STOCKHOLM SCHOOL
OF ECONOMICS
THE ECONOMIC RESEARCH INSTITUTE

A Second-Order Approximation to the Bias of OLS Estimates in Bivariate VAR Models

Tomas Brännström

Working Paper No. 81
November 1995

Working Paper Series
in Economics and Finance

A Second-Order Approximation to the Bias of OLS Estimates in Bivariate VAR Models

Tomas Brännström
Stockholm School of Economics
P.O. Box 6501
S-113 83 Stockholm, Sweden

ABSTRACT: Two terms of a second-order approximation to the bias of the multivariate OLS estimate are derived using the same technique as in Nicholls and Pope (1988). The resulting second-order bias approximation is then tested against first-order alternatives on two bivariate Monte Carlo simulated VAR models. Even though the second-order terms do not completely remedy the bias left unadjusted by the first-order approximations, it makes a significant change in the right direction where the first-order approximations fail, notably in the smallest samples and when systems approach non-stationarity.

KEY WORDS: Vector autoregressive models – Least-squares estimates – Bias approximation

JEL CLASSIFICATION: C 32

1. Introduction

Although the bias of least squares estimates of VAR parameters can be sizeable, it is still a method frequently used by analysts. In most cases no attention is paid to the effects of this bias on the results of the analysis, but if it were, there are at least two approximations to the bias (Tjøstheim and Paulsen 1983; Nicholls and Pope 1988) capable of reducing the bias problem. Monte Carlo studies (Brännström and Karlsson 1993; Brännström 1994*a*; Brännström 1994*b*) have examined the properties of these two approximations and found them to approximate the bias well, at least in simple, bivariate first- and second-order VAR models. The results also suggest that bias-reduced estimates may be obtained using any one of these approximations, thereby reducing not only the bias but also the mean square error of the parameter estimates.

However, the approximations do not perform equally well for all sample sizes and for all eigenvalue combinations. More precisely, in samples as small as 25 or 50 observations, the approximations perform worse than in samples of size 100 or 200. Furthermore they perform better when the system is stationary than when one of the characteristic roots approaches unity. These differences are probably (at least in part) due to the fact that the approximations are first order and therefore include no $O(T^{-2})$ terms. The present paper attempts to add second-order terms to the approximations and investigate to what extent such an expansion improves the performance of the approximations in the cases outlined above. The approximations will be evaluated using estimated as well as true parameters.

It is however important to emphasize that the analysis is partial in the sense that only two second-order terms are derived, although presumably the two most important terms. Also, Monte Carlo simulations of two bivariate VAR systems are used to evaluate the approximations, which of course means that the analysis will be very restricted but, as Lütkepohl (1993) points out, in small samples analytical results are difficult to obtain and so one has to resort to Monte Carlo methods.

The paper is organized as follows. The VAR models and the two first-order approximations are presented in Section 2. Second-order terms are derived in Section 3. Section 4 holds a comparison between a first-order and a second-order approximation for Monte Carlo simulated data, and bias-reduced estimates are constructed and evaluated in Section 5. Section 6 concludes.

2. Preliminaries

Consider the following m-dimensional VAR(p) process:

$$\mathbf{x}_t = \mu + \mathbf{A}_1\mathbf{x}_{t-1} + \mathbf{A}_2\mathbf{x}_{t-2} + \cdots + \mathbf{A}_p\mathbf{x}_{t-p} + \varepsilon_t \tag{2.1}$$

where $\mathbf{x}$, μ and ε are $m\times 1$ vectors and the parameter matrices $\mathbf{A}_i$ are $m\times m$. Assuming (for simplicity) that μ is a zero vector and stacking lagged $\mathbf{x}$ vectors in $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$ makes it possible to write (2.1) as a VAR(1) process:

$$\mathbf{x}_t^* = \mathbf{A}\mathbf{x}_{t-1}^* + \varepsilon_t^* \tag{2.2}$$

$$\text{where } \mathbf{A} = \left[\begin{array}{ccc|c} \mathbf{A}_1 & \cdots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \hline & \mathbf{I}_{m(p-1)} & & \mathbf{0} \end{array}\right]$$

is an $mp\times mp$ matrix and $\mathbf{x}_t^* = \begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \\ \vdots \\ \mathbf{x}_{t-p+1} \end{bmatrix}$ and $\varepsilon_t^* = \begin{bmatrix} \varepsilon_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$ are $mp\times 1$ vectors.

It is a well-known fact that estimating the parameter matrix $\mathbf{A}$ in (2.2) by means of OLS introduces a bias (in finite samples). There are however a number of ways to approximate the bias of the OLS estimate (henceforth denoted by $\hat{\mathbf{A}}$), for instance that suggested by Tjøstheim and Paulsen (1983):

$$E(\hat{\mathbf{A}}-\mathbf{A}) \approx$$
$$-\frac{1}{T}\left[2\mathbf{A}+(\mathbf{I}-\mathbf{A})\mathbf{C}_0(\mathbf{I}-\mathbf{A}')^{-1}\mathbf{C}_0^{-1}+\mathbf{G}\mathbf{A}'(\mathbf{I}-\mathbf{A}'\mathbf{A}')^{-1}\mathbf{C}_0^{-1}+\mathbf{G}\sum_{i=1}^{mp}\rho_i(\mathbf{I}-\rho_i\mathbf{A})^{-1}\mathbf{C}_0^{-1}\right]$$
$$+\frac{1}{T-p+1}\left[\begin{array}{ccc}\mathbf{A}_1 & \cdots & \mathbf{A}_p \\ \hline \mathbf{0} & \cdots & \mathbf{0}\end{array}\right] \tag{2.3}$$

where T is the sample size, $\mathbf{I}$ is the identity matrix of dimension mp, $\mathbf{C}_0$ is the autocovariance function of $\mathbf{x}_t$, $\mathbf{G}$ is the autocovariance function of ε_t, and ρ_i (i=1,...,mp) are the eigenvalues of $\mathbf{A}$. Nicholls and Pope (1988) attack the same problem in a different manner and derive the following approximation:

$$E(\hat{\mathbf{A}}-\mathbf{A}) \approx -\frac{1}{T}\mathbf{G}\left[(\mathbf{I}-\mathbf{A}')^{-1}+\mathbf{A}'(\mathbf{I}-\mathbf{A}'\mathbf{A}')^{-1}+\sum_{i=1}^{mp}\rho_i(\mathbf{I}-\rho_i\mathbf{A}')^{-1}\right]\mathbf{C}_0^{-1} \tag{2.4}$$

The two approximations may appear to bear but faint resemblance with each other, but it is not difficult to show that they are in fact equivalent for first-order VAR models (but not for $p \geq 2$ since the denominator in the last term of (2.3) will then be less than
T). Both or them are of a first-order character and therefore include no $O(T^{-2})$ terms. Nevertheless, Monte Carlo studies (Brännström and Karlsson 1993; Brännström 1994*a*; Brännström 1994*b*) have demonstrated that for a first-order bivariate VAR model (m=2, p=1) the approximations perform excellently, whereas for a second-order bivariate VAR model (m=p=2) both approximations perform worse and that (2.3) is then slightly outperformed by (2.4). The conclusions appear to hold even under the more realistic assumption that $\mathbf{A}$, $\mathbf{G}$, $\mathbf{C}_0$ and ρ_i are unknown and replaced by their estimates, even though both approximations then perform slightly worse. It was also demonstrated in Brännström (1994*a*) and Brännström (1994*b*) how these bias approximations can be used to construct bias-reduced estimates, and that such estimates have smaller mean square error than the original estimates.

On the other hand, the Monte Carlo simulations produce less impressive results when one (or more) eigenvalue approaches unity (ρ=0.9 or ρ=1.0) and for the smallest sample sizes (T=25 or T=50) than for stationary processes and sample sizes of 100 or 200. In an earlier paper it was proposed that by adding second-order terms, such behaviour on the part of the bias functions might be better modelled. In the next section, a second-order term will be derived and tested against the first-order alternative (2.3).

3. Derivation of second-order terms

Both approximations are based on the Yule-Walker estimate $\hat{\mathbf{A}} = \hat{\mathbf{C}}_{-1}\hat{\mathbf{C}}_0^{-1}$, where $\mathbf{C}_{-1}$ is the covariance function between $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$ and $\mathbf{C}_0$ is the autocovariance function of $\mathbf{x}_t$, but unlike Tjøstheim and Paulsen (1983), the discussion in Nicholls and Pope (1988) centres around the equivalent expression $\hat{\mathbf{A}} = (\mathbf{A}+\mathbf{P})(\mathbf{I}+\mathbf{Q})^{-1}$, where
$\mathbf{P} = (\hat{\mathbf{C}}_{-1} - \mathbf{C}_{-1})\mathbf{C}_0^{-1}$ and $\mathbf{Q} = (\hat{\mathbf{C}}_0 - \mathbf{C}_0)\mathbf{C}_0^{-1}$.
Since $(\mathbf{I}+\mathbf{Q})^{-1} = \mathbf{I} - \mathbf{Q} + \mathbf{Q}^2 - \mathbf{Q}^3 + \ldots$, this approach leads to the following alternative expression:

$$\hat{\mathbf{A}} = (\mathbf{A}+\mathbf{P})\sum_{k=0}^{\infty}(-1)^k\mathbf{Q}^k \qquad (3.1)$$

The precision of the resulting approximation will thus (basically) depend on how many terms in this sum are evaluated. Nicholls and Pope derive (2.4) as an approximate expression for the expectation of the first five terms $\mathbf{A} + \mathbf{P} - \mathbf{AQ} - \mathbf{PQ} + \mathbf{AQ}^2$, ignoring all remaining terms of (3.1). However, the expectation of $\mathbf{PQ}^2 - \mathbf{AQ}^3$, the two terms to follow next, is not very difficult to evaluate:

$$E\left(\mathbf{PQ}^2 - \mathbf{AQ}^3\right) = E\left((\mathbf{P}-\mathbf{AQ})\mathbf{Q}^2\right) = E\left[\left(\hat{\mathbf{C}}_{-1}\mathbf{C}_0^{-1} - \mathbf{A} - \mathbf{A}\left(\hat{\mathbf{C}}_0\mathbf{C}_0^{-1} - \mathbf{I}\right)\right)\mathbf{Q}^2\right] = E\left[\left(\hat{\mathbf{C}}_{-1} - \mathbf{A}\hat{\mathbf{C}}_0\right)\mathbf{C}_0^{-1}\mathbf{Q}^2\right]$$
$$= E\left(\hat{\mathbf{C}}_{-1} - \mathbf{A}\hat{\mathbf{C}}_0\right)\mathbf{C}_0^{-1}E\left(\mathbf{Q}^2\right) + E\left[\left(\hat{\mathbf{C}}_{-1} - E\left(\hat{\mathbf{C}}_{-1}\right) - \mathbf{A}\left(\hat{\mathbf{C}}_0 - E\left(\hat{\mathbf{C}}_0\right)\right)\right)\mathbf{C}_0^{-1}\left(\mathbf{Q}^2 - E\left(\mathbf{Q}^2\right)\right)\right]$$

Since $\mathbf{Q}^2 - E(\mathbf{Q}^2)$ is $O_p(T^{-2})$ and $\hat{\mathbf{C}} - E(\hat{\mathbf{C}})$ is $O_p(T^{-1})$ as demonstrated by Nicholls and Pope (1988) and $\mathbf{A}$ and $\mathbf{C}_0^{-1}$ are $O(1)$, the second term on the last row above is $O_p(T^{-3})$ and can be ignored from this point. The first term can of course be decomposed into $E(\hat{\mathbf{C}}_{-1})\mathbf{C}_0^{-1}E(\mathbf{Q}^2) - \mathbf{A}E(\hat{\mathbf{C}}_0)\mathbf{C}_0^{-1}E(\mathbf{Q}^2)$, making it clear that three expectations must be evaluated. Fortunately, these three approximations were derived as well in Nicholls and Pope (1988):

$$E(\hat{\mathbf{C}}_{-1})\mathbf{C}_0^{-1} = \mathbf{A} - \frac{1}{T}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{G}(\mathbf{I}-\mathbf{A}')^{-1}\mathbf{C}_0^{-1} + O\left(T^{-2}\right)$$

$$E(\hat{\mathbf{C}}_{0})\mathbf{C}_0^{-1} = \mathbf{I} - \frac{1}{T}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{G}(\mathbf{I}-\mathbf{A}')^{-1}\mathbf{C}_0^{-1} + O\left(T^{-2}\right)$$

$$(T-1)E\left(\mathbf{Q}^2\right) = (m+1)\mathbf{I} + \mathbf{A}^2\left(\mathbf{I}-\mathbf{A}^2\right)^{-1} + \mathbf{C}_0\mathbf{A}'\mathbf{A}'\mathbf{C}_0^{-1}\left(\mathbf{I}-\mathbf{C}_0\mathbf{A}'\mathbf{A}'\mathbf{C}_0^{-1}\right)^{-1} + \\ + \sum_{i=1}^{mp}\rho_i\left(\mathbf{A}\left(\mathbf{I}-\rho_i\mathbf{A}\right)^{-1} + \mathbf{C}_0\mathbf{A}'\mathbf{C}_0^{-1}\left(\mathbf{I}-\rho_i\mathbf{C}_0\mathbf{A}'\mathbf{C}_0^{-1}\right)^{-1}\right) + O\left(T^{-1}\right)$$

After some manipulation, an extended version of (2.3) or (2.4) is thus found to be

$$E(\hat{\mathbf{A}}-\mathbf{A}) \approx \text{ (2.3) or (2.4)} \\ -\frac{1}{T^2}\mathbf{G}(\mathbf{I}-\mathbf{A}')^{-1}\mathbf{C}_0^{-1}\Big[(m+1)\mathbf{I} + \mathbf{C}_0\mathbf{A}'\mathbf{A}'\mathbf{C}_0^{-1}\left(\mathbf{I}-\mathbf{C}_0\mathbf{A}'\mathbf{A}'\mathbf{C}_0^{-1}\right)^{-1} \\ + \mathbf{A}^2\left(\mathbf{I}-\mathbf{A}^2\right)^{-1} + \sum_{i=1}^{mp}\rho_i\left(\mathbf{A}\left(\mathbf{I}-\rho_i\mathbf{A}\right)^{-1} + \mathbf{C}_0\mathbf{A}'\mathbf{C}_0^{-1}\left(\mathbf{I}-\rho_i\mathbf{C}_0\mathbf{A}'\mathbf{C}_0^{-1}\right)^{-1}\right)\Big] \tag{3.2}$$

It should however be kept in mind that this approximation does not include all second-order terms but presumably the two leading terms. Including all second-order terms would mean having to evaluate each of the numerous $O(T^{-2})$ terms that arise as residual terms at various stages in Nicholls and Pope's analysis as well as possible second-order terms attributable to terms succeeding $\mathbf{PQ}^2 - \mathbf{AQ}^3$.

4. Performance of second-order vs. first-order approximation

In this section the performance of (3.2) (meaning (2.4) with second-order terms) will be compared with the performance of (2.3) and (2.4) by means of Monte Carlo simulation. Data were generated by two bivariate versions of (2.1); a first-order (m=2, p=1) and a second-order (m=p=2) model, both of which without constants ($\mu=\mathbf{0}$). For reasons of clarity, the parameters will be indexed a_1 through a_4 for the first-order system, thus $\mathbf{A} = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}$, but a_{11}

through a_{24} in the second-order system, thus $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$ in that case.

Eight eigenvalues, ρ={1.0, 0.9, 0.5, 0.3, 0, –0.3, –0.5, –0.9}, were used, leading to 64 eigenvalue combinations for the first-order and 1296 * combinations for the second-order system to be simulated for each of the four sample sizes T={25, 50, 100, 200}. The number of replications was 10,000. The parameter matrix **A** was OLS estimated (as was μ) after each replication, and after all 10,000 replications average estimates and their variances were computed. The resulting bias was then regressed on the corresponding approximate value provided by inserting the true or estimated values of **A**, **G**, $\mathbf{C}_0$, ρ and T into (2.3), (2.4) or (3.2), the underlying idea being that for an approximation to work well, not only should it be strongly positively correlated with the observed bias, but the intercept in the regression should also be zero and the slope should be one. Table 1 below holds estimated parameters and squared correlations from such regressions for the first-order approximation (2.3 or 2.4, since they are equivalent here) and for the second-order approximation (3.2) based on true parameters for 49 stationary cases, i.e. excluding (ρ_1 =1.0 $\cup$ ρ_2 =1.0).

* Of which 882 were actually simulated. For a detailed account of how the Monte Carlo experiments were designed, see Brännström and Karlsson (1993) or Brännström (1994*a*).

		(2.3 / 2.4)			(3.2)		
Parameter	Sample size	a	b	R^2	a	b	R^2
	25	–0.0069 (–4.04)	0.9683 (–1.80)	0.985	–0.0027 (–2.95)	0.9641 (–3.95)	0.996
a_1	50	–0.0019 (–2.56)	1.0090 (0.59)	0.989	–0.0008 (–1.43)	1.0077 (0.69)	0.994
	100	–0.0004 (–1.18)	1.0174 (1.27)	0.992	–0.0001 (–0.31)	1.0178 (1.45)	0.993
	200	–0.0002 (–1.23)	1.0186 (1.20)	0.989	–0.0001 (–0.78)	1.0183 (1.17)	0.989
	25	0.0002 (0.38)	0.9580 (–1.87)	0.975	0.0002 (0.75)	0.9616 (–3.44)	0.994
a_2	50	0.0000 (0.07)	1.0023 (0.13)	0.986	0.0000 (0.10)	1.0044 (0.35)	0.993
	100	0.0000 (0.24)	1.0115 (0.69)	0.987	0.0000 (0.27)	1.0124 (0.82)	0.990
	200	0.0000 (0.66)	1.0157 (0.91)	0.987	0.0000 (0.67)	1.0157 (0.91)	0.987
	25	–0.0008 (–0.38)	0.9580 (–1.87)	0.975	–0.0003 (–0.21)	0.9414 (–4.12)	0.989
a_3	50	–0.0001 (–0.07)	1.0023 (0.13)	0.986	–0.0001 (–0.11)	1.0007 (0.07)	0.994
	100	–0.0001 (–0.24)	1.0115 (0.69)	0.987	–0.0001 (–0.28)	1.0118 (0.82)	0.991
	200	–0.0001 (–0.59)	1.0159 (0.93)	0.987	–0.0001 (–0.59)	1.0156 (0.90)	0.987
	25	–0.0062 (–3.57)	0.9610 (–2.18)	0.984	–0.0015 (–1.52)	0.9674 (–3.24)	0.995
a_4	50	–0.0020 (–2.82)	1.0008 (0.06)	0.990	–0.0008 (–1.49)	1.0049 (0.46)	0.995
	100	–0.0004 (–1.07)	1.0081 (0.58)	0.991	–0.0001 (–0.27)	1.0095 (0.76)	0.993
	200	–0.0001 (–0.76)	1.0055 (0.40)	0.991	–0.0001 (–0.32)	1.0056 (0.42)	0.992

Table 1: Fitted OLS regressions of the type $y = \alpha + \beta x + \varepsilon$, where y = observed bias, x = approximation to the bias (as defined by (2.3), (2.4) or (3.2), based on true parameters), and t ratios are reported in brackets (tests of zero intercept and unit slope, respectively). Based on 49 stationary eigenvalue combinations, each simulated 10,000 times.

It is clear from Table 1 that the effects of adding second-order terms to (2.3) or (2.4) (remember that they are equivalent in this case since p=1) are minimal in most cases, no doubt due to the fact that the first-order approximation already performs so well in this case. But (3.2) is more strongly correlated with the observed bias than the first-order approximation is. This effect is quite substantial for the smallest samples but fades out rapidly since it is second-order only. As a somewhat discomforting result, sample correlations actually decline as the sample size increases for (3.2), whereas they increase with increasing sample size, a more intuitively appealing result, for (2.3) and (2.4). Also the estimated slopes are significantly (at the 5 % significance level) smaller than unity for a sample size of 25 in all four cases for (3.2), but only in one case ($\hat{a}_4$) for the first-order approximations.

So judging by Table 1, the efforts of deriving (3.2) are poorly rewarded in terms of fit, but let us recall for a moment that the underlying objective of deriving these second-order terms was to approximate the bias more accurately when the first-order approximation performs less satisfactorily, notably for small sample sizes and for eigenvalues close to unity.

		(2.3 / 2.4)			(3.2)		
Parameter	Sample size	*a*	*b*	R^2	*a*	*b*	R^2
	25	–0.0084 (–4.78)	1.1328 (6.18)	0.983	–0.0054 (–2.39)	1.0262 (1.05)	0.973
a_1	50	–0.0021 (–2.89)	1.0919 (5.55)	0.991	–0.0013 (–1.83)	1.0326 (2.23)	0.991
	100	–0.0004 (–1.27)	1.0582 (4.03)	0.991	–0.0002 (–0.59)	1.0161 (1.33)	0.993
	200	–0.0002 (–1.28)	1.0385 (2.40)	0.989	0.0003 (1.55)	1.0127 (0.82)	0.989
	25	0.0002 (0.41)	1.0936 (3.94)	0.978	0.0008 (1.39)	0.8703 (–5.07)	0.961
a_2	50	0.0000 (0.07)	1.0730 (4.20)	0.988	0.0003 (1.11)	0.9663 (–1.21)	0.962
	100	0.0000 (0.25)	1.0463 (2.69)	0.987	0.0001 (0.59)	0.9765 (–0.77)	0.956
	200	0.0000 (0.68)	1.0328 (1.85)	0.986	–0.0000 (–0.39)	0.9585 (–1.41)	0.959
	25	–0.0008 (–0.41)	1.0936 (3.94)	0.978	–0.0042 (–2.35)	0.8876 (–6.67)	0.983
a_3	50	–0.0001 (–0.07)	1.0730 (4.20)	0.988	–0.0003 (–0.54)	0.9857 (–1.35)	0.995
	100	–0.0001 (–0.25)	1.0463 (2.69)	0.987	–0.0001 (–0.29)	0.9950 (–0.36)	0.991
	200	–0.0002 (–0.79)	1.0316 (1.74)	0.986	–0.0002 (–0.78)	1.0040 (0.22)	0.986
	25	–0.0078 (–4.43)	1.1222 (5.73)	0.983	–0.0038 (–2.64)	0.9988 (–0.08)	0.989
a_4	50	–0.0020 (–3.29)	1.0834 (5.52)	0.991	–0.0010 (–2.14)	1.0180 (1.92)	0.996
	100	–0.0004 (–1.17)	1.0481 (3.29)	0.991	–0.0000 (–0.14)	1.0054 (0.48)	0.994
	200	–0.0001 (–0.79)	1.0251 (1.77)	0.991	–0.0000 (–0.26)	1.0019 (0.14)	0.992

Table 2: Fitted OLS regressions of the type $y = \alpha + \beta x + \varepsilon$, where y = observed bias, x = approximation to the bias (as defined by (2.3), (2.4) or (3.2), based on estimates), and t ratios are reported in brackets (tests of zero intercept and unit slope, respectively). Based on 49 stationary eigenvalue combinations, each simulated 10,000 times.

Table 2 holds the corresponding results for the bias approximations based on estimates in the same 49 stationary cases. It demonstrates that using estimates instead of true parameters has a very adverse effect on the performance of the first-order approximation, whose estimated slopes significantly (at the 5 % significance level) exceed unity for all sample sizes smaller than 200, signalling that (2.3) and (2.4) understate the bias when based on estimates. In

Table 1, only one of the estimated slopes is significantly greater than unity ($\hat{a}_4$ when T=25). Estimated intercepts and correlations are not affected much by the estimation of $\mathbf{A}$, $\mathbf{G}$, $\mathbf{C}_0$ and ρ though.

Turning to the right-hand half of Table 2, the second-order approximation appears to be able to handle the bias better. Correlations are slightly lower than in Table 1 and some of the previously insignificant intercept estimates turn out to be significant in Table 2 but more importantly, estimated slopes have not deteriorated the way they do for (2.3) and (2.4). In fact, in terms of inference they are qualitatively unchanged; (3.2) significantly overstates the bias in a sample of 25 observations (except for $\hat{a}_1$) but not in samples of 50 observations or more, based on parameters as well as based on estimates.

Let us now focus on the 15 eigenvalue combinations involving one or two unit roots, which were not included in Tables 1 and 2. In fact, all three approximations are intended for stationary combinations only, but here they will be evaluated for unit-root combinations as well. The 15 unit-root cases are of special interest here because with this particular experiment design, one unit root in the bivariate first-order system corresponds to the two variables in $\mathbf{x}_t$ being first-order cointegrated, and two unit roots correspond to them being $I(1)$ but not cointegrated.

It is however impossible to insert unit roots into the formulæ since that will make ($\mathbf{I}$–$\mathbf{A}$) noninvertible, causing all three approximations to break down. Thus it is not possible to evaluate the approximations in the presence of unit roots by regressing bias on approximations based on true parameters as in Table 1. However, basing the results on estimates is possible for unit-root combinations as well, at least as long as no eigenvalue is estimated exactly at unity. Therefore Table 3 below, which holds regression results for the same 49 stationary eigenvalue combinations as the two previous tables plus 15 combinations involving one or two unit roots, should be related to Table 2 rather than Table 1.

It should however be pointed out that (3.2) appears to be highly sensitive to eigenvalues estimated close to unity. As a consequence it occasionally produces values which are clearly nonsensical while at the same time the first-order approximations yield realistic results. Such nonsensical results (a total 12 out of 4·15=60 unit-root combinations) are of no use and have been discarded from the table below, not only for (3.2) but for the first-order approximations as well for reasons of comparison. Therefore, instead of 64 eigenvalue combinations (49 stationary, 14 $CI(1,1)$, and one $I(1)$ but not cointegrated), the results in Table 3 are based on around 60 combinations for each sample size.

		(2.3 / 2.4)			(3.2)		
Parameter	Sample size	a	b	R^2	a	b	R^2
a_1	25	–0.0117	1.2564	0.953	–0.0060	1.0362	0.970
		(–3.58)	(6.91)		(–2.25)	(1.54)	
	50	–0.0026	1.2000	0.963	–0.0016	1.0666	0.981
		(–1.76)	(6.40)		(–1.45)	(3.41)	
	100	–0.0009	1.1662	0.960	–0.0004	1.0504	0.984
		(–1.13)	(5.24)		(–0.70)	(2.90)	
	200	0.0001	1.1151	0.967	0.0002	1.0335	0.989
		(0.23)	(4.04)		(1.15)	(2.29)	
a_2	25	0.0004	1.1425	0.956	–0.0010	0.8505	0.954
		(0.73)	(4.34)		(–0.90)	(–3.29)	
	50	0.0000	1.1448	0.948	0.0004	0.9901	0.963
		(0.02)	(4.05)		(1.22)	(–0.39)	
	100	0.0001	1.1044	0.934	0.0000	0.9872	0.935
		(0.34)	(2.68)		(0.01)	(–0.38)	
	200	–0.0000	1.1093	0.917	–0.0000	0.9953	0.931
		(–0.01)	(2.06)		(–0.80)	(–0.13)	
a_3	25	–0.0023	1.1533	0.970	–0.0037	0.9304	0.983
		(–0.93)	(5.66)		(–1.77)	(–4.14)	
	50	0.0009	1.1552	0.971	–0.0009	1.0023	0.988
		(0.65)	(5.82)		(–1.01)	(0.16)	
	100	–0.0002	1.1157	0.957	–0.0000	1.0159	0.981
		(–0.31)	(3.70)		(–0.02)	(0.84)	
	200	0.0001	1.1057	0.959	–0.0002	1.0168	0.990
		(0.28)	(3.35)		(–1.02)	(1.21)	
a_4	25	–0.0089	1.2099	0.971	–0.0043	0.9958	0.984
		(–3.41)	(7.71)		(–2.16)	(–0.25)	
	50	–0.0020	1.1723	0.975	–0.0009	1.0455	0.992
		(–1.62)	(6.93)		(–1.25)	(3.68)	
	100	–0.0007	1.1601	0.969	–0.0001	1.0342	0.989
		(–0.92)	(5.83)		(–0.25)	(2.34)	
	200	–0.0002	1.0899	0.971	–0.0001	1.0164	0.991
		(–0.54)	(3.55)		(–0.40)	(1.26)	

Table 3: Fitted OLS regressions of the type $y = \alpha + \beta x + \varepsilon$, where y = observed bias, x = approximation to the bias (as defined by (2.3), (2.4) or (3.2), based on estimates), and t ratios are reported in brackets (tests of zero intercept and unit slope, respectively). Based on around 60 eigenvalue combinations, each simulated 10,000 times.

The table clearly illustrates that the introduction of unit-root combination makes the performance of all three approximations considerably less impressive. Correlations are much lower than before, especially for $\hat{a}_2$, and estimated slopes are even more in excess of unity than before, this time even for T=200. The fact that intercept estimates are still mainly insignificant suggests that it is basically a scale problem which could be solved by inflating the approximations based on estimates by some unknown factor. As for the relative performance of the first- and second-order approximations, (3.2) again

	Sample	(2.3)			(2.4)			(3.2)		
Parameter	size	a	b	R^2	a	b	R^2	a	b	R^2
	25	–.0010 (–.83)	1.3998 (22.3)	.886	–.0017 (–1.57)	1.3798 (22.7)	.896	.0027 (2.54)	1.2092 (15.0)	.905
a_{11}	50	–.0003 (–.77)	1.2489 (21.5)	.937	–.0005 (–1.33)	1.2387 (21.3)	.940	.0008 (2.32)	1.1662 (17.6)	.951
	100	–.0001 (–1.17)	1.1309 (16.8)	.964	–.0002 (–1.57)	1.1259 (16.4)	.965	.0001 (.98)	1.0934 (13.8)	.971
	200	–.0000 (–.40)	1.0684 (8.71)	.959	–.0000 (–.59)	1.0659 (8.42)	.959	.0000 (.57)	1.0501 (6.67)	.961
	25	–.0002 (–.79)	1.0944 (2.07)	.424	–.0002 (–.83)	1.1365 (3.23)	.478	.0001 (.50)	.8943 (–3.87)	.642
a_{12}	50	–.0001 (–.62)	1.1051 (3.26)	.599	–.0001 (–.63)	1.1088 (3.50)	.618	.0000 (.11)	.9895 (–.43)	.733
	100	–.0000 (–.26)	1.0628 (1.94)	.579	–.0000 (–.25)	1.0604 (1.89)	.584	.0000 (.36)	1.0017 (.06)	.670
	200	–.0001 (–1.68)	1.0158 (.38)	.432	–.0001 (–1.68)	1.0131 (.32)	.433	–.0000 (–.60)	.9736 (–.65)	.490
	25	.0252 (28.1)	1.1664 (23.1)	.971	.0243 (27.2)	1.1572 (22.0)	.971	.0262 (22.5)	1.0699 (8.16)	.952
a_{13}	50	.0075 (23.7)	1.1135 (22.6)	.984	.0073 (23.1)	1.1092 (21.8)	.984	.0080 (23.5)	1.0720 (13.7)	.982
	100	.0016 (14.0)	1.0534 (14.4)	.990	.0016 (13.5)	1.0514 (13.9)	.990	.0018 (15.5)	1.0349 (9.67)	.991
	200	.0003 (4.54)	1.0186 (4.28)	.986	.0003 (4.37)	1.0176 (4.06)	.986	.0004 (5.21)	1.0098 (2.28)	.986
	25	.0000 (.21)	1.1820 (22.6)	.965	.0000 (.21)	1.1728 (21.6)	.965	.0008 (3.23)	1.0879 (9.53)	.947
a_{14}	50	–.0001 (–.70)	1.1170 (16.0)	.967	–.0001 (–.70)	1.1128 (15.5)	.967	.0001 (1.31)	1.0783 (10.8)	.966
	100	.0000 (.07)	1.0473 (5.95)	.957	.0000 (.07)	1.0454 (5.72)	.957	.0000 (.98)	1.0306 (3.96)	.958
	200	–.0000 (–.72)	1.0102 (.99)	.924	–.0000 (–.72)	1.0093 (.90)	.924	–.0000 (–.38)	1.0024 (.24)	.924
	25	.0005 (.38)	1.0877 (2.08)	.457	.0005 (.39)	1.1293 (3.31)	.516	–.0025 (–2.57)	.8852 (–5.11)	.664
a_{21}	50	.0002 (.41)	1.1015 (3.86)	.691	.0002 (.43)	1.1064 (4.26)	.714	–.0007 (–2.16)	.9828 (–1.00)	.806
	100	.0000 (.14)	1.0782 (4.35)	.821	.0000 (.14)	1.0761 (4.35)	.828	–.0002 (–1.70)	1.0122 (.87)	.868
	200	.0000 (.03)	1.0321 (1.72)	.796	.0000 (.03)	1.0299 (1.61)	.797	–.0001 (–.75)	.9958 (–.24)	.809
	25	–.0009 (–.80)	1.4134 (23.0)	.887	–.0017 (–1.56)	1.3931 (23.4)	.898	.0063 (5.58)	1.1931 (13.7)	.902
a_{22}	50	–.0003 (–.93)	1.2569 (22.5)	.939	–.0005 (–1.51)	1.2465 (22.4)	.942	.0017 (5.20)	1.1653 (18.0)	.954
	100	–.0001 (–1.11)	1.1370 (18.6)	.968	–.0002 (–1.54)	1.1318 (18.2)	.969	.0004 (3.50)	1.0963 (15.3)	.975
	200	–.0001 (–2.46)	1.0619 (8.45)	.964	–.0002 (–2.67)	1.0594 (8.15)	.964	–.0000 (–.44)	1.0425 (6.08)	.966
	25	–.0002 (–0.26)	1.1887 (28.7)	.976	–.0002 (–.26)	1.1794 (27.4)	.976	–.0040 (–3.64)	1.0935 (11.3)	.957
a_{23}	50	.0001 (.44)	1.1227 (25.9)	.986	.0001 (.44)	1.1184 (25.1)	.986	–.0008 (–2.54)	1.0833 (16.8)	.984
	100	–.0000 (–.08)	1.0601 (16.1)	.990	–.0000 (–.08)	1.0581 (15.6)	.990	–.0002 (–2.07)	1.0428 (12.0)	.991
	200	.0000 (.70)	1.0223 (5.19)	.986	.0000 (.70)	1.0213 (4.97)	.986	–.0000 (–.12)	1.0141 (3.36)	.987
	25	.0258 (26.9)	1.1710 (22.2)	.967	.0248 (26.5)	1.1618 (21.2)	.967	.0285 (19.7)	1.0478 (4.61)	.929
a_{24}	50	.0073 (22.9)	1.1136 (22.4)	.984	.0070 (22.4)	1.1093 (21.7)	.984	.0085 (21.8)	1.0632 (10.7)	.976
	100	.0016 (13.7)	1.0550 (14.3)	.990	.0016 (13.3)	1.0530 (13.8)	.990	.0020 (15.9)	1.0333 (8.52)	.989
	200	.0003 (4.29)	1.0218 (4.97)	.986	.0003 (4.11)	1.0209 (4.76)	.986	.0004 (5.63)	1.0117 (2.69)	.986

Table 4: Fitted OLS regressions of the type $y = \alpha + \beta x + \varepsilon$, where y = observed bias, x = approximation to the bias (based on true parameters), and t ratios are reported in brackets. Based on 784 stationary eigenvalue combinations.

appears to approximate the bias better than the first-order approximations when based on estimates. Although it significantly understates the bias of $\hat{a}_1$ and $\hat{a}_4$ in most cases and overstates the bias of $\hat{a}_2$ and $\hat{a}_3$ in a sample of 25, it does so less (in absolute terms as well as relative terms) than (2.3) and (2.4). Furthermore, (3.2) correlates more strongly with bias.

Tables 4, 5 and 6 hold the corresponding results for bivariate second-order models (m=p=2). Increasing the order p to 2 doubles the number of parameters in $\mathbf{A}$ to eight and increases the number of unique eigenvalue combinations to 784 stationary and 512 unit-root combinations. Every stationary combination was simulated, but only 98 of the 512 unit-root combinations.

The results of regressing the bias on the three approximations based on true parameter values can be found in Table 4 for stationary eigenvalue combinations. These results are the counterparts of the results in Table 1, and even though different sets of data have been generated, it should be safe to compare the results between tables since the number of replicates is so large. The general impression after such a comparison is that all three approximations perform markedly worse in the second-order than in the first-order case. Slope estimates significantly in excess of unity (except for $\hat{a}_{12}$ and $\hat{a}_{21}$, but that is probably due to the very poor fit) signal that all three approximations severely understate the bias of $\hat{\mathbf{A}}$ even when based on true parameters. Squared correlations are lower than in Table 1 (in the case of $\hat{a}_{12}$ and $\hat{a}_{21}$ much lower) but still generally high. Furthermore, for the second-order cross-term estimates $\hat{a}_{13}$ and $\hat{a}_{24}$ the estimated intercepts are highly significant for all sample sizes, which was never the case for the first-order system. The implications of using (2.3), (2.4) or (3.2) to approximate the bias of these two estimates is that they will show no bias when in fact $\hat{a}_{13}$ and $\hat{a}_{24}$ are biased, and vice versa.

As far as the relative performance of the three approximations is concerned, the two first-order approximations perform almost identically, which is not surprising considering that they differ only by $(T-1)^{-1}\mathbf{A}$ when p is 2. Nevertheless, (2.4) appears to correlate more strongly with the bias, in particular for the smaller sample sizes, and in addition its estimated slopes are often closer (in absolute terms) to unity than in the case of (2.3). The second-order approximation, on the other hand, is easier to rank since correlations generally are about as high or higher than for (2.3) and (2.4) and slope

Parameter	Sample size	(2.3) a	(2.3) b	(2.3) R^2	(2.4) a	(2.4) b	(2.4) R^2	(3.2) a	(3.2) b	(3.2) R^2
a_{11}	25	–.0005 (–.44)	1.8254 (38.2)	.913	–.0024 (–2.85)	1.7455 (44.2)	.936	.0039 (4.29)	1.5217 (34.0)	.931
	50	–.0002 (–.40)	1.4119 (30.6)	.942	–.0006 (–1.80)	1.3875 (34.2)	.952	.0012 (4.06)	1.2932 (31.3)	.962
	100	–.0001 (–.55)	1.2032 (22.2)	.962	–.0001 (–1.19)	1.1950 (24.9)	.968	.0003 (2.93)	1.1516 (22.7)	.975
	200	.0000 (.02)	1.1049 (11.1)	.959	.0000 (.04)	1.1045 (11.1)	.958	.0001 (1.82)	1.0836 (10.6)	.960
a_{12}	25	–.0006 (–2.40)	1.6468 (12.5)	.592	–.0004 (–2.05)	1.6748 (16.1)	.689	–.0002 (–.96)	1.0226 (.75)	.618
	50	–.0002 (–1.91)	1.2846 (7.55)	.625	.0000 (.27)	1.2958 (8.62)	.656	.0001 (.80)	1.0646 (2.44)	.684
	100	–.0001 (–1.31)	1.1443 (3.78)	.564	–.0000 (–.33)	1.1633 (4.63)	.583	–.0000 (–.39)	1.0605 (1.93)	.596
	200	–.0001 (–1.85)	.9617 (–.81)	.375	–.0000 (–1.48)	1.0788 (1.77)	.429	–.0000 (–1.48)	1.0350 (.83)	.434
a_{13}	25	.0254 (18.7)	1.6272 (41.6)	.945	.0219 (16.6)	1.6026 (39.9)	.939	.0244 (14.7)	1.4794 (27.4)	.907
	50	.0069 (17.3)	1.3080 (41.5)	.979	.0067 (15.2)	1.3143 (37.5)	.970	.0078 (16.1)	1.2672 (30.1)	.964
	100	.0014 (10.8)	1.1414 (30.9)	.989	.0016 (9.78)	1.1603 (27.0)	.980	.0020 (11.6)	1.1392 (23.4)	.979
	200	.0003 (3.85)	1.0607 (12.4)	.986	.0004 (4.49)	1.0802 (12.6)	.974	.0005 (5.47)	1.0698 (11.1)	.974
a_{14}	25	.0003 (1.05)	1.6522 (38.4)	.935	.0001 (.35)	1.5714 (38.0)	.937	.0003 (.76)	1.3789 (20.3)	.881
	50	–.0000 (–.14)	1.3146 (31.7)	.964	–.0001 (–1.60)	1.2809 (30.2)	.962	–.0001 (–1.23)	1.2072 (20.6)	.950
	100	–.0000 (–.05)	1.1386 (14.9)	.958	–.0000 (–.84)	1.1342 (15.3)	.956	–.0000 (–.87)	1.1134 (13.0)	.955
	200	–.0000 (–1.24)	1.0524 (4.59)	.928	–.0001 (–2.33)	1.0583 (5.38)	.924	–.0001 (–2.32)	1.0499 (4.64)	.924
a_{21}	25	.0012 (1.21)	1.6204 (13.2)	.625	.0003 (.34)	1.6843 (19.6)	.766	.0001 (.08)	1.2864 (12.8)	.821
	50	.0004 (1.25)	1.2746 (9.78)	.743	–.0002 (–.79)	1.3002 (12.5)	.798	–.0004 (–1.52)	1.1366 (8.63)	.874
	100	.0001 (.86)	1.1513 (7.88)	.834	–.0000 (–.21)	1.1687 (9.56)	.850	–.0000 (–.28)	1.0873 (6.23)	.886
	200	.0000 (.49)	1.0702 (3.51)	.800	.0000 (.14)	1.0714 (3.79)	.804	.0000 (.16)	1.0301 (1.73)	.816
a_{22}	25	–.0016 (–1.33)	1.8551 (37.3)	.908	–.0029 (–3.51)	1.7335 (43.1)	.934	.0042 (4.94)	1.5148 (35.7)	.938
	50	–.0007 (–1.59)	1.4225 (30.4)	.941	–.0008 (–2.57)	1.3649 (33.5)	.954	.0011 (4.20)	1.2749 (33.5)	.970
	100	–.0002 (–1.52)	1.2065 (23.6)	.966	–.0003 (–2.29)	1.1858 (24.5)	.969	.0002 (1.87)	1.1438 (22.4)	.967
	200	–.0002 (–2.46)	1.0702 (3.51)	.960	–.0002 (–2.78)	1.0852 (11.1)	.962	–.0001 (–1.00)	1.0642 (8.79)	.964
a_{23}	25	–.0017 (–1.27)	1.6437 (41.6)	.944	–.0002 (–.26)	1.5841 (42.6)	.948	–.0004 (–.29)	1.4305 (26.6)	.914
	50	–.0002 (–.50)	1.3153 (43.3)	.980	.0001 (.44)	1.2966 (45.1)	.981	.0004 (1.03)	1.2478 (33.1)	.973
	100	–.0000 (–.13)	1.1482 (23.5)	.990	–.0000 (–.08)	1.1420 (23.3)	.989	.0001 (.63)	1.1235 (20.5)	.988
	200	.0001 (1.47)	1.0646 (13.6)	.987	.0000 (.70)	1.0627 (13.6)	.986	.0001 (1.92)	1.0542 (12.1)	.986
a_{24}	25	.0279 (19.3)	1.6396 (40.3)	.939	.0227 (17.9)	1.5821 (41.0)	.944	.0168 (10.8)	1.3583 (23.2)	.912
	50	.0072 (7.90)	1.3120 (41.7)	.978	.0054 (15.3)	1.2776 (41.9)	.980	.0060 (15.7)	1.2230 (32.5)	.977
	100	.0016 (11.6)	1.1458 (31.4)	.989	.0013 (10.4)	1.1372 (31.5)	.989	.0016 (12.4)	1.1147 (26.4)	.988
	200	.0003 (3.90)	1.0671 (14.1)	.986	.0002 (3.14)	1.0652 (14.7)	.987	.0003 (4.57)	1.0555 (12.6)	.986

Table 5: Fitted OLS regressions of the type $y = \alpha + \beta x + \varepsilon$, where y = observed bias, x = approximation to the bias (based on estimates), and t ratios are reported in brackets. Based on 784 stationary eigenvalue combinations.

estimates are closer to unity in most cases, in absolute as well as relative terms. In a number of cases however, estimated intercepts are significant for (3.2) but not for (2.3) or (2.4), causing the same sort of confusion as over $\hat{a}_{13}$ and $\hat{a}_{24}$ for all three approximations. It is nonetheless safe to say that (3.2) outperforms (2.3) and (2.4) in correlations as well as in the unit slope-zero intercept respect. Its performance is particularly impressive in the case of the first-order cross terms $\hat{a}_{12}$ and $\hat{a}_{21}$, where the first-order approximations fail completely. Not only does the second-order approximation correlate much more strongly with the bias (albeit not at all like for the remaining six estimates), slopes are also insignificantly close to unity for T=50 or more.

Not surprisingly, using estimates instead of true parameter values makes the results look even worse. Most slope estimates are much greater in Table 5 than in Table 4, which means that the three approximations understate the bias even more when based on estimates than when based on true parameters, just like in the first-order case. Intercept estimates are basically unaffected by the move from true to estimated parameters, and squared correlations actually increase in several cases (notably for the smallest sample sizes). Just like in Table 4, the two first-order approximations perform more or less equally well, the second-order approximation considerably better.

Finally, Table 6 below illustrates the effects of adding some unit-root combinations to the 784 stationary combinations in Table 5. Owing to the design of the Monte Carlo experiment, one unit root in the bivariate second-order model corresponds to the two variables in $\mathbf{x}_t$ being $CI(1,1)$; if there are two unit roots the variables can be either $I(1)$ without cointegration or $CI(2,2)$, depending on the allocation of the two unit roots. Three unit roots correspond to the two variables being $CI(2,1)$, and four unit roots to them being $I(2)$ without cointegrating vectors.

Just like in the first-order case (cf. Table 3), eigenvalue estimates exactly at unity cause all three approximations to break down; hence there can be no results based on true parameters for unit-root combinations. Furthermore, eigenvalues estimated close to unity appear to cause (3.2) to produce highly implausible values while at the same time the two first-order approximations do not. Whenever this occurred (around 30 times out of 98 for each sample size), the results were discarded not only for (3.2) but for (2.3) and (2.4) as well in order to enhance comparison between approximations. Therefore Table

6 is based on the same 784 eigenvalue combinations as in Table 5 plus the unit-root combinations for which all three approximations produced reasonable results (normally 50 to 60 additional cases).

Not surprisingly, adding unit-root cases causes the results in Table 5 to deteriorate. The three approximations understate the bias even more for unit-root cases than for stationary cases, making the estimated slopes even greater. However, just like in the first-order case this is merely a scale effect; the approximations work in the right direction but approximate only part of the bias. Also parallel to the first-order case, correlations are generally lower than in Table 5; with the exception of $\hat{a}_{12}$ and $\hat{a}_{21}$ most of them are reduced by around 0.10.

Once again the second-order approximation compares very favourably to the two first-order approximations. Correlations are generally higher and slope estimates closer to unity for (3.2) than for (2.3) and (2.4).

Parameter	Sample size	(2.3) a	(2.3) b	(2.3) R^2	(2.4) a	(2.4) b	(2.4) R^2	(3.2) a	(3.2) b	(3.2) R^2
a_{11}	25	–.0009	1.9407	.856	–.0023	1.9022	.863	.0035	1.5737	.887
		(–.61)	(33.2)		(–1.62)	(33.4)		(2.67)	(28.7)	
	50	–.0003	1.4962	.869	–.0006	1.4821	.871	.0014	1.3547	.901
		(–.54)	(24.2)		(–.98)	(24.0)		(2.51)	(22.3)	
	100	–.0001	1.3034	.833	–.0002	1.2969	.833	.0004	1.2206	.897
		(–.35)	(15.0)		(–.55)	(14.7)		(1.62)	(15.3)	
	200	.0000	1.1901	.823	.0000	1.1927	.833	–.0002	1.0544	.857
		(.20)	(9.90)		(.29)	(10.4)		(–1.69)	(3.62)	
a_{12}	25	–.0007	1.7155	.455	–.0006	1.7206	.494	–.0002	1.0166	.585
		(–2.12)	(10.6)		(–2.07)	(11.5)		(–.77)	(.54)	
	50	–.0002	1.2386	.395	–.0002	1.2382	.408	–.0001	.9831	.553
		(–1.13)	(4.39)		(–1.15)	(4.52)		(–.45)	(–.54)	
	100	–.0001	1.1249	.364	–.0001	1.1241	.369	–.0000	.9634	.511
		(–.70)	(2.41)		(–.68)	(2.42)		(–.70)	(–1.11)	
	200	–.0001	.9364	.250	–.0001	.9430	.254	–.0000	.9596	.438
		(–2.08)	(–1.13)		(–2.09)	(–1.01)		(–1.28)	(–1.19)	
a_{13}	25	.0254	1.6272	.945	.0301	1.7387	.873	.0268	1.5050	.877
		(18.7)	(41.6)		(14.2)	(31.4)		(13.1)	(25.2)	
	50	.0069	1.3080	.979	.0116	1.4534	.903	.0112	1.3390	.912
		(17.3)	(41.5)		(12.9)	(27.2)		(13.1)	(23.3)	
	100	.0014	1.1414	.989	.0045	1.2936	.892	.0030	1.1795	.940
		(10.8)	(30.9)		(9.89)	(18.8)		(9.22)	(17.3)	
	200	.0003	1.0607	.986	.0012	1.1619	.914	.0009	1.1080	.945
		(3.85)	(12.4)		(6.22)	(13.0)		(6.11)	(11.6)	
a_{14}	25	.0010	1.7911	.870	.0010	1.7763	.871	.0011	1.3976	.850
		(2.25)	(32.4)		(2.15)	(32.1)		(2.22)	(19.1)	
	50	.0000	1.4520	.906	.0000	1.4460	.906	–.0001	1.2242	.904
		(.01)	(27.5)		(.02)	(27.4)		(–.49)	(16.0)	
	100	–.0000	1.2581	.916	–.0001	1.2607	.910	–.0001	1.1139	.914
		(–.62)	(19.5)		(–.77)	(18.9)		(–1.51)	(9.61)	
	200	–.0000	1.1470	.897	–.0000	1.1493	.900	–.0001	1.0446	.885
		(–.74)	(10.9)		(–.96)	(11.3)		(–1.64)	(3.41)	
a_{21}	25	.0023	1.6556	.436	.0021	1.6508	.477	.0011	1.3156	.684
		(1.47)	(9.66)		(1.44)	(10.4)		(.96)	(9.78)	
	50	–.0001	1.3180	.547	–.0001	1.3152	.564	–.0006	1.0994	.765
		(–.20)	(7.43)		(–.20)	(7.64)		(–1.46)	(4.57)	
	100	–.0001	1.0839	.439	–.0001	1.0824	.444	–.0002	1.1111	.751
		(–.19)	(1.96)		(–.20)	(1.95)		(–.97)	(4.98)	
	200	–.0001	1.0400	.612	–.0001	1.0441	.620	–.0000	1.0032	.751
		(–.74)	(1.38)		(–.70)	(1.54)		(–.29)	(.16)	
a_{22}	25	–.0023	1.9462	.816	–.0035	1.9084	.825	.0003	1.4661	.875
		(–1.34)	(28.8)		(–2.15)	(29.0)		(.21)	(23.6)	
	50	–.0010	1.5214	.806	–.0012	1.5082	.810	–.0011	1.2046	.897
		(–1.22)	(20.0)		(–1.54)	(19.9)		(–1.91)	(14.3)	
	100	–.0004	1.2858	.787	–.0005	1.2809	.789	.0001	1.1806	.889
		(–1.17)	(12.3)		(–1.30)	(12.2)		(.26)	(12.5)	
	200	–.0002	1.1501	.830	–.0003	1.1480	.828	–.0003	1.0377	.917
		(–1.64)	(8.27)		(–1.76)	(8.14)		(–2.82)	(3.47)	
a_{23}	25	–.0054	1.8131	.880	–.0054	1.7967	.879	–.0017	1.5081	.887
		(–2.49)	(34.1)		(–2.47)	(33.6)		(–.82)	(26.5)	
	50	–.0002	1.4454	.929	–.0003	1.4395	.929	.0006	1.3404	.940
		(–.32)	(31.8)		(–.34)	(31.5)		(.85)	(28.8)	
	100	–.0004	1.2796	.926	–.0004	1.2773	.928	.0003	1.1881	.954
		(–1.01)	(22.3)		(–1.01)	(22.4)		(.96)	(20.8)	
	200	.0000	1.1468	.942	.0000	1.1495	.933	.0001	1.1081	.962
		(.00)	(14.8)		(.21)	(13.9)		(.90)	(14.2)	
a_{24}	25	.0449	1.8270	.860	.0435	1.8108	.859	.0305	1.4872	.881
		(19.0)	(31.7)		(18.5)	(31.3)		(15.0)	(25.2)	
	50	.0128	1.4302	.909	.0126	1.4247	.909	.0089	1.2705	.938
		(14.9)	(27.0)		(14.7)	(26.8)		(13.1)	(23.5)	
	100	.0050	1.2789	.915	.0049	1.2767	.915	.0042	1.2097	.948
		(12.5)	(20.7)		(12.3)	(20.5)		(13.9)	(21.4)	
	200	.0013	1.1540	.934	.0013	1.1534	.932	.0008	1.0841	.965
		(8.39)	(14.4)		(8.11)	(14.1)		(7.03)	(11.7)	

Table 6: Fitted OLS regressions of the type $y = \alpha + \beta x + \varepsilon$, where y = observed bias, x = approximation to the bias (based on estimates), and t ratios are reported in brackets. Based on around 840 eigenvalue combinations.

5. Bias-reduced estimates

The results in the previous section may not appear to invite to further use of any of the three bias approximations, at least not based on estimates since they were all found to perform worse when based on estimates than when based on true parameter values. Two important properties to keep in mind however are that they can be expected to work in the right direction and that they correlate strongly with bias (again, except for $\hat{a}_{12}$ and $\hat{a}_{21}$ of the second-order model). Using them to construct bias-reduced estimates should therefore prove fruitful, even if only part of the bias were to be reduced.

To illustrate this point, the bias before and after bias reduction will be plotted against the eigenvalues (in the case of the second-order model against eigenvalue products or eigenvalues sums) and surfaces will be smoothed using distance-weighted least squares. Since it would hardly be meaningful to speak of bias-reduced estimates when approximations are based true parameters, all results in this section are based on estimates.

Starting with the four estimates of the first-order model, the following figures show that deducting (2.3), (2.4) or (3.2) based on estimates from $\hat{\mathbf{A}}$ really does reduce its bias considerably. Note the often dramatic change of scales from the first figure on each row to the second and third figures – in most cases only a fraction of the original bias remains after bias reduction.

Another important conclusion from the bias plots is that more bias remains along the surface edges than in their interior. This effect is particularly apparent along the edges closest to the reader (i.e. in the presence of unit roots) and stronger for the first-order approximations than for (3.2), which is an encouraging result considering that one intention of the latter approximation was to improve the performance as eigenvalues tend to unity. The second-order terms really appear to handle ”near-integration” and integration better, but fail to completely approximate the bias, leaving surfaces that still bend more as eigenvalues tend to unity. In addition the second-order terms have smaller impact at the (1,1) corner, where the system is first-order integrated but not cointegrated, than when it is *CI*(1,1). There is also virtually no effect far away from the unit-root regions; in particular the remaining bias left by (2.3) and (2.4) along the two back edges is not adjusted at all.

<<< FIGURES 1 to 48 ABOUT HERE >>>

Turning to the corresponding plots for the second-order model, it should again be noted that increasing the order p to 2 introduces a difference between (2.3) and (2.4), but as the regression results in Tables 4, 5 and 6 showed, this difference is marginal for such a low-order model. Considering that there are four bias plots (original bias and three remaining bias plots) for each of the eight parameter estimates and for each of the four sample sizes, displaying all results would be unmanageable. Therefore results are given for two estimates only, $\hat{a}_{11}$ and $\hat{a}_{14}$, where $\hat{a}_{11}$ is taken to represent all four autoregressive estimates ($\hat{a}_{11}$, $\hat{a}_{13}$, $\hat{a}_{22}$ and $\hat{a}_{24}$) and $\hat{a}_{14}$ to represent all four cross-term estimates ($\hat{a}_{12}$, $\hat{a}_{14}$, $\hat{a}_{21}$ and $\hat{a}_{23}$). In both cases an alternative projection than in the first-order case must be used since there are now four eigenvalues against which to plot bias. In the case of $\hat{a}_{11}$ this problem is solved by projecting the bias against the eigenvalue sums instead of the eigenvalues, and in the case of $\hat{a}_{14}$ the horizontal scales represent eigenvalue products. Again, all results are based on estimates.

<<< FIGURES 49 to 80 ABOUT HERE >>>

As the figures above show, all three bias approximations reduce a smaller fraction of the bias in the cases of $\hat{a}_{11}$ and $\hat{a}_{14}$ than of the bias of the four estimates in the first-order model, a result which is consistent with the regression results presented in Section 4. In addition, the biases of $\hat{a}_{11}$ and of $\hat{a}_{14}$ appear to be more difficult to approximate, with sharp bias increases along the edges (where the two variables in $\mathbf{x}_t$ are second-order integrated, with or without cointegration). These are also the regions in which the approximations, in particular the first-order approximations, perform the worst. As intended, the second-order approximation performs considerably better in these regions, but still leaves sizeable remaining bias (which, as we shall see, is likely to be significant in most cases). It should be noted that it often leaves more or less as much bias in the corner where there is second-order integration but no cointegration (i.e. in the (2,2) corner in the case of $\hat{a}_{11}$ and the (1,1) corner in the case of $\hat{a}_{14}$) as the first-order approximations do.

As expected, the two first-order approximations perform almost identically, which again is consistent with the tables in Section 4. Furthermore, in the interior of the figures the performance of (3.2) is more or less the same as the performance of (2.3) and (2.4), just like in the case of the four estimates in the first-order model.

So far the results in this section appear to support the idea of bias reduction, because even if the three approximations perform less well when based on estimates, they can still be expected to reduce at least part of the bias. In addition, the second-order approximation appears to outperform the first-order approximations in situations that the latter cannot handle very well, notably in the presence of unit roots. But even the second-order approximation leaves bias unadjusted in such situations. Also, all three approximations work better when based on an estimated first-order than a second-order model.

However, before concluding this section it is important to study the relative biases after bias reduction and the effects on mean square error of bias reduction. In both cases it is vital to make sure that an absolute bias reduction is not accompanied by an increased relative bias, nor by increasing MSE. Both these effects are conceivable provided the bias reduction is not too great, because the standard error of the bias-reduced estimate may or may not exceed the original bias (depending on the correlation between estimate and approximation based on estimates). However, as the figures to follow will make clear, the bias effect dominates the standard error effect in most cases, hence bias reduction can be expected to lead to smaller *t* ratios as well as smaller mean square errors.

Starting with the relative biases, it is furthermore interesting to determine for which eigenvalue combinations the bias-reduced estimates can be expected to be insignificant and whether or not bias reduction has a qualitative impact in terms of inference, i.e. if insignificant bias obtains in the same cases as before bias.

When computing relative biases before bias reduction, each bias is simply divided by its standard error, using sample variances over the 10,000 replicates rather than the (biased) average least-squares variance estimates. When computing relative biases for the bias-reduced estimates, standard errors will differ since the variance of a bias-reduced estimate will then be the sum of the original variance and the variance of the approximation based on estimates, less the covariance between the two. Since the variance of the approximation tends to be of a smaller order of magnitude than the sample variance of the estimate and the two will often be positively correlated, standard errors will sometimes be smaller for bias-reduced estimates than for the original estimates. In case of zero or negative correlation though, standard errors will of course increase following bias reduction.

<<< FIGURES 81 to 116 ABOUT HERE >>>

As the above figures demonstrate for sample sizes 25 and 200, relative biases can be huge without bias reduction, in particular in small samples and for autoregressive estimates ($\hat{a}_1$, $\hat{a}_4$, $\hat{a}_{11}$, $\hat{a}_{13}$, $\hat{a}_{22}$ and $\hat{a}_{24}$). In fact, it appears that autoregressive parameters are likely to be estimated without significant bias (at the 5 per cent significance level, meaning values absolutely smaller than 1.96) only in a very small region where the bias plane crosses the zero plane, i.e. as one eigenvalue is around –0.3 in the case of $\hat{a}_1$ and $\hat{a}_4$ and for intermediate eigenvalue sums in the case of $\hat{a}_{11}$. Reducing the bias adds a number of interior eigenvalue combinations for which no significant bias is likely to remain, but around the four corners and in particular as eigenvalues tend to unity, highly significant bias is likely to persist even in a sample of 200 observations.

The situation looks brighter for cross-term estimates. Even before bias reduction they are insignificantly biased for a number of eigenvalue combinations along the main diagonal connecting the (1,1) with the (–0.9,–0.9) corner. The number of such combinations also increases as the sample size grows larger. After bias reduction the region of no significant bias is expanded along the main diagonal, in particular if the second-order approximation is used to reduce the bias. Significant bias is still likely to remain at the (1,1) corner in the case $\hat{a}_2$ and $\hat{a}_3$ and for the rear *CI*(2,1) and *CI*(2,2) combinations in the case of $\hat{a}_{14}$ though.

The effects of bias reduction on mean square errors can be found in the figures to follow, again only for sample sizes 25 and 200. Each figure depicts the difference between the MSE of the original estimate and the MSE of the bias-reduced estimate, so that positive graph segments indicate greater MSE before than after bias reduction, while negative segments indicate that the mean square error of the bias-reduced estimate exceeds the mean square error of the unadjusted estimate.

However as the figures demonstrate, MSE differences are mainly non-negative for all six estimates and for both sample sizes, indicating that bias reduction does not add to mean square error but may in fact reduce it. Owing to the large number of replicates, the bias component dominates MSE before as well as after bias reduction, causing the MSE difference plots to (basically) reflect bias reduction effects. Where bias reduction has little effect (i.e. for intermediate negative eigenvalues in the case of autoregressive estimates and for equal eigenvalues for cross terms) there is virtually no mean square error difference, but where bias reduction is really effective it also appears to reduce MSE. It should also be noted that the slightly higher MSE differences for (3.2) than for the first-order approximations (of which only (2.4) is displayed in the case of $\hat{a}_{11}$ and $\hat{a}_{14}$) are due to its superior performance as eigenvalues tend to unity.

<<< FIGURES 117 to 140 ABOUT HERE >>>

5. Conclusions

It appears to be possible, and is certainly desirable in several interesting cases, to improve the performance of first-order bias approximations such as (2.3) and (2.4) by adding second-order terms. In this paper two such terms have been derived and added to (2.4) to form the second-order approximation (3.2). Though probably not a complete second-order approximation, it turned out to work well where the first-order approximations fail, for instance as eigenvalues approach unity and for the smallest sample sizes. As for the simple regressions in Tables 1 and 4, adding second-order terms to the approximations based on true parameters did not improve estimates or correlations in the VAR(1) case since the fit was already excellent there, but in the VAR(2) case it raised correlations, reduced the general underadjustment and changed the results completely for the two cross-terms $\hat{a}_{12}$ and $\hat{a}_{21}$. The improved results for these two parameter estimates alone should merit the use of second-order rather than first-order bias approximations, since the performance of (2.3) and (2.4) is not satisfactory in those two cases (cf. Brännström 1994*b*).

The virtues of (3.2) appear even more evident when it is based on estimates and for eigenvalues close to unity. Even though all three approximations perform worse when based on estimates, generally understating the bias, the second-order approximation leaves less bias unadjusted, in absolute as well as relative terms. In integrated and ”near-integrated” cases in particular, it handles the very sizeable resulting bias much better than the first-order approximations. On the other hand, in the very same cases (3.2) occasionally tends to explode, making it useless for bias reduction.

Bias reduction appears to reduce relative biases as well, but leads to few qualitative changes in terms of inference; at least for the smallest sample size, basically the same eigenvalue combinations will hold insignificant biases before and after bias reduction. As the sample size increases though, more insignificant biases are added.

Bias reduction also appears to reduce mean square errors for most eigenvalue combinations but never causes MSE increases.

6. References

Brännström, T. and S. Karlsson (1993), Least squares estimates in bivariate VAR systems; Stockholm: EFI Research Paper 6512

Brännström, T. (1994*a*), Reducing the bias of OLS estimates in bivariate VAR models; Stockholm: EFI Research Paper 6533

Brännström, T. (1994*b*), A comparison between two multivariate bias approximations; Stockholm: Stockholm School of Economics Working Paper 22

Lütkepohl, H. (1993), Introduction to Multiple Time Series Analysis (2nd ed.); Berlin: Springer-Verlag

Nicholls, D.F. and A.L. Pope (1988), Bias in the estimation of multivariate autoregressions; *Australian Journal of Statistics*, **30A**, 296-309

Tjøstheim, D. and J. Paulsen (1983), Bias of some commonly-used time series estimates; *Biometrika*, **70**, 389-399