The Finite-Sample Effects of VAR Dimensions on OLS Bias, OLS Variance, and Minimum MSE Estimators: Purely Nonstationary Case

STEVE LAWFORD* and MICHALIS P. STAMATOGIANNIS
Brunel University and University of York

This draft: February 2004. First draft: March 2002.

Abstract

Vector autoregressions (VARs) are an important tool 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 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 required for the least squares estimator to attain minimum mean squared error (MSE). Our results improve and extend one of the main finite-sample analytical bias results of Abadir, Hadri and Tzavalis (Econometrica 67 (1999) 163), generalize the univariate variance and MSE findings of Abadir (Econ. Lett. 47 (1995) 263), and complement various asymptotic studies.

Keywords: Finite-sample bias; Monte Carlo simulation; Nonstationary time series; Response surfaces; Vector autoregression

JEL classification: C15; C22; C32

*Corresponding author: Steve Lawford, Department of Economics and Finance, Brunel University, Uxbridge, UB8 3PH, UK; steve_lawford@yahoo.co.uk; fax: +44 1895 203 384.

1 Introduction

Vector autoregressions (VARs) 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 recent paper, Abadir, Hadri and Tzavalis (1999) (AHT) study nonstationary multiple autoregressive series, and derive an approximate expression for the bias of the ordinary least squares (OLS) 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).[1]

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, ch. 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 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 nonlinear least squares. Well-specified and parsimonious response surfaces are chosen following diagnostic testing, and are shown to perform extremely 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. To our knowledge, no other small sample approximations – analytic or otherwise – were previously available in the over-parameterized cases, or for excess lags.

Secondly, we focus attention on the variance and mean squared error (MSE) of the least squares estimator, and generalize the heuristic univariate variance approximation of Abadir (1995a) to the multivariate and over-

[1] Throughout this paper, we refer to mean-bias as "bias". Median-bias is defined as the difference between the median of an estimator and the true parameter value; see Andrews (1993) for a discussion of median-bias in the context of unit-root/autoregressive models.

parameterized setting. 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.

The paper is organized as follows. Section 2 introduces the possibly overparameterized 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 bias and variance, following an extensive series of Monte Carlo experiments. Section 4 concludes the paper. Proofs of Theorems 2.1 and 2.2 are available from the authors on request. The notation generally follows Abadir and Magnus (2002). We represent scalar, vector and matrix quantities as a, $\boldsymbol{a}$ and $\boldsymbol{A}$ respectively: these have typical elements $\boldsymbol{a} = (\boldsymbol{a})_j$ and $\boldsymbol{A} = (\boldsymbol{A})_{ij}$. When a vector or matrix reduces to a scalar, we write $\boldsymbol{a} = a$ or $\boldsymbol{A} = a$. Special vectors and matrices include the $k \times 1$ zero vector $\mathbf{0}_k$ and the $k \times k$ identity matrix $\boldsymbol{I}_k$.

2 Models and background

Let $\{\boldsymbol{x}_t\}_1^T$ be a $k \times 1$ discrete time series that follows a purely nonstationary Gaussian VAR(1), where T is the sample size, and $\boldsymbol{\Omega}$ is positive-definite:

$$\boldsymbol{x}_t = \boldsymbol{A}_1 \boldsymbol{x}_{t-1} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathrm{IN}\left(\mathbf{0}_k, \boldsymbol{\Omega}\right), \quad \boldsymbol{A}_1 = \boldsymbol{I}_k. \tag{1}$$

We examine the finite-sample bias, variance and MSE of the least squares (maximum likelihood) estimator of $\boldsymbol{A}_1$, for each of the following estimated VAR(p) models:

$$\boldsymbol{x}_t = \sum_{j=1}^{p} \widehat{\boldsymbol{A}}_j \boldsymbol{x}_{t-j} + \widehat{\boldsymbol{\varepsilon}}_t, \tag{A}$$

$$\boldsymbol{x}_t = \overline{\boldsymbol{\mu}} + \sum_{j=1}^{p} \overline{\boldsymbol{A}}_j \boldsymbol{x}_{t-j} + \overline{\boldsymbol{\varepsilon}}_t, \tag{B}$$

$$\boldsymbol{x}_t = \widetilde{\boldsymbol{\mu}} + \widetilde{\boldsymbol{\delta}} t + \sum_{j=1}^{p} \widetilde{\boldsymbol{A}}_j \boldsymbol{x}_{t-j} + \widetilde{\boldsymbol{\varepsilon}}_t, \tag{C}$$

where overparameterization arises through inclusion of a constant (B), a constant and time trend (C), and when there are multiple lags; that is, when $p > 1$ (A,B,C).[2] Throughout, we assume the following:

[2] We are very grateful to one of the referees, who suggested that we consider $p \neq 1$.

Assumption 2.1. $\boldsymbol{x}_{-j} = \mathbf{0}_k, \quad j = 0, 1, \ldots, p-1.$

Assumption 2.2. $\boldsymbol{\Omega} = \boldsymbol{I}_k$.

Assumption 2.1 (zero initial values) is chosen for simplicity, and to avoid potential problems of bias nonmonotonicity that can arise when non-zero initial values are considered.[3] Assumption 2.2 (spherical errors) may be imposed in simulations without loss of generality, following invariance Theorems 2.1 and 2.2 (proofs of which are available from the authors on request):

Theorem 2.1. Given Assumption 2.1, the bias matrix $\boldsymbol{B} = \mathrm{E}(\widehat{\boldsymbol{A}}_1) - \boldsymbol{A}_1$ of $\widehat{\boldsymbol{A}}_1$ is scalar, and bias is invariant to $\boldsymbol{\Omega}$, for the Models (A)–(C).

Theorem 2.2. Given Theorem 2.1, the variances of each of the diagonal elements of $\widehat{\boldsymbol{A}}_1$ are identical, and variance is invariant to $\boldsymbol{\Omega}$.

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 $\widehat{a}_1$ given Model (A), $k = p = 1$, and with $|a_1| = 1$. The final expression is based upon parabolic cylinder functions, and is computationally very efficient. Abadir further shows that bias may be described rather more simply in terms of exponential functions in polynomials of T^{-1}, and develops the heuristic approximation

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

where -1.7814 is the expected value of the limiting distribution of $T(\widehat{a}_1 - 1)$; e.g. see Le Breton and Pham (1989, p. 562).[4] Heuristic fits have been used elsewhere in the literature, e.g. Dickey and Fuller (1981, p. 1064), and we distinguish here between "heuristic" approximations and the response surface approach used in this paper. Despite the fact that only 5 datapoints are used in the derivation of (2), it is accurate in-sample to 5 decimal places

[3] 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 $\widehat{a}_1$ 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 (A) by Lawford (2001), with $k \leq 6$, $p = 1$ and $\boldsymbol{x}_0 \neq \mathbf{0}_k$, leads to the interesting conjecture that bias nonmonotonicity also disappears when $k > 1$.

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

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 bias and 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.

More recently, AHT consider Model (A), $k \geq 1$, $p = 1$, and prove that $\boldsymbol{B}$ is exactly a *scalar matrix*, i.e. diagonal with equal diagonal elements: $\boldsymbol{B} = \text{diag}(b, \ldots, b)$, and that $\boldsymbol{B}$ is not a function of $\boldsymbol{\Omega}$. Furthermore, they develop a simple approximation to *multivariate* finite-sample bias (especially AHT, p. 166, and Abadir, 1995a, p. 264):

$$\boldsymbol{B}^{\text{AHT}} \approx b^{\text{UNIV}} k \boldsymbol{I}_k \equiv b^{\text{AHT}} \boldsymbol{I}_k, \tag{3}$$

for $T > k + 2$ (this existence requirement varies with the density). It is clear that bias is invariant to $\boldsymbol{\Omega}$, and is approximately proportional to the dimension of the VAR, even when $\boldsymbol{\Omega} = \boldsymbol{I}_k$. To facilitate a subsequent discussion of cointegrating relations, AHT formulate their maintained model as $\triangle \boldsymbol{x}_t = \boldsymbol{C}\boldsymbol{x}_{t-1} + \boldsymbol{\varepsilon}_t$, where the difference operator satisfies $\triangle \boldsymbol{x}_t = \boldsymbol{x}_t - \boldsymbol{x}_{t-1}$, and $\boldsymbol{C} \equiv \boldsymbol{A}_1 - \boldsymbol{I}_k$. Since the bias of $\widehat{\boldsymbol{C}}$ is equivalent to the bias of $\widehat{\boldsymbol{A}}_1$, our results may be directly compared to those in AHT.

Abadir (1995a, 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 $\widehat{\alpha}_1$ 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:

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

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 $\widehat{\boldsymbol{A}}_1$ are identical, we may use $\text{MSE}(\widehat{\alpha}_1) = b^2 + v$ to compute the mean squared error.

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) for an introduction. Briefly, a statistic τ is modelled as a function (response surface) $\Psi(\cdot)$ of relevant variables, that is usually formulated in line with known analytical results. Monte Carlo simulation is used to generate estimates τ_n^* of τ, $n = 1, 2, \ldots, N$, based upon $M(n)$ replications respectively; and $\widehat{\Psi}(\cdot)$ is estimated using an appropriate procedure, depending upon the form chosen for $\Psi(\cdot)$. The method can be computationally intensive, since $M(n)$ and (especially) N must be large if $\widehat{\Psi}(\cdot)$ is to be accurately specified. To avoid problems of specificity, $\widehat{\Psi}(\cdot)$ must be subjected to diagnostic testing, and its out-of-sample performance assessed.

3.1 Monte Carlo design and simulation

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

$$T \in \{25, 30, \ldots, 80, 90, 100, 150, 200\}, \ k \in \{1, 2, 3, 4\}, \ p \in \{1, 2, 3, 4\}, \quad (5)$$

giving 256 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. We calculate the OLS estimate for each combination of (T, k, p) in the parameter space, from which we directly derive the bias. Since $\boldsymbol{B}$ is a scalar matrix, we may estimate the scalar b by averaging over the estimated diagonal elements of $\boldsymbol{B}$. This results in a further reduction in the number of replications needed for a given accuracy as k increases.[5] We simulate variance v similarly.

[5] We experimented with a pseudo-antithetic variate technique, based upon Abadir's (1995b) univariate "AV4", and were able to increase the speed of the 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.

We generate random numbers from the $\mathrm{N}(0,1)$ using a similar technique to MacKinnon's (1994, p. 170) long-period algorithm. We create two independent series of $\mathrm{U}_{(0,1)}$ pseudorandom numbers $\{u_j\}$ and $\{v_j\}$, using multiplicative linear congruential uniform generators due to L'Ecuyer (1988). We then generate a single series of $\mathrm{N}(0,1)$ variates using Marsaglia and Bray's (1964, p. 263, correcting for a typo in point 4) mixture and rejection method. The period of our procedure is much larger than our total random number requirement. All simulations were performed on a Pentium 4 machine, with a 2GHz processor and 256MB of RAM, running GAUSS under Microsoft Windows XP.

Wherever 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 $2^{-1/2}T\,(\widehat{\alpha}_1 - \alpha_1)$ 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).

3.2 Post-simulation analysis

We regressed the Monte Carlo estimates of bias and variance under Models (A)–(C) on functions of sample size, VAR dimension and lag-order, to reflect the dependence of b and v upon these parameters, and the degree of overparameterization. Following extensive experimentation, we chose to fit the following nonlinear bias response surface for each of the models:[6]

$$\begin{aligned} b\,(T_i,k_i,p_i) \quad &= \quad \left(\beta_1 + \beta_2\,k_i + \beta_3\,p_i + \beta_4\,p_i^2 + \beta_5\,k_i^3 p_i^6\right) T_i^{-1} \\ &\quad\quad \times \exp\left[\left(\beta_6 + \beta_7\,k_i + \beta_8\,p_i + \beta_9\,p_i^2 + \beta_{10}\,k_i p_i^2\right) T_i^{-1}\right] + u_i. \end{aligned} \tag{6}$$

The dependent variable $b\,(T_i,k_i,p_i)$ is the simulated finite-sample bias with sample size T_i, VAR dimension k_i, and lag-order p_i, which take values from

[6] Some early motivation for numerical refinement of (3), when $p = 1$, 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 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 (3).

(5); u_i is an error term. We denote the estimated response surface by b^{RS}, and estimated coefficients are reported in Table 1. Convergence of the nonlinear least squares routine was very fast, and required few iterations. Selection criteria included small residual variance, parsimony, and satisfactory diagnostic performance. The response surface fits are very good, and the diagnostic test statistics are insignificant at the 5% level. The sign of each estimated coefficient remains the same across the models. Moreover, the absolute value of each estimated coefficient changes monotonically, as additional deterministic terms are included in the estimated model. The estimated coefficient $\widehat{\beta}_2$ does not change significantly across the models.

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 $(T, k) = (25, 8)$, the absolute bias of each of the estimated parameters on the diagonal of $\widehat{\boldsymbol{A}}_1$ 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 $p \neq 1$, and also when additional deterministics are included. 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 $T \to \infty$ in Tb^{RS}, from (6), which gives $Tb^{\mathrm{RS}} = \sum_{i=1}^{5} \widehat{\beta}_i$ of approximately -1.7, -5.2 and -9.7 in Models (A), (B) and (C) respectively.

Kiviet and Phillips (2003, equation (2.19), and Figure 2.1) consider univariate Model (B), where the DGP 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 DGP, while $g_1(T)$ appears as the bias increment due to non-zero drift. Our equation (6) simplifies (when $k = p = 1$) to $g_0(T) \approx -5.1577\, T^{-1} \exp\left(-2.3134\, T^{-1}\right)$, which provides a convenient means of calculating $g_0(T)$ without further simulations.

Using (4) as motivation, we fit the variance response surface

$$\begin{aligned} v_i\left(T_i, k_i, p_i\right) \;=\; & \left(\gamma_1 + \gamma_2\ k_i + \gamma_3\ k_i^2 + \gamma_4\ p_i + \gamma_5\ p_i^2 + \gamma_6\ p_i^3\right) T_i^{-2} \\ & \times \exp[\left(\gamma_7 + \gamma_8\ p_i + \gamma_9\ p_i^2 + \gamma_{10}\ p_i^3 + \gamma_{11}\ k_i p_i\right) T_i^{-1} + \\ & \left(\gamma_{12} + \gamma_{13}\ p_i + \gamma_{14}\ p_i^2 + \gamma_{15}\ p_i^3 + \gamma_{16}\ k_i^2 p_i^2\right) T_i^{-2}] + u_i, \end{aligned} \tag{7}$$

where $v\left(T_i, k_i, p_i\right)$ is the simulated finite-sample variance. 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 remains the same across the models, although their absolute values do not change monotonically. While the RESET4 and RESET2 tests give conflicting results, v^{RS} provides a good approximation across the parameter space (5), and we note that no variance approximations were previously available for overparameterized models, excess lags, or $k > 1$. Equations (6) and (7) may be combined to give an approximation to MSE. 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 and k, for Models (A)–(C), and $p \in \{1, 2, 3, 4\}$.

Bias and variance are not the only criteria to be used in comparison of time series estimates, and the mean squared error, $\mathrm{MSE}(\widehat{\alpha}_1) = b^2 + v$, is often of interest. For univariate Model A ($p = 1$), Abadir (1995a) defines λ as a correction factor such that $\mathrm{MSE}(\lambda\widehat{\alpha}_1)$ is minimized, b^m and v^m as the bias and variance of the corrected OLS estimator $\lambda\widehat{\alpha}_1$, and shows that

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

when $\alpha_1 = 1$. We are now in a position to substitute simulated values for bias and variance into (8) in order to calculate λ for various T, k, p. As an illustration, correction factors are reported in Table 4, for $p = 1$, which displays qualitatively similar results to those in Abadir (1995a, Tables 2,3). 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 completely reversed following the minimum MSE correction; and that MSE efficiency is generally decreasing in T and k.

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 overparameterized VAR(p), for small sample sizes, and various VAR dimensions and lag lengths. Although the univariate framework has been the subject of much previous research in econometrics, a comprehensive multivariate simulation study has not previously been performed. We estimate computationally convenient response surfaces for bias and variance, that are generally much more accurate than existing approximations. 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 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.

Acknowledgements The authors are indebted to Karim Abadir, for his insightful comments. For helpful feedback and discussions, we are grateful to two anonymous referees, whose comments greatly improved the paper; Walter Distaso, Les Godfrey, Jan Magnus, and Patrick Marsh; and seminar participants at Brunel, Liverpool, Tilburg and York. We also thank participants at the Econometric Study Group Meeting (Bristol, July 2002) and the 6th ERC/METU Conference in Economics (Ankara, September 2002), 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). Steve Lawford is very grateful to Marc Hallin and ECARES for their hospitality (2000–2002). This paper was typed in Scientific Word and numerical results were derived using GAUSS and E–Views.

References

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

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

3. Abadir, K. M., 1995b. On efficient simulations in dynamic models, Discussion Paper in Economics 95/21, University of Exeter.

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

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

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

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

8. Abadir, K. M., Magnus, J. R., 2002. Notation in econometrics: A proposal for a standard, Econometrics Journal 5, 76-90.

9. Andrews, D. W. K., 1993. Exactly median-unbiased estimation of first order autoregressive/unit root models, Econometrica 61, 139-165.

10. 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.

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

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

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

14. 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.

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

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

17. Hendry, D. F., 1984. Monte Carlo experimentation in econometrics, in: Griliches, Z., Intriligator, M. D. eds., Handbook of Econometrics, Volume 2 (North-Holland, Amsterdam) 937-976.

18. Jarque, C. M., Bera, A. K., 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals, Economics Letters 6, 255-259.

19. Kiviet, J. F., Phillips, G. D. A., 2003. Moment approximations for least squares estimators in dynamic regression models with a unit root, Discussion Paper 2003/03, Department of Quantitative Economics, Universiteit van Amsterdam.

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

21. Le Breton, A., Pham, D. T., 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.

22. L'Ecuyer, P., 1988. Efficient and portable combined random number generators, Communications of the ACM 31, 742-749, 774.

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

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

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

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

27. 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.

28. Marsaglia, G., Bray, T. A., 1964. A convenient method for generating normal variables, SIAM Review 6, 260-264.

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

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

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

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

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

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

35. Ramsey, J. B., Schmidt, P., 1976. Some further results on the use of OLS and BLUS residuals in specification error tests, Journal of the American Statistical Association 71, 389-390.

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

37. 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.

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

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

40. Tsui, A. K., Ali, M. M., 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.

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

42. White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, Econometrica 48, 817-838.

Table 1			
Estimated bias response surfaces b^{RS} (6) for Models A, B and C[7]			
	Model A	Model B	Model C
$\widehat{\beta}_1$	-0.5920	-4.8260	-11.1301
	(0.0485)	(0.1069)	(0.2428)
$\widehat{\beta}_2$	-1.9972	-1.9827	-1.9541
	(0.0094)	(0.0188)	(0.0399)
$\widehat{\beta}_3$	1.0400	1.9973	4.1048
	(0.0497)	(0.0964)	(0.2150)
$\widehat{\beta}_4$	-0.1750	-0.3463	-0.7001
	(0.0101)	(0.0199)	(0.0442)
$\widehat{\beta}_5$	1.95×10^{-6}	2.64×10^{-6}	4.35×10^{-6}
	$\left(2.09 \times 10^{-7}\right)$	$\left(3.88 \times 10^{-7}\right)$	$\left(7.85 \times 10^{-7}\right)$
$\widehat{\beta}_6$	-1.6710	-3.5992	-4.4288
	(0.4850)	(0.5476)	(0.7987)
$\widehat{\beta}_7$	-1.1296	-1.3918	-1.4934
	(0.1034)	(0.1289)	(0.1596)
$\widehat{\beta}_8$	1.3006	3.3222	3.9751
	(0.3647)	(0.4163)	(0.7417)
$\widehat{\beta}_9$	-0.5663	-0.9621	-1.0562
	(0.0744)	(0.0895)	(0.1566)
$\widehat{\beta}_{10}$	0.3173	0.3175	0.3230
	(0.0114)	(0.0138)	(0.0189)
$\overline{R}^2$	0.9995	0.9992	0.9981
RSS	0.000523	0.001403	0.006277
JB	$\chi^2(2) = 2.08$	$\chi^2(2) = 4.69$	$\chi^2(2) = 1.84$
RESET4	$\mathrm{F}(3, 243) = 0.78$	$\mathrm{F}(3, 243) = 1.89$	$\mathrm{F}(3, 243) = 0.39$
RESET2	$\mathrm{F}(1, 245) = 1.41$	$\mathrm{F}(1, 245) = 1.94$	$\mathrm{F}(1, 245) = 1.18$

[7] Response surfaces (6) were estimated using nonlinear least squares, in E-Views. White's (1980) heteroscedasticity-consistent standard errors are given in parentheses. $\overline{R}^2$ is the degrees-of-freedom adjusted coefficient of determination. RSS is the residual sum of squares. JB is the Jarque-Bera (1980) test for normality, asymptotically distributed as $\chi^2(2)$. RESET4 and RESET2 are Ramsey-Schmidt (1976) tests for omitted variables/correct functional form, distributed as $\mathrm{F}(s, N-K-s)$, where $s+1 = 4, 2$ is the highest power of fitted bias included in the auxiliary regression, $N = 256$ and $K = 10$.

Table 2									
Simulated and estimated scaled bias in Models A, B, C ($p = 1$)[8]									
		VAR dimension (k)							
T		1	2	3	4	5	6	7	8
25	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)
	b^{RS}	[6.4]	[13.4]	[20.0]	[26.1]	[31.8]	[37.1]	[42.1]	[46.7]
	$\bar{b}$	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
50	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)
	b^{RS}	[3.3]	[7.1]	[10.7]	[14.2]	[17.6]	[20.9]	[24.0]	[27.1]
	$\bar{b}$	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
100	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)
	b^{RS}	[1.7]	[3.6]	[5.5]	[7.4]	[9.2]	[11.0]	[12.8]	[14.6]
	$\bar{b}$	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
200	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)
	b^{RS}	[0.9]	[1.8]	[2.8]	[3.8]	[4.7]	[5.7]	[6.6]	[7.6]
	$\bar{b}$	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
400	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)
	b^{RS}	[0.4]	[0.9]	[1.4]	[1.9]	[2.4]	[2.9]	[3.4]	[3.9]
	$\bar{b}$	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
800	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)
	b^{RS}	[0.2]	[0.5]	[0.7]	[1.0]	[1.2]	[1.5]	[1.7]	[1.9]
	$\bar{b}$	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

[8] All reported bias values have been multiplied by -100; b : simulated Model A bias; b^{AHT} : AHT approximation (3) to Model A bias; b^{RS} : response surface approximation (6) to Model A bias; $\bar{b}$: simulated Model B bias; $\widetilde{b}$: simulated Model C bias.

Table 3			
Estimated variance response surfaces v^{RS} (7) for Models A, B and C[9]			
	Model A	Model B	Model C
$\widehat{\gamma}_1$	-577.3455	-475.2821	-599.9928
	(67.9299)	(64.6035)	(66.6310)
$\widehat{\gamma}_2$	14.8429	11.8024	11.8949
	(1.5446)	(1.3549)	(1.3234)
$\widehat{\gamma}_3$	-1.2515	-0.8579	-0.9851
	(0.2118)	(0.2323)	(0.2337)
$\widehat{\gamma}_4$	835.9746	699.5872	926.2761
	(110.7923)	(104.9714)	(107.7841)
$\widehat{\gamma}_5$	-293.3134	-237.3200	-338.4298
	(49.6442)	(46.8486)	(47.7701)
$\widehat{\gamma}_6$	33.2823	26.3554	39.3300
	(6.6201)	(6.2138)	(6.3091)
$\widehat{\gamma}_7$	204.7508	156.7636	195.9282
	(33.4742)	(30.3960)	(27.0088)
$\widehat{\gamma}_8$	-317.0862	-248.6802	-311.9480
	(47.0654)	(44.2159)	(41.2558)
$\widehat{\gamma}_9$	111.6797	84.8254	113.3971
	(19.8419)	(18.8541)	(17.9374)
$\widehat{\gamma}_{10}$	-12.7168	-9.4667	-13.1610
	(2.5656)	(2.4405)	(2.3538)
$\widehat{\gamma}_{11}$	0.9210	0.6972	0.6001
	(0.1081)	(0.0988)	(0.1064)
$\widehat{\gamma}_{12}$	-2580.397	-1926.864	-2996.315
	(582.2628)	(554.1242)	(497.8064)
$\widehat{\gamma}_{13}$	3804.827	2807.966	4422.204
	(829.8897)	(808.5275)	(764.8489)
$\widehat{\gamma}_{14}$	-1361.316	-955.2007	-1646.161
	(353.6673)	(345.7677)	(334.5968)
$\widehat{\gamma}_{15}$	157.2569	107.3854	194.8083
	(45.9910)	(44.8123)	(44.0648)
$\widehat{\gamma}_{16}$	1.4373	1.4001	1.6215
	(0.1178)	(0.1190)	(0.1350)
$\overline{R}^2$	0.9970	0.9974	0.9979
RSS	0.000260	0.000226	0.000233
JB	$\chi^2(2) = 5.37$	$\chi^2(2) = 8.56^{*}$	$\chi^2(2) = 1.74$
RESET4	$\mathrm{F}(3,237) = 21.20^{**}$	$\mathrm{F}(3,237) = 29.21^{**}$	$\mathrm{F}(3,237) = 25.59^{**}$
RESET2	$\mathrm{F}(1,239) = 3.73$	$\mathrm{F}(1,239) = 3.22$	$\mathrm{F}(1,239) = 1.03$

[9] * and ** indicate that the statistic is significant at the 5% and 1% levels respectively.

Table 4									
Minimum MSE correction in Model A ($p = 1$)[10]									
		VAR dimension (k)							
T		1	2	3	4	5	6	7	8
25	λ	1.05	1.12	1.18	1.25	1.32	1.38	1.44	1.49
	br	0.23	0.24	0.26	0.29	0.32	0.35	0.39	0.44
	vr	1.11	1.25	1.40	1.57	1.74	1.91	2.08	2.23
	bc	24/1	42/3	53/5	61/7	66/10	70/13	73/17	75/20
	me	86	75	69	67	66	66	68	70
50	λ	1.03	1.07	1.11	1.15	1.19	1.23	1.27	1.31
	br	0.11	0.11	0.12	0.12	0.13	0.13	0.14	0.15
	vr	1.06	1.14	1.23	1.31	1.41	1.51	1.62	1.73
	bc	24/0.4	42/1	53/1	61/2	66/2	71/3	74/3	76/4
	me	81	67	58	52	48	46	44	43
100	λ	1.02	1.04	1.06	1.08	1.10	1.12	1.14	1.16
	br	0.06	0.05	0.06	0.06	0.06	0.06	0.06	0.06
	vr	1.03	1.07	1.12	1.16	1.20	1.25	1.30	1.35
	bc	23/0.1	42/0.2	53/0.3	61/0.4	66/0.5	70/0.6	74/0.8	76/0.9
	me	79	62	53	46	41	37	34	32
200	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
	vr	1.02	1.04	1.06	1.08	1.10	1.12	1.14	1.17
	bc	24/0.0	42/0.1	54/0.1	60/0.1	66/0.1	71/0.1	74/0.1	76/0.2
	me	77	60	49	43	37	33	30	28
400	λ	1.00	1.01	1.01	1.02	1.02	1.03	1.03	1.04
	br	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	vr	1.01	1.02	1.03	1.04	1.05	1.06	1.07	1.08
	bc	21/0.0	40/0.0	51/0.0	59/0.0	65/0.0	70/0.0	73/0.0	76/0.1
	me	80	61	51	43	37	32	29	26
800	λ	1.00	1.00	1.01	1.01	1.01	1.02	1.02	1.02
	br	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	vr	1.00	1.01	1.01	1.02	1.02	1.03	1.03	1.04
	bc	17/0.0	45/0.0	49/0.0	63/0.0	64/0.0	71/0.0	72/0.0	77/0.0
	me	84	55	51	38	37	29	29	24

[10] λ : correction factor, such that $\lambda\widehat{\alpha}_1$ attains minimum MSE; br: bias ratio≡corrected bias/OLS bias; vr: variance ratio≡corrected variance/OLS variance $\left(\text{vr} \equiv \lambda^2\right)$; bc: "$x/y$" indicates that b^2 forms x% of MSE under OLS, and *corrected* b^2 forms y% of *minimized* MSE; me: MSE efficiency≡MSE following correction/MSE under OLS.

Figure 1: scaled bias response surface approximations, b^{RS}

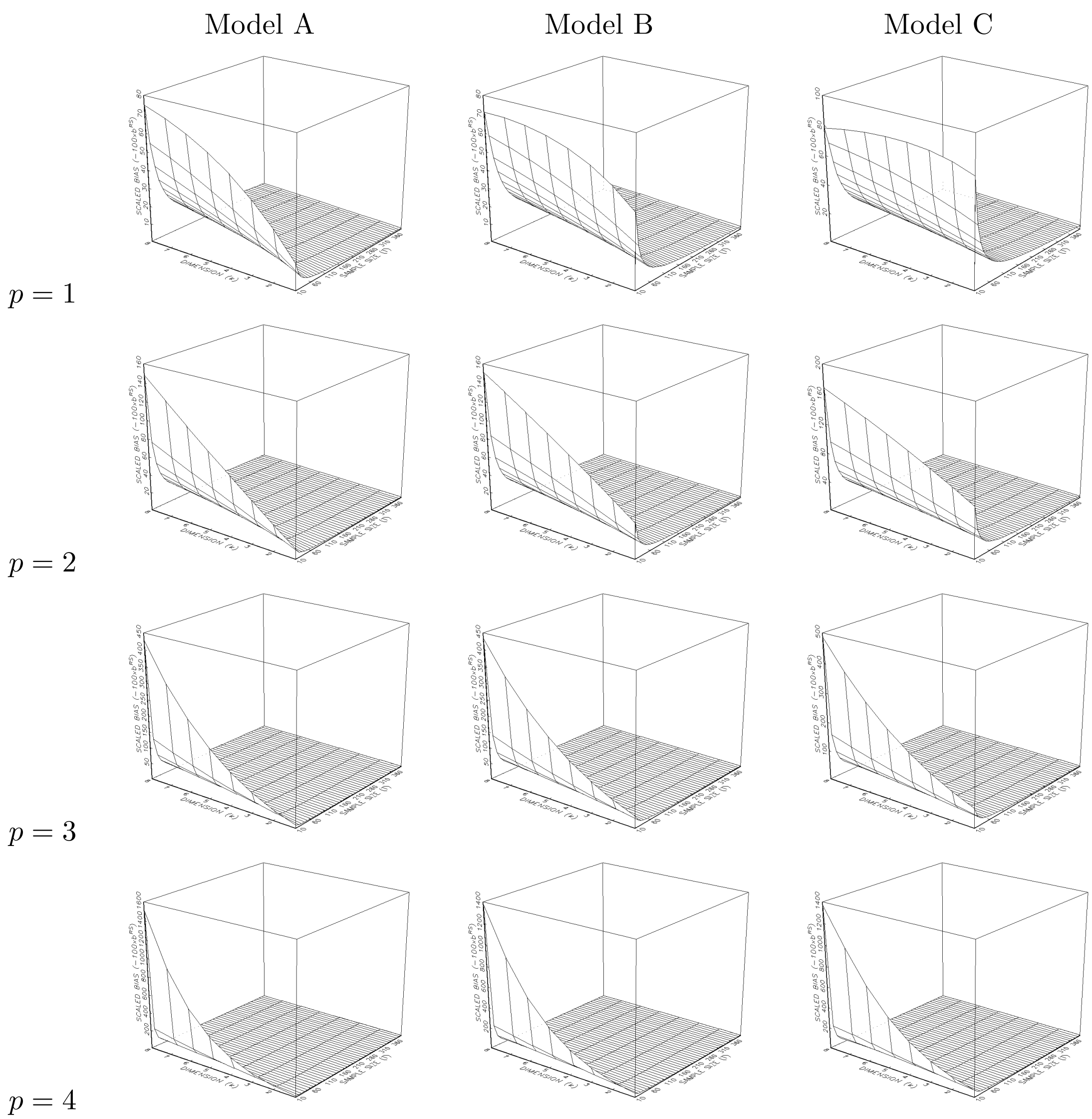

Figure 2: scaled variance response surface approximations, v^{RS}