THE RIMINI CENTRE FOR ECONOMIC ANALYSIS

WP 13-08

Steve Lawford
ENAC, France,
University of Nottingham, UK,
Philips College, Cyprus.
and
The Rimini The Rimini Center for Economic Analysis, Italy

Michalis P. Stamatogiannis

"THE FINITE-SAMPLE E ECTS OF VAR DIMENSIONS ON OLS BIAS, OLS VARIANCE, AND MINIMUM MSE ESTIMATORS"

The *Rimini Centre for Economic Analysis* (RCEA) was established in March 2007. RCEA is a private, non-profit organization dedicated to independent research in Applied and Theoretical Economics and related fields. RCEA organizes seminars and workshops, sponsors a general interest journal *The Review of Economic Analysis*, and organizes a biennial conference: *Small Open Economies in the Globalized World* (SOEGW). Scientific work contributed by the RCEA Scholars is published in the RCEA Working Papers series.

The Finite-Sample Effects of VAR Dimensions on OLS Bias, OLS Variance, and Minimum MSE Estimators*

Steve Lawford† and Michalis P. Stamatogiannis

Department of Economics and Econometrics, ENAC, France;
and School of Economics, University of Nottingham, UK,
and Department of Business Studies, Philips College, Cyprus.

Abstract

Vector autoregressions (VARs) are important tools in time series analysis. However, relatively little is known about the finite-sample behaviour of parameter estimators. We address this issue, by investigating ordinary least squares (OLS) estimators given a data generating process that is a purely nonstationary first-order VAR. Specifically, we use Monte Carlo simulation and numerical optimization to derive response surfaces for OLS bias and variance, in terms of VAR dimensions, given correct specification and several types of over-parameterization of the model: we include a constant, and a constant and trend, and introduce excess lags. We then examine the correction factors that are required for the least squares estimator to attain minimum mean squared error (MSE). Our results improve and extend one of the main finite-sample multivariate analytical bias results of Abadir, Hadri and Tzavalis (Econometrica 67 (1999) 163), generalize the univariate variance and MSE findings of Abadir (Economics Letters 47 (1995) 263) to the multivariate setting, and complement various asymptotic studies.

*JEL classification: C15, C22, C32. Keywords: Finite-sample bias, Monte Carlo simulation, nonstationary time series, response surfaces, vector autoregression.

†Corresponding author: Steve Lawford, LH/ECO, ENAC, 7 avenue Edouard Belin, BP 54005, 31055, Toulouse, Cedex 4, France. Tel: +33 5 62 17 43 37. Fax: +33 5 62 17 40 17. Email: steve_lawford@yahoo.co.uk.

1 Introduction

Vector autoregressions have been extensively studied in econometrics and continue to be one of the most frequently used tools in time series analysis. However, little is currently known about the properties of parameter estimators when applied to finite samples of data, and especially in nonstationary frameworks. In particular, the form and extent of estimator bias and variance have not yet been fully investigated. In a paper that is central to this issue, Abadir, Hadri and Tzavalis (1999) (AHT) study nonstationary multivariate autoregressive series, and derive an approximate expression for the mean bias of the ordinary least squares estimator of the matrix of autoregressive parameters, in terms of the sample size T and VAR dimension k. They consider estimation of a correctly-parameterized first-order vector autoregression (a VAR(1)), with no constant or trend, given that the data generating process is a k-dimensional Gaussian random walk. Using Monte Carlo simulation, they show that their "analytic approximation" provides a good representation of bias in finite samples, and for small k (AHT, Table I).

The purposes of this paper are twofold. Firstly, we extend the results given by AHT in a number of directions, building upon previous studies by Stamatogiannis (1999) and Lawford (2001, chapter 4). In broadening the scope of AHT, we assess over-parameterization of the estimated VAR *model*, by including a constant, and a constant and deterministic trend. This creates additional bias problems, as was suggested by simulation results for the univariate case in Abadir and Hadri (2000, p. 97) and Tanizaki (2000, Table 1). We also assess the effects of introducing $p-1$ excess lags into the estimated model. We use Monte Carlo methods to simulate small sample bias, and then fit a series of response surfaces

using weighted nonlinear least squares. Well-specified and parsimonious response surfaces are chosen following diagnostic testing, and are shown to perform very well in out-of-sample prediction. In the correctly-parameterized setting, the prediction error of our response surface is substantially less than that of the AHT form, across the parameter space under investigation.

Secondly, we focus attention on the variance and MSE of the least squares estimator, and generalize the heuristic univariate variance approximation of Abadir (1995) to rigorous response surfaces. We develop response surfaces for variance, and show that multiplying the OLS estimator by a scalar correction factor achieves minimum MSE and removes most of the bias, at the expense of a small increase in estimator variance. To our knowledge, no other finite-sample approximations (analytic or otherwise), and few simulations, were previously available for bias in the multivariate over-parameterized cases, or for excess lags, or for variance in the multivariate setting.

The paper is organized as follows. Section 2 introduces the possibly over-parameterized VAR model and briefly reviews existing finite-sample results. Section 3 outlines the response surface methodology, presents the experimental design, and proposes response surfaces for multivariate bias and variance, based upon an extensive series of Monte Carlo experiments. Section 4 concludes the paper. We represent vector (and scalar) and matrix quantities as a and A respectively. Special vectors and matrices include the $k \times 1$ zero vector 0_k and the $k \times k$ identity matrix I_k.

2 Models and background

Let $\{x_t\}_1^T$ be a $k \times 1$ discrete time series that follows a purely nonstationary VAR(1), where T is the sample size, the innovations are independently and identically distributed with distribution D, and Ω is positive-definite:

$$x_t = x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d.D}\left(0_k, \Omega\right), \quad t = 1, 2, \ldots, T. \tag{1}$$

We examine the finite-sample bias, variance and MSE of the least squares estimator of (1), for each of the following estimated VAR(p) models:

$$\begin{aligned}
\text{Model A}: \quad x_t &= \widehat{\Phi} x_{t-1} + \sum_{j=1}^{p-1} \widehat{\Gamma}_j \Delta x_{t-j} + \widehat{\varepsilon}_t, \\
\text{Model B}: \quad x_t &= \overline{\mu} + \overline{\Phi} x_{t-1} + \sum_{j=1}^{p-1} \overline{\Gamma}_j \Delta x_{t-j} + \overline{\varepsilon}_t, \\
\text{Model C}: \quad x_t &= \widetilde{\mu} + \widetilde{\delta} t + \widetilde{\Phi} x_{t-1} + \sum_{j=1}^{p-1} \widetilde{\Gamma}_j \Delta x_{t-j} + \widetilde{\varepsilon}_t,
\end{aligned}$$

where Δ is the backward-difference operator, and over-parameterization arises through inclusion of a constant (Model B), a constant and time trend (Model C), and when there are multiple lags, with $p > 1$ (Models A, B, and C).[1] Zero initial values are chosen for simplicity ($x_{-j} = 0_k$, $j = 0, 1, \ldots, p-1$), and to avoid the problems of bias nonmonotonicity that can potentially arise when non-zero initial values are considered.[2]

[1] We are very grateful to the referees, who suggested that we generalize our original models.

[2] The correctly-parameterized univariate Model A, with $k = p = 1$, was examined by Abadir and Hadri (2000), given a (nearly) nonstationary data generating process, and non-zero initial values. They show, using numerical integration, that the bias of $\widehat{\phi}$ can be increasing in sample size T, due to the effect of $|x_0|$. This nonmonotonicity disappears under estimation of univariate Models B and C, at the expense of higher bias. A small simulation study of (1) and Model A by Lawford (2001), with $k \leq 6$, $p = 1$ and $x_0 \neq 0_k$, leads to the interesting conjecture that bias

PROPOSITION 1: *The bias matrix* $B = E(\widehat{\Phi}) - I_k$ *is scalar, and bias is invariant to* Ω*, for Models A, B, and C, if the error distribution D is symmetric, and* Ω *is positive-definite. Furthermore, the variances of each of the diagonal elements of* $\widehat{\Phi}$ *are identical, and variance is invariant to* Ω*, for Models A, B, and C, if D is symmetric, and* Ω *is both positive-definite and diagonal. A proof is available from the authors on request.*

Abadir (1993) uses some results on moment generating functions to derive a high-order closed form (integral-free) analytical approximation to the *univariate* finite-sample bias of $\widehat{\phi}$ given Model A, $k = p = 1$, and with $|\phi| = 1$. The final expression is based upon parabolic cylinder functions, and is computationally very efficient. Abadir further shows that bias may be described more simply in terms of exponential functions in polynomials of T^{-1}, and develops the following heuristic approximation:

$$b^{\text{UNIV}} \approx -1.7814\, T^{-1} \exp\left(-2.6138\, T^{-1}\right), \tag{2}$$

where -1.7814 is the expected value of the limiting distribution of $T(\widehat{\phi} - 1)$, e.g. see Le Breton and Pham (1989, p. 562).[3] Heuristic fits such as (2) have been used elsewhere in the literature, e.g. Dickey and Fuller (1981, p. 1064), and we distinguish here between these approximations and the rigorous response surface approach that is used in this paper. Despite the fact that only 5 datapoints are

nonmonotonicity also disappears when $k > 1$.

[3]This constant can be calculated conveniently by using the expression $1 - \frac{1}{2}\int_0^\infty u\,(\cosh u)^{-1/2}\,\mathrm{d}u = 1 - 2\sqrt{2}\ {}_3F_2\,(1/4, 1/4, 1/2; 5/4, 5/4; -1) \approx -1.7814$, where ${}_3F_2$ is a hypergeometric function.

used in the derivation of (2), it is accurate in-sample to 5 decimal places for bias, and is more accurate than the special function expression (see Abadir, 1993, Table 1). We found that (2) also performs very well out-of-sample, at least to 1 decimal place of $-100\times$bias. Other studies that examine the exact moments of OLS in univariate autoregressive models, with a variety of disturbances, include Evans and Savin (1981), Nankervis and Savin (1988), Tsui and Ali (1994), and Vinod and Shenton (1996); see also Maeshiro (1999) and Tanizaki (2000), and references therein.

In the multivariate setting, AHT consider Model A, $k \geq 1$, $p = 1$, and prove that B is exactly a *scalar matrix*, i.e. diagonal with equal diagonal elements: $B = \text{diag}(b, \ldots, b)$, and that B is invariant to Ω, given only a symmetric error distribution. Furthermore, they develop a simple quantitative approximation to *multivariate* finite-sample bias (especially AHT, p. 166, and Abadir, 1995, p. 264):

$$B^{\text{AHT}} \approx b^{\text{UNIV}} k I_k \equiv b\,{}^{\text{AHT}} I_k. \tag{3}$$

It is clear that bias is approximately proportional to the dimension of the VAR, even when Ω is diagonal. To facilitate discussion of cointegrating relations, AHT formulate their model as $\triangle x_t = \Psi x_{t-1} + \varepsilon_t$, where $\Psi \equiv \Phi - I_k$. Since the bias of $\widehat{\Psi}$ is equivalent to the bias of $\widehat{\Phi}$, our results may be compared directly to those in AHT, for $p = 1$, and no deterministics.

Abadir (1995, p. 265) uses the univariate Model A ($p = 1$) variance definition $v = 2T^{-2}\text{SD}^2$, with values for standard deviation "SD" of normalized $\widehat{\phi}$ taken from Evans and Savin (1981, Table III), and performs a similar heuristic process to that used in derivation of (2) for bias. This gives a variance approximation for

$k = p = 1$:

$$v^{\text{UNIV}} \approx 10.1124\, T^{-2} \exp\left(-5.4462\, T^{-1} + 14.519\, T^{-2}\right), \tag{4}$$

which is shown to be accurate to at least 7 decimal places in small samples. Since the bias and variance of each of the diagonal elements of $\widehat{\Phi}$ are respectively identical, we may use $\text{MSE}(\widehat{\phi}) = b^2 + v$ directly, to compute the MSE.

In the following section, we present the Monte Carlo experimental design, develop very accurate response surface approximations to multivariate bias and variance, and consider a simple correction for the OLS estimator to have minimum MSE.

3 Structure of Monte Carlo analysis

Response surfaces are numerical-analytical approximations, which can be very useful when summarizing and interpreting the small sample behaviour of tests and estimators. They have been applied to a variety of econometric problems by, *inter alia*, Engle, Hendry and Trumble (1985), Campos (1986), Ericsson (1991), MacKinnon (1994, 1996), Cheung and Lai (1995), MacKinnon, Haug and Michelis (1999) and Ericsson and MacKinnon (2002). See Hendry (1984) and Doornik and Hendry (2007, chapter 15) for good introductions. Briefly, a statistic τ is modelled as a (response surface) function $f(\cdot)$ of relevant variables, that is usually formulated in line with known analytical results. Monte Carlo is used to generate simulated estimates τ_i^* of τ, $i = 1, 2, \ldots, N$, for N experiments, where each τ_i^* is based upon M replications, and the parameters of $f(\cdot)$ are estimated using an appropriate procedure, depending upon the functional form.

The method can be computationally intensive, since M and N must be large if $f(\cdot)$ is to be accurately specified. To avoid problems of specificity, the estimated $f(\cdot)$ must be subjected to testing, and its out-of-sample performance assessed.

3.1 Monte Carlo design and simulation

The data generating process and models were introduced in (1) and Models A, B, and C. We adopt a minimal complete factorial design, which covers all triples (T, k, p) from:

$$T \in \{20, 21, \ldots, 30, 35, \ldots, 80, 90, 100, 150, 200\}, \; k \in \{1, 2, 3, 4\}, \; p \in \{1, 2, 3, 4\}, \tag{5}$$

giving $N = 400$ datapoints. The sample sizes that we have chosen are representative of those that are commonly used in practice, and our design includes small k and p, so that the effects of changes in VAR dimension and model lag can be explored. From Proposition 1, and with no loss of generality, we set $\varepsilon_t \sim$ i.i.d.N$(0_k, I_k)$ in the simulations. We calculate the OLS estimate for each combination of (T, k, p) in the parameter space, from which we directly derive the bias. Since B is a scalar matrix, we may estimate the scalar b by averaging over the estimated diagonal elements of B. This results in a further increase in accuracy as k increases. We simulate variance v similarly.[4] The period of our pseudo-random number procedure is much larger than the total random number requirement. All simulations were performed most recently on Pentium 4 machines, with 2.5GHz processors and 512MB of RAM, running GAUSS and/or Python under Microsoft

[4]We experimented with a pseudo-antithetic variate technique, based upon Abadir and Paruolo's (2007) univariate "AV4", and were able to increase the speed of the bias simulations by roughly 50%, for a given precision [Model A, $p = 1$]. While conventional antithetics are not generally applicable to the nonstationary setting, the pseudo-antithetic is not valid either for some of the models considered above, and is therefore not used in this paper.

Windows XP.

Where possible, our numerical results were checked with partial exact and approximate results in the literature. These include MacKinnon and Smith (1998, Figure 1), who plot bias functions under Model B ($k = p = 1$), and Pere (2000, Table 3), who reports values that correspond to variances in the same model, in his study of adjusted profile likelihood. Evans and Savin (1981, Table 3) give bias and standard deviation for $2^{-1/2}T(\widehat{\phi}-\phi)$ under Model A ($k = p = 1$), which agree closely (3 to 5 decimal places) with our simulation results. Roy and Fuller (2001, Tables 1 and 6) report bias and MSE for $T = 100$, under univariate Models B and C, for $p = 1$.

3.2 Post-simulation analysis

We regressed the Monte Carlo estimates of bias and variance under Models A, B, and C, on functions of sample size, VAR dimension and lag order, to reflect the dependence of b and v upon these parameters, and on the degree of over-parameterization. Following extensive experimentation, and motivated by (2), we fit the following nonlinear bias response surface for each of the models:[5]

$$(s_i^b)^{-1}\, b\,(T_i, k_i, p_i) = (s_i^b)^{-1}\,(\beta_1 + \beta_2\, k_i)\, T_i^{-1} \exp\left[(\beta_3 + \beta_4\, k_i + \beta_5\, k_i p_i + \beta_6\, k_i^{p_i})\, T_i^{-1}\right] + u_i. \tag{6}$$

[5]Some early motivation for numerical refinement of (3), for Model A, with $p = 1$, came from consideration of low-order partial derivatives of b^{AHT} . Straightforward algebra gives (for $T \geq 1$) $b^{\mathrm{AHT}} < 0$, $\partial b^{\mathrm{AHT}}/\partial k < 0$, $\partial^2 b^{\mathrm{AHT}}/\partial k^2 = 0$, (for $T \geq 3$) $\partial b^{\mathrm{AHT}}/\partial T > 0$, $\partial^2 b^{\mathrm{AHT}}/\partial k \partial T > 0$, (for $T \geq 5$) $\partial^2 b^{\mathrm{AHT}}/\partial T^2 < 0$. Upon comparing these theoretical partials with approximate numerical partial derivatives from simulated data, it is found that each holds, except for $\partial^2 b/\partial k^2 = 0$ (simulations suggest that $\partial^2 b/\partial k^2 > 0$, for T not too large). This finding suggested that improvements were possible over (3), and especially that k entered the formula in a more complicated manner than in (3).

The dependent variable $b\left(T_i, k_i, p_i\right)$ is the simulated finite-sample bias for sample size T_i, VAR dimension k_i, and lag order p_i, which take values from (5), and u_i is an error term. We correct for Monte Carlo sampling heteroscedasticity using the term s_i^b, which is the simulated sampling error standard deviation of bias over replications (see Doornik and Hendry, 2007, chapter 15, for details). We denote the fitted values of the estimated response surface by b^{RS}, and estimated coefficients are reported in Table 1. Convergence of the weighted nonlinear least squares routine was very fast, and required few iterations. Selection criteria included small residual variance and good in-sample fit, parsimony, and satisfactory diagnostic performance. The response surface fits are extremely good in-sample, and the Jarque-Bera statistic for normality is small. The signs of all estimated coefficients apart from the constant β_1 remain the same across the different models. Note that the asymptotic bias $T_i b$ (as $T_i \to \infty$) is a linear function of k_i alone, which agrees with numerical observations, and that $\beta_1 + \beta_2 \, k_i$ can be interpreted as the asymptotic component of bias, with the exponential representing the (analytically intractable) finite-sample "adjustment", which depends on k_i and p_i (and T_i).

(Table 1 about here)

We recalculate Table I in AHT as Table 2 in this paper, with increased accuracy, with additional results reported for $T = 400, 800$ and $k = 6, 7, 8$, and correcting for a typo in AHT Table I: $(T, k) = (25, 5)$. It is convenient to interpret the scaled bias values as percentages of the true parameter value, e.g. in Model A, given $(T, k) = (25, 8)$, and $p = 1$, the absolute bias of each of the estimated parameters on the diagonal of $\widehat{\Phi}$ is 46.7% of the true value (unity). Clearly, absolute bias is strictly increasing in k and decreasing in T. As T increases, bias

goes to zero, as is well-known from asymptotic theory. We see that b^{AHT} gives a good approximation to bias for k small, and especially for $k = 1$, where (3) reduces to the excellent heuristic approximation (2). However, as k increases, b^{RS} provides much closer approximations to bias, even for T quite large. Out-of-sample points reported in Table 2 for b^{RS} are combinations of $k = 5, 6, 7, 8$, and $T = 400, 800$. While b^{AHT} is only applicable for correctly-parameterized Model A, our response surfaces can be used when $p > 1$, and also when deterministics are included. The out-of-sample fit appears to be excellent for all T, and up to about $k = p = 6$ (as k and p jointly become large, with small T, the term k^p will dominate the bias approximation, and out-of-sample predictions should be used with particular caution). Although the response surfaces are developed with small sample rather than asymptotic considerations in mind, it is interesting to approximate univariate asymptotic bias by setting $k = p = 1$ and letting $T_i \to \infty$ in $T_i b^{\mathrm{RS}}$, from (6), which gives $T_i b^{\mathrm{RS}} = \widehat{\beta}_1 + \widehat{\beta}_2$ of approximately -1.7, -5.4 and -10.3 in Models A, B and C respectively.

(Table 2 about here)

Kiviet and Phillips (2005, equation (14), and Figure 1) consider univariate Model B, where the data generating process can have a non-zero drift, and write autoregressive bias in terms of "g-functions " $g_0(T)$ and $g_1(T)$, which they calculate using simulations. The function $g_0(T)$ represents least squares bias when there is a zero drift in the data generating process, while $g_1(T)$ appears as the bias increment due to non-zero drift. Our equation (6) simplifies (when $k = p = 1$) to $g_0(T) \approx -5.3654\, T^{-1} \exp\left(-2.6513\, T^{-1}\right)$, which provides a convenient means of calculating $g_0(T)$ without further simulations.

Using (4) to motivate the choice of functional form, we fit the variance response surface:

$$\begin{aligned}(s_i^v)^{-1} v\,(T_i, k_i, p_i) \;&=\; (s_i^v)^{-1}\,(\gamma_1 + \gamma_2\ k_i)\,T_i^{-2} \exp[(\gamma_3 + \gamma_4\ k_i + \gamma_5\ p_i + \gamma_6\ k_i p_i)\,T_i^{-1} \\ &\quad + (\gamma_7\ k_i p_i + \gamma_8\ k_i^{p_i})\,T_i^{-2}] + u_i, \qquad (7)\end{aligned}$$

where $v\,(T_i, k_i, p_i)$ is the simulated finite-sample variance, and s_i^v is the simulated sampling error standard deviation of the variance over replications. In estimating (7), we did not use datapoints for which $T_i = 20, \ldots, 24$ (and so $N = 320$), since variance becomes very large for such small sample sizes, which makes it very difficult to specify good response surfaces across the full parameter space. Estimated response surfaces v^{RS} are given in Table 3, and are seen to fit very well. The signs of each of the estimated coefficients, except for γ_1, remains the same across the models, the Jarque-Bera statistic is relatively low, and v^{RS} provides a very good approximation in-sample. The out-of-sample variance approximation should be used with caution as k and p jointly exceed about 5 or 6, with small T, again due to the effect of the term k^p. We note that no variance approximations were previously available for over-parameterized models, excess lags, or even for $k > 1$. Similarly to the bias response surfaces, the asymptotic variance $T_i^2 v$ (as $T_i \to \infty$) is a linear function of k_i alone, and $\gamma_1 + \gamma_2\, k_i$ can be interpreted as the asymptotic component of variance, with the exponential term again representing the finite-sample "adjustment", which depends upon k_i and p_i (and T_i). The dependencies of bias and variance on T, k, and p are depicted in Figures 1 and 2, which plot scaled response surfaces $-100 \times b^{\mathrm{RS}}$ and $10,000 \times v^{\mathrm{RS}}$, against T, for Models A, B, and C, with $k = 1, 2$ and $p = 1, 2$.

(Table 3 about here)

(Figures 1 and 2 about here)

Bias and variance are not the only criteria to be used in comparison of time series estimates, and the mean squared error, $\mathrm{MSE}(\widehat{\phi}) = b^2 + v$, is often of interest. For univariate Model A ($p = 1$), Abadir (1995) defines λ as a correction factor such that $\mathrm{MSE}(\lambda\widehat{\phi})$ is minimized, and b^m and v^m as the bias and variance of the corrected OLS estimator $\lambda\widehat{\phi}$, with:

$$\lambda = \frac{1+b}{v+(1+b)^2}, \quad b^m = \frac{-v}{v+(1+b)^2}, \quad v^m = \lambda^2 v, \tag{8}$$

when $\phi = 1$. Equations (6) and (7) may be now combined to give an approximation to MSE, and by substitution of response surface values for bias and variance into (8), we are able to calculate λ for various T, k, and p. As an illustration, correction factors are reported in Table 4, for $p = 1$ and Model A, which displays qualitatively similar results to those in Abadir (1995, Tables 2 and 3). It is clear that OLS ($\lambda = 1$) does not achieve minimum MSE. It is also shown that the corrected OLS is almost unbiased, unlike OLS. From Table 4, λ increases monotonically with k and decreases monotonically with T (asymptotically, the OLS achieves minimum MSE). The correction can be particularly large for small T, e.g. $(T, k) = (25, 5)$ implies a correction of 32%. The corrected estimator is much less biased than the OLS, and b^m tends to zero more rapidly than b. However, this reduction in bias comes at the expense of a small increase in the variance of the corrected estimator, v^m. It is seen that b^2 forms a much larger proportion of MSE than variance for $k \geq 3$, although this is reversed following the minimum MSE correction; and that MSE efficiency is generally decreasing

in T and k.

(Table 4 about here)

4 Concluding comments

We have performed an extensive set of Monte Carlo experiments on the bias and variance of the OLS of the autoregressive parameters, given a data generating process that is a purely nonstationary VAR(1), where the estimated model is a possibly over-parameterized VAR(p), for small sample sizes, and various VAR dimensions and lag lengths. Although the univariate framework has been the subject of much research, a comprehensive multivariate simulation study has not previously been performed. We estimate parsimonious and computationally convenient response surfaces for bias and variance, that are much more accurate and more general than existing approximations. In this way, we improve numerically upon existing finite-sample analytical bias results, and extend them to $p > 1$ and deterministics, and also extend existing finite-sample variance results to $k > 1$, $p > 1$, and to deterministics. Finally, we investigate the correction factors required for the OLS to achieve minimum MSE and show that this correction can significantly reduce bias, at the expense of a small increase in estimator variance. Our results may provide guidelines for applied researchers as to the behaviour of VAR models, given that relatively short samples and nonstationary data are often relevant in empirical work.

Our work complements important asymptotic treatments by Phillips (1987a) in the univariate framework, and Park and Phillips (1988, 1989), Phillips (1987b), and Tsay and Tiao (1990) in the multivariate setting. Our results may also be

useful when studying the derivation of exact formulae (for instance, in conjunction with work by Abadir and Larsson, 1996, 2001, who derive the exact finite-sample moment generating function of the quadratic forms that create the basis for the sufficient statistic in a discrete Gaussian vector autoregression). Exact analytical bias expressions may involve multiple infinite series of matrix-argument hypergeometric functions (generalizing, e.g. Abadir, 1993). When such series arise in other areas of econometrics, they are generally complicated and may be difficult to implement for numerical evaluation. We may, therefore, need to rely upon approximations in practice, even when the exact formulae are available.

The authors are indebted to Karim Abadir, for his insightful comments. For helpful feedback and discussions, we are grateful to the co-editor Peter Robinson and two referees, whose comments greatly improved the paper, and to Walter Distaso, Les Godfrey, Steve Leybourne, Tassos Magdalinos, Jan Magnus, and Patrick Marsh, and seminar participants at Brunel, Liverpool, Tilburg and York. We also thank participants at the 4th Econometric Study Group Meeting (Bristol) and the 6th ERC/METU Conference in Economics (Ankara), where this paper was awarded the CBRT Prize for Best Paper by a Young Economist. Partial financial support was provided by a Marie Curie Individual Fellowship (Lawford) and a University of York Teaching Fellowship (Stamatogiannis). Lawford is very grateful to Marc Hallin and ECARES, and Luc Bauwens and CORE, for their hospitality at various times, and to Nathalie Lenoir and ENAC, where the final version was prepared. Stamatogiannis thanks Philippos Constantinou and the Philips Research Centre for support during part of the research. This paper was typed in MiKTeX and numerical results were derived using GAUSS, Python and E–Views.

Abadir, K. M., 1993, OLS bias in a nonstationary autoregression, Econometric Theory 9, 81-93.

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

Abadir, K. M. and K. Hadri, 2000, Is more information a good thing? Bias nonmonotonicity in stochastic difference equations, Bulletin of Economic Research 52, 91-100.

Abadir, K. M., Hadri, K. and E. Tzavalis, 1999, The influence of VAR dimensions on estimator biases, Econometrica 67, 163-181.

Abadir, K. M. and R. Larsson, 1996, The joint moment generating function of quadratic forms in multivariate autoregressive series, Econometric Theory 12, 682-704.

Abadir, K. M. and R. Larsson, 2001, The joint moment generating function of quadratic forms in multivariate autoregressive series: The case with deterministic components, Econometric Theory 17, 222-246.

Abadir, K. M. and P. Paruolo, 2007, On efficient simulations in dynamic models, unpublished manuscript.

Campos, J., 1986, Finite-sample properties of the instrumental-variables estimator for dynamic simultaneous-equation subsystems with ARMA disturbances, Journal of Econometrics 32, 333-366.

Cheung, Y. -W. and K. S. Lai, 1995, Lag order and critical values of the augmented Dickey-Fuller test, Journal of Business and Economic Statistics 13, 277-280.

Dickey, D. A. and W. A. Fuller, 1981, Likelihood ratio statistics for autoregressive time series with a unit root, Econometrica 43, 1057-1072.

Doornik, J. A. and D. F. Hendry, 2007, Interactive Monte Carlo experimentation in econometrics using PcNaive4, Timberlake Consultants, London.

Engle, R. F., Hendry, D. F. and D. Trumble, 1985, Small-sample properties of ARCH estimators and tests, Canadian Journal of Economics 18, 66-93.

Ericsson, N. R., 1991, Monte Carlo methodology and the finite sample properties of instrumental variables statistics for testing nested and non-nested hypotheses, Econometrica 59, 1249-1277.

Ericsson, N. R. and J. G. MacKinnon, 2002, Distributions of error correction tests for cointegration, Econometrics Journal 5, 285-318.

Evans, G. B. A. and N. E. Savin, 1981, Testing for unit roots: 1, Econometrica 49, 753-779.

Hendry, D. F., 1984, Monte Carlo experimentation in econometrics, in: Griliches, Z., Intriligator, M. D. eds., Handbook of Econometrics, Volume 2, North-Holland, Amsterdam.

Kiviet, J. F. and G. D. A. Phillips, 2005, Moment approximation for least-squares estimators in dynamic regression models with a unit root, Econometrics Journal 8, 115-142.

Lawford, S., 2001, Improved modelling in finite-sample and nonlinear frameworks, unpublished D.Phil. thesis, University of York.

Le Breton, A. and D. T. Pham, 1989, On the bias of the least squares estimator for the first order autoregressive process, Annals of the Institute of Statistical Mathematics 41, 555-563.

MacKinnon, J. G., 1994, Approximate asymptotic distribution functions for unit-root and cointegration tests, Journal of Business and Economic Statistics 12, 167-176.

MacKinnon, J. G., 1996, Numerical distribution functions for unit root and cointegration tests, Journal of Applied Econometrics 11, 601-618.

MacKinnon, J. G., Haug, A. A. and L. Michelis, 1999, Numerical distribution functions of likelihood ratio tests for cointegration, Journal of Applied Econometrics 14, 563-577.

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

Maeshiro, A., 1999, A lagged dependent variable, autocorrelated disturbances, and unit root tests – peculiar OLS bias properties – a pedagogical note, Applied Economics 31, 381-396.

Nankervis, J. C. and N. E. Savin, 1988, The exact moments of the least squares estimator for the autoregressive model: Corrections and extensions, Journal of Econometrics 37, 381-388.

Park, J. Y. and P. C. B. Phillips, 1988, Statistical inference in regressions with integrated processes: Part 1, Econometric Theory 4, 468-497.

Park, J. Y. and P. C. B. Phillips, 1989, Statistical inference in regressions with integrated processes: Part 2, Econometric Theory 5, 95-131.

Pere, P., 2000, Adjusted estimates and Wald statistics for the AR(1) model with constant, Journal of Econometrics 98, 335-363.

Phillips, P. C. B., 1987a, Time series regression with a unit root, Econometrica 55, 277-301.

Phillips, P. C. B., 1987b, Asymptotic expansions in nonstationary vector autoregressions, Econometric Theory 3, 45-68.

Roy, A. and W. A. Fuller, 2001, Estimation for autoregressive time series with a root near 1, Journal of Business and Economic Statistics 19, 482-493.

Stamatogiannis, M. P., 1999, A study on the finite sample bias of an unstable VAR with a drift, unpublished M.Sc. thesis, University of York.

Tanizaki, H., 2000, Bias correction of OLSE in the regression model with lagged dependent variables, Computational Statistics and Data Analysis 34, 495-511.

Tsay, R. S. and G. C. Tiao, 1990, Asymptotic properties of multivariate nonstationary processes with applications to autoregressions, Annals of Statistics 18, 220-250.

Tsui, A. K. and M. M. Ali, 1994, Exact distributions, density functions and moments of the least-squares estimator in a first-order autoregressive model, Computational Statistics and Data Analysis 17, 433-454.

Vinod, H. D. and L. R. Shenton, 1996, Exact moments for autoregressive and random walk models for a zero or stationary initial value, Econometric Theory 12, 481-499.

	Model A	Model B	Model C
$\widehat{\beta}_1$	0.320104	-3.475124	-8.522213
	(0.010386)	(0.013364)	(0.053317)
$\widehat{\beta}_2$	-2.044023	-1.890281	-1.743529
	(0.004011)	(0.005275)	(0.018449)
$\widehat{\beta}_3$	-1.123620	-1.788271	-1.410424
	(0.136479)	(0.093583)	(0.227607)
$\widehat{\beta}_4$	-1.860624	-1.907111	-2.632320
	(0.039287)	(0.029559)	(0.080549)
$\widehat{\beta}_5$	0.998913	1.037856	1.404018
	(0.010382)	(0.009215)	(0.019767)
$\widehat{\beta}_6$	0.008005	0.006210	0.002401
	(0.000712)	(0.000496)	(0.000822)
$\overline{R}^2$	0.9995	0.9996	0.9976
JB	1.35	$8.95^{\star}$	$8.92^{\star}$

Table 1: Estimated bias response surfaces b^{RS} for Models A, B, and C. Response surfaces (6) were estimated using weighted nonlinear least squares. White's heteroscedasticity-consistent standard errors are given in parentheses, $\overline{R}^2$ is the degrees-of-freedom adjusted coefficient of determination, JB is the Jarque-Bera test statistic for normality, asymptotically distributed as $\chi^2(2)$, and $\star$ denotes significance at the 5% level.

		VAR dimension (k)							
T		1	2	3	4	5	6	7	8
	b	6.4	13.5	20.0	26.1	31.8	37.1	42.1	46.7
	b^{AHT}	(6.4)	(12.8)	(19.3)	(25.7)	(32.1)	(38.5)	(44.9)	(51.3)
25	b^{RS}	[6.4]	[13.5]	[20.1]	[26.2]	[31.9]	[37.2]	[42.1]	[46.7]
	$\overline{b}$	19.2	25.0	30.6	35.9	40.9	45.7	50.2	54.5
	$\widetilde{b}$	35.3	40.0	44.5	49.0	53.2	57.3	61.2	64.9
	b	3.4	7.2	10.8	14.3	17.6	20.9	24.0	27.0
	b^{AHT}	(3.4)	(6.8)	(10.1)	(13.5)	(16.9)	(20.3)	(23.7)	(27.1)
50	b^{RS}	[3.3]	[7.1]	[10.8]	[14.3]	[17.8]	[21.1]	[24.3]	[27.3]
	$\overline{b}$	10.1	13.4	16.7	19.9	23.0	26.0	28.9	31.8
	$\widetilde{b}$	19.0	21.8	24.7	27.5	30.3	33.0	35.7	38.3
	b	1.7	3.7	5.6	7.5	9.3	11.1	12.9	14.6
	b^{AHT}	(1.7)	(3.5)	(5.2)	(6.9)	(8.7)	(10.4)	(12.1)	(13.9)
100	b^{RS}	[1.7]	[3.7]	[5.6]	[7.5]	[9.4]	[11.2]	[13.0]	[14.8]
	$\overline{b}$	5.2	7.0	8.7	10.5	12.2	14.0	15.7	17.3
	$\widetilde{b}$	9.9	11.4	13.0	14.6	16.3	17.9	19.5	21.1
	b	0.9	1.9	2.9	3.8	4.8	5.8	6.7	7.6
	b^{AHT}	(0.9)	(1.8)	(2.6)	(3.5)	(4.4)	(5.3)	(6.2)	(7.0)
200	b^{RS}	[0.9]	[1.9]	[2.9]	[3.8]	[4.8]	[5.8]	[6.8]	[7.7]
	$\overline{b}$	2.6	3.6	4.5	5.4	6.3	7.3	8.2	9.1
	$\widetilde{b}$	5.0	5.8	6.7	7.6	8.4	9.3	10.2	11.1
	b	0.4	0.9	1.4	1.9	2.4	2.9	3.4	3.9
	b^{AHT}	(0.4)	(0.9)	(1.3)	(1.8)	(2.2)	(2.7)	(3.1)	(3.5)
400	b^{RS}	[0.4]	[0.9]	[1.4]	[1.9]	[2.4]	[2.9]	[3.4]	[3.9]
	$\overline{b}$	1.3	1.8	2.3	2.7	3.2	3.7	4.2	4.6
	$\widetilde{b}$	2.5	3.0	3.4	3.9	4.3	4.8	5.2	5.7
	b	0.2	0.5	0.7	1.0	1.2	1.5	1.7	2.0
	b^{AHT}	(0.2)	(0.4)	(0.7)	(0.9)	(1.1)	(1.3)	(1.6)	(1.8)
800	b^{RS}	[0.2]	[0.5]	[0.7]	[1.0]	[1.2]	[1.5]	[1.7]	[2.0]
	$\overline{b}$	0.7	0.9	1.1	1.4	1.6	1.9	2.1	2.4
	$\widetilde{b}$	1.3	1.5	1.7	1.9	2.2	2.4	2.6	2.9

Table 2: Simulated scaled bias in Models A, B, and C, for $p = 1$, and AHT and Model A approximations. All reported bias values have been multiplied by -100, b is the simulated Model A bias, b^{AHT} is the AHT approximation (3) to Model A bias, b^{RS} is the response surface approximation (6) to Model A bias, $\overline{b}$ is the simulated Model B bias, and $\widetilde{b}$ is the simulated Model C bias. In-sample points correspond to $k = 1, 2, 3, 4$ and $T = 25, 50, 100, 200$.

	Model A	Model B	Model C
$\widehat{\gamma}_1$	-0.345264	10.42961	26.23034
	(0.054616)	(0.081527)	(0.150349)
$\widehat{\gamma}_2$	10.40041	9.895323	10.10400
	(0.040182)	(0.048754)	(0.087121)
$\widehat{\gamma}_3$	-4.469397	-9.679804	-17.05139
	(0.203317)	(0.191752)	(0.249821)
$\widehat{\gamma}_4$	-5.302022	-4.979473	-4.800818
	(0.077115)	(0.083314)	(0.114279)
$\widehat{\gamma}_5$	1.245448	2.058910	4.750621
	(0.092856)	(0.075961)	(0.102145)
$\widehat{\gamma}_6$	2.925115	2.957002	2.969689
	(0.041124)	(0.035076)	(0.046760)
$\widehat{\gamma}_7$	13.23278	11.64608	14.66810
	(0.883959)	(0.767140)	(0.966050)
$\widehat{\gamma}_8$	0.993006	0.888653	0.923219
	(0.041359)	(0.033083)	(0.045075)
$\overline{R}^2$	0.9991	0.9990	0.9982
JB	91.03**	46.38**	30.22**

Table 3: Estimated variance response surfaces v^{RS} for Models A, B, and C. Response surfaces (7) were estimated using weighted nonlinear least squares. White's heteroscedasticity-consistent standard errors are given in parentheses, $\overline{R}^2$ is the degrees-of-freedom adjusted coefficient of determination, JB is the Jarque-Bera test statistic for normality, asymptotically distributed as $\chi^2(2)$, and $\star\star$ denotes significance at the 1% level.

		VAR dimension (k)							
T		1	2	3	4	5	6	7	8
	λ	1.05	1.12	1.19	1.26	1.32	1.39	1.46	1.52
	br	0.23	0.24	0.26	0.28	0.31	0.34	0.37	0.40
25	vr	1.11	1.25	1.41	1.58	1.75	1.94	2.13	2.32
	bc	24/1	42/3	54/5	61/7	67/10	71/13	74/15	77/19
	me	86	75	69	66	65	64	65	66
	λ	1.03	1.07	1.11	1.15	1.19	1.23	1.28	1.32
	br	0.12	0.11	0.12	0.12	0.13	0.13	0.14	0.15
50	vr	1.06	1.14	1.23	1.32	1.41	1.52	1.63	1.74
	bc	23/0.4	42/1	53/1	61/2	67/2	71/3	74/3	77/4
	me	82	67	58	52	48	45	43	42
	λ	1.02	1.04	1.06	1.08	1.10	1.12	1.14	1.16
	br	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
100	vr	1.03	1.07	1.12	1.16	1.20	1.25	1.30	1.35
	bc	23/0.1	41/0.2	53/0.3	61/0.4	66/0.5	71/0.7	74/0.8	77/0.9
	me	80	63	53	46	41	37	34	32
	λ	1.01	1.02	1.03	1.04	1.05	1.06	1.07	1.08
	br	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
200	vr	1.02	1.04	1.06	1.08	1.10	1.12	1.15	1.17
	bc	23/0.0	41/0.1	53/0.1	60/0.1	66/0.1	70/0.2	73/0.2	76/0.2
	me	78	61	50	43	38	34	30	28
	λ	1.00	1.01	1.01	1.02	1.02	1.03	1.04	1.04
	br	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
400	vr	1.01	1.02	1.03	1.04	1.05	1.07	1.07	1.08
	bc	23/0.0	41/0.0	52/0.0	60/0.0	66/0.0	70/0.0	73/0.0	76/0.1
	me	78	60	49	41	36	32	29	26
	λ	1.00	1.00	1.01	1.01	1.01	1.01	1.02	1.02
	br	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
800	vr	1.00	1.01	1.01	1.02	1.02	1.03	1.04	1.04
	bc	23/0.0	41/0.0	52/0.0	60/0.0	66/0.0	70/0.0	73/0.0	76/0.0
	me	77	60	48	41	35	31	28	25

Table 4: Minimum MSE correction in Model A, for $p = 1$. λ is a correction factor, such that $\lambda\widehat{\phi}$ attains minimum MSE, br is the bias ratio≡corrected bias/OLS bias, vr is the variance ratio≡corrected variance/OLS variance $\left(\text{vr} \equiv \lambda^2\right)$, bc and "$x/y$" indicate that b^2 forms x% of MSE under OLS, and *corrected* b^2 forms y% of *minimized* MSE, me is the MSE efficiency≡MSE after correction/MSE under OLS ($\times 100$). All values are computed using the appropriate response surface approximations (6) and (7). In-sample points correspond to $k = 1, 2, 3, 4$ and $T = 25, 50, 100, 200$.

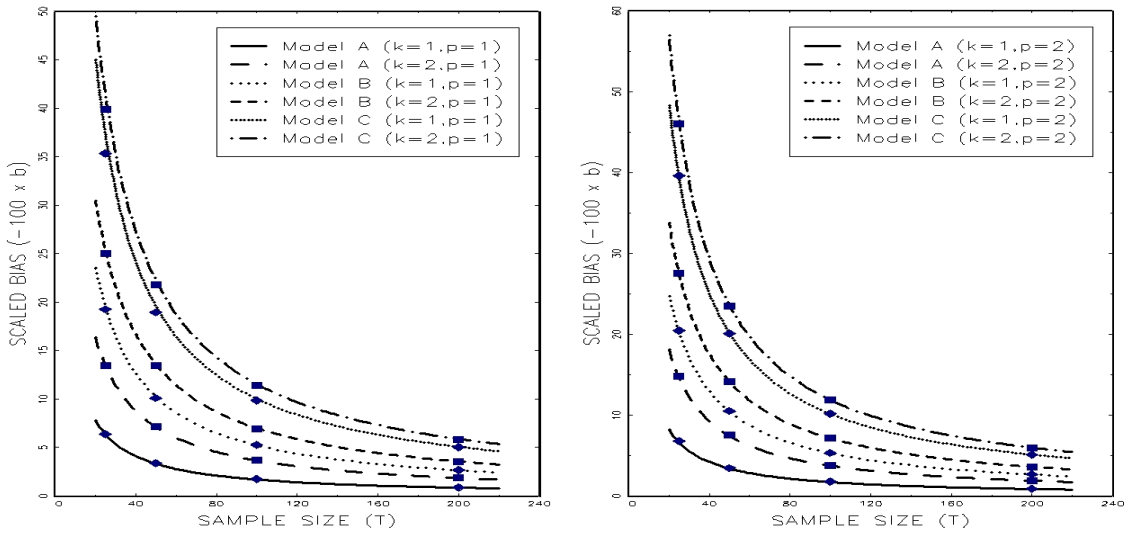

Figure 1:

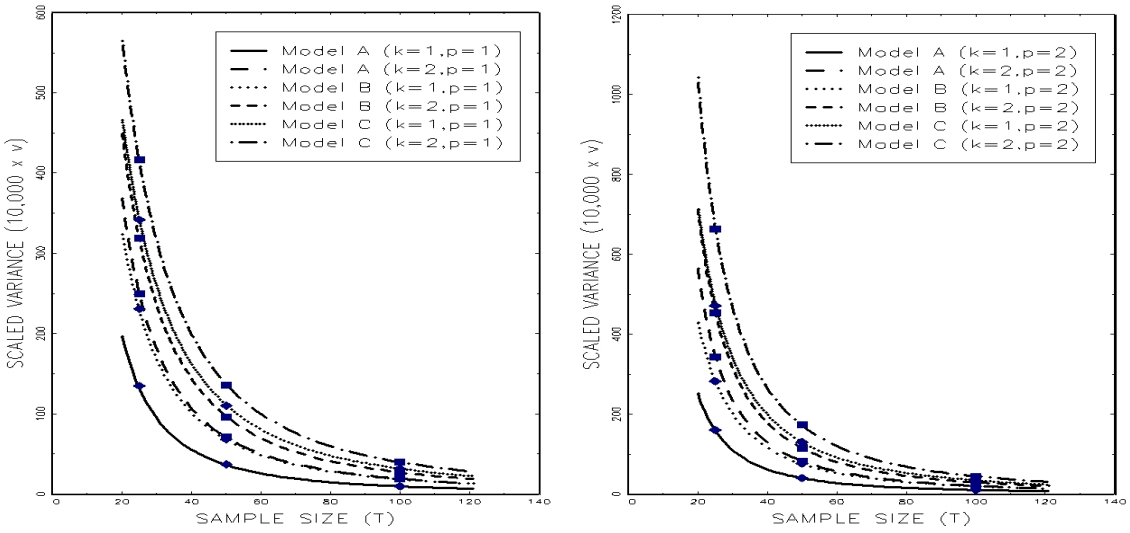

Figure 2:

Figure 1. Bias response surfaces: scaled bias against T, for $k = 1, 2$ and $p = 1, 2$, for Models A, B, and C. Simulated values are represented by diamonds ($k = 1$) and squares ($k = 2$).

Figure 2. Variance response surfaces: scaled variance against T, for $k = 1, 2$ and $p = 1, 2$, for Models A, B, and C. Simulated values are represented by diamonds ($k = 1$) and squares ($k = 2$).