

Improved Variance Estimation of Maximum Likelihood Estimators in Stable First-Order Dynamic Regression Models

Jan F. KIVIET and Garry D. A. PHILLIPS

31 October 2012

EGC Report No: 2012/06

HSS-04-88
Tel: +65 67905689
Email: D-EGC@ntu.edu.sg
<http://egc.hss.ntu.edu.sg>



The author(s) bear sole responsibility for this paper.

Views expressed in this paper are those of the author(s) and not necessarily those of the
Economic Growth Centre, NTU.

Improved Variance Estimation of Maximum Likelihood Estimators in Stable First-Order Dynamic Regression Models

JAN F. KIVIET* and GARRY D.A. PHILLIPS†

31 October 2012

JEL-code: C13, C22

Keywords: higher-order asymptotic expansions, bias correction, efficiency gains, lagged dependent variables, finite sample moments, size improvement

Abstract

In dynamic regression models conditional maximum likelihood (least-squares) coefficient and variance estimators are biased. From expansions of the coefficient variance and its estimator we obtain an approximation to the bias in variance estimation and a bias corrected variance estimator, for both the standard and a bias corrected coefficient estimator. These enable a comparison of their mean squared errors to second order. We formally derive sufficient conditions for admissibility of these approximations. Illustrative numerical and simulation results are presented on bias reduction of coefficient and variance estimation for three relevant classes of first-order autoregressive models, supplemented by effects on mean squared errors, test size and size corrected power. These indicate that substantial biases do occur in moderately large samples, but these can be mitigated substantially and may also yield mean squared error reduction. Crude asymptotic tests are cursed by huge size distortions. However, operational bias corrections of both the estimates of coefficients and their estimated variance are shown to curb type I errors reasonably well.

* *corresponding author* Division of Economics, School of Humanities and Social Sciences, Nanyang Technological University, 14 Nanyang Drive, Singapore 637332 (jfkiviet@ntu.edu.sg; +65 65921535) and Amsterdam School of Economics, University of Amsterdam, Valckenierstraat 65, 1018 XE Amsterdam, The Netherlands (j.f.kiviet@uva.nl; +31 20 5254224).

† Cardiff Business School, Aberconway Building, Colum Drive, CF10 3EU, Cardiff, Wales, UK (phillipsgd1@cardiff.ac.uk).

1. Introduction

Using conditioning on initial observations maximum likelihood estimation corresponds to least squares estimation in normal dynamic regression models. It is well known that the resulting estimators of the regression coefficients and of the disturbance variance, and consequently the coefficient variance estimator, are all biased in finite samples. With regard to the estimation of the disturbance variance, asymptotic approximations can be used to show that deflating the sum of squared residuals not by the sample size (T), but by the smaller number known as the degrees of freedom, reduces the order of its bias by a factor of T^{-1} . Kiviet and Phillips (1998) shows that this bias can be reduced by another factor T^{-1} when employing a much more sophisticated expression for degrees of freedom correction. A natural extension to this work is to examine the bias in estimators for the variance of the coefficient estimates. In addition to this, we shall also examine the variance and mean squared error of a bias corrected estimator of the full vector of coefficients, from which conclusions can be drawn on the effectiveness of bias correction and on appropriate variance estimation of (bias corrected) estimators and their effects on test size in stable normal ARX(1) models. In this class of model the dependent variable is determined linearly by an arbitrary number of strongly exogenous stationary or nonstationary regressor variables, the one period lagged dependent variable with coefficient smaller than one in absolute value, and by normally distributed i.i.d. disturbances.

We obtain our approximations to finite sample moments by extending the approach followed by Nagar (1959) in such a way that the approximation errors of the results are of order T^{-1} or T^{-2} , as in Mikhail (1972), or even smaller. This requires the development of a Taylor-type expansion and then the analytical evaluation of the expectation of expressions which involve terms consisting of products of up to four quadratic forms in standard normal vectors. In Kiviet and Phillips (2000) this methodology has been used to show that 2SLS variance estimation in static simultaneous models is upwards biased. The approximation of the moments of statistical estimators in stable autoregressive models by use of asymptotic expansions has already been undertaken for well over half a century. Most early work is particularly concerned with the estimator of the serial correlation coefficient in a first-order autoregressive Gaussian process, see Bartlett (1946), Hurwicz (1950), Kendall (1954), Marriott and Pope (1954), White (1961), Shenton and Johnson (1965) and Sawa (1978). In the latter three studies, which focus on the AR(1) model (with no, a known, or an unknown intercept) an analysis is also given of the variance of the least-squares coefficient estimator. Results on its mean squared error for the model with unknown intercept and nonnormal errors have recently been obtained in Bao (2007). No work has been done yet, however, to find out how well the usual standard deviation estimator estimates the true standard errors in general linear

dynamic econometric models, and what the options are for bias correction of variance estimators of (possibly bias corrected) coefficient estimators. Nor have we seen a rather simple proof of the general admissibility of the Nagar approach as presented here in Appendix A.

Our results concern a more general model than the AR(1), because we allow for any number of arbitrary exogenous regressors in the autoregressive model. As is usual in that context, we will condition on the initial value of the dependent variable. In Kiviet and Phillips (1993, 2012)¹ the focus of attention is the bias of ordinary least-squares (OLS) estimation of all the regression coefficients in the first-order normal linear dynamic regression model. Here the focus is their second moment. The model is

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

where $y = (y_1, \dots, y_T)'$ is a $T \times 1$ vector of observations on a dependent variable, y_{-1} is the y vector lagged one period, i.e. $y_{-1} = (y_0, \dots, y_{T-1})'$, and X is a full column-rank $T \times K$ matrix of observations on K fixed or strongly exogenous regressors (such as a constant, a linear trend, step/impulse/seasonal dummy variables or any other covariates not affected by feedbacks from the dependent variable). The scalar coefficient λ (with $|\lambda| < 1$) and $K \times 1$ coefficient vector β are unknown, and u is a $T \times 1$ vector of independent Gaussian disturbances with zero mean and constant variance σ^2 . The approach that we follow might allow a relaxation of the normality assumption and enable results explicit in cumulants (up to a certain order, see Bao and Ullah, 2007) of the disturbances to be obtained, but this would clutter the already quite complex expressions substantially and is therefore avoided here; the same can be said regarding increasing the order of the dynamics (including y_{-2} etc.) or allowing for a random start-up value². Thus, the present derivations should be seen as a necessary stepping stone in establishing more general results. Although for some special cases of the class of models considered presently exact results are already available, whereas application of the bootstrap may implicitly achieve some of the purposes of our analytical procedures, we nevertheless think that it is useful to follow and develop the present approach, because eventually it should enable a better understanding of the actual nature of finite sample problems especially in cases where it seems impossible to ever obtain exact results and for situations where the bootstrap may break down (as for instance, close to the unit circle and in the presence of further weakly exogenous regressors which would complicate the design of appropriate resampling schemes).

We first reiterate earlier results on the finite sample bias of OLS and next focus on the usual estimator of the (asymptotic) variance of the OLS estimator $\hat{\alpha}$ of the full

¹For related work in stable ARX models see also Kiviet and Phillips (1994) and Kiviet et al. (1995). The finite sample characteristics of the first two moments of the least-squares coefficient estimators in dynamic models with a unit root is analyzed in Kiviet and Phillips (2005).

²Results for variance estimation in ARX(1) models with random y_0 can be found in KP (2010).

coefficient vector $\alpha = (\lambda, \beta')'$ and its bias, and we shall develop a bias corrected variance estimator. We shall also consider a bias corrected estimator $\tilde{\alpha}$ of α and examine its mean squared error and compare with $\hat{\alpha}$, both analytically and experimentally in simulations. Rewriting (1.1) as

$$y = Z\alpha + u, \quad (1.2)$$

where $Z = (y_{-1}, X)$, the OLS estimator of the $(K + 1) \times 1$ vector α is

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

and, based on regularity conditions and asymptotic and finite sample arguments, its variance $V(\hat{\alpha}) = E[\hat{\alpha} - E(\hat{\alpha})][\hat{\alpha} - E(\hat{\alpha})]'$ is usually estimated by

$$\hat{V}(\hat{\alpha}) = s^2 (Z'Z)^{-1}, \quad (1.4)$$

where

$$s^2 = \frac{(y - Z\hat{\alpha})'(y - Z\hat{\alpha})}{T - K - 1}. \quad (1.5)$$

Occasionally the degrees of freedom correction is omitted and σ^2 is estimated by the ML estimator $\hat{\sigma}^2 = (y - Z\hat{\alpha})'(y - Z\hat{\alpha})/T$. The coefficient variance estimator $\hat{\sigma}^2(Z'Z)^{-1}$ disregards finite sample considerations. Note that in this model the derivation of moments such as $E(\hat{\alpha})$, $V(\hat{\alpha})$ and $E[\hat{V}(\hat{\alpha})]$ is non-trivial, because Z is stochastic and depends linearly on u , whereas $\hat{\alpha}$ depends on u and nonlinearly on Z , so these moments involve expressions which are all highly nonlinear in u .

Below in Section 2 we first rewrite Z in such a way that its dependence on u becomes fully explicit, and next we produce, for the various moments of interest, expansions consisting of individual terms whose expectations can be obtained analytically upon using particular basic results. From these we obtain approximations to the MSE (mean squared error) and the true variance of $\hat{\alpha}$ in the general ARX(1) model, and also to the expectation of estimators of this variance. Even though we do not start off from an explicit representation for the true variance (but only from a higher-order asymptotic approximation), these results can be used to develop a bias correction to the standard asymptotic variance estimator. In Appendix A we prove (which is a novelty in this line of research) the validity of the claimed order of magnitude of the approximation errors, and hence the admissibility of the Nagar approach in this model. Next, in Section 3, we examine the first and second moments of an implementation of a bias corrected estimator, which is unbiased to order T^{-1} . In Section 4 we specialize the general results and examine their implications for the specific case of a simple AR(1) model with an unknown intercept. Here some remarkably simple analytic results on the scope for bias and mean squared error reduction are obtained. In Section 5 we verify by Monte Carlo simulation the numerical magnitude of the bias of alternative coefficient and variance estimators and their effects on test size and power for a range of particular cases. Finally,

in Section 6, we summarize our main conclusions. Proofs are indicated in a series of Appendices; more detailed proofs can be found in Kiviet and Phillips (2010).

2. Bias of variance estimators in ARX(1) models

The starting point for our analysis is summarized as follows.

ASSUMPTION 2.1: *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 and $\mathbf{E}(Z'Z)$ is nonsingular; (iv) the regressors in X are strongly exogenous; (v) the disturbances follow $u \sim \mathbf{N}(0, \sigma^2 I_T)$.*

For the derivations to follow the assumption (ii), which excludes a linear trend or any $I(1)$ regressors, yields a worst case analysis. The approximations we obtain can be shown to be valid as well when some of the regressors are nonstationary, see Kiviet and Phillips (2012). The only difference is that they then have a higher degree of accuracy than under (ii). In what follows moments will always be taken conditional on X and y_0 (without indicating this explicitly).

In order to distinguish the fixed (conditional on X and y_0) and Zero-mean stochastic elements of the regressor matrix Z , we decompose $Z = \bar{Z} + \tilde{Z}$, where \bar{Z} is defined as the mathematical expectation of Z , i.e.

$$\bar{Z} = \mathbf{E}(Z) = [\mathbf{E}(y_{-1}), X] = (\bar{y}_{-1}, X) \quad (2.1)$$

$$\tilde{Z} = Z - \bar{Z} = (y_{-1} - \bar{y}_{-1}, X - X) = (\tilde{y}_{-1}, O) = \tilde{y}_{-1}e'_1, \quad (2.2)$$

where $e_1 = (1, 0, \dots, 0)'$ is a unit vector of $K+1$ elements. It follows directly from model (1.1) that

$$\bar{y}_{-1} = \begin{pmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ \lambda & 1 & \cdot & & & \cdot \\ \lambda^2 & \lambda & 1 & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & & & \cdot & \cdot & 0 \\ \lambda^{T-1} & \cdot & \cdot & \cdot & \lambda & 1 \end{pmatrix} \begin{pmatrix} y_0 \\ x'_1\beta \\ x'_2\beta \\ \cdot \\ \cdot \\ x'_{T-1}\beta \end{pmatrix}, \quad (2.3)$$

where $X' = (x_1, \dots, x_T)$, and hence we find that \bar{Z} is determined by X , y_0 , β and λ .

Defining

$$F = \begin{pmatrix} 1 \\ \lambda \\ \lambda^2 \\ \cdot \\ \cdot \\ \lambda^{T-1} \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 1 & 0 & & & & \cdot \\ \lambda & 1 & 0 & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & \cdot & \cdot \\ \lambda^{T-2} & \cdot & \cdot & \lambda & 1 & 0 \end{pmatrix}, \quad (2.4)$$

one can easily verify that

$$\tilde{y}_{-1} = Cu \quad \text{and} \quad \tilde{Z} = Cue'_1. \quad (2.5)$$

We find

$$\begin{aligned} \mathbf{E}(Z'Z) &= \mathbf{E}(\bar{Z} + Cue'_1)'(\bar{Z} + Cue'_1) \\ &= \bar{Z}'\bar{Z} + \sigma^2 \text{tr}(C'C)e_1e'_1 \end{aligned} \quad (2.6)$$

and we shall denote the inverse of $\mathbf{E}(Z'Z)$ by Q , whereas q_1 denotes the first column of Q , and q_{11} the first element of q_1 , hence:

$$Q = [\mathbf{E}(Z'Z)]^{-1}, \quad q_1 = Qe_1, \quad q_{11} = e'_1Qe_1. \quad (2.7)$$

Using the same notation the following results have been proved in Kiviet and Phillips (2012).

THEOREM 2.1: *Under Assumption 2.1 the bias of the least-squares estimator (1.3) can be approximated to first order as $\mathbf{E}(\hat{\alpha} - \alpha) = \mathbf{B}_1(\hat{\alpha}) + o(T^{-1})$, where*

$$\mathbf{B}_1(\hat{\alpha}) = -\sigma^2[\text{tr}(Q\bar{Z}'C\bar{Z})q_1 + Q\bar{Z}'C\bar{Z}q_1 + 2\sigma^2q_{11} \text{tr}(CC'C)q_1].$$

COROLLARY 2.1: *Under Assumption 2.1 the bias of the least-squares estimator $\hat{\lambda}$ can be approximated to first order as*

$$\mathbf{E}(\hat{\lambda} - \lambda) = -\sigma^2[q_{11} \text{tr}(Q\bar{Z}'C\bar{Z}) + q'_1\bar{Z}'C\bar{Z}q_1 + 2\sigma^2q_{11}^2 \text{tr}(CC'C)] + o(T^{-1}).$$

In fact, Kiviet and Phillips (2012) presents a more accurate and complicated second order approximation to the bias of $\hat{\alpha}$. However, for our present purposes the $O(T^{-1})$ bias approximation of Theorem 2.1 suffices. Notice that the bias approximation in Theorem 2.1 involves the term $\text{tr}(CC'C)$ which is $O(T)$. This could be replaced by an approximation in λ , correct to $O(1)$, without changing the order of the bias approximation. Such an approximation is given in Appendix G but we prefer not to introduce it in Theorem 2.1 because it is somewhat cumbersome. In later theorems and corollaries other trace terms involving the C matrix will appear and suitable approximations for them can also be found in Appendix G.

In order to obtain the bias approximation in Theorem 2.1 one has to find an expansion (in this particular case of the estimation error) in such a form that successive terms are of decreasing stochastic order so that the order of the remainder term is known, whereas the individual terms in the expansion have an expectation which can be derived analytically. Such results have been presented for various settings and special cases in the literature already but usually little attention is being paid to their validity, in particular to the existence of the expectation of the remainder term in the expansion. That the type of approximations that we use in the present model are admissible is demonstrated for the result in Theorem 2.1 in Appendix A. Irrespective of whether one wants to approximate (the bias in) the first or the second moment of estimators for the coefficients (or for the disturbance variance), the typical expansion will involve terms in which particular types of expressions occur frequently. For many of these typical expressions their expectation can be found in Kiviet and Phillips (2010, 2012), see also Bao and Ullah (2010) and Ullah (2004).

We shall present results now that are relevant in order to obtain further insight into matters of interest regarding (the estimation of) the second moment of the full vector of least-squares coefficient estimators. In Appendix B we derive:

THEOREM 2.2: *Under Assumption 2.1 we find for the variance of the estimator $\hat{\alpha}$ the approximation $V(\hat{\alpha}) = E\{[\hat{\alpha} - E(\hat{\alpha})][\hat{\alpha} - E(\hat{\alpha})]'\} =$*

$$\begin{aligned}
& \sigma^2 Q \\
& + \sigma^4 \{ [\text{tr}(Q\bar{Z}'CC'\bar{Z}) - 2\text{tr}(Q\bar{Z}'CC\bar{Z}) + \text{tr}(Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z})]q_1q_1' \\
& \quad - Q\bar{Z}'(CC + C'C + C'C')\bar{Z}q_1q_1' - q_1q_1'\bar{Z}'(CC + C'C + C'C')\bar{Z}Q \\
& \quad + Q\bar{Z}'C\bar{Z}q_1q_1'\bar{Z}'C\bar{Z}Q + Q\bar{Z}'(C + C')\bar{Z}q_1q_1'\bar{Z}'C'\bar{Z}Q \\
& \quad + q_1q_1'\bar{Z}'(C + C')\bar{Z}Q\bar{Z}'C\bar{Z}Q + q_1q_1'\bar{Z}'C'\bar{Z}Q\bar{Z}'(C + C')\bar{Z}Q \\
& \quad + Q\bar{Z}'(C + C')\bar{Z}Q\bar{Z}'C\bar{Z}q_1q_1' + Q\bar{Z}'C'\bar{Z}Q\bar{Z}'(C + C')\bar{Z}q_1q_1' \\
& \quad + \text{tr}(Q\bar{Z}'C\bar{Z})(q_1q_1'\bar{Z}'C\bar{Z}Q + Q\bar{Z}'C'\bar{Z}q_1q_1') + (q_1'\bar{Z}'C\bar{Z}q_1)Q\bar{Z}'(C + C')\bar{Z}Q \\
& \quad + q_{11}[\text{tr}(Q\bar{Z}'C\bar{Z})Q\bar{Z}'(C + C')\bar{Z}Q + Q\bar{Z}'(CC' - CC - C'C')\bar{Z}Q \\
& \quad + Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}Q + Q\bar{Z}'C'\bar{Z}Q\bar{Z}'C'\bar{Z}Q] \} \\
& + 2\sigma^6 \{ 6q_1'\bar{Z}'C\bar{Z}q_1 \text{tr}(CC'C)q_1q_1' \\
& \quad + q_{11}[\text{tr}(CC'CC') - 4\text{tr}(CC'CC) - 2\text{tr}(CC'C'C) + 2\text{tr}(CC'C)\text{tr}(Q\bar{Z}'C\bar{Z})]q_1q_1' \\
& \quad + q_{11}\text{tr}(CC'C)(2Q\bar{Z}'C\bar{Z}q_1q_1' + 2q_1q_1'\bar{Z}'C'\bar{Z}Q + 3Q\bar{Z}'C'\bar{Z}q_1q_1' + 3q_1q_1'\bar{Z}'C\bar{Z}Q) \\
& \quad + q_{11}^2\text{tr}(CC'C)Q\bar{Z}'(C + C')\bar{Z}Q \} \\
& + 20\sigma^8 q_{11}^2 [\text{tr}(CC'C)]^2 q_1q_1' + o(T^{-2}).
\end{aligned}$$

COROLLARY 2.2: *Under Assumption 2.1 we find from Theorem 2.2 for the variance of*

the estimator $\hat{\lambda}$ the approximation $\mathbf{V}(\hat{\lambda}) = \mathbf{E}[\hat{\lambda} - \mathbf{E}(\hat{\lambda})]^2 =$

$$\begin{aligned} & \sigma^2 q_{11} \\ & + \sigma^4 \{ 5(q'_1 \bar{Z}' C \bar{Z} q_1)^2 \\ & \quad + q_{11} [6q'_1 \bar{Z}' C \bar{Z} Q \bar{Z}' C \bar{Z} q_1 + 4q'_1 \bar{Z}' C' \bar{Z} Q \bar{Z}' C \bar{Z} q_1 \\ & \quad + q'_1 \bar{Z}' (3CC' - 6CC - 4C'C) \bar{Z} q_1 + 4q'_1 \bar{Z}' C \bar{Z} q_1 \text{tr}(Q \bar{Z}' C \bar{Z})] \\ & \quad + q_{11}^2 [\text{tr}(Q \bar{Z}' C C' \bar{Z}) - 2 \text{tr}(Q \bar{Z}' C C \bar{Z}) + \text{tr}(Q \bar{Z}' C \bar{Z} Q \bar{Z}' C \bar{Z})] \} \\ & + 2\sigma^6 \{ 18q_{11}^2 q'_1 \bar{Z}' C \bar{Z} q_1 \text{tr}(CC' C) \\ & \quad + q_{11}^3 [\text{tr}(CC' C C') - 4 \text{tr}(CC' C C) - 2 \text{tr}(CC' C' C) + 2 \text{tr}(CC' C) \text{tr}(Q \bar{Z}' C \bar{Z})] \} \\ & + 20\sigma^8 q_{11}^4 [\text{tr}(CC' C)]^2 + o(T^{-2}). \end{aligned}$$

Next we shall examine how closely the above rather complex approximation to the actual variance of the coefficient estimator corresponds to the expectation of the usual estimator for the true variance. In Appendix C we prove:

THEOREM 2.3: *Under Assumption 2.1 we find for the expectation of the usual estimator $\hat{\mathbf{V}}(\hat{\alpha})$, given in (1.4), of $\mathbf{V}(\hat{\alpha})$ the approximation $\mathbf{E}[\hat{\mathbf{V}}(\hat{\alpha})] = \mathbf{E}[s^2(Z'Z)^{-1}] =$*

$$\begin{aligned} & \sigma^2 Q + \sigma^4 \{ [\text{tr}(Q \bar{Z}' C C' \bar{Z}) - 2T^{-1} \text{tr}(C' C)] q_1 q'_1 + Q \bar{Z}' C C' \bar{Z} q_1 q'_1 + (q_1 q'_1 + q_{11} Q) \bar{Z}' C C' \bar{Z} Q \} \\ & + 2\sigma^6 q_{11} \text{tr}(CC' C C') q_1 q'_1 + o(T^{-2}). \end{aligned}$$

COROLLARY 2.3: *Under Assumption 2.1 the expectation of the usual estimator of the variance of the estimator $\hat{\lambda}$ can be approximated as $\mathbf{E}[\hat{\mathbf{V}}(\hat{\lambda})] = \mathbf{E}[s^2 e'_1 (Z'Z)^{-1} e_1] =$*

$$\begin{aligned} & \sigma^2 q_{11} + \sigma^4 \{ 3q_{11} (q'_1 \bar{Z}' C C' \bar{Z} q_1) + q_{11}^2 [\text{tr}(Q \bar{Z}' C C' \bar{Z}) - 2T^{-1} \text{tr}(C' C)] \} \\ & + 2\sigma^6 q_{11}^3 \text{tr}(CC' C C') + o(T^{-2}). \end{aligned}$$

Note that the approximation to order T^{-1} (the leading term) of both $\mathbf{V}(\hat{\alpha})$ and $\mathbf{E}[\hat{\mathbf{V}}(\hat{\alpha})]$ is simply $\sigma^2 Q$. However, the second-order approximations of $\mathbf{V}(\hat{\alpha})$ and $\mathbf{E}[\hat{\mathbf{V}}(\hat{\alpha})]$ differ markedly with respect to contributions of order T^{-2} . Note that Theorem 2.3 implies that the first-order approximation to $\mathbf{E}[\hat{\sigma}^2(Z'Z)^{-1}]$, the estimator which omits a degrees of freedom correction, is given by $\sigma^2 Q$ too; self-evidently, the degrees of freedom correction does not affect the leading term. Since the second-order approximation to $\mathbf{E}[\hat{\sigma}^2(Z'Z)^{-1}]$ equals the expression given in Theorem 2.3 plus the term $-\frac{K+1}{T} \sigma^2 Q$ we find that this differs from both the expressions given in Theorems 2.2 and 2.3. Whether or not these differences have an actual magnitude that is worth bothering about has to be found out by numerical evaluation of these expressions for given values of X , y_0 , α and σ^2 at relevant sample sizes T , and by comparing these approximate expressions with estimates of the true variance which can be obtained from Monte Carlo experiments. If these differences can be substantial it would seem interesting to develop a corrected estimator of $\mathbf{V}(\hat{\alpha})$,

say $\check{V}(\hat{\alpha})$, which adds particular terms to the standard estimator $\hat{V}(\hat{\alpha})$, such that $E[\check{V}(\hat{\alpha})]$ is equivalent to $V(\hat{\alpha})$ to second order. We shall elaborate on the issue of bias reduction of variance and of coefficient estimators in the next section.

A more focussed comparison of the above analytical results on variance matrices is possible if we limit ourselves to the simpler scalar results for the single lagged dependent variable coefficient λ . Again we note that the two approximations given in Corollaries 2.2 and 2.3 differ substantially with respect to their order T^{-2} terms, which may be an indication that there is scope for developing a second-order unbiased estimator $\check{V}(\hat{\lambda})$ for $V(\hat{\lambda})$.

3. The efficiency of bias corrected coefficient estimators

The approach already laid out in the foregoing section consists of three stages: (i) approximate the second moment of a coefficient estimator to second order and next, (ii) obtain to second order the expectation of a variance estimator of that coefficient estimator in order to, (iii) exploit these results to correct the variance estimator such that it will become unbiased to second order. This can also be applied to a bias corrected least-squares estimator in which the result of Theorem 2.1 has been exploited such that the corrected estimator is unbiased to order T^{-1} . For the expression $\hat{\alpha} - \mathbf{B}_1(\hat{\alpha})$, with $\mathbf{B}_1(\hat{\alpha})$ as given in Theorem 2.1, it is obvious that this has expectation $\alpha + o(T^{-1})$, but it is not an operational estimator, because σ^2 , Q , \bar{Z} , and C are unobservable. However, consider the operational corrected ordinary least-squares (COLS) estimator defined as

$$\begin{aligned}\check{\alpha} &\equiv \hat{\alpha} - \hat{\mathbf{B}}_1(\hat{\alpha}), \text{ with} \\ \hat{\mathbf{B}}_1(\hat{\alpha}) &= -s^2[\text{tr}(P\hat{Z}'\hat{C}\hat{Z})p_1 + P\hat{Z}'\hat{C}\hat{Z}p_1 + 2s^2p_{11}\text{tr}(\hat{C}\hat{C}'\hat{C})p_1],\end{aligned}\tag{3.1}$$

where $\hat{\alpha}$ and s^2 are the usual least-squares estimators, $P = (Z'Z)^{-1}$, which has first column p_1 with first element p_{11} , $\hat{Z} = (\hat{F}y_0 + \hat{C}X\hat{\beta}, X)$ and \hat{C} corresponds to C (as \hat{F} corresponds to F) with the unknown λ replaced by $\hat{\lambda}$. In Appendix D we show that the difference between the corresponding terms in the non-operational and operational forms of the COLS estimator are of stochastic order $T^{-3/2}$; hence both estimators have the same expected value to order T^{-1} . Thus we have:

THEOREM 3.1: *Under Assumption 2.1 the COLS estimator $\check{\alpha} = \hat{\alpha} - \hat{\mathbf{B}}_1(\hat{\alpha})$ given in (3.1) is unbiased to order T^{-1} , i.e. $E(\check{\alpha}) = \alpha + o(T^{-1})$.*

COROLLARY 3.1: *Under Assumption 2.1 the COLS estimator $\check{\lambda}$ for λ defined as*

$$\check{\lambda} \equiv \hat{\lambda} - \mathbf{B}_1(\hat{\lambda}) = \hat{\lambda} + s^2[p_1'\hat{Z}'\hat{C}\hat{Z}p_1 + p_{11}\text{tr}(P\hat{Z}'\hat{C}\hat{Z}) + 2s^2p_{11}^2\text{tr}(\hat{C}\hat{C}'\hat{C})]$$

is unbiased to order T^{-1} .

For estimator $\check{\alpha}$ we obtain in Appendix E:

THEOREM 3.2: *Under Assumption 2.1 we find for the variance of the bias corrected estimator $\check{\alpha}$ given in (3.1) the approximation $\mathbf{V}(\check{\alpha}) = \mathbf{E}\{[\check{\alpha} - \mathbf{E}(\check{\alpha})][\check{\alpha} - \mathbf{E}(\check{\alpha})]'\} =$*

$$\begin{aligned} & \sigma^2 Q \\ & + \sigma^4 \{ [\text{tr}(Q\bar{Z}'CC'\bar{Z}) + \text{tr}(Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z})]q_1q_1' \\ & \quad + Q\bar{Z}'CC'\bar{Z}q_1q_1' + q_1q_1'\bar{Z}'CC'\bar{Z}Q + Q\bar{Z}'C\bar{Z}q_1q_1'\bar{Z}'C'\bar{Z}Q \\ & \quad + Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}q_1q_1' + q_1q_1'\bar{Z}'C'\bar{Z}Q\bar{Z}'C'\bar{Z}Q + q_{11}Q\bar{Z}'CC'\bar{Z}Q \} \\ & + 2\sigma^6 \{ 2q_1'\bar{Z}'C\bar{Z}q_1 \text{tr}(CC'C)q_1q_1' + q_{11} \text{tr}(CC'CC')q_1q_1' \\ & \quad + q_{11} \text{tr}(CC'C)[Q\bar{Z}'C\bar{Z}q_1q_1' + q_1q_1'\bar{Z}'C'\bar{Z}Q] \} \\ & + 4\sigma^8 q_{11}^2 [\text{tr}(CC'C)]^2 q_1q_1' + o(T^{-2}). \end{aligned}$$

COROLLARY 3.2: *Under Assumption 2.1 we find $\mathbf{V}(\check{\lambda}) = \mathbf{E}[\check{\lambda} - \mathbf{E}(\check{\lambda})]^2 =$*

$$\begin{aligned} & \sigma^2 q_{11} + \\ & + \sigma^4 \{ (q_1'\bar{Z}'C\bar{Z}q_1)^2 + q_{11}(2q_1'\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}q_1 + 3q_1'\bar{Z}'CC'\bar{Z}q_1) \\ & \quad + q_{11}^2 [\text{tr}(Q\bar{Z}'CC'\bar{Z}) + \text{tr}(Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z})] \} \\ & + 2\sigma^6 [4q_{11}^2 q_1'\bar{Z}'C\bar{Z}q_1 \text{tr}(CC'C) + q_{11}^3 \text{tr}(CC'CC')] \\ & + 4\sigma^8 q_{11}^4 [\text{tr}(CC'C)]^2 + o(T^{-2}). \end{aligned}$$

Note that the expression for the second order contribution to the variance of the corrected estimator is substantially simpler than for the uncorrected least-squares estimators given in Theorem 2.2 and Corollary 2.2.

In deriving Theorem 3.2 and its Corollary 3.2 we have also obtained an approximation for the MSE of the corrected estimator, because the variance and MSE are equivalent up to the order of the approximation. Comparison of these with the MSE of the uncorrected estimator, which of course differs from the variance in the $O(T^{-2})$ terms due to the $O(T^{-1})$ coefficient bias, yields information on any possible efficiency gains or losses through bias correction.

From the results in Theorems 2.3 and 3.2 it follows that the variance of the COLS estimator $\mathbf{V}(\check{\alpha})$ can be estimated unbiasedly to order T^{-1} by $\hat{\mathbf{V}}(\check{\alpha}) \equiv s^2(Z'Z)^{-1} = s^2P$, i.e. the standard OLS estimator, because $\mathbf{V}(\check{\alpha})$ and $\mathbf{E}[s^2(Z'Z)^{-1}]$ both have leading term $\sigma^2 Q$. However, from the same Theorems it also follows (proof in Appendix F) that an estimator which is unbiased to the order of T^{-2} can be constructed as follows.

THEOREM 3.3: *Under Assumption 2.1 the estimator for the variance $\check{\mathbf{V}}(\check{\alpha})$, of the bias corrected estimator $\check{\alpha}$ given in (3.1), has $\mathbf{E}[\check{\mathbf{V}}(\check{\alpha}) - \mathbf{V}(\check{\alpha})] = o(T^{-2})$, and hence is almost*

unbiased, if we define $\check{V}(\check{\alpha}) \equiv$

$$\begin{aligned} & s^2 P \\ & + s^4 \{ [\text{tr}(P\hat{Z}'\hat{C}\hat{Z}P\hat{Z}'\hat{C}\hat{Z}) + 2T^{-1} \text{tr}(\hat{C}'\hat{C})] p_1 p_1' \\ & \quad + P\hat{Z}'\hat{C}\hat{Z}p_1 p_1' \hat{Z}'\hat{C}'\hat{Z}P + P\hat{Z}'\hat{C}\hat{Z}P\hat{Z}'\hat{C}\hat{Z}p_1 p_1' + p_1 p_1' \hat{Z}'\hat{C}'\hat{Z}P\hat{Z}'\hat{C}'\hat{Z}P \} \\ & + 2s^6 \text{tr}(\hat{C}\hat{C}'\hat{C}) [2(p_1' \hat{Z}'\hat{C}\hat{Z}p_1) p_1 p_1' + p_{11} (P\hat{Z}'\hat{C}\hat{Z}p_1 p_1' + p_1 p_1' \hat{Z}'\hat{C}'\hat{Z}P)] \\ & + 4s^8 p_{11}^2 [\text{tr}(\hat{C}\hat{C}'\hat{C})]^2 p_1 p_1'. \end{aligned}$$

COROLLARY 3.3: *Under Assumption 2.1 the variance estimator $\check{V}(\check{\lambda})$ of the corrected estimator $\check{\lambda}$ is unbiased to second order for its true variance $V(\check{\lambda})$ when defining $\check{V}(\check{\lambda}) \equiv$*

$$\begin{aligned} & s^2 p_{11} + \\ & + s^4 \{ (p_1' \hat{Z}'\hat{C}\hat{Z}p_1)^2 + 2p_{11} (p_1' \hat{Z}'\hat{C}\hat{Z}P\hat{Z}'\hat{C}\hat{Z}p_1) + p_{11}^2 [\text{tr}(P\hat{Z}'\hat{C}\hat{Z}P\hat{Z}'\hat{C}\hat{Z}) + 2T^{-1} \text{tr}(\hat{C}'\hat{C})] \} \\ & + 8s^6 p_{11}^2 p_1' \hat{Z}'\hat{C}\hat{Z}p_1 \text{tr}(\hat{C}\hat{C}'\hat{C}) \\ & + 4s^8 p_{11}^4 [\text{tr}(\hat{C}\hat{C}'\hat{C})]^2. \end{aligned}$$

It is relatively easy now to obtain, in the same spirit as the result of Theorem 3.3, an operational bias corrected estimator $\check{V}(\hat{\alpha})$ for the variance of the uncorrected OLS estimator. In the next sections we shall just encounter its element $\check{V}(\hat{\lambda})$, and for the sake of completeness we simply present therefore (without further derivations) its formula

$$\begin{aligned} \check{V}(\hat{\lambda}) \equiv & s^2 p_{11} + \\ & + s^4 \{ 5(p_1' \hat{Z}'\hat{C}\hat{Z}p_1)^2 \\ & \quad + p_{11} [6p_1' \hat{Z}'\hat{C}\hat{Z}P\hat{Z}'\hat{C}\hat{Z}p_1 + 4p_1' \hat{Z}'\hat{C}'\hat{Z}P\hat{Z}'\hat{C}\hat{Z}p_1 \\ & \quad - 6p_1' \hat{Z}'\hat{C}\hat{C}\hat{Z}p_1 - 4p_1' \hat{Z}'\hat{C}'\hat{C}\hat{Z}p_1 + 4p_1' \hat{Z}'\hat{C}\hat{Z}p_1 \text{tr}(P\hat{Z}'\hat{C}\hat{Z})] \\ & \quad + p_{11}^2 [\text{tr}(P\hat{Z}'\hat{C}\hat{Z}P\hat{Z}'\hat{C}\hat{Z}) - 2\text{tr}(P\hat{Z}'\hat{C}\hat{C}\hat{Z}) + 2T^{-1} \text{tr}(\hat{C}'\hat{C})] \} \\ & + s^6 \{ 36p_{11}^2 p_1' \hat{Z}'\hat{C}\hat{Z}p_1 \text{tr}(\hat{C}\hat{C}'\hat{C}) \\ & \quad - p_{11}^3 [8\text{tr}(\hat{C}\hat{C}'\hat{C}\hat{C}) + 4\text{tr}(\hat{C}\hat{C}'\hat{C}'\hat{C}) - 4\text{tr}(\hat{C}\hat{C}'\hat{C}) \text{tr}(P\hat{Z}'\hat{C}\hat{Z})] \} \\ & + 20s^8 p_{11}^4 [\text{tr}(\hat{C}\hat{C}'\hat{C})]^2. \end{aligned} \tag{3.2}$$

4. Results for the AR(1) model with intercept

In this section we focus on the variance of the OLS and COLS estimators for the lagged dependent variable coefficient λ in the model of Assumption 2.1 with an intercept as the one and only exogenous regressor, hence

$$y_t = \lambda y_{t-1} + \beta + u_t. \tag{4.1}$$

In order to obtain specific results for this special model from our general formulas given in the earlier sections, it is helpful to rescale the model. Defining

$$y_t^* \equiv \frac{1}{\sigma} \left(y_t - \frac{\beta}{1-\lambda} \right), \quad t = 0, \dots, T \tag{4.2}$$

and substituting in (4.1), we obtain $y_t^* = \lambda y_{t-1}^* + u_t/\sigma$, indicating that general results for the AR(1) model with unknown intercept and arbitrary start-up are obtained by taking in fact a unit disturbance variance, zero intercept and start-up value y_0^* , indicating that all results will be invariant with respect to β and σ .

Corollary 2.1 easily reduces for this AR(1) model to the well-known Kendall (1954) approximation (proofs for all results in this section can be found in Appendix G)

$$\mathbb{E}(\hat{\lambda} - \lambda) = -\frac{1}{T} (1 + 3\lambda) + o(T^{-1}). \quad (4.3)$$

Hence, this approximation proves to be valid irrespective of β , σ and the nature of the start-up value y_0 . From Corollary 2.2 we find for the true variance

$$\mathbb{V}(\hat{\lambda}) = \frac{1 - \lambda^2}{T} - \frac{1 - \lambda^2}{T^2} y_0^{*2} - \frac{1 - 4\lambda - 14\lambda^2}{T^2} + o(T^{-2}), \quad (4.4)$$

where the leading term $(1 - \lambda^2)/T$ is simply the asymptotic variance of $\hat{\lambda}$. Notice that the variance $\mathbb{V}(\hat{\lambda})$ decreases the more y_0 deviates from its stationary mean $\beta/(1 - \lambda)$. Obviously, by adding to $\mathbb{V}(\hat{\lambda})$ the square of the first-order bias, we obtain

$$\text{MSE}(\hat{\lambda}) = \frac{1 - \lambda^2}{T} - \frac{1 - \lambda^2}{T^2} y_0^{*2} + \frac{\lambda(10 + 23\lambda)}{T^2} + o(T^{-2}). \quad (4.5)$$

For the expectation of the standard variance estimator, which is here

$$\hat{\mathbb{V}}(\hat{\lambda}) = s^2 p_{11} = s^2 \left[\sum_{t=1}^T \left(y_{t-1} - \frac{1}{T} \sum_{t=1}^T y_{t-1} \right)^2 \right]^{-1}, \quad (4.6)$$

we find, using Corollary 2.3,

$$\mathbb{E}[\hat{\mathbb{V}}(\hat{\lambda})] = \frac{1 - \lambda^2}{T} - \frac{1 - \lambda^2}{T^2} y_0^{*2} + \frac{2 + 2\lambda + 5\lambda^2}{T^2} + o(T^{-2}). \quad (4.7)$$

It is obvious that $\hat{\mathbb{V}}(\hat{\lambda})$ is unbiased to order T^{-1} , but biased to order T^{-2} , since its second order term differs from the corresponding one of (4.4). So, even though we do not know $\mathbb{V}(\hat{\lambda})$ exactly, we find from its approximation that the standard estimator is biased to second order, viz.

$$\mathbb{E}[\hat{\mathbb{V}}(\hat{\lambda}) - \mathbb{V}(\hat{\lambda})] = \frac{3 - 2\lambda - 9\lambda^2}{T^2} + o(T^{-2}), \quad (4.8)$$

implying that $\hat{\mathbb{V}}(\hat{\lambda})$ overstates (omitting $o(T^{-2})$ terms) whenever $-0.699 < \lambda < 0.477$ and understates otherwise. Employing the same type of reasoning as in Section 3 we can obtain an almost unbiased estimator of $\mathbb{V}(\hat{\lambda})$, viz.

$$\check{\mathbb{V}}(\hat{\lambda}) \equiv \hat{\mathbb{V}}(\hat{\lambda}) - \frac{3 - 2\hat{\lambda} - 9\hat{\lambda}^2}{T^2}, \quad (4.9)$$

which is unbiased to $O(T^{-2})$, and is a special case of (3.2).

For the simple model (4.1) our implementation of COLS leads to (*cf* Sawa, 1978, p.165)

$$\check{\lambda} \equiv \hat{\lambda} + \frac{1}{T}(1 + 3\hat{\lambda}) = \frac{T+3}{T}\hat{\lambda} + \frac{1}{T}. \quad (4.10)$$

Specializing now the result of Corollary 3.2 for this AR(1) model we obtain

$$\begin{aligned} V(\check{\lambda}) &= \frac{1 - \lambda^2}{T} - \frac{1 - \lambda^2}{T^2} y_0^{*2} + \frac{5 + 4\lambda + 8\lambda^2}{T^2} + o(T^{-2}) \\ &= V(\hat{\lambda}) + 6 \left(\frac{1 - \lambda^2}{T^2} \right) + o(T^{-2}). \end{aligned} \quad (4.11)$$

Note that this reflects the fact that correcting an estimator for bias will invariably lead to an increase in variance. However, from (4.5) and (4.11), which is a second order approximation to $MSE(\check{\lambda})$ too, it also follows that

$$MSE(\check{\lambda}) = MSE(\hat{\lambda}) + \frac{5 - 6\lambda - 15\lambda^2}{T^2} + o(T^{-2}), \quad (4.12)$$

from which a rather precise result follows, viz.:

THEOREM 4.1: *In the AR(1) model with intercept the OLS estimator $\hat{\lambda}$ is likely to be more efficient than the bias corrected estimator $\check{\lambda}$ only when $-0.811 < \lambda < 0.411$, because to order T^{-2} such λ values imply $MSE(\hat{\lambda}) < MSE(\check{\lambda})$, and the reverse otherwise.*

Hence, if $\lambda > 0.411$ it seems always beneficial to use the COLS estimator in this model. In a similar way, it can be derived for the AR(1) model with no (or known) intercept that bias correction yields a MSE reduction when $|\lambda| > 0.707$.

An asymptotically valid estimator for the variance $V(\check{\lambda})$ is provided by $\hat{V}(\hat{\lambda})$, but an estimator unbiased to order T^{-2} follows simply from (4.11) and (4.8), viz.

$$\check{V}(\check{\lambda}) \equiv \hat{V}(\hat{\lambda}) + \frac{3 + 2\hat{\lambda} + 3\hat{\lambda}^2}{T^2}, \quad (4.13)$$

although the correction could also be evaluated in $\check{\lambda}$. Note that $\hat{V}(\hat{\lambda})$ is negatively biased for $V(\check{\lambda})$ to second order, because $3 + 2\lambda + 3\lambda^2 > 0$.

Bias correction of AR(1) models has been entertained in the literature in many studies, see *inter alia* Copas (1966), Orcutt and Winokur (1969), Rudebusch (1992) and MacKinnon and Smith (1998). All these studies based their bias correction on the Kendall (1954) approximation to the bias (4.3), although, instead of using (4.10), all the studies just referred to used a bias corrected estimator $\dot{\lambda}$ which is obtained by solving

$$\hat{\lambda} = \dot{\lambda} - \frac{1}{T}(1 + 3\dot{\lambda}).$$

This yields a bias corrected estimator which slightly differs from (4.10), viz.

$$\dot{\lambda} \equiv \frac{T}{T-3}\hat{\lambda} + \frac{1}{T-3}, \quad (4.14)$$

leading to the following relationships:

THEOREM 4.2: *In the AR(1) model with intercept the corrected estimators $\check{\lambda}$ and $\dot{\lambda}$ are both unbiased to order $O(T^{-1})$, but $\dot{\lambda}$ has uniformly (for any $|\lambda| < 1$) smaller second order bias than $\check{\lambda}$, whereas the latter is uniformly more efficient than $\dot{\lambda}$ because $\text{MSE}(\check{\lambda}) < \text{MSE}(\dot{\lambda})$ to order T^{-2} .*

Note that there is no straightforward generalization of $\dot{\lambda}$ for general ARX models unless one is willing to solve highly non-linear equations. In the next section we shall examine the actual numerical significance and accuracy of all the above analytical findings.

5. Numerical results

We shall examine the estimators $\hat{\lambda}$, $\check{\lambda}$ and $\dot{\lambda}$, their actual bias and efficiency, the accuracy of the approximations to their first two moments and the qualities of their respective (bias corrected) variance estimators for three types of models, viz.: (i) model (4.1), i.e. the AR(1) model with intercept; (ii) the AR(1) with both intercept and linear trend; and (iii) the autoregressive model with an intercept and one strongly exogenous regressor generated by an AR(1) process itself. For that purpose we perform various numerical evaluations and execute a series of Monte Carlo experiments in which we shall also examine the effects on test size when bias corrected coefficient or variance estimators are used.

In what follows we simply write $V(\hat{\lambda})$ for what in fact is the Monte Carlo estimate of $V(\hat{\lambda})$. Because we generated a great number of replications (10^5) for each design the Monte Carlo estimates will be very close to the actual population characteristics. Also for the mean over the Monte Carlo replications of $\hat{\lambda}$ we simply write $E(\hat{\lambda})$ and the Monte Carlo estimate of the bias $E(\hat{\lambda}) - \lambda$ of $\hat{\lambda}$ is indicated as $B(\hat{\lambda})$, and likewise for $\check{\lambda}$ and $\dot{\lambda}$. For the mean over the simulations of $\hat{V}(\hat{\lambda})$ we write $E[\hat{V}(\hat{\lambda})]$, and similarly for the estimated expectations of $\check{V}(\hat{\lambda})$ and $\check{V}(\check{\lambda})$. Often the results are given as ratios. Self-evidently, when $E[\hat{V}(\hat{\lambda})]/V(\hat{\lambda})$ is unity this indicates unbiasedness of $\hat{V}(\hat{\lambda})$ and values smaller (greater) than one are found in case of negative (positive) bias, whereas the ratio $\text{MSE}(\check{\lambda})/\text{MSE}(\hat{\lambda})$ indicates an efficiency gain due to bias correction if it is smaller than unity, and so on. In accordance with Assumption 2.1 we restrict ourselves to $|\lambda| < 1$ in all experiments, with an emphasis on positive values. As we also want to explore in particular where in the parameter space the (higher-order) asymptotic approximations break down (which they naturally will do in extreme cases) we examine rather small values of T and a range of λ values including 0.99.

5.1. Results for the AR(1) model with intercept

In the AR(1) model with intercept only, we have two versions of all asymptotic approximations, viz. the general formulae of Sections 2 and 3 after substituting $X = \iota$, which we call the untrimmed expressions, and those of Section 4, which we call the trimmed expressions, because all higher-order terms have been eliminated here. In Tables 1 through 3 we examine the trimmed expressions for the AR(1) model with intercept, assuming mean-stationarity, hence $y_0^* = 0$, giving $E(y_t) = \beta/(1 - \lambda)$. Note that all results for the mean-stationary AR(1) model given here are invariant with respect to β and σ . We write $V_1(\hat{\lambda})$ for the leading term of the asymptotic variance of $\hat{\lambda}$, which is $(1 - \lambda^2)/T$. The second-order asymptotic approximation to $V(\hat{\lambda})$ is denoted as $V_2(\hat{\lambda})$. Using (4.4) we find

$$V_2(\hat{\lambda}) = \frac{1 - \lambda^2}{T} - \frac{1 - \lambda^2}{T^2} y_0^{*2} - \frac{1 - 4\lambda - 14\lambda^2}{T^2}. \quad (5.1)$$

Apart from the bias in $\hat{V}(\hat{\lambda})$ we also examine the bias of two alternative estimators of $V(\hat{\lambda})$, viz. $\hat{V}_2(\hat{\lambda})$ and $\check{V}(\hat{\lambda})$. By $\hat{V}_2(\hat{\lambda})$ we denote the estimator obtained by replacing λ by $\hat{\lambda}$ in (5.1), and the bias corrected estimator $\check{V}(\hat{\lambda})$ is given by (4.9).

In Table 1 we find for the mean-stationary fixed start-up model the following. $B(\hat{\lambda})$, the actual bias of $\hat{\lambda}$, its variance $V(\hat{\lambda})$ and the bias in the standard estimator $\hat{V}(\hat{\lambda})$ are given in columns (2) through (4) respectively, for various values of λ and T . In fact, column (4) presents the relative bias of $\hat{V}(\hat{\lambda})$ plus unity. In relative terms the bias of $\hat{\lambda}$ is especially serious for small values of λ , and in absolute terms the situation is worst for large positive λ . When $T = 20$ the coefficient bias is large (for positive λ about -25%) and also at $T = 50$ it is still substantial (about -10% for $\lambda \geq 0.2$). Note that for substantial λ the bias of $\hat{\lambda}$ and the standard deviation of $\hat{\lambda}$ are about of equal magnitude. The Kendall formula suggests that the bias changes sign at $\lambda = -0.33$. The actual bias results are found to be in agreement with that, and as a matter of course the bias is small around $\lambda = -0.3$. From column (5) of this table we also see that the Kendall formula for the bias, evaluated at the true (but in practice unknown) value of λ , is rather accurate at $T = 50$ for negative and non-extreme positive values of λ . For λ close to one it understates the actual bias, but oddly enough it is more accurate for positive λ when T is smaller. Column (3) shows how $V(\hat{\lambda})$ changes with λ and T . Column (6) shows that its first-order asymptotic approximation $V_1(\hat{\lambda})$ (evaluated for the true λ values) is very inaccurate, especially at extreme λ values, also at $T = 50$. However, from column (7) we see that the second-order approximation that we developed here is much better, especially at $T = 50$. Column (4) shows that, in agreement with our conclusions from (4.8), the standard estimator for $V(\hat{\lambda})$ systematically overstates for moderate and negative (but not extremely negative) λ values, whereas it is much too optimistic for positive substantial values of λ and even more so at $T = 50$ than at a smaller sample size. Column (9) shows that at small sample size our corrected estimator $\check{V}(\hat{\lambda})$ is actually

very good and at $T = 50$ it is in fact almost unbiased, even for extreme values of λ . For smaller values of T it behaves adequately, though slightly conservatively. From column (8) we see that the alternative estimator $\hat{V}_2(\hat{\lambda})$ in (5.1) is less satisfactory for extreme λ values. Hence, Table 1 shows that in this model the bias corrected variance estimator $\check{V}(\hat{\lambda})$ is a much better variance estimator than the standard expression $\hat{V}(\hat{\lambda})$. However, for correcting the bias in $\hat{\lambda}$ itself, the situation seems less promising, because even when evaluated at the true λ value the bias approximation shows some defects and, especially for large λ and $T = 50$, results are not as accurate as regarding $\check{V}(\hat{\lambda})$. Hence, having an almost unbiased estimator for its variance seems cold comfort. Therefore we proceed to examine a bias correction of $\hat{\lambda}$.

In Table 2 columns (3) and (4) show that the bias corrected estimator $\check{\lambda}$ is much less biased than $\hat{\lambda}$, in general. From column (7) we see that efficiency gains (losses) can be substantial when λ is 0.8 or larger (0.2 or smaller); this is in close agreement with the first-order results of Theorem 4.1. Surprisingly, for larger T the potential efficiency losses decrease, whereas the potential gains increase. Next we examine whether we have an adequate estimator of $V(\check{\lambda})$. Column (5) shows that estimator (4.13) tends to have a negative bias. In column (6) we tried an estimator, denoted $\check{V}^*(\check{\lambda})$, which differs from (4.13), because we evaluated the correction term in $\check{\lambda}$ rather than in $\hat{\lambda}$. Although equivalent to (4.13) to second order, we see that this leads to more satisfactory results. Note that in finite samples there is a non-zero probability that $|\hat{\lambda}| \geq 1$. The frequency of such occurrences is reported in column (2). In practice, where one has a single sample, one might be tempted to adjust such an estimate in one way or another, in order to satisfy the stability assumption $|\lambda| < 1$. This would of course not only affect the bias, since the resulting distribution would be truncated and have a correspondingly different variance. In Table 2 we did not use any such adjustments. Note that Theorem 3.1 holds because $\hat{\lambda} = \lambda + O_p(T^{-1/2})$, with $|\lambda| < 1$, which does not exclude the occurrence of $|\hat{\lambda}| \geq 1$ in the substitution (3.1) that generates $\check{\lambda}$. From column (8) we see that values of $\check{\lambda}$ in the non-stationarity region occur very frequently, especially when λ is extreme and T is small.

The sharp increase of the frequency of values of $\check{\lambda}$ in the non-stationarity region in comparison to this happening with the OLS estimator $\hat{\lambda}$ is slightly worrying. This increase is due to the fact that $\hat{\lambda}$ values which are large in absolute value, induce $\check{\lambda}$ values that are even farther away from zero after bias correction. Although asymptotically valid in general, the estimation of our correction terms is really meant for $\hat{\lambda}$ values which are absolutely smaller than one. These asymptotic properties are not jeopardized, however, if we redefine $\check{\lambda}$ and $\check{V}(\check{\lambda})$ such that the involved corrections of $\hat{\lambda}$ and $\hat{V}(\hat{\lambda})$ are only performed when $|\hat{\lambda}| < 1$ and leaving $\hat{\lambda}$ and $\hat{V}(\hat{\lambda})$ unchanged otherwise. In Table 2A we have done so in columns (2) through (4). We have omitted the frequencies of estimators $\hat{\lambda}$ and $\check{\lambda}$ falling into the non-stationarity region as these are unaffected, also

for $\check{\lambda}$, because the only effect of this re-definition is that when $\hat{\lambda}$ is already in the non-stationarity region, the correction is not pushing it into non-stationarity any further. Self-evidently this adapted procedure has an effect for extreme λ values only, where $|\hat{\lambda}| \geq 1$ actually occurs, and then it leads, as we see, to a slightly less successful bias correction of the coefficient, but also to slightly less bias in the variance estimator and to a minor improvement in MSE.

To really avoid large frequencies of corrected estimators in the non-stationarity region, we could be more drastic in our re-definition of $\check{\lambda}$, viz. not correcting $\hat{\lambda}$ if the correction would lead to a non-stationary value. The estimator $\check{V}(\check{\lambda})$ of such a $\check{\lambda}$ estimator could then be taken as either the original formula when $\hat{\lambda}$ is corrected and $\hat{V}(\hat{\lambda})$ otherwise³. We have examined this procedure in Table 2A columns (5) through (7). Now there is as a rule again less bias reduction in the coefficient estimator, but the variance estimator is always conservative and actually very good, even for very small sample size, whereas there is a substantial improvement again in the MSE ratio's. An obvious attraction of this correction procedure is that it does not aggravate (but neither precludes) the occurrence of estimators in the non-stationarity region, which after all are a natural phenomenon for least-squares in finite samples as this technique does not impose the stability restriction.

In Table 2B we present results for the alternative bias corrected estimator $\dot{\lambda}$. Theorem 4.2 predicts that this estimator is less biased than $\check{\lambda}$, but less efficient too, neglecting the effects of higher order terms. We find at the sample sizes examined (and for versions of $\dot{\lambda}$ and $\check{\lambda}$ that have been corrected in all replications) that the first property generally holds except for λ close to -1 . Especially for moderate values of λ the estimator $\dot{\lambda}$ has much less bias. It is remarkable that $\dot{\lambda}$ is also a bit more efficient than $\check{\lambda}$ when λ is large and positive, but this goes with a higher probability to produce estimates in the non-stationarity region. Because it is computationally cumbersome to generalize $\dot{\lambda}$ for models with more regressors, we will stick here to the type of correction as performed in $\check{\lambda}$.

5.2. Results for the AR(1) model with intercept and another regressor

First we present a few results for the AR(1) model with intercept and trend⁴

$$y_t = \lambda y_{t-1} + \beta_1 + \beta_2 t + u_t, \quad (5.2)$$

³In the latter instance one could also take $\check{V}(\hat{\lambda})$, but some initial experimentation showed that this will yield a heavily biased variance estimator when λ values are close to the unit circle.

⁴Generalizations of this model (also allowing $\lambda = 1$ and higher-order dynamics) are examined in Roy et al. (2004) where the focus is not on estimating and testing λ , but on the trend coefficient.

which we rescale by defining

$$y_t^* \equiv \frac{1}{\sigma} \left[y_t - \frac{1}{1-\lambda} \left(\beta_1 + \frac{\lambda}{1-\lambda} \beta_2 \right) - \frac{1}{1-\lambda} \beta_2 t \right], \quad t = 0, \dots, T. \quad (5.3)$$

Substitution yields $y_t^* = \lambda y_{t-1}^* + u_t/\sigma$, indicating that general results are obtained by taking a unit disturbance variance and a zero intercept and zero trend coefficient, with start-up value y_0^* . We examine the mean-stationary fixed start-up case, with $y_0^* = 0$ i.e. $y_0 = [\beta_1 + \lambda\beta_2/(1-\lambda)]/(1-\lambda)$, for which all the results are invariant with respect to β_1 , β_2 and σ . We found that in models with more regressors than just a constant our original correction procedures work only reasonably well for moderate values of λ with T not too small. However, if correction is performed only conditional on either $\hat{\lambda}$ or $\check{\lambda}$ not in the non-stationarity region, then results of practical interest are also obtained in more extreme cases.

From Table 3 it is seen how badly biased $\hat{\lambda}$ is in the model with trend when T is small. It is also seen that $\hat{V}(\hat{\lambda})$ has serious shortcomings, whereas the corrected variance estimator $\check{V}(\hat{\lambda})$ already works pretty well for $T = 50$. The first order bias approximation $B_1(\hat{\lambda})$ calculated for the true parameter values (not presented in the Table) is not very accurate for small T , especially not for large λ , and consequently the estimator $\check{\lambda}$ shows still substantial bias. Nevertheless column (7) shows that it is more efficient than OLS for $\lambda \geq 0.4$, whereas its variance can be assessed remarkably well by $\check{V}(\check{\lambda})$. Column (9) shows with what probability the corrected estimator would have obtained values in the non-stationarity region if we had not followed again the same strategy as in the last columns of Table 2A, which assures that $\check{\lambda}$ is only in the non-stationarity region when $\hat{\lambda}$ is.

Next we shall perform some experiments for the stationary autoregressive model with intercept and one strongly exogenous regressor, which itself is stationary and first-order autoregressive, i.e.

$$\left. \begin{aligned} y_t &= \lambda y_{t-1} + \beta_1 + \beta_2 x_t + \sigma \varepsilon_t, \\ x_t &= \rho x_{t-1} + \xi_t. \end{aligned} \right\} \quad (5.4)$$

Here ε_t and ξ_t are both mutually independent i.i.d. $N(0, 1)$ series. In addition to $|\lambda| < 1$ we assume $|\rho| < 1$ and strong stationarity implies

$$V(y_t) = \frac{1}{1-\lambda^2} \left(\sigma^2 + \frac{\beta_2^2}{1-\rho^2} \frac{1+\lambda\rho}{1-\lambda\rho} \right). \quad (5.5)$$

In our simulations we shall only compare models with parameter values such that they have all equivalent signal-to-noise ratio

$$SNR = \frac{V(y_t) - \sigma^2}{\sigma^2}, \quad (5.6)$$

and thus, normalizing $\sigma^2 = 1$, we chose

$$\beta_2 = \left| \sqrt{[SNR(1 - \lambda^2) - \lambda^2](1 - \rho^2) \frac{1 - \lambda\rho}{1 + \lambda\rho}} \right|,$$

which requires $\lambda^2 \leq SNR/(SNR + 1)$. We will fix $SNR = 19$ (i.e. the population $R^2 = 0.95$) and therefore have to restrict our calculations to $|\lambda| \leq 0.975$. For the series x_t we chose $\rho = 0.95$ so that a relatively smooth time-series results. We generated only one arbitrary x_t series, which has been used in all Monte Carlo replications (numbering again 10^5). Hence our results do not pertain to the whole family of models (5.4), but only to a very particular case. Note that, in agreement with this, all our theorems are about moments of coefficient estimators conditional on X .

In Table 4 we show results for $\lambda \geq 0$ only, but also include $T = 100$. Note that the bias is again substantial in small samples, that for large λ values $\hat{V}(\hat{\lambda})$ is negatively biased, but so is $\check{V}(\hat{\lambda})$, although this behaves notably better. Because the first order bias approximation $B_1(\hat{\lambda})$ is reasonably accurate (if λ is not very large whilst T very small) $\check{\lambda}$ shows substantial bias reduction, especially for larger T and moderate λ . Therefore $\check{\lambda}$ is much more efficient than $\hat{\lambda}$, especially for larger λ values. Even when $T = 100$ and the least-squares bias is not very large (although note that it is not much smaller than the standard deviation of $\hat{\lambda}$) still substantial efficiency gains are achieved by bias correction. However, $\check{V}(\check{\lambda})$ underestimates the variance when there is an efficiency improvement, though it is less biased than $\check{V}(\hat{\lambda})$ and $\hat{V}(\hat{\lambda})$ are. The final column shows that leaving OLS uncorrected did not occur very often for larger T .

5.3. Results on test statistics

Finally we examine how bias correction of either or both coefficient and variance estimators affect operational test procedures in which any bias corrections have been estimated. We consider various one-sided tests of $H_0 : \lambda = \lambda_0$ both against the left-hand and against the right-hand side alternative, all at nominal significance level 5%. They are the habitual test statistic $t \equiv (\hat{\lambda} - \lambda_0)/[\hat{V}(\hat{\lambda})]^{1/2}$, and three corrected statistics where the estimators $\check{\lambda}$, $\check{V}(\hat{\lambda})$ and $\check{V}(\check{\lambda})$ are exploited, i.e. $t^* \equiv (\check{\lambda} - \lambda_0)/[\hat{V}(\hat{\lambda})]^{1/2}$ where the bias corrected coefficient is used, $t_* \equiv (\hat{\lambda} - \lambda_0)/[\check{V}(\hat{\lambda})]^{1/2}$ where only the denominator $\check{\lambda}$ has been corrected, and $t_*^* \equiv (\check{\lambda} - \lambda_0)/[\check{V}(\check{\lambda})]^{1/2}$ where both the coefficient estimate and its variance have been corrected. Table 5 presents results on actual test size when using Student critical values for the mean-stationary AR(1) model with intercept and trend, for which only untrimmed corrections have been derived. For both alternative hypotheses we find that the size distortions for the standard t -test are extremely severe. Correcting just the numerator (shifting the location of the distribution) does mitigate the problems against right-hand side alternatives and for positive λ_0 also against left-hand-side alternatives. Correcting just the denominator (adjusting the scale of the distribution)

does not help much to cure the size problems. The results for t_*^* show that adjusting both numerator and denominator leads to a test that converges faster to an exact test, especially for moderate and negative values of λ_0 . Performing the correction always (also when $|\check{\lambda}| \geq 1$) does not seem harmful for the tail probabilities.

For models such as those examined here Rudebusch (1992, 1993) suggests tests which mimic t^* . The results in Table 5 suggest that a much more successful finite sample correction of test statistics will be achieved by not just applying bias correction to parameter estimators but also to the associated variance estimators. The required bias approximation to the variance of the already bias corrected coefficient estimator has been derived here for any first-order autoregressive model. Now the remaining question is what the effects of such corrections are for test power. For this we consider again the simple AR(1) model with intercept. This shows substantial size distortions too, especially for large values of λ , but again much less so for the t_*^* statistic. From calculated size-corrected rejection frequencies for a range of λ values under the alternative hypothesis we established that the power seems hardly affected by giving special treatment to estimates outside the stationarity region. Therefore Table 6 just presents results for the standard statistics. These show that the better size control achieved by correcting both numerator and denominator has no detrimental effects on power. On the contrary, when testing values close to one we note an improvement in rejection frequency of false null hypotheses by bias correction of both the estimated coefficient and its estimated variance.

6. Conclusions

By adapting and extending techniques we employed in some earlier papers to approximate to an accuracy of order $O(T^{-2})$ the bias of the least-squares estimators for all the parameters (both coefficients and disturbance variance) in linear regression models with a lagged dependent explanatory variable, we find here an approximation to the same order for the mean squared error and for the true variance of the least-squares coefficient estimator. For the latter approximation we find that its algebraic expression differs substantially from an approximation to the same order of accuracy for the expectation of the expression that is usually employed to estimate the variance on the basis of standard asymptotic reasoning. This means that the usual variance estimator, although asymptotically valid, has a bias in finite samples which can be reduced by employing alternative estimators derived in this paper. We employed similar techniques to approximate the variance of bias corrected coefficient estimators and to develop bias corrected estimators for the variance of such bias corrected coefficient estimators. The analytic results presented in this paper may be illuminating as such (showing how complicated basic notions such as variance and its estimation are in a single linear dynamic regression

model), but should be of use primarily for improving the accuracy of inference in finite samples of single dynamic regression models. Earlier, see Rudebusch (1992, 1993), test statistics were adapted by just using bias corrected coefficient estimates; the present results allow one to adapt their variance estimates as well. We undertook in this study some numerical and simulation analysis to obtain insight into the seriousness of the finite sample inaccuracies of first-order asymptotic expressions for first and second moments of estimators and also into the ability of the higher-order asymptotic analytical approximations to assess and to correct such discrepancies and to improve the control over type I errors in inference.

Our bias corrected coefficient and variance estimators are found to work surprisingly well in the AR(1) model with intercept, even for small sample sizes, but less so close to the unit circle. In models with an extra explanatory variable, such as a linear trend or an arbitrary strongly exogenous AR(1) regressor, we still find surprisingly accurate results after correction, especially when the sample size is not too small. However, it is also found that there may be some adverse effects on the accuracy of higher-order approximations due to the occurrence of estimates of the autoregressive parameter close to or outside the stationarity region, but we show how these effects can be mitigated. Note that it might be possible to achieve even better results by slightly adapting the implementations of our bias corrected versions $\check{\lambda}$, $\check{V}(\hat{\lambda})$ and $\check{V}(\check{\lambda})$, by not taking \hat{C} in the respective formulas, but by iterating at least once and using \check{C} (the same for \hat{Z}). Also σ^2 could be re-estimated on the basis of residuals obtained by employing $\check{\lambda}$ and $\check{\beta}$. Whether such modifications have further positive effects on accuracy and efficiency has not been explored here. We find strong analytical evidence for simple AR(1) models when bias correction will increase the mean squared error ($-0.81 < \lambda < 0.41$) and when it will decrease it (otherwise). In addition, the simulation experiments show more generally that exploiting our higher-order asymptotic findings can often substantially reduce bias and mean squared error at the same time, and almost resolve test size problems without deteriorating power. More particularly, we found that bias correction may be more effective from a mean squared error point of view when the sample size is moderate rather than in smaller samples, where the coefficient bias is usually much larger but harder to assess accurately due to larger variances.

Hopefully the present results for single first-order dynamic models will form a stepping stone to achieve useful results for more general models as for instance the multivariate higher-order VAR model, where the determinants of bias have been studied, see Abadir et al. (1999), but no results are available yet on variance and MSE, which of course are much more relevant (see Mikhail, 1972, footnote 1).

Acknowledgements

Over the years, we benefited from remarks made by various colleagues. In particular we are grateful to Don Andrews, who stimulated us to obtain a proof (see Appendix A) for the general validity of Nagar-type higher-order asymptotic approximations to moments.

References

- Abadir, K.M., Kaddour, H., Tzavalis, E., 1999. The influence of VAR dimensions on estimator biases. *Econometrica* 67, 163-181.
- 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.
- Bao, Y., Ullah, A., 2010. Expectation of quadratic forms in normal and nonnormal variables with applications. *Journal of Statistical Planning and Inference*. 140, 1193-1205.
- Bartlett, M.S., 1946. On the theoretical specification and sampling properties of autocorrelated time-series. *Supplement to the 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.
- Grubb, D., Symons, J., 1987. Bias in regressions with a lagged dependent variable. *Econometric Theory* 3, 371-386.
- Hurwicz, L., 1950. Least squares bias in time series. In: Koopmans, T. (ed.) *Statistical inference in dynamic economic models*. Wiley, New York, 365-383.
- Kendall, M.G., 1954. Note on 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., 2000. The bias of the 2SLS variance estimator. *Economics Letters* 66, 7-15.
- Kiviet, J.F., Phillips, G.D.A., 2005. 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., 2010. Improved variance estimation of coefficients in stable first-order dynamic regression models. *UvA-Econometrics working paper 2002/02*.

Kiviet, J.F., Phillips, G.D.A., 2012. Higher-order asymptotic expansions of the least-squares estimation bias in first-order dynamic regression models. *Computational Statistics and Data Analysis* 56, 3705-3729.

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.

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* 67, 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.

Orcutt, G.H., Winokur Jr., H.S., 1969. First order autoregression: inference, estimation, and prediction. *Econometrica* 37, 1-14.

Roy, A., Falk, B., Fuller, W.A., 2004. Testing for trend in the presence of autoregressive error. *Journal of the American Statistical Association* 99, 1082-1091.

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 GNP. *American Economic Review* 83, 264-272.

Sawa, T., 1978. The exact moments of the least squares estimator for the autoregressive model. *Journal of Econometrics* 8, 159-172.

Shenton, L.R., Johnson, W.L., 1965. Moments of a serial correlation coefficient. *Journal of the Royal Statistical Society, Series B* 27, 308-320.

Ullah, A., 2004. *Finite Sample Econometrics*. Oxford University Press.

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

Appendices

A. Validity of the approximations

Let Assumption 2.1* comprise all elements of Assumption 2.1, including $u_t \sim \text{i.i.d.}(0, \sigma^2)$ for $t = 1, \dots, T$, but excluding the normality of u . Here we examine by what Assumption 2.1* should be supplemented in order to guarantee admissibility of the Nagar approach.

Premultiplying model (1.2) by Z' and taking expectations (and hence assuming that these exist on both sides) gives

$$\mathbf{E}(Z'y) = \mathbf{E}(Z'Z)\alpha, \quad (\text{A.1})$$

because $\mathbf{E}(Z'u) = 0$. That $\mathbf{E}(Z'u \mid X)$ exists and is zero is implied by (2.5) and Assumption 2.1*, since the latter entails the existence of the first two moments of u . Because we assumed that $\mathbf{E}(Z'Z)$ is nonsingular the set of equations (A.1) in α can be solved and leads to

$$\alpha = [\mathbf{E}(Z'Z)]^{-1}\mathbf{E}(Z'y). \quad (\text{A.2})$$

Notice that $\hat{\alpha}$, given in (1.3), is the same function, though of the components of $Z'Z$ and $Z'y$, instead of $\mathbf{E}(Z'Z)$ and $\mathbf{E}(Z'y)$. Hence, upon collecting all distinct components of the data cross-products $Z'Z$ and $Z'y$ in a vector d , we have $\hat{\alpha} = f(d)$ and $\alpha = f(\delta)$, where $\delta = E(d)$.

We shall now focus on the i -th scalar component $f_i(\cdot)$ of the vector valued function $f(\cdot)$ and assume that $f_i(\cdot)$ is differentiable up to third order with derivatives that are uniformly bounded in a neighbourhood of δ as $T \rightarrow \infty$. We also assume that the components of d do not just have finite first moment, as already assumed above, but finite moments up to (at least) third order. Then expanding $\hat{\alpha}_i = f_i(d)$ in a Taylor series about the point δ yields

$$\hat{\alpha}_i = \alpha_i + (d - \delta)' f_i^{(1)}(\delta) + \frac{1}{2!} (d - \delta)' f_i^{(2)}(\delta) (d - \delta) + \frac{1}{3!} \sum_j (d_j - \delta_j) (d - \delta)' f_{ij}^{(3)}(\delta^*) (d - \delta), \quad (\text{A.3})$$

where $f_i^{(1)} = \frac{\partial f_i}{\partial \delta}$, $f_i^{(2)} = \frac{\partial^2 f_i}{\partial \delta \partial \delta'}$ and $f_{ij}^{(3)} = \frac{\partial f_{ij}^{(2)}}{\partial \delta_j}$. The derivatives in $f_{ij}^{(3)}$ are evaluated at δ^* which lies between δ and d . Because all derivatives in (A.3) are bounded as $T \rightarrow \infty$ and Assumption 2.1* implies $d - \delta = O_p(T^{-1/2})$ the successive terms in the expansion are of decreasing order of stochastic magnitude and its remainder term is $O_p(T^{-3/2})$, whereas the expectation of this remainder term is bounded and $o(T^{-1})$. The expectation of the first term being zero, the expectation of the second term in the expansion therefore provides a bias approximation to $O(T^{-1})$. This approximation will coincide with that obtained by the Nagar expansion.

Similar arguments can be used to show that an expansion up to q -th order with $q + 1$ -th order remainder term requires bounded $q + 1$ -th derivatives of f and then,

assuming in addition the existence of finite moments of d up to $q + 1$ -th order, which in the present model requires the existence of the moments of u up to order $2(q + 1)$, an approximation to $E(\hat{\alpha})$ of order $O(T^{-q/2})$ follows with the remainder term having bounded expectation of order $O(T^{-(q+1)/2})$. The same line of reasoning leads to the requirements for admissibility of approximations to the second moment of $\hat{\alpha}$ as obtained in this study.

From the above it follows immediately that the normality that we adopted in Assumption 2.1 is sufficient for the admissibility of the Nagar approach up to any order, because it ensures the existence of the moments of any order of d . What remains are the additional mild assumptions regarding the derivatives of f .

B. An approximation to $V(\hat{\alpha})$

For the variance $V(\hat{\alpha})$ of the least-squares estimator $\hat{\alpha}$ we have

$$\begin{aligned} V(\hat{\alpha}) &= E[\hat{\alpha} - E(\hat{\alpha})][\hat{\alpha} - E(\hat{\alpha})]' \\ &= E[\hat{\alpha} - \alpha][\hat{\alpha} - \alpha]' - [E(\hat{\alpha}) - \alpha][E(\hat{\alpha}) - \alpha]'. \end{aligned} \quad (\text{B.1})$$

We want to approximate this to the order of $O(T^{-2})$. We shall make use of

$$Z'u = \bar{Z}'u + \tilde{Z}'u = \bar{Z}'u + (u'Cu)e_1 = O_p(T^{1/2}), \quad (\text{B.2})$$

The first term of (B.1) is $\text{MSE}(\hat{\alpha})$. For this we find

$$E(\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)' = E(Z'Z)^{-1}Z'uu'Z(Z'Z)^{-1}. \quad (\text{B.3})$$

We first develop an expansion of $(Z'Z)^{-1}$. Referring to (2.6) and (2.7) we have $E(Z'Z) = Q^{-1} = \bar{Z}'\bar{Z} + E(\tilde{Z}'\tilde{Z})$, and so

$$\begin{aligned} Z'Z &= (\bar{Z} + \tilde{Z})'(\bar{Z} + \tilde{Z}) \\ &= E(Z'Z) - E(\tilde{Z}'\tilde{Z}) + \bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z} + \tilde{Z}'\tilde{Z} \\ &= \{I_{K+1} + (\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q + [\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q\}Q^{-1}. \end{aligned} \quad (\text{B.4})$$

Hence,

$$(Z'Z)^{-1} = Q\{I_{K+1} + (\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q + [\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q\}^{-1}, \quad (\text{B.5})$$

where the stochastic terms $(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q$ and $[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q$ are both $O_p(T^{-1/2})$. The inverse matrix of the form $[I_n + A]^{-1}$, with $A = O_p(T^{-1/2})$ an $n \times n$ matrix, may be expanded in $[I_n - A + A^2 - A^3 + \dots]$, whereby successive terms are of decreasing order in probability. The expansion retains terms up to a certain order and in this way an expansion is obtained which includes terms up to any desired order. For an expansion

of $(Z'Z)^{-1}$ to order T^{-2} we require

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

whereas the expansion to order $T^{-3/2}$ amounts to

$$(Z'Z)^{-1} = Q - Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q - Q[\tilde{Z}'\tilde{Z} - \mathbf{E}(\tilde{Z}'\tilde{Z})]Q + o_p(T^{-3/2}), \tag{B.7}$$

and to order T^{-1} we simply have

$$(Z'Z)^{-1} = Q + o_p(T^{-1}). \tag{B.8}$$

The expansion (B.6) for $(Z'Z)^{-1}$ can be written as

$$(Z'Z)^{-1} = Q(I_{K+1} - W_1 - W_2 + W_1W_1 + W_1W_2 + W_2W_1 + W_2W_2) + o_p(T^{-2}) \tag{B.9}$$

where we introduced some further shorthand notation, viz.

$$W_1 = (\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q = \bar{Z}'Cuq'_1 + e_1u'C'\bar{Z}Q = O_p(T^{-1/2}) \tag{B.10}$$

and

$$W_2 = [\tilde{Z}'\tilde{Z} - \mathbf{E}(\tilde{Z}'\tilde{Z})]Q = [u'C'Cu - \sigma^2 \text{tr}(C'C)]e_1q'_1 = O_p(T^{-1/2}). \tag{B.11}$$

Note that after premultiplication by Q we have seven terms in (B.9). Of these the first is $O(T^{-1})$, the second and the third are $O_p(T^{-3/2})$, and the remaining four are all $O_p(T^{-2})$.

This yields the following expansion for the squared estimation errors:

$$\begin{aligned}
(\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)' &= (Z'Z)^{-1}Z'uu'Z(Z'Z)^{-1} \\
&= QZ'uu'ZQ - QW_1Z'uu'ZQ - QW_2Z'uu'ZQ + Q(W_1 + W_2)^2Z'uu'ZQ \\
&\quad - QZ'uu'ZW_1'Q - QZ'uu'ZW_2'Q + QZ'uu'Z(W_1' + W_2')^2Q \\
&\quad + QW_1Z'uu'ZW_1'Q + QW_1Z'uu'ZW_2'Q + QW_2Z'uu'ZW_1'Q \\
&\quad + QW_2Z'uu'ZW_2'Q + o_p(T^{-2}).
\end{aligned} \tag{B.12}$$

Note that

$$\begin{aligned}
Z'uu'Z &= [\bar{Z}'u + (u'Cu)e_1][u'\bar{Z} + (u'Cu)e'_1] \\
&= \bar{Z}'uu'\bar{Z} + \bar{Z}'u(u'Cu)e'_1 + (u'Cu)e_1u'\bar{Z} + (u'Cu)^2e_1e'_1.
\end{aligned} \tag{B.13}$$

We now derive the expectation of the eleven terms of (B.12). For the first one we obtain

$$\mathbf{E}(QZ'uu'ZQ) = \mathbf{E}Q\bar{Z}'uu'\bar{Z}Q + \mathbf{E}Q(u'Cu)^2e_1e'_1Q = \sigma^2Q.$$

For the expectation of the second term of (B.12) we find

$$\begin{aligned}
& \mathbf{E} (QW_1 Z' uu' ZQ) \\
&= \sigma^4 Q \bar{Z}' (CC' + CC) \bar{Z} q_1 q_1' + \sigma^4 q_{11} Q \bar{Z}' (CC' + CC) \bar{Z} Q \\
&\quad + \sigma^4 q_1 q_1' \bar{Z}' (CC' + CC) \bar{Z} Q + \sigma^4 \text{tr}(Q \bar{Z}' CC \bar{Z}) q_1 q_1' + \sigma^4 \text{tr}(Q \bar{Z}' CC' \bar{Z}) q_1 q_1',
\end{aligned} \tag{B.14}$$

for the third

$$\begin{aligned}
& \mathbf{E} (QW_2 Z' uu' ZQ) \\
&= 2\sigma^4 q_1 q_1' \bar{Z}' C' C \bar{Z} Q + 2\sigma^6 q_{11} [2 \text{tr}(CC' CC) + \text{tr}(CC' C' C) + \text{tr}(CC' CC')] q_1 q_1',
\end{aligned} \tag{B.15}$$

and for the fourth

$$\begin{aligned}
\mathbf{E}[Q(W_1 + W_2)^2 Z' uu' ZQ] &= \mathbf{E}(QW_1 W_1 Z' uu' ZQ) + \mathbf{E}(QW_1 W_2 Z' uu' ZQ) \\
&\quad + \mathbf{E}(QW_2 W_1 Z' uu' ZQ) + \mathbf{E}(QW_2 W_2 Z' uu' ZQ).
\end{aligned} \tag{B.16}$$

We examine these four terms separately. Removing terms that are $o(T^{-2})$ and using $\bar{Z}' \bar{Z} Q = I - \sigma^2 \text{tr}(C' C) e_1 q_1'$ gives

$$\begin{aligned}
\mathbf{E}(QW_1 W_1 Z' uu' ZQ) &= \\
&\sigma^4 (q_1' \bar{Z}' C \bar{Z} q_1) Q \bar{Z}' C \bar{Z} Q + \sigma^4 Q \bar{Z}' C C' \bar{Z} q_1 q_1' + \sigma^4 Q \bar{Z}' C \bar{Z} q_1 q_1' \bar{Z}' C \bar{Z} Q \\
&+ \sigma^4 q_{11} \text{tr}(Q \bar{Z}' C \bar{Z}) Q \bar{Z}' C \bar{Z} Q + \sigma^4 q_{11} Q \bar{Z}' C C' \bar{Z} Q + \sigma^4 q_{11} Q \bar{Z}' C \bar{Z} Q \bar{Z}' C \bar{Z} Q \\
&+ \sigma^4 \text{tr}(Q \bar{Z}' C C' \bar{Z}) q_1 q_1' + \sigma^4 \text{tr}(Q \bar{Z}' C \bar{Z}) q_1 q_1' \bar{Z}' C \bar{Z} Q + \sigma^4 q_1 q_1' \bar{Z}' C C' \bar{Z} Q \\
&+ 2\sigma^4 q_1 q_1' \bar{Z}' C' \bar{Z} Q \bar{Z}' C \bar{Z} Q + \sigma^4 q_1 q_1' \bar{Z}' C \bar{Z} Q \bar{Z}' C \bar{Z} Q + o(T^{-2})
\end{aligned} \tag{B.17}$$

for the first term. For the second term of (B.16) we find

$$\begin{aligned}
& \mathbf{E}(QW_1 W_2 Z' uu' ZQ) \\
&= 2\sigma^6 q_{11} \text{tr}(CC' C) Q \bar{Z}' C \bar{Z} q_1 q_1' + 2\sigma^6 \text{tr}(CC' C) (q_1' \bar{Z}' C \bar{Z} q_1) q_1 q_1' \\
&\quad + 2\sigma^6 q_{11}^2 \text{tr}(CC' C) Q \bar{Z}' C \bar{Z} Q + 2\sigma^6 q_{11} \text{tr}(CC' C) q_1 q_1' \bar{Z}' C \bar{Z} Q + o(T^{-2}),
\end{aligned} \tag{B.18}$$

for the third

$$\begin{aligned}
& \mathbf{E}(QW_2 W_1 Z' uu' ZQ) \\
&= 2\sigma^6 \text{tr}(CC' C) (q_1' \bar{Z}' C \bar{Z} q_1) q_1 q_1' + 2\sigma^6 q_{11} \text{tr}(CC' C) \text{tr}(Q \bar{Z}' C \bar{Z}) q_1 q_1' \\
&\quad + 4\sigma^6 q_{11} \text{tr}(CC' C) q_1 q_1' \bar{Z}' C \bar{Z} Q + o(T^{-2}),
\end{aligned} \tag{B.19}$$

and for the fourth term of (B.16) we obtain

$$\begin{aligned}
& \mathbf{E}(QW_2 W_2 Z' uu' ZQ) \\
&= 2\sigma^6 q_{11} \text{tr}(C' C C' C) q_1 q_1' + 8\sigma^8 q_{11}^2 \text{tr}(CC' C) \text{tr}(CC' C) q_1 q_1' + o(T^{-2}).
\end{aligned} \tag{B.20}$$

Collecting the four terms of (B.16), i.e. the expectation of the fourth term of (B.12), we get

$$\begin{aligned}
& \mathbb{E}[Q(W_1 + W_2)^2 Z'uu'ZQ] \\
= & \sigma^4(q_1' \bar{Z}'C\bar{Z}q_1)Q\bar{Z}'C\bar{Z}Q + \sigma^4Q\bar{Z}'CC'\bar{Z}q_1q_1' + \sigma^4Q\bar{Z}'C\bar{Z}q_1q_1'\bar{Z}'C\bar{Z}Q \\
& + \sigma^4q_{11} \text{tr}(Q\bar{Z}'C\bar{Z})Q\bar{Z}'C\bar{Z}Q + \sigma^4q_{11}Q\bar{Z}'CC'\bar{Z}Q + \sigma^4q_{11}Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}Q \\
& + \sigma^4 \text{tr}(Q\bar{Z}'CC'\bar{Z})q_1q_1' + \sigma^4 \text{tr}(Q\bar{Z}'C\bar{Z})q_1q_1'\bar{Z}'C\bar{Z}Q \\
& + 2\sigma^4q_1q_1'\bar{Z}'C'\bar{Z}Q\bar{Z}'C\bar{Z}Q + \sigma^4q_1q_1'\bar{Z}'CC'\bar{Z}Q + \sigma^4q_1q_1'\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}Q \\
& + 2\sigma^6q_{11} \text{tr}(CC'C)Q\bar{Z}'C\bar{Z}q_1q_1' + 4\sigma^6 \text{tr}(CC'C)(q_1'\bar{Z}'C\bar{Z}q_1)q_1q_1' \\
& + 2\sigma^6q_{11}^2 \text{tr}(CC'C)Q\bar{Z}'C\bar{Z}Q + 6\sigma^6q_{11} \text{tr}(CC'C)q_1q_1'\bar{Z}'C\bar{Z}Q \\
& + 2\sigma^6q_{11} \text{tr}(CC'C) \text{tr}(Q\bar{Z}'C\bar{Z})q_1q_1' + 2\sigma^6q_{11} \text{tr}(C'CC'C)q_1q_1' \\
& + 8\sigma^8q_{11}^2 \text{tr}(CC'C) \text{tr}(CC'C)q_1q_1' + o(T^{-2}).
\end{aligned} \tag{B.21}$$

For the expectation of the fifth term of (B.12) we find

$$\mathbb{E}(QZ'uu'ZW_1'Q) = \mathbb{E}(QW_1Z'uu'ZQ)', \tag{B.22}$$

which is just the transpose of the result for the second term (B.14). For the sixth we find

$$\mathbb{E}(QZ'uu'ZW_2'Q) = \mathbb{E}(QW_2Z'uu'ZQ)', \tag{B.23}$$

which follows easily from (B.15). Likewise the expectation of the seventh term of (B.12) equals the transpose of (B.21), hence

$$\mathbb{E}[QZ'uu'Z(W_1' + W_2')^2Q] = \mathbb{E}[Q(W_1 + W_2)^2Z'uu'ZQ']. \tag{B.24}$$

The expectation of the eighth term of (B.12) is

$$\begin{aligned}
& \mathbb{E}(QW_1Z'uu'ZW_1'Q) \\
= & \sigma^4(q_1' \bar{Z}'\bar{Z}q_1)Q\bar{Z}'CC'\bar{Z}Q + 2\sigma^4Q\bar{Z}'C\bar{Z}q_1q_1'\bar{Z}'C'\bar{Z}Q + \sigma^4 \text{tr}(Q\bar{Z}'C\bar{Z})Q\bar{Z}'C\bar{Z}q_1q_1' \\
& + \sigma^4Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}q_1q_1' + \sigma^4Q\bar{Z}'CC'\bar{Z}Q\bar{Z}'\bar{Z}q_1q_1' + \sigma^4 \text{tr}(Q\bar{Z}'C\bar{Z})q_1q_1'\bar{Z}'C'\bar{Z}Q \\
& + \sigma^4q_1q_1'\bar{Z}'C'\bar{Z}Q\bar{Z}'C'\bar{Z}Q + \sigma^4q_1q_1'\bar{Z}'\bar{Z}Q\bar{Z}'CC'\bar{Z}Q + \sigma^4 \text{tr}(Q\bar{Z}'C\bar{Z}) \text{tr}(Q\bar{Z}'C\bar{Z})q_1q_1' \\
& + \sigma^4 \text{tr}(Q\bar{Z}'\bar{Z}Q\bar{Z}'CC'\bar{Z})q_1q_1' + \sigma^4 \text{tr}(Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z})q_1q_1' \\
& + \sigma^6q_{11}^2 \text{tr}(C'C)Q\bar{Z}'CC'\bar{Z}Q + \sigma^6q_{11} \text{tr}(C'C)Q\bar{Z}'CC'\bar{Z}q_1q_1' \\
& + \sigma^6q_{11} \text{tr}(C'C)q_1q_1'\bar{Z}'CC'\bar{Z}Q + \sigma^6 \text{tr}(C'C)(q_1'\bar{Z}'CC'\bar{Z}q_1)q_1q_1' + o(T^{-2}).
\end{aligned}$$

Substituting $Q\bar{Z}'\bar{Z} = I - \sigma^2 \text{tr}(C'C)q_1e_1'$ and $q_1'\bar{Z}'\bar{Z}q_1 = q_{11} - \sigma^2q_{11}^2 \text{tr}(C'C)$ this yields

$$\begin{aligned}
& \mathbb{E}(QW_1Z'uu'ZW_1'Q) \\
= & \sigma^4q_{11}Q\bar{Z}'CC'\bar{Z}Q + 2\sigma^4Q\bar{Z}'C\bar{Z}q_1q_1'\bar{Z}'C'\bar{Z}Q + \sigma^4 \text{tr}(Q\bar{Z}'C\bar{Z})Q\bar{Z}'C\bar{Z}q_1q_1' \\
& + \sigma^4Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}q_1q_1' + \sigma^4Q\bar{Z}'CC'\bar{Z}q_1q_1' + \sigma^4 \text{tr}(Q\bar{Z}'C\bar{Z})q_1q_1'\bar{Z}'C'\bar{Z}Q \\
& + \sigma^4q_1q_1'\bar{Z}'C'\bar{Z}Q\bar{Z}'C'\bar{Z}Q + \sigma^4q_1q_1'\bar{Z}'CC'\bar{Z}Q + \sigma^4 \text{tr}(Q\bar{Z}'C\bar{Z}) \text{tr}(Q\bar{Z}'C\bar{Z})q_1q_1' \\
& + \sigma^4 \text{tr}(Q\bar{Z}'CC'\bar{Z})q_1q_1' + \sigma^4 \text{tr}(Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z})q_1q_1' + o(T^{-2}).
\end{aligned} \tag{B.25}$$

For the expectation of the ninth term of (B.12) we find

$$\begin{aligned} & \mathbb{E}(QW_1Z'uu'ZW_2'Q) \\ &= 4\sigma^6 q_{11} \text{tr}(CC'C)Q\bar{Z}'C\bar{Z}q_1q_1' + 2\sigma^6 q_{11} \text{tr}(Q\bar{Z}'C\bar{Z}) \text{tr}(CC'C)q_1q_1' \\ & \quad + 2\sigma^6 \text{tr}(CC'C)(q_1'\bar{Z}'C\bar{Z}q_1)q_1q_1' + o(T^{-2}), \end{aligned} \quad (\text{B.26})$$

and for the tenth

$$\mathbb{E}(QW_2Z'uu'ZW_1'Q) = \mathbb{E}(QW_1Z'uu'ZW_2'Q)', \quad (\text{B.27})$$

which is just the transpose of the former term. Finally, the eleventh yields

$$\begin{aligned} & \mathbb{E}(QW_2Z'uu'ZW_2'Q) \\ &= 2\sigma^6 q_{11} \text{tr}(CC'CC')q_1q_1' + 8\sigma^8 q_{11}^2 \text{tr}(CC'C) \text{tr}(CC'C)q_1q_1' + o(T^{-2}). \end{aligned} \quad (\text{B.28})$$

Assembling the various contributions to the mean squared error we obtain after some simplification

$$\begin{aligned} \text{MSE}(\hat{\alpha}) &= \mathbb{E}[(\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)'] = \\ & \sigma^2 Q + \\ & + \sigma^4 Q\bar{Z}'(CC' - CC - 2C'C - C'C')\bar{Z}q_1q_1' \\ & + \sigma^4 q_1q_1'\bar{Z}'(CC' - CC - 2C'C - C'C')\bar{Z}Q \\ & + \sigma^4 q_{11}Q\bar{Z}'(CC' - CC - C'C')\bar{Z}Q + \sigma^4 \text{tr}(Q\bar{Z}'CC'\bar{Z})q_1q_1' + \sigma^4 [\text{tr}(Q\bar{Z}'C\bar{Z})]^2 q_1q_1' \\ & + \sigma^4 \text{tr}(Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z})q_1q_1' - 2\sigma^4 \text{tr}(Q\bar{Z}'CC\bar{Z})q_1q_1' \\ & + \sigma^4 q_{11} \text{tr}(Q\bar{Z}'C\bar{Z})Q\bar{Z}'(C + C')\bar{Z}Q + \sigma^4 (q_1'\bar{Z}'C\bar{Z}q_1)Q\bar{Z}'(C + C')\bar{Z}Q \\ & + \sigma^4 Q\bar{Z}'C\bar{Z}q_1q_1'\bar{Z}'(C + C')\bar{Z}Q + \sigma^4 Q\bar{Z}'(C + C')\bar{Z}q_1q_1'\bar{Z}'C'\bar{Z}Q \\ & + \sigma^4 q_{11}Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}Q + \sigma^4 q_{11}Q\bar{Z}'C'\bar{Z}Q\bar{Z}'C'\bar{Z}Q \\ & + \sigma^4 \text{tr}(Q\bar{Z}'C\bar{Z})q_1q_1'\bar{Z}'(C + C')\bar{Z}Q + \sigma^4 \text{tr}(Q\bar{Z}'C\bar{Z})Q\bar{Z}'(C + C')\bar{Z}q_1q_1' \\ & + \sigma^4 q_1q_1'\bar{Z}'(C + C')\bar{Z}Q\bar{Z}'C\bar{Z}Q + \sigma^4 q_1q_1'\bar{Z}'C'\bar{Z}Q\bar{Z}'(C + C')\bar{Z}Q \\ & + \sigma^4 Q\bar{Z}'(C + C')\bar{Z}Q\bar{Z}'C\bar{Z}q_1q_1' + \sigma^4 Q\bar{Z}'C'\bar{Z}Q\bar{Z}'(C + C')\bar{Z}q_1q_1' \\ & + 6\sigma^6 q_{11} \text{tr}(CC'C)Q\bar{Z}'(C + C')\bar{Z}q_1q_1' + 6\sigma^6 q_{11} \text{tr}(CC'C)q_1q_1'\bar{Z}'(C + C')\bar{Z}Q \\ & + 2\sigma^6 q_{11}^2 \text{tr}(CC'C)Q\bar{Z}'(C + C')\bar{Z}Q \\ & + 12\sigma^6 q_1'\bar{Z}'C\bar{Z}q_1 \text{tr}(CC'C)q_1q_1' + 8\sigma^6 q_{11} \text{tr}(CC'C) \text{tr}(Q\bar{Z}'C\bar{Z})q_1q_1' \\ & + \sigma^6 q_{11} [2 \text{tr}(CC'CC') - 8 \text{tr}(CC'CC) - 4 \text{tr}(CC'C'C)] q_1q_1' \\ & + 24\sigma^8 q_{11}^2 \text{tr}(CC'C) \text{tr}(CC'C)q_1q_1' + o(T^{-2}). \end{aligned} \quad (\text{B.29})$$

From Theorem 2.1 we easily find for the squared bias, the second term of (B.1):

$$\begin{aligned}
& [\mathbf{E}(\hat{\alpha}) - \alpha][\mathbf{E}(\hat{\alpha}) - \alpha]' \tag{B.30} \\
&= \sigma^4 \{ [\text{tr}(Q\bar{Z}'C\bar{Z})]^2 q_1 q_1' + Q\bar{Z}'C\bar{Z}q_1 q_1' \bar{Z}'C'\bar{Z}Q + \text{tr}(Q\bar{Z}'C\bar{Z})[Q\bar{Z}'C\bar{Z}q_1 q_1' + q_1 q_1' \bar{Z}'C'\bar{Z}Q] \} \\
&+ \sigma^6 \{ 4q_{11} \text{tr}(CC'C) \text{tr}(Q\bar{Z}'C\bar{Z})q_1 q_1' + 2q_{11} \text{tr}(CC'C) [Q\bar{Z}'C\bar{Z}q_1 q_1' + q_1 q_1' \bar{Z}'C'\bar{Z}Q] \} \\
&+ \sigma^8 \{ 4q_{11}^2 [\text{tr}(CC'C)]^2 q_1 q_1' \} + o(T^{-2}).
\end{aligned}$$

This result has to be subtracted from the MSE approximation (B.29) to find the required approximation to $\mathbf{V}(\hat{\alpha})$ of Theorem 2.2.

C. An approximation to $\mathbf{E}[s^2(Z'Z)^{-1}]$

For the numerator of the estimator s^2 , given in (1.5), we have, upon using (B.8),

$$\begin{aligned}
(y - Z\hat{\alpha})'(y - Z\hat{\alpha}) &= u'u - u'Z(Z'Z)^{-1}Z'u \tag{C.1} \\
&= u'u - u'(\bar{Z} + \tilde{Z})Q(\bar{Z} + \tilde{Z})'u + o_p(1).
\end{aligned}$$

First we shall examine an approximation to the expectation of the coefficient variance estimator $\hat{\sigma}^2(Z'Z)^{-1}$, where $\hat{\sigma}^2 = (y - Z\hat{\alpha})'(y - Z\hat{\alpha})/T$ and (C.1) yields

$$\hat{\sigma}^2 = T^{-1}(u'u - u'\bar{Z}Q\bar{Z}'u - u'\bar{Z}Q\tilde{Z}'u - u'\tilde{Z}Q\bar{Z}'u - u'\tilde{Z}Q\tilde{Z}'u) + o_p(T^{-1}). \tag{C.2}$$

An order T^{-2} approximation to

$$\mathbf{E}[\hat{\sigma}^2(Z'Z)^{-1}] = \mathbf{E}[(\hat{\sigma}^2 - \sigma^2)(Z'Z)^{-1}] + \sigma^2 \mathbf{E}[(Z'Z)^{-1}] \tag{C.3}$$

is now obtained by employing (C.2) and an expansion for $(Z'Z)^{-1}$ to an appropriate order, upon noting that $(\hat{\sigma}^2 - \sigma^2) = O_p(T^{-1/2})$. The first right-hand term of (C.3) amounts to:

$$\begin{aligned}
& \mathbf{E}[(\hat{\sigma}^2 - \sigma^2)(Z'Z)^{-1}] \tag{C.4} \\
&= \mathbf{E}[(T^{-1}u'u - \sigma^2)(Z'Z)^{-1}] \\
&\quad - T^{-1} \mathbf{E}[(u'\bar{Z}Q\bar{Z}'u + u'\bar{Z}Q\tilde{Z}'u + u'\tilde{Z}Q\bar{Z}'u + u'\tilde{Z}Q\tilde{Z}'u)(Z'Z)^{-1}] + o(T^{-2}) \\
&= -T^{-1}[\sigma^2(K+1)Q + 2\sigma^4 \text{tr}(C'C)q_1 q_1'] + o(T^{-2}).
\end{aligned}$$

An approximation for the second right-hand term of (C.3) can be obtained from (B.6). Note that of the terms in curly brackets the second and the third term have zero mean, while the fifth and sixth term involve factors with zero mean and products of an odd number of zero-mean normal random variables. Hence, when expected values are taken these terms may be ignored. We then have

$$\begin{aligned}
\mathbf{E}[(Z'Z)^{-1}] &= Q + \mathbf{E}[Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q] \tag{C.5} \\
&\quad + \mathbf{E}\{Q[\tilde{Z}'\tilde{Z} - \mathbf{E}(\tilde{Z}'\tilde{Z})]Q[\tilde{Z}'\tilde{Z} - \mathbf{E}(\tilde{Z}'\tilde{Z})]Q\} + o_p(T^{-2}).
\end{aligned}$$

The second term of (C.5) is

$$\begin{aligned} & \mathbf{E}[Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q] \\ &= \sigma^2[Q\bar{Z}'CC'\bar{Z}q_1q_1' + q_{11}Q\bar{Z}'CC'\bar{Z}Q + \text{tr}(Q\bar{Z}'CC'\bar{Z})q_1q_1' + q_1q_1'\bar{Z}'CC'\bar{Z}Q], \end{aligned} \quad (\text{C.6})$$

and the third

$$\mathbf{E}\{Q[\tilde{Z}'\tilde{Z} - \mathbf{E}(\tilde{Z}'\tilde{Z})]Q[\tilde{Z}'\tilde{Z} - \mathbf{E}(\tilde{Z}'\tilde{Z})]Q\} = 2\sigma^4q_{11}\text{tr}(CC'CC')q_1q_1'. \quad (\text{C.7})$$

Gathering terms yields the result

$$\begin{aligned} \sigma^2\mathbf{E}[(Z'Z)^{-1}] &= \sigma^2Q \\ &+ \sigma^4[\text{tr}(Q\bar{Z}'CC'\bar{Z})q_1q_1' + Q\bar{Z}'CC'\bar{Z}q_1q_1' + (q_1q_1' + q_{11}Q)\bar{Z}'CC'\bar{Z}Q] \\ &+ 2\sigma^6q_{11}\text{tr}(CC'CC')q_1q_1' + o_p(T^{-2}). \end{aligned} \quad (\text{C.8})$$

Adding up the terms (C.4) and (C.8) we obtain for (C.3) the approximation $\mathbf{E}[\hat{\sigma}^2(Z'Z)^{-1}] =$

$$\begin{aligned} & T^{-1}(T - K - 1)\sigma^2Q \\ &+ \sigma^4\{[\text{tr}(Q\bar{Z}'CC'\bar{Z}) - 2T^{-1}\text{tr}(C'C)]q_1q_1' + Q\bar{Z}'CC'\bar{Z}q_1q_1' + (q_1q_1' + q_{11}Q)\bar{Z}'CC'\bar{Z}Q\} \\ &+ 2\sigma^6q_{11}\text{tr}(CC'CC')q_1q_1' + o(T^{-2}). \end{aligned} \quad (\text{C.9})$$

From this the result of Theorem 2.3 follows upon multiplying by $T/(T - K - 1)$. The latter affects the leading term, but not the remaining terms to the order of T^{-2} .

D. The bias of the COLS Estimator

The bias of the COLS estimator (3.1) is given by

$$\begin{aligned} \mathbf{E}(\tilde{\alpha} - \alpha) &= \mathbf{E}(\hat{\alpha} - \hat{\mathbf{B}}_1(\hat{\alpha}) - \alpha) \\ &= \mathbf{E}(\mathbf{B}_1(\hat{\alpha}) - \hat{\mathbf{B}}_1(\hat{\alpha}) + \hat{\alpha} - \alpha - \mathbf{B}_1(\hat{\alpha})) \\ &= -\mathbf{E}(\hat{\mathbf{B}}_1(\hat{\alpha}) - \mathbf{B}_1(\hat{\alpha})) + o(T^{-1}). \end{aligned} \quad (\text{D.1})$$

From Theorem 2.1 and (3.1) it follows that

$$\begin{aligned} \hat{\mathbf{B}}_1(\hat{\alpha}) - \mathbf{B}_1(\hat{\alpha}) &= \sigma^2\text{tr}(Q\bar{Z}'C\bar{Z})q_1 - s^2\text{tr}(P\hat{Z}'\hat{C}\hat{Z})p_1 \\ &+ \sigma^2Q\bar{Z}'C\bar{Z}q_1 - s^2P\hat{Z}'\hat{C}\hat{Z}p_1 + 2[\sigma^4q_{11}\text{tr}(CC'CC')q_1 - s^4p_{11}\text{tr}(\hat{C}\hat{C}'\hat{C})p_1]. \end{aligned} \quad (\text{D.2})$$

We shall examine the three pairs of terms of (D.2) in turn by exploiting a series of intermediate results, which have to be developed first. We do that to a level of generality that makes these intermediate results useful for derivations in the next Appendix as well.

From (B.7) we obtain for $P = (Z'Z)^{-1}$ that

$$\begin{aligned} P &= Q + P^* + o_p(T^{-3/2}), \text{ with} \\ P^* &= -Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q - Q[\tilde{Z}'\tilde{Z} - \mathbf{E}(\tilde{Z}'\tilde{Z})]Q \\ &= -Q\bar{Z}'C\bar{Z}q_1' - q_1u'C'\bar{Z}Q - [u'C'C\bar{Z} - \sigma^2\text{tr}(C'C)]q_1q_1' = O_p(T^{-3/2}). \end{aligned} \quad (\text{D.3})$$

From this, it straightforwardly follows that $p_1 = q_1 + p_1^* + o_p(T^{-3/2})$ and $p_{11} = q_{11} + p_{11}^* + o_p(T^{-3/2})$, with $p_1^* = P^*e_1$ and $p_{11}^* = e_1'P^*e_1$ both $O_p(T^{-3/2})$.

In order to find the leading term of $\hat{Z}'\hat{C}\hat{Z} - \bar{Z}'C\bar{Z}$ we have to produce some auxiliary results. First note that for both C and \hat{C} the $(i, j)^{th}$ element is zero for $i \leq j$ and for the elements $i > j$ they are such that, employing a first order Taylor expansion,

$$\begin{aligned} (\hat{C} - C)_{i,j} &= \hat{\lambda}^{i-j-1} - \lambda^{i-j-1} = (\hat{\lambda} - \lambda) \frac{\partial}{\partial \lambda} \lambda^{i-j-1} + o_p(T^{-1/2}) \\ &= (\hat{\lambda} - \lambda)(i - j - 1) \lambda^{i-j-2} + o_p(T^{-1/2}). \end{aligned} \quad (D.4)$$

In fact, because it is easily verified that

$$\frac{\partial}{\partial \lambda} C = CC, \quad (D.5)$$

we can simply write

$$\hat{C} = C + (\hat{\lambda} - \lambda)CC + o_p(T^{-1/2}), \quad (D.6)$$

and similarly, because $\frac{\partial}{\partial \lambda} F = CF$, we have

$$\hat{F} = F + (\hat{\lambda} - \lambda)CF + o_p(T^{-1/2}). \quad (D.7)$$

Further,

$$\begin{aligned} \hat{Z} - \bar{Z} &= (\hat{y}_{-1}, X) - (\bar{y}_{-1}, X) \\ &= (y_0\hat{F} + \hat{C}X\hat{\beta}, X) - (y_0F + CX\beta, X) \\ &= [y_0(\hat{F} - F) + (\hat{C} - C + C)X(\hat{\beta} - \beta + \beta) - CX\beta]e_1'. \end{aligned} \quad (D.8)$$

Substitution of (D.7) and (D.6) yields

$$\begin{aligned} \hat{Z} - \bar{Z} &= \hat{Z}^* + o_p(T^{-1/2}), \text{ with} \\ \hat{Z}^* &= [y_0(\hat{\lambda} - \lambda)CF + CX(\hat{\beta} - \beta) + (\hat{\lambda} - \lambda)CCX\beta]e_1' \\ &= [C\bar{Z}(\hat{\alpha} - \alpha)]e_1' = O_p(T^{-1/2}). \end{aligned} \quad (D.9)$$

Now we obtain

$$\begin{aligned} \hat{Z}'\hat{C}\hat{Z} - \bar{Z}'C\bar{Z} &= \hat{Z}'\hat{C}\hat{Z} - \bar{Z}'\hat{C}\bar{Z} + \bar{Z}'\hat{C}\bar{Z} - \bar{Z}'C\bar{Z} \\ &= (\hat{\lambda} - \lambda)\hat{Z}^*CC\hat{Z}^* + \hat{Z}^*C\hat{Z}^* + (\hat{\lambda} - \lambda)\hat{Z}^*CC\bar{Z} + \hat{Z}^*C\bar{Z} \\ &\quad + (\hat{\lambda} - \lambda)\bar{Z}'CC\hat{Z}^* + \bar{Z}'C\hat{Z}^* + (\hat{\lambda} - \lambda)\bar{Z}'CC\bar{Z} + o_p(T^{1/2}). \end{aligned} \quad (D.10)$$

Noting that only a few of these terms are $O_p(T^{1/2})$ we find, using (D.9),

$$\begin{aligned} \hat{Z}'\hat{C}\hat{Z} &= \bar{Z}'C\bar{Z} + A^* + o_p(T^{1/2}), \text{ with} \\ A^* &= e_1(\hat{\alpha} - \alpha)' \bar{Z}'C' C \bar{Z} + \bar{Z}'CC\bar{Z}(\hat{\alpha} - \alpha)e_1' + (\hat{\lambda} - \lambda)\bar{Z}'CC\bar{Z} = O_p(T^{1/2}). \end{aligned} \quad (D.11)$$

Next we develop a result regarding $\tau = \text{tr}(CC'C) = O(T)$ and $\hat{\tau} = \text{tr}(\hat{C}\hat{C}'\hat{C})$. From (D.5) we find

$$\frac{\partial}{\partial \lambda} \text{tr}(CC'C) = \text{tr} \left[\frac{\partial}{\partial \lambda} (CC'C) \right] = \text{tr}(C'C'CC) + 2 \text{tr}(C'CCC), \quad (\text{D.12})$$

because

$$\frac{\partial}{\partial \lambda} (CC'C) = \left(\frac{\partial C}{\partial \lambda} \right) C'C + C \left(\frac{\partial C'}{\partial \lambda} \right) = CCC'C + CC'C'C + CC'CC.$$

Since $\hat{\tau} = \text{tr}(CC'C) + (\hat{\lambda} - \lambda) \frac{\partial}{\partial \lambda} \text{tr}(CC'C) + o_p(T^{1/2})$ we may write

$$\begin{aligned} \hat{\tau} &= \tau + \hat{\tau}^* + o_p(T^{1/2}), \text{ with} \\ \hat{\tau}^* &= (\hat{\lambda} - \lambda) [\text{tr}(C'C'CC) + 2 \text{tr}(C'CCC)] = O_p(T^{1/2}). \end{aligned} \quad (\text{D.13})$$

Next we consider s^2 . From Kiviet and Phillips (1998) we have

$$s^2 = \sigma^2 + s_*^2 + o_p(T^{-1/2}), \text{ with } s_*^2 = O_p(T^{-1/2}), \quad (\text{D.14})$$

and a Taylor expansion yields

$$\begin{aligned} s^4 &= \sigma^4 + 2(s^2 - \sigma^2)\sigma^2 + o_p(T^{-1/2}) \\ &= \sigma^4 + 2s_*^2\sigma^2 + o_p(T^{-1/2}). \end{aligned} \quad (\text{D.15})$$

This completes the intermediate results which allow to examine the three pairs of terms of (D.2).

For the first pair we find

$$\begin{aligned} &\sigma^2 \text{tr}(Q\bar{Z}'C\bar{Z})q_1 - s^2 \text{tr}(P\hat{Z}'\hat{C}\hat{Z})p_1 \\ &= -\sigma^2 \text{tr}(Q\bar{Z}'C\bar{Z})p_1^* - \sigma^2 \text{tr}(QA^*)q_1 - \sigma^2 \text{tr}(P^*\bar{Z}'C\bar{Z})q_1 - s_*^2 \text{tr}(Q\bar{Z}'C\bar{Z})q_1 + o_p(T^{-3/2}). \end{aligned} \quad (\text{D.16})$$

For the second pair we obtain

$$\begin{aligned} &\sigma^2 Q\bar{Z}'C\bar{Z}q_1 - s^2 P\hat{Z}'\hat{C}\hat{Z}p_1 \\ &= -\sigma^2 Q\bar{Z}'C\bar{Z}p_1^* - \sigma^2 QA^*q_1 - \sigma^2 P^*\bar{Z}'C\bar{Z}q_1 - s_*^2 Q\bar{Z}'C\bar{Z}q_1 + o_p(T^{-3/2}), \end{aligned} \quad (\text{D.17})$$

and for the third

$$\begin{aligned} &\sigma^4 q_{11} \text{tr}(CC'C)q_1 - s^4 p_{11} \text{tr}(\hat{C}\hat{C}'\hat{C})p_1 \\ &= -\sigma^4 q_{11} \tau p_1^* - \sigma^4 q_{11} \hat{\tau}^* q_1 - \sigma^4 p_{11}^* \tau q_1 - 2s_*^2 \sigma^2 q_{11} \tau q_1 + o_p(T^{-3/2}). \end{aligned} \quad (\text{D.18})$$

Upon noting that $\mathbf{E}(P^*) = O$, $\mathbf{E}(A^*) = O(T^{1/2})$, $\mathbf{E}(s_*^2) = 0$ and $\mathbf{E}(\hat{\tau}^*) = O(T^{-1})$ it is now obvious that $\mathbf{E}(\hat{\mathbf{B}}_1(\hat{\alpha}) - \mathbf{B}_1(\hat{\alpha})) = 0 + o(T^{-3/2})$, thus (D.1) implies $\mathbf{E}(\check{\alpha} - \alpha) = o(T^{-1})$, as stated in Theorem 3.1.

E. The variance of the COLS Estimator

The variance $V(\check{\alpha})$ and the $MSE(\check{\alpha})$ of the COLS estimator are the same to order T^{-2} since the squared bias is $o(T^{-2})$. We have

$$\begin{aligned} MSE(\check{\alpha}) &= E[(\check{\alpha} - \alpha)(\check{\alpha} - \alpha)'] = E[(\hat{\alpha} - \hat{B}_1(\hat{\alpha}) - \alpha)(\hat{\alpha} - \hat{B}_1(\hat{\alpha}) - \alpha)'] \\ &= MSE(\hat{\alpha}) + E[\hat{B}_1(\hat{\alpha})\hat{B}_1(\hat{\alpha})'] - E[\hat{B}_1(\hat{\alpha})(\hat{\alpha} - \alpha)'] - E[(\hat{\alpha} - \alpha)\hat{B}_1(\hat{\alpha})']. \end{aligned} \quad (E.1)$$

Since $E(\hat{B}_1(\hat{\alpha}) - B_1(\hat{\alpha})) = 0 + o(T^{-1})$ with $B_1(\hat{\alpha}) = O(T^{-1})$ it is apparent that

$$E(\hat{B}_1(\hat{\alpha})\hat{B}_1(\hat{\alpha})') = B_1(\hat{\alpha})B_1(\hat{\alpha})' + o(T^{-2}). \quad (E.2)$$

From

$$\begin{aligned} E[\hat{B}_1(\hat{\alpha})(\hat{\alpha} - \alpha)'] &= E[B_1(\hat{\alpha})(\hat{\alpha} - \alpha)'] + E[(\hat{B}_1(\hat{\alpha}) - B_1(\hat{\alpha}))(\hat{\alpha} - \alpha)'] \\ &= B_1(\hat{\alpha})B_1(\hat{\alpha}) + E[(\hat{B}_1(\hat{\alpha}) - B_1(\hat{\alpha}))(\hat{\alpha} - \alpha)'] + o(T^{-2}) \end{aligned} \quad (E.3)$$

it follows that on substituting these results into (E.1) we may write

$$\begin{aligned} MSE(\check{\alpha}) &= MSE(\hat{\alpha}) - B_1(\hat{\alpha})B_1(\hat{\alpha})' \\ &\quad - E[(\hat{B}_1(\hat{\alpha}) - B_1(\hat{\alpha}))(\hat{\alpha} - \alpha)'] - E[(\hat{\alpha} - \alpha)(\hat{B}_1(\hat{\alpha}) - B_1(\hat{\alpha}))'] + o(T^{-2}) \\ &= V(\hat{\alpha}) - E[(\hat{B}_1(\hat{\alpha}) - B_1(\hat{\alpha}))(\hat{\alpha} - \alpha)'] - E[(\hat{\alpha} - \alpha)(\hat{B}_1(\hat{\alpha}) - B_1(\hat{\alpha}))'] + o(T^{-2}). \end{aligned} \quad (E.4)$$

An approximation for $V(\hat{\alpha})$ to order T^{-2} is given in Theorem 2.2. Hence, to establish an approximation to $MSE(\check{\alpha})$, i.e. to $V(\check{\alpha})$, we have to find an approximation to order T^{-2} of $E[(\hat{B}_1(\hat{\alpha}) - B_1(\hat{\alpha}))(\hat{\alpha} - \alpha)']$ and its transpose. Note that its two factors are $O_p(T^{-3/2})$ and $O_p(T^{-1/2})$ respectively, hence we only have to obtain the expectation of the product of their leading terms. For $(\hat{\alpha} - \alpha)$ these are $Q\bar{Z}'u + (u'Cu)q_1$, whereas the leading $O_p(T^{-3/2})$ terms of $(\hat{B}_1(\hat{\alpha}) - B_1(\hat{\alpha}))$ have already been obtained in Appendix E, notably in the formulas (D.16), (D.17) and (D.18). Gathering these and regrouping we obtain

$$\begin{aligned} &(\hat{B}_1(\hat{\alpha}) - B_1(\hat{\alpha}))(\hat{\alpha} - \alpha)' \\ &= -\sigma^2 \{ [\text{tr}(Q\bar{Z}'C\bar{Z})p_1^* + Q\bar{Z}'C\bar{Z}p_1^* + 2\sigma^2q_{11}\tau p_1^*](u'\bar{Z}Q + u'Cuq_1') \\ &\quad + [\text{tr}(QA^*)q_1 + QA^*q_1](u'\bar{Z}Q + u'Cuq_1') \\ &\quad + [\text{tr}(P^*\bar{Z}'C\bar{Z})q_1 + P^*\bar{Z}'C\bar{Z}q_1](u'\bar{Z}Q + u'Cuq_1') \\ &\quad + [\text{tr}(Q\bar{Z}'C\bar{Z})q_1 + Q\bar{Z}'C\bar{Z}q_1 + 4\sigma^2q_{11}\tau q_1]s_*^2(u'\bar{Z}Q + u'Cuq_1') \\ &\quad + 2\sigma^2[q_{11}q_1\hat{\tau}^*(u'\bar{Z}Q + u'Cuq_1') + \tau q_1p_{11}^*(u'\bar{Z}Q + u'Cuq_1')] \} + o_p(T^{-2}). \end{aligned} \quad (E.5)$$

To obtain the expectation of the latter expression, we first derive a few auxiliary results. Exploiting (D.3) we have

$$E[p_1^*(u'\bar{Z}Q + u'Cuq_1')] = -\sigma^2[q_{11}Q\bar{Z}'C\bar{Z}Q + q_1q_1'\bar{Z}'C\bar{Z}Q + 2\sigma^2q_{11}\text{tr}(CC'C)q_1q_1'],$$

from which it follows that

$$\mathbf{E}[p_{11}^*(u' \bar{Z}Q + u'Cuq'_1)] = -2\sigma^2[q_{11}q'_1 \bar{Z}'C \bar{Z}Q + \sigma^2 q_{11}^2 \text{tr}(CC'C)q'_1].$$

Using (D.11) we find

$$\begin{aligned} & \mathbf{E}\{[\text{tr}(QA^*)q_1 + QA^*q_1](\hat{\alpha} - \alpha)'\} \\ &= \sigma^2[q_1q'_1 \bar{Z}'(CC + 2C'C) \bar{Z}Q + q_{11}Q \bar{Z}'CC \bar{Z}Q + \text{tr}(Q \bar{Z}'CC \bar{Z})q_1q'_1 + Q \bar{Z}'CC \bar{Z}q_1q'_1] + o(T^{-2}). \end{aligned}$$

and again substituting (D.3)

$$\begin{aligned} & \mathbf{E}[\text{tr}(P^* \bar{Z}'C \bar{Z})q_1 + P^* \bar{Z}'C \bar{Z}q_1](u' \bar{Z}Q + u'Cuq'_1) \\ &= -\sigma^2\{q_1q'_1 \bar{Z}'(C + 2C') \bar{Z}Q \bar{Z}'C \bar{Z}Q + q'_1 \bar{Z}'C \bar{Z}q_1Q \bar{Z}'C \bar{Z}Q\} - 4\sigma^4q'_1 \bar{Z}'C \bar{Z}q_1 \text{tr}(CC'C)q_1q'_1. \end{aligned}$$

With (D.14) we find

$$\begin{aligned} \mathbf{E}[s_*^2(u' \bar{Z}Q + u'Cuq'_1)] &= \mathbf{E}[(s^2 - \sigma^2)(u' \bar{Z}Q + u'Cuq'_1)] + o(T^{-1}) \\ &= (T - K - 1)^{-1} \mathbf{E}\{[u'u - u'Z(Z'Z)^{-1}Z'u]u'Cuq'_1\} + o(T^{-1}) \\ &= T^{-1} \mathbf{E}(u'uu'Cu)q'_1 + o(T^{-1}) = o(T^{-1}), \end{aligned}$$

and employing (D.13) we obtain

$$\mathbf{E}[\hat{\tau}^*(u' \bar{Z}Q + u'Cuq'_{11})] = \sigma^2[\text{tr}(C'C'CC) + 2 \text{tr}(C'CCC)]q'_1 + o(1).$$

Taking the expectation of (E.5) by substitution of the above results yields

$$\begin{aligned} & \mathbf{E}[(\hat{\mathbf{B}}_1(\hat{\alpha}) - b\mathbf{B}_1(\hat{\alpha}))(\hat{\alpha} - \alpha)'] \tag{E.6} \\ &= \sigma^4\{\text{tr}(Q \bar{Z}'C \bar{Z})q_1q'_1 \bar{Z}'C \bar{Z}Q + Q \bar{Z}'C \bar{Z}q_1q'_1 \bar{Z}'C \bar{Z}Q \\ & \quad + q_1q'_1 \bar{Z}'(C + 2C') \bar{Z}Q \bar{Z}'C \bar{Z}Q + q'_1 \bar{Z}'C \bar{Z}q_1Q \bar{Z}'C \bar{Z}Q \\ & \quad - \text{tr}(Q \bar{Z}'CC \bar{Z})q_1q'_1 - Q \bar{Z}'CC \bar{Z}q_1q'_1 - q_1q'_1 \bar{Z}'(CC + 2C'C) \bar{Z}Q \\ & \quad + q_{11}[\text{tr}(Q \bar{Z}'C \bar{Z})Q \bar{Z}'C \bar{Z}Q + Q \bar{Z}'C \bar{Z}Q \bar{Z}'C \bar{Z}Q - Q \bar{Z}'CC \bar{Z}Q]\} \\ & \quad + 2\sigma^6\{2q'_1 \bar{Z}'C \bar{Z}q_1 \text{tr}(CC'C)q_1q'_1 \\ & \quad + q_{11} \text{tr}(CC'C)[\text{tr}(Q \bar{Z}'C \bar{Z})q_1q'_1 + Q \bar{Z}'C \bar{Z}q_1q'_1 + 3q_1q'_1 \bar{Z}'C \bar{Z}Q] \\ & \quad - q_{11}[\text{tr}(C'C'CC) + 2 \text{tr}(C'CCC)]q_1q'_1 + q_{11}^2 \text{tr}(CC'C)Q \bar{Z}'C \bar{Z}Q\} \\ & \quad + 8\sigma^8q_{11}^2[\text{tr}(CC'C)]^2q_1q'_1 + o(T^{-2}). \end{aligned}$$

Finally, we substitute the results of Theorem 2.2 and (E.6) in (E.4). Exploiting the equivalence regarding their leading order T terms, as proved in Kiviet and Phillips (2012, Appendix C), of respectively $\text{tr}(C'C'CC)$ and $\text{tr}(CC'C'C)$ and of $\text{tr}(C'CCC)$ and $\text{tr}(CC'CC')$ yields the required approximation to $\text{MSE}(\tilde{\alpha})$ and, hence, to the variance $\text{V}(\tilde{\alpha})$ as stated in Theorem 3.2.

F. Estimating the variance of the COLS estimator

Combining Theorems 2.3 and 3.2 we may show that

$$\begin{aligned}
& \hat{V}(\hat{\alpha}) \\
& + \sigma^4 \{ [\text{tr}(Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}) + 2(1 - \lambda^2)^{-1}] q_1 q_1' \\
& \quad + Q\bar{Z}'C\bar{Z}q_1 q_1' \bar{Z}'C'\bar{Z}Q + Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}q_1 q_1' + q_1 q_1' \bar{Z}'C'\bar{Z}Q\bar{Z}'C'\bar{Z}Q \} \\
& + 2\sigma^6 \{ 2q_1' \bar{Z}'C\bar{Z}q_1 \text{tr}(CC'C) q_1 q_1' + q_{11} \text{tr}(CC'C) [Q\bar{Z}'C\bar{Z}q_1 q_1' + q_1 q_1' \bar{Z}'C'\bar{Z}Q] \} \\
& + 4\sigma^8 q_{11}^2 [\text{tr}(CC'C)]^2 q_1 q_1'
\end{aligned} \tag{F.1}$$

is unbiased for $V(\tilde{\alpha})$ to order T^{-2} . However, this is not an estimator because the terms in σ^4 , σ^6 and σ^8 , which are $O(T^{-2})$, are unknown. It follows that if these unknown terms are replaced with estimates which have the same expected value to order T^{-2} , the resulting estimator will also be unbiased to order T^{-2} . Using the results of Appendix D, we find that we may replace Q with P , $\bar{Z}'C\bar{Z}$ with $\hat{Z}'\hat{C}\hat{Z}$, and σ^4 , σ^6 and σ^8 with s^4 , s^6 and s^8 respectively, $\text{tr}(CC'C)$ with $\text{tr}(\hat{C}\hat{C}'\hat{C})$, and λ with $\hat{\lambda}$ such that the resulting expression, given in Theorem 3.3, will have the same expectation to order T^{-2} .

G. Special results for the AR(1) model

Taking $\bar{Z} = (y_0^* F, \iota)$ we can obtain

$$\begin{aligned}
q_{11} &= (1 - \lambda^2)T^{-1} + [1 - (1 - \lambda^2)y_0^{*2}]T^{-2} + o(T^{-2}) \\
q_{12} &= -y_0^*(1 + \lambda)T^{-2} + o(T^{-2}) \\
q_{22} &= T^{-1} + o(T^{-2}),
\end{aligned} \tag{G.1}$$

and moreover

$$\begin{aligned}
\text{tr}(Q\bar{Z}'C\bar{Z}) &= (1 - \lambda)^{-1} + O(T^{-1}) \\
q_1' \bar{Z}'C\bar{Z}q_1 &= y_0^{*2}\lambda T^{-2} + o(T^{-2}) = O(T^{-2}).
\end{aligned} \tag{G.2}$$

Some of these results show orders smaller than expected, due to the typical nature of the first column of \bar{Z} , which has the effect that only one element of $\bar{Z}'\bar{Z}$ is $O(T)$ while the other three are $O(1)$.

Result (4.3) follows using $\text{tr}(CC'C) = \lambda(1 - \lambda^2)^{-2}T + O(1)$, which is proved in Kiviet and Phillips (2012, formula C.8). Kiviet and Phillips (2012) also gives

$$\begin{aligned}
\text{tr}(CC'CC') &= (1 + \lambda^2)(1 - \lambda^2)^{-3}T + O(1) \\
\text{tr}(CC'CC) &= \lambda^2(1 - \lambda^2)^{-3}T + O(1) \\
\text{tr}(CC'C'C) &= (1 + \lambda^2)(1 - \lambda^2)^{-3}T + O(1)
\end{aligned} \tag{G.3}$$

and in this special model we further have

$$\begin{aligned}
q_1' \bar{Z}' C C' \bar{Z} q_1 &= o(T^{-1}) & \text{tr}(Q \bar{Z}' C C' \bar{Z}) &= (1 - \lambda)^{-2} + o(1) \\
q_1' \bar{Z}' C C \bar{Z} q_1 &= o(T^{-1}) & \text{tr}(Q \bar{Z}' C C \bar{Z}) &= (1 - \lambda)^{-2} + o(1) \\
q_1' \bar{Z}' C' C \bar{Z} q_1 &= o(T^{-1}) & \text{tr}(Q \bar{Z}' C \bar{Z} Q \bar{Z}' C \bar{Z}) &= (1 - \lambda)^{-2} + o(1) \\
q_1' \bar{Z}' C \bar{Z} Q \bar{Z}' C \bar{Z} q_1 &= o(T^{-2}) & q_1 \bar{Z}' C' \bar{Z} Q \bar{Z}' C \bar{Z} q_1 &= o(T^{-2}).
\end{aligned} \tag{G.4}$$

Substituting the above in Corollary 2.2 yields

$$\mathbf{V}(\hat{\lambda}) = (1 - \lambda^2)T^{-1} - [(1 - \lambda^2)y_0^{*2}]T^{-2} - (1 - 4\lambda - 14\lambda^2)T^{-2} + o(T^{-2}),$$

given in (4.4), and adding $[(1+3\lambda)T^{-1}]^2$ yields (4.5). Evaluating Corollary 2.3 upon using $\text{tr}(C'C) = T(1 - \lambda^2)^{-1} + O(1)$ gives (4.7) from which (4.8) and (4.9) straightforwardly follow. Evaluation of Corollary 3.2 produces (4.11) and then it is easily established that the roots of $5 - 6\lambda - 15\lambda^2 = 0$, which are 0.4110101 and 0.8110101, determine the sign of $\text{MSE}(\hat{\lambda}) - \text{MSE}(\check{\lambda})$ as stated in Theorem 4.1.

In Kiviet and Phillips (2012) it has been derived that

$$\mathbf{E}(\hat{\lambda}) = \lambda - \frac{1}{T}(1 + 3\lambda) - \frac{1}{T^2} \left(\frac{1 - 3\lambda + 9\lambda^2}{1 - \lambda} \right) + o(T^{-2}). \tag{G.5}$$

This implies

$$\mathbf{E}(\check{\lambda} - \lambda) = -\frac{1}{T^2} \left(\frac{4 + 3\lambda}{1 - \lambda} \right) + o(T^{-2}) \tag{G.6}$$

and

$$\mathbf{E}(\dot{\lambda} - \lambda) = -\frac{1}{T^2} \left(\frac{1 + 6\lambda}{1 - \lambda} \right) + o(T^{-2}). \tag{G.7}$$

Because $1 + 6\lambda < 4 + 3\lambda$ for any $|\lambda| < 1$ estimator $\dot{\lambda}$ has smaller second order bias. However, since $\mathbf{V}(\check{\lambda}) = [(T + 3)/T]^2 \mathbf{V}(\hat{\lambda})$ and $\mathbf{V}(\dot{\lambda}) = [T/(T - 3)]^2 \mathbf{V}(\hat{\lambda})$, $\text{MSE}(\check{\lambda}) < \text{MSE}(\dot{\lambda})$ follows from $(T + 3)/T - T/(T - 3) = -9/[T(T - 3)] < 0$ for $T > 3$. Note that from (G.6) and (G.7) a higher-order bias correction is immediately available.

Table 1: Mean-stationary AR(1) model with unknown intercept

	λ	$B(\hat{\lambda})$	$V(\hat{\lambda})$	$\frac{E[\hat{V}(\hat{\lambda})]}{V(\hat{\lambda})}$	$\frac{B_1(\hat{\lambda})}{B(\hat{\lambda})}$	$\frac{V_1(\hat{\lambda})}{V(\hat{\lambda})}$	$\frac{V_2(\hat{\lambda})}{V(\hat{\lambda})}$	$\frac{E[\hat{V}_2(\hat{\lambda})]}{V(\hat{\lambda})}$	$\frac{E[\check{V}(\hat{\lambda})]}{V(\hat{\lambda})}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$T = 20$	-0.9	0.067	0.023	0.89	1.27	0.42	1.16	1.25	1.09
	-0.6	0.030	0.035	1.12	1.34	0.91	1.03	1.04	1.05
	-0.3	-0.011	0.043	1.18	0.45	1.05	0.99	0.97	1.04
	0	-0.053	0.048	1.18	0.94	1.05	1.00	0.97	1.04
	0.2	-0.082	0.048	1.15	0.97	1.00	1.02	0.99	1.04
	0.4	-0.113	0.047	1.11	0.97	0.90	1.05	1.04	1.04
	0.6	-0.148	0.044	1.03	0.95	0.73	1.10	1.10	1.04
	0.8	-0.193	0.040	0.91	0.88	0.46	1.16	1.20	1.03
	0.9	-0.221	0.036	0.82	0.84	0.26	1.22	1.28	1.02
	0.99	-0.235	0.033	0.68	0.84	0.03	1.28	1.37	0.98
$T = 50$	-0.9	0.031	0.006	0.89	1.09	0.61	1.05	1.17	1.03
	-0.6	0.015	0.013	1.04	1.06	0.96	1.01	1.02	1.01
	-0.3	-0.002	0.018	1.07	0.92	1.02	1.00	0.99	1.01
	0	-0.020	0.020	1.07	1.01	1.02	1.00	0.99	1.01
	0.2	-0.032	0.019	1.06	1.00	1.00	1.01	1.00	1.01
	0.4	-0.044	0.018	1.03	0.99	0.95	1.02	1.02	1.01
	0.6	-0.058	0.015	0.99	0.97	0.85	1.02	1.07	1.01
	0.8	-0.074	0.011	0.89	0.92	0.63	1.03	1.15	1.00
	0.9	-0.087	0.009	0.79	0.85	0.41	1.02	1.22	0.99
	0.99	-0.103	0.007	0.63	0.77	0.06	1.01	1.38	0.97

$B(\hat{\lambda})$: Monte Carlo estimate of the actual bias of the ML estimator $\hat{\lambda}$

$V(\hat{\lambda})$: Monte Carlo estimate of the actual variance of the ML estimator $\hat{\lambda}$

$\hat{V}(\hat{\lambda})$: standard estimator (4.6) of $V(\hat{\lambda})$

$B_1(\hat{\lambda})$: first-order approximation to $B(\hat{\lambda})$, $-(1 + 3\lambda)/T$

$V_1(\hat{\lambda})$: first-order approximation to $V(\hat{\lambda})$, $(1 - \lambda^2)/T$

$V_2(\hat{\lambda})$: second-order approximation (5.1) to $V(\hat{\lambda})$

$\hat{V}_2(\hat{\lambda})$: $V_2(\hat{\lambda})$ evaluated in $\hat{\lambda}$

$\check{V}(\hat{\lambda})$: bias corrected estimator (4.9) of $V(\hat{\lambda})$

Table 2: Mean-stationary AR(1) model with unknown intercept

	λ	$ \hat{\lambda} \geq 1$	$B(\check{\lambda})$	$\frac{B(\check{\lambda})}{B(\hat{\lambda})}$	$\frac{E[\check{V}(\check{\lambda})]}{V(\hat{\lambda})}$	$\frac{E[\check{V}^*(\check{\lambda})]}{V(\hat{\lambda})}$	$\frac{MSE(\check{\lambda})}{MSE(\hat{\lambda})}$	$ \check{\lambda} \geq 1$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$T = 20$	-0.9	0.063	-.008	-0.12	0.94	0.96	1.11	0.346
	-0.6	0.001	-.006	-0.19	0.98	0.99	1.29	0.006
	-0.3	0.000	-.008	0.70	1.02	1.03	1.32	0.000
	0	0.000	-.011	0.21	1.02	1.02	1.25	0.000
	0.2	0.000	-.015	0.18	0.99	1.00	1.16	0.000
	0.4	0.000	-.020	0.18	0.95	0.97	1.05	0.001
	0.6	0.000	-.030	0.21	0.89	0.92	0.90	0.013
	0.8	0.003	-.052	0.27	0.83	0.87	0.72	0.113
	0.9	0.012	-.069	0.31	0.82	0.87	0.62	0.229
	0.99	0.044	-.072	0.31	0.85	0.91	0.56	0.400
$T = 50$	-0.9	0.003	-.001	-0.03	0.97	0.98	0.97	0.054
	-0.6	0.000	-.000	-0.00	1.00	1.00	1.11	0.000
	-0.3	0.000	-.000	0.14	1.01	1.01	1.12	0.000
	0	0.000	-.001	0.05	1.01	1.01	1.10	0.000
	0.2	0.000	-.002	0.06	1.00	1.01	1.07	0.000
	0.4	0.000	-.003	0.07	0.99	1.00	1.01	0.000
	0.6	0.000	-.005	0.09	0.98	0.98	0.92	0.000
	0.8	0.000	-.010	0.14	0.93	0.95	0.77	0.004
	0.9	0.001	-.018	0.21	0.91	0.92	0.64	0.084
	0.99	0.029	-.030	0.29	0.93	0.96	0.50	0.380

$\check{\lambda}$: the bias corrected estimator $\hat{\lambda} + (1 + 3\hat{\lambda})/T$

$B(\check{\lambda})$: Monte Carlo estimate of the actual bias of $\check{\lambda}$

$B(\hat{\lambda})$: Monte Carlo estimate of the actual bias of $\hat{\lambda}$

$V(\check{\lambda})$: Monte Carlo estimate of the actual variance of $\check{\lambda}$

$\check{V}(\check{\lambda})$: the bias corrected estimator (4.13) of $V(\check{\lambda})$

$\check{V}^*(\check{\lambda})$: bias corrected estimator of $V(\check{\lambda})$ given by $\hat{V}(\hat{\lambda}) + (3 + 2\check{\lambda} + 3\check{\lambda}^2)/T^2$

Table 2A: Effects of alternative definitions of $\check{\lambda}$ in AR(1) model

		$ \hat{\lambda} \geq 1$ no correction			$ \check{\lambda} \geq 1$ no correction		
	λ	$\frac{B(\check{\lambda})}{B(\hat{\lambda})}$	$\frac{E[\check{V}^*(\check{\lambda})]}{V(\hat{\lambda})}$	$\frac{MSE(\check{\lambda})}{MSE(\hat{\lambda})}$	$\frac{B(\check{\lambda})}{B(\hat{\lambda})}$	$\frac{E[\check{V}^*(\check{\lambda})]}{V(\hat{\lambda})}$	$\frac{MSE(\check{\lambda})}{MSE(\hat{\lambda})}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$T = 20$	-0.9	-0.02	0.96	1.02	0.37	1.16	0.82
	-0.6	-0.19	0.99	1.29	-0.17	1.00	1.28
	-0.3	0.70	1.03	1.32	0.70	1.02	1.32
	0	0.21	1.02	1.25	0.21	1.02	1.25
	0.2	0.18	1.00	1.16	0.18	0.99	1.16
	0.4	0.18	0.97	1.05	0.18	0.96	1.04
	0.6	0.21	0.92	0.90	0.22	0.96	0.87
	0.8	0.27	0.87	0.71	0.38	1.07	0.62
	0.9	0.32	0.87	0.61	0.50	1.13	0.53
	0.99	0.35	0.91	0.53	0.63	1.10	0.52
$T = 50$	-0.9	-0.03	0.98	0.97	0.04	1.03	0.92
	-0.6	-0.00	1.00	1.11	-0.00	1.00	1.11
	-0.3	0.14	1.01	1.12	0.14	1.01	1.12
	0	0.05	1.01	1.10	0.05	1.01	1.10
	0.2	0.06	1.01	1.07	0.06	1.00	1.07
	0.4	0.07	1.00	1.01	0.07	0.99	1.01
	0.6	0.09	0.98	0.92	0.09	0.98	0.92
	0.8	0.14	0.95	0.77	0.14	0.95	0.76
	0.9	0.21	0.93	0.64	0.28	1.03	0.57
	0.99	0.31	0.97	0.48	0.58	1.16	0.46

Table 2B: Effects of alternative bias correction in AR(1)

	λ	$B(\dot{\lambda})$	$\frac{B(\dot{\lambda})}{B(\check{\lambda})}$	$\frac{MSE(\dot{\lambda})}{MSE(\check{\lambda})}$	$ \dot{\lambda} \geq 1$
	(1)	(2)	(3)	(4)	(5)
$T = 20$	-0.9	-0.005	0.61	0.98	0.422
	-0.6	-0.014	2.32	1.05	0.011
	-0.3	-0.007	0.91	1.05	0.000
	0.0	-0.003	0.29	1.05	0.000
	0.2	-0.002	0.16	1.04	0.000
	0.4	-0.004	0.19	1.04	0.002
	0.6	-0.011	0.36	1.02	0.027
	0.8	-0.034	0.66	0.96	0.173
	0.9	-0.055	0.80	0.91	0.311
	0.99	-0.063	0.88	0.89	0.489
$T = 50$	-0.9	0.002	-1.94	1.00	0.103
	-0.6	-0.000	2.92	1.01	0.000
	-0.3	-0.000	0.85	1.01	0.000
	0.0	-0.000	0.16	1.01	0.000
	0.2	-0.000	0.10	1.01	0.000
	0.4	-0.001	0.19	1.01	0.000
	0.6	-0.002	0.38	1.01	0.000
	0.8	-0.006	0.63	1.00	0.009
	0.9	-0.015	0.82	0.98	0.132
	0.99	-0.029	0.95	0.95	0.464
$\dot{\lambda}$: estimator $(T\hat{\lambda} + 1)/(T - 3)$ of (4.14) $\check{\lambda}$: estimator $[(T + 3)\hat{\lambda} + 1]/T$ of (4.10)					

Table 3: ARX(1) model with intercept and trend (no correction if $|\check{\lambda}| \geq 1$)

	λ	$B(\hat{\lambda})$	$V(\hat{\lambda})$	$\frac{E[\hat{V}(\hat{\lambda})]}{V(\hat{\lambda})}$	$\frac{E[\check{V}(\hat{\lambda})]}{V(\hat{\lambda})}$	$\frac{E[\check{V}(\check{\lambda})]}{V(\check{\lambda})}$	$\frac{MSE(\check{\lambda})}{MSE(\hat{\lambda})}$	$ \hat{\lambda} \geq 1$	$ \check{\lambda} \geq 1$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$T = 20$	-0.9	0.057	0.021	0.98	1.94	1.22	0.92	0.070	0.340
	-0.6	0.007	0.033	1.22	1.28	1.12	1.41	0.001	0.009
	-0.3	-0.048	0.041	1.27	1.04	1.09	1.34	0.000	0.000
	0.0	-0.105	0.047	1.26	1.00	1.04	1.16	0.000	0.000
	0.2	-0.146	0.048	1.21	1.04	1.00	1.04	0.000	0.000
	0.4	-0.191	0.049	1.15	1.11	0.95	0.89	0.000	0.003
	0.6	-0.244	0.048	1.06	1.18	0.94	0.73	0.000	0.020
	0.8	-0.310	0.047	0.93	1.18	0.94	0.62	0.001	0.066
	0.9	-0.355	0.046	0.86	1.09	0.92	0.63	0.003	0.088
	0.99	-0.417	0.047	0.80	0.98	0.88	0.66	0.007	0.092
$T = 50$	-0.9	0.029	0.006	0.93	1.21	1.07	0.98	0.003	0.057
	-0.6	0.007	0.013	1.08	1.04	1.02	1.16	0.000	0.000
	-0.3	-0.016	0.0017	1.10	1.01	1.02	1.14	0.000	0.000
	0.0	-0.040	0.020	1.09	1.00	1.01	1.08	0.000	0.000
	0.2	-0.057	0.020	1.07	1.01	1.00	1.01	0.000	0.000
	0.4	-0.074	0.018	1.03	1.02	0.98	0.92	0.000	0.000
	0.6	-0.094	0.016	0.97	1.04	0.95	0.79	0.000	0.000
	0.8	-0.120	0.013	0.85	1.04	0.91	0.62	0.000	0.006
	0.9	-0.141	0.012	0.76	1.00	0.91	0.53	0.000	0.047
	0.99	-0.180	0.011	0.65	0.87	0.89	0.54	0.005	0.106

Table 4: ARX(1) model with AR(1) regressor (no correction if $|\check{\lambda}| \geq 1$)

	λ	$B(\hat{\lambda})$	$V(\hat{\lambda})$	$\frac{E[\hat{V}(\hat{\lambda})]}{V(\hat{\lambda})}$	$\frac{E[\check{V}(\hat{\lambda})]}{V(\hat{\lambda})}$	$\frac{E[\check{V}(\check{\lambda})]}{V(\check{\lambda})}$	$\frac{MSE(\check{\lambda})}{MSE(\hat{\lambda})}$	$ \hat{\lambda} \geq 1$	$ \check{\lambda} \geq 1$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$T = 20$	0.0	-0.018	0.013	1.06	1.00	1.01	1.09	0.000	0.000
	0.2	-0.032	0.015	1.08	1.00	1.01	1.07	0.000	0.000
	0.4	-0.050	0.016	1.06	1.01	1.01	0.99	0.000	0.000
	0.6	-0.075	0.018	0.94	0.99	0.97	0.85	0.000	0.001
	0.8	-0.118	0.026	0.70	0.84	0.87	0.67	0.003	0.027
	0.9	-0.171	0.037	0.58	0.73	0.79	0.63	0.026	0.109
	0.95	-0.236	0.045	0.59	0.72	0.78	0.65	0.047	0.156
$T = 50$	0.0	-0.009	0.005	1.03	1.00	1.01	1.03	0.000	0.000
	0.2	-0.014	0.006	1.03	1.00	1.00	1.02	0.000	0.000
	0.4	-0.020	0.005	1.03	1.01	1.00	0.99	0.000	0.000
	0.6	-0.029	0.005	0.99	1.01	1.00	0.91	0.000	0.000
	0.8	-0.044	0.005	0.84	0.96	0.95	0.72	0.000	0.000
	0.9	-0.065	0.006	0.66	0.84	0.88	0.58	0.001	0.014
	0.95	-0.099	0.009	0.58	0.76	0.86	0.53	0.014	0.096
$T = 100$	0.0	-0.004	0.003	1.01	1.00	1.01	1.01	0.000	0.000
	0.2	-0.006	0.003	1.02	1.01	1.00	1.01	0.000	0.000
	0.4	-0.010	0.003	1.02	1.01	1.00	1.00	0.000	0.000
	0.6	-0.016	0.003	1.00	1.01	1.00	0.95	0.000	0.000
	0.8	-0.024	0.002	0.92	1.00	0.99	0.81	0.000	0.000
	0.9	-0.034	0.002	0.79	0.95	0.95	0.67	0.000	0.000
	0.95	-0.044	0.003	0.67	0.88	0.92	0.56	0.001	0.033

Table 5: Test size in mean-stationary AR(1) model with intercept and trend

		left-hand alternative, corrected:						right-hand alternative, corrected:					
		always				if $ \check{\lambda} < 1$		always				if $ \check{\lambda} < 1$	
	λ_0	t	t^*	t_*	t_*^*	t^*	t_*^*	t	t^*	t_*	t_*^*	t^*	t_*^*
$T = 20$	-0.9	0.03	0.14	0.02	0.03	0.03	0.03	0.04	0.04	0.05	0.03	0.04	0.03
	-0.6	0.04	0.09	0.04	0.05	0.09	0.05	0.02	0.04	0.04	0.03	0.04	0.03
	-0.3	0.06	0.08	0.07	0.06	0.08	0.06	0.02	0.04	0.03	0.03	0.04	0.03
	0.0	0.09	0.08	0.13	0.07	0.08	0.07	0.01	0.04	0.01	0.03	0.04	0.03
	0.2	0.11	0.08	0.16	0.07	0.08	0.07	0.01	0.05	0.00	0.02	0.05	0.02
	0.4	0.15	0.10	0.19	0.09	0.10	0.09	0.00	0.05	0.00	0.02	0.05	0.02
	0.6	0.21	0.12	0.23	0.11	0.12	0.11	0.00	0.04	0.00	0.01	0.02	0.01
	0.8	0.34	0.18	0.34	0.16	0.18	0.16	0.00	0.02	0.00	0.00	0.00	0.00
	0.9	0.47	0.25	0.44	0.22	0.25	0.22	0.00	0.01	0.00	0.00	0.00	0.00
	0.99	0.67	0.37	0.60	0.32	0.37	0.33	0.00	0.00	0.00	0.00	0.00	0.00
$T = 50$	-0.9	0.03	0.11	0.01	0.05	0.08	0.04	0.06	0.04	0.06	0.03	0.04	0.03
	-0.6	0.05	0.07	0.05	0.06	0.07	0.06	0.04	0.04	0.04	0.04	0.04	0.04
	-0.3	0.06	0.06	0.07	0.05	0.06	0.05	0.03	0.05	0.03	0.04	0.05	0.04
	0.0	0.08	0.06	0.09	0.06	0.06	0.06	0.02	0.05	0.02	0.04	0.05	0.04
	0.2	0.09	0.06	0.10	0.06	0.06	0.06	0.02	0.06	0.02	0.04	0.06	0.04
	0.4	0.11	0.06	0.12	0.06	0.06	0.06	0.01	0.06	0.01	0.04	0.06	0.04
	0.6	0.15	0.07	0.15	0.06	0.07	0.06	0.01	0.06	0.00	0.04	0.06	0.04
	0.8	0.23	0.10	0.21	0.08	0.10	0.08	0.00	0.05	0.00	0.02	0.05	0.02
	0.9	0.36	0.14	0.30	0.12	0.14	0.12	0.00	0.03	0.00	0.01	0.00	0.00
	0.99	0.68	0.29	0.55	0.24	0.29	0.24	0.00	0.00	0.00	0.00	0.00	0.00

$$t \equiv (\hat{\lambda} - \lambda_0)/[\hat{V}(\hat{\lambda})]^{1/2}, t^* \equiv (\check{\lambda} - \lambda_0)/[\hat{V}(\hat{\lambda})]^{1/2}, t_* \equiv (\hat{\lambda} - \lambda_0)/[\check{V}(\hat{\lambda})]^{1/2}, t_*^* \equiv (\check{\lambda} - \lambda_0)/[\check{V}(\check{\lambda})]^{1/2}$$

Table 6: Test size (*italics*) and size-corrected power in mean-stationary AR(1) with intercept, $H_0 : \lambda = \lambda_0$, λ is true value, $T = 50$

λ_0	λ	left-hand alternative:				right-hand alternative:			
		t	t^*	t_*	t_*^*	t	t^*	t_*	t_*^*
0.0	0.0	<i>0.06</i>	<i>0.06</i>	<i>0.07</i>	<i>0.05</i>	<i>0.03</i>	<i>0.05</i>	<i>0.03</i>	<i>0.04</i>
	0.2	0.00	0.00	0.00	0.00	0.40	0.40	0.40	0.40
	0.4	0.00	0.00	0.00	0.00	0.86	0.86	0.86	0.86
0.4	0.4	<i>0.08</i>	<i>0.06</i>	<i>0.09</i>	<i>0.05</i>	<i>0.02</i>	<i>0.06</i>	<i>0.02</i>	<i>0.05</i>
	0.2	0.37	0.37	0.37	0.37	0.00	0.00	0.00	0.00
	0.6	0.00	0.00	0.00	0.00	0.48	0.48	0.48	0.48
0.8	0.8	<i>0.13</i>	<i>0.06</i>	<i>0.12</i>	<i>0.05</i>	<i>0.01</i>	<i>0.08</i>	<i>0.00</i>	<i>0.04</i>
	0.6	0.42	0.42	0.42	0.42	0.00	0.00	0.00	0.00
	0.99	0.00	0.00	0.00	0.00	0.68	0.69	0.67	0.68
0.99	0.99	<i>0.39</i>	<i>0.10</i>	<i>0.25</i>	<i>0.07</i>	<i>0.00</i>	<i>0.09</i>	<i>0.00</i>	<i>0.01</i>
	0.6	0.89	0.93	0.92	0.94	0.00	0.00	0.00	0.00
	0.8	0.32	0.38	0.37	0.40	0.00	0.00	0.00	0.00