UvA ECONOMETRICS

Discussion Paper: 2009/09

Higher-order asymptotic expansions of the least-squares estimation bias in first-order dynamic regression models

Jan F. Kiviet and Garry D.A. Phillips

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

Amsterdam School of Economics
Department of Quantitative Economics
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

Higher-order asymptotic expansions of the least-squares estimation bias in first-order dynamic regression models

JAN F. KIVIET* and GARRY D.A. PHILLIPS†
(University of Amsterdam) (University of Cardiff)

9 December 2009

JEL-code: C13, C22

Keywords: ARX-model, asymptotic expansions, bias approximation, lagged dependent variables, Monte Carlo simulation

Abstract

An approximation to order T^{-2} is obtained for the bias of the full vector of least-squares estimates in general stable but not necessarily stationary ARX(1) models with normal disturbances. This yields generalizations, allowing for various forms of initial conditions, of Kendall's and White's classic results for stationary AR(1) models. The accuracy of various alternative approximations is examined and compared by simulation for particular parametrizations of AR(1) and ARX(1) models. The results show that often the second-order approximation is considerably better than its first order counterpart and hence opens perspectives for improved bias correction. However, we also find that order T^{-2} approximations are more vulnerable in the near unit root case than the much simpler order T^{-1} approximations.

1. Introduction and framework

The statistical literature concerned with the use of asymptotics for approximating statistical phenomena is vast. The overview by Pierce and Peters (1992) is one of a number of important contributions and while this article and many others focus on the use of higher-order asymptotics to improve inference, there is also considerable interest in their application to analysing the bias of ML estimators; see, for example, Cox and Snell (1968) and Copas (1988), who discuss a general method for approximating the ML estimation bias to the order of T^{-1}, where T is the sample size, using an asymptotic expansion of the score function (see also Firth's contribution to the discussion in Pierce and Peters, 1992). While Firth (1993), on noting that bias corrected ML estimators are, quite generally, second-order efficient, shows that in regular parametric problems this first-order term is removed by a suitable modification of the score function, Kass (1992) commented that when the first-order asymptotic approximation to a density is poor but not horrible, the higher-order approximation usually mops up most of

* Tinbergen Institute and Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands (j.f.kiviet@uva.nl)
† Cardiff Business School, Aberconway Building, Colum Drive, CF10 3EU, Cardiff, Wales, UK (phillipsgd1@cardiff.ac.uk).

the error. One purpose of this paper is to examine this type of phenomenon in the context of bias approximation in autoregressive models by comparing the first-order and the second-order approximations in a number of cases.

The use of asymptotic expansions in approximating the moments of estimators in stable autoregressive models has a relatively long history. The early work focused on the least-squares estimator of the serial correlation coefficient in the simplest autoregressive Gaussian process. See, for example, Bartlett (1946), who found a first-order variance approximation, and Hurwicz (1950), who obtained moment approximations for the case $T = 3$. Later White (1960) and Shenton and Johnson (1965) found higher-order approximations in terms of powers of T for the first two moments in the AR(1) model. For the case of an AR(1) model with an intercept, Kendall (1954) and Marriott 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}. Higher-order approximations to the bias in the vector of the least-squares coefficient estimator in normal autoregressive models with or without an intercept or with any further exogenous explanatory variables were obtained by us in a very early version of this paper[1], but remained unpublished because until recently we couldn't prove the general validity of these approximations. In Kiviet and Phillips (2009), however, which focuses on improved variance estimation in autoregressive models, we provide a general proof in which the order in a power of T is established of the remainder term in any higher-order expansion yielding an approximation to first or higher-order moments of a linear least-squares estimator. The proof only requires assumptions on the existence of particular data moments and the differentiability of the non-linear function of the data moments which identifies and establishes the least-squares estimator. These assumptions are rather mild and will hold in the dynamic regression model to be examined here.

Research on the accuracy of the approximations published thus far has shown that the higher-order results of White are very accurate and also that Kendall's first-order approximation is often surprisingly good. For evidence on these points, see Sawa (1978) and Nankervis and Savin (1988). Their exact results both confirm the severity of the bias problem and demonstrate the quality of some of the approximations. In the context of the AR(1) model with intercept, Monte Carlo results by Orcutt and Winokur (1969) provide both additional evidence on these matters and an illustration of how bias correction based on Kendall's approximation can be effective in not only reducing bias but in lowering the mean-squared error (MSE) as well. This latter point has been noted too by Rudebusch (1993), who uses Kendall's approximation and an approximation for higher-order AR models in bias corrected estimators when investigating whether real GNP is trend-stationary or difference-stationary. The first-order estimation bias in higher-order autoregressive processes has been examined by Shaman and Stine (1988), and in multivariate autoregressive processes by Tjøstheim and Paulsen (1983) and by Nicholls and Pope (1988). Naturally, the accuracy of asymptotic approximations is limited and depends on the order of the approximation, the actual size of the sample, but usually also on the model parameters and design, and on initial conditions. If the accuracy of a first-order approximation falls short for a specific case, then it seems recommended to examine a higher-order approximation, although considerable analytic problems may be incurred. Evans and Savin (1981) demonstrate the effectiveness of particular higher-order results in the AR(1) model without intercept.

For multi-parameter static simultaneous equations models the seminal paper of Nagar (1959) provided approximations to the moments of consistent k-class estimators. In particular they include a bias approximation to the order of T^{-1}. The results were later confirmed by Kadane (1971) using the approach of small disturbance asymptotics. Mikhail (1972) suggested that the first-order approximation to the bias may be inaccurate in some cases and he ex-

[1] This paper (same title) was presented at the Econometric Society World Conference 1995 held in Tokyo.

tended Nagar's approximation to the order T^{-2}. Hadri and Phillips (1999) showed that the higher-order approximation often yields a considerable improvement.

More recently, a number of papers have examined the small sample bias of the ordinary least-squares (OLS) estimator in single dynamic regression models. Of particular relevance for the current work are the papers by Grubb and Symons (1987), who derived – under normality of the disturbances – the bias to the order T^{-1} for the lagged dependent variable coefficient estimator in a stable first-order dynamic regression model with fixed regressors, and that of Kiviet and Phillips (1993) – henceforth KP – who gave the T^{-1} approximation of the full coefficient vector as well as the small disturbance asymptotic counterpart. KP (1994) give extensions of these results to higher-order dynamic models, Kiviet et al. (1995) to systems of seemingly unrelated regressions, and Iglesias and Phillips (2006) to QML estimation of autoregressive models with ARCH disturbances. Recently, the effects of non-normal disturbances on asymptotic approximations to the bias of the lagged dependent variable coefficient in regression models has been examined. Bao and Ullah (2007) find that the bias to order T^{-1} is only affected by the skewness, while the fact that the bias to order T^{-2} is also affected by the kurtosis has been derived in Bao (2007). Though, this latter study does not examine the actual accuracy in finite samples of this higher-order approximation.

A general finding in the above mentioned studies is that $O(T^{-1})$ bias approximations can be most helpful in constructing bias corrected estimators, but they do not always work well. This is particularly so in first-order dynamic models when the autoregressive coefficient approaches unity, the so-called unit root case. This is in agreement with Kendall (1954, p.404) who remarked for his first-order approximation in the simple AR(1) model with intercept that it seemed of doubtful validity for an autoregressive parameter value close to unity; the numerical results for that model in Nankervis and Savin (1988) corroborate this suspicion, especially for really small sample sizes. It may thus be of interest to derive and to verify the accuracy by simulation of higher-order approximations (similar to those of White, Mikhail and Bao) for the full coefficient vector of general though stable dynamic regression models. Higher-order bias approximations in case the lagged dependent variable coefficient is equal to unity have been derived already in KP (2005) and proved to be highly accurate.

In this paper we focus on the stable first-order dynamic regression model with normally distributed disturbances and expressions are obtained for the bias of the least-squares coefficient estimators to the order of T^{-2} by extending Nagar's approach. Attention is paid to models with either a fixed or random start-up. So, the focus of interest is the bias of the OLS estimator of all the regression coefficients in the model

$$y = \lambda y_{-1} + X\beta + u, \tag{1.1}$$

where $y = (y_1, ..., y_T)'$ is a $T \times 1$ vector of observations on a dependent variable, y_{-1} is the y vector lagged one period, i.e. $y_{-1} = (y_0, ..., y_{T-1})'$, X is a full column-rank $T \times K$ matrix of observations on K fixed or strongly exogenous regressors with $K \times 1$ coefficient vector β, and u is a $T \times 1$ vector of independent disturbances with zero mean and constant variance. We shall not only examine the fixed start-up case – as in KP (1993, 1994) – but also the more general case where y_0 may be random. We shall find it convenient to rewrite (1.1) as

$$y = Z\alpha + u, \tag{1.2}$$

where $\alpha' = (\lambda, \beta')$ and $Z = (y_{-1}, X)$. The OLS estimator of α is

$$\hat{\alpha} = (Z'Z)^{-1}Z'y = \alpha + (Z'Z)^{-1}Z'u, \tag{1.3}$$

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

$$B_\alpha = E(\hat{\alpha} - \alpha) = E[(Z'Z)^{-1}Z'u]. \tag{1.4}$$

Below, higher-order approximations up to $O(T^{-2})$ are derived for B_α by expanding the right-hand side of (1.3). All proofs are presented in Appendices. The first Appendix A contains a Lemma with frequently employed results on expectations of certain products of quadratic forms in vectors of independent normal random variables. The bias approximation for the general case is presented in Section 2. Then in Section 3 we specialize this result for an AR(1) model without or with an intercept, and compare our results with those already given in the literature. In Section 4 we use empirical data to present numerical results for general autoregressive distributed lag models of the ARX(1) type, and in the final Section 5 we summarize the conclusions.

2. Second-order bias approximation

The starting point for our analysis is the following:

ASSUMPTION 1: *In the first-order dynamic regression model $y = \lambda y_{-1} + X\beta + u$, where the scalar λ and the $K \times 1$ vector β are unknown coefficients, we have: (i) stability, i.e. $|\lambda| < 1$; (ii) the matrix $Z = (y_{-1}, X)$ is such that $Z'Z = O_p(T)$; (iii) the $T \times (K+1)$ matrix Z has* $\text{rank}(Z) = K+1$ *with probability one; (iv) the regressors in X are strongly exogenous; (v) the disturbances follow $u \sim N(0, \sigma^2 I_T)$, with $0 < \sigma < \infty$; (vi) the start-up value has $y_0 \sim N(\bar{y}_0, \omega^2\sigma^2)$, with $0 \le \omega < \infty$; (vii) y_0 and u are mutually independent.*

Note that $\omega = 0$ represents the fixed start-up case, and for $\omega > 0$ the start-up is random; if $\omega = (1-\lambda^2)^{-1/2}$ then y_t has constant variance conditional on X. At the end of this section we shall demonstrate that our results are still applicable when relaxing item (ii) of this assumption and allowing for non-stationary regressors.

In what follows we shall condition on the observed matrix X. In order to distinguish the fixed and stochastic elements of the lagged dependent explanatory variable y_{-1}, we define the $T \times 1$ vector $\bar{y}^*_{-1}$ and the $(T+1) \times (T+1)$ diagonal matrix Ω as

$$\bar{y}^*_{-1} = \begin{pmatrix} \bar{y}_0 \\ x_1'\beta \\ \cdot \\ \cdot \\ x_{T-1}'\beta \end{pmatrix} \text{ and } \Omega = \begin{pmatrix} \omega & 0 & \cdot & \cdot & 0 \\ 0 & 1 & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & 1 \end{pmatrix}, \tag{2.1}$$

where $(x_1, ..., x_T) = X'$. We also define the $(T+1) \times 1$ random vector

$$v = (u_0, u') = (u_0, ..., u_T)' \sim N(0, \sigma^2 I_{T+1}), \tag{2.2}$$

and introduce the $T \times T$ matrix

$$\Lambda = \begin{pmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ -\lambda & 1 & & & & \cdot \\ 0 & -\lambda & 1 & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & -\lambda & 1 \end{pmatrix} \text{ with } \Lambda^{-1} = \begin{pmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ \lambda & 1 & \cdot & & & \cdot \\ \lambda^2 & \lambda & 1 & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & & \cdot & \cdot & \cdot & 0 \\ \lambda^{T-1} & \cdot & \cdot & \lambda^2 & \lambda & 1 \end{pmatrix}. \tag{2.3}$$

Employing (2.1) through (2.3) we find from (1.1) that we may write

$$\Lambda y_{-1} = \bar{y}^*_{-1} + (I_T, 0)\Omega v,$$

where $\omega u_0 = y_0 - \bar{y}_0$. Premultiplying by Λ^{-1} yields

$$y_{-1} = \Lambda^{-1}\bar{y}^*_{-1} + \Lambda^{-1}(I_T, 0)\Omega v = \bar{y}_{-1} + Gv, \tag{2.4}$$

where we introduced the $T \times 1$ vector

$$\bar{y}_{-1} = E(y_{-1}) = \Lambda^{-1}\bar{y}^*_{-1}, \tag{2.5}$$

and the $T \times (T+1)$ matrix

$$G = \Lambda^{-1}(I_T, 0)\Omega. \tag{2.6}$$

The vector $\bar{y}_{-1}$ denotes the deterministic part of y_{-1} (taken to be the mathematical expectation conditional on X). The second term of (2.4), Gv, is the remaining stochastic part of y_{-1}, which has mean zero.

If we write $\bar{Z}$ for the deterministic part of Z, then

$$\bar{Z} = E(Z) = (\bar{y}_{-1}, X), \tag{2.7}$$

while the zero-mean stochastic part of Z can now be expressed as

$$\tilde{Z} = Z - \bar{Z} = (Gv, O) = Gve_1', \tag{2.8}$$

where G is given in (2.6) and e_i denotes the $K+1$ element unit vector with i^{th} component unity. The decomposition $Z = \bar{Z} + \tilde{Z}$ will be used extensively below.

In the earlier KP papers, where we focused on the fixed start-up case, we had

$$\tilde{Z} = Cue_1', \tag{2.9}$$

with C the $T \times T$ matrix

$$C = \begin{pmatrix} 0 & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & 0 & \cdot & & & \cdot \\ \lambda & 1 & 0 & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & & \cdot & \cdot & \cdot & 0 \\ \lambda^{T-2} & \cdot & \cdot & \lambda & 1 & 0 \end{pmatrix} = \Lambda^{-1}L, \text{ where } L = \begin{pmatrix} 0 & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & 0 & \cdot & & & \cdot \\ 0 & 1 & 0 & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & 1 & 0 \end{pmatrix}. \tag{2.10}$$

It is obvious that the present more general setup, where $\tilde{Z}$ is given by (2.8), simplifies to (2.9) when we take $\omega = 0$. We get for $\omega = 0$

$$G = \Lambda^{-1}(I_T, 0)\Omega = \Lambda^{-1}(0, L) = (0, C), \tag{2.11}$$

which yields $Gv = (0, C)v = Cu$ in the fixed start-up case. Obviously, the fixed part $\bar{Z}$ of Z is unaffected by allowing for a random start-up (except that the expected value $\bar{y}_0$ instead of y_0 is put in the top-left position).

Before we proceed, we derive a simple result which is valid for any value of ω and allows some simplification of the expressions to be evaluated below, viz.:

$$G(0, I_T)' = \Lambda^{-1}(I_T, 0)\Omega(0, I_T)' = \Lambda^{-1}L = C. \tag{2.12}$$

Upon defining now the $(K+1) \times (K+1)$ matrix D as

$$D = Z'Z \tag{2.13}$$

and exploiting the results given above, we find that the deterministic part of D is

$$\bar{D} = E(D) = E(\bar{Z} + Gve_1')'(\bar{Z} + Gve_1') = \bar{Z}'\bar{Z} + \sigma^2 \operatorname{tr}(GG')e_1e_1'. \tag{2.14}$$

In order to keep the expressions in the results to follow as compact as possible we introduce some further simplifying notation. We use the matrix Q to denote the $(K+1)\times(K+1)$ matrix $(\bar{D})^{-1}$, and q_1 denotes the first column of Q, whereas q_1 has first element q_{11}, hence:

$$Q = (\bar{D})^{-1}, \qquad q_1 = (\bar{D})^{-1}e_1, \qquad q_{11} = e_1'(\bar{D})^{-1}e_1. \tag{2.15}$$

The following result is proved in Appendix B.

Theorem 1: *Under Assumption 1, which according to KP (2009, Appendix A) guarantees the order of the approximation error, the bias B_α of the least-squares estimator $\hat{\alpha}$ in (1.3) can be approximated by $B_\alpha(T^{-2})$, where $B_\alpha = B_\alpha(T^{-2}) + o(T^{-2})$, with $B_\alpha(T^{-2}) =$*

$$\begin{aligned}
&-\sigma^2[\operatorname{tr}(Q\bar{Z}'C\bar{Z})q_1 + Q\bar{Z}'C\bar{Z}q_1] \\
&+\sigma^4\{[-2q_{11}\operatorname{tr}(GG'C) + 2q_{11}\operatorname{tr}(Q\bar{Z}'GG'C\bar{Z}) + 2q_{11}\operatorname{tr}(Q\bar{Z}'GG'C'\bar{Z}) \\
&\qquad -2q_{11}\operatorname{tr}(Q\bar{Z}'GG'\bar{Z}Q\bar{Z}'C\bar{Z}) - q_{11}\operatorname{tr}(Q\bar{Z}'C\bar{Z})\operatorname{tr}(Q\bar{Z}'GG'\bar{Z}) + 4q_1'\bar{Z}'GG'C\bar{Z}q_1 \\
&\qquad +2q_1'\bar{Z}'GG'C'\bar{Z}q_1 - 4q_1'\bar{Z}'GG'\bar{Z}Q\bar{Z}'C\bar{Z}q_1 - 2q_1'\bar{Z}'GG'\bar{Z}Q\bar{Z}'C'\bar{Z}q_1 \\
&\qquad -q_1'\bar{Z}'GG\bar{Z}q_1\operatorname{tr}(Q\bar{Z}'C\bar{Z}) - 2q_1'\bar{Z}'C\bar{Z}q_1\operatorname{tr}(Q\bar{Z}'GG'\bar{Z})]q_1 \\
&\quad -[q_{11}\operatorname{tr}(Q\bar{Z}'GG'\bar{Z}) + q_1'\bar{Z}'GG\bar{Z}q_1]Q\bar{Z}'C\bar{Z}q_1 \\
&\quad -2[q_{11}\operatorname{tr}(Q\bar{Z}'C\bar{Z}) + q_1'\bar{Z}'C\bar{Z}q_1]Q\bar{Z}'GG'\bar{Z}q_1 \\
&\quad +2q_{11}Q\bar{Z}'[GG'C + CGG' + GG'C']\bar{Z}q_1 \\
&\quad -2q_{11}Q\bar{Z}'GG'\bar{Z}Q\bar{Z}'[C + C']\bar{Z}q_1 - 2q_{11}Q\bar{Z}'C\bar{Z}Q\bar{Z}'GG'\bar{Z}q_1\} \\
&+\sigma^6\{[8q_{11}^2\operatorname{tr}(GG'GG'C) - 2q_{11}^2\operatorname{tr}(GG'GG')\operatorname{tr}(Q\bar{Z}'C\bar{Z}) - 4q_{11}^2\operatorname{tr}(GG'C)\operatorname{tr}(Q\bar{Z}'GG'\bar{Z}) \\
&\qquad -12q_{11}(q_1'\bar{Z}'GG'\bar{Z}q_1)\operatorname{tr}(GG'C) - 8q_{11}(q_1'\bar{Z}'C\bar{Z}q_1)\operatorname{tr}(GG'GG')]q_1 \\
&\quad -2q_{11}^2\operatorname{tr}(GG'GG')\,Q\bar{Z}'C\bar{Z}q_1 - 8q_{11}^2\operatorname{tr}(GG'C)\,Q\bar{Z}'GG'\bar{Z}q_1\} \\
&-\sigma^8[12q_{11}^3\operatorname{tr}(GG'C)\operatorname{tr}(GG'GG')\,q_1].
\end{aligned}$$

Compared with $B_\alpha(T^{-1})$, the bias to order T^{-1} derived in KP (1993, Theorem 7) and also given in (B.10), we see that the approximation to order T^{-2} is far more complex.

When interest centers on one of the coefficients of α, the required bias approximation can be obtained on noting that $\alpha_i = e_i'\alpha$, $i = 1, ..., K+1$. In particular $\lambda = e_1'\alpha$ and $E(\hat{\lambda} - \lambda) = e_1'E(\hat{\alpha} - \alpha)$. Writing $B_\lambda(T^{-2}) = e_1'B_\alpha(T^{-2})$ for the bias approximation of $\hat{\lambda}$, we may deduce the following result.

Corollary 1: *Under Assumption 1, the bias B_λ of the least-squares estimator $\hat{\lambda}$ in (1.1) can be approximated by $B_\lambda(T^{-2})$, where $B_\lambda = B_\lambda(T^{-2}) + o(T^{-2})$, with $B_\lambda(T^{-2}) =$*

$$\begin{aligned}
&-\sigma^2[q_{11}\operatorname{tr}(Q\bar{Z}'C\bar{Z}) + q_1'\bar{Z}'C\bar{Z}q_1] \\
&+\sigma^4[-2q_{11}^2\operatorname{tr}(GG'C) + 2q_{11}^2\operatorname{tr}(Q\bar{Z}'GG'C\bar{Z}) + 2q_{11}^2\operatorname{tr}(Q\bar{Z}'GG'C'\bar{Z}) \\
&\quad -2q_{11}^2\operatorname{tr}(Q\bar{Z}'GG'\bar{Z}Q\bar{Z}'C\bar{Z}) - q_{11}^2\operatorname{tr}(Q\bar{Z}'C\bar{Z})\operatorname{tr}(Q\bar{Z}'GG'\bar{Z}) \\
&\quad +6q_{11}q_1'\bar{Z}'GG'(C + C')\bar{Z}q_1 - 6q_{11}q_1'\bar{Z}'GG'\bar{Z}Q\bar{Z}'C\bar{Z}q_1 \\
&\quad -6q_{11}q_1'\bar{Z}'GG'\bar{Z}Q\bar{Z}'C'\bar{Z}q_1 - 3q_{11}q_1'\bar{Z}'GG\bar{Z}q_1\operatorname{tr}(Q\bar{Z}'C\bar{Z}) \\
&\quad -3q_{11}q_1'\bar{Z}'C\bar{Z}q_1\operatorname{tr}(Q\bar{Z}'GG'\bar{Z}) - 3q_1'\bar{Z}'GG\bar{Z}q_1q_1'\bar{Z}'C\bar{Z}q_1] \\
&+\sigma^6[8q_{11}^3\operatorname{tr}(GG'GG'C) - 2q_{11}^3\operatorname{tr}(GG'GG')\operatorname{tr}(Q\bar{Z}'C\bar{Z}) - 4q_{11}^3\operatorname{tr}(GG'C)\operatorname{tr}(Q\bar{Z}'GG'\bar{Z}) \\
&\quad -20q_{11}^2q_1'\bar{Z}'GG'\bar{Z}q_1\operatorname{tr}(GG'C) - 10q_{11}^2q_1'\bar{Z}'C\bar{Z}q_1\operatorname{tr}(GG'GG')] \\
&-\sigma^8[12q_{11}^4\operatorname{tr}(GG'C)\operatorname{tr}(GG'GG')].
\end{aligned}$$

Again only three terms of this formula constitute the bias approximation to the order of T^{-1}, viz.

$$B_\lambda(T^{-1}) = -\sigma^2[q_{11}\,\mathrm{tr}(Q\bar{Z}'C\bar{Z}) + q_1'\bar{Z}'C\bar{Z}q_1] - 2\sigma^4 q_{11}^2\,\mathrm{tr}\,(GG'C)\,. \tag{2.16}$$

Due to the way in which we present the various terms in the bias approximations given above, one could easily get the impression that they also establish small disturbance approximations. That is not the case, since all elements of Q depend on σ. However, small disturbance asymptotic results can be readily obtained. From (2.14) and (2.15) we have

$$Q = [\bar{Z}'\bar{Z} + \sigma^2\,\mathrm{tr}(GG')e_1e_1']^{-1}. \tag{2.17}$$

It is easily verified that this can be expressed equivalently as

$$Q = (\bar{Z}'\bar{Z})^{-1} - \frac{\sigma^2\,\mathrm{tr}(GG')}{1+\sigma^2\,\mathrm{tr}(GG')e_1'(\bar{Z}'\bar{Z})e_1}(\bar{Z}'\bar{Z})^{-1}e_1e_1'(\bar{Z}'\bar{Z})^{-1}. \tag{2.18}$$

Introducing the notation

$$P = (\bar{Z}'\bar{Z})^{-1},\ p_1 = Pe_1,\ p_{11} = e_1'p_1,\ A = \mathrm{tr}(GG')p_1p_1',\ a = \mathrm{tr}(GG')p_{11}, \tag{2.19}$$

Q can also be written as

$$Q = P - \frac{\sigma^2}{1+\sigma^2 a}A = P - (\sigma^2 - \sigma^4 a + \sigma^6 a^2 - \sigma^8 a^3 + ...)A. \tag{2.20}$$

Substituting (2.20) in the above approximation yields small disturbance approximations. Since $P = O(T^{-1})$, $A = O(T^{-1})$ and $a = O(1)$ all terms in the expansion (2.20) are of order T^{-1}. This implies that all terms in the large sample approximations presented above contain terms of low up to infinitely high order in σ. Correspondingly, a finite order small-σ approximation will omit terms that cannot be neglected from a large-T perspective. Therefore, we do not expect it to be fruitful to further pursue small disturbance asymptotic results in the context of stable dynamic models. Also note, that in the simplest case, where $K = 0$ and $\bar{y}_0 = 0$, and hence $\bar{Z} = 0$, the small disturbance approximation is not defined. This is also the case when $K > 0$ and $\beta = 0$.

Before we analyse in the next sections the usefulness of the approximation in Theorem 1, we want to discuss the restrictiveness of Assumption 1. The examination of any differences of the present result with similar results but derived under alternative or weaker conditions, such as non-stable dynamic relationships with $\lambda = 1$ or $\lambda > 1$, models with higher-order dynamics or which include weakly exogenous regressors and may have non-normal disturbance terms is deferred to future research. It is easy to show, however, that the effects of relaxing (ii), i.e. the inclusion of non-stationary regressors which involve deterministic or stochastic trends, are quite straightforward. Assume that for $i = 1, ..., K+1$ the series of real positive constants d_i is given, such that $(K+1)\times(K+1)$ diagonal matrix $\Phi = \mathrm{diag}(\phi_i)$ can be constructed which for $\phi_i = T^{-d_i}$ yields $Z_* = Z\Phi$ such that $Z_*'Z_* = O_p(T)$. Now $\Phi^{-1}(\hat{\alpha} - \alpha) = (Z_*'Z_*)^{-1}Z_*'u$ and $E[(Z_*'Z_*)^{-1}Z_*'u]$ can be approximated to the order of $O(T^{-2})$ by the formula of Theorem 1, upon making the required substitutions. Note, however, that $\hat{\alpha} - \alpha = (Z'Z)^{-1}Z'u = \Phi(Z_*'Z_*)^{-1}Z_*'u$, so in order to find the bias in $\hat{\alpha}$ we have to premultiply by Φ. This leads to the following result.

Corollary 2: *The formula of Theorem 1 also applies when Z contains non-stationary regressors, but then the i^{th} element of B_α is approximated by terms of order $O(T^{-1-d_i})$ and $O(T^{-2-d_i})$ respectively, with remainder term $o(T^{-2-d_i})$, where the values $d_i \geq 0$ are as given above.*

Note that in the presence of non-stationary regressors the approximation indicated by $B_\alpha(T^{-2})$ still has a remainder term of $o(T^{-2})$ but for particular elements it may be smaller. Hence, item (ii) of Assumption 1 refers actually to a worst case situation, and relaxing it does not weaken our results.

3. Bias approximations for AR(1) models

In this section, we shall examine and compare the accuracy of our approximations of order T^{-1} and order T^{-2} in various simple AR(1) models. We consider the bias of the least-squares estimator for data that have been generated by a first-order Markov process, both for the case where the data have known mean and for the AR(1) model with unknown mean. In the latter case the least-squares estimator is obtained from a regression where an intercept term has been included. We compare the results of our formulas with estimates of the true bias obtained from extensive simulation experiments and also with the results obtained by the approximation formulae that have been derived and published in the past for some special members of this simple class of first-order autoregressive models. Some of these classic bias approximations pertain to the AR(1) model with a fixed initial value of the process so that the series is not strictly stationary; others have been derived on the basis of a random initial value with a variance such that the series is sometimes covariance-stationary and sometimes not. Our framework encompasses all these different situations.

In the various results below we shall affix a superscript F to expressions referring to the bias in models with a fixed start-up ($\omega = 0$). So, the random start-up is the default case, and then the results involve some choice for ω. We (also) put the letter M in the superscript when the start-up value is such that the process is mean-stationary, and replace this by S if the process is covariance stationary too, which can only occur in the random start-up case. For clarity we also add the superscript label NC for results that pertain to the least-squares bias in the AR(1) model with no constant term (which is equivalent to the model with known mean), and C for the model with an intercept. The model with no intercept will be investigated first, followed by the model with an intercept.

3.1. The AR(1) model without intercept

In this model the expectation of $\hat{\lambda}$ exists for $T > 3$, see Evans and Savin (1981, p.767). White (1961) presents the first extensive analysis of first- and higher-order approximations of the expectation and the variance of coefficient estimators in the AR(1) model with no intercept. Although he considers the case where the initial value of the series may be either fixed or random, he also excludes this initial value from the least-squares estimation procedure. This is irrelevant, of course, when $y_0 = 0$, but otherwise it is not. White obtains his results by integrating term wise an expansion of a complicated integrand and this yields a power series in λ, which he then reduces to a power series in T^{-1}. For the model with zero fixed start-up (i.e. $y_0 = 0$, $\omega = 0$, $K = 0$) White (1961, pp.87-89) finds (note that $|\lambda| < 1$) the small-λ asymptotic approximation:

$$\begin{aligned} W_{\lambda}^{FM,NC}(\lambda^5) &= \frac{-2(T-2)}{(T-1)(T+1)}\lambda + \frac{12}{(T+1)(T+3)(T+5)}\lambda^3 \\ &\quad + \frac{18(T+8)}{(T+3)(T+5)(T+7)(T+9)}\lambda^5, \end{aligned} \tag{3.1}$$

where $B_{\lambda}^{FM,NC} = W_{\lambda}^{FM,NC}(\lambda^5) + o(\lambda^5)$. Since $\frac{-2(T-2)}{(T-1)(T+1)} = -2T^{-1} + 4T^{-2} + O(T^{-3})$ and supposing that (3.1) is accurate to order T^{-2}, White (1996, formula 9) derives a result from (3.1) which implies

$$W_{\lambda}^{FM,NC}(T^{-2}) = -2\frac{\lambda}{T} + 4\frac{\lambda}{T^2}, \tag{3.2}$$

where $B_{\lambda}^{FM,NC} = W_{\lambda}^{FM,NC}(T^{-2}) + o(T^{-2})$. Shenton and Johnson (1965) examine the zero-mean AR(1) model also and focus exclusively on the fixed start-up case. They distinguish explicitly between approximations in ascending powers of λ and in descending powers of T respectively and present results which are accurate to even higher orders of approximation than White's. They obtain (see their formula 18a for $T > 6$) the ninth-order small-λ approximation

$$\begin{aligned} SJ_\lambda^{FM,NC}(\lambda^9) &= \frac{-2(T-2)}{(T+1)^{[2]}}\lambda + \frac{12}{(T+5)^{[3]}}\lambda^3 + \frac{18(T+8)}{(T+9)^{[4]}}\lambda^5 \\ &+ \frac{24(T+12)^{[2]}}{(T+13)^{[5]}}\lambda^7 + \frac{30(T+16)^{[3]}}{(T+17)^{[6]}}\lambda^9, \end{aligned} \tag{3.3}$$

where $x^{[n]} = x(x-2)...(x-2n+2)$ and $B_\lambda^{FM,NC} = SJ_\lambda^{FM,NC}(\lambda^9) + o(\lambda^9)$. Their separately derived large-T result (formula 21a) generalizes (3.3). It says

$$\begin{aligned} SJ_\lambda^{FM,NC}(T^{-6}) &= -2\frac{\lambda}{T} + 4\frac{\lambda}{T^2} - 2\frac{\lambda}{T^3}\frac{1-8\lambda^2+4\lambda^4}{(1-\lambda^2)^2} + 4\frac{\lambda}{T^4}\frac{1-30\lambda^2+12\lambda^4-4\lambda^6}{(1-\lambda^2)^3} \\ &-2\frac{\lambda}{T^5}\frac{1-352\lambda^2-204\lambda^4-64\lambda^6+16\lambda^8}{(1-\lambda^2)^4} \\ &+4\frac{\lambda}{T^6}\frac{1-995\lambda^2-2780\lambda^4-1240\lambda^6+80\lambda^8-16\lambda^{10}}{(1-\lambda^2)^5}, \end{aligned} \tag{3.4}$$

where $B_\lambda = SJ_\lambda^{FM,NC}(T^{-6}) + o(T^{-6})$.

Specializing our results of Corollary 1 for this particular case ($\bar{y}_0 = 0$, $\omega = 0$, $K = 0$), where the matrix $\bar{Z}$ simplifies to a vector of zero elements and the matrix $\bar{D}$ to the scalar $\sigma^2 \operatorname{tr}(C'C)$, we find

$$B_\lambda^{FM,NC}(T^{-1}) = -2\frac{\operatorname{tr}(CC'C)}{[\operatorname{tr}(C'C)]^2} \tag{3.5}$$

and

$$B_\lambda^{FM,NC}(T^{-2}) = -2\frac{\operatorname{tr}(CC'C)}{[\operatorname{tr}(C'C)]^2} + 8\frac{\operatorname{tr}(CC'CC'C)}{[\operatorname{tr}(C'C)]^3} - 12\frac{\operatorname{tr}(CC'C)\operatorname{tr}(C'CC'C)}{[\operatorname{tr}(C'C)]^4}. \tag{3.6}$$

Note that it is not possible to compare (3.5) and (3.6) directly with the corresponding terms in (3.3) or (3.4), because our results are not explicit in powers of T^{-1} or λ. In fact, (3.5) also contains terms of order $o(T^{-1})$ and similarly, we could remove $o(T^{-2})$ terms from (3.6). This is called "filtering" below, and involves exploiting the basic results collected in Appendix C. Upon doing so for the case involving arbitrary values of $\bar{y}_0$ and ω, we find for model (1.1) with $K = 0$ (see Appendix D for the proof):

THEOREM 2: *Under the conditions (i), (v), (vi) and (vii) of Assumption 1, the bias B_λ^{NC} of the OLS estimator $\hat{\lambda}$, obtained from a sample ($t = 1, ..., T$) of the AR(1) model with no intercept $y_t = \lambda y_{t-1} + \varepsilon_t$, can be approximated by the expression $KP_\lambda^{NC}(T^{-2})$, where $B_\lambda^{NC} = KP_\lambda^{NC}(T^{-2}) + o(T^{-2})$, with*

$$KP_\lambda^{NC}(T^{-2}) = -2\frac{\lambda}{T} + 2\frac{\lambda}{T^2}\left(2 + \frac{\bar{y}_0^2}{\sigma^2} + \omega^2\right).$$

From this result we note that the order T^{-1} bias is not at all affected by the nature (stochastic or not) and the (expected) value (or variance) of the initial observation y_0. That the order T^{-1} result is rather robust is also illustrated by the fact that Marriott and Pope (1954) already found the bias approximation $-2\lambda/T$ for an estimator of the first-order serial correlation coefficient (which differs slightly from our estimator $\hat{\lambda}$) in the Markov scheme or stationary zero-mean

AR(1) process. Note that for $\bar{y}_0 = 0$ and $\omega = 0$ Theorem 2 re-establishes White's specific second-order result (3.2).

We shall now make some numerical comparisons between the various approximations given above. Table 1 contains results on the bias of $\hat{\lambda}$ in the zero start-up mean-stationary AR(1) model estimated without intercept. We present a Monte Carlo estimate of the true bias. All Monte Carlo estimates presented in this study have been obtained from 500,000 simulation experiments and therefore their accuracy will be such that we simply label them as "true bias". Our estimates of $B_{\lambda}^{FM,NC}$ conform to three decimal places with corresponding values published in Tsui and Ali (1992, 1994) and in the study by Vinod and Shenton (1996), who calculate the exact bias for this specific case by Gaussian quadrature methods (and therefore constitute approximations to the true bias too). We see that the two $O(T^{-1})$ approximations $-2\lambda/T$ and (3.5) are very close to each other for the smaller λ values, especially for larger T values. For high values of λ they show a substantial difference, even at $T = 50$. As a rule, both $O(T^{-1})$ approximations overstate the severity of the (negative) bias, except for λ close to the unit circle. In this area White's simple formula still overstates the bias (by at least 15%), and our formula understates it (by 15% or less). Hence, White's simple formula is often reasonable, but the more involved order T^{-1} formula (3.5) seems slightly better. A similar relationship is not found for the two $O(T^{-2})$ approximations. On the whole, White's $O(T^{-2})$ approximation (3.2) is remarkably good, but less so close to the unit circle. In the model with sample size $T = 10$ it overstates the bias for $\lambda > 0.5$, whereas this happens for $\lambda > 0.8$ in the $T = 50$ case. The unfiltered second-order approximation (3.6) is less accurate, especially so close to the unit root, where it substantially overstates the actual bias and is even worse than White's first-order formula. Hence, we find that for the near unit root case the $o(T^{-2})$ terms that have not been removed from $B_{\lambda}^{FM,NC}(T^{-2})$ do more harm than good. Furthermore, Table 1 shows that the higher-order large-T Shenton and Johnson formulas give very poor results for large values of λ. Only when λ and T are both small do the extra terms in the SJ formulas sometimes lead to an improved approximation. This deterioration of the higher-order results may seem surprising but it is less so when one considers the detrimental effects for the near unit root case of the $(1-\lambda)$ terms in the denominators of the higher-order terms of (3.4). It is striking, however, how accurate the ninth-order small-λ approximation (3.3) is over the whole range of λ and T values examined here. Evans and Savin (1981, p.770) have already noted the high precision of White's fifth-order formula (3.1). Here we re-establish the impressive accuracy of the small-λ approximation, which is found to be superior especially when λ is not at all small, viz. when λ is close to one. This phenomenon can be explained following similar arguments as used below (2.20) when small disturbance approximations were disqualified for dynamic models. Formula (3.2) has been obtained from (3.1) and obviously the $O(\lambda)$ approximation implicit in (3.1) is accurate also to order $O(T^{-1})$ and even to order $O(T^{-2})$, because the T^{-1} and T^{-2} terms happen to be of order $O(\lambda)$. For the reverse, however, we immediately observe that the $O(T^{-1})$ approximation is not accurate to $O(\lambda)$, and nor is the $O(T^{-2})$ approximation. Even for the $O(T^{-6})$ approximation (3.4) we find after collecting all terms of order λ that it involves:

$$\begin{aligned}\lambda\left(-\frac{2}{T}+\frac{4}{T^2}\right)\left(1+\frac{1}{T^2}+\frac{1}{T^4}\right) &\neq -2\lambda\frac{T-2}{(T-1)(T+1)} \\ &= \lambda\left(-\frac{2}{T}+\frac{4}{T^2}\right)\left(1+\frac{1}{T^2}+\frac{1}{T^4}+...\right).\end{aligned}$$

The above shows that it requires a large-T approximation of infinitely large order to obtain a small-λ approximation which is correct to first-order in λ. Hence, in this very special model, where small disturbance asymptotics is not even defined, large sample asymptotics does not seem very well suited either, because a large-T approximation of finite order omits terms of order λ, which may be substantial when T is moderate and λ not small.

We shall now consider the mean-stationary AR(1) model with random start-up and no intercept. Because White always excludes the initial value from the least-squares estimation formula, his results with fixed start-up $y_0 = 0$ can also be interpreted as having a sample size of $T-1$ and a random start-up with $\bar{y}_0 = 0$ and $\omega = 1$. So replacing his T with $T+1$ we find, upon removing terms of smaller order than T^{-2}, from his formula (9)

$$W_\lambda^{M,NC}(T^{-2}) = -2\frac{\lambda}{T+1} + 4\frac{\lambda}{(T+1)^2} = -2\frac{\lambda}{T} + 6\frac{\lambda}{T^2} + o(T^{-2}), \qquad (\omega = 1) \qquad (3.7)$$

where $B_\lambda^{M,NC} = W_\lambda^{M,NC}(T^{-2}) + o(T^{-2})$. This conforms indeed to the approximation formula obtained from the general Theorem 2 for $\bar{y}_0 = 0$ and $\omega = 1$.

From Theorem 2 it is also simply found that the bias in the strongly stationary case, where $\bar{y}_0 = 0$ and $\omega^2 = (1-\lambda^2)^{-1}$, is

$$KP_\lambda^{S,NC}(T^{-2}) = \frac{-2\lambda}{T} + \frac{2\lambda}{T^2}\left(2 + \frac{1}{1-\lambda^2}\right). \qquad (3.8)$$

Unlike (3.4) and (3.7), we see that now also the order T^{-2} term may be problematic for λ values close to unity. White analyzed the strongly stationary case too. In order to avoid confusion when citing his results, we shall again replace his T with $T+1$ for models where $y_0 \neq 0$, and so translate his bias approximations in terms of our framework. Thus, White (1961, p.90) yields

$$\begin{aligned} W_\lambda^{S,NC}(\lambda^5) &= -2\frac{T-1}{(T+2)^{[2]}}\lambda + 2\frac{T^2+10T-13}{(T+6)^{[4]}}\lambda^3 \\ &\quad +4\frac{T^4+28T^3+180T^2+37T+24}{(T+10)^{[6]}}\lambda^5, \qquad (3.9) \end{aligned}$$

where $B_\lambda^{S,NC} = W_\lambda^{S,NC}(\lambda^5) + o(\lambda^5)$. Note that in (3.9) the term of order λ does contain all $O(T^{-1})$ contributions, but not all those of order $O(T^{-2})$, as was the case in (3.1). Now all coefficients of the power series in λ involve contributions of order $O(T^{-2})$ and an infinite power series in λ is required in order to achieve accuracy of order $O(T^{-2})$. White (1961, formula 11) reduces the terms of (3.9) to a large-T approximation, and then conjectures

$$W_\lambda^{S,NC}(T^{-2}) = \frac{-2\lambda}{T+1} + \frac{2\lambda}{(T+1)^2} + \frac{2\lambda}{(T+1)^2}\left(1+\lambda^2+\lambda^4+\ldots\right).$$

From this we deduce

$$\begin{aligned} W_\lambda^{S,NC}(T^{-2}) &= \frac{-2\lambda}{T} + \frac{4\lambda}{T^2} + \frac{2\lambda}{T^2}\frac{1}{1-\lambda^2} + o(T^{-2}) \\ &= \frac{-2\lambda}{T} + \frac{2\lambda}{T^2}\left(2+\frac{1}{1-\lambda^2}\right) + o(T^{-2}), \qquad (3.10) \end{aligned}$$

which is in agreement indeed with our (3.8).

Table 2 contains numerical results for the random start-up strongly stationary model with known mean (no intercept). Upon comparing our simulated $B_\lambda^{S,NC}$ values, which correspond to 3 decimal places with Sawa's (1978) exact results, with those of Table 1, we note that the bias in the random start-up model is less serious, especially so for large λ values and for smaller sample sizes. The simple $-2\lambda/T$ approximation, which is the same as for the fixed start-up case (where it already involved an overstatement of the actual bias), is not very accurate now, especially when T is small. Our unfiltered $O(T^{-1})$ formula $B_\lambda^{S,NC}(T^{-1})$, which is obtained by substituting $K=0$, $\bar{y}_0 = 0$ and $\omega = (1-\lambda^2)^{-1/2}$ in (2.16), is now much better, although it is extremely poor for the near unit root case. The same quality difference is found for the two $O(T^{-2})$ approximations. The unfiltered second-order approximation $B_\lambda^{S,NC}(T^{-2})$ is better than $KP_\lambda^{S,NC}(T^{-2})$ given in (3.8), which does not contain any $o(T^{-2})$ terms. The latter is extremely bad for λ close to unity (as already predicted); then it is even worse than the first-order approximations. The unfiltered second-order approximation, however, behaves quite satisfactorily, even close to the unit circle, where it is much better in this model than the small-λ approximation (3.9). The latter is only better when λ is really small, but for λ close to unity it becomes obvious that in this model any finite order small-λ approximation has an approximation error of order $O(T^{-2})$. It appears that our unfiltered approximation $B_\lambda^{S,NC}(T^{-2})$ is to be preferred here, because its approximation errors are $o(T^{-2})$ whereas its accuracy with respect to powers of λ does not seem to be bad either.

Hence, from Tables 1 and 2 we find mixed evidence on the superiority of a second-order large-T approximation over its first-order component. Even in the simplest AR(1) model the bias does not depend exclusively on T, but also on λ (and possibly $\bar{y}_0$ and ω). We established that the accuracy of the large-T approximation may seriously deteriorate in a particular area of the parameter space of λ when an extra term of the power series with respect to T^{-1} is taken into account.

We did not perform calculations for cases where the process is not mean-stationary, i.e. $\bar{y}_0 \neq 0$, nor for non covariance-stationary cases with random start-up, i.e. $\omega \neq 0$ and $\omega^2 \neq (1-\lambda^2)^{-1}$. For these, the filtered T^{-1} and T^{-2} approximations are given by Theorem 2, and unfiltered approximations follow directly from Corollary 1. Since for these settings the order T^{-2} term of $KP_\lambda^{NC}(T^{-2})$ does not have a factor $(1-\lambda)$ in its denominator, we expect that the filtered approximation of Theorem 2 behaves reasonably well here. White (1961, p.92) presents an approximation of order $O(\lambda)$ for the special case $\omega = 0$ and $y_0 = \bar{y}_0 = c\sigma$, viz.

$$W_\lambda^{NC}(\lambda) = \frac{-2\lambda}{T+2+c^2}, \tag{3.11}$$

where $B_\lambda^{NC} = W_\lambda^{NC}(\lambda) + o(\lambda)$. This result can be rewritten as

$$W_\lambda^{NC}(\lambda) = \frac{-2\lambda}{T}\left[1+\frac{1}{T}\left(2+\frac{\bar{y}_0^2}{\sigma^2}\right)\right]^{-1} = \frac{-2\lambda}{T} + \frac{2\lambda}{T^2}\left(2+\frac{\bar{y}_0^2}{\sigma^2}\right) + o(T^{-2})$$

and from Theorem 2 we find that this precisely yields the $O(T^{-2})$ approximation of the bias in $\hat{\lambda}$.

3.2. The AR(1) model with an intercept

Turning now to the only slightly more general AR(1) model with unknown mean, we find that the older literature provided only very few results, viz. the order T^{-1} approximation for $\hat{\lambda}$ given by Marriott and Pope (1954, p.394) and by Kendall (1954, p.404), apparently in the strongly stationary model. Strictly speaking, they did not examine the bias in the regression estimator but in various estimators of the serial correlation coefficient. As they mention, however, this bias is equivalent to the least-squares bias to the order of T^{-1}. They find

$$MP_\lambda^{S,C}(T^{-1}) = -\frac{1+3\lambda}{T}, \tag{3.12}$$

where $B_\lambda^{S,C} = MP_\lambda^{S,C}(T^{-1}) + o(T^{-1})$. This result has been confirmed by several authors, see for example Maekawa (1983). The first-order bias in the estimator for the intercept in the strongly stationary model has been obtained by Tanaka (1983, p.1226). He presents

$$T_\beta^{S,C}(T^{-1}) = \frac{\beta}{T}\frac{1+3\lambda}{1-\lambda}, \tag{3.13}$$

where $B_\beta^{S,C} = T_\beta^{S,C}(T^{-1}) + o(T^{-1})$.

It is obvious that Theorem 1 provides approximations to order T^{-2} for any random or fixed start-up parametrization of this model ($K = 1$) upon substitution of $X = (1, ..., 1)'$ and the appropriate values of $\alpha' = (\lambda, \beta)$, σ, ω and $\bar{y}_0$. By introducing

$$y_t^* = \frac{1}{\sigma}\left(y_t - \frac{\beta}{1-\lambda}\right), \qquad t = 0, 1, ..., T \tag{3.14}$$

it is easy to see that the general AR(1) model with intercept can be rewritten as

$$y_t^* = \lambda y_{t-1}^* + \beta^* + \frac{1}{\sigma}\varepsilon_t, \qquad \text{where } \alpha^* = (\lambda, \beta^*)' \text{ with } \beta^* \equiv 0.$$

Clearly the distribution of $\hat{\alpha}^* = (\hat{\lambda}, \hat{\beta}^*)'$ and thus its bias (approximation) are determined by the actual values of λ, ω and the mean of the transformed start-up value

$$\bar{y}_0^* = \frac{1}{\sigma}\left(\bar{y}_0 - \frac{\beta}{1-\lambda}\right) \tag{3.15}$$

only. Hence, this implies invariance of $\hat{\alpha}^*$ with respect to β, σ and $\bar{y}_0$ in the mean-stationary case, where $\bar{y}_t = \beta/(1-\lambda)$ and $\bar{y}_t^* = 0$ for $t = 0, ..., T$.

Focusing on the transformed mean-stationary case (3.14), and specializing the result of Theorem 1 for $\sigma = 1$, $\beta = 0$, $\bar{y}_0 = 0$ and $\bar{Z} = (0, \iota)$, where ι is a column of unit elements, we find that Q is a diagonal matrix now, so that $\bar{Z}q_1 = 0$ and $e_2'q_1 = 0$. Therefore all terms of $e_2'B_{\alpha^*}^{M,C}(T^{-2})$ vanish and we find that the least-squares estimator of β^* is unbiased to the order of T^{-2} in the mean-stationary AR(1) model with unknown intercept, i.e. $B_{\beta^*}^{M,C}(T^{-2}) = 0$. From this we can show that the approximations (3.12) and (3.13) match, even in the mean-stationary model, for the following reasons. Note that

$$\hat{\beta}^* = \frac{1}{T}\sum_{t=1}^{T} y_t^* - \hat{\lambda}\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^* = \frac{1}{\sigma}\hat{\beta} - \frac{1-\hat{\lambda}}{1-\lambda}\frac{\beta}{\sigma} = \frac{\hat{\beta}-\beta}{\sigma} - \frac{\hat{\lambda}-\lambda}{1-\lambda}\frac{\beta}{\sigma}, \tag{3.16}$$

from which it is easy to derive

$$B_\beta^{M,C} = \sigma E(\hat{\beta}^*) - \frac{\beta}{1-\lambda}B_\lambda^{M,C} = -\frac{\beta}{1-\lambda}B_\lambda^{M,C} + o(T^{-2}). \tag{3.17}$$

In the mean-stationary case, the bias approximation formula for $\hat{\lambda}$ simplifies considerably as well. This is due again to $\bar{Z}q_1 = 0$, but also to $q_{11} = [\mathrm{tr}(GG')]^{-1}$, $q_{22} = T^{-1}$ and $\bar{Z}Q\bar{Z}' = T^{-1}\iota\iota'$. From Corollary 1 we obtain

$$\begin{aligned} B_\lambda^{M,C}(T^{-2}) \;=\; & -T^{-1}[\mathrm{tr}(GG')]^{-1}\iota'C\iota - 2[\mathrm{tr}(GG')]^{-2}\,\mathrm{tr}(GG'C) \\ & +T^{-1}[\mathrm{tr}(GG')]^{-2}[2\iota'GG'C\iota + 2\iota'GG'C'\iota - 3T^{-1}\iota'GG'\iota\iota'C\iota] \\ & -T^{-1}[\mathrm{tr}(GG')]^{-3}[2\iota'C\iota\,\mathrm{tr}(GG'GG') + 4\iota'GG'\iota\,\mathrm{tr}(GG'C)] \\ & +8[\mathrm{tr}(GG')]^{-3}\,\mathrm{tr}(GG'GG'C) - 12[\mathrm{tr}(GG')]^{-4}\,\mathrm{tr}(GG'C)\,\mathrm{tr}(GG'GG'), \end{aligned} \tag{3.18}$$

which for $\omega = 0$ yields the counterpart of (3.6) for the model with unknown intercept. From Corollary 1 we can also obtain for the general AR(1) model with unknown intercept a comprehensive result as in Theorem 2, from which all $o(T^{-2})$ contributions have been removed by exploiting the basic results collected in Appendix C. It is given below (see Appendix E for the proof).

Theorem 3: *Under Assumption 1, the bias B_λ^C of the OLS estimator $\hat{\lambda}$, obtained from a sample ($t = 1, ..., T$) of the AR(1) model $y_t = \beta + \lambda y_{t-1} + \varepsilon_t$, can be approximated by the* *expression* $KP_\lambda^C(T^{-2})$, *where* $B_\lambda^C = KP_\lambda^C(T^{-2}) + o(T^{-2})$, *with*

$$KP_\lambda^C(T^{-2}) = -\frac{1+3\lambda}{T} - \frac{1-3\lambda+9\lambda^2}{T^2(1-\lambda)} + \frac{1+3\lambda}{T^2}\left[\frac{1}{\sigma^2}\left(\bar{y}_0 - \frac{\beta}{1-\lambda}\right)^2 + \omega^2\right].$$

From this result, which is in agreement with formula (8) from Bao (2007), we directly obtain the special result for the mean-stationary model with fixed start-up

$$KP_\lambda^{FM,C}(T^{-2}) = -\frac{1+3\lambda}{T} - \frac{1-3\lambda+9\lambda^2}{T^2(1-\lambda)}, \tag{3.19}$$

where $B_\lambda^{FM,C} = KP_\lambda^{FM,C}(T^{-2}) + o(T^{-2})$. In the random start-up model we find for the strongly stationary case

$$KP_\lambda^{S,C}(T^{-2}) = -\frac{1+3\lambda}{T} - \frac{\lambda}{T^2(1-\lambda)}\left(9\lambda - 3 - \frac{2}{1+\lambda}\right), \tag{3.20}$$

where $B_\lambda^{S,C} = KP_\lambda^{S,C}(T^{-2}) + o(T^{-2})$. Note that from these and (3.17) filtered second-order approximations for the bias in the intercept readily follow.

From the second-order terms of all the approximations for B_λ^C given above (and even for the first-order terms of $B_\beta^{M,C}$), it seems obvious that the factor $(1-\lambda)$ will again lead to poor approximations when λ is close to unity. Note that randomness of the initial value and an expectation of the initial value that deviates from the mean-stationarity value both produce a positive contribution (if $\lambda > -0.33$) to the order T^{-2} term of the bias. Hence, the bias, which is generally negative for positive λ, is expected to be smaller when $\bar{y}_0 \neq \beta/(1-\lambda)$ and also when the start-up value is random (as we already found in the model with no intercept). That among mean-stationary processes those with a fixed start-up have a slightly more serious bias than those with a covariance-stationary random start-up is also suggested by the difference in distribution function established in Evans and Savin (1984, p.1265). The order T^{-1} bias which follows from Theorem 3 re-establishes the classic result by Marriott and Pope and by Kendall for the strongly stationary case, but it shows that this approximation is much more general and is valid irrespective of the values of ω, $\bar{y}_0$, β and σ; only the second-order bias is found to be affected by ω and by $[\bar{y}_0 - \beta/(1-\lambda)]/\sigma$.

In Tables 3 and 4 we present numerical results for the mean-stationary AR(1) model with unknown intercept for the fixed start-up case and the strongly stationary random start-up case respectively, and choose $\beta = 0$. Above we established that in those cases the estimator of the intercept is unbiased to order T^{-2} and indeed, the actual bias of the intercept found (but not tabulated) was extremely close to zero over the whole range of λ and T values (note that this would not necessarily have been the case for $\bar{y}_0 \neq \beta/(1-\lambda)$ or $\beta \neq 0$). We notice that the bias in $\hat{\lambda}$ is substantially bigger than in the model with no intercept, and ranges from about 10% (at $T = 50$) to 40% and above (at $T = 10$) of the actual value of λ. As predicted by the approximations, the bias is slightly more serious in the fixed start-up model. From Table 3, which contains results on the fixed start-up model, we find that the filtered order T^{-1} approximation, i.e. the classic $-(1+3\lambda)/T$ result of (3.10), understates the actual bias, especially when the bias is really serious (this is contrary to our finding in the known intercept model, where White's first-order approximation produces overstatements). Again, the unfiltered T^{-1} approximation gives smaller values than its filtered counterpart, and hence it performs rather poorly here, especially for large λ values. The two alternative order T^{-2} approximations usually produce an overstatement of the actual bias though they are quite accurate when λ is rather small and T is not too small. However, as expected, the filtered approximation is very bad for λ close to unity, whereas the unfiltered version behaves appropriately over the whole range, although for the very small sample size the simple order T^{-1} approximation is to be preferred for large λ values.

From Table 4 we see that the classic first-order result is exceptionally good in the strongly stationary random start-up model. Especially for λ close to unity, it is much better than the unfiltered order T^{-1} approximation and even better than the filtered second-order approximation. From the two second order approximations the unfiltered one has to be preferred. Filtering of the second-order approximation again ruins the accuracy close to the unit root. Note, however, that for λ close to one the strongly stationary AR(1) model can be expected to show odd behaviour (and does not approach the standard unit root with drift process) since $\lambda \to 1$ involves here an infinitely large value for the variance of all elements of the process.

4. Bias approximations for ARX(1) models

In this section we examine the actual bias and the quality of our approximations in models of more practical interest. We consider two types of stylized ARX(1) models and start with the trend-stationary model

$$y_t = \lambda y_{t-1} + \beta_1 + \beta_2 t + \sigma\varepsilon_t, \tag{4.1}$$

where $\varepsilon_t \sim \text{i.i.d.}N(0,1)$, $t = 1, ..., T$. For this model one can derive

$$\begin{array}{l} \bar{y}_t = \beta_1^* + \beta_2^* t, \\ \text{with } \beta_1^* = \bar{y}_0 = \frac{\beta_1 - \lambda\beta_2^*}{1-\lambda} \text{ and } \beta_2^* = \frac{\beta_2}{1-\lambda}. \end{array} \tag{4.2}$$

We shall examine this model for a range of λ values in two different settings, viz. for (A) $\beta_1^* = 0$, $\beta_2^* = 0$, $y_0 = 0$, $\sigma = 1$ and for (B) $\beta_1^* = 4.64$, $\beta_2^* = 0.04$, $y_0 = 4.76$, $\sigma = 0.05$. In the first setting the data generating process conforms to that of Tables 1 and 3; only the estimation equation differs. The second setting mimics some of the empirical findings for US real GNP for the annual time-series from 1908 on. Model (4.1) is certainly not a perfect specification for this data, but some of its characteristics are nevertheless reasonably well captured, especially when the parameter values are $\lambda = 0.9$, $\beta_1 = 0.5$, $\beta_2 = 0.004$, $\sigma = 0.05$ with initial observation $y_0 = 4.76$. So, by varying

$$\lambda,\ \beta_2 = (1-\lambda)\beta_2^*,\ \beta_1 = \beta_1^* + \lambda(\beta_2^* - \beta_1^*), \tag{4.3}$$

we can now examine the bias of least-squares estimators in a family of first-order trend-stationary models with common deterministic trend pattern (4.2). Note that due to the presence of this linear trend in model (4.1), it does not satisfy Assumption 1; nevertheless we can exploit Theorem 1 under the interpretation given by Corollary 2. The present framework, where $|\lambda| < 1$, does not permit a similar analysis of difference-stationarity for first-order integrated processes; the bias of least-squares estimators in such models has been examined in KP (2001).

In setting (A) we obviously have $\bar{y}_{-1} = 0$, so $\bar{Z} = (0, X)$; hence, the matrix Q is block-diagonal and therefore $\bar{Z}q_1 = 0 = e_i' q_1$ for $i = 2, 3$. From this it follows that $B_\beta^{CT} = o(T^{-2})$ and so for setting (A) we will only examine B_λ^{CT}. Note that we use the superscript label CT to indicate that the present first-order autoregressive model includes both a constant and a linear trend. Table 5 presents the results on B_λ^{CT} for both settings (A) and (B). With respect to $\hat{\lambda}$ these two settings give almost similar results. We find that for high values of λ and very small values of T the bias is extremely high. For $\lambda < 0.9$ the second-order approximation is strikingly accurate, and much better than the first-order approximation. For values of λ closer to one, the second-order approximation understates the actual bias, but much less so than the first-order formula, even at sample size $T = 50$. For this type of model, which is often used to analyse the alternatives of trend-stationarity and difference-stationarity, high values of λ are usually very relevant. We see that standard asymptotic inference may then be misleading and some form of bias correction seems most appropriate. However, from the simulation results it is obvious, that a bias correction to merely first-order will yield estimates that are still defective.

Table 6 shows the results for the estimators of the other coefficients in model (4.1) under setting (B). There is a severe positive bias in the estimate of the intercept, especially when λ is large, and again the second-order approximation is substantially better. The same holds for the estimate of the linear trend coefficient. Note that the first- and second-order approximations for $\hat{\beta}_2$ are actually accurate to a higher-order here, due to Corollary 2, and so is $\hat{\lambda}$.

Next we turn to a different type of model and make use of a data set analyzed and published in Davidson and MacKinnon (1985). They use the ARX(1) specification to regress the natural logarithm of housing starts (hs) on its first lag and on the first lag of the natural logarithm of gross national expenditure in 1971\$ (y), the first lag of real interest rates (RR) and a constant. These are Canadian quarterly data and the length of the series is 113. Although we know that the ARX(1) model probably involves a misspecification for these data, see Kiviet and Dufour (1995), we can perfectly well use them and the ARX(1) specification here merely for illustrative purposes. This is because we shall again use the empirical data only to obtain a realistic matrix X and start-up value y_0, and to make a relevant choice for the coefficient values. In line with (4.2), we use now stylized values for the long-run multipliers of the two exogenous regressors and for the standard deviation of the disturbances. Over the period 1956(1) to 1982(4) ordinary least-squares yields:

$$\begin{aligned} hs &= \underset{(0.070)}{0.610hs_{-1}} + \underset{(0.538)}{2.48} + \underset{(0.055)}{0.183y_{-1}} - \underset{(0.009)}{0.041RR_{-1}} \end{aligned} \tag{4.4}$$

Hence, the long-run multipliers with respect to y and RR are 0.55 and -0.08 respectively, and in the longrun relationship the intercept is 5.34. The estimate of σ in (4.4) amounts to 0.142.

These results prompt us to examine the true and the approximated bias of the least-squares coefficient estimator in the data generating process:

$$\begin{aligned} &hs_t = \lambda hs_{t-1} + 5(1-\lambda) + 0.5(1-\lambda)y_{t-1} - 0.1(1-\lambda)RR_{t-1} + 0.15\varepsilon_t \\ &\text{with } \varepsilon_t \sim \text{i.i.d.}N(0,1), \qquad t = 1, ..., T. \end{aligned} \tag{4.5}$$

We do this for a range of values for λ and T. Tables 7 and 8 present the results, which again indicate both the need for and the accuracy of asymptotic expansion methods to assess the finite sample bias in empirical ARX models when the sample size is moderate or small.

For this data generating process again very substantial relative magnitudes of the bias are found in a very small sample of $T = 10$. Even when $\lambda < 0.8$, bias values are found that are several times larger than the actual coefficient value. Also for larger samples and λ close to the unit circle, the least-squares estimate of λ shows a very substantial bias. We have to conclude that the standard type of analysis in such a model (where σ is such that R^2 is about 0.85) is almost useless. In the $T = 50$ case the relative bias is found to be moderate for $\lambda < 0.7$. Only when at least 100 observations are used is reliance on consistency vindicated for the least-squares estimator provided that λ is not too close to the unit circle.

From the Tables 7 and 8 we also see that even in the $T = 10$ case alternative but least-squares based methods can be developed and employed to get improved estimators of the unknown parameter values. For $\lambda < 0.9$, the $B_\alpha(T^{-2})$ approximation works extremely well and provides a substantial improvement over the first-order approximation. It seems worthwhile to develop and examine bias correction procedures in which the approximation formula is evaluated on the basis of the original least-squares estimator which, eventually in an iterative procedure, is then used in an attempt to remove the bias while maintaining or even improving the efficiency (MSE) of the resulting estimator. When T is large (≥ 100) and λ substantial (around 0.8) then the relative bias may still be rather large (in the present model for the β_2 estimate and for the intercept this bias is above 10%), while the T^{-2} approximation is extremely accurate (except close to the unit circle) and much better than the T^{-1} approximation. So, even in ARX(1) models of moderately large sample size, the $B_\alpha(T^{-2})$ based approximation procedures can be used at least to signal possible bias problems. When minor, such information may supplement any other (test) evidence on the adequacy of both the model specification and the inference procedures used. When substantial, then finite sample bias problems are diagnosed, and the application of some bias correction procedure seems advisable.

5. Conclusions

After considerable analytic efforts we have obtained an expression that includes the $O(T^{-1})$ and the $O(T^{-2})$ terms of the bias in the full vector of least-squares estimators of the coefficients of a stable ARX(1) model with independent and identically distributed normal disturbances and any given (normal) distribution of the initial start-up value. This estimator coincides with the (under standard regularity assumptions consistent but biased) maximum likelihood estimator conditional on the initial value of the dependent variable. In the ARX(1) class of regression models the regressor matrix includes an arbitrary number of strongly exogenous explanatory variables in addition to one weakly exogenous regressor: the one period lagged dependent variable. For very special cases, viz. those with no exogenous regressors or with just a constant term, approximations for the bias (mainly of first-order) were obtained some decades ago. Most of these earlier results are explicit in powers of T^{-1}, whereas our approximations, because of their generality, are in more compound terms and the dependence of the order of magnitude of the various terms on powers of T is implicit. As a consequence our $O(T^{-1})$ approximation also includes elements of terms that are actually $o(T^{-1})$, and our $O(T^{-2})$ approximation, in its most general form given here in Theorem 1, contains elements of terms that are actually $o(T^{-2})$. These $o(T^{-2})$ elements are, of course, not necessarily negligible when evaluated for finite T. Therefore, evaluation of our approximations in the just mentioned very special models yields results that are different from those provided by the earlier approximations. Only in special cases can all smaller order contributions be removed. By doing so, we have been able to re-establish and extend some classic results. Notably in Theorem 3, we produced an explicit second-order bias approximation for any variant of the normal AR(1) model with unknown intercept.

In a number of numerical experiments we find that the finite sample bias of least-squares estimators in ARX(1) models may be very substantial, especially when either or both the sample size is small and the dynamic adjustment process captured by the model is slow. In general, the second-order bias approximation is found to be very accurate in ARX(1) models, and it is also found to yield an, often very substantial, improvement over the first-order approximation. In the more specific types of models which contain no arbitrary regressor vectors, but only an intercept with either a known or unknown intercept value, the picture is different. We came across some pathological cases, where higher-order approximations are found to be more vulnerable in the neighbourhood of the non-stationarity region of the parameter space (λ close to unity) than first-order approximations are. This is especially true of the filtered approximations, from which any smaller order contributions have been removed, but which then have terms involving $(1-\lambda)$ factors in the denominator, giving unstable results for λ close to one. For some of these models expansions in powers of λ are available, and these are then to be preferred to expansions in powers of T^{-1}.

References

Bao, Y., 2007. The approximate moments of the least squares estimator for the stationary autoregressive model under a general error distribution. *Econometric Theory* 23, 1013-1021.

Bao, Y., Ullah, A., 2007. The second-order bias and mean squared error of estimators in time series models. *Journal of Econometrics* 140, 650-669.

Bartlett, M.S., 1946. On the theoretical specification and sampling properties of autocorrelated time-series. *Journal of the Royal Statistical Society* 8, 27-41.

Copas, J.B., 1966. Monte Carlo results for estimation in a stable Markov time series. *Journal of the Royal Statistical Society*, Series A 129, 110-116.

Cox, D.R., Snell, E.J., 1968. A general definition of residuals. *Journal of the Royal Statistical Society*, Series B 30, 248-275.

Davidson, R., MacKinnon, J.G., 1985. Testing linear and loglinear regressions against Box-Cox alternatives. *Canadian Journal of Economics* 18, 499-517.

Evans, G.B.A., Savin, N.E., 1981. Testing for unit roots: 1. *Econometrica* 149, 753-779.

Evans, G.B.A., Savin, N.E., 1984. Testing for unit roots: 2. *Econometrica* 152, 1241-1269.

Firth, D., 1993. Bias reduction of maximum likelihood estimates. *Biometrika* 180, 27-38.

Grubb, D., Symons, J., 1987. Bias in regressions with a lagged dependent variable. *Econometric Theory* 3, 371-386.

Hadri, K., Phillips, G.D.A., 1999. The accuracy of the higher-order bias approximation for the 2SLS estimator. *Economics Letters* 62, 167-174.

Hurwicz, L., 1950. Least squares bias in time series. In: Koopmans, T. (ed.) *Statistical inference in dynamic economic models.* Wiley, New York, 365-383.

Iglesias, E.M., Phillips, G.D.A., 2008. Finite sample theory of QMLE in ARCH models with dynamics in the mean equation. *Journal of Time Series Analysis* 29, 719-737.

Kadane, J.B., 1971. Comparison of k-class estimators when the disturbances are small. *Econometrica* 139, 723-737.

Kass, R.E., 1988. Discussion on saddlepoint methods and statistical inference (by N. Reid). *Statistical Science* 13, 213-238.

Kendall, M.G., 1954. Note on the bias in the estimation of autocorrelation. *Biometrika* 61, 403-404.

Kiviet, J.F., Phillips, G.D.A., 1993. Alternative bias approximations in regressions with a lagged dependent variable. *Econometric Theory* 9, 62-80.

Kiviet, J.F., Phillips, G.D.A., 1994. Bias assessment and reduction in linear error-correction models. *Journal of Econometrics* 63, 215-243.

Kiviet, J.F., Phillips, G.D.A., 1998. Degrees of freedom adjustment for disturbance variance estimators in dynamic regression models. *The Econometrics Journal* 1, 44-70.

Kiviet, J.F., Phillips, G.D.A., 2001. Moment approximation for least squares estimators in dynamic regression models with a unit root. *The Econometrics Journal* 8, 115-142.

Kiviet, J.F., Phillips, G.D.A., 2009. Improved variance estimation of coefficients in stable first-order dynamic regression models. mimeo.

Kiviet, J.F., Phillips, G.D.A., Schipp, B., 1995. The bias of OLS, GLS and ZEF estimators in dynamic seemingly unrelated regression models. *Journal of Econometrics* 69, 241-266.

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

Magnus, J.R., 1978. The moments of products of quadratic forms in normal variables. *Statistica Neerlandica* 132, 201-210.

Maekawa, K., 1983. An approximation to the distribution of the least squares estimator in an autoregressive model with exogenous variables. *Econometrica* 151, 229-238.

Marriott, F.H.C., Pope, J.A., 1954. Bias in the estimation of autocorrelations. *Biometrika* 61, 393-403.

Mikhail, W.M., 1972. The bias of the two-stage least squares estimator. *Journal of the American Statistical Association* 167, 625-627.

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.

Nankervis, J.C., Savin, N.E., 1988. The exact moments of the least squares estimator for the autoregressive model: corrections and extensions. *Journal of Econometrics* 137, 381-388.

Nicholls, D.F., Pope, A.L., 1988. Bias in the estimation of multivariate autoregressions. *Australian Journal of Statistics* 130A, 296-309.

Orcutt, G.H., Winokur, H.S., 1969. First order regression, estimation and prediction. *Econometrica* 37, 1-14.

Pierce, D.A., Peters, D., 1992. Practical use of higher order asymptotics for multiparameter exponential families. *Journal of the Royal Statistical Society*, Series B 54, 701-738.

Rudebusch, G.D., 1992. Trends and random walks in macroeconomic time series: A re-examination. *International Economic Review* 33, 661-680.

Rudebusch, G.D., 1993. The uncertain unit root in real GDP. *American Economic Review* 83, 264-272.

Sawa, T., 1978. The exact moments of the least squares estimator for the autoregressive model. *Journal of Econometrics* 18, 159-172.

Shaman, P., Stine, R.A., 1988. The bias of autoregressive coefficient estimators. *Journal of the American Statistical Association* 183, 842-848.

Shenton, L.R., Johnson, W.L., 1965. Moments of a serial correlation coefficient. *Journal of the Royal Statistical Society*, Series B 27, 308-318.

Tanaka, K., 1983. Asymptotic expansions associated with the AR(1) model with unknown mean. *Econometrica* 151, 1221-1231.

Tjøstheim, D., Paulsen, J., 1983. Bias of some commonly-used time series estimates. *Biometrika* 170, 389-399.

Tsui, A.K., Ali, M.M., 1992. Numerical computation of exact moments of the least squares estimator in a first order stationary autoregressive model. *Journal of Statistical Computation and Simulation* 142, 65-78.

Tsui, A.K., Ali, M.M., 1994. Exact distributions, density functions and moments of the least squares estimator in a first-order autoregressive model. *Computational Statistics and Data Analysis* 117, 433- 454.

Vinod, H.D., Shenton, L.R., 1996. Exact Moments for autoregressive and random walk models for a zero or stationary initial value. *Econometric Theory* 12, 481-499.

White, J.S., 1961. Asymptotic expansions for the mean and variance of the serial correlation coefficient. *Biometrika* 48, 85-94.

Appendices

A. Auxiliary results on expectations involving quadratic forms

In Appendix B we shall present a proof of Theorem 1. The analysis is extensive and involves numerous evaluations of expectations of products up to four quadratic forms in normal variables. We commence by stating in this appendix some essential basic results used repeatedly in the subsequent analysis.

Lemma 1: *Let A_1, A_2 and A_3 be a symmetric $T \times T$ matrices and B an arbitrary $T \times T$ matrix. Let the $T \times 1$ random vector ε be such that $\varepsilon \sim N(0, \sigma^2 I_T)$, then the following results hold:*

$$E(\varepsilon' A_1 \varepsilon)(\varepsilon' B \varepsilon) = \sigma^4 [\operatorname{tr}(A_1) \operatorname{tr}(B) + 2 \operatorname{tr}(A_1 B)]; \tag{A.1}$$

$$E[\varepsilon' A_1 \varepsilon - \sigma^2 \operatorname{tr}(A_1)](\varepsilon' B \varepsilon) = 2\sigma^4 \operatorname{tr}(A_1 B); \tag{A.2}$$

$$E(\varepsilon\varepsilon' B \varepsilon\varepsilon') = E(\varepsilon' B \varepsilon)\varepsilon\varepsilon' = \sigma^4 [\operatorname{tr}(B) I_T + B + B']; \tag{A.3}$$

$$E[\varepsilon' A_1 \varepsilon - \sigma^2 \operatorname{tr}(A_1)]\varepsilon\varepsilon' = 2\sigma^4 A_1; \tag{A.4}$$

$$\begin{aligned} E(\varepsilon' A_1 \varepsilon)(\varepsilon' A_2 \varepsilon)(\varepsilon' B \varepsilon) = \sigma^6 [&\operatorname{tr}(A_1) \operatorname{tr}(A_2) \operatorname{tr}(B) + 2 \operatorname{tr}(A_1) \operatorname{tr}(A_2 B) \\ &+2 \operatorname{tr}(A_2) \operatorname{tr}(A_1 B) + 2 \operatorname{tr}(B) \operatorname{tr}(A_1 A_2) + 4 \operatorname{tr}(A_1 A_2 B) + 4 \operatorname{tr}(A_2 A_1 B)]; \end{aligned} \tag{A.5}$$

$$\begin{aligned} E[\varepsilon' A_1 \varepsilon - \sigma^2 \operatorname{tr}(A_1)](\varepsilon' A_2 \varepsilon)(\varepsilon' B \varepsilon) = \sigma^6 [&2 \operatorname{tr}(A_2) \operatorname{tr}(A_1 B) + 2 \operatorname{tr}(B) \operatorname{tr}(A_1 A_2) \\ &+4 \operatorname{tr}(A_1 A_2 B) + 4 \operatorname{tr}(A_2 A_1 B)]; \end{aligned} \tag{A.6}$$

$$E[\varepsilon' A_1 \varepsilon - \sigma^2 \operatorname{tr}(A_1)]^2 \varepsilon' B_1 \varepsilon = \sigma^6 [2 \operatorname{tr}(B) \operatorname{tr}(A_1 A_1) + 8 \operatorname{tr}(A_1 A_1 B)]; \tag{A.7}$$

$$\begin{aligned} E(\varepsilon\varepsilon' A_1 \varepsilon\varepsilon' B \varepsilon\varepsilon') = E(\varepsilon' A_1 \varepsilon)(\varepsilon' B \varepsilon)\varepsilon\varepsilon' = \sigma^6 [&\operatorname{tr}(A_1) \operatorname{tr}(B) I_T + \operatorname{tr}(A_1)(B + B') \\ &+2 \operatorname{tr}(B) A_1 + 2 \operatorname{tr}(A_1 B) I_T + 2(A_1 B + B A_1 + A_1 B' + B' A_1)]; \end{aligned} \tag{A.8}$$

$$\begin{aligned} &E[\varepsilon' A_1 \varepsilon - \sigma^2 \operatorname{tr}(A_1)]\varepsilon\varepsilon' B \varepsilon\varepsilon' = E[\varepsilon' A_1 \varepsilon - \sigma^2 \operatorname{tr}(A_1)](\varepsilon' B \varepsilon)\varepsilon\varepsilon' = \\ &\quad \sigma^6 [2 \operatorname{tr}(B) A_1 + 2 \operatorname{tr}(A_1 B) I_T + 2(A_1 B + B A_1 + A_1 B' + B' A_1)]; \end{aligned} \tag{A.9}$$

$$E[\varepsilon' A_1 \varepsilon - \sigma^2 \operatorname{tr}(A_1)]^2 \varepsilon\varepsilon' = \sigma^6 [2 \operatorname{tr}(A_1 A_1) I_T + 8 A_1 A_1]; \tag{A.10}$$

$$\begin{aligned} E(\varepsilon' B \varepsilon) \textstyle\prod_{j=1}^{3} (\varepsilon' A_j \varepsilon) = \sigma^8 \{&\operatorname{tr}(A_1) \operatorname{tr}(A_2) \operatorname{tr}(A_3) \operatorname{tr}(B) \\ &+2[\operatorname{tr}(A_1) \operatorname{tr}(A_2) \operatorname{tr}(A_3 B) + \operatorname{tr}(A_1) \operatorname{tr}(A_3) \operatorname{tr}(A_2 B) + \operatorname{tr}(A_1) \operatorname{tr}(B) \operatorname{tr}(A_2 A_3) \\ &\quad + \operatorname{tr}(A_2) \operatorname{tr}(A_3) \operatorname{tr}(A_1 B) + \operatorname{tr}(A_2) \operatorname{tr}(B) \operatorname{tr}(A_1 A_3) + \operatorname{tr}(A_3) \operatorname{tr}(B) \operatorname{tr}(A_1 A_2)] \\ &+4[\operatorname{tr}(A_1) \operatorname{tr}(A_2 A_3 B) + \operatorname{tr}(A_1) \operatorname{tr}(A_3 A_2 B) + \operatorname{tr}(A_2) \operatorname{tr}(A_1 A_3 B) + \operatorname{tr}(A_2) \operatorname{tr}(A_3 A_1 B) \\ &\quad + \operatorname{tr}(A_3) \operatorname{tr}(A_1 A_2 B) + \operatorname{tr}(A_3) \operatorname{tr}(A_2 A_1 B) + 2 \operatorname{tr}(B) \operatorname{tr}(A_1 A_2 A_3)] \\ &+4[\operatorname{tr}(A_1 A_2) \operatorname{tr}(A_3 B) + \operatorname{tr}(A_1 A_3) \operatorname{tr}(A_2 B) + \operatorname{tr}(A_1 B) \operatorname{tr}(A_2 A_3)] \\ &+8[\operatorname{tr}(A_1 A_2 A_3 B) + \operatorname{tr}(A_3 A_1 A_2 B) + \operatorname{tr}(A_1 A_3 A_2 B) + \operatorname{tr}(A_3 A_2 A_1 B) \\ &\quad + \operatorname{tr}(A_2 A_1 A_3 B) + \operatorname{tr}(A_2 A_3 A_1 B)]\}; \end{aligned} \tag{A.11}$$

$$\begin{aligned} E[\varepsilon' A_1 \varepsilon - \sigma^2 \operatorname{tr}(A_1)]^3 \varepsilon' B \varepsilon \;=\; \sigma^8 [&8 \operatorname{tr}(A_1 A_1 A_1) \operatorname{tr}(B) + 12 \operatorname{tr}(A_1 A_1) \operatorname{tr}(A_1 B) \\ &+48 \operatorname{tr}(A_1 A_1 A_1 B)]. \end{aligned} \tag{A.12}$$

PROOF OF LEMMA 1: It is well-known that $E\varepsilon' B\varepsilon = \sigma^2 \operatorname{tr}(B)$. In Magnus (1978) the results (A.1), (A.5) and (A.11) are proved for a symmetric matrix B. Our slightly more general results easily follow upon using that $\varepsilon' B\varepsilon = \varepsilon'[\frac{1}{2}(B + B')]\varepsilon = \varepsilon' A_0\varepsilon$, where $A_0 = \frac{1}{2}(B + B')$ is symmetric. Now result (A.1) is easily proved since $\operatorname{tr}(A_0) = \operatorname{tr}(B)$ and $\operatorname{tr}(A_1A_0) = \operatorname{tr}(A_1B)$, due to the symmetry of A_1. Results (A.5) and (A.11) follow in a similar manner.

Result (A.3) is found as follows. Note that the (i, j)-component of $\varepsilon\varepsilon'$ is $\varepsilon_i\varepsilon_j$, which can be written as a symmetric quadratic form in the $T \times T$ matrix J, defined as:

$$\varepsilon_i\varepsilon_j = \varepsilon' e_i\varepsilon' e_j = \varepsilon' e_i e_j'\varepsilon = \varepsilon'[\frac{1}{2}e_i e_j' + \frac{1}{2}e_j e_i']\varepsilon = \varepsilon' J\varepsilon.$$

Hence, the (i, j)-component of $E(\varepsilon' B\varepsilon)\varepsilon\varepsilon'$ can be expressed as $E(\varepsilon' B\varepsilon)\varepsilon' J\varepsilon$, and applying result (A.1), we find that this is equal to $\sigma^4[\operatorname{tr}(J)\operatorname{tr}(B) + 2\operatorname{tr}(JB)]$, which simplifies to $\sigma^4[\operatorname{tr}(B)\delta_{ij} + e_i'(B + B')e_j]$, where $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ otherwise. Now (A.3) directly follows. Similarly, upon using (A.5), we find that the (i, j)-component of $E(\varepsilon' A_1\varepsilon)(\varepsilon' B\varepsilon)\varepsilon\varepsilon'$ can be expressed as $E(\varepsilon' A_1\varepsilon)(\varepsilon' B\varepsilon)\varepsilon' J\varepsilon = \sigma^6[\operatorname{tr}(A_1)\operatorname{tr}(B)\delta_{ij} + \operatorname{tr}(A_1)e_i'(B + B')e_j + 2\operatorname{tr}(A_1B)\delta_{ij} + 2\operatorname{tr}(B)e_i' A_1 e_j + 2e_i'(A_1B + BA_1 + A_1B' + B'A_1)e_j]$, which yields (A.8). Result (A.4) follows from (A.3) and the symmetry of A_1, (A.6) and (A.7) follow from (A.5) and (A.1), and so on.

B. Proof of Theorem 1

We proceed by applying a Nagar type expansion to the estimation error

$$\hat{\alpha} - \alpha = (Z'Z)^{-1}Z'u. \tag{B.1}$$

First we note that on putting, according to (2.7) and (2.8), $Z = \bar{Z} + \tilde{Z}$, we find for D, introduced in (2.13):

$$D = Z'Z = (\bar{Z} + \tilde{Z})'(\bar{Z} + \tilde{Z}) = \bar{Z}'\bar{Z} + \bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z} + \tilde{Z}'\tilde{Z}. \tag{B.2}$$

Now $E(Z'Z) = \bar{D} = \bar{Z}'\bar{Z} + E(\tilde{Z}'\tilde{Z})$, since $E(\bar{Z}'\tilde{Z}) = O$, and so

$$\begin{aligned} Z'Z &= \bar{D} + \bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z} + \tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z}) \\ &= \{I_{K+1} + (\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})(\bar{D})^{-1} + [\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})](\bar{D})^{-1}\}\bar{D}. \end{aligned} \tag{B.3}$$

Hence,

$$(Z'Z)^{-1} = (\bar{D})^{-1}\{I_{K+1} + (\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})(\bar{D})^{-1} + [\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})](\bar{D})^{-1}\}^{-1}, \tag{B.4}$$

where the stochastic terms $(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})(\bar{D})^{-1}$ and $[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})](\bar{D})^{-1}$ 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 $(I-A+A^2-A^3+....)$, whereby successive terms are of decreasing order in probability. The required expansion retains terms up to $O_p(T^{-3/2})$ and these terms, premultiplied by $(\bar{D})^{-1} = O(T^{-1})$, are then combined in (B.1), the estimation error, with those of

$$Z'u = \bar{Z}'u + \tilde{Z}'u. \tag{B.5}$$

Here both terms on the right-hand side are $O_p(T^{1/2})$. In this way a Nagar expansion is obtained which includes terms up to $O_p(T^{-2})$. The required bias approximation to the order of T^{-2} is then found by summing the expected values of all the retained terms.

Proceeding in this way, and upon using from now on the notation introduced in (2.15), we find for $(Z'Z)^{-1}$ the expression

$$
\begin{aligned}
Q\{I_{K+1} &- (\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q - [\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q \\
&+(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q + (\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q \\
&+[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q + [\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q \\
&-(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q \\
&-(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q \\
&-(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q \\
&-[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q \\
&-(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q \\
&-[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q \\
&-[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q \\
&-[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q\} + o_p(T^{-5/2}).
\end{aligned}
\tag{B.6}
$$

Since we have fifteen terms here, multiplication by the two terms of (B.5) will yield thirty terms, but fifteen of them involve products of an odd number of zero-mean normal random variables, and such products have a zero expected value. Ignoring those terms, we seek to evaluate the expectation below. The terms of interest establish $E(Z'Z)^{-1}Z'u$, and they are:

$$
\begin{aligned}
E\{Q\tilde{Z}'u &- Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q\bar{Z}'u - Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q\tilde{Z}'u \\
&+Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q\tilde{Z}'u + Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q\bar{Z}'u \\
&+Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q\bar{Z}'u + Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q\tilde{Z}'u \\
&-Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q\bar{Z}'u \\
&-Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q\tilde{Z}'u \\
&-Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q\tilde{Z}'u \\
&-Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q\tilde{Z}'u \\
&-Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q\bar{Z}'u \\
&-Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q\bar{Z}'u \\
&-Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q\bar{Z}'u \\
&-Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q\tilde{Z}'u\} + o(T^{-2}).
\end{aligned}
\tag{B.7}
$$

To find the bias approximation to order T^{-2} requires the evaluation of the fifteen separate terms of (B.7). We indicate these below as (i) through (xv), and shall evaluate each in turn using, where necessary, the results of Lemma 1 which is proved in Appendix A. To do this, we make substitutions that follow from Section 2, viz.:

$$u = (0, I_T)v,\ \tilde{Z} = Gve_1',\ G = \Lambda^{-1}(I_T, 0)\Omega,\ G(0, I_T)' = C,$$

which lead to

$$
\begin{aligned}
\tilde{Z}'u &= e_1v'G'(0, I_T)v = e_1v'Hv, \\
\tilde{Z}'\bar{Z} &= e_1v'G'\bar{Z}, \\
\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z}) &= [v'G'Gv - \sigma^2]\operatorname{tr}(G'G)e_1e_1'.
\end{aligned}
\tag{B.8}
$$

Here we have introduced the shorthand notation

$$
\begin{aligned}
H &= G'(0, I_T), \text{ with} \\
\operatorname{tr}(H) &= \operatorname{tr}[(0, I_T)'G] = \operatorname{tr}[G(0, I_T)'] = \operatorname{tr}(C) = 0,
\end{aligned}
\tag{B.9}
$$

because C has diagonal elements zero, see (2.10).

The first three terms of (B.7) yield the bias approximation to order T^{-1} which was evaluated by KP (1993, p.69) for the fixed start-up case ($\omega = 0$). We now obtain, using $v \sim N(0, \sigma^2 I_{T+1})$ and the results of Lemma 1:

$$EQ\tilde{Z}'u = EQe_1 v'Hv = \text{tr}(C)q_1 = 0; \tag{i}$$

$$\begin{aligned} EQ(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q\bar{Z}'u &= EQ(\bar{Z}'Gve_1' + e_1 v'G'\bar{Z})Q\bar{Z}'(0, I_T)v \\ &= EQ\bar{Z}'Gve_1'Q\bar{Z}'(0, I_T)v + EQe_1 v'G'\bar{Z}Q\bar{Z}'(0, I_T)v) \\ &= EQ\bar{Z}'Gvv'(0, I_T)'\bar{Z}q_1 + \sigma^2 \,\text{tr}[Q\bar{Z}'(0, I_T)G'\bar{Z}]q_1 \\ &= \sigma^2 Q\bar{Z}'G(0, I_T)'\bar{Z}q_1 + \sigma^2 \,\text{tr}[Q\bar{Z}'C'\bar{Z}]q_1 \\ &= \sigma^2 Q\bar{Z}'C\bar{Z}q_1 + \sigma^2 \,\text{tr}(Q\bar{Z}'C\bar{Z})q_1; \end{aligned} \tag{ii}$$

$$\begin{aligned} EQ[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q\tilde{Z}'u &= EQ[v'G'Gv - \sigma^2 \,\text{tr}(G'G)]e_1 e_1'Qe_1 v'G'(0, I_T)v \\ &= q_{11}E[v'G'Gv - \sigma^2 \,\text{tr}(G'G)]v'G'(0, I_T)vq_1 \\ &= 2\sigma^4 q_{11} \,\text{tr}[G'GG'(0, I_T)]q_1 \\ &= 2\sigma^4 q_{11} \,\text{tr}(GG'C)q_1. \end{aligned} \tag{iii}$$

In (i) we made use of (B.9). Note that (i) and (ii) yield in fact the same results as in the fixed start-up case, whereas the expectation for (iii) is different. It has been obtained by using (A.2).

From (i), (ii) and (iii) we find

$$B_\alpha(T^{-1}) = \sigma^2[Q\bar{Z}'C\bar{Z}q_1 + \text{tr}(Q\bar{Z}'C\bar{Z})q_1 - 2\sigma^2 q_{11} \,\text{tr}(GG'C)q_1]. \tag{B.10}$$

In order to derive the order T^{-2} bias the next term of interest is:

$$\begin{aligned} & EQ(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q\tilde{Z}'u \\ = \;& EQ(\bar{Z}'Gve_1' + e_1 v'G'\bar{Z})Q(\bar{Z}'Gve_1' + e_1 v'G'\bar{Z})Qe_1'Hv \\ = \;& EQ\bar{Z}'Gve_1'Q\bar{Z}'Gve_1'Qe_1'Hv + EQ\bar{Z}'Gve_1'Qe_1 v'G'\bar{Z}Qe_1'Hv \\ & + EQe_1 v'G'\bar{Z}Q\bar{Z}'Gve_1'Qe_1'Hv + EQe_1 v'G'\bar{Z}Qe_1 v'G'\bar{Z}Qe_1'Hv \\ = \;& q_{11}Q\bar{Z}'GE(v'Hv)vv'G'\bar{Z}q_1 + q_{11}Q\bar{Z}'GE(v'Hv)vv'G'\bar{Z}q_1 \\ & + q_{11}E(v'G'\bar{Z}Q\bar{Z}'Gv)(v'Hv)q_1 + E(v'G'\bar{Z}q_1 q_1'\bar{Z}'Gv)(v'Hv)q_1 \\ = \;& 2\sigma^4[q_{11}Q\bar{Z}'(GG'C' + CGG')\bar{Z}q_1 + q_{11} \,\text{tr}(Q\bar{Z}'GG'C'\bar{Z})q_1 + (q_1'\bar{Z}'GG'C'\bar{Z}q_1)q_1]; \end{aligned} \tag{iv}$$

This result is obtained by substitution of (B.8) and (B.9), followed by rearrangements that simply involve transposing scalar factors in such a way that the resulting expressions are of a format whose expectation has already been obtained in Lemma 1 of Appendix A; in (iv) we used (A.1) and (A.3).

Now the remaining terms will be listed and evaluated (neglecting for the moment their sign) in similar fashion[2].

$$EQ(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q\bar{Z}'u = 2\sigma^4[q_{11}Q\bar{Z}'GG'C\bar{Z}q_1 + (q_1'\bar{Z}'GG'C\bar{Z}q_1)q_1]; \tag{v}$$

$$EQ[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q\bar{Z}'u = 2\sigma^4[(q_1'\bar{Z}'GG'C\bar{Z}q_1)q_1 + q_{11} \,\text{tr}(Q\bar{Z}'GG'C\bar{Z})q_1]; \tag{vi}$$

[2] We omitted a detailed derivation for the remaining terms, which follow upon using the same principles. However, full proofs can be obtained from the authors on request.

$$EQ[\tilde{Z}'\tilde{Z}-E(\tilde{Z}'\tilde{Z})]Q[\tilde{Z}'\tilde{Z}-E(\tilde{Z}'\tilde{Z})]Q\tilde{Z}'u=8\sigma^6 q_{11}^2\,\mathrm{tr}(GG'GG'C)q_1; \qquad \text{(vii)}$$

$$\begin{aligned}
& EQ(\bar{Z}'\tilde{Z}+\tilde{Z}'\bar{Z})Q(\bar{Z}'\tilde{Z}+\tilde{Z}'\bar{Z})Q(\bar{Z}'\tilde{Z}+\tilde{Z}'\bar{Z})Q\bar{Z}'u \qquad \text{(viii)}\\
= \ & \sigma^4[2(q_1'\bar{Z}'GG'\bar{Z}q_1)Q\bar{Z}'C\bar{Z}q_1+2(q_1'\bar{Z}'C\bar{Z}q_1)Q\bar{Z}'GG'\bar{Z}q_1\\
& +2q_{11}\,\mathrm{tr}(Q\bar{Z}'C\bar{Z})Q\bar{Z}'GG'\bar{Z}q_1+4q_{11}Q\bar{Z}'GG'\bar{Z}Q\bar{Z}'C'\bar{Z}q_1\\
& +2q_{11}Q\bar{Z}'C\bar{Z}Q\bar{Z}'GG'\bar{Z}q_1+2q_{11}Q\bar{Z}'GG'\bar{Z}Q\bar{Z}'C\bar{Z}q_1\\
& +q_{11}\,\mathrm{tr}(Q\bar{Z}'GG'\bar{Z})Q\bar{Z}'C\bar{Z}q_1+(q_1'\bar{Z}'GG'\bar{Z}q_1)\,\mathrm{tr}(Q\bar{Z}'C\bar{Z})q_1\\
& +2(q_1'\bar{Z}'C\bar{Z}q_1)\,\mathrm{tr}(Q\bar{Z}'GG'\bar{Z})q_1+4(q_1'\bar{Z}'GG'\bar{Z}Q\bar{Z}'C'\bar{Z}q_1)q_1\\
& +q_{11}\,\mathrm{tr}(Q\bar{Z}'C\bar{Z})\,\mathrm{tr}(Q\bar{Z}'GG'\bar{Z})q_1+2q_{11}\,\mathrm{tr}(Q\bar{Z}'GG'\bar{Z}Q\bar{Z}'C'\bar{Z})q_1];
\end{aligned}$$

$$\begin{aligned}
& EQ(\bar{Z}'\tilde{Z}+\tilde{Z}'\bar{Z})Q(\bar{Z}'\tilde{Z}+\tilde{Z}'\bar{Z})Q[\tilde{Z}'\tilde{Z}-E(\tilde{Z}'\tilde{Z})]Q\tilde{Z}'u \qquad \text{(ix)}\\
= \ & \sigma^6\{4q_{11}^2\,\mathrm{tr}(GG'C)Q\bar{Z}'GG'\bar{Z}q_1+2q_{11}^2\,\mathrm{tr}(GG'C)\,\mathrm{tr}(Q\bar{Z}'GG'\bar{Z})q_1\\
& +4q_{11}^2Q\bar{Z}'(GG'GG'C'+GG'C'GG'+GG'CGG'+CGG'GG')\bar{Z}q_1\\
& +4q_{11}^2\,\mathrm{tr}(Q\bar{Z}'GG'CGG'\bar{Z})q_1+4q_{11}^2\,\mathrm{tr}(Q\bar{Z}'CGG'GG'\bar{Z})q_1\\
& +2q_{11}(q_1'\bar{Z}'GG'\bar{Z}q_1)\,\mathrm{tr}(GG'C)q_1\\
& +4q_{11}[q_1'\bar{Z}'(GG'CGG'+CGG'GG')\bar{Z}q_1]q_1\};
\end{aligned}$$

$$\begin{aligned}
& EQ(\bar{Z}'\tilde{Z}+\tilde{Z}'\bar{Z})Q[\tilde{Z}'\tilde{Z}-E(\tilde{Z}'\tilde{Z})]Q(\bar{Z}'\tilde{Z}+\tilde{Z}'\bar{Z})Q\tilde{Z}'u \qquad \text{(x)}\\
= \ & \sigma^6\{4q_{11}^2\,\mathrm{tr}(GG'C)Q\bar{Z}'GG'\bar{Z}q_1+4q_{11}(q_1'\bar{Z}'GG'\bar{Z}q_1)\,\mathrm{tr}(GG'C)q_1\\
& +4q_{11}^2Q\bar{Z}'(GG'GG'C'+GG'C'GG'+GG'CGG'+CGG'GG')\bar{Z}q_1\\
& +8q_{11}[q_1'\bar{Z}'(GG'CGG'+CGG'GG')\bar{Z}q_1]q_1\};
\end{aligned}$$

$$\begin{aligned}
& EQ[\tilde{Z}'\tilde{Z}-E(\tilde{Z}'\tilde{Z})]Q(\bar{Z}'\tilde{Z}+\tilde{Z}'\bar{Z})Q(\bar{Z}'\tilde{Z}+\tilde{Z}'\bar{Z})Q\tilde{Z}'u \qquad \text{(xi)}\\
= \ & \sigma^6\{6q_{11}(q_1'\bar{Z}'GG'\bar{Z}q_1)\,\mathrm{tr}(GG'C)q_1\\
& +12q_{11}[q_1'\bar{Z}'(GG'CGG'+CGG'GG')\bar{Z}q_1]q_1\\
& +2q_{11}^2\,\mathrm{tr}(GG'C)\,\mathrm{tr}(Q\bar{Z}'GG'\bar{Z})q_1\\
& +4q_{11}^2\,\mathrm{tr}(Q\bar{Z}'CGG'GG'\bar{Z})q_1+4q_{11}^2\,\mathrm{tr}(Q\bar{Z}'GG'CGG'\bar{Z})q_1\};
\end{aligned}$$

$$\begin{aligned}
& EQ(\bar{Z}'\tilde{Z}+\tilde{Z}'\bar{Z})Q[\tilde{Z}'\tilde{Z}-E(\tilde{Z}'\tilde{Z})]Q[\tilde{Z}'\tilde{Z}-E(\tilde{Z}'\tilde{Z})]Q\bar{Z}'u \qquad \text{(xii)}\\
= \ & \sigma^6[2q_{11}^2\,\mathrm{tr}(G'GG'G)Q\bar{Z}'C\bar{Z}q_1+8q_{11}^2Q\bar{Z}'GG'GG'C\bar{Z}q_1\\
& +2q_{11}(q_1'\bar{Z}'C\bar{Z}q_1)\,\mathrm{tr}(G'GG'G)q_1+8q_{11}(q_1'\bar{Z}'GG'GG'C\bar{Z}q_1)q_1];
\end{aligned}$$

$$\begin{aligned}
& EQ[\tilde{Z}'\tilde{Z}-E(\tilde{Z}'\tilde{Z})]Q(\bar{Z}'\tilde{Z}+\tilde{Z}'\bar{Z})Q[\tilde{Z}'\tilde{Z}-E(\tilde{Z}'\tilde{Z})]Q\bar{Z}'u \qquad \text{(xiii)}\\
= \ & \sigma^6q_{11}[4(q_1'\bar{Z}'C\bar{Z}q_1)\,\mathrm{tr}(G'GG'G)+16(q_1'\bar{Z}'GG'GG'C\bar{Z}q_1)]q_1;
\end{aligned}$$

$$\begin{aligned}
& EQ[\tilde{Z}'\tilde{Z}-E(\tilde{Z}'\tilde{Z})]Q[\tilde{Z}'\tilde{Z}-E(\tilde{Z}'\tilde{Z})]Q(\bar{Z}'\tilde{Z}+\tilde{Z}'\bar{Z})Q\bar{Z}'u \qquad \text{(xiv)}\\
= \ & \sigma^6[2q_{11}(q_1'\bar{Z}'C\bar{Z}q_1)\,\mathrm{tr}(G'GG'G)+8q_{11}(q_1'\bar{Z}'GG'GG'C\bar{Z}q_1)\\
& +2q_{11}^2\,\mathrm{tr}(G'GG'G)\,\mathrm{tr}(Q\bar{Z}'C\bar{Z})+8q_{11}^2\,\mathrm{tr}(Q\bar{Z}'GG'GG'C\bar{Z})]q_1;
\end{aligned}$$

$$\begin{aligned}
& EQ[\tilde{Z}'\tilde{Z}-E(\tilde{Z}'\tilde{Z})]Q[\tilde{Z}'\tilde{Z}-E(\tilde{Z}'\tilde{Z})]Q[\tilde{Z}'\tilde{Z}-E(\tilde{Z}'\tilde{Z})]Q\tilde{Z}'u \qquad \text{(xv)}\\
= \ & \sigma^8q_{11}^3[12\,\mathrm{tr}(GG'C)\,\mathrm{tr}(G'GG'G)+48\,\mathrm{tr}(GG'GG'GG'C)]q_1;
\end{aligned}$$

Now gathering the terms from (i) to (xv), taking their sign into account, starting with those explicitly involving σ^2 (note that Q also involves σ^2 implicitly), and next those in σ^4, σ^6 and σ^8 respectively, and upon removing the terms that are $o(T^{-2})$, yields Theorem 1. Contributions that are $O(T^{-3})$, and hence do not belong to an $O(T^{-2})$ approximation, stem for example from the second term in square brackets of (xv), which is $O(T)$, and hence, because q_{11} and q_1 are $O(T^{-1})$, can be omitted; the second term in square brackets of (xiv) is another example. That the aggregate of all individual $o(T^{-2})$ contributions which have been left out of the $O(T^{-2})$ approximation is still $o(T^{-2})$ follows from the general proof given in KP (2009, Appendix A).

C. Auxiliary results on the order of frequently occurring expressions

Here we state a number of separate results, which are proved by summing numerous - mainly geometric - series and then omitting terms of relatively small order. In Appendix D these results are used to prove Theorems 2 and 3.

$$\operatorname{tr}(C'C) = T\left(\frac{1}{1-\lambda^2}\right) - \left(\frac{1}{1-\lambda^2}\right)^2 + o(1); \tag{C.1}$$

$$\operatorname{tr}(CC'C) = T\lambda\left(\frac{1}{1-\lambda^2}\right)^2 - 2\lambda\left(\frac{1}{1-\lambda^2}\right)^3 + o(1); \tag{C.2}$$

$$\operatorname{tr}(C'CC'C) = T\left(\lambda^2+1\right)\left(\frac{1}{1-\lambda^2}\right)^3 + o(T); \tag{C.3}$$

$$\operatorname{tr}(C'CC'CC) = T\lambda\left(\lambda^2+2\right)\left(\frac{1}{1-\lambda^2}\right)^4 + o(T). \tag{C.4}$$

From () we have $G = (\omega F, C)$, with $F' = \left(1, \lambda, \lambda^2, ..., \lambda^{T-1}\right)$, for which we have

$$F'F = \left(\frac{1}{1-\lambda^2}\right) + o(1); \tag{C.5}$$

$$\begin{aligned}\operatorname{tr}(GG') &= \omega^2 F'F + \operatorname{tr}(C'C) \\ &= T\left(\frac{1}{1-\lambda^2}\right) + \omega^2\left(\frac{1}{1-\lambda^2}\right) - \left(\frac{1}{1-\lambda^2}\right)^2 + o(1);\end{aligned} \tag{C.6}$$

$$F'CF = \lambda\left(\frac{1}{1-\lambda^2}\right)^2 + o(1); \tag{C.7}$$

$$\begin{aligned}\operatorname{tr}(GG'C) &= \omega^2 F'CF + \operatorname{tr}(CC'C) \\ &= T\left(\frac{1}{1-\lambda^2}\right)^2 + \omega^2\lambda\left(\frac{1}{1-\lambda^2}\right)^2 - 2\lambda\left(\frac{1}{1-\lambda^2}\right)^3 + o(1);\end{aligned} \tag{C.8}$$

$$F'CC'F = \lambda^2\left(\frac{1}{1-\lambda^2}\right)^3 + o(1); \tag{C.9}$$

$$\operatorname{tr}(GG'GG') = \omega^4(F'F)^2 + 2\omega^2 F'CC'F + \operatorname{tr}(C'CC'C) = \operatorname{tr}(C'CC'C) + o(T); \tag{C.10}$$

$$\begin{aligned}\operatorname{tr}(GG'GG'C) &= \omega^4(F'F)F'CF + \omega^2 F'CC'CF + \omega^2 F'CCC'F + \operatorname{tr}(C'CC'CC) \\ &= \operatorname{tr}(C'CC'CC) + o(T).\end{aligned} \tag{C.11}$$

For regressions with an intercept the results below are relevant, where ι is a $T \times 1$ vector with all elements equal to unity.

$$F'\iota = \frac{1}{1-\lambda} + o(1); \tag{C.12}$$

$$\iota' C \iota = T\left(\frac{1}{1-\lambda}\right) - \left(\frac{1}{1-\lambda}\right)^2 + o(1); \tag{C.13}$$

$$F'C'\iota = \left(\frac{1}{1-\lambda}\right)^2 + o(1); \tag{C.14}$$

$$F'C\iota = \frac{\lambda}{1+\lambda}\left(\frac{1}{1-\lambda}\right)^2 + o(1); \tag{C.15}$$

$$F'CC'\iota = \frac{\lambda}{1+\lambda}\left(\frac{1}{1-\lambda}\right)^3 + o(1); \tag{C.16}$$

$$\begin{aligned} \iota' GG' \iota &= \iota'\left(\omega^2 FF' + CC'\right)\iota = \omega^2 \left(F'\iota\right)^2 + \iota' CC' \iota = \iota' CC' \iota + o(T) \\ &= T\left(\frac{1}{1-\lambda}\right)^2 + o(1); \end{aligned} \tag{C.17}$$

$$\iota' GG' C \iota = \iota'\left(\omega^2 FF'C + CC'C\right)\iota = \iota' CC'C\iota + o(T) = T\left(\frac{1}{1-\lambda}\right)^3 + o(1); \tag{C.18}$$

$$\iota' GG' C' \iota = \iota' CC'C'\iota + o(T) = T\left(\frac{1}{1-\lambda}\right)^3 + o(1). \tag{C.19}$$

D. Proof of Theorem 2

We evaluate the formula of Corollary 1 for the case $K = 0$, i.e. $\bar{Z} = \bar{y}_0 F$, with $F' = \left(1, \lambda, \lambda^2, ..., \lambda^{T-1}\right)$. It follows from (), (C.5) and (C.6) that in this case

$$\begin{aligned} \bar{D} &= \bar{y}_0^2 FF' + \sigma^2 \operatorname{tr}(GG') = \left(\omega^2\sigma^2 + \bar{y}_0^2\right) FF' + \sigma^2 \operatorname{tr}(GG') \\ &= \sigma^2 \left[T\left(\frac{1}{1-\lambda^2}\right) + \left(\omega^2 + \bar{y}_0^2/\sigma^2\right)\left(\frac{1}{1-\lambda^2}\right) - \left(\frac{1}{1-\lambda^2}\right)^2\right] + o(1). \end{aligned} \tag{D.1}$$

Hence,

$$\begin{aligned} (\bar{D})^{-1} &= Q = q_1 = q_{11} \\ &= \sigma^{-2}\left(1-\lambda^2\right)T^{-1} - \sigma^{-2}\left[\left(1-\lambda^2\right)\left(\omega^2 + \bar{y}_0^2/\sigma^2\right) - 1\right]T^{-2} + o(T^{-2}). \end{aligned} \tag{D.2}$$

Now, upon using (D.2) and various results given in Appendix C, we evaluate the nineteen terms of the approximation formula of Corollary 1, as far as they contain $O(T^{-1})$ and $O(T^{-2})$ elements. We obtain

$$\sigma^2 q_{11} \operatorname{tr}(Q\bar{Z}'C\bar{Z}) = \sigma^2 q_{11}^2 \bar{y}_0^2 F'CF = \left(y_0/\sigma\right)^2 \lambda T^{-2} + o(T^{-2}); \tag{i}$$

$$\sigma^2 q_1' \bar{Z}'C\bar{Z} q_1 = \sigma^2 q_{11}^2 \bar{y}_0^2 F'CF = \left(y_0/\sigma\right)^2 \lambda T^{-2} + o(T^{-2}); \tag{ii}$$

$$\sigma^4 q_{11}^2 \operatorname{tr}(GG'C) = \lambda T^{-1} - \lambda\left[\omega^2 + 2\left(y_0/\sigma\right)^2\right]T^{-2} + o(T^{-2}); \tag{iii}$$

$$\sigma^6 q_{11}^3 \operatorname{tr}(GG'GG'C) = \lambda\frac{2+\lambda^2}{1-\lambda^2}T^{-2} + o(T^{-2}); \tag{xiv}$$

$$\sigma^8 q_{11}^4 \operatorname{tr}(GG'C)\operatorname{tr}(GG'GG') = \lambda\frac{1+\lambda^2}{1-\lambda^2}T^{-2} + o(T^{-2}). \tag{xix}$$

All other terms are found to be $o(T^{-2})$. Collecting the above terms (and taking into account their sign and integer coefficient) we obtain the results of Theorem 2.

E. Proof of Theorem 3

Now we evaluate all the $O(T^{-1})$ and $O(T^{-2})$ elements in the nineteen terms of the approximation formula presented in Corollary 1 for the case where $K = 1$ and $X = \iota$. Given the invariance results found above (3.15), we may restrict ourselves to the special case where $\beta = 0$ and $\sigma = 1$ when we take $y_0^* \sim N(\bar{y}_0^*, \omega^2)$ for the start-up value. From the bias approximation for this particular case, we can find the result for the untransformed general model by simply substituting $\bar{y}_0^* = [\bar{y}_0 - \beta/(1-\lambda)]/\sigma$.

Note that now $\bar{Z} = (\bar{y}_0^* F, \iota)$. For simplicities sake we first consider the special case where $\omega = 0$ (fixed start-up). Then the 2×2 matrix $\bar{D}$ is

$$\bar{D} = Q^{-1} = \begin{pmatrix} \bar{y}_0^{*2} F'F + \mathrm{tr}(C'C) & \bar{y}_0^* F'\iota \\ \bar{y}_0^* F'\iota & T \end{pmatrix}. \tag{E.1}$$

Hence, when using (C.1), (C.5) and (C.12), we obtain for the elements of Q

$$\begin{aligned} q_{11} &= \left[\bar{y}_0^{*2} F'F + \mathrm{tr}(C'C) - \bar{y}_0^{*2} (F'\iota)^2/T\right]^{-1} \\ &= \left(1-\lambda^2\right) T^{-1} + \left[1 - \bar{y}_0^{*2}\left(1-\lambda^2\right)\right] T^{-2} + o(T^{-2}); \end{aligned} \tag{E.2}$$

$$q_{12} = -\bar{y}_0^* (1+\lambda) T^{-2} + o(T^{-2}); \tag{E.3}$$

$$q_{22} = T^{-1} + o(T^{-2}). \tag{E.4}$$

We now evaluate the terms of the Corollary, and find

$$\begin{aligned} q_{11}\,\mathrm{tr}(Q\bar{Z}'C\bar{Z}) &= q_{11}^2 \bar{y}_0^{*2} F'CF + \bar{y}_0^* q_{11} q_{12} (F'C\iota + \iota'CF) + q_{11} q_{22} \iota'C\iota \\ &= (1+\lambda) T^{-1} - \left[\bar{y}_0^{*2} + \lambda/(1-\lambda)\right] T^{-2} + o(T^{-2}); \end{aligned} \tag{i}$$

$$q_1'\bar{Z}'C\bar{Z}q_1 = q_{11}^2 \bar{y}_0^2 F'CF + o(T^{-2}) = \lambda \bar{y}_0^{*2} T^{-2} + o(T^{-2}); \tag{ii}$$

$$q_{11}^2\,\mathrm{tr}(GG'C) = \lambda T^{-1} - 2\lambda \bar{y}_0^{*2} T^{-2} + o(T^{-2}); \tag{iii}$$

$$q_{11}^2\,\mathrm{tr}(Q\bar{Z}'CC'C\bar{Z}) = q_{11}^2 q_{22} \iota'CC'C\iota + o(T^{-2}) = (1+\lambda)^2/(1-\lambda) T^{-2} + o(T^{-2}); \tag{iv}$$

$$q_{11}^2\,\mathrm{tr}(Q\bar{Z}'CC'C'\bar{Z}) = (1+\lambda)^2/(1-\lambda) T^{-2} + o(T^{-2}); \tag{v}$$

$$q_{11}^2\,\mathrm{tr}(Q\bar{Z}'CC'\bar{Z}Q\bar{Z}'C\bar{Z}) = q_{11}^2 q_{22}^2 (\iota'CC'\iota)(\iota'C\iota) + o(T^{-2}) = (1+\lambda)^2/(1-\lambda) T^{-2} + o(T^{-2}); \tag{vi}$$

$$q_{11}^2\,\mathrm{tr}(Q\bar{Z}'C\bar{Z})\,\mathrm{tr}(Q\bar{Z}'CC'\bar{Z}) = (1+\lambda)^2/(1-\lambda) T^{-2} + o(T^{-2}); \tag{vii}$$

the terms (viii) through (xiii) are $o(T^{-2})$;

$$q_{11}^3\,\mathrm{tr}(CC'CC'C) = \lambda \frac{2+\lambda^2}{1-\lambda^2} T^{-2} + o(T^{-2}); \tag{xiv}$$

$$q_{11}^3\,\mathrm{tr}(CC'CC')\,\mathrm{tr}(Q\bar{Z}'C\bar{Z}) = \frac{1+\lambda^2}{1-\lambda^2} T^{-2} + o(T^{-2}); \tag{xv}$$

$$q_{11}^3\,\mathrm{tr}(CC'C)\,\mathrm{tr}(Q\bar{Z}'CC'\bar{Z}) = \lambda \frac{1+\lambda}{1-\lambda^2} T^{-2} + o(T^{-2}); \tag{xvi}$$

the terms (xvii) and (xviii) are $o(T^{-2})$;

$$q_{11}^4\,\mathrm{tr}(CC'C)\,\mathrm{tr}(CC'CC') = \lambda \frac{1+\lambda^2}{1-\lambda^2} T^{-2} + o(T^{-2}). \tag{xix}$$

Collecting the terms (whilst taking into account their sign and integer coefficient) and making the substitution for $\bar{y}_0^*$ we obtain

$$KP_\lambda^F(T^{-2}) = -(1+3\lambda)T^{-1} - \frac{1-3\lambda+9\lambda^2}{1-\lambda}T^{-2} + (1+3\lambda)\left(\frac{\bar{y}_0}{\sigma} - \frac{\beta}{\sigma(1-\lambda)}\right)^2 T^{-2}. \quad \text{(E.5)}$$

Hence, for the AR(1) model with an intercept and arbitrary fixed start-up $\bar{y}_0$, the bias B_λ^F of the least-squares estimator λ can be approximated by (E.5), where $B_\lambda^F = KP_\lambda^F(T^{-2}) + o(T^{-2})$.

From this it is easy to obtain the second-order bias of $\hat{\lambda}$ in the random start-up model where $y_0 \sim N(\bar{y}_0, \omega^2\sigma^2)$ is independent of $(u_1, ..., u_T)$. It is obvious that in that model the bias conditional on y_0 is

$$E_u(\hat{\lambda} \mid y_0) - \lambda = -(1+3\lambda)T^{-1} - \frac{1-3\lambda+9\lambda^2}{1-\lambda}T^{-2} + (1+3\lambda)\left(\frac{y_0}{\sigma} - \frac{\beta}{\sigma(1-\lambda)}\right)^2 T^{-2} + o(T^{-2}).$$

Hence, for the unconditional bias we find the approximation $KP_\lambda(T^{-2})$ of the theorem.

Table 1. Bias in the AR(1) model with no intercept and fixed zero start-up

λ	$W^*_\lambda(T^{-1})$ (-2λ/T)	$B^*_\lambda(T^{-1})$ (3.5)	$W^*_\lambda(T^{-2})$ (3.7)	$B^*_\lambda(T^{-2})$ (3.6)	$SJ^*_\lambda(T^{-3})$ (3.4)	$SJ^*_\lambda(T^{-6})$ (3.4)	$W^*_\lambda(\lambda^9)$ (3.3)	B^*_λ
$T = 10$								
.0	-.0000	-.0000	.0000	.0000	.0000	.0000	.0000	-.0002
.1	-.0200	-.0197	-.0160	-.0154	-.0162	-.0162	-.0162	-.0164
.2	-.0400	-.0395	-.0320	-.0307	-.0323	-.0323	-.0323	-.0325
.3	-.0600	-.0591	-.0480	-.0458	-.0482	-.0484	-.0483	-.0486
.4	-.0800	-.0785	-.0640	-.0608	-.0638	-.0647	-.0642	-.0645
.5	-.1000	-.0976	-.0800	-.0756	-.0787	-.0820	-.0799	-.0803
.6	-.1200	-.1159	-.0960	-.0906	-.0920	-.1060	-.0952	-.0956
.7	-.1400	-.1320	-.1120	-.1079	-.1015	-.1834	-.1098	-.1102
.8	-.1600	-.1426	-.1280	-.1336	-.0974	-.9939	-.1233	-.1233
.9	-.1800	-.1404	-.1440	-.1731	-.0016	-42.1584	-.1347	-.1333
.99	-.1980	-.1216	-.1584	-.2035	+14.8333	-5957894	-.1419	-.1374
$T = 25$								
.0	-.0000	-.0000	.0000	.0000	.0000	.0000	.0000	-.0002
.1	-.0080	-.0080	-.0074	-.0073	-.0074	-.0074	-.0074	-.0076
.2	-.0160	-.0160	-.0147	-.0146	-.0147	-.0147	-.0147	-.0150
.3	-.0240	-.0239	-.0221	-.0219	-.0221	-.0221	-.0221	-.0224
.4	-.0320	-.0319	-.0294	-.0292	-.0294	-.0294	-.0294	-.0297
.5	-.0400	-.0399	-.0368	-.0365	-.0367	-.0368	-.0368	-.0371
.6	-.0480	-.0478	-.0442	-.0436	-.0439	-.0441	-.0440	-.0443
.7	-.0560	-.0556	-.0515	-.0507	-.0508	-.0514	-.0512	-.0515
.8	-.0640	-.0630	-.0589	-.0575	-.0569	-.0608	-.0582	-.0584
.9	-.0720	-.0677	-.0662	-.0671	-.0571	-.2118	-.0650	-.0642
.99	-.0792	-.0548	-.0729	-.0897	+.8866	-23971	-.0705	-.0649
$T = 50$								
.0	-.0000	-.0000	.0000	.0000	.0000	.0000	.0000	.0002
.1	-.0040	-.0040	-.0038	-.0038	-.0038	-.0038	-.0038	-.0040
.2	-.0080	-.0080	-.0077	-.0077	-.0077	-.0077	-.0077	-.0078
.3	-.0120	-.0120	-.0115	-.0115	-.0115	-.0115	-.0115	-.0117
.4	-.0160	-.0160	-.0154	-.0153	-.0154	-.0154	-.0154	-.0155
.5	-.0200	-.0200	-.0192	-.0192	-.0192	-.0192	-.0192	-.0193
.6	-.0240	-.0240	-.0230	-.0230	-.0230	-.0230	-.0230	-.0231
.7	-.0280	-.0280	-.0269	-.0268	-.0268	-.0268	-.0268	-.0270
.8	-.0320	-.0319	-.0307	-.0305	-.0305	-.0306	-.0306	-.0307
.9	-.0360	-.0355	-.0346	-.0340	-.0334	-.0360	-.0343	-.0342
.99	-.0396	-.0298	-.0380	-.0458	+.0819	-363.895	-.0376	-.0347

* for all entries the superscript label is "FM,NC", because $\omega = 0$, $y_0 = 0$; $K = 0$

Table 2. Bias in the AR(1) model with no intercept and random strongly stationary start-up

λ	$W^*_\lambda(T^{-1})$ (-2λ/T)	$B^*_\lambda(T^{-1})$ (2.16)	$KP^*_\lambda(T^{-2})$ (3.8)	$B^*_\lambda(T^{-2})$ (Corollary 1)	$W^*_\lambda(\lambda^5)$ (3.9)	B^*_λ
$T = 10$						
.0	-.0000	-.0000	.0000	.0000	.0000	-.0004
.1	-.0200	-.0180	-.0140	-.0144	-.0150	-.0154
.2	-.0400	-.0358	-.0278	-.0287	-.0299	-.0303
.3	-.0600	-.0534	-.0414	-.0428	-.0446	-.0451
.4	-.0800	-.0705	-.0545	-.0566	-.0589	-.0594
.5	-.1000	-.0867	-.0667	-.0700	-.0725	-.0733
.6	-.1200	-.1013	-.0772	-.0830	-.0852	-.0861
.7	-.1400	-.1126	-.0845	-.0960	-.0963	-.0973
.8	-.1600	-.1161	-.0836	-.1124	-.1052	-.1049
.9	-.1800	-.0968	-.0493	-.1360	-.1111	-.1033
.99	-.1980	-.0168	-.8366	-.0461	-.1128	-.0555
$T = 25$						
.0	-.0000	-.0000	.0000	.0000	.0000	-.0002
.1	-.0080	-.0077	-.0070	-.0071	-.0071	-.0073
.2	-.0160	-.0153	-.0141	-.0141	-.0142	-.0144
.3	-.0240	-.0229	-.0210	-.0211	-.0212	-.0215
.4	-.0320	-.0305	-.0279	-.0280	-.0282	-.0285
.5	-.0400	-.0379	-.0347	-.0349	-.0350	-.0353
.6	-.0480	-.0450	-.0412	-.0415	-.0416	-.0420
.7	-.0560	-.0516	-.0471	-.0476	-.0479	-.0482
.8	-.0640	-.0569	-.0518	-.0530	-.0537	-.0534
.9	-.0720	-.0569	-.0511	-.0579	-.0588	-.0556
.99	-.0792	-.0163	-.0863	-.0395	-.0625	-.0353
$T = 50$						
.0	-.0000	-.0000	.0000	.0000	.0000	.0002
.1	-.0040	-.0039	-.0038	-.0038	-.0038	-.0039
.2	-.0080	-.0078	-.0075	-.0075	-.0075	-.0076
.3	-.0120	-.0117	-.0113	-.0113	-.0113	-.0114
.4	-.0160	-.0156	-.0150	-.0150	-.0150	-.0151
.5	-.0200	-.0195	-.0187	-.0187	-.0187	-.0188
.6	-.0240	-.0232	-.0223	-.0223	-.0223	-.0225
.7	-.0280	-.0269	-.0258	-.0258	-.0259	-.0260
.8	-.0320	-.0302	-.0289	-.0291	-.0293	-.0293
.9	-.0360	-.0322	-.0308	-.0313	-.0325	-.0315
.99	-.0396	-.0144	+.0018	-.0290	-.0352	-.0227

* for all entries the superscript label is "S,NC", because $\omega^2 = 1/(1-\lambda^2)$, $\bar{y}_0 = 0$; $K = 0$

Table 3. Bias in the mean-stationary AR(1) model with unknown intercept and fixed start-up

λ	$MP_{\lambda}^{S,C}(T^{-1})$ (3.12)	$B_{\lambda}^{*}(T^{-1})$ (2.16)	$KP_{\lambda}^{*}(T^{-2})$ (3.19)	$B_{\lambda}^{*}(T^{-2})$ (3.18)	B_{λ}^{*}
$T=10$					
.0	-.1000	-.1000	-.1100	-.1102	-.1111
.1	-.1300	-.1285	-.1388	-.1377	-.1386
.2	-.1600	-.1567	-.1695	-.1665	-.1670
.3	-.1900	-.1843	-.2030	-.1973	-.1964
.4	-.2200	-.2110	-.2407	-.2307	-.2272
.5	-.2500	-.2361	-.2850	-.2680	-.2598
.6	-.2800	-.2584	-.3410	-.3117	-.2944
.7	-.3100	-.2750	-.4203	-.3658	-.3312
.8	-.3400	-.2799	-.5580	-.4345	-.3689
.9	-.3700	-.2636	-.9290	-.5065	-.4015
.99	-.3970	-.2242	-7.2479	-.5291	-.4135
$T=25$					
.0	-.0400	-.0400	.0416	.0416	-.0419
.1	-.0520	-.0518	-.0534	-.0533	-.0536
.2	-.0640	-.0636	-.0655	-.0653	-.0656
.3	-.0760	-.0752	-.0781	-.0777	-.0779
.4	-.0880	-.0868	-.0913	-.0906	-.0906
.5	-.1000	-.0982	-.1056	-.1043	-.1040
.6	-.1120	-.1092	-.1218	-.1193	-.1185
.7	-.1240	-.1195	-.1417	-.1367	-.1347
.8	-.1360	-.1279	-.1709	-.1593	-.1541
.9	-.1480	-.1281	-.2374	-.1960	-.1785
.99	-.1588	-.0978	-1.2549	-.2286	-.1938
$T=50$					
.0	-.0200	-.0200	.0204	.0204	.0206
.1	-.0260	-.0260	-.0264	-.0263	-.0265
.2	-.0320	-.0319	-.0324	-.0324	-.0325
.3	-.0380	-.0378	-.0385	-.0385	-.0386
.4	-.0440	-.0437	-.0448	-.0447	-.0449
.5	-.0500	-.0496	-.0514	-.0512	-.0513
.6	-.0560	-.0554	-.0584	-.0581	-.0581
.7	-.0620	-.0610	-.0664	-.0657	-.0656
.8	-.0680	-.0662	-.0767	-.0750	-.0746
.9	-.0740	-.0695	-.0964	-.0896	-.0874
.99	-.0794	-.0529	-.3534	-.1182	-.1032

* for all entries the superscript label is "FM,C", because $\omega=0$, $\bar{y}_0=\beta_1/(1-\lambda)$; $K=1$

Table 4. Bias in the strongly stationary AR(1) model with unknown intercept

λ	$MP^*_\lambda(T^{-1})$ (3.12)	$B^*_\lambda(T^{-1})$ (2.16)	$KP^*_\lambda(T^{-2})$ (3.20)	$B^*_\lambda(T^{-2})$ (3.18)	B^*_λ
$T=10$					
.0	-.1000	-.0900	.1000	.0990	-.1003
.1	-.1300	-.1158	-.1236	-.1243	-.1257
.2	-.1600	-.1408	-.1487	-.1506	-.1518
.3	-.1900	-.1648	-.1755	-.1781	-.1786
.4	-.2200	-.1871	-.2050	-.2072	-.2063
.5	-.2500	-.2076	-.2383	-.2385	-.2352
.6	-.2800	-.2215	-.2785	-.2727	-.2654
.7	-.3100	-.2275	-.3321	-.3110	-.2975
.8	-.3400	-.2157	-.4191	-.3524	-.3322
.9	-.3700	-.1630	-.6395	-.3680	-.3702
.99	-.3970	-.0255	-4.2580	-.0952	-.4085
$T=25$					
.0	-.0400	-.0384	-.0400	-.0399	-.0401
.1	-.0520	-.0497	-.0510	-.0512	-.0514
.2	-.0640	-.0609	-.0622	-.0627	-.0629
.3	-.0760	-.0720	-.0737	-.0744	-.0746
.4	-.0880	-.0827	-.0856	-.0866	-.0866
.5	-.1000	-.0931	-.0981	-.0992	-.0991
.6	-.1120	-.1026	-.1118	-.1127	-.1122
.7	-.1240	-.1105	-.1275	-.1275	-.1264
.8	-.1360	-.1145	-.1487	-.1448	-.1426
.9	-.1480	-.1047	-.1911	-.1663	-.1628
.99	-.1588	-.0252	-.7766	-.0853	-.1884
$T=50$					
.0	-.0200	-.0196	-.0200	-.0200	-.0201
.1	-.0260	-.0254	-.0257	-.0258	-.0260
.2	-.0320	-.0312	-.0315	-.0317	-.0318
.3	-.0380	-.0370	-.0374	-.0376	-.0378
.4	-.0440	-.0427	-.0434	-.0437	-.0438
.5	-.0500	-.0483	-.0495	-.0499	-.0500
.6	-.0560	-.0537	-.0559	-.0564	-.0564
.7	-.0620	-.0586	-.0629	-.0634	-.0633
.8	-.0680	-.0626	-.0712	-.0713	-.0711
.9	-.0740	-.0626	-.0848	-.0819	-.0813
.99	-.0794	-.0227	-.2338	-.0670	-.0984

* for all entries the superscript label is "S,C", because $\omega^2 = 1/(1-\lambda^2)$, $\bar{y}_0 = \beta_1/(1-\lambda)$; $K=1$

Table 5. Bias of $\hat{\lambda}$ in the fixed start-up trend-stationary ARX(1) model (4.1)

	(A): $\beta_1^* = 0$, $\beta_2^* = 0$, $y_0 = 0$, $\sigma = 1$			(B): $\beta_1^* = 4.64$, $\beta_2^* = 0.04$, $y_0 = 4.76$, $\sigma = 0.05$		
λ	$B_\lambda^*(T^{-1})$	$B_\lambda^*(T^{-2})$	B_λ^*	$B_\lambda^*(T^{-1})$	$B_\lambda^*(T^{-2})$	B_λ^*
$T = 10$						
.0	-.1778	-.2112	-.2171	-.1232	-.1371	-.1386
.1	-.2103	-.2492	-.2552	-.1500	-.1678	-.1699
.2	-.2410	-.2891	-.2948	-.1772	-.2022	-.2049
.3	-.2693	-.3313	-.3359	-.2050	-.2412	-.2447
.4	-.2940	-.3759	-.3790	-.2331	-.2865	-.2909
.5	-.3136	-.4229	-.4242	-.2609	-.3397	-.3455
.6	-.3257	-.4720	-.4719	-.2866	-.4024	-.4100
.7	-.3266	-.5222	-.5229	-.3048	-.4749	-.4838
.8	-.3109	-.5711	-.5792	-.3040	-.5510	-.5631
.9	-.2717	-.6036	-.6455	-.2712	-.6012	-.6433
.99	-.2147	-.5794	-.7220	-.2147	-.5794	-.7220
$T = 25$						
.0	-.0767	-.0826	-.0832	-.0636	-.0672	-.0674
.1	-.0915	-.0988	-.0995	-.0764	-.0809	-.0812
.2	-.1061	-.1155	-.1163	-.0893	-.0954	-.0957
.3	-.1203	-.1330	-.1338	-.1021	-.1108	-.1112
.4	-.1339	-.1514	-.1523	-.1149	-.1275	-.1280
.5	-.1465	-.1712	-.1722	-.1275	-.1461	-.1469
.6	-.1575	-.1927	-.1941	-.1397	-.1677	-.1691
.7	-.1653	-.2168	-.2192	-.1504	-.1940	-.1967
.8	-.1662	-.2436	-.2491	-.1570	-.2269	-.2334
.9	-.1498	-.2702	-.2878	-.1478	-.2648	-.2829
.99	-.0981	-.2599	-.3441	-.0981	-.2599	-.3441
$T = 50$						
.0	-.0392	-.0407	-.0410	-.0353	-.0365	-.0367
.1	-.0469	-.0488	-.0491	-.0424	-.0438	-.0440
.2	-.0546	-.0571	-.0573	-.0494	-.0514	-.0516
.3	-.0621	-.0656	-.0658	-.0564	-.0591	-.0593
.4	-.0695	-.0744	-.0747	-.0634	-.0673	-.0675
.5	-.0767	-.0837	-.0840	-.0702	-.0760	-.0762
.6	-.0835	-.0938	-.0942	-.0769	-.0856	-.0860
.7	-.0894	-.1051	-.1059	-.0831	-.0968	-.0975
.8	-.0932	-.1187	-.1206	-.0881	-.1111	-.1130
.9	-.0899	-.1360	-.1421	-.0876	-.1316	-.1378
.99	-.0546	-.1383	-.1810	-.0546	-.1383	-.1810

* for all entries the superscript label is "F,CT", because $\omega = 0$; $K = 2$

Table 6. Bias of the estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ in the fixed start-up trend-stationary ARX(1) model (4.1) in setting (B)

λ	β_1	$B^*_{\beta_1}(T^{-1})$	$B^*_{\beta_1}(T^{-2})$	$B^*_{\beta_1}$	β_2	$B^*_{\beta_2}(T^{-1})$	$B^*_{\beta_2}(T^{-2})$	$B^*_{\beta_2}$
$T = 10$								
.0	4.64	.5724	.6373	.6445	.040	.0042	.0046	.0046
.1	4.18	.6972	.7811	.7911	.036	.0051	.0056	.0056
.2	3.72	.8249	.9420	.9550	.032	.0059	.0066	.0066
.3	3.26	.9553	1.1256	1.1420	.028	.0067	.0077	.0077
.4	2.80	1.0877	1.3390	1.3603	.024	.0074	.0089	.0089
.5	2.34	1.2197	1.5906	1.6188	.020	.0081	.0103	.0103
.6	1.88	1.3430	1.8882	1.9251	.016	.0086	.0119	.0119
.7	1.42	1.4322	2.2336	2.2772	.012	.0089	.0138	.0139
.8	.96	1.4323	2.5974	2.6564	.008	.0090	.0162	.0165
.9	.50	1.2803	2.8391	3.0388	.004	.0087	.0193	.0207
.99	.086	1.0137	2.7353	3.4087	.0004	.0083	.0225	.0281
$T = 25$								
.0	4.64	.2938	.3102	.3114	.040	.0025	.0026	.0026
.1	4.18	.3532	.3741	.3754	.036	.0030	.0031	.0031
.2	3.72	.4128	.4411	.4427	.032	.0035	.0037	.0037
.3	3.26	.4725	.5126	.5145	.028	.0039	.0043	.0043
.4	2.80	.5321	.5905	.5929	.024	.0044	.0049	.0049
.5	2.34	.5911	.6777	.6813	.020	.0049	.0055	.0056
.6	1.88	.6483	.7793	.7857	.016	.0053	.0063	.0063
.7	1.42	.6999	.9034	.9166	.012	.0056	.0072	.0073
.8	.96	.7329	1.0606	1.0917	.008	.0057	.0083	.0085
.9	.50	.6942	1.2445	1.3307	.004	.0053	.0095	.0101
.99	.086	.4629	1.2268	1.6244	.0004	.0038	.0101	.0134
$T = 50$								
.0	4.64	.1629	.1682	.1690	.040	.0014	.0014	.0015
.1	4.18	.1955	.2021	.2030	.036	.0017	.0017	.0017
.2	3.72	.2279	.2369	.2378	.032	.0020	.0020	.0020
.3	3.26	.2603	.2729	.2738	.028	.0022	.0023	.0023
.4	2.80	.2924	.3106	.3116	.024	.0025	.0027	.0027
.5	2.34	.3243	.3510	.3522	.020	.0028	.0030	.0030
.6	1.88	.3554	.3957	.3975	.016	.0030	.0034	.0034
.7	1.42	.3848	.4482	.4517	.012	.0033	.0038	.0038
.8	.96	.4090	.5160	.5248	.008	.0034	.0043	.0044
.9	.50	.4090	.6147	.6442	.004	.0034	.0050	.0053
.99	.086	.2575	.6526	.8542	.0004	.0021	.0054	.0071

* for all entries the superscript label is "F,CT", because $\omega = 0$; $K = 2$

Table 7. Bias of the estimators $\hat{\lambda}$ and $\hat{\beta}_1$ in the ARX(1) model with empirical X matrix generated according to (4.5) obtained from (B.10) and Theorem 1

	λ	$B_\lambda(T^{-1})$	$B_\lambda(T^{-2})$	B_λ	β_1	$B_{\beta_1}(T^{-1})$	$B_{\beta_1}(T^{-2})$	B_{β_1}
$T = 10$								
	.0	-.0481	-.0508	-.0506	5.00	1.6353	1.7640	1.7444
	.1	-.0612	-.0651	-.1650	4.50	2.2634	2.4588	2.4404
	.2	-.0773	-.0833	-.1831	4.00	3.1096	3.4217	3.4033
	.3	-.0978	-.1076	-.1072	3.50	4.2696	4.7911	4.7669
	.4	-.1248	-.1413	-.1404	3.00	5.8890	6.7939	6.7437
	.5	-.1606	-.1895	-.1873	2.50	8.1361	9.7545	9.6278
	.6	-.2039	-.2574	-.2523	2.00	10.9377	13.9306	13.6150
	.7	-.2420	-.3439	-.3348	1.50	13.2999	18.9668	18.2095
	.8	-.2560	-.4342	-.4345	1.00	13.3310	22.3798	21.8124
	.9	-.2424	-.5214	-.5887	.50	9.7240	20.6552	21.7824
	.99	-.2030	-.5535	-.7752	.05	2.9278	7.9424	10.5093
$T = 25$								
	.0	-.0442	-.0460	-.0460	5.00	.9590	1.0187	1.0224
	.1	-.0528	-.0554	-.0555	4.50	1.2298	1.3177	1.3235
	.2	-.0622	-.0661	-.0662	4.00	1.5602	1.6940	1.7040
	.3	-.0725	-.0786	-.0789	3.50	1.9711	2.1809	2.1995
	.4	-.0844	-.0938	-.0945	3.00	2.4942	2.8340	2.8707
	.5	-.0982	-.1134	-.1150	2.50	3.1725	3.7464	3.8231
	.6	-.1146	-.1402	-.1437	2.00	4.0588	5.0752	5.2429
	.7	-.1348	-.1792	-.1868	1.50	5.2195	7.0766	7.4404
	.8	-.1588	-.2360	-.2509	1.00	6.6076	9.9393	10.5882
	.9	-.1609	-.2970	-.3238	.50	6.4982	11.9107	12.6925
	.99	-.1156	-.3056	-.4017	.05	1.9237	5.0368	6.2212
$T = 50$								
	.0	-.0291	-.0298	-.0297	5.00	.2453	.2530	.2553
	.1	-.0338	-.0347	-.0347	4.50	.2942	.3048	.3072
	.2	-.0384	-.0398	-.0397	4.00	.3469	.3621	.3647
	.3	-.0430	-.0449	-.0449	3.50	.4047	.4269	.4299
	.4	-.0476	-.0503	-.0503	3.00	.4691	.5020	.5058
	.5	-.0521	-.0560	-.0560	2.50	.5416	.5918	.5975
	.6	-.0565	-.0623	-.0626	2.00	.6246	.7043	.7144
	.7	-.0610	-.0701	-.0710	1.50	.7276	.8611	.8834
	.8	-.0664	-.0821	-.0847	1.00	.8927	1.1455	1.2090
	.9	-.0728	-.1070	-.1177	.50	1.2831	1.9216	2.1767
	.99	-.0540	-.1377	-.1903	.05	.8587	2.1762	2.9227
$T = 100$								
	.0	-.0180	-.0183	-.0182	5.00	.0965	.0982	.0978
	.1	-.0206	-.0210	-.0209	4.50	.1100	.1122	.1117
	.2	-.0231	-.0236	-.0235	4.00	.1225	.1256	.1250
	.3	-.0255	-.0262	-.0261	3.50	.1342	.1382	.1377
	.4	-.0278	-.0288	-.0287	3.00	.1448	.1503	.1498
	.5	-.0301	-.0314	-.0313	2.50	.1546	.1622	.1618
	.6	-.0323	-.0342	-.0341	2.00	.1642	.1750	.1748
	.7	-.0345	-.0374	-.0374	1.50	.1758	.1924	.1927
	.8	-.0369	-.0417	-.0420	1.00	.1976	.2277	.2303
	.9	-.0391	-.0497	-.0515	.50	.2711	.3546	.3739
	.99	-.0314	-.0727	-.0939	.05	.3994	.9242	1.1989

Table 8. Bias of the estimators $\hat{\beta}_2$ and $\hat{\beta}_3$ in the ARX(1) model with empirical X matrix generated according to (4.5) obtained from (B.10) and Theorem 1

	λ	β_2	$B^*_{\beta_2}(T^{-1})$	$B^*_{\beta_2}(T^{-2})$	$B^*_{\beta_2}$	β_3	$B^*_{\beta_3}(T^{-1})$	$B^*_{\beta_3}(T^{-2})$	$B^*_{\beta_3}$
$T = 10$									
	.0	.50	-.1230	-.1338	-.1319	-.10	-.0090	-.0096	-.0094
	.1	.45	-.1760	-.1926	-.1909	-.09	-.0114	-.0123	-.0121
	.2	.40	-.2491	-.2758	-.2742	-.08	-.0143	-.0157	-.0154
	.3	.35	-.3509	-.3961	-.3939	-.07	-.0180	-.0199	-.0197
	.4	.30	-.4949	-.5738	-.5694	-.06	-.0226	-.0255	-.0251
	.5	.25	-.6965	-.8384	-.8273	-.05	-.0280	-.0327	-.0320
	.6	.20	-.9498	-1.2124	-1.1840	-.04	-.0324	-.0401	-.0392
	.7	.15	-1.1632	-1.6493	-1.5883	-.03	-.0302	-.0428	-.0418
	.8	.10	-1.1545	-1.9327	-1.8700	-.02	-.0177	-.0314	-.0331
	.9	.05	-.7849	-1.6575	-1.7026	-.01	-.0019	-.0058	-.0116
	.99	.01	-.0919	-.2461	-.2773	-.00	+.0010	+.0024	+.0014
$T = 25$									
	.0	.50	-.0555	-.0599	-.0603	-.10	-.0064	-.0067	-.0067
	.1	.45	-.0753	-.0818	-.0824	-.09	-.0075	-.0080	-.0079
	.2	.40	-.1006	-.1107	-.1116	-.08	-.0086	-.0093	-.0093
	.3	.35	-.1335	-.1495	-.1511	-.07	-.0097	-.0106	-.0106
	.4	.30	-.1769	-.2032	-.2064	-.06	-.0108	-.0120	-.0121
	.5	.25	-.2350	-.2801	-.2866	-.05	-.0115	-.0132	-.0133
	.6	.20	-.3128	-.3943	-.4086	-.04	-.0113	-.0136	-.0138
	.7	.15	-.4167	-.5686	-.5994	-.03	-.0088	-.0115	-.0118
	.8	.10	-.5420	-.8178	-.8714	-.02	-.0021	-.0037	-.0043
	.9	.05	-.5284	-.9659	-1.0202	-.01	+.0054	+.0090	+.0072
	.99	.01	-.0801	-.2065	-.2281	-.00	+.0010	+.0026	+.0024
$T = 50$									
	.0	.50	+.0045	+.0044	+.0041	-.10	-.0039	-.0040	-.0039
	.1	.45	+.0043	+.0041	+.0038	-.09	-.0045	-.0046	-.0046
	.2	.40	+.0036	+.0033	+.0030	-.08	-.0050	-.0053	-.0052
	.3	.35	+.0023	+.0019	+.0015	-.07	-.0056	-.0059	-.0058
	.4	.30	+.0002	-.0005	-.0009	-.06	-.0060	-.0064	-.0064
	.5	.25	-.0029	-.0041	-.0046	-.05	-.0063	-.0068	-.0068
	.6	.20	-.0072	-.0095	-.0103	-.04	-.0062	-.0069	-.0069
	.7	.15	-.0137	-.0183	-.0198	-.03	-.0054	-.0064	-.0065
	.8	.10	-.0262	-.0365	-.0405	-.02	-.0039	-.0051	-.0053
	.9	.05	-.0603	-.0924	-.1081	-.01	-.0036	-.0045	-.0048
	.99	.01	-.0326	-.0819	-.1044	-.00	-.0009	-.0021	-.0017
$T = 100$									
	.0	.50	.0083	.0084	.0084	-.10	-.0017	-.0018	-.0018
	.1	.45	.0095	.0097	.0097	-.09	-.0019	-.0020	-.0020
	.2	.40	.0108	.0110	.0110	-.08	-.0021	-.0021	-.0021
	.3	.35	.0120	.0123	.0122	-.07	-.0022	-.0023	-.0023
	.4	.30	.0132	.0136	.0136	-.06	-.0023	-.0023	-.0023
	.5	.25	.0144	.0149	.0149	-.05	-.0022	-.0023	-.0023
	.6	.20	.0155	.0163	.0163	-.04	-.0021	-.0022	-.0022
	.7	.15	.0165	.0176	.0176	-.03	-.0018	-.0020	-.0020
	.8	.10	.0165	.0182	.0182	-.02	-.0014	-.0016	-.0016
	.9	.05	.0112	.0132	.0129	-.01	-.0010	-.0012	-.0012
	.99	.01	-.0087	-.0200	-.0262	-.00	-.0001	-.0003	-.0003