Improved coefficient and variance estimation in stable first-order dynamic regression models

Kiviet, J.F.; Phillips, G.D.A.

UvA ECONOMETRICS

Discussion Paper: 2002/02
revised: June 2003

Improved coefficient and variance estimation in stable first-order dynamic regression models

Jan F. Kiviet and Garry D.A. Phillips

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

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

UvA UNIVERSITEIT VAN AMSTERDAM

Improved Coefficient and Variance Estimation in Stable First-Order Dynamic Regression Models*

Jan F. Kiviet† and Garry D.A. Phillips‡
(University of Amsterdam) (Cardiff University)

30 June 2003
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 the least-squares coefficient estimators are biased in finite samples, and so are the usual estimators for the disturbance variance and for the variance of the coefficient estimators. By deriving the expectation of the sum of the initial terms in an expansion of the usual expression for the coefficient variance estimator and by comparing this with an approximation to the true variance we find an approximation to the bias in variance estimation from which a bias corrected estimator for the variance readily follows. This is also achieved for a bias corrected coefficient estimator which enables one to compare analytically the second-order approximation to the mean squared error of the ordinary least-squares estimator and its counterpart after bias correcting the coefficient estimator to first order. Illustrative numerical and simulation results on the magnitude of bias in coefficient and variance estimation and on the options for bias reduction are presented for three particularly relevant cases of the ARX(1) class of models. These show that substantial efficiency gains and test size improvements can easily be realized.

1. Introduction

In dynamic regression models it is well known that the least squares estimators of the regression coefficients and of the disturbance variance, and consequently the coefficient variance estimator, are 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 by the degrees of freedom reduces the order of the bias by a factor of T, where T is the sample size. In fact, the bias of such an estimator can be reduced even further. Kiviet and Phillips (1998b) show that this bias can be reduced by another factor T when employing a

*A very early version of this paper has been presented at the Econometric Society World Meeting of 2000 in Seattle. The first author wants to thank the universities of Illinois (Champaign-Urbana, USA) and Melbourne (Victoria, Australia) where he enjoyed hospitality when elaborating this project.

† Tinbergen Institute and Faculty of Economics & Econometrics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands (J.F.Kiviet@UvA.NL)

‡ Cardiff Business School, Cardiff University, Colum Drive, Cardiff CF10 3EU, Wales UK (PhillipsGD1@cardiff.ac.uk)

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 coefficients. In this paper we have a closer look at the second moment of the least-squares estimator for the full vector of coefficients. In addition, we also examine the variance and mean squared error of a bias corrected estimator of the coefficients. From the results various conclusions can be drawn on the effectiveness of bias correction and on appropriate variance estimation of (bias corrected) least-squares 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 (non-)stationary 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} or even smaller, where T is the sample size. 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. The approximation of the moments of statistical estimators in stable autoregressive models by use of asymptotic expansions has already been undertaken for 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) and White (1961). In the latter study, which focuses on the AR(1) model with no (or a known) intercept, an analysis is also given of the bias in the variance estimator of the coefficients, but generally speaking very little work has been done to find out how well the usual standard deviation estimator estimates the true standard errors in linear dynamic econometric models.

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 and any form of pre-sample initial condition of the dependent variable of this dynamic system. As in Kiviet and Phillips (1998b)[1] the focus of attention here is the bias of ordinary least-squares (OLS) estimation (which is also maximum likelihood conditional on y_0 and X) of all the regression coefficients in the first-order normal linear dynamic regression model

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

where $y = (y_1, ..., y_T)'$ is a $T \times 1$ vector of observations on a dependent variable, y_{-1} is the y vector lagged one period, i.e. $y_{-1} = (y_0, ..., y_{T-1})'$, 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. Below we shall give further attention to the precise assumptions made on the initial conditions, i.e. concerning y_0.

We first focus on an examination of the finite sample bias of the usual estimator of the (asymptotic) variance of the OLS estimator $\hat{\alpha}$ of the full coefficient vector $\alpha = (\lambda, \beta')'$, and we shall develop a bias corrected variance estimator. We shall also consider a bias corrected estimator $\check{\alpha}$ of α and examine its relative efficiency, both analytically and experimentally, in simulations. Rewriting (1.1) as

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

[1]Henceforth we refer to the present authors by KP. For related work in stable ARX models see also KP (1993, 1994, 1998a) and Kiviet et al. (1995). We have analyzed the finite sample characteristics of the first two moments of the least-squares coefficient estimators in dynamic models with a unit root in KP 2003.

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

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

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

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

where

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

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 the derivation of moments such as $\mathsf{E}(\hat{\alpha})$, $\mathsf{V}(\hat{\alpha})$ and $\mathsf{E}[\hat{\mathsf{V}}(\hat{\alpha})]$ is non-trivial, because Z is stochastic and depends linearly on u, whereas $\hat{\alpha}$ depends nonlinearly on Z, so these are moments of 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 some basic results which are collected in Appendix A. 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 have an explicit representation for the true variance (but only a higher-order asymptotic approximation), these results can be used to develop a bias correction to the standard asymptotic variance estimator. 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. Some remarkably simple analytic results on the scope for bias reduction and efficiency gains are obtained. In Section 5 we verify the numerical magnitude of the bias of alternative coefficient and variance estimators and their effects on test size by Monte Carlo simulation for a range of particular cases. Finally, in Section 6, we summarize our main conclusions. Proofs can be found in a series of Appendices.

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) stationarity, i.e. the matrix $Z = (y_{-1}, X)$ is such that $Z'Z = O_p(T)$; (iii) the $T \times (K+1)$ matrix Z has $\mathrm{rank}(Z) = K + 1$ with probability one; (iv) the regressors in X are strongly exogenous; (v) the disturbances follow $u \sim \mathsf{N}(0, \sigma^2 I_T)$, with $0 < \sigma^2 < \infty$; (vi) the start-up value has $y_0 \sim \mathsf{N}(\bar{y}_0, \omega^2\sigma^2)$, with $0 \leq \omega < \infty$; (vii) y_0 and u are mutually independent.*

Note that $\omega = 0$ represents the fixed start-up case. For any $\omega \neq 0$ the start-up is random, and if $\omega^2 = (1 - \lambda^2)^{-1}$ then $\{y_t\}$ has a constant variance. Also note that (ii) excludes a linear trend or any $I(1)$ regressors. However, the presence of such variables will not change our approximation formulas as such, as is shown in KP (1998a), but will only render them more accurate, because it reduces their order of magnitude and also the order of the remainder terms.

In order to distinguish the fixed 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 conditional on X and $\bar{y}_0$, i.e.

$$\begin{aligned} \bar{Z} &= \mathsf{E}(Z) = [\mathsf{E}(y_{-1}), X] = (\bar{y}_{-1}, X) && (2.1) \\ \tilde{Z} &= Z - \bar{Z} = (y_{-1} - \bar{y}_{-1}, X - X) = (\tilde{y}_{-1}, O) = \tilde{y}_{-1} e_1', && (2.2) \end{aligned}$$

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

$$\bar{y}_{-1} = \begin{pmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ \lambda & 1 & \cdot & & & \cdot \\ \lambda^2 & \lambda & 1 & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & & & \cdot & \cdot & 0 \\ \lambda^{T-1} & \cdot & \cdot & \cdot & \lambda & 1 \end{pmatrix} \begin{pmatrix} \bar{y}_0 \\ x_1'\beta \\ x_2'\beta \\ \cdot \\ \cdot \\ x_{T-1}'\beta \end{pmatrix}, \qquad (2.3)$$

where $X' = (x_1, ..., x_T)$, and hence we find that $\bar{Z}$ is determined by: X, $\bar{y}_0$, β and λ. Defining $v \equiv (u_0, u_1, ..., u_T)'$ such that

$$v \sim \mathsf{N}(0, \sigma^2 I_{T+1}) \text{ with } y_0 = \bar{y}_0 + \omega u_0, \qquad (2.4)$$

and introducing the $T \times (T + 1)$ matrix G such that

$$G = (\omega F, C), \text{ with } F = \begin{pmatrix} 1 \\ \lambda \\ \lambda^2 \\ \cdot \\ \cdot \\ \lambda^{T-1} \end{pmatrix}, \; C = \begin{pmatrix} 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 1 & 0 & & & & \cdot \\ \lambda & 1 & 0 & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & \cdot & \cdot \\ \lambda^{T-2} & \cdot & \cdot & \lambda & 1 & 0 \end{pmatrix}, \qquad (2.5)$$

one can easily verify that

$$\tilde{y}_{-1} = Gv \qquad \text{and} \qquad \tilde{Z} = Gve_1'. \qquad (2.6)$$

Note, that in the fixed start-up case ($\omega = 0$), we simply have $\tilde{Z} = Cue_1'$. In the derivations to follow we shall stick to the more general case where the value of ω is arbitrary.

We find

$$\begin{aligned} \mathsf{E}(Z'Z) &= \mathsf{E}(\bar{Z} + Gve_1')'(\bar{Z} + Gve_1') && (2.7) \\ &= \bar{Z}'\bar{Z} + \sigma^2 \operatorname{tr}(G'G) e_1 e_1' \end{aligned}$$

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

$$Q = [\mathsf{E}(Z'Z)]^{-1}, \; q_1 = Qe_1, \; q_{11} = e_1'Qe_1. \qquad (2.8)$$

Using the same notation the following result has been proved in KP (1998a).

Theorem 2.1: *Under Assumption 2.1 the bias of the least-squares estimator (1.3) can be approximated to first order as*

$$\mathsf{E}(\hat{\alpha} - \alpha) = -\sigma^2 [\operatorname{tr}(Q\bar{Z}'C\bar{Z})q_1 + Q\bar{Z}'C\bar{Z}q_1 + 2\sigma^2 q_{11} \operatorname{tr}(GG'C)\, q_1] + o(T^{-1}).$$

In fact, KP (1998a) 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. In order to obtain such an approximation one has to find an expansion (in this particular case of the estimation error) in such a form that successive terms are of decreasing 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. 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 Appendix A provides their expectation.

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* $\mathsf{V}(\hat{\alpha}) = \mathsf{E}\{[\hat{\alpha} - \mathsf{E}(\hat{\alpha})][\hat{\alpha} - \mathsf{E}(\hat{\alpha})]'\} =$

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

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 of* $\mathsf{V}(\hat{\alpha})$ *given in (1.4) the approximation* $\mathsf{E}[\hat{\mathsf{V}}(\hat{\alpha})] = \mathsf{E}[s^2(Z'Z)^{-1}] =$

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

Note that the approximation to order T^{-1} (the leading term) of both $\mathsf{V}(\hat{\alpha})$ and $\mathsf{E}[\hat{\mathsf{V}}(\hat{\alpha})]$ is simply $\sigma^2 Q$. However, the second-order approximations of $\mathsf{V}(\hat{\alpha})$ and $\mathsf{E}[\hat{\mathsf{V}}(\hat{\alpha})]$ differ markedly with respect to contributions of order T^{-2}. Note that Theorem 2.3 implies that the first-order approximation to $\mathsf{E}[\hat{\sigma}^2(Z'Z)^{-1}]$, the estimator which omits a degrees of freedom correction, is given by $\sigma^2 Q$ too; selfevidently, the degrees of freedom correction does not affect the leading term. Since the second-order approximation to $\mathsf{E}[\hat{\sigma}^2(Z'Z)^{-1}]$ 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, $\bar{y}_0$, ω, α and σ^2 at relevant sample sizes T, and by comparing these approximative 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 $\mathsf{V}(\hat{\alpha})$, say $\check{\mathsf{V}}(\hat{\alpha})$, which adds particular terms to the standard estimator $\hat{\mathsf{V}}(\hat{\alpha})$, such that $\mathsf{E}[\check{\mathsf{V}}(\hat{\alpha})]$ is equivalent to second order to $\mathsf{V}(\hat{\alpha})$. 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. From Theorem 2.1 one easily obtains:

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

$$\mathsf{E}(\hat{\lambda}-\lambda) = -\sigma^2[q_{11}\operatorname{tr}(Q\bar{Z}'C\bar{Z}) + q_1'\bar{Z}'C\bar{Z}q_1 + 2\sigma^2 q_{11}^2\operatorname{tr}(GG'C)] + o(T^{-1}).$$

From Theorem 2.2 we obtain after pre- and postmultiplication by e_1:

Corollary 2.2: *Under Assumption 2.1 we find for the variance of the estimator* $\hat{\lambda}$ *the approximation* $\mathsf{V}(\hat{\lambda}) = \mathsf{E}[\hat{\lambda}-\mathsf{E}(\hat{\lambda})]^2 =$

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

From Theorem 2.3 we obtain:

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* $\mathsf{E}[\hat{\mathsf{V}}(\hat{\lambda})] = \mathsf{E}[s^2 e_1'(Z'Z)^{-1}e_1] =$

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

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 $\check{\mathsf{V}}(\hat{\lambda})$ for $\mathsf{V}(\hat{\lambda})$.

3. The efficiency of bias corrected coefficient estimators

The approach already laid out in the foregoing section consists of three stages: (i) assess 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 $\hat{\alpha} - \mathsf{B}_\alpha$ with

$$\mathsf{B}_\alpha = -\sigma^2[\mathrm{tr}(Q\bar{Z}'C\bar{Z})q_1 + Q\bar{Z}'C\bar{Z}q_1 + 2\sigma^2 q_{11}\,\mathrm{tr}\,(GG'C)\,q_1] \tag{3.1}$$

it is obvious that this has expectation $\alpha + o(T^{-1})$, but it is not an operational estimator, because σ^2, Q, $\bar{Z}$, C, G, ω and $\bar{y}_0$ are unobservable. However, consider the operational corrected ordinary least-squares (COLS) estimator defined as

$$\begin{array}{rcl}\check{\alpha} & \equiv & \hat{\alpha} - \hat{\mathsf{B}}_\alpha, \text{ with} \\ \hat{\mathsf{B}}_\alpha & = & -s^2[\mathrm{tr}(P\hat{Z}'\hat{C}\hat{Z})p_1 + P\hat{Z}'\hat{C}\hat{Z}p_1 + 2s^2 p_{11}\,\mathrm{tr}(\hat{C}\hat{C}'\hat{C})p_1],\end{array} \tag{3.2}$$

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 $\check{\alpha} = \hat{\alpha} - \hat{\mathsf{B}}_\alpha$ given in (3.2) is unbiased to order T^{-1}, i.e.* $\mathsf{E}(\check{\alpha}) = \alpha + o(T^{-1})$.

For this bias corrected 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 (3.2) the approximation* $\mathsf{V}(\check{\alpha}) = \mathsf{E}\{[\check{\alpha} - \mathsf{E}(\check{\alpha})][\check{\alpha} - \mathsf{E}(\check{\alpha})]'\} =$

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

It is noteworthy that the expression for the second order contribution to the variance of the corrected estimator is in fact substantially simpler than for the uncorrected least-squares estimator given in Theorem 2.2.

From the results in Theorems 2.3 and 3.2 it follows that the variance of the COLS estimator $\mathsf{V}(\check{\alpha})$ can be estimated unbiasedly to order T^{-1} by $\hat{\mathsf{V}}(\check{\alpha}) \equiv s^2(Z'Z)^{-1} = s^2P$, i.e. the standard OLS estimator, because $\mathsf{V}(\check{\alpha})$ and $\mathsf{E}[s^2(Z'Z)^{-1}]$ have both leading term σ^2Q. 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* $\check{\mathsf{V}}(\check{\alpha})$ *of the bias corrected estimator* $\check{\alpha}$ *given in (3.2) has* $\mathsf{E}[\check{\mathsf{V}}(\check{\alpha}) - \mathsf{V}(\check{\alpha})] = o(T^{-2})$, *and hence is almost unbiased, if we define:*

$$\begin{aligned}
\check{\mathsf{V}}(\check{\alpha}) \equiv\; & s^2 P \\
& + s^4\{[\mathrm{tr}(P\hat{Z}'\hat{C}\hat{Z}P\hat{Z}'\hat{C}\hat{Z}) + 2T^{-1}\,\mathrm{tr}(\hat{C}'\hat{C})]p_1p_1' \\
& \quad + P\hat{Z}'\hat{C}\hat{Z}p_1p_1'\hat{Z}'\hat{C}'\hat{Z}P + P\hat{Z}'\hat{C}\hat{Z}P\hat{Z}'\hat{C}\hat{Z}p_1p_1' + p_1p_1'\hat{Z}'\hat{C}'\hat{Z}P\hat{Z}'\hat{C}'\hat{Z}P\} \\
& + 2s^6\,\mathrm{tr}(\hat{C}\hat{C}'\hat{C})[2(p_1'\hat{Z}'\hat{C}\hat{Z}p_1)p_1p_1' + p_{11}(P\hat{Z}'\hat{C}\hat{Z}p_1p_1' + p_1p_1'\hat{Z}'\hat{C}'\hat{Z}P)] \\
& + 4s^8 p_{11}^2[\mathrm{tr}(\hat{C}\hat{C}'\hat{C})]^2 p_1p_1'.
\end{aligned}$$

From (3.1) and (3.2) we easily find B_λ and $\check{\lambda}$, for which we have:

Corollary 3.1: *Under Assumption 2.1 the* COLS *estimator* $\check{\lambda}$ *for* λ *defined as* $\check{\lambda} \equiv \hat{\lambda} + s^2[p_1'\hat{Z}'\hat{C}\hat{Z}p_1 + p_{11}\,\mathrm{tr}(P\hat{Z}'\hat{C}\hat{Z}) + 2s^2p_{11}^2\,\mathrm{tr}(\hat{C}\hat{C}'\hat{C})]$ *is unbiased to order* T^{-1}.

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

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

and

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

$$\begin{aligned}
\check{\mathsf{V}}(\check{\lambda}) \equiv\; & s^2 p_{11} + \\
& + s^4\{(p_1'\hat{Z}'\hat{C}\hat{Z}p_1)^2 + 2p_{11}(p_1'\hat{Z}'\hat{C}\hat{Z}P\hat{Z}'\hat{C}\hat{Z}p_1) \\
& \quad + p_{11}^2[\mathrm{tr}(P\hat{Z}'\hat{C}\hat{Z}P\hat{Z}'\hat{C}\hat{Z}) + 2T^{-1}\,\mathrm{tr}(\hat{C}'\hat{C})]\} \\
& + 8s^6 p_{11}^2 p_1'\hat{Z}'\hat{C}\hat{Z}p_1\,\mathrm{tr}(\hat{C}\hat{C}'\hat{C}) \\
& + 4s^8 p_{11}^4[\mathrm{tr}(\hat{C}\hat{C}'\hat{C})]^2.
\end{aligned}$$

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.

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

$$\begin{aligned}
\check{\mathsf{V}}(\hat{\lambda}) \equiv\; & s^2 p_{11} + \\
& + s^4\{5(p_1'\hat{Z}'\hat{C}\hat{Z}p_1)^2
\end{aligned} \tag{3.3}$$

$$
\begin{aligned}
&\quad +p_{11}[6p_1'\hat{Z}'\hat{C}\hat{Z}P\hat{Z}'\hat{C}\hat{Z}p_1 + 4p_1'\hat{Z}'\hat{C}'\hat{Z}P\hat{Z}'\hat{C}\hat{Z}p_1 \\
&\qquad -6p_1'\hat{Z}'\hat{C}\hat{C}\hat{Z}p_1 - 4p_1'\hat{Z}'\hat{C}'\hat{C}\hat{Z}p_1 + 4p_1'\hat{Z}'\hat{C}\hat{Z}p_1 \operatorname{tr}(P\hat{Z}'\hat{C}\hat{Z})] \\
&\quad +p_{11}^2[\operatorname{tr}(P\hat{Z}'\hat{C}\hat{Z}P\hat{Z}'\hat{C}\hat{Z}) - 2\operatorname{tr}(P\hat{Z}'\hat{C}\hat{C}\hat{Z}) + 2T^{-1}\operatorname{tr}(\hat{C}'\hat{C})]\} \\
&+s^6\{36p_{11}^2 p_1'\hat{Z}'\hat{C}\hat{Z}p_1 \operatorname{tr}(\hat{C}\hat{C}'\hat{C}) \\
&\quad -p_{11}^3[8\operatorname{tr}(\hat{C}\hat{C}'\hat{C}\hat{C}) + 4\operatorname{tr}(\hat{C}\hat{C}'\hat{C}'\hat{C}) - 4\operatorname{tr}(\hat{C}\hat{C}'\hat{C})\operatorname{tr}(P\hat{Z}'\hat{C}\hat{Z})]\} \\
&+20s^8 p_{11}^4[\operatorname{tr}(\hat{C}\hat{C}'\hat{C})]^2.
\end{aligned}
$$

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

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

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 and a zero intercept, but with start-up mean value

$$\bar{y}_0^* = \frac{1}{\sigma}\left(\bar{y}_0 - \frac{\beta}{1-\lambda}\right). \tag{4.3}$$

Note that when $\bar{y}_0 = \beta/(1-\lambda)$, i.e. the mean start-up value is right on the mean-stationary track, all results are 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)

$$\mathsf{E}(\hat{\lambda} - \lambda) = -\frac{1}{T}(1+3\lambda) + o(T^{-1}). \tag{4.4}$$

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

$$\mathsf{V}(\hat{\lambda}) = \frac{1-\lambda^2}{T} - \frac{1-\lambda^2}{T^2}\left(\bar{y}_0^{*2} + \omega^2\right) - \frac{1-4\lambda-14\lambda^2}{T^2} + o(T^{-2}), \tag{4.5}$$

where the leading term $(1-\lambda^2)/T$ is simply the asymptotic variance of $\hat{\lambda}$. Notice that the variance $\mathsf{V}(\hat{\lambda})$ decreases the larger the variance of the initial value ω^2 and also the more $\bar{y}_0$ deviates from the mean-stationary track $\beta/(1-\lambda)$. Obviously, by adding to $\mathsf{V}(\hat{\lambda})$ the square of the first-order bias, we obtain

$$\operatorname{MSE}(\hat{\lambda}) = \frac{1-\lambda^2}{T} - \frac{1-\lambda^2}{T^2}\left(\bar{y}_0^{*2} + \omega^2\right) + \frac{\lambda(10+23\lambda)}{T^2} + o(T^{-2}). \tag{4.6}$$

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

$$\hat{\mathsf{V}}(\hat{\lambda}) = s^2 p_{11} = s^2 \left[\sum_{t=1}^{T} \left(y_{t-1} - \frac{1}{T} \sum_{t=1}^{T} y_{t-1} \right)^2 \right]^{-1}, \tag{4.7}$$

we find, using Corollary 2.3,

$$\mathsf{E}[\hat{\mathsf{V}}(\hat{\lambda})] = \frac{1-\lambda^2}{T} - \frac{1-\lambda^2}{T^2} \left(\bar{y}_0^{*2} + \omega^2 \right) + \frac{2+2\lambda+5\lambda^2}{T^2} + o(T^{-2}). \tag{4.8}$$

It is obvious that $\hat{\mathsf{V}}(\hat{\lambda})$ is unbiased to order T^{-1}, but biased to order T^{-2}, since its second order term differs from the corresponding one of (4.5). So, even though we do not know $\mathsf{V}(\hat{\lambda})$ exactly, we find from its approximation that the standard estimator is biased to second order, viz.

$$\mathsf{E}[\hat{\mathsf{V}}(\hat{\lambda}) - \mathsf{V}(\hat{\lambda})] = \frac{3-2\lambda-9\lambda^2}{T^2} + o(T^{-2}), \tag{4.9}$$

implying that $\hat{\mathsf{V}}(\hat{\lambda})$ 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 $\mathsf{V}(\hat{\lambda})$, viz.

$$\check{\mathsf{V}}(\hat{\lambda}) \equiv \hat{\mathsf{V}}(\hat{\lambda}) - \frac{3-2\hat{\lambda}-9\hat{\lambda}^2}{T^2}, \tag{4.10}$$

which is unbiased to $O(T^{-2})$, and is a special case of (3.3).

For the simple model (4.1) our implementation of COLS leads to

$$\check{\lambda} \equiv \hat{\lambda} + \frac{1}{T}(1+3\hat{\lambda}) = \frac{T+3}{T}\hat{\lambda} + \frac{1}{T}. \tag{4.11}$$

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

$$\begin{aligned} \mathsf{V}(\check{\lambda}) &= \frac{1-\lambda^2}{T} - \frac{1-\lambda^2}{T^2} \left(\bar{y}_0^{*2} + \omega^2 \right) + \frac{5+4\lambda+8\lambda^2}{T^2} + o(T^{-2}) \\ &= \mathsf{V}(\hat{\lambda}) + 6\left(\frac{1-\lambda^2}{T^2} \right) + o(T^{-2}). \end{aligned} \tag{4.12}$$

Note that this reflects the fact that correcting an estimator for bias will invariably lead to an increase in variance. However, from (4.6) and (4.12), which is a second order approximation to $\mathrm{MSE}(\check{\lambda})$ too, it also follows that

$$\mathrm{MSE}(\check{\lambda}) = \mathrm{MSE}(\hat{\lambda}) + \frac{5-6\lambda-15\lambda^2}{T^2} + o(T^{-2}), \tag{4.13}$$

from which a rather precise result follows, viz.:

Theorem 4.1: *In the* AR(1) *model with intercept the* OLS *estimator* $\hat{\lambda}$ *is 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* $\mathrm{MSE}(\hat{\lambda}) < \mathrm{MSE}(\check{\lambda})$, *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 that in the AR(1) model with no (or known) intercept bias correction yields a MSE reduction when $|\lambda| > 0.707$.

An asymptotically valid estimator for the variance $\mathsf{V}(\check{\lambda})$ is provided by $\hat{\mathsf{V}}(\hat{\lambda})$, but an estimator unbiased to order T^{-2} follows simply from (4.12) and (4.9), viz.

$$\check{\mathsf{V}}(\check{\lambda}) \equiv \hat{\mathsf{V}}(\hat{\lambda}) + \frac{3 + 2\hat{\lambda} + 3\hat{\lambda}^2}{T^2}, \tag{4.14}$$

although the correction could also be evaluated in $\check{\lambda}$. Note that $\hat{\mathsf{V}}(\hat{\lambda})$ is negatively biased for $\mathsf{V}(\check{\lambda})$ 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 (4.4), although, instead of using (4.11), 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 an estimator which slightly differs from (4.11), viz.

$$\dot{\lambda} \equiv \frac{T}{T-3}\hat{\lambda} + \frac{1}{T-3}, \tag{4.15}$$

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* $\mathrm{MSE}(\check{\lambda}) < \mathrm{MSE}(\dot{\lambda})$ *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 $\mathsf{V}(\hat{\lambda})$ for what in fact is the Monte Carlo estimate of $\mathsf{V}(\hat{\lambda})$. 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 $\mathsf{E}(\hat{\lambda})$, and likewise for $\check{\lambda}$ and $\dot{\lambda}$. For the mean over the simulations of $\hat{\mathsf{V}}(\hat{\lambda})$ we write $\mathsf{E}[\hat{\mathsf{V}}(\hat{\lambda})]$, and similarly for the estimated expectations of $\check{\mathsf{V}}(\hat{\lambda})$ and $\check{\mathsf{V}}(\check{\lambda})$. Often the results are given as ratios. Self-evidently, when $\mathsf{E}[\hat{\mathsf{V}}(\hat{\lambda})]/\mathsf{V}(\hat{\lambda})$ is unity this indicates unbiasedness of $\hat{\mathsf{V}}(\hat{\lambda})$ and values smaller (greater) than one are found in case of negative (positive) bias, whereas the ratio $\mathrm{MSE}(\check{\lambda})/\,\mathrm{MSE}(\hat{\lambda})$ 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 (what they naturally will do in extreme cases) we examine rather small values of the sample size T and a range of λ values including 0.99.

In the AR(1) model with intercept only, we have two versions of all asymptotic approximations, viz. the general formulas 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. Regarding the first-order approximation to the bias of $\hat{\lambda}$ the trimmed expression is the Kendall formula, i.e. $\mathsf{B}_\lambda = -(1+3\lambda)/T$, and its untrimmed counterpart follows from the first element of (3.1) upon substituting $\bar{y}_0 = \bar{y}_0^*$, $\beta = 0$, $\sigma = 1$, i.e. $\bar{Z} = (\bar{y}_0^* F, \iota)$ and $Q = [\bar{Z}'\bar{Z} + \mathrm{tr}(G'G)e_1e_1']^{-1}$. In Tables 1 through 7 we examine the mean-stationary AR(1) model with intercept, hence $\bar{y}_0^* = 0$, $\bar{Z} = (0, \iota)$ and $q_{11} = [\mathrm{tr}(G'G)]^{-1}$, $q_{22} = T^{-1}$ and $q_{12} = q_{21} = 0$. In Tables 1 and 2, where $\omega = 0$ and $\omega^2 = (1-\lambda^2)^{-1}$ respectively, we examine whether we find any systematic differences in accuracy between trimmed and untrimmed expressions, not only regarding B_λ, but also regarding approximations and estimators of $\mathsf{V}(\hat{\lambda})$. We write $\mathsf{V}_1(\hat{\lambda})$ for the leading term of the asymptotic variance of $\hat{\lambda}$, i.e. for the trimmed expression we have $\mathsf{V}_1(\hat{\lambda}) = (1-\lambda^2)/T$ and for the untrimmed expression $\mathsf{V}_1(\hat{\lambda}) = [\mathrm{tr}(G'G)]^{-1}$. The latter also includes some (but certainly not all) $o(T^{-1})$ contributions. The second-order asymptotic approximation to $\mathsf{V}(\hat{\lambda})$ is denoted as $\mathsf{V}_2(\hat{\lambda})$. For the trimmed expression we have from (4.5)

$$\mathsf{V}_2(\hat{\lambda}) = \frac{1-\lambda^2}{T} - \frac{1-\lambda^2}{T^2}\left(\bar{y}_0^{*2} + \omega^2\right) - \frac{1-4\lambda-14\lambda^2}{T^2}, \tag{5.1}$$

and the untrimmed expression for $\mathsf{V}_2(\hat{\lambda})$ is found from Corollary 2.2. Apart from the bias in $\hat{\mathsf{V}}(\hat{\lambda})$ we also examine the bias of two alternative estimators of $\mathsf{V}(\hat{\lambda})$, viz. $\hat{\mathsf{V}}_2(\hat{\lambda})$ and $\check{\mathsf{V}}(\hat{\lambda})$. By $\hat{\mathsf{V}}_2(\hat{\lambda})$ we denote the estimator obtained by replacing λ by $\hat{\lambda}$ in either the trimmed or untrimmed expressions for $\mathsf{V}_2(\hat{\lambda})$. The trimmed version of $\check{\mathsf{V}}(\hat{\lambda})$ is given by (4.10) and the untrimmed version by (3.3).

In Table 1 we find for the mean-stationary fixed start-up model (i.e. $y_0^* = 0$) the following. The actual bias of $\hat{\lambda}$, its variance and the bias in the standard estimator $\hat{\mathsf{V}}(\hat{\lambda})$ are given in columns (2) through (4) respectively, for various values of λ and T. Column (2) shows that at the very small sample size of $T = 10$ the least-squares estimator is generally badly biased (at least -40% for positive λ). In relative terms the bias 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 smaller but still 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 $\hat{\lambda}$ is about of equal magnitude to the standard deviation of $\hat{\lambda}$. 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 (10) 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. From column (5) we see that as a rule the untrimmed approximation is less accurate, especially for small T and when λ is large. Column (3) shows how $\mathsf{V}(\hat{\lambda})$ changes with λ and T. Columns (6) and (11) show that the first-order asymptotic approximations (evaluated for the true λ values), both untrimmed and trimmed, are very inaccurate, especially at extreme λ values, also at $T = 50$. However, from columns (7) and (12) we see that the second-order approximations are much better, and especially the trimmed version is very good at $T = 50$. At much smaller sample size both seriously overstate the actual variance when λ is not very close to zero. Column (4) shows that, in agreement with our conclusions from (4.9), the standard estimator for $\mathsf{V}(\hat{\lambda})$ 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 smaller sample sizes. Columns (9) and (14) show that at very small sample size the trimmed version of our corrected estimator $\check{\mathsf{V}}(\hat{\lambda})$ is much better than its untrimmed counterpart. At $T = 50$ the trimmed version of $\check{\mathsf{V}}(\hat{\lambda})$ is in fact almost unbiased, also for extreme values of λ, and even for smaller values of T it behaves adequately, though slightly conservative. From columns (8) and (13) we see that the alternative estimator $\check{\mathsf{V}}_2(\hat{\lambda})$ is less satisfactory for extreme λ values. Hence, Table 1 shows that in this model the trimmed version of $\check{\mathsf{V}}(\hat{\lambda})$ is a much better variance estimator than the standard expression $\hat{\mathsf{V}}(\hat{\lambda})$. However, for correcting the bias in $\hat{\lambda}$ itself, the situation seems less promising, because even when evaluated at the true λ value both the trimmed and untrimmed approximations show some defects, and especially for large λ and $T = 50$ results are not as accurate as regarding $\check{\mathsf{V}}(\hat{\lambda})$.

Table 1: Simulations and evaluations for mean-stationary AR(1) model with unknown intercept and fixed start-up

	standard OLS results			untrimmed approximations					trimmed approximations				
λ (1)	$E(\hat{\lambda}-\lambda)$ (2)	$V(\hat{\lambda})$ (3)	$\frac{E[\hat{V}(\hat{\lambda})]}{V(\hat{\lambda})}$ (4)	$\frac{B_\lambda}{E(\hat{\lambda}-\lambda)}$ (5)	$\frac{V_1(\hat{\lambda})}{V(\hat{\lambda})}$ (6)	$\frac{V_2(\hat{\lambda})}{V(\hat{\lambda})}$ (7)	$\frac{E[\hat{V}_2(\hat{\lambda})]}{V(\hat{\lambda})}$ (8)	$\frac{E[\check{V}(\hat{\lambda})]}{V(\hat{\lambda})}$ (9)	$\frac{B_\lambda}{E(\hat{\lambda}-\lambda)}$ (10)	$\frac{V_1(\hat{\lambda})}{V(\hat{\lambda})}$ (11)	$\frac{V_2(\hat{\lambda})}{V(\hat{\lambda})}$ (12)	$\frac{E[\hat{V}_2(\hat{\lambda})]}{V(\hat{\lambda})}$ (13)	$\frac{E[\check{V}(\hat{\lambda})]}{V(\hat{\lambda})}$ (14)
T = 10													
-0.9	0.100	0.063	1.01	1.22	0.57	1.51	4.13	3.50	1.70	0.30	1.38	1.38	1.29
-0.6	0.039	0.075	1.29	1.82	1.02	1.14	2.66	2.30	2.04	0.86	1.08	1.11	1.21
-0.3	-0.034	0.086	1.39	0.40	1.19	0.99	1.37	1.22	0.30	1.06	0.95	0.99	1.17
0	-0.112	0.094	1.39	0.90	1.18	0.99	0.99	0.90	0.90	1.07	0.96	0.96	1.15
0.2	-0.167	0.097	1.35	0.94	1.10	1.06	1.14	1.02	0.96	0.99	1.02	0.98	1.14
0.4	-0.227	0.099	1.28	0.93	0.96	1.18	1.53	1.34	0.97	0.85	1.13	1.03	1.13
0.6	-0.294	0.099	1.19	0.88	0.76	1.36	2.05	1.77	0.95	0.64	1.29	1.11	1.12
0.8	-0.368	0.097	1.06	0.76	0.51	1.58	2.41	2.04	0.92	0.37	1.52	1.22	1.10
0.9	-0.400	0.096	0.96	0.66	0.37	1.54	2.35	1.95	0.93	0.20	1.66	1.30	1.08
0.99	-0.412	0.096	0.82	0.55	0.24	1.25	2.10	1.71	0.96	0.02	1.75	1.35	1.03
T = 20													
-0.9	0.067	0.023	0.89	1.13	0.57	1.26	2.13	1.80	1.27	0.42	1.16	1.25	1.09
-0.6	0.030	0.035	1.12	1.29	0.99	1.04	1.31	1.23	1.34	0.91	1.03	1.04	1.05
-0.3	-0.011	0.043	1.18	0.52	1.11	1.00	1.01	1.02	0.45	1.05	0.99	0.97	1.04
0	-0.053	0.048	1.18	0.94	1.11	1.01	0.95	0.97	0.94	1.05	1.00	0.97	1.04
0.2	-0.082	0.048	1.15	0.96	1.05	1.03	1.01	1.01	0.97	1.00	1.02	0.99	1.04
0.4	-0.113	0.047	1.11	0.95	0.96	1.06	1.14	1.08	0.97	0.90	1.05	1.04	1.04
0.6	-0.148	0.044	1.03	0.91	0.79	1.11	1.35	1.21	0.95	0.73	1.10	1.10	1.04
0.8	-0.193	0.040	0.91	0.81	0.53	1.18	1.61	1.36	0.88	0.46	1.16	1.20	1.03
0.9	-0.221	0.036	0.82	0.70	0.35	1.22	1.71	1.39	0.84	0.26	1.22	1.28	1.02
0.99	-0.235	0.033	0.68	0.51	0.18	0.95	1.65	1.27	0.84	0.03	1.28	1.37	0.98
T = 50													
-0.9	0.031	0.006	0.89	1.06	0.69	1.07	1.38	1.19	1.09	0.61	1.05	1.17	1.03
-0.6	0.015	0.013	1.04	1.05	0.99	1.01	1.05	1.04	1.06	0.96	1.01	1.02	1.01
-0.3	-0.002	0.018	1.07	0.96	1.05	1.00	0.99	1.00	0.92	1.02	1.00	0.99	1.01
0	-0.020	0.020	1.07	1.01	1.05	1.00	0.98	1.00	1.01	1.02	1.00	0.99	1.01
0.2	-0.032	0.019	1.06	1.00	1.02	1.01	1.00	1.01	1.00	1.00	1.01	1.00	1.01
0.4	-0.044	0.018	1.03	0.99	0.97	1.02	1.04	1.02	0.99	0.95	1.02	1.02	1.01
0.6	-0.058	0.015	0.99	0.96	0.88	1.03	1.11	1.04	0.97	0.85	1.02	1.07	1.01
0.8	-0.074	0.011	0.89	0.90	0.67	1.03	1.24	1.08	0.92	0.63	1.03	1.15	1.00
0.9	-0.087	0.009	0.79	0.80	0.46	1.02	1.35	1.11	0.85	0.41	1.02	1.22	0.99
0.99	-0.103	0.007	0.63	0.51	0.16	0.83	1.45	1.07	0.77	0.06	1.01	1.38	0.97

In Table 2 similar results are presented for the model with a random start-up where the y_t series is both mean and covariance stationary. This increases the variance of the initial observations of the series substantially, in comparison to the fixed start-up model, for λ away from zero. Columns (2) and (3) show that $\hat{\lambda}$ is slightly less biased and marginally more efficient now, whereas $\hat{\mathsf{V}}(\hat{\lambda})$ has similar shortcomings as in Table 1. The approximations based on the trimmed and untrimmed formulas have about the same relative qualities as in Table 1.

Table 2: Simulations and evaluations for fully-stationary AR(1) model with unknown intercept and random start-up

	standard OLS results			untrimmed approximations					trimmed approximations				
λ	$\mathsf{E}(\hat{\lambda}-\lambda)$	$\mathsf{V}(\hat{\lambda})$	$\frac{\mathsf{E}[\hat{\mathsf{V}}(\hat{\lambda})]}{\mathsf{V}(\hat{\lambda})}$	$\frac{\mathsf{B}_\lambda}{\mathsf{E}(\hat{\lambda}-\lambda)}$	$\frac{\mathsf{V}_1(\hat{\lambda})}{\mathsf{V}(\hat{\lambda})}$	$\frac{\mathsf{V}_2(\hat{\lambda})}{\mathsf{V}(\hat{\lambda})}$	$\frac{\mathsf{E}[\hat{\mathsf{V}}_2(\hat{\lambda})]}{\mathsf{V}(\hat{\lambda})}$	$\frac{\mathsf{E}[\check{\mathsf{V}}(\hat{\lambda})]}{\mathsf{V}(\hat{\lambda})}$	$\frac{\mathsf{B}_\lambda}{\mathsf{E}(\hat{\lambda}-\lambda)}$	$\frac{\mathsf{V}_1(\hat{\lambda})}{\mathsf{V}(\hat{\lambda})}$	$\frac{\mathsf{V}_2(\hat{\lambda})}{\mathsf{V}(\hat{\lambda})}$	$\frac{\mathsf{E}[\hat{\mathsf{V}}_2(\hat{\lambda})]}{\mathsf{V}(\hat{\lambda})}$	$\frac{\mathsf{E}[\check{\mathsf{V}}(\hat{\lambda})]}{\mathsf{V}(\hat{\lambda})}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
T = 10													
-0.9	0.084	0.041	0.97	1.03	0.46	1.26	3.39	2.78	2.01	0.46	1.84	1.83	1.38
-0.6	0.040	0.065	1.26	1.59	0.99	1.11	2.32	2.03	2.00	0.99	1.09	1.13	1.15
-0.3	-0.027	0.078	1.35	0.41	1.17	0.99	1.27	1.17	0.37	1.17	0.92	0.95	1.10
0	-0.100	0.086	1.35	0.90	1.16	1.00	0.97	0.91	1.00	1.16	0.93	0.92	1.08
0.2	-0.151	0.090	1.31	0.93	1.07	1.05	1.10	1.01	1.06	1.07	1.00	0.95	1.08
0.4	-0.206	0.092	1.24	0.91	0.92	1.13	1.43	1.28	1.07	0.92	1.12	1.02	1.08
0.6	-0.264	0.093	1.13	0.84	0.69	1.21	1.80	1.59	1.06	0.69	1.28	1.10	1.08
0.8	-0.331	0.094	0.98	0.65	0.38	1.14	1.91	1.68	1.03	0.38	1.46	1.18	1.06
0.9	-0.369	0.095	0.89	0.44	0.20	0.83	1.73	1.54	1.00	0.20	1.56	1.22	1.05
0.99	-0.407	0.097	0.81	0.06	0.02	0.09	1.46	1.31	0.98	0.02	1.64	1.25	1.03
T = 20													
-0.9	0.058	0.018	0.89	1.07	0.54	1.17	1.94	1.61	1.48	0.54	1.36	1.46	1.15
-0.6	0.028	0.032	1.11	1.27	0.99	1.04	1.28	1.21	1.41	0.99	1.04	1.05	1.04
-0.3	-0.010	0.041	1.17	0.52	1.10	1.00	1.00	1.02	0.49	1.10	0.98	0.96	1.02
0	-0.051	0.046	1.17	0.94	1.10	1.01	0.95	0.97	0.99	1.10	0.99	0.96	1.02
0.2	-0.078	0.046	1.15	0.96	1.05	1.02	1.01	1.00	1.02	1.05	1.01	0.99	1.03
0.4	-0.108	0.044	1.10	0.95	0.95	1.05	1.13	1.08	1.02	0.95	1.05	1.04	1.03
0.6	-0.139	0.041	1.02	0.90	0.78	1.08	1.31	1.18	1.01	0.78	1.11	1.11	1.03
0.8	-0.177	0.037	0.88	0.77	0.49	1.07	1.50	1.28	0.96	0.49	1.18	1.22	1.02
0.9	-0.202	0.035	0.78	0.60	0.27	0.91	1.48	1.23	0.92	0.27	1.21	1.27	1.01
0.99	-0.230	0.033	0.67	0.11	0.03	0.14	1.28	1.05	0.86	0.03	1.21	1.31	0.98
T = 50													
-0.9	0.029	0.005	0.90	1.06	0.69	1.06	1.36	1.17	1.19	0.69	1.11	1.24	1.06
-0.6	0.015	0.013	1.04	1.05	0.99	1.01	1.05	1.04	1.10	0.99	1.01	1.02	1.01
-0.3	-0.002	0.017	1.07	0.96	1.05	1.00	0.99	1.00	0.94	1.05	1.00	0.99	1.01
0	-0.019	0.019	1.07	1.01	1.04	1.00	0.98	1.00	1.03	1.04	1.00	0.99	1.01
0.2	-0.031	0.019	1.06	1.00	1.02	1.01	1.00	1.01	1.03	1.02	1.01	1.00	1.01
0.4	-0.043	0.017	1.03	0.99	0.97	1.02	1.04	1.02	1.02	0.97	1.02	1.02	1.01
0.6	-0.056	0.015	0.99	0.96	0.88	1.02	1.11	1.04	1.00	0.88	1.03	1.07	1.01
0.8	-0.070	0.011	0.89	0.89	0.67	1.02	1.23	1.07	0.97	0.67	1.05	1.17	1.01
0.9	-0.081	0.009	0.78	0.78	0.44	0.96	1.31	1.08	0.92	0.44	1.05	1.25	1.00
0.99	-0.098	0.007	0.60	0.23	0.06	0.28	1.19	0.92	0.81	0.06	0.97	1.31	0.95

Note that all results for the mean-stationary AR(1) model given here are invariant with respect to β and σ. For both fixed and random start-up we found that $\hat{\mathsf{V}}(\hat{\lambda})$ can be badly biased, that the first-order asymptotic approximation $\mathsf{V}_1(\hat{\lambda})$ is even worse, but that our corrected estimator $\check{\mathsf{V}}(\hat{\lambda})$, especially when based on the trimmed approximations, is very satisfactory when T is not very small. However, $\hat{\lambda}$ itself is badly biased when T is small or moderate, 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 3 results are presented for the model with fixed start-up. Columns (3) and (7) show that the bias corrected estimator $\check{\lambda}$ is much less biased than $\hat{\lambda}$, in general. Using the Kendall formula instead of the untrimmed formula seems only beneficial for large positive λ values. From columns (5) and (10) 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 $\mathsf{V}(\check{\lambda})$. From column (4) we see that the untrimmed estimator according to Corollary 3.3 improves for larger T, but understates the variance where efficiency gains are made. Column (8) shows that the trimmed estimator (4.14) tends to have a negative bias, but the quality difference between the trimmed and untrimmed in estimating $\mathsf{V}(\check{\lambda})$ is much smaller than we found in Tables 1 and 2 with respect to bias corrected estimation of $\mathsf{V}(\hat{\lambda})$. In column (9) we tried an estimator, denoted $\check{\mathsf{V}}(\check{\lambda})^*$, which slightly differs from (4.14), because we evaluated the correction term in $\check{\lambda}$

rather than in $\hat{\lambda}$ (as is also done in the untrimmed variance estimator). Although equivalent to second order, we see that this leads to more satisfactory results here. Note that in finite sample there is a non-zero probability that $|\hat{\lambda}| \geq 1$. The frequency of such occurrences is reported in column (2). In practice one has a single sample and one might be tempted to adjust such an estimate in one way or another, in order to satisfy the stationarity assumption. This would of course not only affect the bias: The resulting distribution would be truncated and have a correspondingly different variance. In Tables 3 and 4 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 (3.2) that generates $\check{\lambda}$. From columns (6) and (11) we see that values of $\check{\lambda}$ in the non-stationarity region occur very frequently, especially when λ is extreme and T is small, and even more so when the trimmed formula is applied.

Table 3: Simulations for mean-stationary AR(1) model with unknown intercept and fixed start-up

		untrimmed approximations				trimmed approximations				
λ	$\lvert\hat{\lambda}\rvert \geq 1$	$\frac{E(\check{\lambda}-\lambda)}{E(\hat{\lambda}-\lambda)}$	$\frac{E[\check{V}(\check{\lambda})]}{V(\check{\lambda})}$	$\frac{MSE(\check{\lambda})}{MSE(\hat{\lambda})}$	$\lvert\check{\lambda}\rvert \geq 1$	$\frac{E(\check{\lambda}-\lambda)}{E(\hat{\lambda}-\lambda)}$	$\frac{E[\check{V}(\check{\lambda})]}{V(\check{\lambda})}$	$\frac{E[\check{V}(\check{\lambda})^*]}{V(\check{\lambda})}$	$\frac{MSE(\check{\lambda})}{MSE(\hat{\lambda})}$	$\lvert\check{\lambda}\rvert \geq 1$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
T = 10										
-0.9	0.186	-0.32	1.22	1.54	0.479	-0.40	0.89	0.94	1.48	0.507
-0.6	0.022	-1.25	1.23	1.79	0.140	-0.74	0.95	0.97	1.67	0.123
-0.3	0.004	1.30	1.19	1.68	0.023	1.00	1.03	1.04	1.68	0.023
0	0.001	0.36	1.11	1.52	0.007	0.41	1.01	1.03	1.51	0.009
0.2	0.001	0.26	1.04	1.40	0.021	0.34	0.95	0.97	1.34	0.013
0.4	0.002	0.25	0.98	1.28	0.060	0.33	0.87	0.90	1.15	0.036
0.6	0.006	0.29	0.92	1.13	0.130	0.35	0.79	0.84	0.96	0.094
0.8	0.017	0.39	0.87	0.94	0.210	0.38	0.76	0.83	0.79	0.201
0.9	0.031	0.45	0.85	0.84	0.252	0.38	0.78	0.86	0.72	0.280
0.99	0.059	0.49	0.81	0.78	0.315	0.34	0.80	0.92	0.69	0.392
T = 20										
-0.9	0.063	-0.10	1.06	1.14	0.337	-0.12	0.94	0.96	1.11	0.346
-0.6	0.001	-0.39	1.06	1.32	0.009	-0.19	0.98	0.99	1.29	0.006
-0.3	0.000	0.95	1.06	1.31	0.000	0.70	1.02	1.03	1.32	0.000
0	0.000	0.19	1.04	1.24	0.000	0.21	1.02	1.02	1.25	0.000
0.2	0.000	0.14	1.02	1.18	0.000	0.18	0.99	1.00	1.16	0.000
0.4	0.000	0.14	0.98	1.09	0.001	0.18	0.95	0.97	1.05	0.001
0.6	0.000	0.18	0.94	0.98	0.022	0.21	0.89	0.92	0.90	0.013
0.8	0.003	0.28	0.88	0.82	0.134	0.27	0.83	0.87	0.72	0.113
0.9	0.012	0.37	0.85	0.71	0.224	0.31	0.82	0.87	0.62	0.229
0.99	0.044	0.44	0.81	0.63	0.337	0.31	0.85	0.91	0.56	0.400
T = 50										
-0.9	0.003	-0.04	1.01	1.00	0.057	-0.03	0.97	0.98	0.97	0.054
-0.6	0.000	-0.06	1.02	1.11	0.000	-0.00	1.00	1.00	1.11	0.000
-0.3	0.000	0.33	1.02	1.12	0.000	0.14	1.01	1.01	1.12	0.000
0	0.000	0.05	1.01	1.10	0.000	0.05	1.01	1.01	1.10	0.000
0.2	0.000	0.05	1.01	1.07	0.000	0.06	1.00	1.01	1.07	0.000
0.4	0.000	0.06	1.00	1.02	0.000	0.07	0.99	1.00	1.01	0.000
0.6	0.000	0.08	0.99	0.94	0.000	0.09	0.98	0.98	0.92	0.000
0.8	0.000	0.15	0.95	0.80	0.005	0.14	0.93	0.95	0.77	0.004
0.9	0.001	0.24	0.91	0.68	0.086	0.21	0.91	0.92	0.64	0.084
0.99	0.029	0.39	0.88	0.53	0.335	0.29	0.93	0.96	0.50	0.380

Table 4 shows similar results for the covariance stationary model, although here we find that using $\check{\lambda}$ for evaluating $\check{V}(\check{\lambda})$, i.e. employing $\check{V}(\check{\lambda})^*$, often increases the bias slightly. From the MSE ratio's in Tables 3 and 4 we see that, although the magnitude of bias is smaller in larger samples, a larger sample size offers in fact more scope for relative efficiency gains through coefficient bias correction, because then the bias assessment is more accurate. Hence, it is not the case that bias correction is called for only when biases are huge. A better effect on efficiency is obtained when we correct for bias in cases where the bias is more moderate, so that the bias approximation is reasonably accurate, and therefore more effective. Note that a MSE reduction of 40%, which is possible at substantial λ values, implies a reduction of some

25% in terms of root mean squared errors (or standard errors). This seems quite attractive against the risks faced when λ is smaller than anticipated, provided one uses the correction only for relatively large $\hat{\lambda}$.

Table 4: Simulations for fully-stationary AR(1) model with unknown intercept and random start-up

		untrimmed approximations				trimmed approximations				
λ	$\lvert\hat{\lambda}\rvert \geq 1$	$\frac{E(\check{\lambda}-\lambda)}{E(\hat{\lambda}-\lambda)}$	$\frac{E[\check{V}(\check{\lambda})]}{V(\check{\lambda})}$	$\frac{MSE(\check{\lambda})}{MSE(\hat{\lambda})}$	$\lvert\check{\lambda}\rvert \geq 1$	$\frac{E(\check{\lambda}-\lambda)}{E(\hat{\lambda}-\lambda)}$	$\frac{E[\check{V}(\check{\lambda})]}{V(\check{\lambda})}$	$\frac{E[\check{V}(\check{\lambda})^*]}{V(\check{\lambda})}$	$\frac{MSE(\check{\lambda})}{MSE(\hat{\lambda})}$	$\lvert\check{\lambda}\rvert \geq 1$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
T = 10										
-0.9	0.129	-0.10	1.26	1.25	0.380	-0.71	1.24	1.33	1.52	0.533
-0.6	0.013	-0.87	1.21	1.64	0.092	-0.70	1.04	1.06	1.66	0.102
-0.3	0.003	1.29	1.17	1.61	0.013	0.93	1.07	1.09	1.68	0.016
0	0.001	0.34	1.10	1.47	0.004	0.30	1.05	1.07	1.52	0.006
0.2	0.001	0.25	1.04	1.36	0.013	0.24	1.00	1.03	1.36	0.011
0.4	0.002	0.25	0.99	1.23	0.045	0.23	0.94	0.98	1.17	0.036
0.6	0.006	0.31	0.94	1.06	0.105	0.24	0.89	0.95	0.99	0.105
0.8	0.022	0.43	0.90	0.87	0.188	0.27	0.89	0.99	0.82	0.241
0.9	0.039	0.52	0.87	0.79	0.226	0.30	0.94	1.06	0.75	0.324
0.99	0.061	0.60	0.84	0.74	0.256	0.32	1.00	1.14	0.69	0.402
T = 20										
-0.9	0.042	-0.02	1.13	1.02	0.240	-0.33	1.15	1.18	1.13	0.342
-0.6	0.000	-0.35	1.07	1.29	0.004	-0.26	1.02	1.02	1.29	0.005
-0.3	0.000	0.93	1.06	1.29	0.000	0.66	1.04	1.04	1.32	0.000
0	0.000	0.19	1.04	1.23	0.000	0.16	1.03	1.04	1.25	0.000
0.2	0.000	0.15	1.02	1.17	0.000	0.13	1.01	1.02	1.17	0.000
0.4	0.000	0.15	0.99	1.08	0.001	0.13	0.98	0.99	1.05	0.001
0.6	0.000	0.19	0.95	0.96	0.016	0.15	0.93	0.96	0.91	0.013
0.8	0.003	0.29	0.90	0.79	0.106	0.19	0.91	0.96	0.73	0.123
0.9	0.015	0.40	0.88	0.68	0.196	0.23	0.94	0.99	0.64	0.265
0.99	0.047	0.53	0.84	0.62	0.282	0.29	1.00	1.08	0.56	0.415
T = 50										
-0.9	0.002	-0.03	1.04	0.96	0.027	-0.13	1.05	1.05	0.98	0.046
-0.6	0.000	-0.06	1.02	1.10	0.000	-0.04	1.01	1.01	1.11	0.000
-0.3	0.000	0.32	1.01	1.12	0.000	0.12	1.01	1.01	1.12	0.000
0	0.000	0.05	1.01	1.10	0.000	0.03	1.01	1.01	1.10	0.000
0.2	0.000	0.05	1.01	1.07	0.000	0.03	1.01	1.01	1.07	0.000
0.4	0.000	0.05	1.00	1.02	0.000	0.04	1.00	1.00	1.01	0.000
0.6	0.000	0.08	0.99	0.94	0.000	0.06	0.99	0.99	0.93	0.000
0.8	0.000	0.15	0.96	0.80	0.002	0.09	0.97	0.98	0.77	0.004
0.9	0.001	0.24	0.94	0.66	0.061	0.14	0.97	0.99	0.65	0.089
0.99	0.035	0.46	0.90	0.52	0.292	0.25	1.06	1.09	0.51	0.410

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 corrections of $\hat{\lambda}$ and $\hat{V}(\hat{\lambda})$ they involve are only performed when $\mid \hat{\lambda} \mid < 1$ and hence leaving $\hat{\lambda}$ and $\hat{V}(\hat{\lambda})$ unchanged otherwise. In Table 5 we have done so for both the models with fixed and random start-up. 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 $\mid \hat{\lambda} \mid \geq 1$ 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 (again $\check{V}(\check{\lambda})^*$ worked markedly better for the fixed start-up model) and to a minor improvement in MSE.

Table 5: Recalculation of Tables 3 and 4 for redefined $\check{\lambda}$ and $\check{\mathsf{V}}(\check{\lambda})$ estimators (no correction when $\mid \hat{\lambda} \mid \geq 1$)

	fixed start-up						random start-up					
	untrimmed			trimmed			untrimmed			trimmed		
λ (1)	$\frac{E(\check{\lambda}-\lambda)}{E(\hat{\lambda}-\lambda)}$ (2)	$\frac{E[\check{V}(\check{\lambda})]}{V(\check{\lambda})}$ (3)	$\frac{MSE(\check{\lambda})}{MSE(\hat{\lambda})}$ (4)	$\frac{E(\check{\lambda}-\lambda)}{E(\hat{\lambda}-\lambda)}$ (5)	$\frac{E[\check{V}(\check{\lambda})^*]}{V(\check{\lambda})}$ (6)	$\frac{MSE(\check{\lambda})}{MSE(\hat{\lambda})}$ (7)	$\frac{E(\check{\lambda}-\lambda)}{E(\hat{\lambda}-\lambda)}$ (8)	$\frac{E[\check{V}(\check{\lambda})]}{V(\check{\lambda})}$ (9)	$\frac{MSE(\check{\lambda})}{MSE(\hat{\lambda})}$ (10)	$\frac{E(\check{\lambda}-\lambda)}{E(\hat{\lambda}-\lambda)}$ (11)	$\frac{E[\check{V}(\check{\lambda})]}{V(\check{\lambda})}$ (12)	$\frac{MSE(\check{\lambda})}{MSE(\hat{\lambda})}$ (13)
T = 10												
-0.9	-0.14	1.27	1.41	0.02	0.92	1.12	-0.05	1.29	1.21	-0.38	1.25	1.19
-0.6	-1.22	1.24	1.77	-0.62	0.97	1.58	-0.86	1.22	1.63	-0.63	1.06	1.61
-0.3	1.30	1.19	1.68	0.98	1.04	1.66	1.29	1.67	1.60	0.91	1.09	1.67
0	0.36	1.11	1.52	0.40	1.03	1.50	0.34	1.10	1.47	0.30	1.07	1.52
0.2	0.26	1.04	1.40	0.35	0.97	1.33	0.25	1.04	1.36	0.25	1.03	1.35
0.4	0.25	0.98	1.28	0.34	0.91	1.14	0.25	0.99	1.23	0.23	0.98	1.16
0.6	0.30	0.92	1.12	0.36	0.84	0.94	0.31	0.94	1.06	0.25	0.95	0.96
0.8	0.40	0.88	0.93	0.40	0.83	0.75	0.43	0.90	0.87	0.30	0.98	0.77
0.9	0.46	0.87	0.83	0.41	0.86	0.68	0.52	0.88	0.79	0.34	1.03	0.69
0.99	0.51	0.83	0.77	0.40	0.91	0.62	0.61	0.86	0.74	0.39	1.10	0.62
T = 20												
-0.9	-0.05	1.09	1.11	-0.02	0.96	1.02	-0.01	1.14	1.01	-0.25	1.17	1.05
-0.6	-0.39	1.06	1.32	-0.19	0.99	1.29	-0.35	1.07	1.29	-0.26	1.02	1.29
-0.3	0.95	1.06	1.31	0.70	1.03	1.32	0.93	1.06	1.29	0.66	1.04	1.32
0	0.19	1.04	1.24	0.21	1.02	1.25	0.19	1.04	1.23	0.16	1.04	1.25
0.2	0.14	1.02	1.18	0.18	1.00	1.16	0.15	1.02	1.17	0.13	1.02	1.17
0.4	0.14	0.98	1.09	0.18	0.97	1.05	0.15	0.99	1.08	0.13	0.99	1.05
0.6	0.18	0.94	0.98	0.21	0.92	0.90	0.19	0.95	0.96	0.15	0.96	0.91
0.8	0.28	0.88	0.82	0.27	0.87	0.71	0.29	0.90	0.79	0.19	0.95	0.72
0.9	0.37	0.86	0.71	0.32	0.87	0.61	0.40	0.88	0.68	0.25	0.99	0.62
0.99	0.45	0.83	0.63	0.35	0.91	0.53	0.54	0.85	0.62	0.33	1.06	0.53
T = 50												
-0.9	-0.04	1.01	0.99	-0.03	0.98	0.97	-0.03	1.04	0.96	-0.12	1.05	0.98
-0.6	-0.06	1.02	1.11	-0.00	1.00	1.11	-0.06	1.02	1.10	-0.04	1.01	1.11
-0.3	0.33	1.02	1.12	0.14	1.01	1.12	0.32	1.01	1.12	0.12	1.01	1.12
0	0.05	1.01	1.10	0.05	1.01	1.10	0.05	1.01	1.10	0.03	1.01	1.10
0.2	0.05	1.01	1.07	0.06	1.01	1.07	0.05	1.01	1.07	0.03	1.01	1.07
0.4	0.06	1.00	1.02	0.07	1.00	1.01	0.05	1.00	1.02	0.04	1.00	1.01
0.6	0.08	0.99	0.94	0.09	0.98	0.92	0.08	0.99	0.94	0.06	0.99	0.93
0.8	0.15	0.95	0.80	0.14	0.95	0.77	0.15	0.96	0.80	0.09	0.98	0.77
0.9	0.24	0.92	0.68	0.21	0.93	0.64	0.24	0.94	0.66	0.14	0.99	0.65
0.99	0.40	0.90	0.53	0.31	0.97	0.48	0.46	0.91	0.52	0.28	1.08	0.48

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 $\check{\mathsf{V}}(\check{\lambda})$ of such a $\check{\lambda}$ estimator could then be taken as either the original formula when $\hat{\lambda}$ is corrected and $\hat{\mathsf{V}}(\hat{\lambda})$ otherwise[2]. We have examined this procedure in Table 6, again omitting the frequencies of estimators in the non-stationarity region, because for this $\check{\lambda}$ these conform to the original frequencies for $\hat{\lambda}$ in Tables 3 and 4. Table 6 shows that 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 and for the untrimmed formula, 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 stationarity restriction.

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

Table 6: Recalculation of Tables 3 and 4 for redefined $\check{\lambda}$ and $\check{V}(\check{\lambda})$ estimators (no correction of $\hat{\lambda}$ when $|\check{\lambda}| \geq 1$)

	fixed start-up						random start-up					
	untrimmed			trimmed			untrimmed			trimmed		
λ	$\frac{E(\check{\lambda}-\lambda)}{E(\hat{\lambda}-\lambda)}$	$\frac{E[\check{V}(\check{\lambda})]}{V(\check{\lambda})}$	$\frac{MSE(\check{\lambda})}{MSE(\hat{\lambda})}$	$\frac{E(\check{\lambda}-\lambda)}{E(\hat{\lambda}-\lambda)}$	$\frac{E[\check{V}(\check{\lambda})]}{V(\check{\lambda})}$	$\frac{MSE(\check{\lambda})}{MSE(\hat{\lambda})}$	$\frac{E(\check{\lambda}-\lambda)}{E(\hat{\lambda}-\lambda)}$	$\frac{E[\check{V}(\check{\lambda})]}{V(\check{\lambda})}$	$\frac{MSE(\check{\lambda})}{MSE(\hat{\lambda})}$	$\frac{E(\check{\lambda}-\lambda)}{E(\hat{\lambda}-\lambda)}$	$\frac{E[\check{V}(\check{\lambda})]}{V(\check{\lambda})}$	$\frac{MSE(\check{\lambda})}{MSE(\hat{\lambda})}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
T = 10												
-0.9	0.69	1.23	0.94	0.59	1.23	0.94	0.60	1.24	0.90	0.47	1.32	0.86
-0.6	-0.14	1.33	1.31	-0.18	1.22	1.40	-0.17	1.28	1.31	-0.25	1.24	1.43
-0.3	1.09	1.22	1.57	0.89	1.09	1.60	1.14	1.19	1.54	0.84	1.11	1.63
0	0.38	1.14	1.47	0.41	1.04	1.47	0.35	1.12	1.44	0.31	1.07	1.49
0.2	0.33	1.13	1.27	0.37	1.05	1.28	0.30	1.10	1.27	0.27	1.07	1.31
0.4	0.40	1.15	1.03	0.38	1.10	1.05	0.37	1.11	1.04	0.29	1.11	1.07
0.6	0.53	1.19	0.83	0.46	1.19	0.82	0.50	1.13	0.84	0.38	1.18	0.82
0.8	0.67	1.19	0.73	0.57	1.25	0.66	0.65	1.11	0.73	0.53	1.21	0.64
0.9	0.72	1.15	0.71	0.62	1.24	0.62	0.71	1.06	0.72	0.61	1.18	0.61
0.99	0.76	1.05	0.72	0.68	1.15	0.61	0.77	1.00	0.73	0.68	1.12	0.61
T = 20												
-0.9	0.46	1.17	0.85	0.37	1.16	0.82	0.35	1.19	0.84	0.23	1.23	0.79
-0.6	-0.34	1.08	1.29	-0.17	1.00	1.28	-0.32	1.07	1.27	-0.25	1.03	1.28
-0.3	0.95	1.06	1.31	0.70	1.02	1.32	0.93	1.06	1.29	0.66	1.04	1.32
0	0.19	1.04	1.24	0.21	1.02	1.25	0.19	1.04	1.23	0.16	1.03	1.25
0.2	0.15	1.02	1.18	0.18	0.99	1.16	0.15	1.02	1.17	0.13	1.01	1.17
0.4	0.15	0.99	1.09	0.18	0.96	1.04	0.15	0.99	1.08	0.13	0.98	1.05
0.6	0.23	0.99	0.91	0.22	0.96	0.87	0.22	0.99	0.91	0.16	0.98	0.88
0.8	0.45	1.06	0.68	0.38	1.07	0.62	0.42	1.03	0.68	0.32	1.08	0.62
0.9	0.59	1.08	0.62	0.50	1.13	0.53	0.57	1.04	0.62	0.48	1.11	0.52
0.99	0.69	1.03	0.62	0.63	1.10	0.52	0.69	0.97	0.63	0.63	1.06	0.52
T = 50												
-0.9	0.05	1.06	0.93	0.04	1.03	0.92	0.01	1.07	0.93	-0.06	1.07	0.93
-0.6	-0.06	1.02	1.11	-0.00	1.00	1.11	-0.06	1.02	1.10	-0.04	1.01	1.11
-0.3	0.33	1.02	1.12	0.14	1.01	1.12	0.32	1.01	1.12	0.12	1.01	1.12
0	0.05	1.01	1.10	0.05	1.01	1.10	0.05	1.01	1.10	0.03	1.01	1.10
0.2	0.05	1.01	1.07	0.06	1.00	1.07	0.05	1.01	1.07	0.03	1.01	1.07
0.4	0.06	1.00	1.02	0.07	0.99	1.01	0.05	1.00	1.02	0.04	1.00	1.01
0.6	0.08	0.99	0.94	0.09	0.98	0.92	0.08	0.99	0.94	0.06	0.99	0.93
0.8	0.15	0.96	0.79	0.14	0.95	0.76	0.15	0.97	0.79	0.10	0.97	0.77
0.9	0.32	1.02	0.61	0.28	1.03	0.57	0.30	1.01	0.62	0.22	1.04	0.57
0.99	0.61	1.11	0.52	0.58	1.16	0.46	0.60	1.04	0.53	0.58	1.10	0.46

In Table 7 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 regularity generally holds except for λ close to -1. Especially for moderate values of λ estimator $\dot{\lambda}$ has much less bias and it even overcorrects when the start-up is random. It is remarkable that $\dot{\lambda}$ is also a bit more efficient than $\check{\lambda}$ when λ is large 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 $\check{\lambda}$.

Table 7: Two alternative trimmed bias corrections in AR(1) with intercept

		fixed start-up			random start-up		
	λ	$\frac{E(\dot\lambda)-\lambda}{E(\check\lambda)-\lambda}$	$\frac{MSE(\dot\lambda)}{MSE(\check\lambda)}$	$\mid\dot\lambda\mid\geq 1$	$\frac{E(\dot\lambda)-\lambda}{E(\check\lambda)-\lambda}$	$\frac{MSE(\dot\lambda)}{MSE(\check\lambda)}$	$\mid\dot\lambda\mid\geq 1$
T = 10	-0.9	0.75	1.01	0.600	2.03	1.35	0.633
	-0.6	2.49	1.14	0.193	2.06	1.23	0.161
	-0.3	0.93	1.18	0.041	0.97	1.21	0.028
	0.0	0.36	1.18	0.019	-0.02	1.20	0.013
	0.2	0.23	1.15	0.034	-0.35	1.20	0.029
	0.4	0.29	1.10	0.084	-0.44	1.19	0.080
	0.6	0.47	1.01	0.185	-0.34	1.18	0.196
	0.8	0.66	0.90	0.330	-0.14	1.15	0.366
	0.9	0.74	0.85	0.421	-0.01	1.12	0.455
	0.99	0.80	0.82	0.531	0.11	1.09	0.528
T = 20	-0.9	0.61	0.98	0.422	1.72	1.08	0.396
	-0.6	2.32	1.05	0.011	1.90	1.05	0.007
	-0.3	0.91	1.05	0.000	0.90	1.05	0.000
	0.0	0.29	1.05	0.000	0.00	1.05	0.000
	0.2	0.16	1.04	0.000	-0.25	1.05	0.000
	0.4	0.19	1.04	0.002	-0.25	1.04	0.002
	0.6	0.36	1.02	0.027	-0.07	1.04	0.022
	0.8	0.66	0.96	0.173	0.23	1.03	0.172
	0.9	0.80	0.91	0.311	0.41	1.01	0.330
	0.99	0.88	0.89	0.489	0.56	0.98	0.482
T = 50	-0.9	-1.94	1.00	0.103	1.59	1.01	0.054
	-0.6	2.92	1.01	0.000	2.05	1.01	0.000
	-0.3	0.85	1.01	0.000	0.85	1.01	0.000
	0.0	0.16	1.01	0.000	-0.13	1.01	0.000
	0.2	0.10	1.01	0.000	-0.28	1.01	0.000
	0.4	0.19	1.01	0.000	-0.18	1.01	0.000
	0.6	0.38	1.01	0.000	0.07	1.01	0.000
	0.8	0.63	1.00	0.009	0.43	1.00	0.005
	0.9	0.82	0.98	0.132	0.63	1.00	0.105
	0.99	0.95	0.95	0.464	0.81	0.98	0.436

Next we present a few results for the AR(1) model with intercept and trend

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

which we rescale by defining

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

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, but with start-up mean value

$$\bar{y}_0^* = \frac{1}{\sigma}\left[\bar{y}_0 - \frac{1}{1-\lambda}\left(\beta_1 + \frac{\lambda}{1-\lambda}\beta_2\right)\right]. \tag{5.4}$$

We examine the mean-stationary fixed start-up case, with $\omega = 0$ and $\bar{y}_0^* = 0$ i.e. $\bar{y}_0 = [\beta_1 + \lambda\beta_2/(1-\lambda)]/(1-\lambda)$, hence all the results are invariant with respect to β_1, β_2 and σ. We found that in models with more regressors our original correction procedures work only reasonably well for moderate values of λ and 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 8 it is seen how badly biased $\hat\lambda$ is in the model with trend when T is small. It is also seen that $\hat{\mathsf{V}}(\hat\lambda)$ has serious shortcomings which (initially) aggravate for increasing T, whereas the variance estimator $\check{\mathsf{V}}(\hat\lambda)$ already works pretty well for $T = 50$. The first order bias approximation B_λ (calculated for the true parameter values) is not very accurate for small T, especially not for large λ, and consequently the estimator $\check\lambda$ shows still substantial bias.

Nevertheless it is more efficient than OLS for $\lambda \geq 0.4$, whereas its variance can be assessed remarkably well by $\check{\mathsf{V}}(\check{\lambda})$. Column (11) shows with what probability the corrected estimator would have obtained values in the non-stationarity region if we had not assured that this does not happen more often than for $\hat{\lambda}$.

Table 8: ARX(1) model with intercept and trend; mean-stationary fixed start-up (no correction if $\mid \check{\lambda} \mid \geq 1$)

λ	$\mathsf{E}(\hat{\lambda})-\lambda$	$\mathsf{V}(\hat{\lambda})$	$\frac{\mathsf{E}[\hat{\mathsf{V}}(\hat{\lambda})]}{\mathsf{V}(\hat{\lambda})}$	$\frac{\mathsf{E}[\check{\mathsf{V}}(\hat{\lambda})]}{\mathsf{V}(\hat{\lambda})}$	$\frac{\mathsf{B}_{\lambda}}{\mathsf{E}(\hat{\lambda})-\lambda}$	$\frac{\mathsf{E}(\check{\lambda})-\lambda}{\mathsf{E}(\hat{\lambda})-\lambda}$	$\frac{\mathsf{E}[\check{\mathsf{V}}(\check{\lambda})]}{\mathsf{V}(\check{\lambda})}$	$\frac{\mathsf{MSE}(\check{\lambda})}{\mathsf{MSE}(\hat{\lambda})}$	$\mid \hat{\lambda} \mid \geq 1$	$\mid \check{\lambda} \mid \geq 1$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
T = 10										
-0.9	0.072	0.055	1.22	4.42	1.50	0.72	1.38	1.04	0.212	0.494
-0.6	-0.012	0.066	1.54	2.84	-2.79	2.87	1.49	1.46	0.031	0.153
-0.3	-0.110	0.078	1.63	1.43	0.67	0.75	1.34	1.55	0.007	0.028
0.0	-0.217	0.088	1.60	1.02	0.82	0.54	1.21	1.25	0.003	0.010
0.2	-0.294	0.094	1.54	1.11	0.82	0.54	1.16	1.03	0.002	0.019
0.4	-0.379	0.099	1.44	1.30	0.78	0.59	1.14	0.86	0.002	0.038
0.6	-0.472	0.105	1.32	1.42	0.69	0.67	1.12	0.77	0.004	0.062
0.8	-0.580	0.112	1.17	1.35	0.54	0.75	1.07	0.75	0.008	0.079
0.9	-0.647	0.116	1.11	1.26	0.42	0.79	1.04	0.76	0.011	0.082
0.99	-0.723	0.118	1.07	1.19	0.30	0.82	1.01	0.78	0.012	0.076
T = 20										
-0.9	0.057	0.021	0.98	1.94	1.21	0.47	1.22	0.92	0.070	0.340
-0.6	0.007	0.033	1.22	1.28	2.68	-2.33	1.12	1.41	0.001	0.009
-0.3	-0.048	0.041	1.27	1.04	0.81	0.46	1.09	1.34	0.000	0.000
0.0	-0.105	0.047	1.26	1.00	0.90	0.29	1.04	1.16	0.000	0.000
0.2	-0.146	0.048	1.21	1.04	0.89	0.28	1.00	1.04	0.000	0.000
0.4	-0.191	0.049	1.15	1.11	0.86	0.32	0.95	0.89	0.000	0.003
0.6	-0.244	0.048	1.06	1.18	0.78	0.41	0.94	0.73	0.000	0.020
0.8	-0.310	0.047	0.93	1.18	0.63	0.57	0.94	0.62	0.001	0.066
0.9	-0.355	0.046	0.86	1.09	0.49	0.66	0.92	0.63	0.003	0.088
0.99	-0.417	0.047	0.80	0.98	0.29	0.74	0.88	0.66	0.007	0.092
T = 50										
-0.9	0.029	0.006	0.93	1.21	1.09	0.04	1.07	0.98	0.003	0.057
-0.6	0.007	0.013	1.08	1.04	1.19	-0.27	1.02	1.16	0.000	0.000
-0.3	-0.016	0.0017	1.10	1.01	0.96	0.15	1.02	1.14	0.000	0.000
0.0	-0.040	0.020	1.09	1.00	0.98	0.10	1.01	1.08	0.000	0.000
0.2	-0.057	0.020	1.07	1.01	0.96	0.11	1.00	1.01	0.000	0.000
0.4	-0.074	0.018	1.03	1.02	0.94	0.13	0.98	0.92	0.000	0.000
0.6	-0.094	0.016	0.97	1.04	0.89	0.18	0.95	0.79	0.000	0.000
0.8	-0.120	0.013	0.85	1.04	0.78	0.30	0.91	0.62	0.000	0.006
0.9	-0.141	0.012	0.76	1.00	0.64	0.45	0.91	0.53	0.000	0.047
0.99	-0.180	0.011	0.65	0.87	0.30	0.64	0.89	0.54	0.005	0.106

Now 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.

$$\left.\begin{aligned} y_t &= \lambda y_{t-1} + \beta_1 + \beta_2 x_t + \sigma\varepsilon_t, \\ x_t &= \rho x_{t-1} + \xi_t. \end{aligned}\right\} \tag{5.5}$$

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

$$\mathsf{V}(y_t) = \frac{1}{1-\lambda^2}\left(\sigma^2 + \frac{\beta_2^2}{1-\rho^2}\frac{1+\lambda\rho}{1-\lambda\rho}\right). \tag{5.6}$$

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

$$SNR = \frac{\mathsf{V}(y_t) - \sigma^2}{\sigma^2}, \tag{5.7}$$

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 (5.5), 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 9 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 substantial λ values $\hat{\mathsf{V}}(\hat{\lambda})$ is negatively biased, but so is $\check{\mathsf{V}}(\hat{\lambda})$, although this behaves significantly better. The first order bias approximation B_λ is reasonably accurate (if λ is not very large whilst T very small) thus $\check{\lambda}$ 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, $\check{\mathsf{V}}(\check{\lambda})$ underestimates the variance when there is an efficiency improvement, though it is less biased than $\check{\mathsf{V}}(\hat{\lambda})$ and $\hat{\mathsf{V}}(\hat{\lambda})$ are. The final column shows that leaving OLS uncorrected did not occur very often for larger T.

Table 9: ARX(1) model with AR(1) regressor; mean-stationary random start-up (no correction if $\mid \check{\lambda} \mid \geq 1$)

λ	$\mathsf{E}(\hat{\lambda}) - \lambda$	$\mathsf{V}(\hat{\lambda})$	$\frac{\mathsf{E}[\hat{\mathsf{V}}(\hat{\lambda})]}{\mathsf{V}(\hat{\lambda})}$	$\frac{\mathsf{E}[\check{\mathsf{V}}(\hat{\lambda})]}{\mathsf{V}(\hat{\lambda})}$	$\frac{\mathsf{B}_\lambda}{\mathsf{E}(\hat{\lambda})-\lambda}$	$\frac{\mathsf{E}(\check{\lambda})-\lambda}{\mathsf{E}(\hat{\lambda})-\lambda}$	$\frac{\mathsf{E}[\check{\mathsf{V}}(\check{\lambda})]}{\mathsf{V}(\check{\lambda})}$	$\frac{\mathsf{MSE}(\check{\lambda})}{\mathsf{MSE}(\hat{\lambda})}$	$\mid \hat{\lambda} \mid \geq 1$	$\mid \check{\lambda} \mid \geq 1$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
T = 10										
0.0	-0.061	0.039	0.98	0.89	1.39	0.27	0.93	1.09	0.000	0.001
0.2	-0.077	0.042	0.94	0.90	1.31	0.25	0.92	1.03	0.000	0.001
0.4	-0.099	0.048	0.88	0.90	1.20	0.29	0.90	0.95	0.001	0.004
0.6	-0.134	0.059	0.79	0.89	1.02	0.41	0.88	0.84	0.005	0.026
0.8	-0.207	0.088	0.68	0.82	0.76	0.62	0.82	0.75	0.038	0.110
0.9	-0.306	0.119	0.66	0.79	0.48	0.73	0.79	0.75	0.080	0.180
0.95	-0.420	0.140	0.70	0.82	0.24	0.79	0.80	0.77	0.080	0.172
T = 20										
0.0	-0.018	0.013	1.06	1.00	1.20	0.07	1.01	1.09	0.000	0.000
0.2	-0.032	0.015	1.08	1.00	1.26	0.09	1.01	1.07	0.000	0.000
0.4	-0.050	0.016	1.06	1.01	1.28	0.13	1.01	0.99	0.000	0.000
0.6	-0.075	0.018	0.94	0.99	1.23	0.21	0.97	0.85	0.000	0.001
0.8	-0.118	0.026	0.70	0.84	0.97	0.40	0.87	0.67	0.003	0.027
0.9	-0.171	0.037	0.58	0.73	0.66	0.58	0.79	0.63	0.026	0.109
0.95	-0.236	0.045	0.59	0.72	0.37	0.68	0.78	0.65	0.047	0.156
T = 50										
0.0	-0.009	0.005	1.03	1.00	0.92	0.02	1.01	1.03	0.000	0.000
0.2	-0.014	0.006	1.03	1.00	0.91	0.03	1.00	1.02	0.000	0.000
0.4	-0.020	0.005	1.03	1.01	0.89	0.05	1.00	0.99	0.000	0.000
0.6	-0.029	0.005	0.99	1.01	0.84	0.08	1.00	0.91	0.000	0.000
0.8	-0.044	0.005	0.84	0.96	0.75	0.17	0.95	0.72	0.000	0.000
0.9	-0.065	0.006	0.66	0.84	0.67	0.32	0.88	0.58	0.001	0.014
0.95	-0.099	0.009	0.58	0.76	0.53	0.52	0.86	0.53	0.014	0.096
T = 100										
0.0	-0.004	0.003	1.01	1.00	0.98	0.02	1.01	1.01	0.000	0.000
0.2	-0.006	0.003	1.02	1.01	0.98	0.02	1.00	1.01	0.000	0.000
0.4	-0.010	0.003	1.02	1.01	0.98	0.02	1.00	1.00	0.000	0.000
0.6	-0.016	0.003	1.00	1.01	0.97	0.04	1.00	0.95	0.000	0.000
0.8	-0.024	0.002	0.92	1.00	0.91	0.10	0.99	0.81	0.000	0.000
0.9	-0.034	0.002	0.79	0.95	0.85	0.19	0.95	0.67	0.000	0.000
0.95	-0.044	0.003	0.67	0.88	0.73	0.34	0.92	0.56	0.001	0.033

Finally we examine how bias correction of either or both coefficient and variance estimators affects test size. We consider various one-sided tests of $\mathsf{H}_0 : \lambda = \lambda_0$ both against the left-hand and against the right-hand side alternative, all at nominal significance level 5%. First we examine again the AR(1) model with intercept only and try the crude asymptotic test statistic $N \equiv (\hat{\lambda} - \lambda_0)/[\mathsf{V}_1(\lambda_0)]^{1/2}$, which is asymptotically standard normal under the null. This can be adapted in the following ways. Correcting only the numerator yields $N^0 \equiv [\hat{\lambda} + (1 + 3\lambda_0)/T - \lambda_0]/[\mathsf{V}_1(\lambda_0)]^{1/2}$, and correcting only the denominator leads to $N_0 \equiv (\hat{\lambda} - \lambda_0)/[\mathsf{V}_2(\lambda_0)]^{1/2}$, whereas correcting both gives $N_0^0 \equiv [\hat{\lambda} + (1 + 3\lambda_0)/T - \lambda_0]/[\mathsf{V}_2(\lambda_0)]^{1/2}$. We shall also examine tests which have the same asymptotic distribution, though we shall use them with critical values from the Student distribution with $T - 2$ degrees of freedom. They are the habitual test statistic $t \equiv (\hat{\lambda} - \lambda_0)/[\hat{\mathsf{V}}(\hat{\lambda})]^{1/2}$, and the corrected statistics $t^0 \equiv [\hat{\lambda} + (1 + 3\lambda_0)/T - \lambda_0]/[\hat{\mathsf{V}}(\hat{\lambda})]^{1/2}$, $t_0 \equiv (\hat{\lambda} - \lambda_0)/[\hat{\mathsf{V}}(\hat{\lambda}) - (3 - 2\lambda_0 - 9\lambda_0^2)/T^2]^{1/2}$, and $t_0^0 \equiv [\hat{\lambda} + (1 + 3\lambda_0)/T - \lambda_0]/[\hat{\mathsf{V}}(\hat{\lambda}) - (3 - 2\lambda_0 - 9\lambda_0^2)/T^2]^{1/2}$. Furthermore, we shall also examine test statistics where the corrections are not evaluated under the null but where the estimators $\check{\lambda}$, $\check{\mathsf{V}}(\hat{\lambda})$ and $\check{\mathsf{V}}(\check{\lambda})$ are exploited, i.e. $t^* \equiv (\check{\lambda} - \lambda_0)/[\hat{\mathsf{V}}(\hat{\lambda})]^{1/2}$, $t_* \equiv (\hat{\lambda} - \lambda_0)/[\check{\mathsf{V}}(\hat{\lambda})]^{1/2}$ and $t_*^* \equiv (\check{\lambda} - \lambda_0)/[\check{\mathsf{V}}(\check{\lambda})]^{1/2}$.

Table 10 contains results for the model with random start-up for the standard uncorrected N and t test statitics and for the statistics where the trimmed corrections are evaluated under the null. The latter exist here because there are no nuisance parameters in the mean-stationary AR(1) model with just an intercept. First focussing on testing against $\mathsf{H}_1 : \lambda < \lambda_0$ we see from Table 10 that a sample of $T = 50$ is still much too small for the crude asymptotic test and the habitual t-test to work well. Their actual size is strongly dependent on λ_0 and is too small for λ_0 substantially negative, and far too large for λ_0 towards one. Correcting just the numerator

Table 10: Test size in mean-stationary AR(1) model with intercept, random start-up

	left-hand alternative						right-hand alternative					
λ_0	N	N_0^0	t	t^0	t_0	t_0^0	N	N_0^0	t	t^0	t_0	t_0^0
T = 10												
-0.9	0.02	0.01	0.03	0.23	0.00	0.02	0.19	0.02	0.04	0.01	0.02	0.01
-0.6	0.01	0.02	0.03	0.07	0.05	0.09	0.08	0.05	0.03	0.02	0.03	0.02
-0.3	0.03	0.05	0.04	0.04	0.09	0.08	0.04	0.06	0.02	0.02	0.03	0.03
0.0	0.08	0.05	0.06	0.03	0.10	0.06	0.02	0.06	0.01	0.02	0.02	0.04
0.2	0.12	0.05	0.07	0.03	0.10	0.04	0.01	0.05	0.01	0.03	0.02	0.05
0.4	0.19	0.05	0.09	0.03	0.10	0.03	0.01	0.03	0.01	0.04	0.01	0.04
0.6	0.29	0.04	0.12	0.03	0.10	0.02	0.01	0.02	0.00	0.05	0.00	0.03
0.8	0.48	0.04	0.18	0.03	0.10	0.02	0.01	0.01	0.00	0.07	0.00	0.02
0.9	0.65	0.04	0.24	0.04	0.11	0.02	0.01	0.01	0.00	0.09	0.00	0.01
0.99	0.88	0.03	0.32	0.04	0.11	0.01	0.03	0.00	0.00	0.10	0.00	0.00
T = 20												
-0.9	0.01	0.00	0.03	0.17	0.00	0.04	0.18	0.04	0.06	0.02	0.04	0.01
-0.6	0.01	0.02	0.04	0.06	0.05	0.07	0.09	0.06	0.04	0.03	0.04	0.03
-0.3	0.03	0.04	0.05	0.05	0.07	0.06	0.05	0.06	0.03	0.03	0.04	0.04
0.0	0.07	0.05	0.06	0.04	0.08	0.05	0.03	0.05	0.02	0.04	0.03	0.05
0.2	0.10	0.05	0.07	0.04	0.09	0.05	0.01	0.04	0.02	0.04	0.02	0.05
0.4	0.14	0.06	0.09	0.04	0.09	0.04	0.01	0.03	0.01	0.05	0.01	0.05
0.6	0.21	0.06	0.11	0.04	0.10	0.03	0.00	0.02	0.01	0.06	0.00	0.05
0.8	0.36	0.06	0.17	0.04	0.12	0.03	0.00	0.01	0.00	0.08	0.00	0.03
0.9	0.52	0.06	0.23	0.05	0.14	0.03	0.00	0.00	0.00	0.09	0.00	0.02
0.99	0.86	0.07	0.38	0.06	0.17	0.03	0.02	0.00	0.00	0.11	0.00	0.01
T = 50												
-0.9	0.00	0.00	0.03	0.11	0.01	0.06	0.15	0.06	0.06	0.03	0.05	0.03
-0.6	0.02	0.03	0.04	0.06	0.05	0.06	0.08	0.06	0.05	0.04	0.05	0.04
-0.3	0.04	0.04	0.05	0.05	0.06	0.06	0.05	0.06	0.04	0.04	0.04	0.05
0.0	0.06	0.05	0.06	0.05	0.07	0.05	0.03	0.05	0.03	0.05	0.04	0.05
0.2	0.08	0.05	0.07	0.04	0.07	0.05	0.02	0.05	0.03	0.05	0.03	0.05
0.4	0.10	0.06	0.08	0.04	0.08	0.04	0.02	0.04	0.02	0.05	0.03	0.06
0.6	0.14	0.06	0.09	0.04	0.09	0.04	0.01	0.03	0.02	0.06	0.02	0.06
0.8	0.23	0.07	0.13	0.05	0.11	0.04	0.00	0.01	0.01	0.08	0.00	0.05
0.9	0.34	0.07	0.18	0.05	0.13	0.04	0.00	0.00	0.00	0.09	0.00	0.03
0.99	0.79	0.09	0.38	0.08	0.20	0.05	0.01	0.00	0.00	0.11	0.00	0.01

or denominator of N helps but does not lead to satisfactory results (not presented in the table) whereas correcting both, though extremely effictive, does not either. A similar pattern emerges for the t^0, t_0 and t_0^0 tests, where the latter behaves quite well for all λ_0 vlues when $T = 50$. Achieving size control through bias correction proves to require larger samples when testing against $\mathsf{H}_1 : \lambda > \lambda_0$.

In Table 11 we examine for the same model the tests where the trimmed correction terms have been estimated instead of evaluated under the null. Against left-hand side alternatives we notice that it is better to correct both numerator and denominator instead of just one, and that the tail probabilities of the test statistics do not improve when corrections are only performed when they do not involve or yield coefficient estimates in the non-stationarity region. Also, test t_*^* seems certainly to be at least as good as test t_0^0, as far as size control is concerned. Similar conclusions apply to the tests against right-hand side alternatives except when very close to the unit root. Because it seems more difficult to achieve size control here, it could be useful for practitioners to know that the test t^* seems to provide a conservative test (actual size larger than nominal level), whereas t proves to be liberal if λ_0 is not extremely negative.

Table 11: Test size in mean-stationary AR(1) model with intercept, random start-up

	left-hand alternative, corrected:					right-hand alternative, corrected:				
	always			if $\mid \check{\lambda} \mid < 1$		always			if $\mid \check{\lambda} \mid < 1$	
λ_0	t^*	t_*	t_*^*	t^*	t_*^*	t^*	t_*	t_*^*	t^*	t_*^*
T = 10										
-0.9	0.25	0.00	0.01	0.03	0.03	0.03	0.06	0.02	0.03	0.02
-0.6	0.11	0.00	0.04	0.06	0.03	0.04	0.05	0.03	0.04	0.03
-0.3	0.09	0.04	0.05	0.08	0.05	0.05	0.03	0.03	0.05	0.03
0.0	0.08	0.08	0.05	0.08	0.05	0.05	0.01	0.03	0.05	0.03
0.2	0.08	0.11	0.06	0.08	0.06	0.05	0.00	0.02	0.05	0.02
0.4	0.08	0.14	0.06	0.08	0.06	0.06	0.00	0.02	0.03	0.01
0.6	0.09	0.17	0.07	0.09	0.07	0.07	0.00	0.01	0.01	0.00
0.8	0.10	0.22	0.07	0.10	0.07	0.08	0.00	0.01	0.00	0.00
0.9	0.11	0.26	0.08	0.11	0.08	0.09	0.00	0.01	0.00	0.00
0.99	0.13	0.30	0.09	0.14	0.10	0.09	0.00	0.01	0.00	0.00
T = 20										
-0.9	0.18	0.00	0.03	0.03	0.03	0.04	0.06	0.03	0.04	0.03
-0.6	0.09	0.02	0.05	0.09	0.05	0.04	0.05	0.04	0.04	0.04
-0.3	0.07	0.05	0.05	0.07	0.05	0.05	0.04	0.04	0.05	0.04
0.0	0.06	0.08	0.05	0.06	0.05	0.05	0.02	0.04	0.05	0.04
0.2	0.06	0.09	0.05	0.06	0.05	0.06	0.02	0.04	0.06	0.04
0.4	0.06	0.11	0.05	0.06	0.05	0.06	0.01	0.04	0.06	0.04
0.6	0.07	0.13	0.05	0.07	0.05	0.07	0.00	0.03	0.07	0.03
0.8	0.08	0.17	0.06	0.08	0.06	0.08	0.00	0.02	0.01	0.00
0.9	0.09	0.21	0.07	0.09	0.07	0.09	0.00	0.01	0.00	0.00
0.99	0.12	0.27	0.08	0.13	0.09	0.10	0.00	0.01	0.00	0.00
T = 50										
-0.9	0.12	0.00	0.05	0.10	0.04	0.04	0.06	0.03	0.04	0.03
-0.6	0.07	0.04	0.06	0.07	0.06	0.05	0.05	0.04	0.05	0.04
-0.3	0.06	0.06	0.05	0.06	0.05	0.05	0.04	0.05	0.05	0.05
0.0	0.06	0.07	0.05	0.06	0.05	0.06	0.04	0.05	0.06	0.05
0.2	0.06	0.08	0.05	0.06	0.05	0.06	0.03	0.05	0.06	0.05
0.4	0.05	0.08	0.05	0.05	0.05	0.06	0.02	0.05	0.06	0.05
0.6	0.05	0.10	0.05	0.05	0.05	0.07	0.01	0.05	0.07	0.05
0.8	0.06	0.12	0.05	0.06	0.05	0.09	0.00	0.04	0.09	0.04
0.9	0.06	0.15	0.05	0.06	0.05	0.10	0.00	0.03	0.02	0.00
0.99	0.10	0.24	0.07	0.10	0.07	0.11	0.00	0.01	0.00	0.00

Table 12 presents results for the mean-stationary AR(1) model with intercept and trend and fixed start-up. Hence, only untrimmed corrections are available now, and we only examined test statistics where the corrections were estimated. Note that for both alternative hypotheses 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 leads to a test that converges faster to an exact test, especially for moderate and negative values of λ_0. Performing the correction always seems beneficial for the tail probabilities.

Table 12: Test size in mean-stationary AR(1) model with intercept and trend, fixed start-up

	left-hand alternative, corrected:						right-hand alternative, corrected:					
		always			if $\mid \check{\lambda} \mid < 1$			always			if $\mid \check{\lambda} \mid < 1$	
λ_0	t	t^*	t_*	t_*^*	t^*	t_*^*	t	t^*	t_*	t_*^*	t^*	t_*^*
T = 10												
-0.9	0.03	0.15	0.02	0.03	0.03	0.03	0.01	0.02	0.03	0.02	0.02	0.02
-0.6	0.04	0.11	0.03	0.04	0.05	0.04	0.01	0.02	0.02	0.01	0.02	0.01
-0.3	0.06	0.10	0.05	0.05	0.08	0.05	0.00	0.02	0.01	0.01	0.02	0.01
0.0	0.09	0.10	0.12	0.07	0.10	0.07	0.00	0.03	0.00	0.01	0.02	0.01
0.2	0.12	0.11	0.18	0.09	0.11	0.09	0.00	0.03	0.00	0.00	0.01	0.00
0.4	0.16	0.14	0.23	0.12	0.14	0.12	0.00	0.03	0.00	0.00	0.00	0.00
0.6	0.23	0.18	0.29	0.16	0.18	0.16	0.00	0.02	0.00	0.00	0.00	0.00
0.8	0.36	0.26	0.40	0.24	0.26	0.24	0.00	0.01	0.00	0.00	0.00	0.00
0.9	0.46	0.32	0.48	0.30	0.32	0.30	0.00	0.00	0.00	0.00	0.00	0.00
0.99	0.57	0.40	0.57	0.37	0.40	0.37	0.00	0.00	0.00	0.00	0.00	0.00
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

Table 13 presents similar results for the model with an exogenous AR(1) regressor. Again we find huge size distortions of the standard t-test. And again convergence of the actual null distribution of the test statistics to Student's seems most rapid as well as monotonic for the t_*^* test, which fully exploits the second-order asymptotic findings of this paper.

Table 13: Test size in mean-stationary ARX(1) model with intercept and AR(1) regressor, random start-up

	left-hand alternative, corrected:						right-hand alternative, corrected:					
		always			if $\mid \breve{\lambda} \mid < 1$			always			if $\mid \breve{\lambda} \mid < 1$	
λ_0	t	t^*	t_*	t^*_*	t^*	t^*_*	t	t^*	t_*	t^*_*	t^*	t^*_*
T = 10												
0.0	0.08	0.07	0.09	0.06	0.07	0.06	0.03	0.04	0.03	0.04	0.04	0.04
0.2	0.09	0.07	0.10	0.07	0.07	0.07	0.02	0.04	0.02	0.04	0.04	0.04
0.4	0.10	0.08	0.11	0.08	0.08	0.08	0.02	0.04	0.02	0.03	0.04	0.03
0.6	0.13	0.09	0.13	0.09	0.09	0.09	0.02	0.03	0.02	0.03	0.03	0.03
0.8	0.18	0.13	0.18	0.13	0.13	0.12	0.01	0.02	0.01	0.02	0.01	0.01
0.9	0.25	0.18	0.25	0.18	0.18	0.17	0.01	0.01	0.01	0.01	0.01	0.01
0.95	0.33	0.24	0.33	0.24	0.24	0.23	0.00	0.01	0.00	0.00	0.00	0.00
T = 20												
0.0	0.07	0.06	0.07	0.06	0.06	0.06	0.03	0.05	0.03	0.04	0.05	0.04
0.2	0.07	0.06	0.08	0.06	0.06	0.06	0.03	0.05	0.03	0.04	0.05	0.04
0.4	0.09	0.06	0.10	0.06	0.06	0.06	0.02	0.05	0.02	0.04	0.05	0.04
0.6	0.11	0.07	0.12	0.06	0.07	0.06	0.02	0.05	0.01	0.04	0.05	0.04
0.8	0.18	0.10	0.17	0.09	0.10	0.09	0.01	0.04	0.01	0.03	0.03	0.02
0.9	0.27	0.14	0.24	0.13	0.14	0.13	0.01	0.01	0.01	0.01	0.01	0.01
0.95	0.37	0.21	0.34	0.19	0.21	0.19	0.00	0.01	0.00	0.00	0.00	0.00
T = 50												
0.0	0.06	0.05	0.06	0.05	0.05	0.05	0.04	0.05	0.04	0.05	0.05	0.05
0.2	0.07	0.05	0.07	0.05	0.05	0.05	0.03	0.05	0.03	0.05	0.05	0.05
0.4	0.08	0.05	0.08	0.05	0.05	0.05	0.03	0.05	0.03	0.05	0.05	0.05
0.6	0.09	0.05	0.10	0.05	0.05	0.05	0.02	0.06	0.02	0.05	0.06	0.05
0.8	0.14	0.06	0.13	0.06	0.06	0.06	0.01	0.06	0.01	0.05	0.06	0.05
0.9	0.22	0.09	0.18	0.08	0.09	0.08	0.01	0.05	0.01	0.03	0.04	0.03
0.95	0.34	0.14	0.27	0.12	0.14	0.12	0.00	0.01	0.00	0.01	0.00	0.00
T = 100												
0.0	0.06	0.05	0.06	0.05	0.05	0.05	0.04	0.05	0.04	0.05	0.05	0.05
0.2	0.06	0.05	0.06	0.05	0.05	0.05	0.04	0.05	0.04	0.05	0.05	0.05
0.4	0.07	0.05	0.07	0.05	0.05	0.05	0.03	0.05	0.03	0.05	0.05	0.05
0.6	0.08	0.05	0.08	0.05	0.05	0.05	0.03	0.05	0.03	0.05	0.06	0.05
0.8	0.11	0.05	0.10	0.05	0.05	0.05	0.02	0.06	0.01	0.05	0.06	0.05
0.9	0.15	0.06	0.13	0.06	0.06	0.06	0.01	0.06	0.01	0.04	0.06	0.04
0.95	0.22	0.08	0.18	0.07	0.08	0.07	0.01	0.05	0.00	0.02	0.03	0.01

For models such as those examined here Rudebusch (1992, 1993) suggests tests which mimic t^*. From the above we conclude that a much more successful finite sample correction of test statistics will be achieved not just by applying bias correction to parameter estimators but also to the associated variance estimators, and the required bias approximation to the variance of the already bias corrected coefficient estimator has been derived here for any first-order autoregressive model.

6. Conclusions

By adapting and extending techniques we employed in some recent 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 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 simple linear dynamic regression model), but should be of use primarily for improving the accuracy of inference in finite samples of dynamic regression models. We undertook in this study some numerical and simulation analysis to produce 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 extremely 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, 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 occurence 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 versions of $\check{\lambda}$, $\check{\mathsf{V}}(\hat{\lambda})$ and $\check{\mathsf{V}}(\check{\lambda})$, 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. Regarding potential efficiency gains or losses due to bias correction we find strong analytical evidence for AR(1) models and the simulation experiments show that exploiting our higher-order asymptotic findings can substantially reduce bias, enhance efficiency and almost resolve test size problems. More particularly we found that bias correction may be more effective from an efficiency 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.

References

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., 1998a. Higher-order asymptotic expansions of the least-squares estimation bias in first-order dynamic regression models. Paper presented at ESWM-95, Tokyo, August 1995; revised October 1998.

Kiviet, J.F., Phillips, G.D.A., 1998b. 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., 2003. Moment Approximation for Least Squares Estimators in Dynamic Regression Models with a Unit Root. Tinbergen Institute discussion paper 2001-118/4; revised version UvA-Econometrics discussion paper 2003/03.

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.

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.

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.

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

A. Some auxiliary results

To prove the results of this paper, we have to obtain the expectation of numerous expressions of various forms, which involve products of up to four quadratic forms in normal variables. In this appendix we state some basic results which are used repeatedly in the subsequent analysis.

Let A be a symmetric $T \times T$ matrix and B_1 and B_2 arbitrary $T \times T$ matrices. Let the $T \times 1$ random vector ε be such that $\varepsilon \sim \mathsf{N}(0, \sigma^2 I_T)$, then the following results hold:

$$\mathsf{E}[(\varepsilon' B_1 \varepsilon)(\varepsilon' B_2 \varepsilon)] = \sigma^4[\operatorname{tr}(B_1)\operatorname{tr}(B_2) + \operatorname{tr}(B_1 B_2) + \operatorname{tr}(B_1 B_2')]; \tag{A.1}$$

$$\mathsf{E}[\varepsilon' A \varepsilon - \sigma^2 \operatorname{tr}(A)](\varepsilon' B_1 \varepsilon) = 2\sigma^4 \operatorname{tr}(AB_1); \tag{A.2}$$

$$\mathsf{E}(\varepsilon\varepsilon' B_1 \varepsilon\varepsilon') = \mathsf{E}(\varepsilon' B_1 \varepsilon)\varepsilon\varepsilon' = \sigma^4[\operatorname{tr}(B_1) I_T + B_1 + B_1']; \tag{A.3}$$

$$\begin{aligned}\mathsf{E}\{[(\varepsilon' A \varepsilon)\textstyle\prod_{j=1}^2(\varepsilon' B_j \varepsilon)]\} = &\\ \sigma^6[&\operatorname{tr}(A)\operatorname{tr}(B_1)\operatorname{tr}(B_2) + \operatorname{tr}(A)\operatorname{tr}(B_1B_2) + \operatorname{tr}(A)\operatorname{tr}(B_1B_2')\\ &+2\operatorname{tr}(B_1)\operatorname{tr}(AB_2) + 2\operatorname{tr}(B_2)\operatorname{tr}(AB_1) + 2\operatorname{tr}(AB_2B_1)\\ &+2\operatorname{tr}(AB_2'B_1) + 2\operatorname{tr}(AB_1B_2) + 2\operatorname{tr}(AB_1B_2')];\end{aligned} \tag{A.4}$$

$$\begin{aligned}\mathsf{E}\{[\varepsilon' A \varepsilon - \sigma^2 \operatorname{tr}(A)]\textstyle\prod_{j=1}^2(\varepsilon' B_j \varepsilon)\} = &\\ 2\sigma^6[&\operatorname{tr}(B_1)\operatorname{tr}(AB_2) + \operatorname{tr}(B_2)\operatorname{tr}(AB_1)\\ &+\operatorname{tr}(AB_2B_1) + \operatorname{tr}(AB_2'B_1) + \operatorname{tr}(AB_1B_2) + \operatorname{tr}(AB_1B_2')];\end{aligned} \tag{A.5}$$

$$\begin{aligned}\mathsf{E}\{[\varepsilon' A \varepsilon - \sigma^2 \operatorname{tr}(A)](\varepsilon\varepsilon' B_1 \varepsilon\varepsilon')\} = &\\ 2\sigma^6[&\operatorname{tr}(AB_1)I_T + \operatorname{tr}(B_1)A + AB_1 + B_1A + AB_1' + B_1'A];\end{aligned} \tag{A.6}$$

$$\mathsf{E}\{[\varepsilon' A \varepsilon - \sigma^2 \operatorname{tr}(A)]^2(\varepsilon' B_1 \varepsilon)\} = \sigma^6[2\operatorname{tr}(B_1)\operatorname{tr}(AA) + 8\operatorname{tr}(AAB_1)]; \tag{A.7}$$

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

$$\begin{aligned} &\mathsf{E}(\varepsilon\varepsilon' B_1 \varepsilon\varepsilon' B_2 \varepsilon\varepsilon') = \\ &\quad \sigma^6\{[\operatorname{tr}(B_1)\operatorname{tr}(B_2) + \operatorname{tr}(B_1 B_2) + \operatorname{tr}(B_1 B_2')]I_T \\ &\qquad + \operatorname{tr}(B_1)B_2 + \operatorname{tr}(B_1)B_2' + \operatorname{tr}(B_2)B_1 + \operatorname{tr}(B_2)B_1' \\ &\qquad + B_1B_2 + B_1'B_2 + B_1B_2' + B_1'B_2' \\ &\qquad + B_2B_1 + B_2'B_1 + B_2B_1' + B_2'B_1'\}; \end{aligned} \tag{A.9}$$

$$\begin{aligned} &\mathsf{E}[(\varepsilon' A\varepsilon)^2(\varepsilon' B_1\varepsilon)^2] = \\ &\quad \sigma^8\{[\operatorname{tr}(A)]^2[\operatorname{tr}(B_1B_1) + \operatorname{tr}(B_1B_1')] \\ &\qquad + [\operatorname{tr}(B_1)]^2[\operatorname{tr}(A)\operatorname{tr}(A) + 2\operatorname{tr}(AA)] \\ &\qquad + 4\operatorname{tr}(A)[2\operatorname{tr}(AB_1B_1) + \operatorname{tr}(AB_1B_1') + \operatorname{tr}(AB_1'B_1)] \\ &\qquad + 8\operatorname{tr}(B_1)[\operatorname{tr}(A)\operatorname{tr}(AB_1) + 2\operatorname{tr}(AAB_1)] \\ &\qquad + 2\operatorname{tr}(AA)[\operatorname{tr}(B_1B_1) + \operatorname{tr}(B_1'B_1)] + 8[\operatorname{tr}(AB_1)]^2 \\ &\qquad + 16\operatorname{tr}(AAB_1B_1) + 8\operatorname{tr}(AAB_1'B_1) + 8\operatorname{tr}(AAB_1B_1') \\ &\qquad + 8\operatorname{tr}(AB_1AB_1) + 8\operatorname{tr}(AB_1AB_1')\}; \end{aligned} \tag{A.10}$$

$$\begin{aligned} &\mathsf{E}\{[\varepsilon' A\varepsilon - \sigma^2\operatorname{tr}(A)]^2(\varepsilon' B_1\varepsilon)^2\} = \\ &\quad 2\sigma^8\{8\operatorname{tr}(B_1)\operatorname{tr}(AAB_1) \\ &\qquad + \operatorname{tr}(AA)[\operatorname{tr}(B_1)\operatorname{tr}(B_1) + \operatorname{tr}(B_1B_1) + \operatorname{tr}(B_1'B_1)] \\ &\qquad + 4[\operatorname{tr}(AB_1)]^2 + 8\operatorname{tr}(AAB_1B_1) + 4\operatorname{tr}(AAB_1'B_1) \\ &\qquad + 4\operatorname{tr}(AAB_1B_1') + 4\operatorname{tr}(AB_1AB_1) + 4\operatorname{tr}(AB_1AB_1')\}. \end{aligned} \tag{A.11}$$

Most of these results are also given in KP (1998a, 1998b). Result (A.1) is obtained upon substituting $\varepsilon' B_2\varepsilon = \varepsilon'[\frac{1}{2}(B_2 + B_2')]\varepsilon = \varepsilon' A_2\varepsilon$, where A_2 is symmetric, in (A.1) of KP (1998a). This substitution also enables to prove (A.4) from KP (1998a, A.5), and (A.9) from KP (1998a, A.8) and (A.10) from KP (1998a, A.11). Result (A.2) follows from (A.1) and from $\mathsf{E}(\varepsilon' B_1\varepsilon) = \operatorname{tr}(B_1)$. The proof of (A.3) is given in KP (1998a). Result (A.5) follows easily from (A.4). Results (A.6), (A.7) and (A.8) can be found in KP (1998a). Finally (A.11) follows easily from (A.10).

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

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

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

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

$$\begin{aligned} Z'u &= \bar{Z}'u + \tilde{Z}'u \\ &= \bar{Z}'(0, I_T)v + e_1 v' G'(0, I_T)v \\ &= \bar{Z}'(0, I_T)v + (v'Hv)e_1 = O_p(T^{1/2}), \end{aligned} \tag{B.2}$$

where H is the non-symmetric matrix

$$H = G'(0, I_T). \tag{B.3}$$

For H we find the useful results

$$\begin{aligned}
\text{tr}(H) &= \text{tr}[(0, I_T)'G] = \text{tr}[G(0, I_T)'] = \text{tr}(C) = 0 && \text{(B.4a)}\\
\text{tr}(HH) &= 0 && \text{(B.4b)}\\
\text{tr}(H'H) &= \text{tr}(G'G) && \text{(B.4c)}\\
G(H + H')(0, I_T)' &= GG' + CC. && \text{(B.4d)}
\end{aligned}$$

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

$$\mathsf{E}(\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)' = \mathsf{E}(Z'Z)^{-1}Z'uu'Z(Z'Z)^{-1}. \tag{B.5}$$

We first develop an expansion of $(Z'Z)^{-1}$. Referring to (2.7) and (2.8) we have $\mathsf{E}(Z'Z) = Q^{-1} = \bar{Z}'\bar{Z} + \mathsf{E}(\tilde{Z}'\tilde{Z})$, and so

$$\begin{aligned}
Z'Z &= (\bar{Z} + \tilde{Z})'(\bar{Z} + \tilde{Z}) && \text{(B.6)}\\
&= \mathsf{E}(Z'Z) - \mathsf{E}(\tilde{Z}'\tilde{Z}) + \bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z} + \tilde{Z}'\tilde{Z}\\
&= \{I_{K+1} + (\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q + [\tilde{Z}'\tilde{Z} - \mathsf{E}(\tilde{Z}'\tilde{Z})]Q\}Q^{-1}.
\end{aligned}$$

Hence,

$$(Z'Z)^{-1} = Q\{I_{K+1} + (\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q + [\tilde{Z}'\tilde{Z} - \mathsf{E}(\tilde{Z}'\tilde{Z})]Q\}^{-1}, \tag{B.7}$$

where the stochastic terms $(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q$ and $[\tilde{Z}'\tilde{Z} - \mathsf{E}(\tilde{Z}'\tilde{Z})]Q$ 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 $[I_n - A + A^2 - A^3 + ...]$, 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

$$\begin{aligned}
(Z'Z)^{-1} = \; & Q\{I_{K+1} - (\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q - [\tilde{Z}'\tilde{Z} - \mathsf{E}(\tilde{Z}'\tilde{Z})]Q && \text{(B.8)}\\
& +(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q + (\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q[\tilde{Z}'\tilde{Z} - \mathsf{E}(\tilde{Z}'\tilde{Z})]Q\\
& +[\tilde{Z}'\tilde{Z} - \mathsf{E}(\tilde{Z}'\tilde{Z})]Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q + [\tilde{Z}'\tilde{Z} - \mathsf{E}(\tilde{Z}'\tilde{Z})]Q[\tilde{Z}'\tilde{Z} - \mathsf{E}(\tilde{Z}'\tilde{Z})]Q\}\\
& +o_p(T^{-2}),
\end{aligned}$$

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

$$(Z'Z)^{-1} = Q - Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q - Q[\tilde{Z}'\tilde{Z} - \mathsf{E}(\tilde{Z}'\tilde{Z})]Q + o_p(T^{-3/2}), \tag{B.9}$$

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

$$(Z'Z)^{-1} = Q + o_p(T^{-1}). \tag{B.10}$$

The expansion (B.8) for $(Z'Z)^{-1}$ can be written as

$$(Z'Z)^{-1} = Q(I_{K+1} - W_1 - W_2 + W_1W_1 + W_1W_2 + W_2W_1 + W_2W_2) + o_p(T^{-2}) \tag{B.11}$$

where we introduced some further shorthand notation, viz.

$$W_1 = (\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q = \bar{Z}'Gvq_1' + e_1v'G'\bar{Z}Q = O_p(T^{-1/2}) \tag{B.12}$$

and

$$W_2 = [\tilde{Z}'\tilde{Z} - \mathsf{E}(\tilde{Z}'\tilde{Z})]Q = [v'G'Gv - \sigma^2 \operatorname{tr}(G'G)]e_1 q_1' = O_p(T^{-1/2}). \tag{B.13}$$

Note that after premultiplication by Q we have seven terms in (B.11). 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:

$$\begin{aligned}
(\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)' &= (Z'Z)^{-1}Z'uu'Z(Z'Z)^{-1} \\
&= QZ'uu'ZQ - Q[W_1 + W_2 - (W_1 + W_2)^2]Z'uu'ZQ \\
&\quad -QZ'uu'Z[W_1' + W_2' - (W_1' + W_2')^2]Q \\
&\quad +Q(W_1 + W_2)Z'uu'Z(W_1' + W_2')Q + o_p(T^{-2}) \\
&= QZ'uu'ZQ - QW_1Z'uu'ZQ - QW_2Z'uu'ZQ + Q(W_1 + W_2)^2 Z'uu'ZQ \\
&\quad -QZ'uu'ZW_1'Q - QZ'uu'ZW_2'Q + QZ'uu'Z(W_1' + W_2')^2 Q \\
&\quad +QW_1Z'uu'ZW_1'Q + QW_1Z'uu'ZW_2'Q + QW_2Z'uu'ZW_1'Q \\
&\quad +QW_2Z'uu'ZW_2'Q + o_p(T^{-2}).
\end{aligned} \tag{B.14}$$

Note that

$$\begin{aligned}
Z'uu'Z &= [\bar{Z}'(0, I_T)v + (v'Hv)e_1][v'(0, I_T)'\bar{Z} + (v'Hv)e_1'] \\
&= \bar{Z}'(0, I_T)vv'(0, I_T)'\bar{Z} + \bar{Z}'(0, I_T)v(v'Hv)e_1' \\
&\quad +(v'Hv)e_1v'(0, I_T)'\bar{Z} + (v'Hv)^2 e_1e_1'.
\end{aligned} \tag{B.15}$$

We now derive the expectation of the eleven terms of (B.14). For the first one we obtain

$$\begin{aligned}
\mathsf{E}\left(QZ'uu'ZQ\right) &= \mathsf{E}Q\bar{Z}'(0, I_T)vv'(0, I_T)'\bar{Z}Q + \mathsf{E}Q(v'Hv)^2 e_1e_1'Q \\
&= \sigma^2 Q\bar{Z}'\bar{Z}Q + \sigma^2 \operatorname{tr}(G'G)Qe_1e_1'Q \\
&= \sigma^2 Q.
\end{aligned}$$

For the expectation of the second term of (B.14) we find

$$\begin{aligned}
&\mathsf{E}\left(QW_1Z'uu'ZQ\right) \\
&= \mathsf{E}[Q\bar{Z}'Gvq_1'\bar{Z}'(0, I_T)v(v'Hv)q_1'] + \mathsf{E}[Q\bar{Z}'Gvq_1'(v'Hv)e_1v'(0, I_T)'\bar{Z}Q] \\
&\quad +\mathsf{E}[q_1v'G'\bar{Z}Q\bar{Z}'(0, I_T)v(v'Hv)q_1'] + \mathsf{E}[q_1v'G'\bar{Z}Q(v'Hv)e_1v'(0, I_T)'\bar{Z}Q] \\
&= Q\bar{Z}'G\mathsf{E}(vv'Hvv')(0, I_T)'\bar{Z}q_1q_1' + q_{11}Q\bar{Z}'G\mathsf{E}(vv'Hvv')(0, I_T)'\bar{Z}Q \\
&\quad +\mathsf{E}[v'G'\bar{Z}Q\bar{Z}'(0, I_T)v](v'Hv)q_1q_1' + q_1q_1'\bar{Z}'G\mathsf{E}(vv'Hvv')(0, I_T)'\bar{Z}Q \\
&= \sigma^4 Q\bar{Z}'G(H + H')(0, I_T)'\bar{Z}q_1q_1' + \sigma^4 q_{11}Q\bar{Z}'G(H + H')(0, I_T)'\bar{Z}Q \\
&\quad +\sigma^4 q_1q_1'\bar{Z}'G(H + H')(0, I_T)'\bar{Z}Q + \sigma^4 \operatorname{tr}[G'\bar{Z}Q\bar{Z}'(0, I_T)(H + H')])q_1q_1' \\
&= \sigma^4 Q\bar{Z}'\left(GG' + CC\right)\bar{Z}q_1q_1' + \sigma^4 q_{11}Q\bar{Z}'\left(GG' + CC\right)\bar{Z}Q \\
&\quad +\sigma^4 q_1q_1'\bar{Z}'\left(GG' + CC\right)\bar{Z}Q + \sigma^4 \operatorname{tr}(Q\bar{Z}'CC\bar{Z})q_1q_1' + \sigma^4 \operatorname{tr}(Q\bar{Z}'GG'\bar{Z})q_1q_1'.
\end{aligned} \tag{B.16}$$

The expectation of the third term of (B.14) is

$$\begin{aligned}
&\mathsf{E}\left(QW_2Z'uu'ZQ\right) \\
&= \mathsf{E}[v'G'Gv - \sigma^2 \operatorname{tr}(G'G)]q_1q_1'\bar{Z}'(0, I_T)vv'(0, I_T)'\bar{Z}Q \\
&\quad +\mathsf{E}[v'G'Gv - \sigma^2 \operatorname{tr}(G'G)]q_1q_1'(v'Hv)^2 e_1e_1'Q \\
&= 2\sigma^4 q_1q_1'\bar{Z}'(0, I_T)G'G(0, I_T)'\bar{Z}Q \\
&\quad +2\sigma^6 q_{11}\left[2\operatorname{tr}\left(G'GHH\right) + \operatorname{tr}\left(G'GH'H\right) + \operatorname{tr}\left(G'GHH'\right)\right]q_1q_1' \\
&= 2\sigma^4 q_1q_1'\bar{Z}'C'C\bar{Z}Q + 2\sigma^6 q_{11}\left[2\operatorname{tr}\left(GG'CC\right) + \operatorname{tr}\left(GG'C'C\right) + \operatorname{tr}\left(GG'GG'\right)\right]q_1q_1'.
\end{aligned} \tag{B.17}$$

For the fourth term of (B.14) we have

$$\mathsf{E}[Q(W_1+W_2)^2 Z'uu'ZQ] = \mathsf{E}\left(QW_1W_1Z'uu'ZQ\right) + \mathsf{E}\left(QW_1W_2Z'uu'ZQ\right) + \mathsf{E}\left(QW_2W_1Z'uu'ZQ\right) + \mathsf{E}\left(QW_2W_2Z'uu'ZQ\right). \tag{B.18}$$

We examine these four terms separately. First we have

$$\begin{aligned}
& \mathsf{E}\left(QW_1W_1Z'uu'ZQ\right) \\
= \; & \mathsf{E}[Q(\bar{Z}'Gvq_1' + e_1v'G'\bar{Z}Q)(\bar{Z}'Gvq_1' + e_1v'G'\bar{Z}Q)Z'uu'ZQ] \\
= \; & \mathsf{E}[Q\bar{Z}'Gvv'G'\bar{Z}q_1q_1'\bar{Z}'(0,I_T)vv'(0,I_T)'\bar{Z}Q] + q_{11}\mathsf{E}[Q\bar{Z}'Gv(v'Hv)^2v'G'\bar{Z}q_1q_1'] \\
& + q_{11}\mathsf{E}[Q\bar{Z}'Gvv'G'\bar{Z}Q\bar{Z}'(0,I_T)vv'(0,I_T)'\bar{Z}Q] + q_{11}\mathsf{E}[Q\bar{Z}'Gv(v'Hv)^2v'G'\bar{Z}q_1q_1'] \\
& + \mathsf{E}[q_1v'G'\bar{Z}Q\bar{Z}'Gvq_1'\bar{Z}'(0,I_T)vv'(0,I_T)'\bar{Z}Q] + q_{11}\mathsf{E}[q_1v'G'\bar{Z}Q\bar{Z}'Gv(v'Hv)^2q_1'] \\
& + \mathsf{E}[q_1q_1'\bar{Z}'Gvv'G'\bar{Z}Q\bar{Z}'(0,I_T)vv'(0,I_T)'\bar{Z}Q] + \mathsf{E}[q_1q_1'\bar{Z}'Gv(v'Hv)^2v'G'\bar{Z}q_1q_1'] \\
= \; & \sigma^4(q_1'\bar{Z}'C\bar{Z}q_1)Q\bar{Z}'C\bar{Z}Q + \sigma^4Q\bar{Z}'GG'\bar{Z}q_1q_1'\bar{Z}'\bar{Z}Q + \sigma^4Q\bar{Z}'C\bar{Z}q_1q_1'\bar{Z}'C\bar{Z}Q \\
& + 2\sigma^6q_{11}\operatorname{tr}(G'G)\,Q\bar{Z}'GG'\bar{Z}q_1q_1' + 4\sigma^6q_{11}Q\bar{Z}'GG'C'C'\bar{Z}q_1q_1' \\
& + 4\sigma^6q_{11}Q\bar{Z}'GCGC'\bar{Z}q_1q_1' + 4\sigma^6q_{11}Q\bar{Z}'GG'GG'\bar{Z}q_1q_1' + 4\sigma^6q_{11}Q\bar{Z}'CCGG'\bar{Z}q_1q_1' \\
& + \sigma^4q_{11}\operatorname{tr}(Q\bar{Z}'C\bar{Z})Q\bar{Z}'C\bar{Z}Q + \sigma^4q_{11}Q\bar{Z}'GG'\bar{Z}Q\bar{Z}'\bar{Z}Q + \sigma^4q_{11}Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}Q \\
& + \sigma^4\operatorname{tr}(Q\bar{Z}'GG'\bar{Z})q_1q_1'\bar{Z}'\bar{Z}Q + 2\sigma^4q_1q_1'\bar{Z}'C'\bar{Z}Q\bar{Z}'C\bar{Z}Q \\
& + \sigma^6q_{11}\operatorname{tr}(Q\bar{Z}'GG'\bar{Z})\operatorname{tr}(G'G)q_1q_1' + 4\sigma^6q_{11}\operatorname{tr}(Q\bar{Z}'CCGG'\bar{Z})q_1q_1' \\
& + 2\sigma^6q_{11}\operatorname{tr}(Q\bar{Z}'GCGC'\bar{Z})q_1q_1' + 2\sigma^6q_{11}\operatorname{tr}(Q\bar{Z}'GG'GG'\bar{Z})q_1q_1' \\
& + \sigma^4\operatorname{tr}(Q\bar{Z}'C\bar{Z})q_1q_1'\bar{Z}'C\bar{Z}Q + \sigma^4q_1q_1'\bar{Z}'GG'\bar{Z}Q\bar{Z}'\bar{Z}Q + \sigma^4q_1q_1'\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}Q \\
& + \sigma^6(q_1'\bar{Z}'GG'\bar{Z}q_1)\operatorname{tr}(G'G)\,q_1q_1' + 4\sigma^6(q_1'\bar{Z}'CCGG'\bar{Z}q_1)q_1q_1' \\
& + 2\sigma^6(q_1'\bar{Z}'GCGC'\bar{Z}q_1)q_1q_1' + 2\sigma^6(q_1'\bar{Z}'GG'GG'\bar{Z}q_1)q_1q_1'.
\end{aligned}$$

Various terms are $o(T^{-2})$ here. Removing these and using $\bar{Z}'\bar{Z}Q = I - \sigma^2\operatorname{tr}(G'G)e_1q_1'$ gives

$$\begin{aligned}
& \mathsf{E}\left(QW_1W_1Z'uu'ZQ\right) = \\
& \quad \sigma^4(q_1'\bar{Z}'C\bar{Z}q_1)Q\bar{Z}'C\bar{Z}Q + \sigma^4Q\bar{Z}'GG'\bar{Z}q_1q_1' + \sigma^4Q\bar{Z}'C\bar{Z}q_1q_1'\bar{Z}'C\bar{Z}Q \\
& \quad + \sigma^4q_{11}\operatorname{tr}(Q\bar{Z}'C\bar{Z})Q\bar{Z}'C\bar{Z}Q + \sigma^4q_{11}Q\bar{Z}'GG'\bar{Z}Q + \sigma^4q_{11}Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}Q \\
& \quad + \sigma^4\operatorname{tr}(Q\bar{Z}'GG'\bar{Z})q_1q_1' + \sigma^4\operatorname{tr}(Q\bar{Z}'C\bar{Z})q_1q_1'\bar{Z}'C\bar{Z}Q + \sigma^4q_1q_1'\bar{Z}'GG'\bar{Z}Q \\
& \quad + 2\sigma^4q_1q_1'\bar{Z}'C'\bar{Z}Q\bar{Z}'C\bar{Z}Q + \sigma^4q_1q_1'\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}Q + o(T^{-2}).
\end{aligned} \tag{B.19}$$

For the second term of (B.18) we find (we immediately remove terms of small order)

$$\begin{aligned}
& \mathsf{E}\left(QW_1W_2Z'uu'ZQ\right) \\
= \; & \mathsf{E}\{Q(\bar{Z}'Gvq_1' + e_1v'G'\bar{Z}Q)[v'G'Gv - \sigma^2\operatorname{tr}(G'G)]e_1q_1'\bar{Z}'(0,I_T)v(v'Hv)q_1'\} \\
& + \mathsf{E}\{Q(\bar{Z}'Gvq_1' + e_1v'G'\bar{Z}Q)[v'G'Gv - \sigma^2\operatorname{tr}(G'G)]e_1q_1'(v'Hv)e_1v'(0,I_T)'\bar{Z}Q\} \\
= \; & q_{11}\mathsf{E}\{Q\bar{Z}'G[v'G'Gv - \sigma^2\operatorname{tr}(G'G)](v'Hv)vv'(0,I_T)'\bar{Z}q_1q_1'\} \\
& + \mathsf{E}\{q_1q_1'\bar{Z}'G[v'G'Gv - \sigma^2\operatorname{tr}(G'G)](v'Hv)vv'(0,I_T)'\bar{Z}q_1q_1'\} \\
& + q_{11}^2\mathsf{E}\{Q\bar{Z}'G[v'G'Gv - \sigma^2\operatorname{tr}(G'G)](v'Hv)vv'(0,I_T)'\bar{Z}Q\} \\
& + q_{11}\mathsf{E}\{q_1q_1'\bar{Z}'G[v'G'Gv - \sigma^2\operatorname{tr}(G'G)](v'Hv)vv'(0,I_T)'\bar{Z}Q\} \\
= \; & 2\sigma^6q_{11}\operatorname{tr}(GG'C)Q\bar{Z}'C\bar{Z}q_1q_1' + 2\sigma^6\operatorname{tr}(GG'C)(q_1'\bar{Z}'C\bar{Z}q_1)q_1q_1' \\
& + 2\sigma^6q_{11}^2\operatorname{tr}(GG'C)Q\bar{Z}'C\bar{Z}Q + 2\sigma^6q_{11}\operatorname{tr}(GG'C)q_1q_1'\bar{Z}'C\bar{Z}Q + o(T^{-2}).
\end{aligned} \tag{B.20}$$

Next we examine the third term of (B.18) and find

$$\begin{aligned}
& \mathsf{E}\,(QW_2W_1Z'uu'ZQ) && \text{(B.21)}\\
= \; & \mathsf{E}\{[v'G'Gv - \sigma^2\operatorname{tr}(G'G)]q_1q_1'(\bar{Z}'Gvq_1' + e_1v'G'\bar{Z}Q)\bar{Z}'(0,I_T)v(v'Hv)q_1'\}\\
& + \mathsf{E}\{[v'G'Gv - \sigma^2\operatorname{tr}(G'G)]q_1q_1'(\bar{Z}'Gvq_1' + e_1v'G'\bar{Z}Q)(v'Hv)e_1v'(0,I_T)'\bar{Z}Q\}\\
= \; & \mathsf{E}\{[v'G'Gv - \sigma^2\operatorname{tr}(G'G)]q_1q_1'\bar{Z}'Gvq_1'\bar{Z}'(0,I_T)v(v'Hv)q_1'\}\\
& + \mathsf{E}\{[v'G'Gv - \sigma^2\operatorname{tr}(G'G)]q_1q_1'e_1v'G'\bar{Z}Q\bar{Z}'(0,I_T)v(v'Hv)q_1'\}\\
& + \mathsf{E}\{[v'G'Gv - \sigma^2\operatorname{tr}(G'G)]q_1q_1'\bar{Z}'Gvq_1'(v'Hv)e_1v'(0,I_T)'\bar{Z}Q\}\\
& + \mathsf{E}\{[v'G'Gv - \sigma^2\operatorname{tr}(G'G)]q_1q_1'e_1v'G'\bar{Z}Q(v'Hv)e_1v'(0,I_T)'\bar{Z}Q\}\\
= \; & \mathsf{E}\{q_1q_1'\bar{Z}'G[v'G'Gv - \sigma^2\operatorname{tr}(G'G)](v'Hv)vv'(0,I_T)'\bar{Z}q_1q_1'\}\\
& + q_{11}\mathsf{E}\{[v'G'Gv - \sigma^2\operatorname{tr}(G'G)]v'G'\bar{Z}Q\bar{Z}'(0,I_T)v(v'Hv)q_1q_1'\}\\
& + 2q_{11}\mathsf{E}\{q_1q_1'\bar{Z}'G[v'G'Gv - \sigma^2\operatorname{tr}(G'G)](v'Hv)vv'(0,I_T)'\bar{Z}Q\}\\
= \; & 2\sigma^6\operatorname{tr}(GG'C)(q_1'\bar{Z}'C\bar{Z}q_1)q_1q_1' + 2\sigma^6q_{11}\operatorname{tr}(GG'C)\operatorname{tr}(Q\bar{Z}'C\bar{Z})q_1q_1'\\
& + 4\sigma^6q_{11}\operatorname{tr}(GG'C)q_1q_1'\bar{Z}'C\bar{Z}Q + o(T^{-2}).
\end{aligned}$$

For the fourth term of (B.18) we obtain

$$\begin{aligned}
& \mathsf{E}\,(QW_2W_2Z'uu'ZQ) && \text{(B.22)}\\
= \; & q_{11}\mathsf{E}\{q_1q_1'\bar{Z}'(0,I_T)[v'G'Gv - \sigma^2\operatorname{tr}(G'G)]^2vv'(0,I_T)'\bar{Z}Q\}\\
& + q_{11}^2\mathsf{E}\{[v'G'Gv - \sigma^2\operatorname{tr}(G'G)]^2(v'Hv)^2\}q_1q_1'\\
= \; & 2\sigma^6q_{11}\operatorname{tr}(G'GG'G)q_1q_1'\bar{Z}'\bar{Z}Q + 8\sigma^6q_{11}q_1q_1'\bar{Z}'C'GG'C\bar{Z}Q\\
& + 2\sigma^8q_{11}^2\operatorname{tr}(G'G)\operatorname{tr}(G'GG'G)q_1q_1' + 8\sigma^8q_{11}^2\operatorname{tr}(GG'C)\operatorname{tr}(GG'C)q_1q_1' + o(T^{-2})\\
= \; & 2\sigma^6q_{11}\operatorname{tr}(G'GG'G)q_1q_1'\bar{Z}'\bar{Z}Q + 2\sigma^8q_{11}^2\operatorname{tr}(G'G)\operatorname{tr}(G'GG'G)q_1q_1'\\
& + 8\sigma^8q_{11}^2\operatorname{tr}(GG'C)\operatorname{tr}(GG'C)q_1q_1' + o(T^{-2})\\
= \; & 2\sigma^6q_{11}\operatorname{tr}(G'GG'G)q_1q_1' + 8\sigma^8q_{11}^2\operatorname{tr}(GG'C)\operatorname{tr}(GG'C)q_1q_1' + o(T^{-2}).
\end{aligned}$$

Collecting the four terms of (B.18), i.e. the expectation of the fourth term of (B.14), we get

$$\begin{aligned}
& \mathsf{E}[Q(W_1+W_2)^2Z'uu'ZQ] && \text{(B.23)}\\
= \; & \sigma^4(q_1'\bar{Z}'C\bar{Z}q_1)Q\bar{Z}'C\bar{Z}Q + \sigma^4Q\bar{Z}'GG'\bar{Z}q_1q_1' + \sigma^4Q\bar{Z}'C\bar{Z}q_1q_1'\bar{Z}'C\bar{Z}Q\\
& + \sigma^4q_{11}\operatorname{tr}(Q\bar{Z}'C\bar{Z})Q\bar{Z}'C\bar{Z}Q + \sigma^4q_{11}Q\bar{Z}'GG'\bar{Z}Q + \sigma^4q_{11}Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}Q\\
& + \sigma^4\operatorname{tr}(Q\bar{Z}'GG'\bar{Z})q_1q_1' + \sigma^4\operatorname{tr}(Q\bar{Z}'C\bar{Z})q_1q_1'\bar{Z}'C\bar{Z}Q\\
& + 2\sigma^4q_1q_1'\bar{Z}'C'\bar{Z}Q\bar{Z}'C\bar{Z}Q + \sigma^4q_1q_1'\bar{Z}'GG'\bar{Z}Q + \sigma^4q_1q_1'\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}Q\\
& + 2\sigma^6q_{11}\operatorname{tr}(GG'C)Q\bar{Z}'C\bar{Z}q_1q_1' + 4\sigma^6\operatorname{tr}(GG'C)(q_1'\bar{Z}'C\bar{Z}q_1)q_1q_1'\\
& + 2\sigma^6q_{11}^2\operatorname{tr}(GG'C)Q\bar{Z}'C\bar{Z}Q + 6\sigma^6q_{11}\operatorname{tr}(GG'C)q_1q_1'\bar{Z}'C\bar{Z}Q\\
& + 2\sigma^6q_{11}\operatorname{tr}(GG'C)\operatorname{tr}(Q\bar{Z}'C\bar{Z})q_1q_1' + 2\sigma^6q_{11}\operatorname{tr}(G'GG'G)q_1q_1'\\
& + 8\sigma^8q_{11}^2\operatorname{tr}(GG'C)\operatorname{tr}(GG'C)q_1q_1' + o(T^{-2}).
\end{aligned}$$

For the expectation of the fifth term of (B.14) we find

$$\mathsf{E}\,(QZ'uu'ZW_1'Q) = \mathsf{E}\,(QW_1Z'uu'ZQ)', \tag{B.24}$$

which is just the transpose of the result for the second term (B.16). For the sixth term of (B.14) we find

$$\mathsf{E}\,(QZ'uu'ZW_2'Q) = \mathsf{E}\,(QW_2Z'uu'ZQ)', \tag{B.25}$$

which follows easily from (B.17). Likewise the expectation of the seventh term of (B.14) equals the transpose of (B.23), hence

$$\mathsf{E}[QZ'uu'Z(W_1'+W_2')^2Q] = \mathsf{E}[Q(W_1+W_2)^2Z'uu'ZQ']. \tag{B.26}$$

The expectation of the eighth term of (B.14) is

$$\begin{aligned}
& \mathsf{E}\,(QW_1Z'uu'ZW_1'Q) \\
= \; & \mathsf{E}[Q(\bar{Z}'Gvq_1' + e_1v'G'\bar{Z}Q)\bar{Z}'(0,I_T)vv'(0,I_T)'\bar{Z}(q_1v'G'\bar{Z} + Q\bar{Z}'Gve_1')Q] \\
& + \mathsf{E}[Q(\bar{Z}'Gvq_1' + e_1v'G'\bar{Z}Q)(v'Hv)^2e_1e_1'(q_1v'G'\bar{Z} + Q\bar{Z}'Gve_1')Q] \\
= \; & \mathsf{E}[Q\bar{Z}'Gvq_1'\bar{Z}'(0,I_T)vv'(0,I_T)'\bar{Z}q_1v'G'\bar{Z}Q] \\
& + \mathsf{E}[Q\bar{Z}'Gvq_1'\bar{Z}'(0,I_T)vv'(0,I_T)'\bar{Z}Q\bar{Z}'Gvq_1'] \\
& + \mathsf{E}[q_1v'G'\bar{Z}Q\bar{Z}'(0,I_T)vv'(0,I_T)'\bar{Z}q_1v'G'\bar{Z}Q] \\
& + \mathsf{E}[q_1v'G'\bar{Z}Q\bar{Z}'(0,I_T)vv'(0,I_T)'\bar{Z}Q\bar{Z}'Gvq_1'] \\
& + q_{11}^2\mathsf{E}[Q\bar{Z}'Gv(v'Hv)^2v'G'\bar{Z}Q] + q_{11}\mathsf{E}[Q\bar{Z}'Gv(v'Hv)^2q_1'\bar{Z}'Gvq_1'] \\
& + q_{11}\mathsf{E}[q_1v'G'\bar{Z}q_1(v'Hv)^2v'G'\bar{Z}Q] + \mathsf{E}[q_1v'G'\bar{Z}(v'Hv)^2q_1q_1'\bar{Z}'Gvq_1'] \\
= \; & \mathsf{E}[Q\bar{Z}'Gvv'(0,I_T)'\bar{Z}q_1q_1'\bar{Z}'(0,I_T)vv'G'\bar{Z}Q] \\
& + \mathsf{E}[Q\bar{Z}'Gvv'(0,I_T)'\bar{Z}Q\bar{Z}'Gvv'(0,I_T)'\bar{Z}q_1q_1'] \\
& + \mathsf{E}[q_1q_1'\bar{Z}'(0,I_T)vv'G'\bar{Z}Q\bar{Z}'(0,I_T)vv'G'\bar{Z}Q] \\
& + \mathsf{E}[v'G'\bar{Z}Q\bar{Z}'(0,I_T)vv'(0,I_T)'\bar{Z}Q\bar{Z}'Gvq_1q_1'] \\
& + q_{11}^2\mathsf{E}[Q\bar{Z}'Gv(v'Hv)^2v'G'\bar{Z}Q] + q_{11}\mathsf{E}[Q\bar{Z}'Gv(v'Hv)^2v'G'\bar{Z}q_1q_1'] \\
& + q_{11}\mathsf{E}[q_1q_1'\bar{Z}'Gv(v'Hv)^2v'G'\bar{Z}Q] + \mathsf{E}[q_1q_1'\bar{Z}'Gv(v'Hv)^2v'G'\bar{Z}q_1q_1'] \\
= \; & \sigma^4(q_1'\bar{Z}'\bar{Z}q_1)Q\bar{Z}'GG'\bar{Z}Q + 2\sigma^4Q\bar{Z}'C\bar{Z}q_1q_1'\bar{Z}'C'\bar{Z}Q + \sigma^4\operatorname{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}'GG'\bar{Z}Q\bar{Z}'\bar{Z}q_1q_1' + \sigma^4\operatorname{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}'GG'\bar{Z}Q + \sigma^4\operatorname{tr}(Q\bar{Z}'C\bar{Z})\operatorname{tr}(Q\bar{Z}'C\bar{Z})q_1q_1' \\
& + \sigma^4\operatorname{tr}(Q\bar{Z}'\bar{Z}Q\bar{Z}'GG'\bar{Z})q_1q_1' + \sigma^4\operatorname{tr}(Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z})q_1q_1' \\
& + \sigma^6q_{11}^2\operatorname{tr}(G'G)Q\bar{Z}'GG'\bar{Z}Q + \sigma^6q_{11}\operatorname{tr}(G'G)Q\bar{Z}'GG'\bar{Z}q_1q_1' \\
& + \sigma^6q_{11}\operatorname{tr}(G'G)q_1q_1'\bar{Z}'GG'\bar{Z}Q + \sigma^6\operatorname{tr}(G'G)(q_1'\bar{Z}'GG'\bar{Z}q_1)q_1q_1' + o(T^{-2}).
\end{aligned}$$

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

$$\begin{aligned}
& \mathsf{E}\,(QW_1Z'uu'ZW_1'Q) \qquad\qquad \text{(B.27)} \\
= \; & \sigma^4q_{11}Q\bar{Z}'GG'\bar{Z}Q + 2\sigma^4Q\bar{Z}'C\bar{Z}q_1q_1'\bar{Z}'C'\bar{Z}Q + \sigma^4\operatorname{tr}(Q\bar{Z}'C\bar{Z})Q\bar{Z}'C\bar{Z}q_1q_1' \\
& + \sigma^4Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}q_1q_1' + \sigma^4Q\bar{Z}'GG'\bar{Z}q_1q_1' + \sigma^4\operatorname{tr}(Q\bar{Z}'C\bar{Z})q_1q_1'\bar{Z}'C'\bar{Z}Q \\
& + \sigma^4q_1q_1'\bar{Z}'C'\bar{Z}Q\bar{Z}'C'\bar{Z}Q + \sigma^4q_1q_1'\bar{Z}'GG'\bar{Z}Q + \sigma^4\operatorname{tr}(Q\bar{Z}'C\bar{Z})\operatorname{tr}(Q\bar{Z}'C\bar{Z})q_1q_1' \\
& + \sigma^4\operatorname{tr}(Q\bar{Z}'GG'\bar{Z})q_1q_1' + \sigma^4\operatorname{tr}(Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z})q_1q_1' + o(T^{-2}).
\end{aligned}$$

For the expectation of the ninth term of (B.14) we find

$$\begin{aligned}
& \mathsf{E}\,(QW_1Z'uu'ZW_2'Q) \qquad\qquad \text{(B.28)} \\
= \; & \mathsf{E}\{Q(\bar{Z}'Gvq_1' + e_1v'G'\bar{Z}Q)\bar{Z}'(0,I_T)v(v'Hv)e_1'[v'G'Gv - \sigma^2\operatorname{tr}(G'G)]q_1q_1'\} \\
& + \mathsf{E}\{Q(\bar{Z}'Gvq_1' + e_1v'G'\bar{Z}Q)(v'Hv)e_1v'(0,I_T)'\bar{Z}[v'G'Gv - \sigma^2\operatorname{tr}(G'G)]q_1q_1'\} \\
= \; & q_{11}\mathsf{E}\{Q\bar{Z}'Gvq_1'\bar{Z}'(0,I_T)v(v'Hv)[v'G'Gv - \sigma^2\operatorname{tr}(G'G)]q_1'\} \\
& + q_{11}\mathsf{E}\{v'G'\bar{Z}Q\bar{Z}'(0,I_T)v(v'Hv)[v'G'Gv - \sigma^2\operatorname{tr}(G'G)]q_1q_1'\} \\
& + q_{11}\mathsf{E}\{Q\bar{Z}'Gv(v'Hv)v'(0,I_T)'\bar{Z}[v'G'Gv - \sigma^2\operatorname{tr}(G'G)]q_1q_1'\}
\end{aligned}$$

$$
\begin{aligned}
& +\mathsf{E}\{q_1 v'G'\bar{Z}q_1(v'Hv)v'(0,I_T)'\bar{Z}[v'G'Gv-\sigma^2\operatorname{tr}(G'G)]q_1q_1'\} \\
= \;& 2q_{11}\mathsf{E}\{Q\bar{Z}'Gvv'(v'Hv)[v'G'Gv-\sigma^2\operatorname{tr}(G'G)](0,I_T)'\bar{Z}q_1q_1'\} \\
& +q_{11}\mathsf{E}\{v'G'\bar{Z}Q\bar{Z}'(0,I_T)v(v'Hv)[v'G'Gv-\sigma^2\operatorname{tr}(G'G)]q_1q_1'\} \\
& +\mathsf{E}\{q_1q_1'\bar{Z}'Gvv'(v'Hv)[v'G'Gv-\sigma^2\operatorname{tr}(G'G)](0,I_T)'\bar{Z}q_1q_1'\} \\
= \;& 4\sigma^6 q_{11}\operatorname{tr}(GG'C)Q\bar{Z}'C\bar{Z}q_1q_1' + 2\sigma^6 q_{11}\operatorname{tr}(Q\bar{Z}'C\bar{Z})\operatorname{tr}(GG'C)q_1q_1' \\
& +2\sigma^6\operatorname{tr}(GG'C)(q_1'\bar{Z}'C\bar{Z}q_1)q_1q_1' + o(T^{-2}).
\end{aligned}
$$

We obtain for the expectation of the tenth term of (B.14)

$$
\mathsf{E}\,(QW_2Z'uu'ZW_1'Q) = \mathsf{E}\,(QW_1Z'uu'ZW_2'Q)', \tag{B.29}
$$

which is just the transpose of the former term. The expectation of the eleventh and final term of (B.14) is

$$
\begin{aligned}
& \mathsf{E}\,(QW_2Z'uu'ZW_2'Q) && \text{(B.30)} \\
= \;& \mathsf{E}\{q_1q_1'\bar{Z}'(0,I_T)v[v'G'Gv-\sigma^2\operatorname{tr}(G'G)]^2 v'(0,I_T)'\bar{Z}q_1q_1'\} \\
& +q_{11}^2\mathsf{E}\{[v'G'Gv-\sigma^2\operatorname{tr}(G'G)]^2(v'Hv)^2 q_1q_1'\} \\
= \;& 2\sigma^6\operatorname{tr}(G'GG'G)(q_1'\bar{Z}'\bar{Z}q_1)q_1q_1' + 2\sigma^8 q_{11}^2\operatorname{tr}(G'GG'G)\operatorname{tr}(G'G)q_1q_1' \\
& +8\sigma^8 q_{11}^2\operatorname{tr}(GG'C)\operatorname{tr}(GG'C)q_1q_1' + o(T^{-2}) \\
= \;& 2\sigma^6 q_{11}\operatorname{tr}(GG'GG')q_1q_1' + 8\sigma^8 q_{11}^2\operatorname{tr}(GG'C)\operatorname{tr}(GG'C)q_1q_1' + o(T^{-2}),
\end{aligned}
$$

where we used $q_1'\bar{Z}'\bar{Z}q_1 = q_{11} - \sigma^2 q_{11}^2\operatorname{tr}(G'G)$.

We may now assemble the various contributions to the mean squared error, and obtain after some simplification

$$
\begin{aligned}
& \mathrm{MSE}(\hat{\alpha}) = \mathsf{E}[(\hat{\alpha}-\alpha)(\hat{\alpha}-\alpha)'] = && \text{(B.31)} \\
& \sigma^2 Q + \\
& +\sigma^4 Q\bar{Z}'\left(GG'-CC-2C'C-C'C'\right)\bar{Z}q_1q_1' \\
& +\sigma^4 q_1q_1'\bar{Z}'\left(GG'-CC-2C'C-C'C'\right)\bar{Z}Q \\
& +\sigma^4 q_{11}Q\bar{Z}'\left(GG'-CC-C'C'\right)\bar{Z}Q + \sigma^4\operatorname{tr}(Q\bar{Z}'GG'\bar{Z})q_1q_1' + \sigma^4[\operatorname{tr}(Q\bar{Z}'C\bar{Z})]^2 q_1q_1' \\
& +\sigma^4\operatorname{tr}(Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z})q_1q_1' - 2\sigma^4\operatorname{tr}(Q\bar{Z}'CC\bar{Z})q_1q_1' \\
& +\sigma^4 q_{11}\operatorname{tr}(Q\bar{Z}'C\bar{Z})Q\bar{Z}'(C+C')\bar{Z}Q + \sigma^4(q_1'\bar{Z}'C\bar{Z}q_1)Q\bar{Z}'(C+C')\bar{Z}Q \\
& +\sigma^4 Q\bar{Z}'C\bar{Z}q_1q_1'\bar{Z}'(C+C')\bar{Z}Q + \sigma^4 Q\bar{Z}'(C+C')\bar{Z}q_1q_1'\bar{Z}'C'\bar{Z}Q \\
& +\sigma^4 q_{11}Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}Q + \sigma^4 q_{11}Q\bar{Z}'C'\bar{Z}Q\bar{Z}'C'\bar{Z}Q \\
& +\sigma^4\operatorname{tr}(Q\bar{Z}'C\bar{Z})q_1q_1'\bar{Z}'(C+C')\bar{Z}Q + \sigma^4\operatorname{tr}(Q\bar{Z}'C\bar{Z})Q\bar{Z}'(C+C')\bar{Z}q_1q_1' \\
& +\sigma^4 q_1q_1'\bar{Z}'(C+C')\bar{Z}Q\bar{Z}'C\bar{Z}Q + \sigma^4 q_1q_1'\bar{Z}'C'\bar{Z}Q\bar{Z}'(C+C')\bar{Z}Q \\
& +\sigma^4 Q\bar{Z}'(C+C')\bar{Z}Q\bar{Z}'C\bar{Z}q_1q_1' + \sigma^4 Q\bar{Z}'C'\bar{Z}Q\bar{Z}'(C+C')\bar{Z}q_1q_1' \\
& +6\sigma^6 q_{11}\operatorname{tr}(GG'C)Q\bar{Z}'(C+C')\bar{Z}q_1q_1' + 6\sigma^6 q_{11}\operatorname{tr}(GG'C)q_1q_1'\bar{Z}'(C+C')\bar{Z}Q \\
& +2\sigma^6 q_{11}^2\operatorname{tr}(GG'C)Q\bar{Z}'(C+C')\bar{Z}Q \\
& +12\sigma^6 q_1'\bar{Z}'C\bar{Z}q_1\operatorname{tr}(GG'C)q_1q_1' + 8\sigma^6 q_{11}\operatorname{tr}(GG'C)\operatorname{tr}(Q\bar{Z}'C\bar{Z})q_1q_1' \\
& +\sigma^6 q_{11}\left[2\operatorname{tr}\left(GG'GG'\right) - 8\operatorname{tr}\left(GG'CC\right) - 4\operatorname{tr}\left(GG'C'C\right)\right]q_1q_1' \\
& +24\sigma^8 q_{11}^2\operatorname{tr}(GG'C)\operatorname{tr}(GG'C)q_1q_1' + o(T^{-2}).
\end{aligned}
$$

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

$$
[\mathsf{E}(\hat{\alpha})-\alpha][\mathsf{E}(\hat{\alpha})-\alpha]' \tag{B.32}
$$

$$
\begin{aligned}
&= \sigma^4[\mathrm{tr}(Q\bar{Z}'C\bar{Z})q_1 + Q\bar{Z}'C\bar{Z}q_1 + 2\sigma^2 q_{11}\,\mathrm{tr}\,(GG'C)\,q_1] \times \\
&\quad [\mathrm{tr}(Q\bar{Z}'C\bar{Z})q_1 + Q\bar{Z}'C\bar{Z}q_1 + 2\sigma^2 q_{11}\,\mathrm{tr}\,(GG'C)\,q_1]' + o(T^{-2}) \\
&= \sigma^4\{[\mathrm{tr}(Q\bar{Z}'C\bar{Z})]^2 q_1 q_1' + Q\bar{Z}'C\bar{Z}q_1 q_1'\bar{Z}'C'\bar{Z}Q \\
&\qquad + \mathrm{tr}(Q\bar{Z}'C\bar{Z})[Q\bar{Z}'C\bar{Z}q_1 q_1' + q_1 q_1'\bar{Z}'C'\bar{Z}Q]\} \\
&\quad +\sigma^6\{4q_{11}\,\mathrm{tr}\,(GG'C)\,\mathrm{tr}(Q\bar{Z}'C\bar{Z})q_1 q_1' \\
&\qquad +2q_{11}\,\mathrm{tr}\,(GG'C)\,[Q\bar{Z}'C\bar{Z}q_1 q_1' + q_1 q_1'\bar{Z}'C'\bar{Z}Q]\} \\
&\quad +\sigma^8\{4q_{11}^2[\mathrm{tr}\,(GG'C)]^2 q_1 q_1'\} + o(T^{-2}).
\end{aligned}
$$

This result has to be subtracted from the MSE approximation (B.31) to find the required approximation to $\mathsf{V}(\hat{\alpha})$ of Theorem 2.2.

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

For the numerator of the estimator s^2, given in (1.5), we have, upon using (B.10),

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

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

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

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:

$$
\begin{aligned}
&\mathsf{E}[(\hat{\sigma}^2 - \sigma^2)(Z'Z)^{-1}] \\
= \;&\mathsf{E}[(T^{-1}u'u - \sigma^2)(Z'Z)^{-1}] \\
&-T^{-1}\mathsf{E}[(u'\bar{Z}Q\bar{Z}'u + u'\bar{Z}Q\tilde{Z}'u + u'\tilde{Z}Q\bar{Z}'u + u'\tilde{Z}Q\tilde{Z}'u)(Z'Z)^{-1}] + o(T^{-2}) \\
= \;&-\mathsf{E}(T^{-1}u'u - \sigma^2)Q[\tilde{Z}'\tilde{Z} - \mathsf{E}(\tilde{Z}'\tilde{Z})]Q - T^{-1}\mathsf{E}(u'\bar{Z}Q\bar{Z}'u + u'\tilde{Z}Q\tilde{Z}'u)Q + o(T^{-2}) \\
= \;&-\mathsf{E}(T^{-1}u'u - \sigma^2)Q[v'G'Gv - \sigma^2\,\mathrm{tr}(G'G)]e_1 e_1'Q \\
&-T^{-1}\mathsf{E}(u'\bar{Z}Q\bar{Z}'u + u'Gve_1'Qe_1v'G'u)Q + o(T^{-2}) \\
= \;&-\mathsf{E}(T^{-1}u'u - \sigma^2)(v'G'Gv)q_1 q_1' - T^{-1}\mathsf{E}(u'\bar{Z}Q\bar{Z}'u + q_{11}u'Gvv'G'u)Q + o(T^{-2}) \\
= \;&-T^{-1}[2\sigma^4\,\mathrm{tr}(C'C)q_1 q_1' + \sigma^2\,\mathrm{tr}(Q\bar{Z}'\bar{Z})Q + q_{11}\sigma^4\,\mathrm{tr}(G'G)Q] + o(T^{-2}) \\
= \;&-T^{-1}[\sigma^2(K+1)Q + 2\sigma^4\,\mathrm{tr}(C'C)q_1 q_1'] + o(T^{-2}).
\end{aligned} \tag{C.4}
$$

An approximation for the second right-hand term of (C.3) can be obtained from (B.8). 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

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

The second term of (C.5) is

$$
\begin{aligned}
& \mathsf{E}[Q(\bar{Z}'\tilde{Z}+\tilde{Z}'\bar{Z})Q(\bar{Z}'\tilde{Z}+\tilde{Z}'\bar{Z})Q] \qquad (C.6)\\
= \; & \mathsf{E}[Q(\bar{Z}'Gve_1'+e_1v'G'\bar{Z})Q(\bar{Z}'Gve_1'+e_1v'G'\bar{Z})Q]\\
= \; & \mathsf{E}[Q\bar{Z}'Gve_1'Q\bar{Z}'Gve_1'Q]+\mathsf{E}[Q\bar{Z}'Gve_1'Qe_1v'G'\bar{Z}Q]\\
& +\mathsf{E}[Qe_1v'G'\bar{Z}Q\bar{Z}'Gve_1'Q]+\mathsf{E}[Qe_1v'G'\bar{Z}Qe_1v'G'\bar{Z}Q]\\
= \; & \mathsf{E}[Q\bar{Z}'Gvv'G'\bar{Z}Qe_1e_1'Q]+q_{11}\mathsf{E}[Q\bar{Z}'Gvv'G'\bar{Z}Q]\\
& +\mathsf{E}[q_1v'G'\bar{Z}Q\bar{Z}'Gvq_1']+\mathsf{E}[q_1e_1'Q\bar{Z}'Gvv'G'\bar{Z}Q]\\
= \; & \sigma^2[Q\bar{Z}'GG'\bar{Z}q_1q_1'+q_{11}Q\bar{Z}'GG'\bar{Z}Q+\operatorname{tr}(Q\bar{Z}'GG'\bar{Z})q_1q_1'+q_1q_1'\bar{Z}'GG'\bar{Z}Q].
\end{aligned}
$$

The third term of (C.5) is

$$
\begin{aligned}
& \mathsf{E}\{Q[\tilde{Z}'\tilde{Z}-\mathsf{E}(\tilde{Z}'\tilde{Z})]Q[\tilde{Z}'\tilde{Z}-\mathsf{E}(\tilde{Z}'\tilde{Z})]Q\} \qquad (C.7)\\
= \; & \mathsf{E}\{Q[v'G'Gv-\sigma^2\operatorname{tr}(G'G)]e_1e_1'Q[v'G'Gv-\sigma^2\operatorname{tr}(G'G)]e_1e_1'Q\}\\
= \; & q_{11}[\mathsf{E}(v'G'Gvv'G'Gv)-2\sigma^2\operatorname{tr}(G'G)\mathsf{E}(v'G'Gv)+\sigma^4\operatorname{tr}(G'G)\operatorname{tr}(G'G)]q_1q_1'\\
= \; & 2\sigma^4q_{11}\operatorname{tr}(GG'GG')q_1q_1'.
\end{aligned}
$$

Gathering terms yields the result

$$
\begin{aligned}
\sigma^2\mathsf{E}[(Z'Z)^{-1}] \;=\; & \sigma^2Q \qquad (C.8)\\
& +\sigma^4[\operatorname{tr}(Q\bar{Z}'GG'\bar{Z})q_1q_1'+Q\bar{Z}'GG'\bar{Z}q_1q_1'+(q_1q_1'+q_{11}Q)\bar{Z}'GG'\bar{Z}Q]\\
& +2\sigma^6q_{11}\operatorname{tr}(GG'GG')q_1q_1'+o_p(T^{-2}).
\end{aligned}
$$

Adding up the terms (C.4) and (C.8) we obtain for (C.3) the approximation

$$
\begin{aligned}
\mathsf{E}[\hat{\sigma}^2(Z'Z)^{-1}] \;=\; & T^{-1}(T-K-1)\sigma^2Q \qquad (C.9)\\
& +\sigma^4\{[\operatorname{tr}(Q\bar{Z}'GG'\bar{Z})-2T^{-1}\operatorname{tr}(C'C)]q_1q_1'\\
& \quad +Q\bar{Z}'GG'\bar{Z}q_1q_1'+(q_1q_1'+q_{11}Q)\bar{Z}'GG'\bar{Z}Q\}\\
& +2\sigma^6q_{11}\operatorname{tr}(GG'GG')q_1q_1'+o(T^{-2}).
\end{aligned}
$$

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 (3.2) is given by

$$
\begin{aligned}
\mathsf{E}(\breve{\alpha}-\alpha) \;=\; & \mathsf{E}(\hat{\alpha}-\hat{\mathsf{B}}_\alpha-\alpha) \qquad (D.1)\\
=\; & \mathsf{E}(\mathsf{B}_\alpha-\hat{\mathsf{B}}_\alpha+\hat{\alpha}-\alpha-\mathsf{B}_\alpha)\\
=\; & -\mathsf{E}(\hat{\mathsf{B}}_\alpha-\mathsf{B}_\alpha)+o(T^{-1}).
\end{aligned}
$$

From (3.1) and (3.2) it follows that

$$
\begin{aligned}
\hat{\mathsf{B}}_\alpha-\mathsf{B}_\alpha \;=\; & \sigma^2\operatorname{tr}(Q\bar{Z}'C\bar{Z})q_1-s^2\operatorname{tr}(P\hat{Z}'\hat{C}\hat{Z})p_1 \qquad (D.2)\\
& +\sigma^2Q\bar{Z}'C\bar{Z}q_1-s^2P\hat{Z}'\hat{C}\hat{Z}p_1\\
& +2[\sigma^4q_{11}\operatorname{tr}(GG'C)q_1-s^4p_{11}\operatorname{tr}(\hat{C}\hat{C}'\hat{C})p_1].
\end{aligned}
$$

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

$$\begin{aligned} P &= Q + P^* + o_p(T^{-3/2}), \text{ with} \\ P^* &= -Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q - Q[\tilde{Z}'\tilde{Z} - \mathsf{E}(\tilde{Z}'\tilde{Z})]Q \\ &= -Q\bar{Z}'Gvq_1' - q_1v'G'\bar{Z}Q - [v'G'Gv - \sigma^2 \operatorname{tr}(G'G)]q_1q_1' = O_p(T^{-3/2}). \end{aligned} \tag{D.3}$$

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,

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

In fact, because it is easily verified that

$$\frac{\partial}{\partial\lambda}C = CC, \tag{D.5}$$

we can simply write

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

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

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

Further,

$$\begin{aligned} \hat{Z} - \bar{Z} &= (\hat{y}_{-1}, X) - (\bar{y}_{-1}, X) \\ &= (y_0\hat{F} + \hat{C}X\hat{\beta}, X) - (\bar{y}_0F + CX\beta, X) \\ &= (y_0\hat{F} - \bar{y}_0F + \hat{C}X\hat{\beta} - CX\beta)e_1' \\ &= [\bar{y}_0(\hat{F} - F) + \tilde{y}_0(\hat{F} - F + F) + (\hat{C} - C + C)X(\hat{\beta} - \beta + \beta) - CX\beta]e_1'. \end{aligned} \tag{D.8}$$

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

$$\begin{aligned} \hat{Z} - \bar{Z} &= \hat{Z}^* + o_p(T^{-1/2}), \text{ with} \\ \hat{Z}^* &= [\bar{y}_0(\hat{\lambda} - \lambda)CF + \tilde{y}_0(\hat{\lambda} - \lambda)CF + \tilde{y}_0F + CX(\hat{\beta} - \beta) + (\hat{\lambda} - \lambda)CCX\beta]e_1' \\ &= [C(\bar{y}_0F + CX\beta)(\hat{\lambda} - \lambda) + CX(\hat{\beta} - \beta) + \tilde{y}_0F + \tilde{y}_0(\hat{\lambda} - \lambda)CF]e_1' \\ &= [C\bar{Z}(\hat{\alpha} - \alpha) + \tilde{y}_0F + \tilde{y}_0(\hat{\lambda} - \lambda)CF]e_1' = O_p(T^{-1/2}). \end{aligned} \tag{D.9}$$

Now we obtain

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

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

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

Next we develop a result regarding $\tau = \mathrm{tr}(GG'C) = O(T)$ and $\hat{\tau} = \mathrm{tr}(\hat{C}\hat{C}'\hat{C})$. In KP (1998a) it has been shown that $\mathrm{tr}(GG'C) = \mathrm{tr}(CC'C) + \omega^2 F'CF = \mathrm{tr}(CC'C) + O(1)$. From (D.5) we find

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

because

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

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

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

Next we consider s^2. From KP (1998b) we have

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

and a Taylor expansion yields

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

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

For the first pair we find

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

For the second pair we obtain

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

and for the third

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

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

E. The variance of the COLS Estimator

The variance $\mathsf{V}(\check{\alpha})$ 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

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

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

$$\mathsf{E}(\hat{\mathsf{B}}_\alpha\hat{\mathsf{B}}_\alpha') = \mathsf{B}_\alpha\mathsf{B}_\alpha' + o(T^{-2}). \tag{E.2}$$

From

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

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

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

An approximation for $\mathsf{V}(\hat{\alpha})$ to order T^{-2} is given in Theorem 2.2. Hence, to establish an approximation to MSE$(\check{\alpha})$, i.e. to $\mathsf{V}(\check{\alpha})$, we have to find an approximation to order T^{-2} of $\mathsf{E}[(\hat{\mathsf{B}}_\alpha-\mathsf{B}_\alpha)(\hat{\alpha}-\alpha)']$ 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 $Q\bar{Z}'u + (v'Hv)q_1$, whereas the leading $O_p(T^{-3/2})$ terms of $(\hat{\mathsf{B}}_\alpha-\mathsf{B}_\alpha)$ have already been obtained in Appendix D, notably in the formulas (D.16), (D.17) and (D.18). Gathering these and regrouping we obtain

$$\begin{aligned}
& (\hat{\mathsf{B}}_\alpha-\mathsf{B}_\alpha)(\hat{\alpha}-\alpha)' && \text{(E.5)}\\
= \;& -[\sigma^2\,\mathrm{tr}(Q\bar{Z}'C\bar{Z})p_1^* + \sigma^2\,\mathrm{tr}(QA^*)q_1 + \sigma^2\,\mathrm{tr}(P^*\bar{Z}'C\bar{Z})q_1 + s_*^2\,\mathrm{tr}(Q\bar{Z}'C\bar{Z})q_1\\
& \quad +\sigma^2 Q\bar{Z}'C\bar{Z}p_1^* + \sigma^2 QA^*q_1 + \sigma^2 P^*\bar{Z}'C\bar{Z}q_1 + s_*^2 Q\bar{Z}'C\bar{Z}q_1\\
& \quad +2\sigma^4 q_{11}\tau p_1^* + 2\sigma^4 q_{11}\hat{\tau}^* q_1 + 2\sigma^4 p_{11}^*\tau q_1 + 4s_*^2\sigma^2 q_{11}\tau q_1](\hat{\alpha}-\alpha)' + o_p(T^{-2})\\
= \;& -\sigma^2\{[\mathrm{tr}(Q\bar{Z}'C\bar{Z})p_1^* + Q\bar{Z}'C\bar{Z}p_1^* + 2\sigma^2 q_{11}\tau p_1^*](u'\bar{Z}Q + v'Hvq_1')\\
& \quad +[\mathrm{tr}(QA^*)q_1 + QA^*q_1](u'\bar{Z}Q + v'Hvq_1')\\
& \quad +[\mathrm{tr}(P^*\bar{Z}'C\bar{Z})q_1 + P^*\bar{Z}'C\bar{Z}q_1](u'\bar{Z}Q + v'Hvq_1')\\
& \quad +[\mathrm{tr}(Q\bar{Z}'C\bar{Z})q_1 + Q\bar{Z}'C\bar{Z}q_1 + 4\sigma^2 q_{11}\tau q_1]s_*^2(u'\bar{Z}Q + v'Hvq_1')\\
& \quad +2\sigma^2[q_{11}q_1\hat{\tau}^*(u'\bar{Z}Q + v'Hvq_{11}') + \tau q_1 p_{11}^*(u'\bar{Z}Q + v'Hvq_1')]\} + o_p(T^{-2}).
\end{aligned}$$

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

$$\begin{aligned}
\mathsf{E}[p_1^*(u'\bar{Z}Q + v'Hvq_1')] &= -\mathsf{E}[q_{11}Q\bar{Z}'Gvv'(0, I_T)'\bar{Z}Q + q_1q_1'\bar{Z}'Gvv'(0, I_T)'\bar{Z}Q] \\
&\quad +\mathsf{E}\{q_{11}q_1q_1'[v'G'Gv - \sigma^2\operatorname{tr}(G'G)](v'Hv)\} \\
&= -\sigma^2[q_{11}Q\bar{Z}'C\bar{Z}Q + q_1q_1'\bar{Z}'C\bar{Z}Q + 2\sigma^2 q_{11}\operatorname{tr}(GG'C)q_1q_1'],
\end{aligned}$$

from which it follows that

$$\mathsf{E}[p_{11}^*(u'\bar{Z}Q + v'Hvq_1')] = -2\sigma^2[q_{11}q_1'\bar{Z}'C\bar{Z}Q + \sigma^2 q_{11}^2\operatorname{tr}(GG'C)q_1'].$$

Using (D.11) we find

$$\begin{aligned}
&\mathsf{E}\{[\operatorname{tr}(QA^*)q_1 + QA^*q_1](\hat{\alpha} - \alpha)'\} \\
= \;& \mathsf{E}[\operatorname{tr}(QA^*)q_1(\hat{\alpha} - \alpha)'] + \mathsf{E}[QA^*q_1(\hat{\alpha} - \alpha)'] \\
= \;& q_1q_1'\bar{Z}'(C + C')C\bar{Z}\mathsf{E}[(\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)'] + \operatorname{tr}(Q\bar{Z}'CC\bar{Z})q_1\mathsf{E}[(\hat{\lambda} - \lambda)(\hat{\alpha} - \alpha)'] \\
& +q_1q_1'\bar{Z}'C'C\bar{Z}\mathsf{E}[(\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)'] + q_{11}Q\bar{Z}'CC\bar{Z}\mathsf{E}[(\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)'] \\
& +Q\bar{Z}'CC\bar{Z}q_1\mathsf{E}[(\hat{\lambda} - \lambda)(\hat{\alpha} - \alpha)'] \\
= \;& \sigma^2[q_1q_1'\bar{Z}'(CC + 2C'C)\bar{Z}Q + q_{11}Q\bar{Z}'CC\bar{Z}Q + \operatorname{tr}(Q\bar{Z}'CC\bar{Z})q_1q_1' + Q\bar{Z}'CC\bar{Z}q_1q_1'] + o(T^{-2}).
\end{aligned}$$

and again substituting (D.3)

$$\begin{aligned}
&\mathsf{E}[\operatorname{tr}(P^*\bar{Z}'C\bar{Z})q_1 + P^*\bar{Z}'C\bar{Z}q_1](u'\bar{Z}Q + v'Hvq_1') \\
= \;& -\mathsf{E}\{\operatorname{tr}[Q\bar{Z}'Gvq_1'\bar{Z}'C\bar{Z} + q_1v'G'\bar{Z}Q\bar{Z}'C\bar{Z}]q_1u'\bar{Z}Q\} \\
& -\mathsf{E}\{[v'G'Gv - \sigma^2\operatorname{tr}(G'G)]\operatorname{tr}(q_1q_1'\bar{Z}'C\bar{Z})q_1v'Hvq_1'\} \\
& -\mathsf{E}[(Q\bar{Z}'Gvq_1' + q_1v'G'\bar{Z}Q)\bar{Z}'C\bar{Z}q_1u'\bar{Z}Q] \\
& -\mathsf{E}\{[v'G'Gv - \sigma^2\operatorname{tr}(G'G)]q_1q_1'\bar{Z}'C\bar{Z}q_1v'Hvq_1'\} \\
= \;& -\mathsf{E}[(q_1'\bar{Z}'C\bar{Z}Q\bar{Z}'Gv + v'G'\bar{Z}Q\bar{Z}'C\bar{Z}q_1)q_1u'\bar{Z}Q] \\
& -2\mathsf{E}\{[v'G'Gv - \sigma^2\operatorname{tr}(G'G)]v'Hvq_1'\bar{Z}'C\bar{Z}q_1q_1q_1'\} \\
& -\mathsf{E}(Q\bar{Z}'Gvq_1'\bar{Z}'C\bar{Z}q_1u'\bar{Z}Q + q_1v'G'\bar{Z}Q\bar{Z}'C\bar{Z}q_1u'\bar{Z}Q) \\
= \;& -\mathsf{E}\{q_1'\bar{Z}'C\bar{Z}Q\bar{Z}'Gvq_1u'\bar{Z}Q + v'G'\bar{Z}Q\bar{Z}'C\bar{Z}q_1q_1u'\bar{Z}Q\} - 4\sigma^4 q_1'\bar{Z}'C\bar{Z}q_1\operatorname{tr}(GG'C)q_1q_1' \\
& -\mathsf{E}\{q_1'\bar{Z}'C\bar{Z}q_1Q\bar{Z}'Gvu'\bar{Z}Q + q_1q_1'\bar{Z}'C'\bar{Z}Q\bar{Z}'Gvu'\bar{Z}Q\} \\
= \;& -\mathsf{E}\{q_1q_1'\bar{Z}'C\bar{Z}Q\bar{Z}'Gvu'\bar{Z}Q + 2q_1q_1'\bar{Z}'C'\bar{Z}Q\bar{Z}'Gvu'\bar{Z}Q\} \\
& -4\sigma^4 q_1'\bar{Z}'C\bar{Z}q_1\operatorname{tr}(GG'C)q_1q_1' - \mathsf{E}\{q_1'\bar{Z}'C\bar{Z}q_1Q\bar{Z}'Gvu'\bar{Z}Q\} \\
= \;& -\sigma^2\{q_1q_1'\bar{Z}'(C + 2C')\bar{Z}Q\bar{Z}'C\bar{Z}Q + q_1'\bar{Z}'C\bar{Z}q_1Q\bar{Z}'C\bar{Z}Q\} - 4\sigma^4 q_1'\bar{Z}'C\bar{Z}q_1\operatorname{tr}(GG'C)q_1q_1'.
\end{aligned}$$

With (D.14) we find

$$\begin{aligned}
\mathsf{E}[s_*^2(u'\bar{Z}Q + v'Hvq_1')] &= \mathsf{E}[(s^2 - \sigma^2)(u'\bar{Z}Q + v'Hvq_1')] + o(T^{-1}) \\
&= (T - K - 1)^{-1}\mathsf{E}\{[u'u - u'Z(Z'Z)^{-1}Z'u]v'Hvq_1'\} + o(T^{-1}) \\
&= T^{-1}\mathsf{E}(u'uv'Hv)q_1' + o(T^{-1}) \\
&= o(T^{-1}),
\end{aligned}$$

and employing (D.13) we obtain

$$\begin{aligned}
\mathsf{E}[\hat{\tau}^*(u'\bar{Z}Q + v'Hvq_{11}')] &= [\operatorname{tr}(C'C'CC) + 2\operatorname{tr}(C'CCC)]\mathsf{E}[(\hat{\lambda} - \lambda)(u'\bar{Z}Q + v'Hvq_{11}')] \\
&= \sigma^2[\operatorname{tr}(C'C'CC) + 2\operatorname{tr}(C'CCC)]q_1' + o(1).
\end{aligned}$$

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

$$
\begin{aligned}
& \mathsf{E}[(\hat{\mathsf{B}}_\alpha - b\mathsf{B}_\alpha)(\hat{\alpha}-\alpha)'] \qquad\qquad \text{(E.6)}\\
= \; & \sigma^4\{\mathrm{tr}(Q\bar{Z}'C\bar{Z})q_1q_1'\bar{Z}'C\bar{Z}Q + Q\bar{Z}'C\bar{Z}q_1q_1'\bar{Z}'C\bar{Z}Q \\
& \quad +q_1q_1'\bar{Z}'(C+2C')\bar{Z}Q\bar{Z}'C\bar{Z}Q + q_1'\bar{Z}'C\bar{Z}q_1Q\bar{Z}'C\bar{Z}Q \\
& \quad -\mathrm{tr}(Q\bar{Z}'CC\bar{Z})q_1q_1' - Q\bar{Z}'CC\bar{Z}q_1q_1' - q_1q_1'\bar{Z}'(CC+2C'C)\bar{Z}Q \\
& \quad +q_{11}[\mathrm{tr}(Q\bar{Z}'C\bar{Z})Q\bar{Z}'C\bar{Z}Q + Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}Q - Q\bar{Z}'CC\bar{Z}Q]\} \\
& +2\sigma^6\{2q_1'\bar{Z}'C\bar{Z}q_1\,\mathrm{tr}(GG'C)q_1q_1' \\
& \quad +q_{11}\,\mathrm{tr}(GG'C)[\mathrm{tr}(Q\bar{Z}'C\bar{Z})q_1q_1' + Q\bar{Z}'C\bar{Z}q_1q_1' + 3q_1q_1'\bar{Z}'C\bar{Z}Q] \\
& \quad -q_{11}[\mathrm{tr}(C'C'CC) + 2\,\mathrm{tr}(C'CCC)]q_1q_1' + q_{11}^2\,\mathrm{tr}(GG'C)Q\bar{Z}'C\bar{Z}Q\} \\
& +8\sigma^8q_{11}^2[\mathrm{tr}(GG'C)]^2q_1q_1' + o(T^{-2}).
\end{aligned}
$$

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 KP (1998, Appendix C), of respectively $\mathrm{tr}(C'C'CC)$ and $\mathrm{tr}(GG'C'C)$ and of $\mathrm{tr}(C'CCC)$ and $\mathrm{tr}(GG'CC')$ yields the required approximation to $\mathrm{MSE}(\breve{\alpha})$ and, hence, to the variance $\mathsf{V}(\breve{\alpha})$ 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

$$
\begin{aligned}
& \hat{\mathsf{V}}(\hat{\alpha}) \qquad\qquad \text{(F.1)}\\
& +\sigma^4\{[\mathrm{tr}(Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}) + 2(1-\lambda^2)^{-1}]q_1q_1' \\
& \quad +Q\bar{Z}'C\bar{Z}q_1q_1'\bar{Z}'C'\bar{Z}Q + Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}q_1q_1' + q_1q_1'\bar{Z}'C'\bar{Z}Q\bar{Z}'C'\bar{Z}Q\} \\
& +2\sigma^6\{2q_1'\bar{Z}'C\bar{Z}q_1\,\mathrm{tr}(GG'C)q_1q_1' + q_{11}\,\mathrm{tr}(GG'C)[Q\bar{Z}'C\bar{Z}q_1q_1' + q_1q_1'\bar{Z}'C'\bar{Z}Q]\} \\
& +4\sigma^8q_{11}^2[\mathrm{tr}(GG'C)]^2q_1q_1'
\end{aligned}
$$

is unbiased for $\mathsf{V}(\breve{\alpha})$ to order T^{-2}. However, this is not an estimator because the terms in σ^4, $\sigma^6$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, $\mathrm{tr}(GG'C)$ with $\mathrm{tr}(\hat{C}\hat{C}'\hat{C})$, 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 $\bar{Z} = (\bar{y}_0^*F, \iota)$ we can obtain

$$
\begin{aligned}
q_{11} &= (1-\lambda^2)T^{-1} + [1-(1-\lambda^2)(\bar{y}_0^{*2}+\omega^2)]T^{-2} + o(T^{-2}) \\
q_{12} &= -\bar{y}_0^*(1+\lambda)T^{-2} + o(T^{-2}) \qquad\qquad \text{(G.1)}\\
q_{22} &= T^{-1} + o(T^{-2}),
\end{aligned}
$$

and moreover

$$
\begin{aligned}
\mathrm{tr}(Q\bar{Z}'C\bar{Z}) &= (1-\lambda)^{-1} + O(T^{-1}) \qquad\qquad \text{(G.2)}\\
q_1'\bar{Z}'C\bar{Z}q_1 &= \bar{y}_0^{*2}\lambda T^{-2} + o(T^{-2}) = O(T^{-2}).
\end{aligned}
$$

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 (4.4) follows using $\mathrm{tr}(GG'C) = \lambda(1-\lambda^2)^{-2}T + O(1)$, which is proved in KP (1998b, formula C.10). KP (1998b) also gives

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

and in this special model we further have

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

Substituting the above in Corollary 2.2 yields

$$\begin{aligned} \mathsf{V}(\hat{\lambda}) &= (1-\lambda^2)T^{-1} + [1-(1-\lambda^2)(\bar{y}_0^{*2}+\omega^2)]T^{-2} \\ &\quad + T^{-2}[-2(1+\lambda^2) - 8\lambda^2 + 4\lambda(1+\lambda) + 20\lambda^2] + o(T^{-2}) \\ &= (1-\lambda^2)T^{-1} - [(1-\lambda^2)(\bar{y}_0^{*2}+\omega^2)]T^{-2} - (1-4\lambda-14\lambda^2)T^{-2} + o(T^{-2}), \end{aligned}$$

given in (4.5), and adding $[(1+3\lambda)T^{-1}]^2$ yields (4.6). Evaluating Corollary 2.3 upon using $\mathrm{tr}(C'C) = T(1-\lambda^2)^{-1} + O(1)$ gives (4.8) from which (4.9) and (4.10) straightforwardly follow. Evaluation of Corollary 3.2 produces (4.12) 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 $\mathrm{MSE}(\hat{\lambda}) - \mathrm{MSE}(\check{\lambda})$ as stated in Theorem 4.1.

In KP (1998a) we derived

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

This implies

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

and

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

Because $1+6\lambda < 4+3\lambda$ for any $|\lambda| < 1$ estimator $\dot{\lambda}$ has smaller second order bias. However, since $\mathsf{V}(\check{\lambda}) = [(T+3)/T]^2\mathsf{V}(\hat{\lambda})$ and $\mathsf{V}(\dot{\lambda}) = [T/(T-3)]^2\mathsf{V}(\hat{\lambda})$, $\mathrm{MSE}(\check{\lambda}) < \mathrm{MSE}(\dot{\lambda})$ 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.