Moment approximation for least squares estimators in dynamic regression models with a unit root

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

Publication date
2003

UvA ECONOMETRICS

Discussion Paper: 2003/03

Moment approximations for least squares estimators in dynamic regression models with a unit root

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

Moment Approximation for Least Squares Estimators in Dynamic Regression Models with a Unit Root*

Jan F. Kiviet† and Garry D.A. Phillips‡

7 April 2003

JEL classification: C13, C22

Keywords: asymptotic expansions, dynamic regression, finite sample bias, moment approximation, unit root

Abstract

To find approximations for bias, variance and mean squared error of least-squares estimators for all coefficients in a linear dynamic regression model with a unit root we derive asymptotic expansions and examine their accuracy by simulation. It is found that in this particular context useful expansions exist only when the autoregressive model contains at least one non-redundant exogenous explanatory variable. Surprisingly, the large-sample and small-disturbance asymptotic techniques give closely related results, which is not the case in stable dynamic regression models. We specialize our general expressions for moment approximations to the case of the random walk with drift model and find that they are unsatisfactory when the drift is small. Therefore, we develop what we call small-drift asymptotics which proves to be very accurate, especially when the sample size is very small.

1. Introduction

In dynamic regression models with normally distributed white noise disturbances least squares (maximum likelihood) estimators may be seriously biased in small samples. Strong evidence for this is provided by Sawa (1978) who used the moment generating function to find exact values for the bias (and variance) of the least squares estimator of the lagged dependent variable coefficient in the case of a constant but no exogenous variables, i.e. the stable AR(1) model. This work was extended by Hoque and Peters (1986) to the case of included exogenous variables under normality assumptions, while Peters (1987) analyzed the same ARX(1) model with non-normal disturbances. These papers provided numerical results for different disturbance structures and exogenous data series.

An alternative approach to investigating the moments of econometric estimators is to find asymptotic approximations. This was the method followed by Grubb and Symons (1987), who used large-T asymptotics in the tradition of Kendall (1954) where T is the sample size. They derived an expression for the bias to the order of T^{-1} of the lagged dependent variable coefficient in the ARX(1) model, while the present authors – henceforth referred to as KP – analyzed the bias of the full coefficient vector, see KP (1993). Later, KP (1994) extended the analysis to the higher-order dynamic regression model, i.e. ARX(p), and Kiviet et al. (1995) to the dynamic

*An early version of this paper was presented at the Econometric Society European Meeting of 1997 in Toulouse, and a later version has been catalogued as Tinbergen Institute discussion paper 2001-118/4. We are greatful for the many constructive remarks we received at various occassions over several years.

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

‡ *Cardiff Business School, Aberconway Building, Colum Drive, CF10 3EU, Cardiff, Wales, UK.*

seemingly unrelated regression model. More recently, KP (1998b) found the bias to the order of T^{-2} in the stable ARX(1) model.

In econometrics there are two main approaches to finding asymptotic approximations to the moments of estimators in models with random regressors. The first was introduced by Nagar (1959), who found large sample approximations to the moments of consistent k-class estimators in a static simultaneous equation model, while a second alternative procedure was employed in the same model by Kadane (1971) based upon small disturbance asymptotics. This yielded small-σ asymptotic approximations which, remarkably, were essentially the same as the large-T ones. However, KP (1993, 1994) compared bias approximations from these two approaches and found that they can produce quite different results in dynamic regression models; in particular, in that context the large-T approximation (which was also used by Grubb and Symons) was found to be superior, both theoretically and numerically.

Asymptotic methods are also used to approximate the distribution of estimators. In the context of stable and unstable dynamic regression models Evans and Savin (1984) employ both large-T and small-σ methods and, because they focus on first-order asymptotic distributions, they establish equivalence. Interesting results on asymptotic distributions in near unit root models have also been obtained by Nabeya and Sørensen (1994), but their results do not yield approximations to the moments of estimators.

The moment approximations in dynamic regression models referred to above were all obtained in stable models with stationary or non-stationary exogenous regressors. Although the large-T approximations, in particular the second-order approximations, are often remarkably accurate, it has also been demonstrated that they are of limited use for models where the AR parameter is close to unity. In KP (1998b) it was even established that for near unit root models an approximation to the order of T^{-2} is generally much more vulnerable than the simpler approximation to order T^{-1}. Given the current interest in non-stationary models, a natural extension of the KP work is to a model in which the stability assumption is relaxed. The need for bias reduction methods in unit root models is also expounded in Abadir (1995).

In this paper we examine the least squares estimator in the normal ARX(1) model when the true coefficient of the lagged dependent variable is unity. This unit root generates stochastic and deterministic trends in the dependent variable which have profound effects on the order of magnitude of the relevant terms in asymptotic expansions. Our major achievements are the derivation of an approximation accurate to order T^{-3} for the bias of the lagged dependent variable coefficient (this bias is of order T^{-2} when the exogenous regressors are stationary) and an approximation accurate to order T^{-4} for the mean squared error (MSE) of this coefficient (when the bias is of order T^{-2}, then the variance and the MSE are of order T^{-3}). In addition, we show that, unlike in the stable case, the large-T and small-σ approximations produce results which are very closely related, with – quite remarkably – the small-σ results potentially superior here because we find that an order T^{-2} approximation does not contain all terms of order σ^2, whereas the order σ^2 approximation does include all contributions of order T^{-2} and also some of the order T^{-3} terms. Attention is also paid to the moments of the full vector of coefficients when the model contains further stationary or non-stationary regressors.

The paper is organized as follows. In Section 2 we distinguish large sample and small disturbance asymptotic methods, focusing on asymptotic expansions for the lagged dependent variable coefficient estimator in ARX(1) models. We identify particular existence problems for these expansions and examine and develop alternatives. In Section 3 we obtain various general large-T and small-σ approximations to finite sample bias, variance and mean squared error in this class of model, and specialize these results for the case of an AR(1) model with a non-zero intercept as the only regressor. Section 4 extends the results for the full coefficient vector of a unit root model with an arbitrary number of possibly non-stationary regressors. In Section 5 we investigate the accuracy of the theoretical results via simulation methods and Section 6 concludes. Proofs are given in a series of Appendices.

2. Expansions for the Unit Root Coefficient

In the Nagar approach to finding moment approximations we commence by expressing the estimation error in terms of stochastic components which are of decreasing order of magnitude in terms of the sample size T. In particular, we determine a positive constant δ such that for an estimator $\widehat{\lambda}$ of the unknown parameter λ we have the expansion

$$T^{\delta}(\widehat{\lambda}-\lambda)=a_0+T^{-1/2}a_1+T^{-1}a_2+T^{-3/2}a_3+\cdots+T^{-q/2}a_q+T^{-(q+1)/2}R_{q+1}, \qquad (2.1)$$

where the a_j, $j=0,...,q$ and the remainder R_{q+1} are all $O_p(1)$ as $T\to\infty$. Notice that the first-order asymptotic distribution is determined by the leading term, i.e. $T^{\delta}(\widehat{\lambda}-\lambda)\overset{L}{\to}a_0$ as $T\to\infty$. Often $\delta=\frac{1}{2}$ but it may take other values in non-stationary models.

The small disturbance approach requires that a normalized estimation error be represented in terms of stochastic components which are of decreasing order of magnitude with respect to the standard deviation of the disturbance term, σ. Typically, the expansion takes the form

$$\frac{1}{\sigma}(\widehat{\lambda}-\lambda)=\dot{a}_0+\sigma\dot{a}_1+\sigma^2\dot{a}_2+\sigma^3\dot{a}_3+\cdots+\sigma^q\dot{a}_q+\sigma^{q+1}\dot{R}_{q+1}, \qquad (2.2)$$

where the $\dot{a}_i$, $i=0,...,q$ and $\dot{R}_{q+1}$ are all bounded in probability as $\sigma\to 0$.

When large-T or small-σ expansions have been found moment approximations can be obtained by dividing the corresponding moments of the retained terms in the expansion by the normalizing constant (i.e. T^{δ} or σ^{-1}). However, there is no standard approach to finding these expansions; we shall return to this point later.

The autoregressive model of our interest will be written

$$y=\lambda y_{-1}+X\beta+u, \qquad (2.3)$$

where the scalar λ and the $k\times 1$ vector β are coefficients whose values are unknown, $y=(y_1,...,y_T)'$ is a $T\times 1$ vector of observations on a dependent variable, y_{-1} is the y vector lagged one time period, i.e. $y_{-1}=(y_0,...,y_{T-1})'$, X is a full column rank $T\times k$ matrix of observations on k strongly exogenous regressors and $u\sim\mathsf{N}(0,\sigma^2)$ is the $T\times 1$ disturbance vector. We shall examine the least-squares estimators of λ and β conditional on X and y_0 (the first observed value for y). In particular we investigate the bias, variance and MSE of these estimators in finite samples. For the moment we shall assume that all components of X are bounded, so $X'X=O(T)$. Assuming that $Z=(y_{-1}:X)$ has rank $k+1$, the least-squares estimator for λ in (2.3) is given by

$$\widehat{\lambda}=\frac{y_{-1}'My}{y_{-1}'My_{-1}}=\lambda+\frac{y_{-1}'Mu}{y_{-1}'My_{-1}}, \qquad (2.4)$$

where $M=I_T-X(X'X)^{-1}X'$. We may write

$$y_{-1}=y_0c(\lambda)+C(\lambda)X\beta+C(\lambda)u, \qquad (2.5)$$

where $c(\lambda)$ is a $T\times 1$ vector and $C(\lambda)$ a $T\times T$ matrix given by

$$c(\lambda)=\begin{pmatrix}1\\ \lambda\\ \lambda^2\\ \cdot\\ \cdot\\ \cdot\\ \lambda^{T-1}\end{pmatrix},\ C(\lambda)=\begin{bmatrix}0&\cdot&\cdot&\cdot&\cdot&\cdot&0\\ 1&0&\cdot&\cdot&\cdot&\cdot&\cdot\\ \lambda&1&\cdot&\cdot&\cdot&\cdot&\cdot\\ \lambda^2&\lambda&\cdot&\cdot&\cdot&\cdot&\cdot\\ \cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot\\ \cdot&\cdot&\cdot&\cdot&\cdot&\cdot&\cdot\\ \lambda^{T-2}&\lambda^{T-3}&\cdot&\cdot&\lambda&1&0\end{bmatrix}. \qquad (2.6)$$

In the unit root case $c(1)$ is just a vector with all components unity, whereas $C(1)$ has zeroes on and above its main diagonal and components unity below. Introducing special notation for this situation we define:

$$\iota := c(1) = \begin{pmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ 1 \end{pmatrix}, \; J := C(1) = \begin{bmatrix} 0 & \cdot & \cdot & \cdot & 0 \\ 1 & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0 & \cdot \\ 1 & \cdot & \cdot & 1 & 0 \end{bmatrix}. \tag{2.7}$$

In Appendix A various properties of expressions in ι and J are collected, which will be used in the derivations to follow. Due to the unit root these differ markedly from those given for $C(\lambda)$ for $|\lambda| < 1$ in Appendix A of KP (1998a).

We shall focus now on the situation where λ is unknown and is estimated by least-squares, but where actually $\lambda \equiv 1$, i.e. model (2.3) has a unit root. Hence, (2.5) specializes into

$$y_{-1} = y_0\iota + JX\beta + Ju, \tag{2.8}$$

and, assuming the presence of a constant in the model so that $M\iota = 0$, we have

$$My_{-1} = MJX\beta + MJu. \tag{2.9}$$

Hence, My_{-1} is free of y_0 so that $\widehat{\lambda}$ in (2.4) does not depend on y_0. This is in sharp contrast to the stable model for which KP (1998b) showed to what extent finite sample bias is affected by the actual value and stochastic properties of the start-up value y_0.

From (2.4) and (2.9) we find for the estimation error of the unit root model

$$\widehat{\lambda} - 1 = \frac{\beta'X'J'Mu + u'J'Mu}{\beta'X'J'MJX\beta + 2\beta'X'J'MJu + u'J'MJu} \tag{2.10}$$

for which an expansion is to be developed. We first focus on obtaining large-T results. To proceed, we have to examine the orders of magnitude of all terms in the above ratio. For the present setting where $X'X = O(T)$, which will be relaxed in Section 4, this is done in Appendix B. Assuming $\beta \neq 0$ (i.e. not all regressors are redundant), we may rewrite the estimation error (2.10) as

$$\widehat{\lambda} - 1 = \left(\frac{\beta'X'J'Mu + u'J'Mu}{\beta'X'J'MJX\beta}\right)\left(1 + \frac{2\beta'X'J'MJu + u'J'MJu}{\beta'X'J'MJX\beta}\right)^{-1}, \tag{2.11}$$

and if β is fixed and finite i.e. $\beta = O(1)$, it follows that the first factor of (2.11) has two terms which are $O_p(T^{-3/2})$ and $O_p(T^{-2})$ respectively, and in the other (inverted) factor there are two ratios which are of order $O_p(T^{-1/2})$ and $O_p(T^{-1})$ respectively. In going from (2.10) to (2.11) we have divided both the numerator and denominator of (2.10) by the ("largest") term $\beta'X'J'MJX\beta$, which we call the base. Notice that it appears in the denominator of the inverse term in (2.11) so that when this inverse is expanded as a power series successive terms are of decreasing order in probability. Using the very simple expansion

$$\left(1 + \frac{2\beta'X'J'MJu + u'J'MJu}{\beta'X'J'MJX\beta}\right)^{-1} = 1 - 2\frac{\beta'X'J'MJu}{\beta'X'J'MJX\beta} + O_p(T^{-1}),$$

and upon omitting in (2.11) all terms of stochastic magnitude $o_p(T^{-2})$, it is easily shown that

$$\widehat{\lambda} - 1 = \frac{\beta'X'J'Mu}{\beta'X'J'MJX\beta} + \frac{u'J'Mu}{\beta'X'J'MJX\beta} - 2\frac{\beta'X'J'Muu'J'MJX\beta}{(\beta'X'J'MJX\beta)^2} + o_p(T^{-2}), \tag{2.12}$$

where the first term is $O_p(T^{-3/2})$ and the remaining two terms are $O_p(T^{-2})$. Hence, the first term of (2.12) determines the first-order asymptotic distribution of the estimator. Under the assumed conditions it is readily shown that the limiting distribution is still normal, i.e.

$$T^{3/2}(\widehat{\lambda}-1) \xrightarrow{L} \mathsf{N}\left(0, \lim_{T\to\infty} \frac{\sigma^2}{\frac{1}{T^3}\beta' X' J' MJX\beta}\right), \tag{2.13}$$

but that the rate of convergence is faster ($\delta = 3/2$) than in the stable model. This surprising result is well-known now, see for example West (1988) and Banerjee et al. (1993, Chapter 6), who give particular attention to the case where there is a constant term but no exogenous variables.

It is of interest to note that in (2.10) the denominator naturally contains a decomposition of the term $y_{-1}' My_{-1}$ into its stochastic and non-stochastic parts such that the non-stochastic part $\beta' X' J' MJX\beta$ is independent of σ^2 whereas it is also the "largest" term and subsequently may then form a suitable base for the expansion of the denominator. This differs from the situation in the stable dynamic model and the approach followed by Kendall (1954), Grubb and Symons (1987) and KP (1993, 1994, 1998a,b) where the base chosen for the expansion was $\mathsf{E}(y_{-1}' My_{-1})$ which is linear (but not generally affine) in σ^2. As a result the large-T and the small-σ expansions yield qualitatively different results, as shown in KP (1993, 1994). In fact the small-σ results were shown to be marked inferior in the stable case because any finite order small-σ approximation omits terms of order T^{-1}.

In the present nonstable model, where the base $\beta' X' J' MJX\beta$ is independent of σ^2, we find a strong correspondence between the large-T and small-σ asymptotic results. To see this, consider the expansion in (2.12). If u is replaced by $\sigma\varepsilon$, where ε is a vector of independent standard normal variables, then the expansion involves terms of increasing order in σ, so that the expansions to $O_p(\sigma)$ and $O_p(\sigma^2)$ coincide with the expansions to $O_p(T^{-3/2})$ and $O_p(T^{-2})$ respectively. However, when focusing on bias, the expansions again show a difference between small-σ and large-T approximations. Below, we will prove that the third term in the expansion (2.12), which is $O_p(T^{-2})$, has an expectation that is $O(T^{-3})$, which is thus omitted from the large-T approximation to order T^{-2}. Therefore the $O(\sigma^2)$ contribution to the bias contains, apart from the components of order of magnitude T^{-2}, some contributions of order T^{-3}. So, surprisingly, in this unit root model the first-order small-σ bias approximation includes a contribution which is of second-order in a large-T sense.

Close correspondence of large-T and small-σ asymptotic results has earlier been established between the findings of Nagar (1959) and Kadane (1971) for consistent k-class estimators in a static simultaneous equation framework. As shown in KP (1996) this equivalence breaks down, however, for the inconsistent least-squares estimator in the static simultaneous equations model, where an appropriate base for the expansion is again linear in σ^2, whereas it is independent of σ^2 for consistent estimators.

One specific finding in KP (1993) is that in the stable model the small-σ expansion is not feasible when $y_0 = 0$ and $\beta = 0$, because the estimator $\widehat{\lambda}$ is invariant with respect to σ in that case. When $\beta = 0$ in the present unit root model small-σ is again not feasible because the estimation error (2.10) reduces then to a simple ratio of quadratic forms in standard normal variates. Hence, its moments can be accurately determined by well-known numerical methods. Nevertheless it is instructive to examine large-T expansions for this case. Using the expectation of the denominator as a base, and substituting $\varepsilon = u/\sigma$, we obtain

$$\begin{aligned} \widehat{\lambda}-1 &= \frac{u' J' Mu}{u' J' MJu} = \frac{\varepsilon' J' M\varepsilon}{\varepsilon' J' MJ\varepsilon} \\ &= \frac{\varepsilon' J' M\varepsilon}{\mathsf{E}(\varepsilon' J' MJ\varepsilon)}\left[1 + \frac{\varepsilon' J' MJ\varepsilon - \mathsf{E}(\varepsilon' J' MJ\varepsilon)}{\mathsf{E}(\varepsilon' J' MJ\varepsilon)}\right]^{-1}. \end{aligned} \tag{2.14}$$

Since $\varepsilon' J' M J \varepsilon - \mathsf{E}(\varepsilon' J' M J \varepsilon) = O_p(T^2)$ and $\mathsf{E}(\varepsilon' J' M J \varepsilon) = O(T^2)$ the random term in the inverse factor is $O_p(1)$ and not $O_p(T^{-\kappa})$ with $\kappa > 0$ as would be required for a converging expansion. An alternative formulation is however

$$\frac{\varepsilon' J' M \varepsilon}{\varepsilon' J' M J \varepsilon} = \frac{\varepsilon' J' M \varepsilon}{\varepsilon' J' J \varepsilon} \left[1 - \frac{\varepsilon' J' X (X'X)^{-1} X J \varepsilon}{\varepsilon' J' J \varepsilon} \right]^{-1} \tag{2.15}$$

and now the random term in the inverse factor is $O_p(T^{-1/2})$ enabling a valid expansion. However, evaluation of moment approximations from this expansion requires the evaluation of products of ratios of stochastic terms, and hence is in fact more involved than obtaining the expectation of the left-hand side simple ratio directly.

As we shall see from the simulations in Section 5, the accuracy of our large-T or small-σ expansions based moment approximations, to be obtained in Sections 3 and 4, deteriorates when β gets close to zero. Therefore it could be worthwhile to develop special approximations for the case where β gets local to zero in some sense. Thus, alongside the cases $\beta = O(1)$ and $\beta = 0$, we examined $\beta = O(T^{-\delta})$ and $\beta = O(\sigma^{\delta})$ for $\delta > 0$. The result is that for $0 < \delta < \frac{1}{2}$ the original expansion is valid and so yields similar unsatisfactory results. For $\delta = \frac{1}{2}$ no valid expansion can be found, while for $\delta > \frac{1}{2}$ the largest term in the numerator of (2.10) is $u' J' M u$ and in the denominator $u' J' M J u$. This implies the same problems as in the $\beta = 0$ case. Hence, expansions for β local to zero in a small-σ or in a large-T sense do either not exist, or are inaccurate, or they cannot be usefully employed in this context.

However, an expansion in which β gets small in its own right will prove to be useful. We will label this small-drift asymptotics, which operates as follows. Consider the random walk with drift model $y_t = \lambda y_{t-1} + \beta + u_t$ with start-up value y_0 and where $\lambda \equiv 1$. Defining $y_t^* = (y_t - y_0)/\sigma$, the model can be rewritten as

$$y_t^* = \lambda y_{t-1}^* + \beta^* + \varepsilon_t, \tag{2.16}$$

with $\lambda = 1$, $\beta^* = \beta/\sigma$, $y_0^* = 0$ and $\varepsilon_t \sim \text{iid}\,\mathsf{N}(0,1)$. Note that β^* is the drift standardized by σ. Defining $A \equiv I_T - \iota\iota'/T$, we can rewrite (2.10) for this special case as

$$\widehat{\lambda} - 1 = \frac{\iota' J' A \varepsilon \beta^* + \varepsilon' J' A \varepsilon}{\iota' J' A J \iota \beta^{*2} + 2\iota' J' A J \varepsilon \beta^* + \varepsilon' J' A J \varepsilon}, \tag{2.17}$$

where now, when β^* gets small, $\varepsilon' J' A J \varepsilon$ is the largest term in the denominator, which, given

$$\widehat{\lambda} - 1 = \frac{\varepsilon' J' A \varepsilon + \iota' J' A \varepsilon \beta^*}{\varepsilon' J' A J \varepsilon} \left[1 + \frac{2\iota' J' A J \varepsilon \beta^* + \iota' J' A J \iota \beta^{*2}}{\varepsilon' J' A J \varepsilon} \right]^{-1},$$

suggests the expansion (accurate to order β^{*2})

$$\widehat{\lambda} - 1 = \frac{\varepsilon' J' A \varepsilon + \iota' J' A \varepsilon \beta^*}{\varepsilon' J' A J \varepsilon} \left[1 - \frac{2\iota' J' A J \varepsilon \beta^* + \iota' J' A J \iota \beta^{*2}}{\varepsilon' J' A J \varepsilon} + 4 \frac{(\iota' J' A J \varepsilon)^2 \beta^{*2}}{(\varepsilon' J' A J \varepsilon)^2} \right] + O_p(\beta^{*3}).$$

Taking expectations we lose all terms in β^* and obtain

$$\begin{aligned} \mathsf{E}(\widehat{\lambda} - 1) &= \mathsf{E}\left(\frac{\varepsilon' J' A \varepsilon}{\varepsilon' J' A J \varepsilon} \right) \\ &\quad - \beta^{*2} \mathsf{E}\left[\frac{(\iota' J' A J \iota) \varepsilon' J' A \varepsilon + 2\varepsilon' A J \iota \iota' J' A J \varepsilon}{(\varepsilon' J' A J \varepsilon)^2} - \frac{4\varepsilon' J' A \varepsilon (\iota' J' A J \varepsilon)^2}{(\varepsilon' J' A J \varepsilon)^3} \right] + o(\beta^{*2}). \end{aligned} \tag{2.18}$$

Note that the two right-hand side expectations are functions of T only, since they do not involve any unknown parameters, and hence have to be calculated only once. The first term is the least-squares bias of $\widehat{\lambda}$ in the unit root model with zero start-up and zero drift, see (2.14), and the second term approximates to order $O(\beta^{*2})$ the incrementation in bias due to a non-zero drift. We can rewrite (2.18) as

$$\mathsf{E}(\widehat{\lambda}-1) = g_0(T) - g_1(T)\beta^{*2} + o(\beta^{*2}), \tag{2.19}$$

and the functions $g_i(T)$, $i = 0, 1$ can be obtained either by sophisticated analytical approximations or by straightforward simulation. We chose the latter and examined $10 \leq T \leq 80$. Using 10^7 replications we inferred from the standard errors of the estimates of these expectations that the precision for $g_0(T)$ is satisfactory (standard error smaller than 0.0001), whereas for $g_1(T)$ these are much larger, especially when T is small, though not exceeding 0.015 when $T \geq 15$. Since $g_1(T)$ is to be multiplied by β^{*2}, which should be small, the overall precision seems satisfactory. Graphs of these functions, to be employed in Section 5, are given in Figure 2.1.

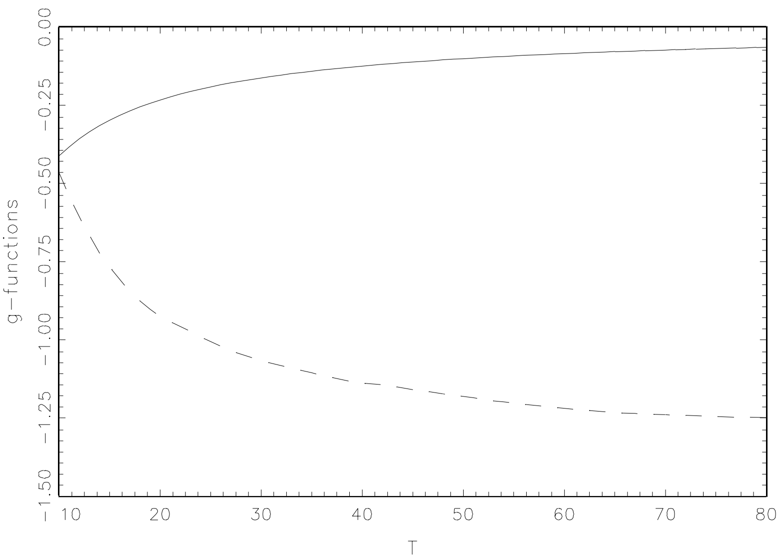

Figure 2.1: Functions $g_0(T)$ [——] and $g_1(T)$ [– – –]

3. The Moments of the Unit Root Coefficient Estimator

We now derive in the model with arbitrary X matrix approximations to the bias, the variance and the mean squared error of the estimator $\widehat{\lambda}$ given in (2.4) according to large-T and small-σ principles. An approximation to the bias accurate to $O(T^{-2})$ is obtained by summing the expected values of the three terms in (2.12). Since the expected value of the first term is zero and that of the third term is of order T^{-3}, just the second term determines the $O(T^{-2})$ bias. Extending this expansion and including all terms of $O_p(T^{-3})$ leads to the following result (proved in Appendix C).

Theorem 1. *In the first-order dynamic regression model (2.3) where the coefficient of the lagged dependent variable* λ *is equal to unity,* $\beta \neq 0$, $\beta = O(1)$ *and* $X'X = O(T)$, *the bias of* *the least squares estimator of λ to the order of T^{-3} is given by:*

$$\begin{aligned}\mathsf{E}(\widehat{\lambda}-1) &= \frac{\sigma^2\left[\operatorname{tr}(MJ)+1\right]}{\beta'X'J'MJX\beta}-\frac{\sigma^4\operatorname{tr}(MJ)\operatorname{tr}(J'MJ)}{\left(\beta'X'J'MJX\beta\right)^2}\\ &\quad+\frac{4\sigma^4\operatorname{tr}(MJ)\,\beta'X'J'MJJ'MJX\beta}{\left(\beta'X'J'MJX\beta\right)^3}+o(T^{-3}). \qquad (3.1)\end{aligned}$$

Note that $\widehat{\lambda}$ is unbiased to order $T^{-3/2}$ and also to order σ. Also note that the bias of $\widehat{\lambda}$ is $O(T^{-2})$ and that the bias to order T^{-2} is given by

$$\mathsf{E}(\widehat{\lambda}-1)=-\sigma^2\frac{\operatorname{tr}\{(X'X)^{-1}X'JX\}}{\beta'X'J'MJX\beta}+o(T^{-2}), \qquad (3.2)$$

whereas an approximation to order σ^2 incorporates an extra $O(T^{-3})$ contribution, viz. the term $\sigma^2\left(\beta'X'J'MJX\beta\right)^{-1}$. Hence, large-$T$ and small-σ asymptotic expansions correspond here more closely than in the stable dynamic model (where small-σ is inferior because any finite order small-σ approximation omits terms of order T^{-1}), but they are not equivalent and the leading term of small-σ incorporates contributions here, which are omitted in the leading term of the large-T approximation.

The case where there is a constant term but no further exogenous variables is of particular interest. The corresponding bias can be obtained by substituting $X=\iota$ in Theorem 1, but the resulting expression will then include also some elements which are $o(T^{-3})$. These can be eliminated; the resulting "trimmed" expression is evaluated in Appendix D, leading to the following result.

Corollary 1. *If in the model of Theorem 1 we have $X=\iota$ with finite intercept $\beta\neq 0$, then the bias simplifies to:*

$$\mathsf{E}(\widehat{\lambda}-1)=-6\left(\frac{\sigma}{\beta}\right)^2\frac{1}{T^2}+18\left(\frac{\sigma}{\beta}\right)^2\frac{1}{T^3}-\frac{84}{5}\left(\frac{\sigma}{\beta}\right)^4\frac{1}{T^3}+o(T^{-3}). \qquad (3.3)$$

From this we see that the bias is always negative to the order T^{-2}, and that the magnitude of the bias crucially depends on the ratio σ/β. From Corollary 1 it is fully evident that an approximation to order $O(\sigma^2)$ incorporates some of the order T^{-3} bias, viz. a positive contribution, whereas a negative order T^{-3} contribution is omitted because it is $O(\sigma^4)$. Note that when $\sigma/\beta=(90/84)^{1/2}\approx 1.15$ the two $O(T^{-3})$ terms cancel.

To obtain an approximation to the MSE of $\widehat{\lambda}$ we use (2.11) and write

$$(\widehat{\lambda}-1)^2=\left(\frac{\beta'X'J'Mu+u'J'Mu}{\beta'X'J'MJX\beta}\right)^2\left(1+\frac{2\beta'X'J'MJu+u'J'MJu}{\beta'X'J'MJX\beta}\right)^{-2}. \qquad (3.4)$$

Expanding the right hand side term as a power series in which successive terms are of increasing powers of $O_p(T^{-1/2})$ yields the following (proof in Appendix E).

Theorem 2. *In the model of Theorem 1 the* MSE *of the least squares estimator of λ to the order of T^{-4} is given by:*

$$\begin{aligned}\mathsf{E}(\widehat{\lambda}-1)^2 &= \frac{\sigma^2}{\beta'X'J'MJX\beta}+\sigma^4\frac{\{[\operatorname{tr}(MJ)]^2+\operatorname{tr}(JMJM)-\operatorname{tr}(J'MJ)\}}{\left(\beta'X'J'MJX\beta\right)^2}\\ &\quad+4\sigma^4\frac{\left(\beta'X'J'MJJ'MJX\beta-\beta'X'J'MJMJMJX\beta\right)}{\left(\beta'X'J'MJX\beta\right)^3}+o(T^{-4}). \qquad (3.5)\end{aligned}$$

Because the squared bias of $\widehat{\lambda}$ is $O(T^{-4})$ the first term in the above expression, which is the only $O(T^{-3})$ contribution to the MSE, establishes also an approximation to $\mathrm{Var}(\widehat{\lambda}) = \mathrm{MSE}(\widehat{\lambda}) - [\mathsf{E}(\widehat{\lambda}-1)]^2$. For the special case of a constant and no further exogenous regressors this yields:

Corollary 2. *If in the model of Theorem 2 we have $X = \iota$ with finite intercept $\beta \neq 0$, then the* MSE *and variance simplify to*

$$\mathsf{E}(\widehat{\lambda}-1)^2 = 12\left(\frac{\sigma}{\beta}\right)^2 \frac{1}{T^3} + \frac{336}{5}\left(\frac{\sigma}{\beta}\right)^4 \frac{1}{T^4} + o(T^{-4}) \tag{3.6}$$

and

$$\mathrm{Var}(\widehat{\lambda}) = 12\left(\frac{\sigma}{\beta}\right)^2 \frac{1}{T^3} + \frac{156}{5}\left(\frac{\sigma}{\beta}\right)^4 \frac{1}{T^4} + o(T^{-4}). \tag{3.7}$$

From the results of the two corollaries it is apparent that for a given sample size T the quality of the approximations will deteriorate as $|\sigma/\beta|$ increases above unity. Thus the smaller the absolute value of the standardized drift $|\beta/\sigma|$ is, the larger the sample size will need to be to achieve a desired accuracy. This point will be addressed in Section 5 where the accuracy of the approximations will be examined and compared with their "untrimmed" counterparts, but also with the specially designed small-drift asymptotic approximations.

4. The Moments of the Full Coefficient Vector

We now approximate the first two moments of the least squares estimator of the full vector of coefficients (λ, β'), and at the same time we shall relax the assumption on the stationarity of the exogenous regressors. We rewrite model (2.3) as

$$y = \lambda y_{-1} + X\beta + u = Z\alpha + u, \tag{4.1}$$

where $Z = (y_{-1} : X)$, $\alpha = (\lambda, \beta')'$ and the least squares estimator

$$\widehat{\alpha} = (\widehat{\lambda}, \widehat{\beta}')' = (Z'Z)^{-1}Z'y \tag{4.2}$$

has estimation error

$$\begin{pmatrix} \widehat{\lambda} - 1 \\ \widehat{\beta} - \beta \end{pmatrix} = \widehat{\alpha} - \alpha = (Z'Z)^{-1}Z'u. \tag{4.3}$$

In the case where all regressors X are stationary the estimation error of $\widehat{\lambda}$ is $O_p(T^{-3/2})$ while that of $\widehat{\beta}$ is $O_p(T^{-1/2})$. If some of the regressors in X are non-stationary this affects the order of probability of both the corresponding coefficients' estimation error and that of $\widehat{\lambda}$. For regressors that are $I(1)$, i.e. integrated of order one, the estimation error will be $O_p(T^{-3/2})$, and if such a regressor has a non-zero coefficient the dependent variable will in principle be $I(2)$, due to the unit root, which reduces the estimation error of $\widehat{\lambda}$ to $O_p(T^{-5/2})$; the same happens when a non-redundant linear deterministic trend occurs in the model.

To facilitate the development of an appropriate asymptotic expansion for general X matrices we shall rescale the regressors and coefficients so that all components of the rescaled estimation error vector are of the same stochastic magnitude. Thus, we consider the $(k+1) \times (k+1)$ diagonal matrix D designed such that:

$$\left.\begin{aligned} D &= diag\,(d_1, \cdots, d_{k+1}) \\ d_i &= T^{\delta_i}\ , \quad (i = 1, ..., k+1) \\ DZ'ZD &= O_p(T) \end{aligned}\right\} \tag{4.4}$$

In the unit root model with stationary X we should have $\delta_1 = -1$ and $\delta_i = 0$ for $i > 1$; in a model with $k = 2$, where the first column of X corresponds to the constant and the second is a linear trend, we should select $\delta_1 = -2$ (if the trend coefficient is nonzero and $\delta_1 = -1$ otherwise), $\delta_2 = 0$ and $\delta_3 = -1$. The model is now

$$y = ZD(D^{-1}\alpha) + u, \tag{4.5}$$

with rescaled coefficients $D^{-1}\alpha$ and estimation error

$$D^{-1}\,(\widehat{\alpha} - \alpha) = (DZ'ZD)^{-1}DZ'u. \tag{4.6}$$

To simplify subsequent analysis, we put

$$W = ZD = \bar{Z}D + \tilde{Z}D = \bar{W} + \tilde{W}, \tag{4.7}$$

where $\bar{W} = \bar{Z}D = \mathsf{E}\,(Z)\,D$ is nonstochastic and $\tilde{W} = \tilde{Z}D = (Z - \bar{Z})D$ is stochastic with zero mean. Since $\tilde{Z} = Jue_1'$, with $e_1 = (1, 0, \cdots, 0)'$ a $(k+1) \times 1$ unit-vector, we may write $\tilde{W} = Jue_1'D$. Now (4.6) can be expressed as

$$D^{-1}\,(\widehat{\alpha} - \alpha) = (\bar{W}'\bar{W} + \bar{W}'\tilde{W} + \tilde{W}'\bar{W} + \tilde{W}'\tilde{W})^{-1}(\bar{W} + \tilde{W})'u. \tag{4.8}$$

Note that D is designed such that $\bar{W}'\bar{W} = D\bar{Z}'\bar{Z}D = O(T)$, $\bar{W}'\tilde{W} = D\bar{Z}'\tilde{Z}D = O_p(T^{1/2})$, $\tilde{W}'\tilde{W} = D\tilde{Z}'\tilde{Z}D = O_p(1)$, $\bar{W}'u = O_p(T^{1/2})$ and $\tilde{W}'u = O_p(1)$. Assuming that $\bar{W}'\bar{W}$ is invertible, and putting

$$R = (\bar{W}'\bar{W})^{-1},\ P = \bar{W}'\tilde{W}R + \tilde{W}'\bar{W}R,\ S = \tilde{W}'\tilde{W}R, \tag{4.9}$$

where $R = O(T^{-1})$, $P = O_p(T^{-1/2})$ and $S = O_p(T^{-1})$, we may write

$$D^{-1}\,(\widehat{\alpha} - \alpha) = R(I + P + S)^{-1}(\bar{W} + \tilde{W})'u, \tag{4.10}$$

and the inverse matrix can be expanded with successive terms being of descending stochastic order. It is our intention here to find a stochastic expansion of (4.10) including terms up to $O_p(T^{-1})$ only. Hence, it will suffice to approximate the inverse matrix by

$$(I + P + S)^{-1} = I - P + o_p(T^{-1/2}).$$

The required expansion is then

$$\begin{aligned} D^{-1}\,(\widehat{\alpha} - \alpha) &= R(I - P)(\bar{W} + \tilde{W})'u + o_p(T^{-1}) \\ &= R\bar{W}'u + R\tilde{W}'u - RP\bar{W}'u + o_p(T^{-1}), \end{aligned} \tag{4.11}$$

from which the following bias approximation readily follows (see Appendix G).

Theorem 3. *In the first-order dynamic regression model (4.1), where the coefficient of the lagged dependent variable λ is equal to unity, the regressor matrix $Z = (y_{-1} : X)$ and the* *scaling matrix* $D = diag\,(d_1, \cdots, d_{k+1})$, *with* $d_i = T^{\delta_i}$ $(i = 1, ..., k+1)$, *is such that* $DZ'ZD = O_p(T)$, *the bias of the least squares estimator of the separate elements of the coefficient vector* $\alpha = (\lambda, \beta')'$ *can be approximated, provided that* $\bar{Z} = (y_0\iota + JX\beta : X)$ *has full column rank and* β *is finite and non-zero, as* $(i = 1, ..., k)$:

$$\begin{aligned} \mathsf{E}(\widehat{\beta}_i - \beta_i) &= -\sigma^2 e'_{i+1}[(\bar{Z}'\bar{Z})^{-1}\bar{Z}'J\bar{Z} + \frac{1}{2}(T-k-1)I_{k+1}](\bar{Z}'\bar{Z})^{-1}e_1 \\ &\quad + o(T^{-1+\delta_{i+1}}), \end{aligned} \tag{4.12}$$

and

$$\mathsf{E}(\widehat{\lambda} - 1) = -\frac{1}{2}(T-k)\sigma^2 e'_1(\bar{Z}'\bar{Z})^{-1}e_1 + o(T^{-1+\delta_1}). \tag{4.13}$$

This bias approximation of order $O(T^{-1+\delta_1})$ for $\widehat{\lambda}$ is equivalent to the $O(T^{-2})$ expression given in (3.2). From this, and more generally from the lines followed in the proof of Theorem 3, it is evident that non-stationarity of the regressors does not change the algebraic form of the approximations; the principal difference is just that the various terms in the approximations may be of smaller order of magnitude. Hence, the full approximation given in Theorem 1 also applies to a model which includes a nonredundant $I(1)$ regressor or a linear trend, but then its accuracy is actually of order $O(T^{-4})$ rather than $O(T^{-3})$.

Finally we shall derive an approximation to the MSE of all elements of the coefficient vector. From (4.10) we obtain the expansion

$$D^{-1}(\widehat{\alpha} - \alpha) = R(I - P - S + PP)\bar{W}'u + R(I-P)\tilde{W}'u + o_p(T^{-3/2}), \tag{4.14}$$

from which an expansion for $D^{-1}(\widehat{\alpha} - \alpha)(\widehat{\alpha} - \alpha)' D^{-1}$ to order T^{-2} easily follows and this yields (proof in Appendix H):

Theorem 4. *In the model of Theorem 3 the* elements of the MSE($\widehat{\alpha}$) *matrix, i.e.* $\mathsf{E}(\widehat{\alpha}_i - \alpha_i)(\widehat{\alpha}_j - \alpha_j)$ *for* $i, j = 1, ..., k+1$, *are given by*

$$\begin{aligned} &\sigma^2 e'_i Q e_j \\ &+\sigma^4[\mathrm{tr}(Q\bar{Z}'JJ'\bar{Z}) - 2\,\mathrm{tr}(Q\bar{Z}'JJ\bar{Z}) - \mathrm{tr}\,(J'J) \\ &\qquad + \mathrm{tr}(Q\bar{Z}'J\bar{Z}Q\bar{Z}'J\bar{Z}) + \mathrm{tr}(Q\bar{Z}'J\bar{Z})\,\mathrm{tr}(Q\bar{Z}'J\bar{Z})](e'_iQe_1)(e'_jQe_1) \\ &+\sigma^4(e'_1Qe_1)(e'_iQ\bar{Z}'[JJ' - JJ - J'J' + J\bar{Z}Q\bar{Z}'J + J'\bar{Z}Q\bar{Z}'J']\bar{Z}Qe_j) \\ &+\sigma^4(e'_1Qe_j)(e'_1Q\bar{Z}'[JJ' - J'J - J'J']\bar{Z}Qe_i) \\ &+\sigma^4(e'_1Qe_i)(e'_1Q\bar{Z}'[JJ' - J'J - JJ]\bar{Z}Qe_j) \\ &+\sigma^4[(e'_1Q\bar{Z}'J\bar{Z}Qe_1) + \mathrm{tr}(Q\bar{Z}'J\bar{Z})(e'_1Qe_1)](e'_iQ\bar{Z}'[J + J']\bar{Z}Qe_j) \\ &+o(T^{-2+\delta_i+\delta_j}), \end{aligned} \tag{4.15}$$

where $Q = (\bar{Z}'\bar{Z})^{-1}$, $\bar{Z} = \mathsf{E}(Z)$ *and* e_i *is the* i^{th} *unit vector.*

From the results in Theorems 3 and 4 approximations to the elements of Var($\widehat{\alpha}$) can be obtained straightforwardly.

5. The Accuracy of the Approximations

In this section the accuracy of the approximations is examined in the context of three types of autoregressive models with a unit root, viz. (i) the AR(1) model with a constant only, (ii) the same model including a linear trend as well, and (iii) the model where this linear trend has been replaced by an arbitrary exogenous regressor generated according to a stationary AR(1) process. Although some of the moments of these least squares estimators can be obtained by numerical integration, see Evans and Savin (1984) and Paolella (2003), we shall estimate them straightforwardly by simulation. With a sufficiently large number of replications, the exact moments can be obtained to a high degree of accuracy so that the estimated moments can be taken as almost exact for the purpose at hand. Our estimates of true bias, variance and MSE presented below are based on 10^5 replications and we also present estimated standard errors of our Monte Carlo estimates (indicated by MCSE).

For the random walk with drift case, the model actually simulated was $y_t^* = \lambda y_{t-1}^* + \beta^* + \varepsilon_t$ $(t = 1, ..., T)$ of (2.16), where $y_0^* = 0$, $\lambda = 1$, $\beta^* = \beta/\sigma \neq 0$. Note that we have already found that $\widehat{\lambda}$ is invariant with respect to y_0, so taking this to be zero has no consequences for our findings on $\widehat{\lambda}$. From (2.10) it is directly seen that the properties of $\widehat{\lambda}$ are not determined by β and σ separately, but only by their ratio, and that is why we scaled the simulation model and gave it unit variance. For $0 \leq |\beta/\sigma| < 1$ the stochastic trend of the random walk with drift model dominates the deterministic trend in a certain sense; for $|\beta/\sigma| > 1$ the deterministic trend dominates. Being especially interested in cases where β is non-negative and given that our approximations are not valid for $\beta = 0$ (for Monte Carlo results on estimator bias in this model when $\beta = 0$ see the $g_0(T)$ function of Figure 2.1 and also MacKinnon and Smith, 1998), we examined cases where $10 \geqq \beta/\sigma \geqq 0.1$. Results for three different sample sizes are given in Tables 1, 2 and 3 respectively. As is to be expected, the bias of $\widehat{\lambda}$ depends strongly on β/σ. For β much larger than σ the bias is very small, even in very small samples. For relatively small values of β the bias is substantial in samples of a limited size and there is a very serious bias problem in small samples when β is much smaller than σ.

The case $\beta < \sigma$ seems to be relevant in practice. Some empirical evidence is provided by Rudebusch (1992, Table 3), where a difference stationary model is fitted by least-squares to 14 time-series. Some care is required in interpreting these results, because they are biased and even inconsistent in case of model misspecification, but also because the random walk with drift model was estimated directly in only a few cases. The more usual case involved augmented equations which implicitly use a transformation to remove serial correlation and, hence, change the constant term. However, one can easily recover an estimate of the original constant. For the 12 cases that have a positive estimate of β/σ, 4 range from 0.19 to 0.26, 7 estimates range from 0.45 to 0.66 and one is 1.18. Empirical evidence is also found in Hylleberg and Mizon (1989, p.227) and Banerjee et al. (1993, p.171) which report β/σ values of 0.25, 0.72, 0.77 and about 1 respectively.

The numerical results for the various approximations to moments derived in this paper are labelled in the tables by the order of their smallest fully included term and also by the formula from which they originate (sometimes by deliberately omitting terms in order to be able to examine the effects of these higher-order terms). Note that non-trimmed approximations may include parts of terms which are of the same or of smaller order as the remainder term.

In the majority of the cases examined the approximations are very good. However, in situations where the bias is very substantial, and these are the cases where β/σ is small (we deliberately included extreme values for β/σ, which may be empirically less relevant, to examine the robustness of the expansions), the quality of the large-T and small-σ approximations is generally poor, or sometimes extremely bad, and in those situations the higher-order approximation is even worse than the approximation established by the leading term only; note for this phenomenon the difference for the result of Corollary 1 when the full $O(T^{-3})$ formula is used or only its $O(T^{-2})$ term. For large β/σ the higher-order approximation is better. We find no systematic quality difference between the trimmed and the untrimmed approximations, and the $O(\sigma^2)$ approximation is not found to be systematically better than the $O(T^{-2})$ approximation. Note that the small-drift approximation (only calculated for $\beta^* \leq 0.5$) does well where the other approximations break down. Especially when T and the drift are both small the $O(\beta^{*2})$ approximation is remarkably accurate. The variance of $\widehat{\lambda}$, and even more so its MSE, increases when β/σ decreases. We find here again that trimming has little effect and that the large-T and small-σ approximations are very bad for very small β/σ values, especially for small T. However, it would be relatively straightforward to develop an adequate small-drift asymptotic approximation for these second moments. Note that the untrimmed approximations for the MSE of $\widehat{\lambda}$ given in Theorems 2 and 4 respectively give slightly different results. This is because they are obtained in different ways and hence the retained terms may include different bits and pieces that are of the same order as the remainder term. The bias in the estimator of the intercept increases when the intercept decreases, hence its relative impact is very substantial for small β/σ and then it does not change much with T (for $20 \leq T \leq 80$). We should keep in mind, however, that the distribution of $\widehat{\beta}$ is not independent of y_0, so choosing $y_0 = 0$ in the Monte Carlo does not provide general results in this respect (it can be shown, though, that only the higher moments of $\widehat{\beta}$ are affected by y_0 and not its bias). The $O(T^{-1})$ approximation to the bias given in Theorem 3 is found to be very accurate as long as the relative bias is less than, say, 50%. For $\beta/\sigma > 0.5$ the approximations to the variance and MSE of $\widehat{\beta}$ are very good, even for samples as small as $T = 20$.

Next we examine the unit root model with a trending drift, i.e.

$$\left.\begin{aligned} y_t^* &= \lambda y_{t-1}^* + \tfrac{\beta_1}{\sigma} + \tfrac{\beta_2}{\sigma} t + \varepsilon_t \\ y_0^* &= 0,\ \lambda = 1,\ \tfrac{\beta_1}{\sigma} \neq 0, \tfrac{\beta_2}{\sigma} \neq 0 \\ \varepsilon_t &\sim \text{iid}\, \mathsf{N}(0,1) \end{aligned}\right\} \quad t = 1, ..., T. \tag{5.1}$$

Note that our approximations are not valid for the case where $\beta_1 = \beta_2 = 0$. We could have included the case where the intercept is redundant ($\beta_1 = 0$) and not the linear trend ($\beta_2 \neq 0$), but we didn't, because this does not seem to be a particularly relevant case. We have to exclude the case where the linear trend is the only redundant regressor (i.e. $\beta_1 \neq 0$, $\beta_2 = 0$) because then we have $X = (\iota : \tau)$ with $\tau_t = t$ so that $JX = (J\iota : J\tau) = (\tau - \iota : J\tau)$ and hence $\bar{Z} = [(\tau - \iota)\, \beta_1/\sigma : \tau - \iota : J\tau]$ does not have full column rank. So this is another case for which the large-T and small-σ expansions do not exist.

The special form of the X matrix implies that My_{-1}, given in (2.9), is invariant with respect to β_1, and so it follows directly from expression (2.4) that the distribution of $\widehat{\lambda}$ will not depend on β_1, and neither will its bias and MSE, nor their approximations. Note that the estimators $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are not invariant with respect to either β_1 or β_2 or y_0. We present numerical results for model (5.1) for parametrizations with $\beta_2/\sigma = 0.1$ only; some support for a value in this range (or smaller) may be obtained again from Rudebusch (1992, Table 9). In Tables 4, 5 and 6 we present some results. For very small samples the bias in $\widehat{\lambda}$ is substantial. Its approximation by the leading term approximation given in Theorem 3 works adequately, even for $T = 20$. Including the $O(T^{-4})$ term, which can be obtained readily from Theorem 1, is found to be counterproductive in a very small sample. Note that the approximation for the MSE of $\widehat{\lambda}$ given in Theorem 4 works well. As always the quality of the approximation of the variance suffers when the bias approximation is poor. Note that especially in small samples the relative biases of $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are very substantial, and that these biases are opposite in sign. Their approximations, even for huge biases, are remarkably good, and also the second moments can be approximated extremely well.

Finally we look into the model

$$\left.\begin{aligned} y_t &= \lambda y_{t-1} + \beta_1 + \beta_2 x_t + u_t \\ x_t &= \gamma_1 + \gamma_2 x_{t-1} + \xi_t \\ \xi_t &\sim \text{iid}\,\mathsf{N}(0, 1-\gamma_2^2) \end{aligned}\right\} \quad t = 1, ..., T, \tag{5.2}$$

where the x_t series is stationary with unit variance and independent of the disturbances. Because $\lambda = 1$, a parsimonious specification would be $\Delta y_t = \beta_1 + \beta_2 x_t + u_t$. We took $\beta_1 = 0$, $\gamma_1 = 0$, $y_0 = 0$ and $\sigma^2 = 1$, which have no consequences for $\widehat{\lambda}$, whereas we chose $\gamma_2 = \{0.5, 0.9\}$, which produces a less and a more smooth x_t series (which were kept constant over the Monte Carlo replications). We varied the signal-noise ratio SN in the parsimonious specification and considered $SN = \{1, 4, 19\}$. Since $SN = \beta_2^2 \operatorname{Var}(x_t)/\sigma^2 = \beta_2^2$ this implies $\beta_2 = \sqrt{SN} = \{1, 2, 4.36\}$, whereas the corresponding population multiple correlation coeffcient R^2 of the model is $\frac{SN}{SN+1} = \{0.5, 0.8, 0.95\}$. Results are given in Tables 7, 8 and 9 (based on 10^6 replications; left panel is for $\gamma_2 = 0.5$, right panel for $\gamma_2 = 0.9$). For the non-zero coefficients λ and β_2 the biases are larger for the model with the smooth regressor, and as a rule the biases decrease with the signal to noise ratio. We chose the design parameters such that the results expose where about the approximations break down. Note that the $O(T^{-4})$ approximation to the bias of $\widehat{\lambda}$ is only better than the $O(T^{-3})$ version when the bias is not too large. The approximations to the bias in the β coefficients all overstate the actual bias, except when this bias is very small. All approximations to second moments are found to be quite satisfactory, except when both the signal to noise ratio and the sample size are very small.

6. Conclusions

The above derivations and calculations shed light on the factors which are important in determining the bias, the variance and MSE in the unit root dynamic regression model. Earlier, in KP (1993, 1998b), we developed bias approximation formulae for the stable model and found that small-σ versions are inferior to large-T versions and that these deteriorate close to the unit root case, especially when higher-order terms are taken into account. In the present study we develop special approximations for the unit root case and establish that the large-T and small-σ versions may both work very well, apart from cases where the regressors are, or are close to being, redundant, but then alternative asymptotic parameter sequences can be exploited. From the numerical experiments we find that for the random walk with drift model the bias is substantial when the sample size is small and the drift is smaller than the standard deviation of the random shock. This may often be a realistic case, and then bias correction may be worth pursuing. The large-T or small-σ approximations obtained in this study can be used for that purpose, but our numerical experiments show that they do not work well for very small relative values of the drift term. For that situation we present an alternative approximation based on what we call a small-drift asymptotic expansion and this proves to be very accurate in the special cases it is meant for. When further exogenous regressors are added to the model the bias may get worse for practically relevant parameter values. We give special attention to the model with an intercept and a linear deterministic trend which is so often applied in practice, viz. when the Dickey-Fuller test is applied. We illustrate that the bias in this model is heavily dependent on the sample size, and that it may be approximated quite accurately. When the model contains another stationary exogenous explanatory variable (instead of a linear trend), we show that the magnitude of the bias depends not only strongly on the sample size, but also on the smoothness of the regressor and on a signal to noise ratio. We expose the (extreme) cases where the bias approximation formulae break down and illustrate that in this respect leading terms in asymptotic expansions are more robust than higher-order terms. The latter are only useful when the error of the leading term does not exceed (roughly) 20%. For all models examined we show that second moments of the least-squares estimators can be approximated quite accurately by our higher-order asymptotic expressions. We conclude that over a substantial and practically very relevant part of the parameter space of autoregressive models the tools developed in this paper can be exploited to improve inference methods when samples are small or only moderately large and a unit root is present.

References

Abadir, K.M.,1995. Unbiased estimation as a solution to testing for random walks. *Economics Letters* 47, 263-268.

Banerjee, A., J. Dolado, J.W. Galbraith,.Hendry, D.F, 1993. *Co-integration, error-correction, and the econometric analysis of non-stationary data.* Advanced texts in Econometrics. Oxford University Press.

Evans. G.R.A., Savin, N.E., 1984. Testing for unit roots 2, *Econometrica* 52, 1241-1269.

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

Hylleberg, S., Mizon, G.E., 1989. A note on the distribution of the least squares estimator of a random walk with drift. *Economics Letters* 29, 225-230.

Hoque, A.S., Peters, T.A., 1986. Finite sample analysis of the ARMAX models. *Sankhya: The Indian Journal of Statistics* 48, Series B, part 2, 266-283.

Kadane, J.B., 1971. Comparison of k-class estimators when the disturbances are small. *Econometrica* 39, 723-737.

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

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

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

Kiviet, J.F., Phillips, G.D.A., 1996. The bias of the ordinary least squares estimator in simultaneous equation models. *Economics Letters* 53, 161-167.

Kiviet, J.F., Phillips, G.D.A., 1998a. 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., 1998b. Higher-order asymptotic expansions of the least-squares estimation bias in first-order dynamic regression models; paper presented at ESWM-95, Tokyo, August 1995; Tinbergen Institute Discussion Paper 96-167/7, revised October 1998.

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.

Nabeya, S., Sørensen, B., 1994. Asymptotic distributions of the least-squares estimators and test statistics in the near unit root model with non-zero initial value and local drift and trend. *Econometric Theory* 10, 937-966.

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.

Paolella, M.S., 2003. Computing moments of ratios of quadratic forms in normal variables. *Computational Statistics & Data Analysis* 42, 313-331.

Peters, T.A., 1987. The exact moments of OLS in dynamic regression models with non-normal errors. *Journal of Econometrics* 17, 279-305.

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

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

West, K.D., 1988. Asymptotic normality when regressors have a unit root. *Econometrica* 56, 1397-1418.

A. Basic results on ι and J

For the $T\times 1$ vector ι and the $T\times T$ matrix J, introduced in (2.7), we have the following results for $t=1,...,T$: $(J\iota)_t = t-1$, $(JJ\iota)_t = (t-1)(t-2)/2$, $(J'J\iota)_t = [T(T-1) - t(t-1)]/2$, $(J'\iota)_t = T-t$ and $(JJ'\iota)_t = (t-1)(T-t/2)$. Making use of the well-known summation results $\sum_{t=1}^{T} t = (T+1)T/2$, $\sum_{t=1}^{T} t^2 = (T+1)(2T+1)T/6$, $\sum_{t=1}^{T} t^3 = [(T+1)T]^2/4$ and $\sum_{t=1}^{T} t^4 = (T+1)(2T+1)(3T^2+3T-1)T/30$ we also find:

$$\iota' J\iota = \sum_{t=1}^{T}(t-1) = \frac{T}{2}(T-1) = \frac{1}{2}T^2 - \frac{1}{2}T, \tag{A.1}$$

$$\iota' J' J\iota = \sum_{t=1}^{T}(t-1)^2 = \frac{1}{3}T^3 - \frac{1}{2}T^2 + \frac{1}{6}T, \tag{A.2}$$

$$\iota' J J'\iota = \sum_{t=1}^{T}(T-t)^2 = \sum_{t=1}^{T}(t-1)^2 = \frac{1}{3}T^3 - \frac{1}{2}T^2 + \frac{1}{6}T, \tag{A.3}$$

$$\iota' J J\iota = \sum_{t=1}^{T}(T-t)(t-1) = \frac{1}{6}T^3 - \frac{1}{2}T^2 + \frac{1}{3}T, \tag{A.4}$$

$$\iota' J J' J\iota = \sum_{t=1}^{T}(t-1)^2 (T-t/2) = \frac{5}{24}T^4 - \frac{5}{12}T^3 + \frac{7}{24}T^2 - \frac{1}{12}T, \tag{A.5}$$

$$\iota' J' J' J\iota = \frac{1}{2}\sum_{t=1}^{T}(t-1)^2 (t-2) = \frac{1}{8}T^4 - \frac{5}{12}T^3 + \frac{3}{8}T^2 - \frac{1}{12}T - 1, \tag{A.6}$$

$$\iota' J J J\iota = \frac{1}{2}\sum_{t=1}^{T}(T-t)(t-1)(t-2) = \frac{1}{24}T^4 - \frac{1}{4}T^3 + \frac{11}{24}T^2 - \frac{1}{4}T, \tag{A.7}$$

$$\iota' J' J J' J\iota = \sum_{t=1}^{T}\left[\frac{T}{2}(T-1) - \frac{t}{2}(t-1)\right]^2 = \frac{2}{15}T^5 + O(T^4), \tag{A.8}$$

$$\iota' J' J' J' J\iota = \frac{1}{2}\sum_{t=1}^{T}\left[\frac{T}{2}(T-1) - \frac{t}{2}(t-1)\right](t-1)(t-2) = \frac{1}{30}T^5 + O(T^4). \tag{A.9}$$

The simple structure of J also leads to the results:

$$\mathrm{tr}(J^i) = 0, \text{ for } i = 1, 2, ... \tag{A.10}$$

$$\mathrm{tr}(J'J) = \frac{1}{2}T^2 - \frac{1}{2}T, \tag{A.11}$$

$$\mathrm{tr}\left(J'JJ'J\right) = \sum_{t=1}^{T}\left[(t-1)(T-t)^2 + \sum_{i=0}^{T-t} i^2\right] = \frac{1}{6}T^4 + O(T^3). \tag{A.12}$$

B. Basic results on orders of magnitude

Here we collect results that support the statements made in Sections 3 and 4 on orders of magnitude of relevant expressions. From $\mathsf{E}(X'u) = 0$ and $\mathrm{Var}(X'u) = \sigma^2(X'X) = O(T)$ follows $X'u = O_p(T^{1/2})$. Since $JX = O(T)$ we have $X'J'JX = O(T^3)$ and $X'J'u = O_p(T^{3/2})$, also giving $X'J'MJX = O(T^3)$ and $X'J'Mu = O_p(T^{3/2})$. Along similar lines $X'J'JJ'JX = O(T^5)$ yields $X'J'Ju = O_p(T^{5/2})$, from which $X'J'MJu = O_p(T^{5/2})$ follows. From $\mathsf{E}(u'Ju) = 0$ and $\mathrm{Var}(u'Ju) = \sigma^4\,\mathrm{tr}(J'J) = O(T^2)$ we find $u'Ju = O_p(T)$, which yields $u'J'Mu = O_p(T)$. Moreover, because $\mathsf{E}(u'J'Ju) = \sigma^2\,\mathrm{tr}(J'J) = O(T^2)$ and $\mathrm{Var}(u'J'Ju) = 2\sigma^4\,\mathrm{tr}(J'JJ'J) = O(T^4)$, we find $u'J'Ju = O_p(T^2)$, from which $u'J'MJu = O_p(T^2)$ follows.

C. Proof of Theorem 1

Expanding the inverse factor of (2.11) further than in (2.12) we obtain

$$\begin{aligned}
&\left(1 + 2\frac{\beta'X'J'MJu}{\beta'X'J'MJX\beta} + \frac{u'J'MJu}{\beta'X'J'MJX\beta}\right)^{-1} \\
= \;& 1 - \frac{2\beta'X'J'MJu}{\beta'X'J'MJX\beta} - \frac{u'J'MJu}{\beta'X'J'MJX\beta} + 4\left(\frac{\beta'X'J'MJu}{\beta'X'J'MJX\beta}\right)^2 \\
& + 4\frac{\left(\beta'X'J'MJu\right)\left(u'J'MJu\right)}{\left(\beta'X'J'MJX\beta\right)^2} - 8\left(\frac{\beta'X'J'MJu}{\beta'X'J'MJX\beta}\right)^3 + o_p(T^{-3/2}).
\end{aligned}$$

Substitution in (2.11) yields

$$\begin{aligned}
\widehat{\lambda} - 1 = \;& \frac{\beta'X'J'Mu}{\beta'X'J'MJX\beta} + \frac{u'J'Mu}{\beta'X'J'MJX\beta} - 2\frac{\left(\beta'X'J'Mu\right)\left(\beta'X'J'MJu\right)}{\left(\beta'X'J'MJX\beta\right)^2} \\
& - 2\frac{\left(\beta'X'J'MJu\right)\left(u'J'Mu\right)}{\left(\beta'X'J'MJX\beta\right)^2} - \frac{\left(\beta'X'J'Mu\right)\left(u'J'MJu\right)}{\left(\beta'X'J'MJX\beta\right)^2} \\
& - \frac{\left(u'J'MJu\right)\left(u'J'Mu\right)}{\left(\beta'X'J'MJX\beta\right)^2} + 4\frac{\left(\beta'X'J'MJu\right)^2\left(\beta'X'J'Mu\right)}{\left(\beta'X'J'MJX\beta\right)^3} \\
& + 4\frac{\left(\beta'X'J'MJu\right)^2\left(u'J'Mu\right)}{\left(\beta'X'J'MJX\beta\right)^3} + 4\frac{\left(\beta'X'J'MJu\right)\left(\beta'X'J'Mu\right)\left(u'J'MJu\right)}{\left(\beta'X'J'MJX\beta\right)^3} \\
& - 8\frac{\left(\beta'X'J'MJu\right)^3\left(\beta'X'J'Mu\right)}{\left(\beta'X'J'MJX\beta\right)^4} + o_p(T^{-3}).
\end{aligned}$$

To approximate the bias we take the expectation of these terms. Terms involving an odd number of zero mean normal random variables can be ignored. Occasionally we can simplify the expressions by using the fact that in traces or in scalars the expression is sometimes unchanged when J is replaced by J', and hence J can be replaced by $[J+J'] = \frac{1}{2}[\iota\iota' - I_T]$. Because $M\iota = 0$, this may lead to some simplification. Using

$$\mathsf{E}\left(u'J'Mu\right) = \sigma^2 \operatorname{tr}(MJ) = \frac{1}{2}\sigma^2 \operatorname{tr}[M(\iota\iota' - I_T)] = -\frac{1}{2}\sigma^2(T-k) = O(T),$$

$$\begin{aligned}\mathsf{E}\left(\beta'X'J'Muu'J'MJX\beta\right) &= \sigma^2\beta'X'J'MJMJX\beta \\ &= -\frac{1}{2}\sigma^2\beta'X'J'MJX\beta = O(T^3),\end{aligned}$$

$$\begin{aligned}\mathsf{E}\left(u'J'MJu\right)\left(u'J'Mu\right) &= \sigma^4\left[\operatorname{tr}(MJ)\operatorname{tr}\left(J'MJ\right) + 2\operatorname{tr}\left(J'MJJ'M\right)\right] \\ &= \sigma^4\left[\operatorname{tr}(MJ)\operatorname{tr}\left(J'MJ\right) - \operatorname{tr}\left(J'MJ\right)\right] \\ &= \sigma^4\operatorname{tr}(MJ)\operatorname{tr}\left(J'MJ\right) + o(T^3),\end{aligned}$$

$$\begin{aligned}&\mathsf{E}\left(\beta'X'J'MJu\right)^2\left(u'J'Mu\right) \\ =\ &\mathsf{E}\left(u'J'MJX\beta\beta'X'J'MJu\right)\left(u'MJu\right) \\ =\ &\sigma^4\left[\operatorname{tr}\left(J'MJX\beta\beta'X'J'MJ\right)\operatorname{tr}(MJ) + 2\operatorname{tr}\left(J'MJX\beta\beta'X'J'MJMJ\right)\right] \\ =\ &\sigma^4\left[\operatorname{tr}(MJ)\,\beta'X'J'MJJ'MJX\beta + 2\beta'X'J'MJMJJ'MJX\beta\right] = O(T^6),\end{aligned}$$

$$\begin{aligned}&\mathsf{E}\left(\beta'X'J'MJu\right)\left(\beta'X'J'Mu\right)\left(u'J'MJu\right) \\ =\ &\mathsf{E}\left(u'J'MJu\right)\left(u'J'MJX\beta\beta'X'J'Mu\right) \\ =\ &\sigma^4\left[\operatorname{tr}\left(J'MJ\right)\beta'X'J'MJ'MJX\beta + 2\beta'X'J'MJ'MJJ'MJX\beta\right] \\ =\ &\sigma^4[-\frac{1}{2}\operatorname{tr}\left(J'MJ\right)\beta'X'J'MJX\beta + 2\beta'X'J'MJ'MJJ'MJX\beta] \\ =\ &2\sigma^4\beta'X'J'MJ'MJJ'MJX\beta + o(T^6),\end{aligned}$$

$$\begin{aligned}&\mathsf{E}\left(\beta'X'J'MJu\right)^3\left(\beta'X'J'Mu\right) \\ =\ &\mathsf{E}\left(u'J'MJX\beta\beta'X'J'MJu\right)\left(u'J'MJX\beta\beta'X'J'Mu\right) \\ =\ &3\sigma^4\left(\beta'X'J'MJJ'MJX\beta\right)\left(\beta'X'J'MJMJX\beta\right) \\ =\ &-\frac{3}{2}\sigma^4\left(\beta'X'J'MJJ'MJX\beta\right)\left(\beta'X'J'MJX\beta\right) = O(T^8),\end{aligned}$$

and removing terms that are of such magnitude that they can be neglected in an $O(T^{-3})$ approximation, yields

$$\begin{aligned}\mathsf{E}(\widehat{\lambda} - 1) = &\frac{\sigma^2\operatorname{tr}(MJ)}{\beta'X'J'MJX\beta} + \frac{\sigma^2}{\beta'X'J'MJX\beta} - \frac{\sigma^4\operatorname{tr}(MJ)\operatorname{tr}(J'MJ)}{\left(\beta'X'J'MJX\beta\right)^2} \\ &+\frac{4\sigma^4\operatorname{tr}(MJ)\,\beta'X'J'MJJ'MJX\beta}{\left(\beta'X'J'MJX\beta\right)^3} + \frac{8\sigma^4\beta'X'J'MJMJJ'MJX\beta}{\left(\beta'X'J'MJX\beta\right)^3} \\ &+\frac{8\sigma^4\beta'X'J'MJ'MJJ'MJX\beta}{\left(\beta'X'J'MJX\beta\right)^3} + o_p(T^{-3}),\end{aligned}$$

where the last two terms can be combined, such that the numerator involves

$$\beta'X'J'(MJ+J')MJJ'MJX\beta = -\frac{1}{2}\beta'X'J'MJJ'MJX\beta = O(T^5),$$

which shows that these two terms can be neglected in the result of the theorem.

D. Proof of Corollary 1

Putting $X = \iota$, $M = I_T - \frac{1}{T}\iota\iota'$ and β scalar in the various terms of Theorem 1 and using results from Appendix A leads to

$$\mathrm{tr}\,(MJ) = -\frac{1}{T}\iota' J' \iota = -\frac{1}{2}\,(T-1)\,,$$

$$X'J'MJX = \iota' J' J \iota - \frac{1}{T}\left(\iota' J \iota\right)^2 = \frac{1}{12} T\,(T-1)\,(T+1)\,,$$

$$\mathrm{tr}\left(J'MJ\right) = \mathrm{tr}\left(J'J\right) - \frac{1}{T}\iota' J J' \iota = \frac{1}{6}(T^2-1),$$

$$X'J'MJJ'MJX = \iota' J' J J' J \iota - \frac{2}{T}\iota' J \iota\iota' J J' J \iota + \frac{1}{T^2}\iota' J' \iota\iota' J J' \iota\iota' J \iota = \frac{1}{120}T^5 + O(T^4),$$

which after substitution lead to the result of the Corollary.

E. Proof of Theorem 2

To find the approximation to the MSE we commence from (3.4). Using an expansion for the inverse factor of the form $(1+x)^{-2} = 1 - 2x + 3x^2 - 4x^3 + ...$ the MSE may be approximated by

$$\begin{aligned}
\mathsf{E}(\widehat{\lambda}-1)^2 \quad = \quad & \frac{\mathsf{E}\left(\beta' X' J' M u\right)^2}{\left(\beta' X' J' M J X \beta\right)^2} + \frac{\mathsf{E}\,(u J' M u)^2}{\left(\beta' X' J' M J X \beta\right)^2} - 2\frac{\mathsf{E}\left(\beta' X' J' M u\right)^2 (u' J' M J u)}{\left(\beta' X' J' M J X \beta\right)^3} \\
& -8\frac{\mathsf{E}\left(\beta' X' J' M u\right)\left(u' J' M u\right)\left(\beta' X' J' M J u\right)}{\left(\beta' X' J' M J X \beta\right)^3} \\
& +12\frac{\mathsf{E}\left(\beta' X' J' M u\right)^2\left(\beta' X' J' M J u\right)^2}{\left(\beta' X' J' M J X \beta\right)^4} + o(T^{-4}).
\end{aligned}$$

Here we have removed terms from the original expansion which involve a product of an odd number of normal random variables with mean zero, together with terms which are $o_p(T^{-4})$. Below we evaluate the expectations in the various numerators and exploit the same simplification as in Appendix C.

$$\mathsf{E}\left(\beta' X' J' M u\right)^2 = \sigma^2 \beta' X' J' M J X \beta,$$

$$\mathsf{E}\left(u J' M u\right)^2 = \sigma^4 \{[\mathrm{tr}\,(MJ)]^2 + \mathrm{tr}\,(JMJM) + \mathrm{tr}\left(J'MJ\right)\},$$

$$\begin{aligned}
& \mathsf{E}\left(\beta' X' J' M u\right)^2\left(u' J' M J u\right) \\
= \;& \mathsf{E}\left(u' M J X \beta\beta' X' J' M u\right)\left(u' J' M J u\right) \\
= \;& \sigma^4\left\{\mathrm{tr}\left(J' M J\right)\beta' X' J' M J X \beta + 2\beta' X' J' M J' M J M J X \beta\right\},
\end{aligned}$$

$$\begin{aligned}
& \mathsf{E}\left(\beta' X' J' M u\right)\left(u' J' M u\right)\left(\beta' X' J' M J u\right) \\
= \;& \mathsf{E}\left(u' J' M J X \beta\beta' X' J' M u\right)\left(u' J' M u\right) \\
= \;& \sigma^4\left\{\mathrm{tr}\,(MJ)\,\beta' X' J' M J' M J X \beta + \beta' X' J' M\left[J' M J' + J J'\right] M J X \beta\right\} \\
= \;& \sigma^4\{-\frac{1}{2}\,\mathrm{tr}\,(MJ)\,\beta' X' J' M J X \beta + \beta' X' J' M\left[J M J + J J'\right] M J X \beta\},
\end{aligned}$$

$$\begin{aligned}
&\mathsf{E}\left(\beta' X' J' M u\right)^2 \left(\beta' X' J' M J u\right)^2 \\
= &\mathsf{E}\left(u' M J X \beta \beta' X' J' M u\right)\left(u' J' M J X \beta \beta' X' J' M J u\right) \\
= &\sigma^4 [\beta' X' J' M J X \beta \times \beta' X' J' M J J' M J X \beta + 2\left(\beta' X' J' M J M J X \beta\right)^2] \\
= &\sigma^4 [\beta' X' J' M J X \beta \times \beta' X' J' M J J' M J X \beta + \frac{1}{2}\left(\beta' X' J' M J X \beta\right)^2].
\end{aligned}$$

Substitution yields

$$\begin{aligned}
\mathsf{E}(\widehat{\lambda}-1)^2 = &\frac{\sigma^2}{\beta' X' J' M J X \beta} + \sigma^4 \frac{[\mathrm{tr}\,(MJ)]^2 + \mathrm{tr}\,(JMJM) + \mathrm{tr}\,(J'MJ)}{\left(\beta' X' J' M J X \beta\right)^2} \\
&-2\sigma^4 \frac{\mathrm{tr}\,(J'MJ)\,\beta' X' J' M J X \beta + 2\beta' X' J' M J' M J M J X \beta}{\left(\beta' X' J' M J X \beta\right)^3} \\
&-8\sigma^4 \frac{\beta' X' J' M\left[JMJ + JJ'\right] M J X \beta - \frac{1}{2}\,\mathrm{tr}\,(MJ)\,\beta' X' J' M J X \beta}{\left(\beta' X' J' M J X \beta\right)^3} \\
&+12\sigma^4 \frac{\beta' X' J' M J X \beta \times \beta' X' J' M J J' M J X \beta + \frac{1}{2}\left(\beta' X' J' M J X \beta\right)^2}{\left(\beta' X' J' M J X \beta\right)^4} + o(T^{-4})
\end{aligned}$$

and removing terms of small order we obtain

$$\begin{aligned}
\mathsf{E}(\widehat{\lambda}-1)^2 = &\frac{\sigma^2}{\beta' X' J' M J X \beta} + \sigma^4 \frac{[\mathrm{tr}\,(MJ)]^2 + \mathrm{tr}\,(JMJM) - \mathrm{tr}\,(J'MJ)}{\left(\beta' X' J' M J X \beta\right)^2} \\
&-4\sigma^4 \frac{\beta' X' J' M J' M J M J X \beta}{\left(\beta' X' J' M J X \beta\right)^3} - 8\sigma^4 \frac{\beta' X' J' M J M J M J X \beta}{\left(\beta' X' J' M J X \beta\right)^3} \\
&+4\sigma^4 \frac{\beta' X' J' M J J' M J X \beta}{\left(\beta' X' J' M J X \beta\right)^3} + o(T^{-4})
\end{aligned}$$

which then, after minor further simplification, gives the result of the theorem.

F. Proof of Corollary 2

Following up on the proof in Appendix D, we have to evaluate a few extra expressions after putting $X = \iota$, $M = I_T - \frac{1}{T}\iota\iota'$ and β scalar. We find

$$\mathrm{tr}\,(JMJM) = \mathrm{tr}\,(JJ) - \frac{2}{T}\iota' JJ\iota + \frac{1}{T^2}\left(\iota' J \iota\right)^2 = -\frac{1}{12}T^2 + \frac{1}{2}T - \frac{5}{12},$$

and

$$\begin{aligned}
X' J' M J M J M J X = &\iota' J' J J J \iota - \frac{1}{T}\iota' J' \iota\iota' J J J \iota - \frac{1}{T}\iota' J' J \iota\iota' J J \iota - \frac{1}{T}\iota' J' J J \iota\iota' J \iota \\
&+\frac{2}{T^2}\left(\iota' J' \iota\right)^2 \iota' J J \iota + \frac{1}{T^2}\iota' J' J \iota\left(\iota' J' \iota\right)^2 - \frac{1}{T^3}\left(\iota' J' \iota\right)^4 = -\frac{1}{720}T^5 + O(T^4).
\end{aligned}$$

Substitution yields

$$\begin{aligned}
\mathsf{E}(\widehat{\lambda}-1)^2 = &\left(\frac{\sigma}{\beta}\right)^2 \frac{12}{T\left(T^2-1\right)} + \left(\frac{\sigma}{\beta}\right)^4 \frac{12^2}{T^4}\left(\frac{1}{4} - \frac{1}{12} - \frac{1}{6}\right) \\
&+4\left(\frac{\sigma}{\beta}\right)^4 \frac{12^3}{T^4}\left(\frac{1}{120} + \frac{1}{720}\right) + o(T^{-4})
\end{aligned}$$

which leads to the result in the corollary.

G. Proof of Theorem 3

The required bias is obtained from the expansion (4.11). Since terms with an odd number of stochastic factors have zero expectation, we have to evaluate

$$\mathsf{E}\left[D^{-1}\left(\widehat{\alpha}-\alpha\right)\right]=\mathsf{E}[R\tilde{W}'u]-R\mathsf{E}[(\bar{W}'\tilde{W}+\tilde{W}'\bar{W})R\bar{W}'u]+o(T^{-1}).$$

We find

$$\mathsf{E}(R\tilde{W}'u)=R\mathsf{E}(De_1u'J'u)=RDe_1\mathsf{E}(u'J'u)=0,$$

$$\mathsf{E}(\bar{W}'\tilde{W}R\bar{W}'u)=\mathsf{E}(\bar{W}'Jue_1'DR\bar{W}'u)=\sigma^2\bar{W}'J\bar{W}RDe_1,$$

$$\mathsf{E}(\tilde{W}'\bar{W}R\bar{W}'u)=\mathsf{E}(De_1u'J'\bar{W}R\bar{W}'u)=\sigma^2\operatorname{tr}(R\bar{W}'J\bar{W})De_1,$$

hence, using $R=D^{-1}(\bar{Z}'\bar{Z})^{-1}D^{-1}$ and $\bar{W}=\bar{Z}D$, we find

$$\begin{aligned}&\mathsf{E}\left[D^{-1}\left(\widehat{\alpha}-\alpha\right)\right]=\\&-\sigma^2D^{-1}[(\bar{Z}'\bar{Z})^{-1}\bar{Z}'J\bar{Z}+\operatorname{tr}\{(\bar{Z}'\bar{Z})^{-1}\bar{Z}'J\bar{Z}\}I_{k+1}](\bar{Z}'\bar{Z})^{-1}e_1+o(T^{-1}).\end{aligned}\tag{G.1}$$

Some further simplification is possible. Note that

$$\operatorname{tr}\{(\bar{Z}'\bar{Z})^{-1}\bar{Z}'J\bar{Z}\}=\frac{1}{2}\operatorname{tr}\{(\bar{Z}'\bar{Z})^{-1}\bar{Z}'\left(J+J'\right)\bar{Z}\}=\frac{1}{2}\operatorname{tr}\{\bar{Z}(\bar{Z}'\bar{Z})^{-1}\bar{Z}'\left(\iota\iota'-I_T\right)\}.$$

Since the regression contains a constant, we have $\bar{Z}(\bar{Z}'\bar{Z})^{-1}\bar{Z}'\iota=\iota$, and hence

$$\operatorname{tr}\{(\bar{Z}'\bar{Z})^{-1}\bar{Z}'J\bar{Z}\}=\frac{1}{2}[\operatorname{tr}\left(\iota\iota'\right)-\operatorname{tr}\left(I_{k+1}\right)]=\frac{1}{2}\left(T-k-1\right).\tag{G.2}$$

Finally consider

$$e_1'(\bar{Z}'\bar{Z})^{-1}\bar{Z}'J\bar{Z}(\bar{Z}'\bar{Z})^{-1}e_1=\frac{1}{2}e_1'(\bar{Z}'\bar{Z})^{-1}\bar{Z}'[\iota\iota'-I]\bar{Z}(\bar{Z}'\bar{Z})^{-1}e_1.$$

Because ι is the second column of $\bar{Z}$ we have $(\bar{Z}'\bar{Z})^{-1}\bar{Z}'\iota=e_2$, and hence, using $e_1'e_2=0$,

$$e_1'(\bar{Z}'\bar{Z})^{-1}\bar{Z}'J\bar{Z}(\bar{Z}'\bar{Z})^{-1}e_1=-\frac{1}{2}e_1'(\bar{Z}'\bar{Z})^{-1}e_1.\tag{G.3}$$

Premultiplying both sides of (G.1) with $e_1'D$ and making use of (G.2) and (G.3) yields the bias of $\widehat{\lambda}$ as stated in the theorem, which can be shown to be equivalent to (3.2) for the case $\delta_1=-1$. Premultiplying (G.1) by $e_{i+1}'D$ yields the bias of the individual elements $\widehat{\beta}_i$, $i=1,...,k$.

H. Proof of Theorem 4

Upon removing the terms which are a product of an odd number of normal random variables with zero mean, we may write

$$\begin{aligned}&\mathsf{E}[D^{-1}\left(\widehat{\alpha}-\alpha\right)\left(\widehat{\alpha}-\alpha\right)'D^{-1}]=\\&R\mathsf{E}[\bar{W}'uu'\bar{W}+\tilde{W}'uu'\tilde{W}+P\bar{W}'uu'\bar{W}P'\\&\quad-P\bar{W}'uu'\tilde{W}-\bar{W}'uu'\tilde{W}P'-P\tilde{W}'uu'\bar{W}-\tilde{W}'uu'\bar{W}P'\\&\quad+PP\bar{W}'uu'\bar{W}+\bar{W}'uu'\bar{W}P'P'-S\bar{W}'uu'\bar{W}-\bar{W}'uu'\bar{W}S']R+o(T^{-2}).\end{aligned}\tag{H.1}$$

The required approximation to the MSE of $\widehat{\alpha}$ is obtained by evaluating this expectation and pre- and post-multiplying the result by D. We make use of the substitutions $\bar{W}'\bar{W} = R^{-1}$, $\tilde{W} = Jue_1'D$, $P = \bar{W}'Jue_1'DR + De_1u'J'\bar{W}R$ and $S = De_1u'J'Jue_1'DR$ and, often using the result $\mathsf{E}(uu'Buu') = \sigma^4[\mathrm{tr}(B)I + B + B']$ for general B matrices, we find for the successive terms:

$$\mathsf{E}(\bar{W}'uu'\bar{W}) = \sigma^2\bar{W}'\bar{W} = \sigma^2 R^{-1}$$

$$\mathsf{E}(\tilde{W}'uu'\tilde{W}) = \mathsf{E}(De_1u'J'uu'Jue_1'D) = \sigma^4\,\mathrm{tr}\left(J'J\right) De_1e_1'D$$

$$\begin{aligned}
&\mathsf{E}(P\bar{W}'uu'\bar{W}P')\\
=\;& \mathsf{E}(\bar{W}'Jue_1'DR\bar{W}'uu'\bar{W}RDe_1u'J'\bar{W} + De_1u'J'\bar{W}R\bar{W}'uu'\bar{W}RDe_1u'J'\bar{W}\\
&+\bar{W}'Jue_1'DR\bar{W}'uu'\bar{W}R\bar{W}'Jue_1'D + De_1u'J'\bar{W}R\bar{W}'uu'\bar{W}R\bar{W}'Jue_1'D)\\
=\;& \mathsf{E}(\bar{W}'Juu'\bar{W}RDe_1e_1'DR\bar{W}'uu'J'\bar{W} + De_1e_1'DR\bar{W}'uu'J'\bar{W}R\bar{W}'uu'J'\bar{W}\\
&+\bar{W}'Juu'\bar{W}R\bar{W}'Juu'\bar{W}RDe_1e_1'D + u'J'\bar{W}R\bar{W}'uu'\bar{W}R\bar{W}'JuDe_1e_1'D)\\
=\;& \sigma^4(e_1'DRDe_1)\bar{W}'JJ'\bar{W} + 2\sigma^4\bar{W}'J\bar{W}RDe_1e_1'DR\bar{W}'J'\bar{W}\\
&+\sigma^4\,\mathrm{tr}(R\bar{W}'J\bar{W})De_1e_1'DR\bar{W}'J'\bar{W} + \sigma^4De_1e_1'DR\bar{W}'J'\bar{W}R\bar{W}'J'\bar{W}\\
&+\sigma^4De_1e_1'DR\bar{W}'JJ'\bar{W} + \sigma^4\,\mathrm{tr}(R\bar{W}'J\bar{W})\bar{W}'J\bar{W}RDe_1e_1'D\\
&+\sigma^4\bar{W}'J\bar{W}R\bar{W}'J\bar{W}RDe_1e_1'D + \sigma^4\bar{W}'JJ'\bar{W}RDe_1e_1'D\\
&+\sigma^4\,\mathrm{tr}(R\bar{W}'J\bar{W})\,\mathrm{tr}(R\bar{W}'J\bar{W})De_1e_1'D\\
&+\sigma^4\,\mathrm{tr}(R\bar{W}'JJ'\bar{W})De_1e_1'D + \sigma^4\,\mathrm{tr}(R\bar{W}'J\bar{W}R\bar{W}'J\bar{W})De_1e_1'D
\end{aligned}$$

$$\begin{aligned}
\mathsf{E}(P\bar{W}'uu'\tilde{W}) &= \mathsf{E}(\bar{W}'Jue_1'DR\bar{W}'uu'Jue_1'D + De_1u'J'\bar{W}R\bar{W}'uu'Jue_1'D)\\
&= \mathsf{E}(\bar{W}'Juu'Juu'\bar{W}RDe_1e_1'D + u'J'\bar{W}R\bar{W}'uu'JuDe_1e_1'D)\\
&= \sigma^4\bar{W}'JJ\bar{W}RDe_1e_1'D + \sigma^4\bar{W}'JJ'\bar{W}RDe_1e_1'D\\
&\quad+\sigma^4\,\mathrm{tr}(R\bar{W}'JJ'\bar{W})De_1e_1'D + \sigma^4\,\mathrm{tr}(R\bar{W}'JJ\bar{W})De_1e_1'D\\
\mathsf{E}(\tilde{W}'uu'\bar{W}P') &= [\mathsf{E}(P\bar{W}'uu'\tilde{W})]'
\end{aligned}$$

$$\begin{aligned}
\mathsf{E}(P\tilde{W}'uu'\bar{W}) &= \mathsf{E}(\bar{W}'Jue_1'DRDe_1u'J'uu'\bar{W} + De_1u'J'\bar{W}RDe_1u'J'uu'\bar{W})\\
&= \mathsf{E}[(e_1'DRDe_1)\bar{W}'Juu'J'uu'\bar{W} + De_1e_1'DR\bar{W}'Juu'J'uu'\bar{W}]\\
&= \sigma^4(e_1'DRDe_1)\bar{W}'JJ'\bar{W} + \sigma^4(e_1'DRDe_1)\bar{W}'JJ\bar{W}\\
&\quad+\sigma^4De_1e_1'DR\bar{W}'JJ'\bar{W} + \sigma^4De_1e_1'DR\bar{W}'JJ\bar{W}\\
\mathsf{E}(\tilde{W}'uu'\bar{W}P') &= [\mathsf{E}(P\bar{W}'uu'\tilde{W})]'
\end{aligned}$$

$$\begin{aligned}
&\mathsf{E}(PP\bar{W}'uu'\bar{W})\\
=\;& \mathsf{E}(\bar{W}'Jue_1'DR\bar{W}'Jue_1'DR\bar{W}'uu'\bar{W} + De_1u'J'\bar{W}R\bar{W}'Jue_1'DR\bar{W}'uu'\bar{W}\\
&+\bar{W}'Jue_1'DRDe_1u'J'\bar{W}R\bar{W}'uu'\bar{W} + De_1u'J'\bar{W}RDe_1u'J'\bar{W}R\bar{W}'uu'\bar{W})\\
=\;& \mathsf{E}[\bar{W}'Juu'J'\bar{W}RDe_1e_1'DR\bar{W}'uu'\bar{W} + De_1e_1'DR\bar{W}'uu'J'\bar{W}R\bar{W}'Juu'\bar{W}\\
&+(e_1'DRDe_1)\bar{W}'Juu'J'\bar{W}R\bar{W}'uu'\bar{W} + De_1e_1'DR\bar{W}'Juu'J'\bar{W}R\bar{W}'uu'\bar{W}]\\
=\;& \sigma^4(e_1'DR\bar{W}'J\bar{W}RDe_1)\bar{W}'J\bar{W} + \sigma^4\bar{W}'JJ'\bar{W}RDe_1e_1'D\\
&+\sigma^4\bar{W}'J\bar{W}RDe_1e_1'DR\bar{W}'J\bar{W} + \sigma^4\,\mathrm{tr}(R\bar{W}'JJ'\bar{W})De_1e_1'D\\
&+2\sigma^4De_1e_1'DR\bar{W}'J'\bar{W}R\bar{W}'J\bar{W} + \sigma^4\,\mathrm{tr}(R\bar{W}'J\bar{W})(e_1'DRDe_1)\bar{W}'J\bar{W}\\
&+\sigma^4(e_1'DRDe_1)\bar{W}'JJ'\bar{W} + \sigma^4(e_1'DRDe_1)\bar{W}'J\bar{W}R\bar{W}'J\bar{W}\\
&+\sigma^4\,\mathrm{tr}(R\bar{W}'J\bar{W})De_1e_1'DR\bar{W}'J\bar{W} + \sigma^4De_1e_1'DR\bar{W}'JJ'\bar{W}\\
&+\sigma^4De_1e_1'DR\bar{W}'J\bar{W}R\bar{W}'J\bar{W}
\end{aligned}$$

$$\mathsf{E}(\bar{W}'uu'\bar{W}P'P') = [\mathsf{E}(PP\bar{W}'uu'\bar{W})]'$$

$$\begin{aligned}
\mathsf{E}(S\bar{W}'uu'\bar{W}) &= \mathsf{E}(De_1u'J'Jue_1'DR\bar{W}'uu'\bar{W}) = \mathsf{E}(De_1e_1'DR\bar{W}'uu'J'Juu'\bar{W}) \\
&= \sigma^4\operatorname{tr}(J'J)De_1e_1'D + 2\sigma^4 De_1e_1'DR\bar{W}'J'J\bar{W} \\
\mathsf{E}(\bar{W}'uu'\bar{W}S') &= [\mathsf{E}(S\bar{W}'uu'\bar{W})]'
\end{aligned}$$

Now we can evaluate the expectation of the term in square brackets in (H.1). This amounts to:

$$\begin{aligned}
&\sigma^2R^{-1} + \sigma^4\operatorname{tr}\left(J'J\right)De_1e_1'D + \sigma^4(e_1'DRDe_1)\bar{W}'JJ'\bar{W} \\
&+2\sigma^4\bar{W}'J\bar{W}RDe_1e_1'DR\bar{W}'J'\bar{W} + \sigma^4\operatorname{tr}(R\bar{W}'J\bar{W})De_1e_1'DR\bar{W}'J'\bar{W} \\
&+\sigma^4De_1e_1'DR\bar{W}'J'\bar{W}R\bar{W}'J'\bar{W} + \sigma^4De_1e_1'DR\bar{W}'JJ'\bar{W} \\
&+\sigma^4\operatorname{tr}(R\bar{W}'J\bar{W})\bar{W}'J\bar{W}RDe_1e_1'D + \sigma^4\bar{W}'J\bar{W}R\bar{W}'J\bar{W}RDe_1e_1'D \\
&+\sigma^4\bar{W}'JJ'\bar{W}RDe_1e_1'D + \sigma^4\operatorname{tr}(R\bar{W}'J\bar{W})\operatorname{tr}(R\bar{W}'J\bar{W})De_1e_1'D \\
&+\sigma^4\operatorname{tr}(R\bar{W}'JJ'\bar{W})De_1e_1'D + \sigma^4\operatorname{tr}(R\bar{W}'J\bar{W}R\bar{W}'J\bar{W})De_1e_1'D \\
&-\sigma^4\bar{W}'JJ\bar{W}RDe_1e_1'D - \sigma^4\bar{W}'JJ'\bar{W}RDe_1e_1'D - 2\sigma^4\operatorname{tr}(R\bar{W}'JJ'\bar{W})De_1e_1'D \\
&-2\sigma^4\operatorname{tr}(R\bar{W}'JJ\bar{W})De_1e_1'D - \sigma^4De_1e_1'DR\bar{W}'J'J'\bar{W} - \sigma^4De_1e_1'DR\bar{W}'JJ'\bar{W} \\
&-2\sigma^4(e_1'DRDe_1)\bar{W}'JJ'\bar{W} - \sigma^4(e_1'DRDe_1)\bar{W}'JJ\bar{W} - \sigma^4De_1e_1'DR\bar{W}'JJ'\bar{W} \\
&-\sigma^4De_1e_1'DR\bar{W}'JJ\bar{W} - \sigma^4(e_1'DRDe_1)\bar{W}'J'J'\bar{W} - \sigma^4\bar{W}'JJ'\bar{W}RDe_1e_1'D \\
&-\sigma^4\bar{W}'J'J'\bar{W}RDe_1e_1'D + \sigma^4(e_1'DR\bar{W}'J\bar{W}RDe_1)\bar{W}'J\bar{W} + \sigma^4\bar{W}'JJ'\bar{W}RDe_1e_1'D \\
&+\sigma^4\bar{W}'J\bar{W}RDe_1e_1'DR\bar{W}'J\bar{W} + 2\sigma^4\operatorname{tr}(R\bar{W}'JJ'\bar{W})De_1e_1'D \\
&+2\sigma^4De_1e_1'DR\bar{W}'J'\bar{W}R\bar{W}'J\bar{W} + \sigma^4\operatorname{tr}(R\bar{W}'J\bar{W})(e_1'DRDe_1)\bar{W}'J\bar{W} \\
&+2\sigma^4(e_1'DRDe_1)\bar{W}'JJ'\bar{W} + \sigma^4(e_1'DRDe_1)\bar{W}'J\bar{W}R\bar{W}'J\bar{W} \\
&+\sigma^4\operatorname{tr}(R\bar{W}'J\bar{W})De_1e_1'DR\bar{W}'J\bar{W} + \sigma^4De_1e_1'DR\bar{W}'JJ'\bar{W} \\
&+\sigma^4De_1e_1'DR\bar{W}'J\bar{W}R\bar{W}'J\bar{W} + \sigma^4(e_1'DR\bar{W}'J\bar{W}RDe_1)\bar{W}'J'\bar{W} \\
&+\sigma^4De_1e_1'DR\bar{W}'JJ'\bar{W} + \sigma^4\bar{W}'J'\bar{W}RDe_1e_1'DR\bar{W}'J'\bar{W} \\
&+2\sigma^4\bar{W}'J'\bar{W}R\bar{W}'J\bar{W}RDe_1e_1'D + \sigma^4\operatorname{tr}(R\bar{W}'J\overline{\bar{W}})(e_1'DRDe_1)\bar{W}'J'\bar{W} \\
&+\sigma^4(e_1'DRDe_1)\bar{W}'J'\bar{W}R\bar{W}'J'\bar{W} + \sigma^4\operatorname{tr}(R\bar{W}'J\bar{W})\bar{W}'J'\bar{W}RDe_1e_1'D \\
&+\sigma^4\bar{W}'JJ'\bar{W}RDe_1e_1'D + \sigma^4\bar{W}'J'\bar{W}R\bar{W}'J'\bar{W}RDe_1e_1'D \\
&-2\sigma^4\operatorname{tr}(J'J)De_1e_1'D - 2\sigma^4De_1e_1'DR\bar{W}'J'J\bar{W} - 2\sigma^4\bar{W}'J'J\bar{W}RDe_1e_1'D
\end{aligned}$$

Exploiting $J + J' = \iota\iota' - I$ yields

$$\begin{aligned}
\bar{W}R\bar{W}'(J+J')\bar{W} &= \bar{W}R\bar{W}'(\iota\iota' - I)\bar{W} = (\iota\iota' - I)\bar{W}, \\
\bar{W}'(J+J')\bar{W}R &= \bar{W}'(\iota\iota' - I)\bar{W}R = \bar{W}'\iota\iota'\bar{W}R - I = \bar{W}'\iota e_2' - I, \\
\bar{W}'(J+J')\bar{W}RDe_1 &= (\bar{W}'\iota e_2' - I)De_1 = -De_1.
\end{aligned}$$

Thus we can simplify the term in square brackets and obtain:

$$\begin{aligned}
&\sigma^2R^{-1} \\
&+\sigma^4[2 - \operatorname{tr}\left(J'J\right) - 2\operatorname{tr}(R\bar{W}'J\bar{W}) - 2\operatorname{tr}(R\bar{W}'JJ\bar{W}) + \operatorname{tr}(R\bar{W}'JJ'\bar{W}) \\
&\qquad + \operatorname{tr}(R\bar{W}'J\bar{W}R\bar{W}'J\bar{W}) + \operatorname{tr}(R\bar{W}'J\bar{W})\operatorname{tr}(R\bar{W}'J\bar{W})]De_1e_1'D \\
&+\sigma^4(e_1'DRDe_1)\bar{W}'(JJ' - JJ - J'J')\bar{W} \\
&+\sigma^4\bar{W}'(JJ' - J'J - J'J')\bar{W}RDe_1e_1'D \\
&+\sigma^4De_1e_1'DR\bar{W}'(JJ' - J'J - JJ)\bar{W} \\
&+\sigma^4[(e_1'DR\bar{W}'J\bar{W}RDe_1) + \operatorname{tr}(R\bar{W}'J\bar{W})(e_1'DRDe_1)]\bar{W}'(J+J')\bar{W} \\
&+\sigma^4(e_1'DRDe_1)(\bar{W}'J\bar{W}R\bar{W}'J\bar{W} + \bar{W}'J'\bar{W}R\bar{W}'J'\bar{W})
\end{aligned}$$

To obtain the required result for (H.1) we should pre- and postmultiply the above by R, but first we may remove terms from it that are $o(1)$. Finally pre- and postmultiplying this expression by D yields the result of the theorem.

Table 1:

Bias, variance and MSE of coefficient estimators in model (2.16) for various values of β/σ

$T = 20$	ref.	10	5	2	1	0.5	0.2	0.1
bias $\widehat{\lambda}$		-0.0001	-0.0005	-0.0033	-0.0151	-0.0774	-0.1916	-0.2229
(MCSE)		(0.0000)	(0.0000)	(0.0001)	(0.0001)	(0.0004)	(0.0006)	(0.0006)
$O(T^{-2})$	(3.3)	-0.0002	-0.0006	-0.0037	-0.0150	-0.0600	-0.3750	-1.5000
$O(T^{-2})$	(3.2)	-0.0001	-0.0006	-0.0036	-0.0143	-0.0571	-0.3571	-1.4286
$O(\sigma^2)$	(3.3)	-0.0001	-0.0005	-0.0032	-0.0127	-0.0510	-0.3187	-1.2750
$O(\sigma^2)$	(3.1)	-0.0001	-0.0005	-0.0032	-0.0128	-0.0511	-0.3195	-1.2782
$O(T^{-3})$	(3.3)	-0.0001	-0.0005	-0.0033	-0.0148	-0.0846	-1.6313	-22.275
$O(T^{-3})$	(3.1)	-0.0001	-0.0005	-0.0033	-0.0148	-0.0834	-1.5803	-21.450
$O(\beta^{*2})$	(2.19)	–	–	–	–	-0.0037	-0.1967	-0.2244
Var $\widehat{\lambda}$		0.0000	0.0001	0.0004	0.0018	0.0138	0.0311	0.0331
$O(T^{-3})$	(3.7)	0.0000	0.0001	0.0004	0.0015	0.0060	0.0375	0.1500
$O(T^{-3})$	(3.5)	0.0000	0.0001	0.0004	0.0015	0.0060	0.0376	0.1504
$O(T^{-4})$	(3.7)	0.0000	0.0001	0.0004	0.0017	0.0091	0.1594	2.1000
MSE $\widehat{\lambda}$		0.0000	0.0001	0.0004	0.0020	0.0198	0.0678	0.0828
$O(T^{-4})$	(3.6)	0.0000	0.0001	0.0004	0.0019	0.0127	0.3000	4.3500
$O(T^{-4})$	(3.5)	0.0000	0.0001	0.0004	0.0019	0.0127	0.3002	4.3519
$O(T^{-4})$	(4.15)	0.0000	0.0001	0.0004	0.0019	0.0121	0.2747	3.9448
bias $\widehat{\beta/\sigma}$		0.0159	0.0331	0.0853	0.1757	0.3075	0.2381	0.1323
(MCSE)		(0.0014)	(0.0014)	(0.0014)	(0.0013)	(0.0015)	(0.0020)	(0.0022)
$O(T^{-1})$	(4.12)	0.0171	0.0343	0.0857	0.1714	0.3429	0.8571	1.7143
Var $\widehat{\beta/\sigma}$		0.1864	0.1862	0.1847	0.1816	0.2131	0.4030	0.4821
$O(T^{-2})$	(4.15)	0.1857	0.1856	0.1848	0.1820	0.1709	0.0929	-0.1857
MSE $\widehat{\beta/\sigma}$		0.1866	0.1873	0.1920	0.2125	0.3077	0.4596	0.4996
$O(T^{-2})$	(4.15)	0.1860	0.1867	0.1921	0.2114	0.2884	0.8276	2.7531

Table 2:

Bias, variance and MSE of coefficient estimators in model (2.16) for various values of β/σ

$T = 40$	ref.	10	5	2	1	0.5	0.2	0.1
bias $\widehat{\lambda}$		-0.0000	-0.0001	-0.0009	-0.0037	-0.0193	-0.0845	-0.1131
(MCSE)		(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0001)	(0.0003)	(0.0003)
$O(T^{-2})$	(3.3)	-0.0000	-0.0002	-0.0009	-0.0037	-0.0150	-0.0938	-0.3750
$O(T^{-2})$	(3.2)	-0.0000	-0.0001	-0.0009	-0.0037	-0.0146	-0.0915	-0.3659
$O(\sigma^2)$	(3.3)	-0.0000	-0.0001	-0.0009	-0.0035	-0.0139	-0.0867	-0.3469
$O(\sigma^2)$	(3.1)	-0.0000	-0.0001	-0.0009	-0.0035	-0.0139	-0.0868	-0.3471
$O(T^{-3})$	(3.3)	-0.0000	-0.0001	-0.0009	-0.0037	-0.0181	-0.2508	-2.9719
$O(T^{-3})$	(3.1)	-0.0000	-0.0001	-0.0009	-0.0037	-0.0180	-0.2472	-2.9136
$O(\beta^{*2})$	(2.19)	–	–	–	–	0.1599	-0.0795	-0.1137
Var $\widehat{\lambda}$		0.0000	0.0000	0.0000	0.0002	0.0014	0.0081	0.0098
$O(T^{-3})$	(3.7)	0.0000	0.0000	0.0000	0.0002	0.0008	0.0047	0.0187
$O(T^{-3})$	(3.5)	0.0000	0.0000	0.0000	0.0002	0.0008	0.0047	0.0188
$O(T^{-4})$	(3.7)	0.0000	0.0000	0.0000	0.0002	0.0009	0.0123	0.1406
MSE $\widehat{\lambda}$		0.0000	0.0000	0.0000	0.0002	0.0018	0.0153	0.0226
$O(T^{-4})$	(3.6)	0.0000	0.0000	0.0000	0.0002	0.0012	0.0211	0.2813
$O(T^{-4})$	(3.5)	0.0000	0.0000	0.0000	0.0002	0.0012	0.0211	0.2813
$O(T^{-4})$	(4.15)	0.0000	0.0000	0.0000	0.0002	0.0011	0.0203	0.2679
bias $\widehat{\beta/\sigma}$		0.0078	0.0171	0.0452	0.0930	0.1895	0.2269	0.1407
(MCSE)		(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0014)	(0.0016)
$O(T^{-1})$	(4.12)	0.0093	0.0185	0.0463	0.0927	0.1854	0.4634	0.9268
Var $\widehat{\beta/\sigma}$		0.0962	0.0962	0.0958	0.0951	0.0993	0.1886	0.2604
$O(T^{-2})$	(4.15)	0.0963	0.0963	0.0961	0.0955	0.0930	0.0758	0.0140
MSE $\widehat{\beta/\sigma}$		0.0963	0.0965	0.0979	0.1037	0.1352	0.2401	0.2802
$O(T^{-2})$	(4.15)	0.0964	0.0967	0.0983	0.1041	0.1274	0.2905	0.8730

Table 3:
Bias, variance and MSE of coefficient estimators in model (2.16) for various values of β/σ

$T=80$	ref.	10	5	2	1	0.5	0.2	0.1
bias $\widehat{\lambda}$		-0.0000	-0.0000	-0.0002	-0.0009	-0.0042	-0.0307	-0.0531
(MCSE)		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0002)
$O(T^{-2})$	(3.3)	-0.0000	-0.0000	-0.0002	-0.0009	-0.0037	-0.0234	-0.0938
$O(T^{-2})$	(3.2)	-0.0000	-0.0000	-0.0002	-0.0009	-0.0037	-0.0231	-0.0926
$O(\sigma^2)$	(3.3)	-0.0000	-0.0000	-0.0002	-0.0009	-0.0036	-0.0226	-0.0902
$O(\sigma^2)$	(3.1)	-0.0000	-0.0000	-0.0002	-0.0009	-0.0036	-0.0226	-0.0902
$O(T^{-3})$	(3.3)	-0.0000	-0.0000	-0.0002	-0.0009	-0.0041	-0.0431	-0.4184
$O(T^{-3})$	(3.1)	-0.0000	-0.0000	-0.0002	-0.0009	-0.0041	-0.0428	-0.4145
$O(\beta^{*2})$	(2.19)	–	–	–	–	–	-0.0149	-0.0523
Var $\widehat{\lambda}$		0.0000	0.0000	0.0000	0.0000	0.0001	0.0016	0.0026
$O(T^{-3})$	(3.7)	0.0000	0.0000	0.0000	0.0000	0.0001	0.0006	0.0023
$O(T^{-3})$	(3.5)	0.0000	0.0000	0.0000	0.0000	0.0001	0.0006	0.0023
$O(T^{-4})$	(3.7)	0.0000	0.0000	0.0000	0.0000	0.0001	0.0011	0.0100
MSE $\widehat{\lambda}$		0.0000	0.0000	0.0000	0.0000	0.0001	0.0025	0.0054
$O(T^{-4})$	(3.6)	0.0000	0.0000	0.0000	0.0000	0.0001	0.0016	0.0187
$O(T^{-4})$	(3.5)	0.0000	0.0000	0.0000	0.0000	0.0001	0.0016	0.0188
$O(T^{-4})$	(4.15)	0.0000	0.0000	0.0000	0.0000	0.0001	0.0016	0.0183
bias $\widehat{\beta}/\sigma$		0.0037	0.0085	0.0230	0.0474	0.0983	0.1817	0.1368
(MCSE)		(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0009)	(0.0011)
$O(T^{-1})$	(4.12)	0.0048	0.0096	0.0241	0.0481	0.0963	0.2407	0.4815
Var $\widehat{\beta}/\sigma$		0.0491	0.0491	0.0491	0.0489	0.0489	0.0763	0.1235
$O(T^{-2})$	(4.15)	0.0491	0.0491	0.0490	0.0489	0.0483	0.0442	0.0295
MSE $\widehat{\beta}/\sigma$		0.0492	0.0492	0.0496	0.0511	0.0586	0.1093	0.1422
$O(T^{-2})$	(4.15)	0.0491	0.0492	0.0496	0.0512	0.0576	0.1021	0.2613

Table 4:
Moments of $\widehat{\lambda}$, $\widehat{\beta}_1$ and $\widehat{\beta}_2$ in model (5.1) for various values of β_1/σ and $\beta_2/\sigma = 0.1$

$T=20$	ref.	10	5	2	1	0.5	0.2	0.1
bias $\widehat{\lambda}$		-0.1776	-0.1776	-0.1776	-0.1776	-0.1776	-0.1776	-0.1776
(MCSE)		(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
$O(T^{-3})$	(4.13)	-0.2051	-0.2051	-0.2051	-0.2051	-0.2051	-0.2051	-0.2051
$O(T^{-4})$	(3.1)	-0.2391	-0.2391	-0.2391	-0.2391	-0.2391	-0.2391	-0.2391
Var $\widehat{\lambda}$		0.0287	0.0287	0.0287	0.0287	0.0287	0.0287	0.0287
$O(T^{-6})$		0.0200	0.0200	0.0200	0.0200	0.0200	0.0200	0.0200
MSE $\widehat{\lambda}$		0.0602	0.0602	0.0602	0.0602	0.0602	0.0602	0.0602
$O(T^{-6})$	(3.5)	0.0708	0.0708	0.0708	0.0708	0.0708	0.0708	0.0708
$O(T^{-6})$	(4.15)	0.0620	0.0620	0.0620	0.0620	0.0620	0.0620	0.0620
bias $\widehat{\beta}_1$		-2.4414	-1.5534	-1.0207	-0.8431	-0.7543	-0.7010	-0.6833
(MCSE)		(0.0065)	(0.0040)	(0.0028)	(0.0024)	(0.0023)	(0.0022)	(0.0022)
$O(T^{-1})$	(4.12)	-2.6824	-1.7710	-1.2242	-1.0419	-0.9508	-0.8961	-0.8779
Var $\widehat{\beta}_1$		4.1908	1.6147	0.7577	0.5869	0.5230	0.4915	0.4822
$O(T^{-2})$		4.3787	1.6642	0.7261	0.5286	0.4513	0.4119	0.3999
MSE $\widehat{\beta}_1$		10.151	4.0278	1.7995	1.2977	1.0919	0.9829	0.9490
$O(T^{-2})$	(4.15)	11.574	4.8007	2.2248	1.6142	1.3553	1.2149	1.1706
bias $\widehat{\beta}_2$		1.9638	1.0759	0.5432	0.3656	0.2768	0.2235	0.2058
(MCSE)		(0.0059)	(0.0032)	(0.0016)	(0.0011)	(0.0008)	(0.0006)	(0.0006)
$O(T^{-2})$	(4.12)	2.0201	1.1087	0.5619	0.3796	0.2884	0.2338	0.2155
Var $\widehat{\beta}_2$		3.4329	1.0138	0.2510	0.1115	0.0633	0.0413	0.0351
$O(T^{-4})$	(4.15)	3.4165	1.0013	0.2429	0.1051	0.0579	0.0364	0.0304
MSE $\widehat{\beta}_2$		7.2895	2.1714	0.5461	0.2452	0.1399	0.0912	0.0774
$O(T^{-4})$	(4.15)	7.4971	2.2305	0.5585	0.2492	0.1411	0.0911	0.0769

Table 5:
Moments of $\widehat{\lambda}$, $\widehat{\beta}_1$ and $\widehat{\beta}_2$ in model (5.1) for various values of β_1/σ and $\beta_2/\sigma = 0.1$

$T = 40$	ref.	10	5	2	1	0.5	0.2	0.1
bias $\widehat{\lambda}$		-0.0131	-0.0131	-0.0131	-0.0131	-0.0131	-0.0131	-0.0131
(MCSE)		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
$O(T^{-3})$	(4.13)	-0.0134	-0.0134	-0.0134	-0.0134	-0.0134	-0.0134	-0.0134
$O(T^{-4})$	(3.1)	-0.0131	-0.0131	-0.0131	-0.0131	-0.0131	-0.0131	-0.0131
Var $\widehat{\lambda}$		0.0007	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
$O(T^{-6})$		0.0007	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
MSE $\widehat{\lambda}$		0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009
$O(T^{-6})$	(3.5)	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009
$O(T^{-6})$	(4.15)	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009	0.0009
bias $\widehat{\beta}_1$		-0.3510	-0.2856	-0.2464	-0.2333	-0.2268	-0.2229	-0.2216
(MCSE)		(0.0022)	(0.0019)	(0.0017)	(0.0016)	(0.0016)	(0.0016)	(0.0016)
$O(T^{-1})$	(4.12)	-0.3476	-0.2841	-0.2460	-0.2333	-0.2270	-0.2232	-0.2219
Var $\widehat{\beta}_1$		0.4999	0.3526	0.2813	0.2605	0.2506	0.2448	0.2429
$O(T^{-2})$		0.5021	0.3533	0.2814	0.2603	0.2503	0.2445	0.2426
MSE $\widehat{\beta}_1$		0.6231	0.4341	0.3420	0.3149	0.3020	0.2945	0.2920
$O(T^{-2})$	(4.15)	0.6229	0.4341	0.3419	0.3148	0.3018	0.2943	0.2918
bias $\widehat{\beta}_2$		0.1588	0.0934	0.0542	0.0411	0.0345	0.0306	0.0293
(MCSE)		(0.0010)	(0.0006)	(0.0003)	(0.0003)	(0.0002)	(0.0002)	(0.0002)
$O(T^{-2})$	(4.12)	0.1542	0.0907	0.0527	0.0400	0.0336	0.0298	0.0285
Var $\widehat{\beta}_2$		0.1027	0.0350	0.0115	0.0065	0.0046	0.0036	0.0033
$O(T^{-4})$	(4.15)	0.1033	0.0351	0.0115	0.0065	0.0046	0.0036	0.0033
MSE $\widehat{\beta}_2$		0.1279	0.0437	0.0144	0.0082	0.0058	0.0045	0.0041
$O(T^{-4})$	(4.15)	0.1271	0.0434	0.0143	0.0081	0.0057	0.0045	0.0041

Table 6:
Moments of $\widehat{\lambda}$, $\widehat{\beta}_1$ and $\widehat{\beta}_2$ in model (5.1) for various values of β_1/σ and $\beta_2/\sigma = 0.1$

$T = 80$	ref.	10	5	2	1	0.5	0.2	0.1
bias $\widehat{\lambda}$		-0.0008	-0.0008	-0.0008	-0.0008	-0.0008	-0.0008	-0.0008
(MCSE)		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$O(T^{-3})$	(4.13)	-0.0009	-0.0009	-0.0009	-0.0009	-0.0009	-0.0009	-0.0009
$O(T^{-4})$	(3.1)	-0.0008	-0.0008	-0.0008	-0.0008	-0.0008	-0.0008	-0.0008
Var $\widehat{\lambda}$		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$O(T^{-6})$		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MSE $\widehat{\lambda}$		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$O(T^{-6})$	(3.5)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$O(T^{-6})$	(4.15)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
bias $\widehat{\beta}_1$		-0.0653	-0.0611	-0.0586	-0.0577	-0.0573	-0.0571	-0.0570
(MCSE)		(0.0012)	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)
$O(T^{-1})$	(4.12)	-0.0641	-0.0599	-0.0574	-0.0566	-0.0562	-0.0559	-0.0558
Var $\widehat{\beta}_1$		0.1440	0.1303	0.1226	0.1201	0.1189	0.1182	0.1179
$O(T^{-2})$		0.1442	0.1305	0.1228	0.1203	0.1191	0.1183	0.1181
MSE $\widehat{\beta}_1$		0.1482	0.1340	0.1260	0.1235	0.1222	0.1214	0.1212
$O(T^{-2})$	(4.15)	0.1483	0.1341	0.1261	0.1235	0.1222	0.1215	0.1212
bias $\widehat{\beta}_2$		0.0120	0.0078	0.0053	0.0045	0.0040	0.0038	0.0037
(MCSE)		(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
$O(T^{-2})$	(4.12)	0.0119	0.0078	0.0052	0.0044	0.0040	0.0037	0.0037
Var $\widehat{\beta}_2$		0.0043	0.0018	0.0008	0.0006	0.0005	0.0004	0.0004
$O(T^{-4})$	(4.15)	0.0043	0.0018	0.0008	0.0006	0.0005	0.0004	0.0004
MSE $\widehat{\beta}_2$		0.0045	0.0019	0.0008	0.0006	0.0005	0.0004	0.0004
$O(T^{-4})$	(4.15)	0.0045	0.0019	0.0008	0.0006	0.0005	0.0004	0.0004

Table 7:
Moments of $\widehat{\lambda}$, $\widehat{\beta}_1$ and $\widehat{\beta}_2$ in model (5.2) for $\gamma_2 = \{0.5, 0.9\}$

$T = 20$	ref.	$SN = 1$	$SN = 4$	$SN = 19$		$SN = 1$	$SN = 4$	$SN = 19$
		$\beta_2 = 1.00$	$\beta_2 = 2.00$	$\beta_2 = 4.36$		$\beta_2 = 1.00$	$\beta_2 = 2.00$	$\beta_2 = 4.36$
bias $\widehat{\lambda}$		-0.0918	-0.0266	-0.0055		-0.1376	-0.0426	-0.0077
(MCSE)		(0.0001)	(0.0001)	(0.0000)		(0.0002)	(0.0001)	(0.0000)
$O(T^{-3})$	(4.13)	-0.1165	-0.0291	-0.0061		-0.1535	-0.0384	-0.0081
$O(T^{-4})$	(3.1)	-0.1316	-0.0277	-0.0055		-0.3270	-0.0460	-0.0077
Var $\widehat{\lambda}$		0.0127	0.0033	0.0007		0.0235	0.0058	0.0009
$O(T^{-6})$		0.0139	0.0033	0.0007		0.0333	0.0053	0.0009
MSE $\widehat{\lambda}$		0.0212	0.0040	0.0007		0.0424	0.0076	0.0010
$O(T^{-6})$	(3.5)	0.0303	0.0043	0.0007		0.0618	0.0071	0.0010
$O(T^{-6})$	(4.15)	0.0274	0.0041	0.0007		0.0569	0.0068	0.0010
bias $\widehat{\beta}_1$		0.1532	0.1312	0.0706		-0.0477	0.0224	0.0320
(MCSE)		(0.0005)	(0.0004)	(0.0003)		(0.0006)	(0.0004)	(0.0003)
$O(T^{-1})$	(4.12)	0.3201	0.1600	0.0734		0.1625	0.0812	0.0373
Var $\widehat{\beta}_1$		0.2325	0.1361	0.1219		0.3832	0.1489	0.0639
$O(T^{-2})$		0.1436	0.1263	0.1218		0.2569	0.1023	0.0616
MSE $\widehat{\beta}_1$		0.2559	0.1534	0.1269		0.3855	0.1494	0.0650
$O(T^{-2})$	(4.15)	0.2460	0.1519	0.1272		0.2832	0.1089	0.0630
bias $\widehat{\beta}_2$		-0.0124	-0.0060	-0.0030		-0.0047	0.0148	0.0124
(MCSE)		(0.0002)	(0.0002)	(0.0002)		(0.0004)	(0.0004)	(0.0004)
$O(T^{-2})$	(4.12)	-0.0129	-0.0064	-0.0030		0.0596	0.0298	0.0137
Var $\widehat{\beta}_2$		0.0624	0.0606	0.0605		0.1832	0.1514	0.1441
$O(T^{-4})$	(4.15)	0.0604	0.0605	0.0606		0.1509	0.1453	0.1438
MSE $\widehat{\beta}_2$		0.0625	0.0606	0.0605		0.1832	0.1516	0.1443
$O(T^{-4})$	(4.15)	0.0606	0.0606	0.0606		0.1545	0.1462	0.1440

Table 8:
Moments of $\widehat{\lambda}$, $\widehat{\beta}_1$ and $\widehat{\beta}_2$ in model (5.2) for $\gamma_2 = \{0.5, 0.9\}$

$T = 40$	ref.	$SN = 1$	$SN = 4$	$SN = 19$		$SN = 1$	$SN = 4$	$SN = 19$
		$\beta_2 = 1.00$	$\beta_2 = 2.00$	$\beta_2 = 4.36$		$\beta_2 = 1.00$	$\beta_2 = 2.00$	$\beta_2 = 4.36$
bias $\widehat{\lambda}$		-0.0369	-0.0094	-0.0019		-0.0275	-0.0072	-0.0015
(MCSE)		(0.0000)	(0.0000)	(0.0000)		(0.0000)	(0.0000)	(0.0000)
$O(T^{-3})$	(4.13)	-0.0380	-0.0095	-0.0020		-0.0307	-0.0077	-0.0016
$O(T^{-4})$	(3.1)	-0.0450	-0.0096	-0.0019		-0.0280	-0.0072	-0.0015
Var $\widehat{\lambda}$		0.0025	0.0005	0.0001		0.0016	0.0004	0.0001
$O(T^{-6})$		0.0024	0.0005	0.0001		0.0014	0.0004	0.0001
MSE $\widehat{\lambda}$		0.0038	0.0006	0.0001		0.0023	0.0004	0.0001
$O(T^{-6})$	(3.5)	0.0039	0.0006	0.0001		0.0024	0.0005	0.0001
$O(T^{-6})$	(4.15)	0.0038	0.0006	0.0001		0.0023	0.0004	0.0001
bias $\widehat{\beta}_1$		-0.0035	-0.0142	-0.0095		-0.0359	-0.0222	-0.0109
(MCSE)		(0.0003)	(0.0002)	(0.0002)		(0.0003)	(0.0002)	(0.0002)
$O(T^{-1})$	(4.12)	-0.0452	-0.0226	-0.0104		-0.0488	-0.0244	-0.0112
Var $\widehat{\beta}_1$		0.1004	0.0439	0.0295		0.0850	0.0470	0.0370
$O(T^{-2})$		0.0902	0.0421	0.0294		0.0844	0.0469	0.0370
MSE $\widehat{\beta}_1$		0.1004	0.0441	0.0296		0.0863	0.0475	0.0371
$O(T^{-2})$	(4.15)	0.0922	0.0426	0.0295		0.0868	0.0475	0.0371
bias $\widehat{\beta}_2$		0.0157	0.0134	0.0070		0.1052	0.0618	0.0295
(MCSE)		(0.0002)	(0.0002)	(0.0002)		(0.0003)	(0.0003)	(0.0003)
$O(T^{-2})$	(4.12)	0.0315	0.0157	0.0072		0.1307	0.0653	0.0300
Var $\widehat{\beta}_2$		0.0344	0.0335	0.0335		0.0658	0.0691	0.0710
$O(T^{-4})$	(4.15)	0.0327	0.0333	0.0335		0.0611	0.0690	0.0711
MSE $\widehat{\beta}_2$		0.0346	0.0337	0.0336		0.0769	0.0730	0.0719
$O(T^{-4})$	(4.15)	0.0337	0.0336	0.0336		0.0782	0.0733	0.0720

Table 9:

Moments of $\widehat{\lambda}$, $\widehat{\beta}_1$ and $\widehat{\beta}_2$ in model (5.2) for $\gamma_2 = \{0.5, 0.9\}$

$T = 80$	ref.	$SN = 1$ $\beta_2 = 1.00$	$SN = 4$ $\beta_2 = 2.00$	$SN = 19$ $\beta_2 = 4.36$		$SN = 1$ $\beta_2 = 1.00$	$SN = 4$ $\beta_2 = 2.00$	$SN = 19$ $\beta_2 = 4.36$
bias $\widehat{\lambda}$		-0.0156	-0.0045	-0.0010		-0.0091	-0.0024	-0.0005
(MCSE)		(0.0000)	(0.0000)	(0.0000)		(0.0000)	(0.0000)	(0.0000)
$O(T^{-3})$	(4.13)	-0.0193	-0.0048	-0.0010		-0.0103	-0.0026	-0.0005
$O(T^{-4})$	(3.1)	-0.0154	-0.0045	-0.0010		-0.0090	-0.0024	-0.0005
Var $\widehat{\lambda}$		0.0004	0.0001	0.0000		0.0002	0.0001	0.0000
$O(T^{-6})$		0.0004	0.0001	0.0000		0.0002	0.0001	0.0000
MSE $\widehat{\lambda}$		0.0007	0.0001	0.0000		0.0003	0.0001	0.0000
$O(T^{-6})$	(3.5)	0.0007	0.0001	0.0000		0.0003	0.0001	0.0000
$O(T^{-6})$	(4.15)	0.0007	0.0001	0.0000		0.0003	0.0001	0.0000
bias $\widehat{\beta}_1$		0.0302	0.0204	0.0106		-0.0319	-0.0194	-0.0095
(MCSE)		(0.0002)	(0.0002)	(0.0001)		(0.0002)	(0.0001)	(0.0001)
$O(T^{-1})$	(4.12)	0.0473	0.0237	0.0109		-0.0423	-0.0211	-0.0097
Var $\widehat{\beta}_1$		0.0448	0.0227	0.0161		0.0324	0.0207	0.0176
$O(T^{-2})$		0.0488	0.0229	0.0161		0.0326	0.0207	0.0176
MSE $\widehat{\beta}_1$		0.0457	0.0231	0.0162		0.0334	0.0211	0.0176
$O(T^{-2})$	(4.15)	0.0510	0.0235	0.0162		0.0344	0.0212	0.0177
bias $\widehat{\beta}_2$		0.0208	0.0136	0.0069		0.0396	0.0216	0.0102
(MCSE)		(0.0001)	(0.0001)	(0.0001)		(0.0001)	(0.0001)	(0.0001)
$O(T^{-2})$	(4.12)	0.0299	0.0149	0.0069		0.0444	0.0222	0.0102
Var $\widehat{\beta}_2$		0.0117	0.0119	0.0120		0.0140	0.0142	0.0143
$O(T^{-4})$	(4.15)	0.0113	0.0118	0.0120		0.0138	0.0142	0.0143
MSE $\widehat{\beta}_2$		0.0122	0.0120	0.0120		0.0156	0.0147	0.0144
$O(T^{-4})$	(4.15)	0.0122	0.0120	0.0120		0.0158	0.0147	0.0144