be obtained by utilising the formula $B_\alpha^{KPh}(T^{-1})$ as shown in (43). As in the previous subsection, the Monte Carlo results are based on 10,000 ($= N$) replications. To examine the effect of the number of observations, two sample sizes $T = 25$ and $T = 50$ are considered.

Insert Tables 3-6 about here.

Tables 3–6 show the simulation results obtained in the Monte Carlo experiments; the first two tables refer to λ, while the last two tables refer to β_1. Since the first-order bias of $\hat{\beta}_0$ is zero, we have that the BC estimator equals the uncorrected estimator, *i.e.* $\breve{\beta}_0 = \hat{\beta}_0$. Hence, the results with respect to the coefficient β_0 are not reported. From Tables 3–6, we observe that $\hat{\lambda}$ is negatively biased, while $\hat{\beta}_1$ is positively biased. The bias of $\hat{\lambda}$ seems to be twice as large in magnitude than the bias of $\hat{\beta}_1$. The bias of $\hat{\lambda}$ for $T = 25$ lies in the range (–0.171,–0.088), while for $T = 50$ it lies in the range (–0.081, –0.050). On average, the BC estimator based on the whole system is slightly less biased than the BC estimator based on the conditional model only. In contrast to the AD(1,0) model, both BC estimators are uniformly more efficient than the uncorrected estimator $\hat{\lambda}$. Comparing this to the results obtained in Section 6.1, the efficiency gain seems to be a consequence of the particular model and parameter values we consider. Surprisingly again, the two BC procedures perform equally well in terms of efficiency, although in 5 out of 17 values for (π, γ) the BC estimator based on the whole system achieves a slightly lower MSE than the BC estimator based on the conditional model only. On average, the reduction in the MSE for both BC estimators seems to be approximately 25%, which is quite high. When the sample size is doubled, the efficiency gain is only slightly less. The bias of $\hat{\beta}_1$ for $T = 25$ lies in the range (0.061, 0.108), while for $T = 50$ it lies in the range (0.029, 0.053). As before, the BC estimators are more efficient than the OLS estimator for all the values of (π, γ) considered in the Monte Carlo experiments. The average reduction in the MSE achieved by the BC estimators equals approximately 15%, which is less than the reduction in MSE achieved by $\breve{\lambda}$. In the majority of the considered (π, γ) values, the BC taking the feedback mechanism into account exhibits less bias than the BC estimator ignoring any feedback effects. Overall, the first-order BC estimators substantially reduce the bias and are more efficient than OLS, at least in the parametrisations we examined.

7 Conclusions

From a practical point of view, it seems interesting to compare the behaviour of a bias-corrected estimator based on only the conditional model to the behaviour of a bias-corrected estimator which takes the marginal model, and hence possibly feedback effects, into account. In order to compare these two estimators, we first derive the least-squares estimation bias in an autoregressive distributed-lag model with weakly exogenous regressors, *i.e.* we allow for the presence of a feedback mechanism. In particular, the estimation is carried out in the AD(1,1)-model

$$y = \lambda y_{-1} + \beta_0 x + \beta_1 x_{-1} + \theta + \varepsilon, \tag{59}$$

where the marginal model for x is given by

$$x = \pi y_{-1} + \gamma x_{-1} + \eta. \tag{60}$$

From the first-order bias formulae, we observe that the bias of the OLS estimators depends explicitly on the value of the parameters of the marginal model (π, γ). Furthermore, we find that the OLS estimator $\hat{\beta}_0$ is unbiased up to first-order. In the AD(1,0)-model, *i.e.* the model without x_{-1}, however, the OLS estimator of β_0 is not unbiased. Hence, we see that a redundant regressor can sometimes reduce the estimation bias in a dynamic regression model.

Next, we have derived an approximation to the first-order bias formula based on strongly exogenous regressors when in fact the regressors are only weakly exogenous. The two bias formulae respond quite differently to changes in the parameters (π, γ). It is found that the sign of the bias formula based on the conditional model only is sometimes erroneous. The finite-sample behaviour of the two bias-corrected estimators is investigated through a small-scale simulation study. In the Monte Carlo experiments, it is found that the bias problem can be rather severe. Fortunately, both bias-corrected estimators reduce the estimation bias considerably. Somewhat surprisingly, none of the two bias-corrected estimators is uniformly more efficient than the other. It turns out that in a particular AD(1,0) model, there is not always an efficiency gain for particular values of (π, γ). When there is an efficiency gain, however, both bias-corrected estimators achieve such a gain. In the parametrisation we consider in the AD(1,1) model, which is designed to mimic situations that are empirically relevant, the efficiency gain can be as large as 25% for the AR(1) coefficient. Judge by the mean squared error criterion, it seems that (on average) the error committed by an practitioner who ignores any possibly feedback effects is small.

Although the bias formulae based on the whole system do not easily generalise to higher-order AD models, one could always obtain a bootstrap approximation of the estimation bias in practice by resampling the whole system. In fact, one could even compare the two bias-corrected estimators by employing different resampling schemes; see Giersbergen and Kiviet (1996) for details. Of course, adoption of the bootstrap approach does not lead to explicit expressions for the bias in the underlying system parameters and is therefore less suitable for the analysis given in this paper.

References

Barlett, M. S. (1946). On the Theoretical Specification and Sampling Properties of Autocorrelated Time-Series. *Journal of the Royal Statistical Society*, **8**, 27–41.

Engle, R. F., D. F. Hendry, and J.-F. Richard (1983). Exogeneity. *Econometrica*, **51**, 277–304.

Giersbergen, N. P. A. v. and J. F. Kiviet (1996). Bootstrapping a Stable AD Model: Weak vs Strong Exogeneity. *Oxford Bulletin of Economics and Statistics*, **58**, 631–656. Reprinted in Banerjee, A. and D. F. Hendry (eds.), 1997, *The Econometrics of Economic Policy*, pp. 61–86. Oxford: Blackwell Publishers.

Grubb, D. and J. Symons (1987). Bias in Regressions with a Lagged Dependent Variable. *Econometric Theory*, **3**, 371–386.

Harvey, A. C. (1981). *Time Series Models*. Philip Allan.

Hendry, D. F. (1983). Econometric Modelling: the 'Consumption Function' in Retrospect. *Scottish Journal of Political Economy*, **30**, 193–220.

Hendry, D. F., A. R. Pagan, and J. D. Sargan (1984). Dynamic Specification. In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics*, Volume 2, pp. 1023–1100. Amsterdam: North-Holland.

Hurwicz, L. (1950). Least Squares Bias in Time Series. In T. Koopmans (Ed.), *Statistical Inference in Dynamic Economic Models*, pp. 365–383. New York: Wiley.

Kendall, M. G. (1954). Note on the Bias in the Estimation of Auto Correlation. *Biometrika*, **61**, 403–404.

Kiviet, J. F. and G. D. A. Phillips (1993). Alternative Bias Approximations in Regressions with a Lagged Dependent Variable. *Econometric Theory*, **9**, 62–80.

Kiviet, J. F. and G. D. A. Phillips (1998). Higher-order Asymptotic Expansions of the Least-squares Estimation Bias in First-order Dynamic Regression Models. Discussion paper TI 96-167/7, Amsterdam: Tinbergen Institute. Revised in October 1998.

Marriot, F. H. C. and J. A. Pope (1954). Bias in the Estimation of Autocorrelations. *Biometrika*, **61**, 393–403.

Nagar, A. L. (1959). The Bias and Moment Matrix of the General k-class Estimators of the Parameters in Simultaneous Equations. *Econometrica*, **27**, 575–595.

White, J. S. (1961). Asymptotic Expansions for the Mean and Variance of the Serial Correlation Coefficient. *Biometrika*, **48**, 85–94.

Wolfram, S. (1991). *Mathematica: a System for Doing Mathematics by Computer*. Addison-Wesley Publishing Company.

Appendix A: Some auxiliary results

Let A be a symmetric $T \times T$ matrix and B_1 and B_2 arbitrary $T \times T$ matrices. In addition, let the $T \times 1$ random vectors ε and η be such that $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2 I_T)$ and $\eta \sim \mathcal{N}(0, \sigma_\eta^2 I_T)$. The following results will be frequently used in Appendix B (here $Tr(\cdot)$ denotes the trace operator):

$$\mathbb{E}[\varepsilon' B_1 \varepsilon] = \sigma_\varepsilon^2 Tr(B_1);$$

$$\mathbb{E}[\varepsilon' A \varepsilon\, \varepsilon' B_1 \varepsilon] = \sigma_\varepsilon^4 [Tr(A)\, Tr(B_1) + 2\, Tr(A B_1)];$$

$$\mathbb{E}[\varepsilon' B_1 \varepsilon\, \varepsilon' B_2 \varepsilon] = \sigma_\varepsilon^4 [Tr(B_1)\, Tr(B_2) + Tr(B_1 B_2) + Tr(B_1' B_2)];$$

$$\mathbb{E}[\eta' B_1 \varepsilon\, \varepsilon' B_2 \eta] = \sigma_\varepsilon^2 \sigma_\eta^2 Tr(B_1 B_2);$$

$$\mathbb{E}[\eta' B_1 \varepsilon\, \eta' B_2 \varepsilon] = \sigma_\varepsilon^2 \sigma_\eta^2 Tr(B_1' B_2).$$

Next, we state a number of trace results. To economize on notation, define

$$\omega \equiv \frac{1}{(1 + \phi - \psi)(1 + \psi)(-1 + \phi + \psi)}.$$

Although the results can be obtained by using the properties of geometric series, we have used the computer algebra system Mathematica 4.0 for computation; see Wolfram (1991).

First, we consider the traces involving two matrices, *viz.* $Tr(P'P)$, $Tr(C'C)$, $Tr(P'V)$, $Tr(C'J)$, $Tr(P'W)$, $Tr(C'S)$, $Tr(V'V)$, $Tr(J'J)$, $Tr(V'W)$, $Tr(J'S)$, $Tr(W'W)$ and $Tr(S'S)$. All these traces can be calculated from the following three basic ones

$$Tr(U'U) = \omega\{-1 + \psi\} T + O(1), \tag{A.1}$$

$$Tr(U'V) = -\omega \phi T + O(1), \tag{A.2}$$

and

$$Tr(U'W) = \psi Tr(U'U) + \phi Tr(U'V). \tag{A.3}$$

For example,

$$\begin{aligned} Tr(C'J) &= Tr((\beta_0 V + \beta_1 W)'(U - \lambda V)) \\ &= \beta_0 Tr(V'U) - \beta_0 \lambda Tr(V'V) + \beta_1 Tr(W'U) - \lambda \beta_1 Tr(W'V) \\ &= \beta_0 Tr(U'V) - \beta_0 \lambda Tr(U'U) + \beta_1 Tr(U'W) - \lambda \beta_1 Tr(U'V) + O(1), \end{aligned}$$

where the last line follows from the fact that

$$Tr(V'V) = Tr(U' L_T' L_T U) = Tr(U'U) + O(1)$$

and

$$Tr(W'V) = Tr(V' L_T' L_T U) = Tr(V'U) + O(1),$$

since $V = L_T U = U L_T$, $W = L_T V = V L_T$ and

$$L_T' L_T = \begin{pmatrix} I_{(T-1)\times(T-1)} & 0_{(T-1)\times 1} \\ 0_{1\times(T-1)} & 0_{1\times 1} \end{pmatrix}.$$

Note that the leading term of (A.1) can be interpreted as the unconditional variance of the AR(2) process $y_t = \phi y_{t-1} + \psi y_{t-2} + \varepsilon_t$ with $\varepsilon_t \sim \mathcal{N}(0, 1)$, while the leading terms of respectively (A.2) and (A.3) are the first and second-order covariances of such a process.

Secondly, we consider traces involving three matrices like $Tr(P'PP')$. All these traces can be calculated from the following ones

$$Tr(V'UU') = Tr(W'VU') + O(1) = -\omega^2\phi(\psi - 1)(1 + \psi)T + O(1), \tag{A.4}$$

$$Tr(V'UV') = \omega^2 \left(\phi^2 + \psi - 2\psi^2 + \psi^3\right) T + O(1), \tag{A.5}$$

$$Tr(W'WU') = Tr(V'VU) + O(1) = \omega^2 \left(1 - 2\psi + \phi^2\psi + \psi^2\right) T + O(1), \tag{A.6}$$

$$Tr(W'UU') = \omega^2 \left(\phi^2 + \psi - 2\psi^2 + \psi^3\right) T + O(1), \tag{A.7}$$

$$Tr(W'UV') = Tr(V'UW') + O(1) = \omega^2 \left(\phi^3 - 2\phi\psi - 2\phi\psi^2\right) T + O(1), \tag{A.8}$$

$$Tr(W'VU') = \omega^2 \left(\phi^3 - 2\phi\psi - 2\phi\psi^2\right) T + O(1), \tag{A.9}$$

$$Tr(W'UW') = \omega^2 \left(\phi^4 + 3\phi^2\psi + \psi^2 - 2\phi^2\psi^2 - 2\psi^3 + \psi^4\right) T + O(1), \tag{A.10}$$

and

$$Tr(V'WU') = -\omega^2 \left(\phi^3 - 2\phi + 2\phi\psi\right) T + O(1). \tag{A.11}$$

For example,

$$\begin{aligned}
Tr(P'CS') &= Tr((V - \gamma W)'(\beta_0 V + \beta_1 W)(V - \lambda W)') \\
&= \beta_0 Tr(V'VV') - \lambda\beta_0 Tr(V'VW') + \beta_1 Tr(V'WV') - \lambda\beta_1 Tr(V'WW') - \gamma\beta_0 Tr(W'VV') \\
&\quad -\gamma\lambda\beta_0 Tr(W'VW') - \gamma\beta_1 Tr(W'WV') + \gamma\lambda\beta_1 Tr(W'WW') \\
&= \beta_0 Tr(V'UU') - \lambda\beta_0 Tr(V'UV') + \beta_1 Tr(V'VU') - \lambda\beta_1 Tr(V'UU') - \gamma\beta_0 Tr(W'UU') \\
&\quad -\gamma\lambda\beta_0 Tr(W'UV') - \gamma\beta_1 Tr(W'VU') + \gamma\lambda\beta_1 Tr(W'UU') + O(1).
\end{aligned}$$

Appendix B: Proof of Theorem 1

First, we focus on the matrix $\bar{D} = \bar{Z}'\bar{Z} + \mathbb{E}[\tilde{Z}'\tilde{Z}]$. Using $\bar{Z} = (\theta P\iota_T : \theta\pi V\iota_T : \theta\pi W\iota_T : \iota_T)$, we have

$$\bar{Z}'\bar{Z} = \begin{bmatrix}
\theta^2\iota_T'P'P\iota_T & \pi\theta^2\iota_T'P'V\iota_T & \pi\theta^2\iota_T'P'W\iota_T & \theta\iota_T'P'\iota_T \\
\pi\theta^2\iota_T'V'P\iota_T & \pi^2\theta^2\iota_T'V'V\iota_T & \pi^2\theta^2\iota_T'V'W\iota_T & \pi\theta\iota_T'V'\iota_T \\
\pi\theta^2\iota_T'W'P\iota_T & \pi\theta^2\iota_T'W'V\iota_T & \pi^2\theta^2\iota_T'W'W\iota_T & \pi\theta\iota_T'W'\iota_T \\
\theta\iota_T'P\iota_T & \theta\pi\iota_T'V\iota_T & \theta\pi\iota_T'W\iota_T & \iota_T'\iota_T
\end{bmatrix}.$$

To evaluate $\bar{Z}'\bar{Z}$, we need to obtain a first-order approximation of the following quantities

$$\iota_T'X'\iota_t \qquad \text{and} \qquad \iota_T'X'Y\iota_T,$$

where the $T \times T$ matrices $X, Y \in \{P, V, W\}$. It appears that the matrix P contributes a multiplicative factor of $(1-\gamma)/(1-\psi-\phi)$, while the matrices V and W both contribute a factor of $1/(1-\psi-\phi)$. For example,

$$\iota_T' P' \iota_T = \frac{(1-\gamma)}{(1-\psi-\phi)} T + O(1) \qquad \text{and} \qquad \iota_T' P' V \iota_T = \frac{(1-\gamma)}{(1-\psi-\phi)^2} T + O(1).$$

Hence, we obtain

$$\bar{Z}'\bar{Z} = ww'T + O(1),$$

where

$$w = \left[\frac{(1-\gamma)\theta}{(1-\phi-\psi)}, \frac{\pi\theta}{(1-\phi-\psi)}, \frac{\pi\theta}{(1-\phi-\psi)}, 1 \right]'.$$

Using $\tilde{Z} = (P\varepsilon + C\eta : \pi V\varepsilon + J\eta : \pi W\varepsilon + S\eta : 0)$, we obtain

$$\begin{aligned}
\mathbb{E}[\tilde{Z}'\tilde{Z}] &= \mathbb{E}\begin{bmatrix} (P\varepsilon+C\eta)'(P\varepsilon+C\eta) & (P\varepsilon+C\eta)'(\pi V\varepsilon+J\eta) & (P\varepsilon+C\eta)'(\pi W\varepsilon+S\eta) & 0 \\ (\pi V\varepsilon+J\eta)'(P\varepsilon+C\eta) & (\pi V\varepsilon+J\eta)'(\pi V\varepsilon+J\eta) & (\pi V\varepsilon+J\eta)'(\pi W\varepsilon+S\eta) & 0 \\ (\pi W\varepsilon+S\eta)'(P\varepsilon+C\eta) & (\pi W\varepsilon+S\eta)'(\pi V\varepsilon+J\eta) & (\pi W\varepsilon+S\eta)'(\pi W\varepsilon+S\eta) & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} \sigma_\varepsilon^2 Tr(P'P) + \sigma_\eta^2 Tr(C'C) & \pi\sigma_\varepsilon^2 Tr(P'V) + \sigma_\eta^2 Tr(C'J) & \pi\sigma_\varepsilon^2 Tr(P'W) + \sigma_\eta^2 Tr(C'S) & 0 \\ \pi\sigma_\varepsilon^2 Tr(V'P) + \sigma_\eta^2 Tr(J'C) & \pi^2\sigma_\varepsilon^2 Tr(V'V) + \sigma_\eta^2 Tr(J'J) & \pi^2\sigma_\varepsilon^2 Tr(V'W) + \sigma_\eta^2 Tr(J'S) & 0 \\ \pi\sigma_\varepsilon^2 Tr(W'P) + \sigma_\eta^2 Tr(S'C) & \pi^2\sigma_\varepsilon^2 Tr(W'V) + \sigma_\eta^2 Tr(S'J) & \pi^2\sigma_\varepsilon^2 Tr(W'W) + \sigma_\eta^2 Tr(S'S) & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\
&= K \cdot T + O(1).
\end{aligned}$$

Although the 4×4 matrix K can be given explicitly using formulae (A.1)-(A.3), this will not be pursued. The (symmetric) inverse of $\bar{D}$, denoted by Q, is given by

$$Q \equiv \begin{bmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{12} & q_{22} & q_{23} & q_{24} \\ q_{13} & q_{23} & q_{33} & q_{34} \\ q_{14} & q_{24} & q_{34} & q_{44} \end{bmatrix} = (\bar{Z}'\bar{Z} + \mathbb{E}[\tilde{Z}'\tilde{Z}])^{-1}.$$

Again, formulae (A.1)-(A.3) can be used to find explicit expressions for the elements of Q, for instance

$$\begin{aligned}
q_{11} = \; & \{((\pi^2\sigma_\varepsilon^2 + (-(1+\lambda)(-1+\gamma\lambda) + \pi\lambda(-\beta_0+\beta_1))\sigma_\eta^2) \\
& (\pi^2\sigma_\varepsilon^2 + ((-1+\lambda)(-1+\gamma\lambda) - \pi\lambda(\beta_0+\beta_1))\sigma_\eta^2))/ \\
& (\pi^2\sigma_\varepsilon^4\sigma_\eta^2 + ((-1+\gamma\lambda)^2 - 2\pi\lambda\beta_0 + \pi\beta_1(-2\gamma\lambda+\pi\beta_1))\sigma_\varepsilon^2\sigma_\eta^4 + (\lambda\beta_0+\beta_1)^2\sigma_\eta^6)\}T^{-1} + o(T^{-1}),
\end{aligned}$$

and

$$q_{12} = -\frac{\pi}{\sigma_\eta^2} T^{-1} + o(T^{-1}).$$

Next, we have to calculate

$$\begin{aligned}
B_\alpha(T^{-1}) = \; & \mathbb{E}[Q\tilde{Z}'\varepsilon] && \text{(B.1a)} \\
& -\mathbb{E}[Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q\tilde{Z}'\varepsilon] && \text{(B.1b)} \\
& -\mathbb{E}[Q(\tilde{Z}'\tilde{Z} - \mathbb{E}[\tilde{Z}'\tilde{Z}])Q\tilde{Z}'\varepsilon] + o(T^{-1}). && \text{(B.1c)}
\end{aligned}$$

When there is no constant, $\bar{Z} = 0$ and the bias formula reduces to

$$B_\alpha^{\text{no constant}}(T^{-1}) = \mathbb{E}[Q\tilde{Z}'\varepsilon] - \mathbb{E}[Q(\tilde{Z}'\tilde{Z} - \mathbb{E}[\tilde{Z}'\tilde{Z}])Q\tilde{Z}'\varepsilon] + o(T^{-1}). \qquad \text{(B.2)}$$

Due to the special structure of the 4 × 4 matrices $\bar{Z}'\bar{Z}$ and $\mathbb{E}[\tilde{Z}'\tilde{Z}]$, the 3 × 3 submatrix $Q_{1\times3,1\times3}$ equals the inverse of the 3 × 3 submatrix $(\mathbb{E}[\tilde{Z}'\tilde{Z}])^{-1}_{1\times3,1\times3}$, *i.e.*

$$E_{34}QE'_{34} = E_{34}(\mathbb{E}[\tilde{Z}'\tilde{Z}])^{-1}E'_{34}, \qquad \text{where} \quad E_{34} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Since the last column of $\tilde{Z}$ equals zero, the bias formula $B_\alpha^{\text{no constant}}(T^{-1})$ only depends on elements of the 3 × 3 submatrix $Q_{1\times3,1\times3}$, *i.e.* $E_{34}QE'_{34}$. Hence, the bias due to the regressors denoted by $B^r_\alpha(T^{-1})$ equals formula (B.2), so that

$$B^c_\alpha(T^{-1}) = -\mathbb{E}[Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q\bar{Z}'\varepsilon] + o(T^{-1})$$

and

$$B^r_\alpha(T^{-1}) = \mathbb{E}[Q\tilde{Z}'\varepsilon] - \mathbb{E}[Q(\tilde{Z}'\tilde{Z} - \mathbb{E}[\tilde{Z}'\tilde{Z}])Q\tilde{Z}'\varepsilon] + o(T^{-1}).$$

Below, we shall calculate each of the components of B_α separately. The first component (B.1a) is equal to

$$\mathbb{E}[Q\tilde{Z}'\varepsilon] = Q\,\mathbb{E}\begin{bmatrix} \varepsilon'P'\varepsilon + \eta'C'\varepsilon \\ \pi\varepsilon'V'\varepsilon + \eta'J'\varepsilon \\ \pi\varepsilon'W'\varepsilon + \eta'S'\varepsilon \\ 0 \end{bmatrix} = Q\begin{bmatrix} \sigma^2_\varepsilon Tr(P') \\ \sigma^2_\varepsilon \pi\, Tr(V') \\ \sigma^2_\varepsilon \pi\, Tr(W') \\ 0 \end{bmatrix} = 0,$$

since the diagonal elements of the matrices P', V' and W' are zero. Next, we turn to the expectation of $Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q\bar{Z}'\varepsilon$. By direct calculation it can be verified that the first part is also zero up to first order, *i.e.*

$$\mathbb{E}[Q\bar{Z}'\tilde{Z}Q\bar{Z}'\varepsilon] = 0 + o(T^{-1}).$$

The expectation of $Q\tilde{Z}'\bar{Z}Q\bar{Z}'\varepsilon$ is calculated as follows. First, we have

$$\begin{aligned}\tilde{Z}'\bar{Z}Q\bar{Z}'\varepsilon \quad = \quad & \begin{bmatrix} (P\varepsilon+C\eta)'\theta P\iota_T & (P\varepsilon+C\eta)'\theta P\iota_T & (P\varepsilon+C\eta)'\theta\pi W\iota_T & (P\varepsilon+C\eta)'\iota_T \\ (\pi V\varepsilon+J\eta)'\theta P\iota_T & (\pi V\varepsilon+J\eta)'\theta P\iota_T & (\pi V\varepsilon+J\eta)'\theta\pi W\iota_T & (\pi V\varepsilon+J\eta)'\iota_T \\ (\pi W\varepsilon+S\eta)'\theta P\iota_T & (\pi W\varepsilon+S\eta)'\theta P\iota_T & (\pi W\varepsilon+S\eta)'\theta\pi W\iota_T & (\pi W\varepsilon+S\eta)'\iota_T \\ 0 & 0 & 0 & 0 \end{bmatrix} \times \\ & Q\begin{bmatrix} \theta\iota'_T P'\varepsilon \\ \theta\iota'_T P'\varepsilon \\ \theta\pi\iota'_T W'\varepsilon \\ \iota'_T\varepsilon \end{bmatrix}.\end{aligned}$$

Denoting $\mathbb{E}[\tilde{Z}'\bar{Z}Q\bar{Z}'\varepsilon] = l$, where $l = (l_1 : l_2 : l_3 : l_4)'$, then the first element of the 4 × 1 vector l is given by

$$\begin{aligned} l_1 \quad = \quad & \sigma^2_\varepsilon\{\theta^2\iota'_T P'P'P\iota_T q_{11} + \pi\theta^2\iota'_T V'P'P\iota_T q_{12} + \pi\theta^2\iota'_T W'P'P\iota_T\varepsilon q_{13} + \theta\iota'_T P'P\iota_T q_{14} \\ & +\pi\theta^2\iota'_T P'P'V\iota_T q_{21} + \pi^2\theta^2\iota'_T V'P'V\iota_T q_{22} + \pi^2\theta^2\iota'_T W'P'V\iota_T q_{23} \\ & +\pi\theta\iota'_T P'V\iota_T q_{24} + \pi\theta^2\iota'_T P'P'W\iota_T q_{31} + \pi^2\theta^2\iota'_T V'P'W\iota_T q_{32} \\ & +\pi^2\theta^2\iota'_T W'P'W\iota_T q_{33} + \pi\theta\iota'_T P'W\iota_T q_{34} + \theta\iota'_T P'P'\iota_T q_{41} \\ & +\pi\theta\iota'_T V'P'\iota_T q_{42} + \pi\theta\iota'_T W'P'\iota_T q_{43} + \iota'_T P'\iota_T q_{44}\}. \end{aligned}$$

Analogous expressions can be derived for l_2 and l_3, while $l_4 = 0$.

To evaluate the vector l, we need to obtain a first-order approximation of the following quantities

$$\iota'_T X'\iota_t, \qquad \iota'_T X'Y'\iota_T \qquad \text{and} \qquad \iota'_T X'Y'Z\iota_T,$$

where the $T \times T$ matrices $X, Y, Z \in \{P, V, W\}$. As before, the matrix P contributes a multiplicative factor of $(1-\gamma)/(1-\psi-\phi)$, while the matrices V and W both contribute a factor of $1/(1-\psi-\phi)$. For example,

$$\iota'_T P'V\iota_T = \frac{(1-\gamma)}{(1-\psi-\phi)^2}T + O(1) \qquad \text{and} \qquad \iota'_T P'P'V'\iota_T = \frac{(1-\gamma)^2}{(1-\psi-\phi)^3}T + O(1).$$

Hence, the first element of the vector l is given by

$$\begin{aligned} l_1 &= \sigma_\varepsilon^2\{\frac{\theta^2(1-\gamma)^3 q_{11}}{(1-\phi-\psi)^3} + \frac{\pi\theta^2(1-\gamma)^2 q_{12}}{(1-\phi-\psi)^3} + \frac{\pi\theta^2(1-\gamma)^2 q_{13}}{(1-\phi-\psi)^3} + \frac{\theta(1-\gamma)^2 q_{14}}{(1-\phi-\psi)^2} + \\ &\frac{\pi\theta^2(1-\gamma)^2 q_{21}}{(1-\phi-\psi)^3} + \frac{\pi^2\theta^2(1-\gamma) q_{22}}{(1-\phi-\psi)^3} + \frac{\pi^2\theta^2(1-\gamma) q_{23}}{(1-\phi-\psi)^3} + \frac{\pi\theta(1-\gamma) q_{24}}{(1-\phi-\psi)^2} + \\ &\frac{\pi\theta^2(1-\gamma)^2 q_{31}}{(1-\phi-\psi)^3} + \frac{\pi^2\theta^2(1-\gamma) q_{32}}{(1-\phi-\psi)^3} + \frac{\pi^2\theta^2(1-\gamma) q_{33}}{(1-\phi-\psi)^3} + \frac{\pi\theta(1-\gamma) q_{34}}{(1-\phi-\psi)^2} + \\ &\frac{\theta(1-\gamma)^2 q_{41}}{(1-\phi-\psi)^2} + \frac{\pi\theta(1-\gamma) q_{42}}{(1-\phi-\psi)^2} + \frac{\pi\theta(1-\gamma) q_{43}}{(1-\phi-\psi)^2} + \frac{(1-\gamma) q_{44}}{1-\phi-\psi}\}T + O(1). \end{aligned}$$

The non-zero elements l_2 and l_3 can be calculated analogously. Substitution of $\phi = \pi\beta_0+\gamma+\lambda$, $\psi = \pi\beta_1-\gamma\lambda$ and the expressions for the elements of the matrix Q into l_1, l_2 and l_3 gives

$$l_1 = \frac{(-1+\gamma)\sigma_\varepsilon^2}{-1+\gamma+\lambda-\gamma\lambda+\pi\beta_0+\pi\beta_1} + o(1),$$

$$l_2 = l_3 + o(1) = -\frac{\pi\sigma_\varepsilon^2}{-1+\gamma+\lambda-\gamma\lambda+\pi\beta_0+\pi\beta_1} + o(1).$$

The bias due to the intercept is found by multiplying the appropriate row of Q with the vector l, *e.g.*

$$B_\lambda^c(T^{-1}) = -q_{11}l_1 - q_{12}l_2 - q_{13}l_3 + o(T^{-1}).$$

The expectation of the term $Q\tilde{Z}'\bar{Z}Q\bar{Z}'\varepsilon$ is calculated as follows

$$\begin{aligned} \tilde{Z}'\tilde{Z}Q\tilde{Z}'\varepsilon &= \begin{bmatrix} (P\varepsilon+C\eta)'(P\varepsilon+C\eta) & (P\varepsilon+C\eta)'(\pi V\varepsilon+J\eta) & (P\varepsilon+C\eta)'(\pi W\varepsilon+S\eta) & 0 \\ (\pi V\varepsilon+J\eta)'(P\varepsilon+C\eta) & (\pi V\varepsilon+J\eta)'(\pi V\varepsilon+J\eta) & (\pi V\varepsilon+J\eta)'(\pi W\varepsilon+S\eta) & 0 \\ (\pi W\varepsilon+S\eta)'(P\varepsilon+C\eta) & (\pi W\varepsilon+S\eta)'(\pi V\varepsilon+J\eta) & (\pi W\varepsilon+S\eta)'(\pi W\varepsilon+S\eta) & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \times \\ & Q \begin{bmatrix} \varepsilon'P'\varepsilon+\eta'C'\varepsilon \\ \pi\varepsilon'V'\varepsilon+\eta'J'\varepsilon \\ \pi\varepsilon'W'\varepsilon+\eta'S'\varepsilon \\ 0 \end{bmatrix}. \end{aligned}$$

Denoting $\mathbb{E}[\tilde{Z}'\tilde{Z}Q\tilde{Z}'\varepsilon] = k$, where $k = (k_1 : k_2 : k_3 : k_4)'$, then the first element of the 4×1 vector k is given by

$$\begin{aligned} k_1 &= 2Tr(P'PP')q_{11}\sigma_\varepsilon^4 + 2\pi Tr(P'PV')q_{12}\sigma_\varepsilon^4 + 2\pi Tr(P'PW')q_{13}\sigma_\varepsilon^4 + \\ &\pi[Tr(P'VP') + Tr(V'PP']q_{21}\sigma_\varepsilon^4 + \pi^2[Tr(P'VV') + Tr(V'PV')]q_{22}\sigma_\varepsilon^4 + \\ &\pi^2[Tr(P'VW') + Tr(V'PW')]q_{23}\sigma_\varepsilon^4 + \pi[Tr(P'WP') + Tr(W'PP')]q_{31}\sigma_\varepsilon^4 + \\ &\pi^2[Tr(P'WV') + Tr(W'PV')]q_{32}\sigma_\varepsilon^4 + \pi^2[Tr(P'WW') + Tr(W'PW')]q_{33}\sigma_\varepsilon^4 + \\ &2Tr(P'CC')q_{11}\sigma_\varepsilon^2\sigma_\eta^2 + 2Tr(P'CJ')q_{12}\sigma_\varepsilon^2\sigma_\eta^2 + 2Tr(P'CS')q_{13}\sigma_\varepsilon^2\sigma_\eta^2 + \\ &Tr(P'JC')q_{21}\sigma_\varepsilon^2\sigma_\eta^2 + \pi Tr(V'CC')q_{21}\sigma_\varepsilon^2\sigma_\eta^2 + Tr(P'JJ')q_{22}\sigma_\varepsilon^2\sigma_\eta^2 + \\ &\pi Tr(V'CJ')q_{22}\sigma_\varepsilon^2\sigma_\eta^2 + Tr(P'JS')q_{23}\sigma_\varepsilon^2\sigma_\eta^2 + \pi Tr(V'CS')q_{23}\sigma_\varepsilon^2\sigma_\eta^2 + \\ &Tr(P'SC')q_{31}\sigma_\varepsilon^2\sigma_\eta^2 + \pi Tr(W'CC')q_{31}\sigma_\varepsilon^2\sigma_\eta^2 + Tr(P'SJ')q_{32}\sigma_\varepsilon^2\sigma_\eta^2 + \\ &\pi Tr(W'CJ')q_{32}\sigma_\varepsilon^2\sigma_\eta^2 + Tr(P'SS')q_{33}\sigma_\varepsilon^2\sigma_\eta^2 + \pi Tr(W'CS')q_{33}\sigma_\varepsilon^2\sigma_\eta^2. \end{aligned}$$

Similar expressions can be derived for k_2 and k_3, while $k_4 = 0$.

Using formulae (A.4)-(A.11), the elements of k are given explicitly by

$$\begin{aligned} k_1 \;=\; & ((-(-1+\gamma^2)(-\gamma-2\lambda+3\gamma\lambda^2)+\pi^2\gamma\beta_0^2+2\pi(-1+3\gamma^2)\lambda\beta_1 \\ & -3\pi^2\gamma\beta_1^2-2\pi\beta_0(1-\gamma^2-2\gamma\lambda+\pi\beta_1))\sigma_\varepsilon^2)/ \\ & ((-1+\gamma\lambda-\pi\beta_1)((-1+\gamma)(-1+\lambda)-\pi\beta_0-\pi\beta_1)(1+\gamma+\lambda+\gamma\lambda+\pi\beta_0-\pi\beta_1))+o(1), \end{aligned}$$

$$k_2 = \frac{-2\pi(\gamma+\lambda+\pi\beta_0)\sigma_\varepsilon^2}{(1+\gamma+\lambda+\gamma\lambda+\pi\beta_0-\pi\beta_1)(-1+\gamma+\lambda-\gamma\lambda+\pi\beta_0+\pi\beta_1)}+o(1),$$

and

$$k_3 = \frac{-(\pi(1+\gamma^2+\lambda^2-3\gamma^2\lambda^2+2\pi(\gamma+\lambda)\beta_0+\pi^2\beta_0^2+2(\pi+3\pi\gamma\lambda)\beta_1-3\pi^2\beta_1^2)\sigma_\varepsilon^2)}{(-1+\gamma\lambda-\pi\beta_1)((-1+\gamma)(-1+\lambda)-\pi\beta_0-\pi\beta_1)(1+\gamma+\lambda+\gamma\lambda+\pi\beta_0-\pi\beta_1)}+o(1).$$

Since $\mathbb{E}[Q\tilde{Z}'\varepsilon]=0$ and $\mathbb{E}[Q\mathbb{E}[\tilde{Z}'\tilde{Z}]Q\tilde{Z}'\varepsilon]=QQ^{-1}Q\mathbb{E}[\tilde{Z}'\varepsilon]=0$, the bias due to the regressors only depends on $\mathbb{E}[Q\tilde{Z}'\tilde{Z}Q\tilde{Z}'\varepsilon]$, which can be calculated by multiplying the appropriate row of Q with the vector k, *e.g.*

$$B_\lambda^r(T^{-1}) = -q_{11}k_1 - q_{12}k_2 - q_{13}k_3 + o(T^{-1}).$$

Appendix C: Some remarks on proving Theorem 2

For the AD(1,0)-model, the parameter vector α equals $(\lambda,\beta,\theta)'$. Furthermore, the matrices Z, $\bar{Z}$ and $\tilde{Z}$ are given by

$$Z=(y_{-1}:x:\iota_T),\qquad \bar{Z}=(\theta P\iota_T:\theta\pi V\iota_T:\iota_T),\qquad \tilde{Z}=(P\varepsilon+C\eta:\pi V\varepsilon+J\eta:0),$$

where the matrices P, V, C, J are all functions of the matrix U given in (11) with the following modified parameters

$$\phi\equiv\pi\beta+\gamma+\lambda \qquad \text{and} \qquad \psi\equiv-\gamma\lambda. \tag{C.1}$$

Now, Theorem 2 follows from the results in Appendix B keeping in mind the two definitions given in (C.1), $(\beta_0,\beta_1)=(\beta,0)$ and the fact that the size of the matrices $Z,\bar{Z},\tilde{Z}$, and hence Q, have changed.

Appendix D: Approximation to the Bias Formula $B_\lambda^{GS}(T^{-1})$

Recall that x_t can be written as an ARMA(2,1) process; see (9). Defining

$$x_t^\varepsilon=\phi x_{t-1}^\varepsilon+\psi x_{t-2}^\varepsilon+\pi\varepsilon_{t-1}$$

and

$$x_t^\eta=\phi x_{t-1}^\eta+\psi x_{t-2}^\eta+\eta_t-\lambda\eta_{t-1}, \tag{D.1}$$

we have

$$\operatorname{cov}(x_t,x_{t-j})=\operatorname{cov}(x_t^\varepsilon,x_{t-j}^\varepsilon)+\operatorname{cov}(x_t^\eta,x_{t-j}^\eta),\qquad j\in\mathbb{N}.$$

If

$$\upsilon(j,\xi,\zeta)=\frac{(-\psi)^j(\lambda_1^j(-\zeta+\xi(\phi+\psi\lambda_1))+\lambda_2^j(\zeta-\xi(\phi+\psi\lambda_2)))}{(1+\phi-\psi)(1+\psi)(-1+\phi+\psi)\sqrt{\phi^2+4\psi}},$$

then

$$\text{cov}(x_t^\varepsilon, x_{t-j}^\varepsilon) = \pi^2 \upsilon(j, \psi - 1, -\phi)\sigma_\varepsilon^2$$

and

$$\text{cov}(x_t^\eta, x_{t-j}^\eta) = \upsilon(j, (1+\lambda^2)(\psi - 1) + 2\lambda\phi, \lambda - \phi - \lambda^2\phi + \lambda\phi^2 - \lambda\psi^2)\sigma_\eta^2. \tag{D.2}$$

The second argument of $\upsilon(\cdot)$ in (D.2) can be interpreted as the unconditional variance of the ARMA process given in (D.1) with $\eta_t \sim \mathcal{N}(0,1)$, while the third argument equals the first-order covariance of the ARMA process.

Approximating sample quantities by their population values, we obtain in the stable AD(1,0) model

$$\begin{aligned} G_0 &= \text{cov}(x_t^\varepsilon, x_t^\varepsilon) + \text{cov}(x_t^\eta, x_t^\eta) + O_p(T^{-1/2}) \\ &= \frac{-\pi^2(1+\gamma\lambda)\sigma_\varepsilon^2 + (-1-\gamma\lambda + 2\lambda(\pi\beta + \gamma + \lambda) - \lambda^2(1+\gamma\lambda))\sigma_\eta^2}{(1-\gamma\lambda)(-1+\pi\beta+\gamma+\lambda-\gamma\lambda)(1+\pi\beta+\gamma+\lambda+\gamma\lambda)} + O_p(T^{-1/2}). \end{aligned}$$

From

$$\sum_{m=1}^{\infty} \lambda^m \upsilon(m, \xi, \zeta) = \frac{-\lambda(\zeta + \lambda\xi\psi)}{(1+\phi-\psi)(1+\psi)(-1+\phi+\psi)(-1+\lambda(\phi+\lambda\psi))}$$

(which is related to the autocovariance-generation function evaluated at λ), it follows that

$$\begin{aligned} \Lambda &= \sum_{m=1}^{\infty} \lambda^m G_m + O_p(T^{-1/2}) \\ &= \pi^2 \sum_{m=1}^{\infty} \lambda^m \upsilon(m, \psi - 1, -\phi)\sigma_\varepsilon^2 + \sum_{m=1}^{\infty} \lambda^m \upsilon(m, (1+\lambda^2)(\psi-1) + 2\lambda\phi, \lambda - \phi - \lambda^2\phi + \lambda\phi^2 - \lambda\psi^2)\sigma_\eta^2 + O_p(T^{-1/2}) \\ &= \lambda \frac{-\pi^2(\pi\beta + \lambda - \gamma(\lambda^2 + \gamma\lambda^3 - 1))\sigma_\varepsilon^2}{(\pi\beta+\gamma+\lambda-\gamma\lambda-1)(\gamma\lambda-1)(1+\pi\beta+\gamma+\lambda+\gamma\lambda)(\lambda(\pi\beta+\gamma+\lambda-\gamma\lambda^2)-1)} \\ &\quad + \lambda \frac{(\pi\beta+\gamma-\gamma\lambda^2)(\lambda(\pi\beta+\gamma+\lambda-\gamma\lambda^2)-1)\sigma_\eta^2}{(\pi\beta+\gamma+\lambda-\gamma\lambda-1)(\gamma\lambda-1)(1+\pi\beta+\gamma+\lambda+\gamma\lambda)(\lambda(\pi\beta+\gamma+\lambda-\gamma\lambda^2)-1)} + O_p(T^{-1/2}). \end{aligned}$$

Using

$$\sum_{m=1}^{\infty} m\lambda^{m-1} \upsilon(m, \xi, \zeta) = \frac{-\zeta + \lambda(\zeta\lambda + \xi(2 - \lambda\phi))\psi}{(1+\phi-\psi)(1+\psi)(-1+\phi+\psi)(-1+\lambda(\phi+\lambda\psi))^2},$$

we obtain after some algebra (carried out by Mathematica)

$$\begin{aligned} \dot{\Lambda} &= \sum_{m=1}^{\infty} m\lambda^{m-1} G_m + O_p(T^{-1/2}) \\ &= \frac{(\pi^2(\gamma + \lambda - 2\gamma\lambda^2 + \gamma^3\lambda^4 + \gamma^2\lambda^3(-2+\lambda^2)) + \pi(\beta + \beta\gamma^2\lambda^4))\sigma_\varepsilon^2}{(-1+\pi\beta+\gamma+\lambda-\gamma\lambda)(1+\pi\beta+\gamma+\lambda-\gamma\lambda)(-1+\gamma\lambda)(\pi\beta\lambda - (\gamma\lambda-1)(\lambda^2-1))^2} \\ &\quad - \frac{(-1+\lambda(\pi\beta+\gamma+\lambda-\gamma\lambda^2))(\pi(\beta+\beta\gamma\lambda^3) + \gamma(1-2\lambda^2+\lambda^4))\sigma_\eta^2}{(-1+\pi\beta+\gamma+\lambda-\gamma\lambda)(1+\pi\beta+\gamma+\lambda-\gamma\lambda)(-1+\gamma\lambda)(\pi\beta\lambda-(\gamma\lambda-1)(\lambda^2-1))^2} + O_p(T^{-1/2}). \end{aligned}$$

The quantity $\dot{\Lambda}$ can be interpreted as the derivative of Λ with respect to λ assuming G_m does not depend on λ. This assumption is only valid when x_t is strongly exogenous, *i.e.* x_t does not depend on lagged values of y_t.

Table 1: Simulation results concerning $\hat{\lambda}$ in the AD(1,0) model without intercept $y = \lambda y_{-1} + \beta x + \varepsilon$ at $(\lambda, \beta) = (0.8, 0.2)$ for various values of (π, γ) and $T = 25$.

	OLS		BC (weak exog.)		BC (strong exog.)	
(π, γ)	Bias($\hat{\lambda}$)	MSE($\hat{\lambda}$)	Bias($\check{\lambda}$)	Rel.Eff.	Bias($\check{\lambda}$)	Rel.Eff.
(–10, 0.258)	0.042	0.440	0.006	1.026	0.058	1.029
(–8,–0.033)	0.046	0.554	0.002	0.995	0.062	1.027
(–6,–0.287)	–0.009	0.425	–0.001	1.014	0.010	1.030
(–4,–0.526)	–0.031	0.291	0.001	1.046	–0.005	1.030
(–2,–0.755)	–0.037	0.201	0.001	1.067	–0.002	1.033
(0,–0.968)	–0.044	0.142	0.001	1.048	–0.001	1.037
(0, 0.345)	–0.041	0.125	–0.003	0.979	–0.002	0.974
(–2, 0.958)	–0.023	0.118	0.004	1.011	–0.005	0.953
(–4, 0.992)	–0.029	0.128	0.005	1.040	–0.016	0.972
(–6, 0.980)	–0.032	0.136	0.005	1.052	–0.021	0.991
(–8, 0.944)	–0.033	0.149	0.004	1.062	–0.022	1.006
(–10, 0.867)	–0.031	0.170	0.004	1.071	–0.019	1.020

Remarks:
(1) The columns under BC (weak exog.) are based on the bias formula $B_{\lambda}^{r}(T^{-1})$ given in Theorem 2, while the columns under BC (strong exog.) are based on the bias formula $B_{\alpha}^{KPh}(T^{-1})$ given in (43).
(2) Rel.Eff.=MSE($\check{\lambda}$)/MSE($\hat{\lambda}$).
(3) First 6 rows are based on the smallest solution (Line 1 in Figure 1) for γ of equation (51), while the last 6 rows are based on the largest solution (Line 2 in Figure 1) for γ.

Table 2: Simulation results concerning $\hat{\lambda}$ in the AD(1,0) model with intercept $y = \lambda y_{-1} + \beta x + \theta + \varepsilon$ at $(\lambda, \beta, \theta) = (0.8, 0.2, 0.0)$ for various values of (π, γ) and $T = 25$.

	OLS		BC (weak exog.)		BC (strong exog.)	
(π, γ)	Bias($\hat{\lambda}$)	MSE($\hat{\lambda}$)	Bias($\check{\lambda}$)	Rel.Eff.	Bias($\check{\lambda}$)	Rel.Eff.
(–10, 0.258)	0.079	0.451	0.009	1.019	0.131	1.061
(–8,–0.033)	0.034	0.570	–0.002	0.988	0.084	1.042
(–6,–0.287)	–0.075	0.446	–0.010	1.008	–0.019	1.030
(–4,–0.526)	–0.108	0.323	–0.013	1.017	–0.040	1.002
(–2,–0.755)	–0.115	0.244	–0.020	0.992	–0.034	0.960
(0,–0.968)	–0.127	0.204	–0.033	0.903	–0.040	0.892
(0, 0.345)	–0.101	0.174	–0.027	0.857	–0.027	0.861
(–2, 0.958)	–0.022	0.122	0.003	1.003	0.018	0.962
(–4, 0.992)	–0.023	0.127	0.006	1.062	0.011	0.979
(–6, 0.980)	–0.023	0.133	0.007	1.087	0.009	1.004
(–8, 0.944)	–0.023	0.145	0.007	1.100	0.011	1.028
(–10, 0.867)	–0.017	0.165	0.007	1.109	0.020	1.057

Remarks:
(1) The columns under BC (weak exog.) are based on the bias formula $B_{\lambda}(T^{-1}) = B_{\lambda}^{c}(T^{-1}) + B_{\lambda}^{r}(T^{-1})$ given in Theorem 2, while the columns under BC (strong exog.) are based on the bias formula $B_{\alpha}^{KPh}(T^{-1})$ given in (43).
(2) and (3) see Table 1.

Table 3: Numerical results concerning $\hat{\lambda}$ in the model $y = \lambda y_{-1} + \beta_0 x + \beta_1 x_{-1} + \theta + \varepsilon$ for $(\lambda, \beta_0, \beta_1, \theta) = (0.8, 0.5, -0.3, 0)$, various values of (π, γ) and $T = 25$.

	OLS		BC (weak exog.)		BC (strong exog.)	
(π, γ)	Bias($\hat{\lambda}$)	MSE($\hat{\lambda}$)	Bias($\check{\lambda}$)	Rel.Eff.	Bias($\check{\lambda}$)	Rel.Eff.
(–0.8, 0.8)	–0.111	0.049	–0.034	0.775	–0.020	0.785
(–0.4, 0.8)	–0.128	0.051	–0.043	0.717	–0.033	0.724
(0.0, 0.8)	–0.174	0.064	–0.067	0.638	–0.070	0.645
(–0.8, 0.4)	–0.119	0.050	–0.035	0.802	–0.029	0.802
(–0.4, 0.4)	–0.131	0.050	–0.041	0.756	–0.038	0.754
(0.0, 0.4)	–0.155	0.056	–0.053	0.699	–0.059	0.697
(–0.8, 0.0)	–0.125	0.051	–0.036	0.828	–0.035	0.819
(–0.4, 0.0)	–0.133	0.050	–0.040	0.785	–0.042	0.778
(0.0, 0.0)	–0.147	0.052	–0.047	0.739	–0.054	0.733
(0.4, 0.0)	–0.171	0.067	–0.061	0.716	–0.071	0.723
(–0.8,–0.4)	–0.129	0.053	–0.037	0.858	–0.039	0.841
(–0.4,–0.4)	–0.135	0.050	–0.040	0.812	–0.045	0.799
(0.0,–0.4)	–0.144	0.050	–0.044	0.771	–0.053	0.761
(0.4,–0.4)	–0.158	0.056	–0.052	0.739	–0.064	0.737
(–0.4,–0.8)	–0.136	0.054	–0.041	0.855	–0.046	0.836
(0.0,–0.8)	–0.143	0.050	–0.043	0.798	–0.053	0.784
(0.4,–0.8)	–0.151	0.052	–0.048	0.771	–0.060	0.765

Remarks:
(1) The columns under BC (weak exog.) are based on the bias formula $B_\lambda(T^{-1}) = B^c_\lambda(T^{-1}) + B^r_\lambda(T^{-1})$ given in Theorem 1, while the columns under BC (strong exog.) are based on the bias formula $B^{KPh}_\alpha(T^{-1})$ given in (43).
(2) Rel.Eff.=MSE($\check{\lambda}$)/MSE($\hat{\lambda}$).

Table 4: Numerical results concerning $\hat{\lambda}$ in the model $y = \lambda y_{-1} + \beta_0 x + \beta_1 x_{-1} + \theta + \varepsilon$ for $(\lambda, \beta_0, \beta_1, \theta) = (0.8, 0.5, -0.3, 0)$, various values of (π, γ) and $T = 50$.

	OLS		BC (weak exog.)		BC (strong exog.)	
(π, γ)	Bias($\hat{\lambda}$)	MSE($\hat{\lambda}$)	Bias($\check{\lambda}$)	Rel.Eff.	Bias($\check{\lambda}$)	Rel.Eff.
(–0.8, 0.8)	–0.050	0.016	–0.008	0.827	0.000	0.843
(–0.4, 0.8)	–0.057	0.015	–0.011	0.770	–0.003	0.784
(0.0, 0.8)	–0.080	0.017	–0.019	0.674	–0.020	0.676
(–0.8, 0.4)	–0.056	0.016	–0.009	0.842	–0.004	0.849
(–0.4, 0.4)	–0.060	0.015	–0.011	0.799	–0.007	0.805
(0.0, 0.4)	–0.072	0.016	–0.014	0.739	–0.016	0.738
(–0.8, 0.0)	–0.060	0.017	–0.010	0.856	–0.007	0.855
(–0.4, 0.0)	–0.063	0.016	–0.011	0.820	–0.010	0.820
(0.0, 0.0)	–0.069	0.016	–0.013	0.777	–0.015	0.774
(0.4, 0.0)	–0.081	0.021	–0.018	0.761	–0.020	0.777
(–0.8,–0.4)	–0.063	0.018	–0.011	0.874	–0.008	0.867
(–0.4,–0.4)	–0.065	0.016	–0.011	0.836	–0.011	0.832
(0.0,–0.4)	–0.069	0.016	–0.012	0.801	–0.015	0.796
(0.4,–0.4)	–0.075	0.017	–0.015	0.776	–0.019	0.779
(–0.4,–0.8)	–0.066	0.018	–0.012	0.865	–0.011	0.860
(0.0,–0.8)	–0.069	0.016	–0.012	0.818	–0.015	0.811
(0.4,–0.8)	–0.072	0.017	–0.014	0.800	–0.017	0.800

Remarks: see Table 3.

Table 5: Numerical results concerning $\hat{\beta}_1$ in the model $y = \lambda y_{-1} + \beta_0 x + \beta_1 x_{-1} + \theta + \varepsilon$ for $(\lambda, \beta_0, \beta_1, \theta) = (0.8, 0.5, -0.3, 0)$, various values of (π, γ) and $T = 25$.

	OLS		BC (weak exog.)		BC (strong exog.)	
(π, γ)	Bias($\hat{\beta}_1$)	MSE($\hat{\beta}_1$)	Bias($\check{\beta}_1$)	Rel.Eff.	Bias($\check{\beta}_1$)	Rel.Eff.
(–0.8, 0.8)	0.065	0.016	0.019	0.845	0.016	0.839
(–0.4, 0.8)	0.079	0.021	0.026	0.823	0.025	0.814
(0.0, 0.8)	0.108	0.037	0.041	0.806	0.043	0.801
(–0.8, 0.4)	0.062	0.013	0.017	0.851	0.018	0.829
(–0.4, 0.4)	0.069	0.015	0.021	0.834	0.023	0.813
(0.0, 0.4)	0.082	0.021	0.027	0.827	0.031	0.811
(–0.8, 0.0)	0.061	0.012	0.017	0.859	0.020	0.831
(–0.4, 0.0)	0.065	0.013	0.019	0.848	0.022	0.823
(0.0, 0.0)	0.072	0.016	0.022	0.845	0.026	0.826
(0.4, 0.0)	0.065	0.043	0.026	0.939	0.011	0.928
(–0.8,–0.4)	0.063	0.013	0.017	0.868	0.023	0.840
(–0.4,–0.4)	0.064	0.013	0.018	0.867	0.023	0.841
(0.0,–0.4)	0.067	0.015	0.020	0.868	0.025	0.849
(0.4,–0.4)	0.069	0.025	0.023	0.908	0.022	0.899
(–0.4,–0.8)	0.071	0.031	0.017	0.929	0.033	0.907
(0.0,–0.8)	0.065	0.019	0.019	0.913	0.024	0.898
(0.4,–0.8)	0.065	0.023	0.021	0.931	0.022	0.920

Remarks:
(1) The columns under BC (weak exog.) are based on the bias formula $B_{\beta_1}(T^{-1}) = B^c_{\beta_1}(T^{-1}) + B^r_{\beta_1}(T^{-1})$ given in Theorem 1, while the columns under BC (strong exog.) are based on the bias formula $B^{KPh}_{\alpha}(T^{-1})$ given in (43).
(2) Rel.Eff.=MSE($\check{\beta}_1$)/MSE($\hat{\beta}_1$).

Table 6: Numerical results concerning β_1 in the model $y = \lambda y_{-1} + \beta_0 x + \beta_1 x_{-1} + \theta + \varepsilon$ for $(\lambda, \beta_0, \beta_1, \theta) = (0.8, 0.5, -0.3, 0)$, various values of (π, γ) and $T = 50$.

	OLS		BC (weak exog.)		BC (strong exog.)	
(π, γ)	Bias($\hat{\beta}_1$)	MSE($\hat{\beta}_1$)	Bias($\check{\beta}_1$)	Rel.Eff.	Bias($\check{\beta}_1$)	Rel.Eff.
(–0.8, 0.8)	0.030	0.006	0.005	0.895	0.002	0.902
(–0.4, 0.8)	0.036	0.007	0.007	0.880	0.005	0.882
(0.0, 0.8)	0.053	0.013	0.012	0.859	0.013	0.855
(–0.8, 0.4)	0.029	0.004	0.005	0.893	0.004	0.887
(–0.4, 0.4)	0.033	0.005	0.006	0.881	0.005	0.873
(0.0, 0.4)	0.039	0.008	0.008	0.876	0.009	0.868
(–0.8, 0.0)	0.029	0.004	0.004	0.897	0.005	0.885
(–0.4, 0.0)	0.031	0.005	0.005	0.890	0.006	0.878
(0.0, 0.0)	0.034	0.006	0.006	0.889	0.007	0.880
(0.4, 0.0)	0.031	0.018	0.008	0.953	–0.002	0.951
(–0.8,–0.4)	0.030	0.005	0.004	0.901	0.006	0.888
(–0.4,–0.4)	0.030	0.005	0.005	0.904	0.006	0.892
(0.0,–0.4)	0.032	0.006	0.005	0.908	0.006	0.900
(0.4,–0.4)	0.033	0.010	0.006	0.938	0.005	0.938
(–0.4,–0.8)	0.034	0.013	0.005	0.948	0.010	0.927
(0.0,–0.8)	0.030	0.007	0.005	0.940	0.006	0.935
(0.4,–0.8)	0.030	0.009	0.005	0.955	0.004	0.954

Remarks: see Table 5.

Figure 1: Region where $R^2 = 0.8$ and $\sigma_\varepsilon^2 = 0.1$ in the AD(1,0) model for $(\lambda, \beta) = (0.8, 0.2)$ and $\pi \in (-11.3, 0.6)$. In the main text, the solid line will be referred to as Line 1, while the dotted line will be referred to as Line 2.

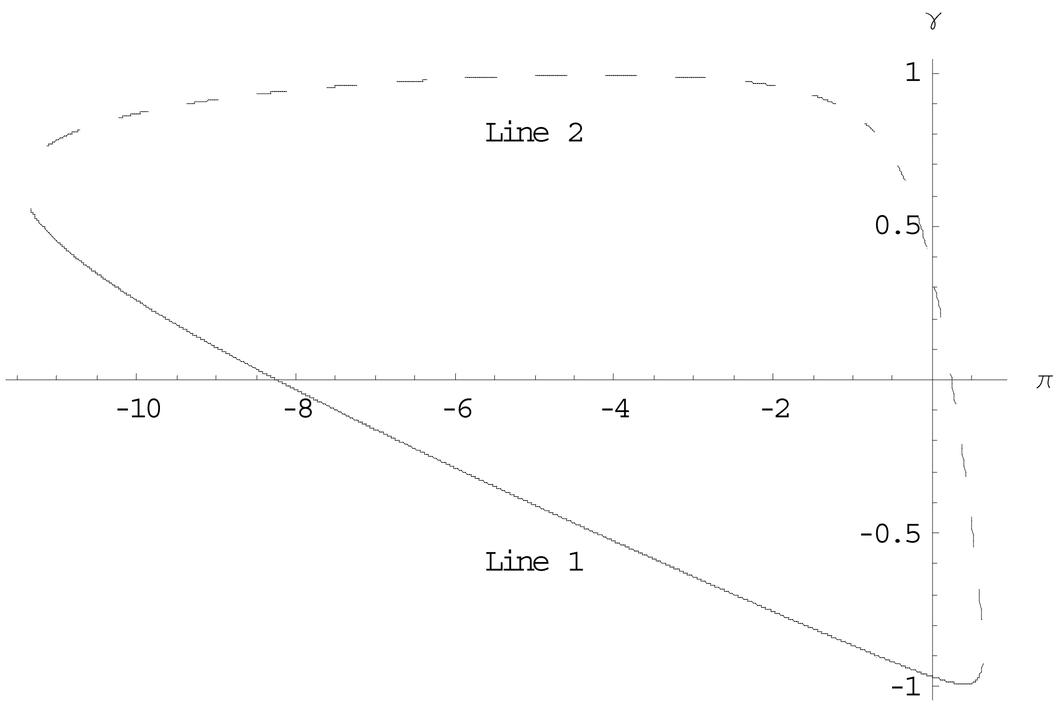

Figure 2: $-T\times$bias in the AD(1,0) model for (π, γ) along Line 1 in Figure 1. Solid line refers to the correct bias approximation $-T\ B_{\lambda}^{r}(T^{-1})$ and the dashed line refers to the incorrect bias approximation $-T\ B_{\lambda}^{GS}(T^{-1})$.

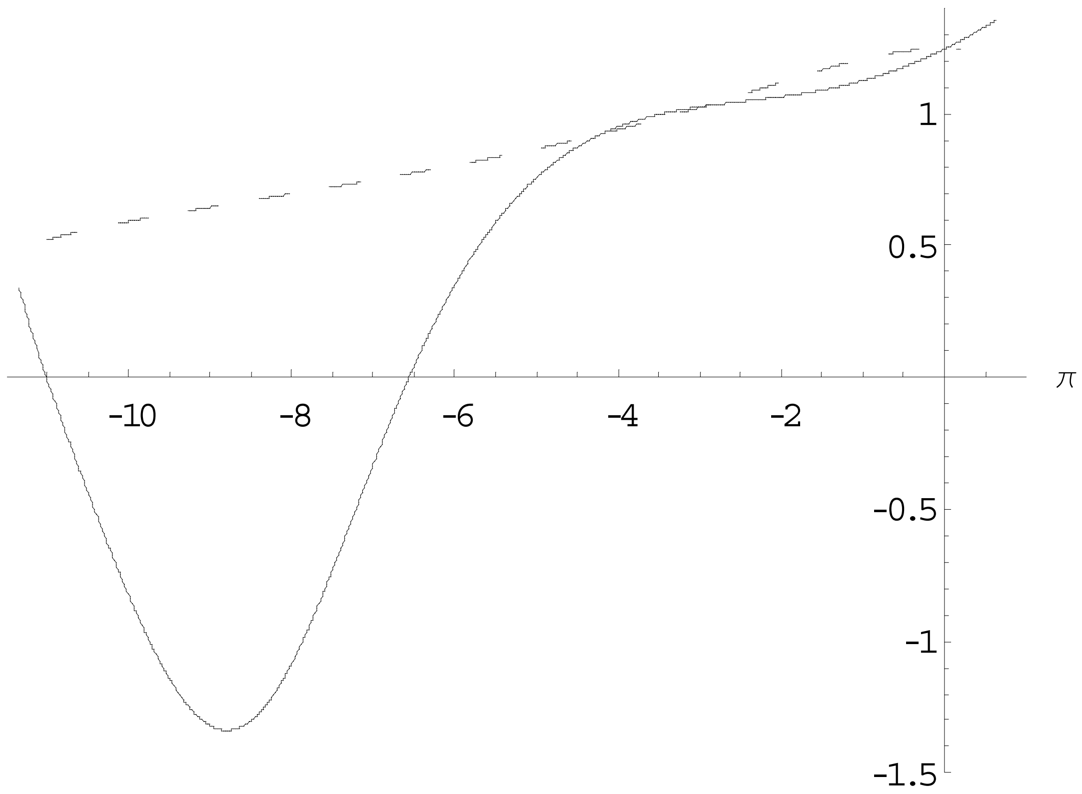

Figure 3: $-T \times$bias in the AD(1,0) model for (π, γ) along Line 2 in Figure 1. Solid line refers to the correct bias approximation $-T\ B_{\lambda}^{r}(T^{-1})$ and the dashed line refers to the incorrect bias approximation $-T\ B_{\lambda}^{GS}(T^{-1})$.

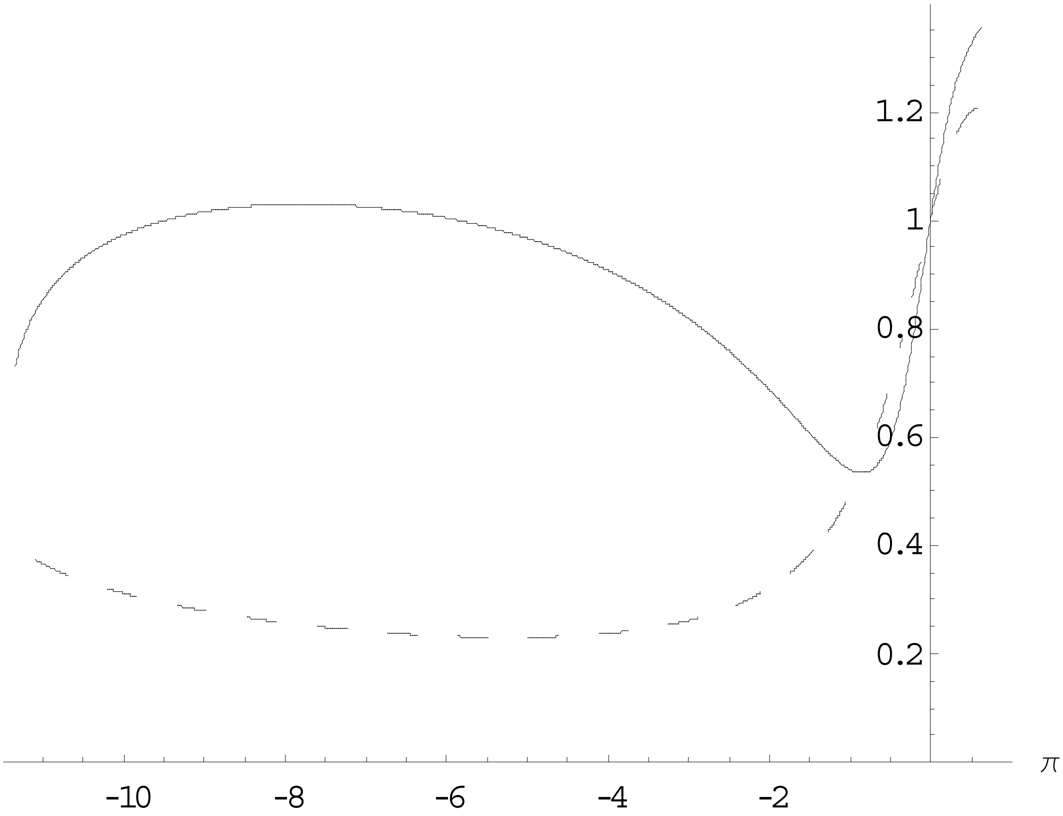

Figure 4: Admissible region of (π, γ) in the AD(1,1) model for $(\lambda, \beta_0, \beta_1, \theta) = (0.8, 0.5, -0.3, 0.0)$.

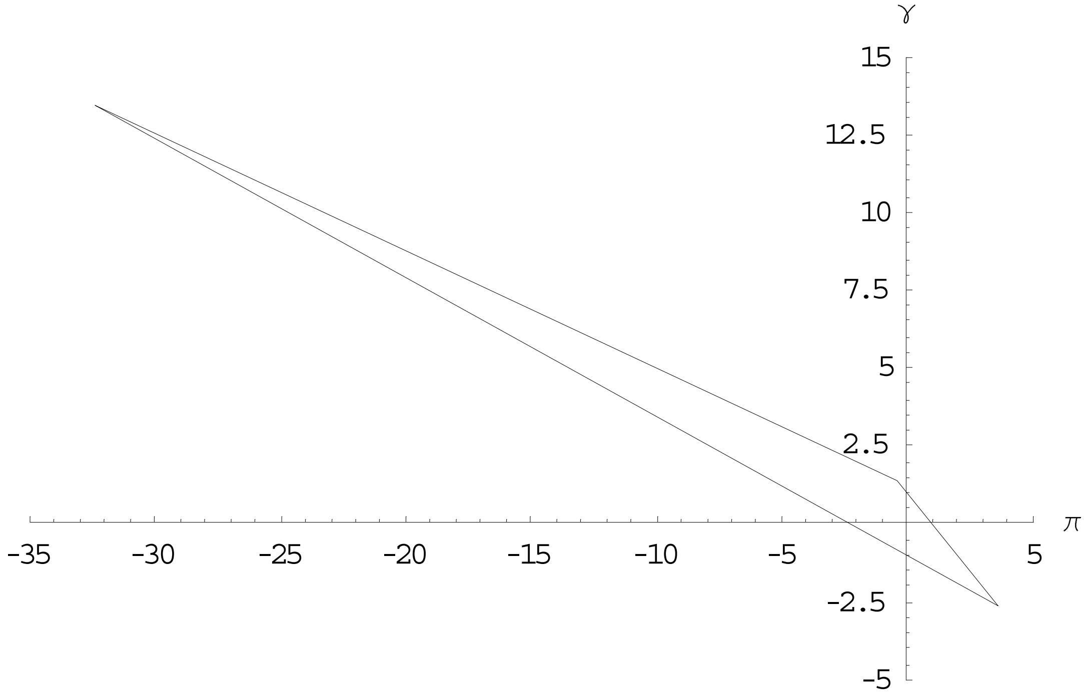

Figure 5: $-T \times B_{\lambda}(T^{-1})$ in the AD(1,1) model at various values of (π, γ) for $(\lambda, \beta_0, \beta_1, \theta) = (0.8, 0.5, -0.3, 0.0)$, $\sigma_{\eta}^2 = 1$ and σ_{ε}^2 such that the population $R^2 = 0.8$.