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

Jan F. Kiviet* (University of Amsterdam) and Garry D.A. Phillips† (University of Exeter)

PRELIMINARY VERSION, 24 January 2000
JEL-code: C13, C22
Keywords: ARX-model, asymptotic expansions, bias correction, efficiency gains, finite sample moments, lagged dependent variables

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 initial terms in an expansion of the usual expression for the asymptotic coefficient variance estimator and by comparing these 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 and allows to compare analytically the second-order approximation to the mean squared error of the least-squares estimator and its counterpart for the first-order bias corrected coefficient estimator. Two rather strong results on efficiency gains through bias correction for AR(1) models follow. Illustrative simulation results on the magnitude of bias in coefficient and variance estimation and on the scope for effective bias correction and efficiency improvement are presented for some relevant particular cases of the ARX(1) class of models.

1. Introduction

In a recent paper, Kiviet and Phillips (1998a), we (henceforth referred to by KP) have obtained a higher-order approximation to the bias in the least-squares estimator of the

* Tinbergen Institute and Faculty of Economics & Econometrics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands (JFK@FEE.UvA.NL)

† School of Business and Economics, University of Exeter, Streatham Court, Exeter EX4 4PU, England (G.D.A.Phillips@exeter.ac.UK)

coefficients of normal stable ARX(1) models. In another recent paper, KP (1998b), we obtained a higher-order approximation to the bias in estimators of the disturbance variance based on the sum of squared least-squares residuals divided by alternative measures for the degrees of freedom. 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. 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 in the first-order stable dynamic regression model. In this class of models the dependent variable is explained linearly by an arbitrary number of strongly exogenous regressor variables and by the one period lagged dependent variable, and it depends on additive normally distributed (i.i.d.) disturbances.

In the stable model the coefficient of the lagged dependent variable is smaller than one in absolute value (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 1999). 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 been undertaken for about 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 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 KP (1998a, 1998b), for related work see also KP (1993, 1994) and Kiviet et al. (1995), the focus of attention is here the bias of ordinary least-squares (OLS) estimation (i.e. Maximum Likelihood conditional on y_0 and X) of all the regression coefficients in the first-order normal linear dynamic regression model

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

where $y = (y_1, \ldots, y_T)'$ is a $T \times 1$ vector of observations on a dependent variable, y_{-1} is the y vector lagged one period, i.e. $y_{-1} = (y_0, \ldots, y_{T-1})'$, 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 α of α and examine its relative efficiency analytically and also experimentally in simulations. Rewriting (1.1) as

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

where $Z = [y_{-1} \vdots X]$; 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 $V(\hat{\alpha}) = E[\hat{\alpha} - E(\hat{\alpha})][\hat{\alpha} - E(\hat{\alpha})]'$ is usually estimated by

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

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

$$\hat{\sigma}^2 = \frac{(y - Z\hat{\alpha})'(y - Z\hat{\alpha})}{T}: \tag{1.6}$$

The coefficient variance estimator $\hat{\sigma}^2(Z'Z)^{-1}$ disregards any finite sample considerations.

Note that the derivation of moments such as $E(\hat{\alpha})$; $V(\hat{\alpha})$ and $E[\hat{V}(\hat{\alpha})]$ is non-trivial, because Z is stochastic and depends linearly on u; whereas $\hat{\alpha}$ depends 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 a representation for the true variance (but only an 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 efficiency gains are obtained. In Section 5 we verify the numerical magnitude of the bias of alternative variance estimators by Monte Carlo simulation. Finally, in Section 6, we summarize our main conclusions. Proofs of our findings 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 $\text{rank}(Z) = k+1$ with probability one; (iv) the regressors in X are strongly exogenous; (v) the disturbances follow $u \sim N(0, \sigma^2 I_T)$, with $0 < \sigma^2 < \infty$; (vi) the start-up value has $y_0 \sim N(\bar{y}_0, \omega^2\sigma^2)$, 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 $\omega > 0$ the start-up is random, and if $\omega^2 = (1-\lambda^2)^{-1}$ then $\{y_t\}$ is a covariance stationary process (but possibly with a non-constant mean). 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 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 y_0, i.e.

$$\bar{Z} = E(Z) = [E(y_{-1}) \vdots X] = [\bar{y}_{-1} \vdots X] \tag{2.1}$$

and

$$\tilde{Z} = Z - \bar{Z} = [y_{-1} - \bar{y}_{-1} \vdots X - X] = [\tilde{y}_{-1} \vdots O] = \tilde{y}_{-1} e_1'; \tag{2.2}$$

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

$$\bar{y}_{-1} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ \lambda & 1 & \ddots & & & \vdots \\ \lambda^2 & \lambda & 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ \lambda^{T-1} & \cdots & \cdots & \cdots & \lambda & 1 \end{bmatrix} \begin{pmatrix} \bar{y}_0 \\ x_1'\beta \\ x_2'\beta \\ \vdots \\ \vdots \\ x_{T-1}'\beta \end{pmatrix}; \tag{2.3}$$

where $X' = [x_1 \vdots \cdots \vdots x_T]$, and hence we find that $\bar{Z}$ is determined by: X; $\bar{y}_0$, β and λ. Also defining $v = (u_0; u_1; \ldots; u_T)'$ such that

$$v \sim N[0; \sigma^2 I_{T+1}] \text{ with } y_0 = \bar{y}_0 + \omega u_0; \tag{2.4}$$

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

$$G = [\omega F \vdots C]; \text{ with } F = \begin{pmatrix} 1 \\ \lambda \\ \lambda^2 \\ \vdots \\ \vdots \\ \lambda^{T-1} \end{pmatrix}; \; C = \begin{bmatrix} 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 1 & 0 & & & & \vdots \\ \lambda & 1 & 0 & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \lambda^{T-2} & \cdots & \cdots & \lambda & 1 & 0 \end{bmatrix}; \tag{2.5}$$

one can easily verify that

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

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

We find

$$\begin{aligned} E(Z'Z) &= E(\bar{Z} + Gve_1')'(\bar{Z} + Gve_1') \\ &= \bar{Z}'\bar{Z} + \sigma^2 \mathrm{tr}(G'G)e_1e_1' \end{aligned} \tag{2.7}$$

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

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

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

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

In fact, KP (1998a) presents a more accurate and complicated approximation to the bias of $\hat\beta$; which gives the bias to second order. 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 some 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 $\hat\beta$ the approximation:

$$
\begin{aligned}
& V(\hat\beta) = E\{[\hat\beta - E(\hat\beta)][\hat\beta - E(\hat\beta)]'\} = \\
& \sigma^2 Q \\
& +\sigma^4\{[tr(Q\bar{Z}'GG'\bar{Z}) - 2tr(Q\bar{Z}'CC\bar{Z}) + tr(Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z})]q_1q_1' \\
& \quad +Q\bar{Z}'(GG' - CC - 2C'C - C'C')\bar{Z}q_1q_1' \\
& \quad +q_1q_1'\bar{Z}'(GG' - CC - 2C'C - C'C')\bar{Z}Q \\
& \quad +Q\bar{Z}'C\bar{Z}q_1q_1'\bar{Z}'C\bar{Z}Q + Q\bar{Z}'C\bar{Z}q_1q_1'\bar{Z}'C'\bar{Z}Q + Q\bar{Z}'C'\bar{Z}q_1q_1'\bar{Z}'C'\bar{Z}Q \\
& \quad +q_1q_1'\bar{Z}'C'\bar{Z}Q\bar{Z}'(C + C')\bar{Z}Q + Q\bar{Z}'(C + C')\bar{Z}Q\bar{Z}'C\bar{Z}q_1q_1' \\
& \quad +q_1q_1'\bar{Z}'(C + C')\bar{Z}Q\bar{Z}'C\bar{Z}Q + Q\bar{Z}'C'\bar{Z}Q\bar{Z}'(C + C')\bar{Z}q_1q_1' \\
& \quad +tr(Q\bar{Z}'C\bar{Z})[q_1q_1'\bar{Z}'C\bar{Z}Q + Q\bar{Z}'C'\bar{Z}q_1q_1'] + (q_1'\bar{Z}'C\bar{Z}q_1)Q\bar{Z}'(C + C')\bar{Z}Q \\
& \quad +q_{11}[tr(Q\bar{Z}'C\bar{Z})Q\bar{Z}'(C + C')\bar{Z}Q + Q\bar{Z}'(GG' - CC - C'C')\bar{Z}Q \\
& \qquad\qquad +Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}Q + Q\bar{Z}'C'\bar{Z}Q\bar{Z}'C'\bar{Z}Q]\} \\
& +\sigma^6\{12tr(GG'C)(q_1'\bar{Z}'C\bar{Z}q_1)q_1q_1' \\
& \quad +q_{11}[2tr(G'GG'G) - 8tr(G'GCC) - 4tr(G'C'CG) \\
& \qquad\qquad +4tr(GG'C)tr(Q\bar{Z}'C\bar{Z})]q_1q_1' \\
& \quad +q_{11}tr(GG'C)[4Q\bar{Z}'C\bar{Z}q_1q_1' + 4q_1q_1'\bar{Z}'C'\bar{Z}Q) \\
& \qquad\qquad\qquad +6Q\bar{Z}'C'\bar{Z}q_1q_1' + 6q_1q_1'\bar{Z}'C\bar{Z}Q)] \\
& \quad +2q_{11}^2 tr(GG'C)Q\bar{Z}'(C + C')\bar{Z}Q\} \\
& +20\sigma^8 q_{11}^2[tr(GG'C)]^2 q_1q_1' \\
& +o(T^{-2}).
\end{aligned}
$$

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 this actual variance. In Appendix C we prove:

Theorem 2.3: Under Assumption 2.1 we find for the expectation of the usual estimator of $V(\hat{\alpha})$ given in (1.4) the approximation

$$
\begin{aligned}
E[\hat{V}(\hat{\alpha})] = \; & E[s^2(Z'Z)^{-1}] = \\
& \sigma^2 Q \\
& + \sigma^4 \{ [\mathrm{tr}(Q\bar{Z}'GG'\bar{Z}) - \frac{2}{T}\mathrm{tr}(C'C)] q_1 q_1' \\
& \qquad + Q\bar{Z}'GG'\bar{Z} q_1 q_1' + q_1 q_1' \bar{Z}'GG'\bar{Z} Q + q_{11} Q\bar{Z}'GG'\bar{Z} Q \} \\
& + 2\sigma^6 q_{11} \mathrm{tr}(G'GG'G) q_1 q_1' + o(T^{-2}).
\end{aligned}
$$

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

If these differences can be substantial it would seem interesting to develop a corrected estimator of $V(\hat{\alpha})$, say $V(\hat{\alpha})$, which adds particular terms to the standard estimator $\hat{V}(\hat{\alpha})$, such that $E[V(\hat{\alpha})]$ is equivalent to second order to $V(\hat{\alpha})$. We return to the issue of bias reduction of variance (and coefficient) estimators later.

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 as:

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

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 $\hat{\beta}$ the approximation:

$$
\begin{aligned}
& V(\hat{\beta}) = E[\hat{\beta} - E(\hat{\beta})]^2 = \\
& \sigma^2 q_{11} \\
& +\sigma^4 \{5(q_1' \bar{Z}' C \bar{Z} q_1)^2 \\
& \qquad +q_{11}[6(q_1' \bar{Z}' C \bar{Z} Q \bar{Z}' C \bar{Z} q_1) + 4(q_1 \bar{Z}' C' \bar{Z} Q \bar{Z}' C \bar{Z} q_1) \\
& \qquad\qquad +(q_1' \bar{Z}'[3GG' - 6CC - 4C'C]\bar{Z} q_1) + 4tr(Q\bar{Z}'C\bar{Z})(q_1'\bar{Z}'C\bar{Z}q_1)] \\
& \qquad +q_{11}^2[tr(Q\bar{Z}'GG'\bar{Z}) - 2tr(Q\bar{Z}'CC\bar{Z}) + tr(Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z})]\} \\
& +\sigma^6 \{36 q_{11}^2 tr(GG'C)(q_1'\bar{Z}'C\bar{Z}q_1)] \\
& \qquad +q_{11}^3[2tr(G'GG'G) - 8tr(G'GCC) - 4tr(G'C'CG) \\
& \qquad\qquad +4tr(GG'C)tr(Q\bar{Z}'C\bar{Z})]\} \\
& +20\sigma^8 q_{11}^4[tr(GG'C)]^2 \\
& +o(T^{-2}).
\end{aligned}
$$

From Theorem 2.3 we obtain:

Corollary 2.3: Under Assumption 2.1 we find for the expectation of the usual estimator of the variance of the estimator $\hat{\beta}$ the approximation:

$$
\begin{aligned}
E[\hat{V}(\hat{\beta})] \; = \; & E[s^2 e_1'(Z'Z)^{-1} e_1] = \\
& \sigma^2 q_{11} \\
& +\sigma^4 \{3 q_{11}(q_1'\bar{Z}'GG'\bar{Z}q_1) + q_{11}^2[tr(Q\bar{Z}'GG'\bar{Z}) - \frac{2}{T} tr(C'C)]\} \\
& +2\sigma^6 q_{11}^3 tr(G'GG'G) + o(T^{-2}).
\end{aligned}
$$

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 some scope for developing a second-order unbiased estimator $V(\hat{\beta})$ for $V(\hat{\beta})$.

3. The efficiency of bias corrected coefficient estimators

The approach layed 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 (iii) to 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

$$\hat{\alpha} + \sigma^2[\mathrm{tr}(Q\bar{Z}'C\bar{Z})q_1 + Q\bar{Z}'C\bar{Z}q_1 + 2\sigma^2 q_{11}\mathrm{tr}(GG'C)q_1]$$

it is obvious that this has expectation $\alpha + o(T^{-2})$; but it is not an operational estimator, because σ^2, C, G and Q are, or depend on, unknown parameters. However, consider the operational corrected least-squares (COLS) estimator

$$\alpha = \hat{\alpha} + s^2\mathrm{tr}(P\hat{Z}'\hat{C}\hat{Z})p_1 + s^2P\hat{Z}'\hat{C}\hat{Z}p_1 + 2s^4p_{11}\mathrm{tr}(\hat{C}\hat{C}'\hat{C})p_1; \qquad (3.1)$$

where $\hat{\alpha}$ and s^2 are the usual least-squares estimators, $P = (Z'Z)^{-1}$, which has first column p_1 with first element p_{11}, $\hat{Z} = [\hat{F}y_0 + \hat{C}X\hat{\lambda} \vdots X]$ and $\hat{C}$ equals C (also $\hat{F}$ equals F) with the unknown λ replaced by $\hat{\lambda}$. In Appendix D we prove:

Theorem 3.1: Under Assumption 2.1 the COLS estimator α given in (3.1) is unbiased to order T^{-1}, i.e.

$$E(\alpha) = \alpha + o(T^{-1}).$$

For this bias corrected estimator we obtain in Appendix E:

Theorem 3.2: Under Assumption 2.1 we find for the variance of the bias corrected estimator α given in (3.1) the approximation:

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

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

From Theorems 2.3 and 3.2 we find (proof in Appendix F):

Theorem 3.3: Under Assumption 2.1 the estimator $V(\alpha)$ of the variance of the bias corrected estimator α given in (3.1) has $E[V(\alpha) - V(\alpha)] = o(T^{-2})$ if we define:

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

Specializing the above results to the variance of the first element of α yields:

Corollary 3.2: Under Assumption 2.1 we find

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

and

Corollary 3.3: Under Assumption 2.1 $V(\lambda)$ is unbiased to second order for $V(\hat{\lambda})$ when defining:

$$
\begin{aligned}
V(\lambda) \equiv\ & \hat{V}(\hat{\lambda}) + \\
& + s^4\{(p_1'\hat{Z}'\hat{C}\hat{Z}p_1)^2 + 2p_{11}(p_1'\hat{Z}'\hat{C}\hat{Z}P\hat{Z}'\hat{C}\hat{Z}p_1) \\
& \qquad + p_{11}^2[\frac{2}{T}tr(\hat{C}'\hat{C}) + tr(P\hat{Z}'\hat{C}\hat{C}'\hat{Z}) + tr(P\hat{Z}'\hat{C}\hat{Z}P\hat{Z}'\hat{C}\hat{Z})]\} \\
& + 8s^6p_{11}^2 tr(\hat{C}\hat{C}'\hat{C})(p_1'\hat{Z}'\hat{C}\hat{Z}p_1) \\
& + 4s^8p_{11}^4[tr(\hat{C}\hat{C}'\hat{C})]^2.
\end{aligned}
$$

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

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

Here 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 only exogenous regressor. Hence, we have here:

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

For this model Corollary 2.1 reduces to the well-known Kendall (1954) approximation restated here as:

Corollary 4.1: Under Assumption 2.1 the bias of the least-squares estimator $\hat\lambda$ for the special case of model (4.1) can be approximated as:

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

Hence, this approximation proves to be valid irrespective of the nature of the start-up value y_0. Defining for model (4.1) the standardized start-up value

$$y_0^* = \frac{1}{\sigma}\left(y_0 - \frac{\beta}{1-\lambda}\right), \tag{4.2}$$

we find the following approximative expression for the true variance, see Appendix G:

Corollary 4.2: Under Assumption 2.1 the variance of the least-squares estimator $\hat\lambda$ for the special case of model (4.1) can be approximated as:

$$\begin{aligned} V(\hat\lambda) &= E[\hat\lambda - E(\hat\lambda)]^2 \\ &= \frac{1-\lambda^2}{T} - \frac{1-4\lambda-14\lambda^2}{T^2} - \frac{1-\lambda^2}{T^2}\left(y_0^{*2} + \omega^2\right) + o(T^{-2}). \end{aligned}$$

Here the leading term $\frac{1-\lambda^2}{T}$ is simply the asymptotic variance of $\hat\lambda$. Notice that in the mean-stationary case, where $y_0 = \frac{\beta}{1-\lambda}$, the approximation does not involve σ^2 nor β. Also note that the variance decreases with the variance of the initial value y_0.

For the expectation of the standard variance estimator we find:

Corollary 4.3: Under Assumption 2.1 the expectation of the estimator for the variance of the least-squares estimator $\hat{\lambda}$ for the special case of model (4.1) can be approximated as:

$$\begin{aligned} E[\hat{V}(\hat{\lambda})] &= E[s^2 e_1'(Z'Z)^{-1}e_1] \\ &= \frac{1-\lambda^2}{T} + \frac{2+2\lambda+5\lambda^2}{T^2} - \frac{1-\lambda^2}{T^2}\left(y_0^{*2} + \omega^2\right) + o(T^{-2}). \end{aligned}$$

It is obvious that this estimator is unbiased to order T^{-1}, but biased to order T^{-2}, since one of its second order terms differs from the corresponding one of Corollary 4.2. So, even though we do not know $V(\hat{\lambda})$ precisely, we find from its approximation that the standard estimator is biased, viz.

$$E[\hat{V}(\hat{\lambda}) - V(\hat{\lambda})] = \frac{3-2\lambda-9\lambda^2}{T^2} + o(T^{-2}). \tag{4.3}$$

If we can modify $\hat{V}(\hat{\lambda})$ to have the same mean to order T^{-2} as $V(\hat{\lambda})$, then an approximately unbiased estimator of $V(\hat{\lambda})$ will result. Thus, for the case $Z = [y_{-1} : \iota]$ where $\iota = (1, \ldots, 1)'$, the statistic

$$s^2 e_1'(Z'Z)^{-1}e_1 - \frac{3-2\lambda-9\lambda^2}{T^2}$$

is unbiased to $O(T^{-2})$ for $V(\hat{\lambda})$ but in practice this estimator is not operational since it depends on λ. However, since $E(\hat{\lambda} - \lambda) = O(T^{-1})$ we find:

Theorem 4.1: Under Assumption 2.1 the corrected estimator for the variance of the least-squares estimator $\hat{\lambda}$ for the special case of model (4.1) given by

$$V(\hat{\lambda}) \equiv \hat{V}(\hat{\lambda}) - \frac{3-2\hat{\lambda}-9\hat{\lambda}^2}{T^2}$$

is unbiased to order T^{-2}, i.e. $E[V(\hat{\lambda}) - V(\hat{\lambda})] = o(T^{-2})$.

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

$$\lambda \equiv \hat{\lambda} + \frac{1}{T}(1+3\hat{\lambda}) = \frac{T+3}{T}\hat{\lambda} + \frac{1}{T}. \tag{4.4}$$

Exploiting now the analytic results of Section 3 on the COLS estimator for the special case of the AR(1) model with intercept we obtain:

Theorem 4.2: Under Assumption 2.1 the COLS estimator λ in the simple model (4.1) has

$$V(\lambda) = V(\hat{\lambda}) + \frac{6-6\lambda^2}{T^2} + o(T^{-2})$$

and

$$MSE(\rho) = MSE(\hat{\rho}) + \frac{5 - 6\rho - 15\rho^2}{T^2} + o(T^{-2});$$

so that, omitting terms of order $o(T^{-2})$; we find

$$MSE(\rho) < MSE(\hat{\rho}) \quad \text{for} \quad -1 < \rho < -0.811 \text{and for } 0.411 < \rho < 1:$$

The first result of this theorem shows that, to the order of the approximation, bias correction will invariably lead to an increase in variance. However, the MSE result indicates that, in the AR(1) model with intercept, bias correction is not beneficial as far as efficiency is concerned when $-0.811 < \rho < 0.411$: In a similar way, it can be derived that in the AR(1) model with no intercept bias correction yields no MSE reduction when $|\rho| < 0.707$: We conjecture that the greater the number of smooth regressors that are included in an ARX(1) model, the more scope there is to improve efficiency through bias correction.

When bias correction has been employed an adequate estimator for the resulting variance is still provided by $\hat{V}(\hat{\rho})$, but an operational estimator which is even unbiased to order T^{-2}; is then given by:

$$V(\rho) \simeq \hat{V}(\hat{\rho}) + \frac{3 + 2\rho + 3\rho^2}{T^2}: \tag{4.5}$$

Bias correction of AR(1) models has been entertained in the literature in many studies, see 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 given in Corollary 4.1, although, instead of using (4.4), in all the studies just referred to a bias corrected estimator $\tilde{\rho}$ has been used which is obtained by solving

$$\hat{\rho} = \tilde{\rho} - \frac{1}{T}(1 + 3\tilde{\rho});$$

yielding

$$\tilde{\rho} \simeq \frac{T}{T - 3}\hat{\rho} + \frac{1}{T - 3}: \tag{4.6}$$

For the bias of this corrected estimator we find

$$\begin{aligned}
E(\tilde{\rho} - \rho) &= \frac{T}{T - 3}E(\hat{\rho}) - \rho + \frac{1}{T - 3} \\
&= E(\hat{\rho} - \rho) + \frac{3}{T - 3}E(\hat{\rho}) + \frac{1}{T - 3} \\
&= -\frac{1}{T}(1 + 3\rho) + \frac{3}{T - 3}\rho + \frac{1}{T - 3} + O(T^{-2})
\end{aligned}$$

$$
\begin{aligned}
&= \lambda\left(\frac{3}{T-3} - \frac{3}{T}\right) + \frac{1}{T-3} - \frac{1}{T} + O(T^{-2}) \\
&= \lambda\frac{9}{T(T-3)} + \frac{3}{T(T-3)} + O(T^{-2}) \\
&= O(T^{-2});
\end{aligned}
$$

hence, both bias corrected estimators remove the $O(T^{-1})$ bias. In fact, however, $\bar{\lambda}$ is a hybrid bias correction estimator, because the correction also involves a non-random term of order $O(T^{-2})$; since $\frac{1}{T-3} = \frac{1}{T} + \frac{3}{T^2} + o(T^{-2})$: We also find:

Theorem 4.3: In the simple model (4.1) the corrected estimators $\tilde{\lambda}$ and $\bar{\lambda}$ are both unbiased to order $O(T^{-1})$; but, omitting terms of order $o(T^{-2})$; $\tilde{\lambda}$ is uniformly (for any $|\lambda| < 1$) more efficient than $\bar{\lambda}$; because

$$
V(\tilde{\lambda}) = \left(\frac{T+3}{T}\right)^2 V(\hat{\lambda}) \quad \text{and} \quad V(\bar{\lambda}) = \left(\frac{T}{T-3}\right)^2 V(\hat{\lambda}):
$$

The superiority of our implementation of bias correction follows from $\frac{T+3}{T} - \frac{T}{T-3} = -\frac{9}{T(T-3)} < 0$:

5. Numerical results

We shall examine the estimators $\hat{\lambda}$ and $\tilde{\lambda}$, their efficiency, the qualities of their respective (bias corrected) variance estimators and the accuracy of the various asymptotic approximations for model (4.1) and also for the AR(1) model with intercept and linear trend. For that purpose we perform various numerical evaluations and execute some series of Monte Carlo experiments. In what follows we write $V[\hat{\lambda}]$ for what in fact is the Monte Carlo estimate of $V(\hat{\lambda})$. Because we generate many replications the Monte Carlo estimates will be very close to the actual population moments. In the model with intercept only, we write $AV_1[\hat{\lambda}]$ for the (leading term) asymptotic variance of $\hat{\lambda}$; which is $(1-\lambda^2)/T$, and $AV_2[\hat{\lambda}]$ for the second-order asymptotic approximation to $V(\hat{\lambda})$; which is given by Corollary 4.2. For the mean over the Monte Carlo replications of $\hat{V}(\hat{\lambda})$ we simply write $E[\hat{V}[\hat{\lambda}]]$; and similarly for $V[\hat{\lambda}]$ of Theorem 4.1 and for $V[\tilde{\lambda}]$ of Corollary 3.3. In the tables we present results for various values of $|\lambda| < 1$ and T; focussing on positive values of λ and $10 \leq T \leq 50$ and the case $\beta = 0$; $y_0 = \beta/(1-\lambda) = 0$; $\omega = 0$; $\sigma = 1$; i.e. the model with a fixed start-up and mean-stationarity (note from the asymptotic approximations that these values often seem to mitigate the magnitude of the second order terms). All results presented in the tables are invariant with respect to β; and most are with respect to σ. Often the results are given as ratios. Then values of unity in the tables may indicate unbiasedness and values smaller (greater) than one negative (positive) bias. In order to compare the effect of bias correction on efficiency we also present the ratio of $MSE[\lambda]$ and $MSE[\hat{\lambda}]$; here values smaller than one indicate an efficiency gain due to bias correction.

Table 5.1:

Estimation in the AR(1) model with intercept; $T = 10$; 250,000 replications

λ	$E[\hat{\lambda}] - \lambda$	$\frac{E[\lambda]-\lambda}{E[\hat{\lambda}]-\lambda}$	$V[\hat{\lambda}]$	$\frac{AV_1[\hat{\lambda}]}{V[\hat{\lambda}]}$	$\frac{AV_2[\hat{\lambda}]}{V[\hat{\lambda}]}$	$\frac{E[\hat{V}[\hat{\lambda}]]}{V[\hat{\lambda}]}$	$\frac{E[V[\hat{\lambda}]]}{V[\hat{\lambda}]}$	$\frac{E[V[\lambda]]}{V[\lambda]}$	$\frac{MSE[\lambda]}{MSE[\hat{\lambda}]}$
0.0	-0.111	0.401	0.095	1.058	0.952	1.374	1.135	1.026	1.513
0.1	-0.139	0.363	0.096	1.026	0.979	1.358	1.130	1.024	1.431
0.2	-0.167	0.343	0.098	0.980	1.016	1.335	1.126	1.022	1.341
0.3	-0.197	0.333	0.099	0.918	1.065	1.305	1.123	1.019	1.247
0.4	-0.227	0.332	0.100	0.841	1.125	1.269	1.120	1.015	1.152
0.5	-0.260	0.338	0.100	0.748	1.197	1.227	1.117	1.012	1.056
0.6	-0.294	0.348	0.100	0.640	1.284	1.179	1.114	1.008	0.962
0.7	-0.331	0.363	0.099	0.515	1.390	1.124	1.110	1.004	0.872
0.8	-0.368	0.377	0.097	0.370	1.516	1.056	1.099	0.997	0.789
0.9	-0.401	0.377	0.096	0.198	1.650	0.954	1.069	0.976	0.721
0.99	-0.413	0.338	0.097	0.021	1.742	0.815	1.024	0.932	0.686

Table 5.2:

Estimation in the AR(1) model with intercept; $T = 20$; 250,000 replications

λ	$E[\hat{\lambda}] - \lambda$	$\frac{E[\lambda]-\lambda}{E[\hat{\lambda}]-\lambda}$	$V[\hat{\lambda}]$	$\frac{AV_1[\hat{\lambda}]}{V[\hat{\lambda}]}$	$\frac{AV_2[\hat{\lambda}]}{V[\hat{\lambda}]}$	$\frac{E[\hat{V}[\hat{\lambda}]]}{V[\hat{\lambda}]}$	$\frac{E[V[\hat{\lambda}]]}{V[\hat{\lambda}]}$	$\frac{E[V[\lambda]]}{V[\lambda]}$	$\frac{MSE[\lambda]}{MSE[\hat{\lambda}]}$
0.0	-0.053	0.200	0.048	1.043	0.991	1.169	1.031	1.009	1.252
0.1	-0.067	0.183	0.048	1.021	0.997	1.157	1.029	1.007	1.213
0.2	-0.082	0.176	0.049	0.989	1.007	1.141	1.028	1.006	1.165
0.3	-0.097	0.175	0.048	0.945	1.021	1.120	1.027	1.004	1.110
0.4	-0.113	0.180	0.047	0.888	1.038	1.092	1.026	1.001	1.047
0.5	-0.130	0.190	0.046	0.814	1.058	1.058	1.024	0.997	0.976
0.6	-0.148	0.207	0.044	0.720	1.082	1.015	1.022	0.992	0.898
0.7	-0.169	0.232	0.042	0.601	1.111	0.961	1.019	0.985	0.813
0.8	-0.193	0.268	0.040	0.451	1.149	0.898	1.016	0.979	0.720
0.9	-0.221	0.313	0.037	0.259	1.210	0.819	1.013	0.974	0.623
0.99	-0.236	0.307	0.034	0.030	1.274	0.682	0.975	0.938	0.557

We focus on the AR(1) model with intercept first. Table 5.1 contains results for $T = 10$: We see from the second column that at such a small sample size the least-squares estimator is badly and negatively biased (often -50%), especially for larger values of λ: The next column indicates that a first-order correction of this estimator reduces the bias by about 60%. Only for small values of λ is the asymptotic variance of the least-squares estimator found to be a reasonable indicator of the actual variance. Especially for λ values close to one, the asymptotic variance is much too

Table 5.3:

Estimation in the AR(1) model with intercept; T = 50; 250,000 replications									
λ	$E[\hat\lambda]-\lambda$	$\frac{E[\lambda]-\lambda}{E[\hat\lambda]-\lambda}$	$V[\hat\lambda]$	$\frac{AV_1[\hat\lambda]}{V[\hat\lambda]}$	$\frac{AV_2[\hat\lambda]}{V[\hat\lambda]}$	$\frac{E[\hat V[\hat\lambda]]}{V[\hat\lambda]}$	$\frac{E[V[\hat\lambda]]}{V[\hat\lambda]}$	$\frac{E[V[\lambda]]}{V[\lambda]}$	$\frac{MSE[\lambda]}{MSE[\hat\lambda]}$
0.0	-0.021	0.103	0.020	1.020	0.999	1.063	1.005	1.002	1.099
0.1	-0.027	0.092	0.020	1.009	1.000	1.058	1.004	1.001	1.084
0.2	-0.033	0.087	0.019	0.994	1.001	1.050	1.004	1.001	1.064
0.3	-0.039	0.086	0.019	0.972	1.003	1.039	1.003	1.000	1.040
0.4	-0.045	0.088	0.018	0.941	1.005	1.025	1.002	0.998	1.009
0.5	-0.052	0.094	0.017	0.900	1.008	1.005	1.001	0.996	0.969
0.6	-0.059	0.104	0.015	0.841	1.010	0.977	0.999	0.993	0.919
0.7	-0.066	0.121	0.013	0.756	1.013	0.936	0.996	0.988	0.853
0.8	-0.075	0.152	0.012	0.624	1.011	0.876	0.990	0.979	0.763
0.9	-0.088	0.216	0.009	0.405	0.999	0.780	0.978	0.964	0.639
0.99	-0.103	0.291	0.007	0.056	1.003	0.620	0.956	0.942	0.498

small. A second-order asymptotic approximation to the variance proves to be more accurate. Where the first-order approximation is much too small, the second-order approximation overshoots, and does so quite seriously for large values of λ. The standard estimator $\hat V(\hat\lambda)$ of the variance of $\hat\lambda$, however, is relatively good, although it can have a bias of ±30% or beyond. Also the corrected variance estimator $V(\hat\lambda)$ shows some remaining bias, but generally this bias is mitigated and always the corrected estimator produces a conservative estimator of the actual variance (it never has a negative bias). Hence, we find that our analytical higher-order asymptotic results on second moments can be used successfully already at a sample size as small as $T = 10$, which seems quite remarkable. Assessing the variance of the bias corrected coefficient estimator λ reasonably accurately is shown to be possible, according to the findings in the last but one column. The final column shows the magnitude in efficiency loss (or gain) due to coefficient bias correction. From the results we see that at this sample size the critical point is actually not $\lambda = 0.411$, as our earlier analysis suggested, but slightly larger. Apparently effects of third-order are sizeable at such a small sample size.

In Table 5.2 we present similar results for $T = 20$. The coefficient bias is smaller now, but still substantial (often -25%). Correcting for bias is more successful, because it yields a reduction to 20 or 30% of the original bias. The accuracy of the asymptotic variance is still appalling, but the second-order approximation seems acceptable now as long as λ is not too close to unity. The standard variance estimator may show a bias of some 20%, but the corrected estimator is really quite an accurate one. From the final column we see that, upon comparing with Table 5.1, the increased sample size offers in fact more scope for relative efficiency gains through coefficient bias correction (all figures in the final column are smaller). The efficiency of the corrected estimator can be estimated very precisely, as is shown by the last but one column, and the value for λ where correction starts to pay off is just above 0.4 now, as was to be expected from the result in Theorem 4.2.

Increasing the sample size further to $T = 50$ yields results presented in Table 5.3 which show more or less the same pattern, apart from the following. We find that the potential efficiency gains from bias correction are much more substantial now than the possible losses that occur in cases where λ is non-negative but small. Especially for large values of λ; say $\lambda = 0.9$ where the bias is only about -10%, efficiency gains of about 40% can be achieved (and for $\lambda > 0$ efficiency losses never exceed 10%), whereas a very accurate estimator of this improved efficiency is available, as can be seen from the last but one column. Note that a MSE reduction of 40% implies a reduction of some 25% in terms of root mean squared errors (or standard errors). This seems quite attractive against the risk of a possible increase by only 5%. Hence, these three tables show that it is not necessarily 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 moderate, so that the bias approximation is reasonably accurate and therefore more effective.

Table 5.4:

Estimation in the AR(1) model with intercept; 250,000 replications

	T = 10			T = 20			T = 50		
λ	$\frac{E[\bar\lambda]-\lambda}{E[\hat\lambda]-\lambda}$	$\frac{E[V[\bar\lambda]]}{V[\bar\lambda]}$	$\frac{MSE[\bar\lambda]}{MSE[\hat\lambda]}$	$\frac{E[\bar\lambda]-\lambda}{E[\hat\lambda]-\lambda}$	$\frac{E[V[\bar\lambda]]}{V[\bar\lambda]}$	$\frac{MSE[\bar\lambda]}{MSE[\hat\lambda]}$	$\frac{E[\bar\lambda]-\lambda}{E[\hat\lambda]-\lambda}$	$\frac{E[V[\bar\lambda]]}{V[\bar\lambda]}$	$\frac{MSE[\bar\lambda]}{MSE[\hat\lambda]}$
0.0	0.145	0.857	1.807	0.059	0.965	1.309	0.045	0.995	1.107
0.1	0.090	0.858	1.703	0.039	0.964	1.266	0.034	0.994	1.092
0.2	0.061	0.858	1.589	0.030	0.963	1.215	0.029	0.994	1.072
0.3	0.048	0.859	1.469	0.030	0.961	1.156	0.028	0.993	1.047
0.4	0.046	0.859	1.346	0.035	0.959	1.088	0.030	0.991	1.015
0.5	0.054	0.860	1.221	0.048	0.957	1.012	0.036	0.989	0.975
0.6	0.069	0.861	1.096	0.067	0.953	0.927	0.046	0.986	0.924
0.7	0.090	0.862	0.973	0.097	0.948	0.832	0.065	0.982	0.856
0.8	0.110	0.862	0.860	0.139	0.943	0.726	0.098	0.973	0.764
0.9	0.110	0.850	0.771	0.191	0.940	0.614	0.166	0.958	0.635
0.99	0.055	0.821	0.742	0.185	0.908	0.543	0.246	0.937	0.487

In Table 5.4 we present results for the alternative bias corrected estimator $\bar\lambda$ for all three sample sizes examined. Surprisingly we find that the less efficient estimator $\bar\lambda$ is much less biased than λ: It seems to be an artifact that the effect that the correction in $\bar\lambda$ has on the order $O(T^{-2})$ term of the bias happens to be such that it mitigates the magnitude of this term. It is easy to show that the estimator

$$V(\bar\lambda) \equiv \hat V(\hat\lambda) + \frac{3 + 2\bar\lambda + 3\bar\lambda^2}{T^2} \qquad (5.1)$$

is unbiased to order $O(T^{-2})$ for $V(\bar\lambda)$: From the simulations we find, however, that this estimator is less accurate than $V(\lambda)$ is for $V(\lambda)$; especially so for the smaller sample sizes. For $T = 10$ the better efficiency of $\hat{\lambda}$ is apparent, but the smaller finite sample bias of $\tilde{\lambda}$ leads to satisfactory MSE results close to the unit circle.

Now we present a few results for the AR(1) model with intercept and trend. We just examine the case where $\beta = 0$ (both intercept and trend are redundant), $y_0 = 0$ and $\omega = 0$ (fixed, mean-stationary start-up), which imply that many characteristics are invariant with respect to σ. We have set $\sigma = 0.1$.

Table 5.5:

Estimation in the AR(1) model with intercept and trend; $T = 20$; 1,000 replications

λ	$E[\hat{\lambda}] - \lambda$	$\frac{E[\tilde{\lambda}] - \lambda}{E[\hat{\lambda}] - \lambda}$	$V[\hat{\lambda}]$	$\frac{E[\hat{V}[\hat{\lambda}]]}{V[\hat{\lambda}]}$	$V[\tilde{\lambda}]$	$\frac{E[\hat{V}[\hat{\lambda}]]}{V[\tilde{\lambda}]}$	$\frac{E[V[\tilde{\lambda}]]}{V[\tilde{\lambda}]}$	$\frac{MSE[\tilde{\lambda}]}{MSE[\hat{\lambda}]}$
0.0	-0.112	0.251	0.045	1.320	0.069	0.848	1.029	1.226
0.1	-0.134	0.227	0.046	1.294	0.071	0.839	1.030	1.124
0.2	-0.155	0.216	0.047	1.263	0.072	0.824	1.025	1.026
0.3	-0.178	0.215	0.047	1.225	0.072	0.806	1.019	0.929
0.4	-0.202	0.222	0.048	1.182	0.072	0.784	1.010	0.836
0.5	-0.227	0.238	0.048	1.133	0.071	0.759	1.001	0.748
0.6	-0.254	0.267	0.047	1.078	0.069	0.738	1.006	0.660
0.7	-0.284	0.311	0.047	1.016	0.066	0.723	1.036	0.578
0.8	-0.318	0.378	0.046	0.946	0.060	0.721	1.118	0.509
0.9	-0.361	0.467	0.046	0.865	0.054	0.732	1.213	0.469
0.99	-0.422	0.558	0.048	0.781	0.053	0.715	1.083	0.478

Table 5.6:

Estimation in the AR(1) model with intercept and trend; $T = 50$; 1,000 replications

λ	$E[\hat{\lambda}] - \lambda$	$\frac{E[\tilde{\lambda}] - \lambda}{E[\hat{\lambda}] - \lambda}$	$V[\hat{\lambda}]$	$\frac{E[\hat{V}[\hat{\lambda}]]}{V[\hat{\lambda}]}$	$V[\tilde{\lambda}]$	$\frac{E[\hat{V}[\hat{\lambda}]]}{V[\tilde{\lambda}]}$	$\frac{E[V[\tilde{\lambda}]]}{V[\tilde{\lambda}]}$	$\frac{MSE[\tilde{\lambda}]}{MSE[\hat{\lambda}]}$
0.0	-0.045	0.149	0.020	1.079	0.023	0.916	0.990	1.073
0.1	-00534	0.135	0.020	1.060	0.024	0.900	0.980	1.036
0.2	-0.061	0.126	0.020	1.042	0.024	0.886	0.972	0.994
0.3	-0.070	0.121	0.020	1.025	0.023	0.871	0.965	0.946
0.4	-0.078	0.121	0.019	1.006	0.022	0.856	0.960	0.892
0.5	-0.087	0.127	0.018	0.982	0.021	0.837	0.952	0.829
0.6	-0.097	0.141	0.017	0.949	0.020	0.809	0.939	0.759
0.7	-0.108	0.170	0.015	0.902	0.018	0.772	0.921	0.675
0.8	-0.123	0.229	0.014	0.842	0.016	0.738	0.918	0.571
0.9	-0.143	0.348	0.011	0.784	0.012	0.756	1.004	0.449
0.99	-0.182	0.526	0.011	0.685	0.009	0.771	1.140	0.424

Table 5.5 contains results for $T = 20$ and 5.6 for $T = 50$. Notice that the coefficient bias is more serious here, and that the reduction by correction is substantial, except for λ close to the unit circle, where more than 50% of the bias remains. The standard

degrees of freedom corrected estimator for the least-squares variance is overstating the true variance when ¸ is small, and this is more serious the closer ¸ is to zero and the sample size smaller. For large values of ¸ the standard expression grossly understates the true variance. Bias correction of the coe±cient estimate makes its variance larger. However, only for ¸ very close to zero it has a detrimental e®ect on the e±ciency. For moderate values of ¸ an improvement of e±ciency can be obtained by bias correction of the coe±cient estimator, and the larger ¸ is the higher the gains will be. The standard expression for the least squares variance, although unbiased to ¯rst-order, is not an appropriate estimator of the variance of the bias corrected coe±cient estimator, because it is too optimistic (understates). However, the specially designed variance estimator which is second-order unbiased, proves to be reasonably accurate, and hence the results are rather positive about the potentials of ¸ and V(¸) to improve on standard ¯rst-order asymptotic inference. Note that it might be possible to achieve still better results by slightly adapting the implementations of our versions of ¸ and V(¸): For instance, ¸ and V(¸) could be made even less biased possibly, by not taking $\hat{C}$ in the respective formulas, but by iterating at least once and using C (the same for $\hat{Z}$): Also ¾^2 could be estimated on the basis of residuals obtained by employing ¸; etc. We plan to examine the e®ects of these factors in simulations yet to be executed.

The above examination should be extended to more general models including some other exogenous explanatory variables. This will be undertaken in a next version of the paper.

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 coe±cients and disturbance variance) in linear regression models with a lagged dependent explanatory variable, we ¯nd here an approximation to the same order for the mean squared error and for the true variance of the least-squares coe±cient estimator. For the latter approximation we ¯nd that its algebraic expression di®ers 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 ¯nite samples that can be assessed by estimating the expression derived in this paper. Its substraction from the standard expression will yield a less biased variance estimator. In that way the analytic results presented in this paper can be of use for producing new methods to improve the accuracy of inference in ¯nite samples of dynamic regression models. Numerical analysis can be undertaken to produce insight into the seriousness of the ¯nite sample inaccuracies of ¯rst-order asymptotic expressions for second moments and also into the ability of the higher-order asymptotic analytical approximations to assess and to correct such discrepancies.

In simple AR(1) models we ¯nd that there seems certainly scope for such improved procedures, because the standard coe±cient variance estimator may understate the true variance of the least-squares estimator by some 30 or 40%, whereas the bias corrected variance estimator is almost unbiased.

The same techniques are also used to approximate the variance of bias corrected coe±cient estimators and to develop accurate estimators for the variance of such corrected estimators. Because the bias correction does not a®ect the leading term of the asymptotic variance of the (corrected) coe±cient estimator, the standard formula can still be used, because it is asymptotically valid. However, the higher-order asymptotic approximations derived here enable the assessment of more accurate (bias corrected) variance estimators, and also produce analytical insight into the potential e±ciency gains or losses due to bias correction. We ¯nd a strong result for AR(1) models regarding the scope for e±ciency improvement. That scope seems to increase with the number of exogenous regressors in the model. The relative magnitude of e±ciency gains is shown to be non-monotonic in the sample size. Hence, bias correction may be more e®ective from an e±ciency point of view when the sample size is moderate than in smaller samples, where the coe±cient bias is usually larger, simply because a moderate coe±cient bias can be assessed more accurately than a huge bias. We also obtain a strong result for the e®ect on e±ciency of di®erent implementations of coe±cient bias correction, and ¯nd that an approach adopted earlier by a good many researchers is sub-optimal from a theoretical point of view.

References

Bartlett, M.S. (1946), On the theoretical speci¯cation and sampling properties of autocorrelated time-series, Journal of the Royal Statistical Society 8, 27-41.

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

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

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

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

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

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

Kiviet, J.F. and G.D.A. Phillips (1998a), Higher-order asymptotic expansions of the least-squares estimation bias in ¯rst-order dynamic regression models, paper presented at ESWM-95, Tokyo, August 1995; revised October 1998.

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

Kiviet, J.F. and G.D.A. Phillips (1999), Bias and mean squared error of least squares estimators in dynamic regression models with a unit root. Paper presented at ESEM-97, Toulouse, France; revised February 1999.

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

MacKinnon, J.G. and A.A. Smith (1998), Approximate bias correction in Econometrics, Journal of Econometrics 85, 205-230.

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

Nagar, A.L. (1959), The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations, Econometrica 27, 575-595.

Orcutt, G.H. and H.S. Winokur (1969), First order regression, 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.

Shenton

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 $\varepsilon \sim N(0; \sigma^2 I_T)$, then the following results hold:

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

$$E(\varepsilon' A \varepsilon - \sigma^2 \mathrm{tr} A)(\varepsilon' B_1 \varepsilon) = 2\sigma^4 \mathrm{tr}(A B_1); \qquad (A.2)$$

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

$$\begin{aligned} E(\varepsilon' A \varepsilon) \textstyle\prod_{j=1}^{2}(\varepsilon' B_j \varepsilon) = {} & \qquad (A.4) \\ \sigma^6[&\mathrm{tr}(A)\mathrm{tr}(B_1)\mathrm{tr}(B_2) + \mathrm{tr}(A)\mathrm{tr}(B_1 B_2) + \mathrm{tr}(A)\mathrm{tr}(B_1 B_2') \\ &+ 2\mathrm{tr}(B_1)\mathrm{tr}(A B_2) + 2\mathrm{tr}(B_2)\mathrm{tr}(A B_1) + 2\mathrm{tr}(A B_2 B_1) \\ &+ 2\mathrm{tr}(A B_2' B_1) + 2\mathrm{tr}(A B_1 B_2) + 2\mathrm{tr}(A B_1 B_2')]; \end{aligned}$$

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

$$E(\varepsilon' A\varepsilon - \sigma^2 \mathrm{tr}A)(\varepsilon\varepsilon' B_1\varepsilon\varepsilon') = \tag{A.6}$$
$$2\sigma^6[\mathrm{tr}(AB_1)I_T + \mathrm{tr}(B_1)A + AB_1 + B_1A + AB_1' + B_1'A];$$

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

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

$$E(\varepsilon\varepsilon' B_1\varepsilon\varepsilon' B_2\varepsilon\varepsilon') = \tag{A.9}$$
$$\sigma^6\{[\mathrm{tr}(B_1)\mathrm{tr}(B_2) + \mathrm{tr}(B_1B_2) + \mathrm{tr}(B_1B_2')]I_T$$
$$+\mathrm{tr}(B_1)B_2 + \mathrm{tr}(B_1)B_2' + \mathrm{tr}(B_2)B_1 + \mathrm{tr}(B_2)B_1'$$
$$+B_1B_2 + B_1'B_2 + B_1B_2' + B_1'B_2'$$
$$+B_2B_1 + B_2'B_1 + B_2B_1' + B_2'B_1'\};$$

$$E(\varepsilon' A\varepsilon)^2(\varepsilon' B_1\varepsilon)^2 = \tag{A.10}$$
$$\sigma^8\{[\mathrm{tr}(A)]^2[\mathrm{tr}(B_1B_1) + \mathrm{tr}(B_1B_1')]$$
$$+[\mathrm{tr}(B_1)]^2[\mathrm{tr}(A)\mathrm{tr}(A) + 2\mathrm{tr}(AA)]$$
$$+4\mathrm{tr}(A)[2\mathrm{tr}(AB_1B_1) + \mathrm{tr}(AB_1B_1') + \mathrm{tr}(AB_1'B_1)]$$
$$+8\mathrm{tr}(B_1)[\mathrm{tr}(A)\mathrm{tr}(AB_1) + 2\mathrm{tr}(AAB_1)]$$
$$+2\mathrm{tr}(AA)[\mathrm{tr}(B_1B_1) + \mathrm{tr}(B_1'B_1)] + 8[\mathrm{tr}(AB_1)]^2$$
$$+16\mathrm{tr}(AAB_1B_1) + 8\mathrm{tr}(AAB_1'B_1) + 8\mathrm{tr}(AAB_1B_1')$$
$$+8\mathrm{tr}(AB_1AB_1) + 8\mathrm{tr}(AB_1AB_1')\};$$

$$E(\varepsilon' A\varepsilon - \sigma^2 \mathrm{tr}A)^2(\varepsilon' B_1\varepsilon)^2 = \tag{A.11}$$
$$2\sigma^8\{8\mathrm{tr}(B_1)\mathrm{tr}(AAB_1)$$
$$+\mathrm{tr}(AA)[\mathrm{tr}(B_1)\mathrm{tr}(B_1) + \mathrm{tr}(B_1B_1) + \mathrm{tr}(B_1'B_1)]$$
$$+4[\mathrm{tr}(AB_1)]^2 + 8\mathrm{tr}(AAB_1B_1) + 4\mathrm{tr}(AAB_1'B_1)$$
$$+4\mathrm{tr}(AAB_1B_1') + 4\mathrm{tr}(AB_1AB_1) + 4\mathrm{tr}(AB_1AB_1')\}.$$

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 one 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 $E(\varepsilon' B_1\varepsilon) = \mathrm{tr}(B_1)$. 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 $V(\hat{\alpha})$

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

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

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

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

where H is the non-symmetric matrix

$$H = G'[0 \vdots I_T]. \qquad (B.3)$$

For H we find the useful results

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

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

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

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

$$
\begin{aligned}
Z'Z &= (\bar{Z} + \tilde{Z})'(\bar{Z} + \tilde{Z}) \qquad (B.6)\\
&= E(Z'Z) - E(\tilde{Z}'\tilde{Z}) + \bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z} + \tilde{Z}'\tilde{Z}\\
&= \left[I_{k+1} + (\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q + [\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q \right] Q^{-1}.
\end{aligned}
$$

Hence,

$$(Z'Z)^{-1} = Q \left[I_{k+1} + (\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q + [\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q \right]^{-1}; \qquad (B.7)$$

where the stochastic terms $(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q$ and $[\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q$ are both $O_p(T^{-1/2})$. The inverse matrix of the form $[I_n + A]^{-1}$, with $A = O_p(T^{-1/2})$ an $n \times n$ matrix, may be expanded in $[I_n - A + A^2 - A^3 + \ldots]$, 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

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

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

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

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.

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

and

$$
\begin{aligned}
W_2 &= [\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z})]Q \\
&= [v'G'Gv - \sigma^2 tr(G'G)]e_1q_1' = O_p(T^{-1/2}).
\end{aligned}
\tag{B.13}
$$

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:

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

Note that

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

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

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

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

$$
\begin{aligned}
&E\,(QW_1Z'uu'ZQ) \qquad \text{(B.16)}\\
= \;& EQ\bar{Z}'Gvq_1'\bar{Z}'[0 \vdots I_T]v(v'Hv)q_1' + EQ\bar{Z}'Gvq_1'(v'Hv)e_1v'[0 \vdots I_T]'\bar{Z}Q \\
& +Eq_1v'G'\bar{Z}Q\bar{Z}'[0 \vdots I_T]v(v'Hv)q_1' + Eq_1v'G'\bar{Z}Q(v'Hv)e_1v'[0 \vdots I_T]'\bar{Z}Q \\
= \;& Q\bar{Z}'GE(vv'Hvv')[0 \vdots I_T]'\bar{Z}q_1q_1' + q_{11}Q\bar{Z}'GE(vv'Hvv')[0 \vdots I_T]'\bar{Z}Q \\
& +E(v'G'\bar{Z}Q\bar{Z}'[0 \vdots I_T]v)(v'Hv)q_1q_1' + q_1q_1'\bar{Z}'GE(vv'Hvv')[0 \vdots I_T]'\bar{Z}Q \\
= \;& \sigma^4Q\bar{Z}'G(H + H')[0 \vdots I_T]'\bar{Z}q_1q_1' + \sigma^4q_{11}Q\bar{Z}'G(H + H')[0 \vdots I_T]'\bar{Z}Q \\
& +\sigma^4q_1q_1'\bar{Z}'G(H + H')[0 \vdots I_T]'\bar{Z}Q + \sigma^4 tr(G'\bar{Z}Q\bar{Z}'[0 \vdots I_T][H + H'])q_1q_1' \\
= \;& \sigma^4Q\bar{Z}'\,(GG' + CC)\,\bar{Z}q_1q_1' + \sigma^4q_{11}Q\bar{Z}'\,(GG' + CC)\,\bar{Z}Q \\
& +\sigma^4q_1q_1'\bar{Z}'\,(GG' + CC)\,\bar{Z}Q + \sigma^4 tr(Q\bar{Z}'CC\bar{Z})q_1q_1' + \sigma^4 tr(Q\bar{Z}'GG'\bar{Z})q_1q_1'.
\end{aligned}
$$

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

$$
\begin{aligned}
&E\,(QW_2Z'uu'ZQ) \qquad \text{(B.17)}\\
= \;& E[v'G'Gv - \sigma^2 tr(G'G)]q_1q_1'\bar{Z}'[0 \vdots I_T]vv'[0 \vdots I_T]'\bar{Z}Q \\
& +E[v'G'Gv - \sigma^2 tr(G'G)]q_1q_1'(v'Hv)^2e_1e_1'Q \\
= \;& 2\sigma^4q_1q_1'\bar{Z}'[0 \vdots I_T]G'G[0 \vdots I_T]'\bar{Z}Q \\
& +2\sigma^6q_{11}\,[2tr\,(G'GHH) + tr\,(G'GH'H) + tr\,(G'GHH')]\,q_1q_1' \\
= \;& 2\sigma^4q_1q_1'\bar{Z}'C'C\bar{Z}Q \\
& +2\sigma^6q_{11}\,[2tr\,(G'GCC) + tr\,(G'C'CG) + tr\,(G'GG'G)]\,q_1q_1',
\end{aligned}
$$

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

$$
\begin{aligned}
&E[Q(W_1 + W_2)^2Z'uu'ZQ] = \qquad \text{(B.18)}\\
&E\,(QW_1W_1Z'uu'ZQ) + E\,(QW_1W_2Z'uu'ZQ) \\
&+E\,(QW_2W_1Z'uu'ZQ) + E\,(QW_2W_2Z'uu'ZQ)\,.
\end{aligned}
$$

We examine these four terms separately. First we have

$$
\begin{aligned}
& E\,(QW_1W_1Z'uu'ZQ) \\
= \; & 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] \\
= \; & E\{Q\bar{Z}'Gvv'G'\bar{Z}q_1q_1'\bar{Z}'[0\vdots I_T]vv'[0\vdots I_T]'\bar{Z}Q\} + q_{11}E[Q\bar{Z}'Gv(v'Hv)^2v'G'\bar{Z}q_1q_1'] \\
& +q_{11}E[Q\bar{Z}'Gvv'G'\bar{Z}Q\bar{Z}'[0\vdots I_T]vv'[0\vdots I_T]'\bar{Z}Q] + q_{11}E[Q\bar{Z}'Gv(v'Hv)^2v'G'\bar{Z}q_1q_1'] \\
& +E[q_1v'G'\bar{Z}Q\bar{Z}'Gvq_1'\bar{Z}'[0\vdots I_T]vv'[0\vdots I_T]'\bar{Z}Q] + q_{11}E[q_1v'G'\bar{Z}Q\bar{Z}'Gv(v'Hv)^2q_1'] \\
& +E[q_1q_1'\bar{Z}'Gvv'G'\bar{Z}Q\bar{Z}'[0\vdots I_T]vv'[0\vdots I_T]'\bar{Z}Q] + 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}\mathrm{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}\mathrm{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\mathrm{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}\mathrm{tr}(Q\bar{Z}'GG'\bar{Z})\mathrm{tr}(G'G)q_1q_1' + 4\sigma^6q_{11}\mathrm{tr}(Q\bar{Z}'CCGG'\bar{Z})q_1q_1' \\
& +2\sigma^6q_{11}\mathrm{tr}(Q\bar{Z}'GCGC'\bar{Z})q_1q_1' + 2\sigma^6q_{11}\mathrm{tr}(Q\bar{Z}'GG'GG'\bar{Z})q_1q_1' \\
& +\sigma^4\mathrm{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)\mathrm{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. If we remove them, and also use $\bar{Z}'\bar{Z}Q = I - \sigma^2\mathrm{tr}(G'G)e_1q_1'$ we obtain

$$
\begin{aligned}
& E\,(QW_1W_1Z'uu'ZQ) \qquad \text{(B.19)} \\
= \; & \sigma^4(q_1'\bar{Z}'C\bar{Z}q_1)Q\bar{Z}'C\bar{Z}Q + \sigma^4Q\bar{Z}'GG'\bar{Z}q_1q_1' + \sigma^4Q\bar{Z}'C\bar{Z}q_1q_1'\bar{Z}'C\bar{Z}Q \\
& +\sigma^4q_{11}\mathrm{tr}(Q\bar{Z}'C\bar{Z})Q\bar{Z}'C\bar{Z}Q + \sigma^4q_{11}Q\bar{Z}'GG'\bar{Z}Q + \sigma^4q_{11}Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}Q \\
& +\sigma^4\mathrm{tr}(Q\bar{Z}'GG'\bar{Z})q_1q_1' + \sigma^4\mathrm{tr}(Q\bar{Z}'C\bar{Z})q_1q_1'\bar{Z}'C\bar{Z}Q + \sigma^4q_1q_1'\bar{Z}'GG'\bar{Z}Q \\
& +2\sigma^4q_1q_1'\bar{Z}'C'\bar{Z}Q\bar{Z}'C\bar{Z}Q + \sigma^4q_1q_1'\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}Q + o(T^{-2}).
\end{aligned}
$$

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

$$
\begin{aligned}
& E\,(QW_1W_2Z'uu'ZQ) \qquad \text{(B.20)} \\
= \; & E\{Q(\bar{Z}'Gvq_1' + e_1v'G'\bar{Z}Q)[v'G'Gv - \sigma^2\mathrm{tr}(G'G)]e_1q_1'\bar{Z}'[0\vdots I_T]v(v'Hv)q_1'\} \\
& +E\{Q(\bar{Z}'Gvq_1' + e_1v'G'\bar{Z}Q)[v'G'Gv - \sigma^2\mathrm{tr}(G'G)]e_1q_1'(v'Hv)e_1v'[0\vdots I_T]'\bar{Z}Q\} \\
= \; & q_{11}E\{Q\bar{Z}'G[v'G'Gv - \sigma^2\mathrm{tr}(G'G)](v'Hv)vv'[0\vdots I_T]'\bar{Z}q_1q_1'\}
\end{aligned}
$$

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

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

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

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

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

Collecting the four terms we find for (B.18), which is the expectation of the fourth term of (B.14):

$$
E[Q(W_1 + W_2)^2 Z'uu'ZQ] \qquad\qquad \text{(B.23)}
$$

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

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

$$
E\,(QZ'uu'ZW_1'Q) = E\,(QW_1Z'uu'ZQ)'; \qquad \text{(B.24)}
$$

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

$$
E\,(QZ'uu'ZW_2'Q) = E\,(QW_2Z'uu'ZQ)'; \qquad \text{(B.25)}
$$

which follows easily from (B.17). Likewise for the expectation of the seventh term of (B.14) we have

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

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

$$
\begin{aligned}
& E\,(QW_1Z'uu'ZW_1'Q) \\
= \quad & EQ(\bar{Z}'Gvq_1' + e_1v'G'\bar{Z}Q)\bar{Z}'[0\vdots I_T]vv'[0\vdots I_T]'\bar{Z}(q_1v'G'\bar{Z} + Q\bar{Z}'Gve_1')Q \\
& + EQ(\bar{Z}'Gvq_1' + e_1v'G'\bar{Z}Q)(v'Hv)^2e_1e_1'(q_1v'G'\bar{Z} + Q\bar{Z}'Gve_1')Q \\
= \quad & EQ\bar{Z}'Gvq_1'\bar{Z}'[0\vdots I_T]vv'[0\vdots I_T]'\bar{Z}q_1v'G'\bar{Z}Q \\
& + EQ\bar{Z}'Gvq_1'\bar{Z}'[0\vdots I_T]vv'[0\vdots I_T]'\bar{Z}Q\bar{Z}'Gvq_1' \\
& + Eq_1v'G'\bar{Z}Q\bar{Z}'[0\vdots I_T]vv'[0\vdots I_T]'\bar{Z}q_1v'G'\bar{Z}Q \\
& + Eq_1v'G'\bar{Z}Q\bar{Z}'[0\vdots I_T]vv'[0\vdots I_T]'\bar{Z}Q\bar{Z}'Gvq_1' \\
& + q_{11}^2EQ\bar{Z}'Gv(v'Hv)^2v'G'\bar{Z}Q + q_{11}EQ\bar{Z}'Gv(v'Hv)^2q_1'\bar{Z}'Gvq_1' \\
& + q_{11}Eq_1v'G'\bar{Z}q_1(v'Hv)^2v'G'\bar{Z}Q + Eq_1v'G'\bar{Z}(v'Hv)^2q_1q_1'\bar{Z}'Gvq_1' \\
= \quad & EQ\bar{Z}'Gvv'[0\vdots I_T]'\bar{Z}q_1q_1'\bar{Z}'[0\vdots I_T]vv'G'\bar{Z}Q \\
& + EQ\bar{Z}'Gvv'[0\vdots I_T]'\bar{Z}Q\bar{Z}'Gvv'[0\vdots I_T]'\bar{Z}q_1q_1' \\
& + Eq_1q_1'\bar{Z}'[0\vdots I_T]vv'G'\bar{Z}Q\bar{Z}'[0\vdots I_T]vv'G'\bar{Z}Q
\end{aligned}
$$

$$
\begin{aligned}
& +Ev'G'\bar{Z}Q\bar{Z}'[0\vdots I_T]vv'[0\vdots I_T]'\bar{Z}Q\bar{Z}'Gvq_1q_1' \\
& +q_{11}^2 EQ\bar{Z}'Gv(v'Hv)^2v'G'\bar{Z}Q + q_{11}EQ\bar{Z}'Gv(v'Hv)^2v'G'\bar{Z}q_1q_1' \\
& +q_{11}Eq_1q_1'\bar{Z}'Gv(v'Hv)^2v'G'\bar{Z}Q + Eq_1q_1'\bar{Z}'Gv(v'Hv)^2v'G'\bar{Z}q_1q_1' \\
= \; & \sigma^4(q_1'\bar{Z}'\bar{Z}q_1)Q\bar{Z}'GG'\bar{Z}Q + 2\sigma^4Q\bar{Z}'C\bar{Z}q_1q_1'\bar{Z}'C'\bar{Z}Q \\
& +\sigma^4\mathrm{tr}(Q\bar{Z}'C\bar{Z})Q\bar{Z}'C\bar{Z}q_1q_1' + \sigma^4Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z}q_1q_1' + \sigma^4Q\bar{Z}'GG'\bar{Z}Q\bar{Z}'\bar{Z}q_1q_1' \\
& +\sigma^4\mathrm{tr}(Q\bar{Z}'C\bar{Z})q_1q_1'\bar{Z}'C'\bar{Z}Q + \sigma^4q_1q_1'\bar{Z}'C'\bar{Z}Q\bar{Z}'C'\bar{Z}Q \\
& +\sigma^4q_1q_1'\bar{Z}'\bar{Z}Q\bar{Z}'GG'\bar{Z}Q + \sigma^4\mathrm{tr}(Q\bar{Z}'C\bar{Z})\mathrm{tr}(Q\bar{Z}'C\bar{Z})q_1q_1' \\
& +\sigma^4\mathrm{tr}(Q\bar{Z}'\bar{Z}Q\bar{Z}'GG'\bar{Z})q_1q_1' + \sigma^4\mathrm{tr}(Q\bar{Z}'C\bar{Z}Q\bar{Z}'C\bar{Z})q_1q_1' \\
& +\sigma^6q_{11}^2\mathrm{tr}(G'G)Q\bar{Z}'GG'\bar{Z}Q + \sigma^6q_{11}\mathrm{tr}(G'G)Q\bar{Z}'GG'\bar{Z}q_1q_1' \\
& +\sigma^6q_{11}\mathrm{tr}(G'G)q_1q_1'\bar{Z}'GG'\bar{Z}Q + \sigma^6\mathrm{tr}(G'G)(q_1'\bar{Z}'GG'\bar{Z}q_1)q_1q_1' + o(T^{-2}).
\end{aligned}
$$

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

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

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

$$
\begin{aligned}
& E\,(QW_1Z'uu'ZW_2'Q) \qquad\qquad \text{(B.28)} \\
= \; & EQ(\bar{Z}'Gvq_1' + e_1v'G'\bar{Z}Q)\bar{Z}'[0\vdots I_T]v(v'Hv)e_1'[v'G'Gv - \sigma^2\mathrm{tr}(G'G)]q_1q_1' \\
& +EQ(\bar{Z}'Gvq_1' + e_1v'G'\bar{Z}Q)(v'Hv)e_1v'[0\vdots I_T]'\bar{Z}[v'G'Gv - \sigma^2\mathrm{tr}(G'G)]q_1q_1' \\
= \; & q_{11}EQ\bar{Z}'Gvq_1'\bar{Z}'[0\vdots I_T]v(v'Hv)[v'G'Gv - \sigma^2\mathrm{tr}(G'G)]q_1' \\
& +q_{11}Ev'G'\bar{Z}Q\bar{Z}'[0\vdots I_T]v(v'Hv)[v'G'Gv - \sigma^2\mathrm{tr}(G'G)]q_1q_1' \\
& +q_{11}EQ\bar{Z}'Gv(v'Hv)v'[0\vdots I_T]'\bar{Z}[v'G'Gv - \sigma^2\mathrm{tr}(G'G)]q_1q_1' \\
& +Eq_1v'G'\bar{Z}q_1(v'Hv)v'[0\vdots I_T]'\bar{Z}[v'G'Gv - \sigma^2\mathrm{tr}(G'G)]q_1q_1' \\
= \; & 2q_{11}EQ\bar{Z}'Gvv'(v'Hv)[v'G'Gv - \sigma^2\mathrm{tr}(G'G)][0\vdots I_T]'\bar{Z}q_1q_1' \\
& +q_{11}Ev'G'\bar{Z}Q\bar{Z}'[0\vdots I_T]v(v'Hv)[v'G'Gv - \sigma^2\mathrm{tr}(G'G)]q_1q_1' \\
& +Eq_1q_1'\bar{Z}'Gvv'(v'Hv)[v'G'Gv - \sigma^2\mathrm{tr}(G'G)][0\vdots I_T]'\bar{Z}q_1q_1' \\
= \; & 4\sigma^6q_{11}\mathrm{tr}(GG'C)Q\bar{Z}'C\bar{Z}q_1q_1' + 2\sigma^6q_{11}\mathrm{tr}(Q\bar{Z}'C\bar{Z})\mathrm{tr}(GG'C)q_1q_1' \\
& +2\sigma^6\mathrm{tr}(GG'C)(q_1'\bar{Z}'C\bar{Z}q_1)q_1q_1' + o(T^{-2}).
\end{aligned}
$$

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

$$
E\,(QW_2Z'uu'ZW_1'Q) = E\,(QW_1Z'uu'ZW_2'Q)'; \qquad\qquad \text{(B.29)}
$$

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

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

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

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

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

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

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

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

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

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

We require an expansion for s^2. For the numerator of this estimator, given in (1.5), we have, upon using (B.10),

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

First we shall examine an approximation to the expectation of the coefficient variance estimator $\hat{\sigma}^2(Z'Z)^{-1}$. Note that (C.1) yields for (1.6) the approximation

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

This can be exploited to obtain an order T^{-2} approximation to

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

by employing (C.2) and an appropriate expansion for $(Z'Z)^{-1}$; see (B.9), and next substituting $\tilde{Z} = Gve_1'$; $\tilde{Z}'\tilde{Z} - E(\tilde{Z}'\tilde{Z}) = [v'G'Gv - \sigma^2\mathrm{tr}(G'G)]e_1e_1'$ while making use of $u = [0 \vdots I_T]v$ and $v \sim N[0; \sigma^2 I_{T+1}]$.

Note that the contribution of the first right-hand term of (C.3) stems from not knowing σ^2 when estimating $V(\hat{\alpha})$. It amounts to:

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

$$
\begin{aligned}
&= -E\left(\frac{1}{T}u'u - \sigma^2\right)Q[v'G'Gv - \sigma^2 tr(G'G)]e_1e_1'Q \\
&\quad - \frac{1}{T}E(u'\bar{Z}Q\bar{Z}'u + u'Gve_1'Qe_1v'G'u)Q + o(T^{-2}) \\
&= -E\left(\frac{1}{T}u'u - \sigma^2\right)(v'G'Gv)q_1q_1' - \frac{1}{T}E(u'\bar{Z}Q\bar{Z}'u + q_{11}u'Gvv'G'u)Q + o(T^{-2}) \\
&= -\frac{2}{T}\sigma^4 tr(C'C)q_1q_1' - \frac{1}{T}\sigma^2 tr(Q\bar{Z}'\bar{Z})Q - q_{11}\frac{1}{T}\sigma^4 tr(G'G)Q + o(T^{-2}) \\
&= -\frac{2}{T}\sigma^4 tr(C'C)q_1q_1' - \frac{1}{T}\sigma^2[k + 1 - \sigma^2 q_{11} tr(G'G)]Q - q_{11}\frac{1}{T}\sigma^4 tr(G'G)Q + o(T^{-2}) \\
&= -\frac{2}{T}\sigma^4 tr(C'C)q_1q_1' - \frac{k+1}{T}\sigma^2 Q + o(T^{-2}).
\end{aligned}
$$

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

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

The second term of (C.5) is

$$
\begin{aligned}
&E[Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q(\bar{Z}'\tilde{Z} + \tilde{Z}'\bar{Z})Q] \qquad (C.6)\\
&= E[Q(\bar{Z}'Gve_1' + e_1v'G'\bar{Z})Q(\bar{Z}'Gve_1' + e_1v'G'\bar{Z})Q] \\
&= E[Q\bar{Z}'Gve_1'Q\bar{Z}'Gve_1'Q] + E[Q\bar{Z}'Gve_1'Qe_1v'G'\bar{Z}Q] \\
&\quad + E[Qe_1v'G'\bar{Z}Q\bar{Z}'Gve_1'Q] + E[Qe_1v'G'\bar{Z}Qe_1v'G'\bar{Z}Q] \\
&= E[Q\bar{Z}'Gvv'G'\bar{Z}Qe_1e_1'Q] + q_{11}E[Q\bar{Z}'Gvv'G'\bar{Z}Q] \\
&\quad + E[q_1v'G'\bar{Z}Q\bar{Z}'Gvq_1'] + E[q_1e_1'Q\bar{Z}'Gvv'G'\bar{Z}Q] \\
&= \sigma^2[Q\bar{Z}'GG'\bar{Z}q_1q_1' + q_{11}Q\bar{Z}'GG'\bar{Z}Q \\
&\quad + tr(Q\bar{Z}'GG'\bar{Z})q_1q_1' + q_1q_1'\bar{Z}'GG'\bar{Z}Q] \\
&= \sigma^2\{[Q\bar{Z}'GG'\bar{Z} + tr(Q\bar{Z}'GG'\bar{Z})I_{k+1}]q_1q_1' + [q_{11}Q + q_1q_1']\bar{Z}'GG'\bar{Z}Q\}.
\end{aligned}
$$

The third term of (C.5) is

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

Gathering terms yields the result

$$\sigma^2 E[(Z'Z)^{-1}] = \sigma^2 Q \qquad \text{(C.8)}$$
$$+\sigma^4 \{tr(Q\bar{Z}'GG'\bar{Z})q_1 q_1'$$
$$+Q\bar{Z}'GG'\bar{Z}q_1 q_1' + q_1 q_1'\bar{Z}'GG'\bar{Z}Q + q_{11}Q\bar{Z}'GG'\bar{Z}Q\}$$
$$+2\sigma^6 q_{11} tr(G'GG'G)q_1 q_1' + o_p(T^{-2}).$$

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

$$\frac{T-k-1}{T}\sigma^2 Q$$
$$+\sigma^4 \{[tr(Q\bar{Z}'GG'\bar{Z}) - \frac{2}{T}tr(C'C)]q_1 q_1'$$
$$+Q\bar{Z}'GG'\bar{Z}q_1 q_1' + q_1 q_1'\bar{Z}'GG'\bar{Z}Q + q_{11}Q\bar{Z}'GG'\bar{Z}Q\}$$
$$+2\sigma^6 q_{11} tr(G'GG'G)q_1 q_1'$$

for $E[\hat{\sigma}^2(Z'Z)^{-1}]$. From this the result of Theorem 2.3 follows upon multiplying the above by $T/(T-k-1)$. This affects the leading term, but not the remaining terms to the order of T^{-2}.

D. The bias of the COLS estimator

Yet to be typed, and so are the appendices E, F, G.