Bias Approximation and Reduction in Vector Autoregressive Models

Tomas Brännström

Dissertation for the Doctor's
Degree in Philosophy

Department of Economic Statistics
Stockholm School of Economics

STOCKHOLM SCHOOL
OF ECONOMICS
THE ECONOMIC RESEARCH INSTITUTE

ISBN 91-7258-405-X

Keywords:

vector autoregression

bias approximation

bias reduction

overparameterization

unit roots

cointegration

Stockholm 1995

ACKNOWLEDGEMENTS

I am indebted to a number of people without whose help and support this Ph.D. thesis could never have been completed or would have been completed at a much later date. Suffice it to mention Sune Karlsson, who initiated this research project and has been fuelling it with new ideas all along. Perhaps more importantly, he has always encouraged and helped me develop my own ideas when promising, while advising me not to pursue less tenable ideas. The relatively efficient research process behind the present thesis can be seen as a result of this very valuable impetus and feedback. At the same time I wish to emphasise that he is in no way responsible for any remaining errors.

Financial support from the Swedish Council for Research in the Humanities and Social Sciences (HSFR) and from the Economic Research Institute (EFI) at the Stockholm School of Economics is gratefully acknowledged.

The Monte Carlo simulations on which most of the results are based were carried out using RATS 4.01. Some of the results were also manipulated using GAUSS 3.01. All graphs in Chapters 3 and 4 were generated in the Sygraph module of Systat 5.02, whereas the figures in Chapter 5 are SPSS 6.0 graphs. EViews 1.1 provided the Johansen LR test statistics in Chapter 5.

Tomas Brännström

Contents

1. Introduction

In the last few decades, vector autoregressive (VAR) models have gained tremendous popularity as an all-purpose tool in econometrics and other disciplines. Some of their most prominent uses are for forecasting, causality tests, tests of economic theories, hypothesis-seeking, data characterisation, innovation accounting, policy analysis, and cointegration analysis (Cooley and LeRoy 1985). Their popularity appears to be attributable to their flexibility relative to other models rather than to their virtues *per se*. In addition, analysts often use VAR models as benchmark models.

Not surprisingly, numerous comparisons have been made between VAR models and their competitors (notably structural models) on simulated as well as real data. Examples include McKnees (1986), Litterman (1986*b*), Boero (1990) and Sarantis and Stewart (1995). Stephen McKnees compared the forecasting accuracy of seventeen different predictors using a real US data set of seven macroeconomic series. So did Robert Litterman, but for four predictors and four economic variables only. Gianna Boero did essentially the same thing for the Italian GDP growth rate, and Nicholas Sarantis and Chris Stewart compared exchange rate forecasts generated by a wide range of different models. The general conclusion appears to be that if there is such a thing as an "ideal" model in a specific situation and that model can be made operational, it is likely to outperform VAR models. In many cases it is necessary to resort to less-than-ideal models, and then it is often better to select VAR models before other compromise solutions. Another general finding seems to be that VAR models may often outperform structural models in the short run but not in the long run.

VAR modelling has not gone uncriticised, though. A list of relevant arguments against VAR modelling can be found in Section 2.3 of this thesis. There is one additional problem which is rarely mentioned though, namely the often heavily biased estimates in VAR models. Although methods to reduce this bias have been available for quite some time, it has probably not been done before, at least not in any systematic way. The present thesis attempts to systematically examine the performance of bias-reduced VAR estimates relative to unadjusted estimates, using two existing and one newly derived approximation to the bias.

The thesis is organised as follows. A brief history of VAR modelling can be found in Chapter 2 together with a review of different representations and a compilation of criticisms against VAR models. Chapter 3 reports the results of very extensive Monte Carlo experiments serving dual purposes: Firstly, the simulations will reveal whether or not bias really poses a serious problem, because if it turns out that biases appear only by exception or are mainly insignificant, there would be little need to reduce the bias. Secondly, the same data as in Chapter 3 will be used in Chapter 4 to evaluate the bias approximations, allowing for direct comparison between bias-reduced and original estimates. Though Monte Carlo methods have been (rightfully) criticised for being too specific to allow for any generalisation, there seems to be no good alternative to analyse small-sample properties of complicated estimators such as these.

Chapter 4 is in a sense the core of the thesis, containing evaluations of three bias approximations. The performance of the bias approximations is evaluated chiefly using single regression equations and 3D surfaces. The only truly new research result in this thesis can also be found in Chapter 4; a second-order approximation to the bias of the parameter matrix in a VAR(p) model. Its performance is compared with the performance of two existing first-order approximations, and all three are used to construct bias-reduced estimators, which are then evaluated.

Chapter 5 holds an application of US money supply and inflation in order to find out whether the results in Chapter 4 can have any real impact. Unfortunately though, bias reduction appears not to make any difference in this particular case. Chapter 6 concludes.

2. Vector Autoregressive Models

2.1 A Brief Historical Background

The close resemblance between VAR models and reduced form equation systems is no coincidence. In fact, the reason VAR models emerged as a serious alternative to simultaneous equation models in the first place was the growing scepticism toward the latter from the 1950's and on, particularly following the failure of large-scale econometric models to predict the economic consequences of the price shocks in the 1970's. Several scholars concluded that the main reason the systems of simultaneous equations failed was that in order for them not to be underidentified, assumptions had to be made and restrictions imposed which are not always easily justified and which may lead to wrong results if incorrect. Furthermore, the reduced form of the system will be more or less the same no matter which restrictions have been imposed: All variables will enter each and every equation of the reduced form even if removed from the structural form for reasons of (over)identification. So why not avoid the unrealistic assumptions and restrictions altogether and move directly to a reduced forms system in which each endogenous variable enters every equation, since that is where one ends up anyway. As Christopher Sims, often referred to as the "inventor" of VAR modelling, declared in 1980, "[---] large-scale models do perform useful forecasting and policy-analysis functions despite their incredible identification; [but] the restrictions imposed in the usual style of identification are neither essential to constructing a model which can perform these functions nor innocuous; an alternative style of identification is available [---]" (Sims 1980*a*). Other pioneers of early VAR modelling include names like Thomas Sargent, William Phillips and Cheng Hsiao. Sargent and Sims (1977) argued in favour of so-called index models (which are essentially just VAR models) when the objective is to benefit from the coherence between almost all economic variables; Sargent (1979*a*) formalised the VAR concept mathematically; Sargent (1979*b*) set up the framework for VAR analysis; Phillips (1978) derived maximum likelihood estimates for different VAR models; and Hsiao (1979) devised a procedure for identification of VAR models. Finally, in these circumstances it would be unjust not to mention Henry Mann and Abraham Wald, who proposed in a paper as early as

in 1943 that without additional information, the only justifiable model is a system of linear autoregressive equations, i.e. a vector autoregression (Mann and Wald 1943).

The "alternative style" Sims (1980*a*) was referring to was unrestricted VAR models, but it was not long before they too turned out to require restrictions and assumptions in order to be informative and give good results not only within the sample but for out-of-sample observations as well. Without such restrictions or assumptions, the in-sample fit is likely to be good but the out-of-sample fit poor due to a lack of degrees of freedom, since the number of parameters to estimate in an unrestricted VAR rapidly grows out of proportion and, hence, several parameters will not be accurately estimated. Another way to express this is that without restrictions, posterior likelihood distributions will be very flat. The only way to overcome this problem of overparameterization is to estimate VAR models subject to restrictions or using Bayesian priors. Important advances with respect to Bayesian VAR modelling have been made by Robert Litterman, Thomas Doan and others. Litterman (1986*a*) presented the technique; Litterman (1986*b*) evaluated its predictive performance as compared to unrestricted VAR models; and Doan *et al.* (1984) introduced the so-called Minnesota prior distribution to handle nonstationary processes.

Unrestricted as well as restricted VAR models rapidly gained massive popularity among empirical economists testing theories and hypotheses in economics, for instance in order to test hypotheses of rational expectations (e.g. Keating 1990). In recent years they have also emerged as the chief instrument for cointegration analysis. James Stock, Mark Watson, Robert Engle, Clive Granger, Anders Warne and Søren Johansen are just some of the many scholars behind important research into integrated and cointegrated systems. Engle and Granger (1987) introduced the concept, derived least-squares estimates of cointegrating vectors and suggested a simple test; Stock (1987) derived the asymptotic properties of estimated cointegrating vectors, Stock and Watson (1988) and Warne (1990) derived tests for so-called common trends; and Johansen (1988) derived ML estimates of cointegrating vectors. Cointegrated and other unit-root systems will be treated briefly in Section 3.3.

2.2 Different VAR Representations

Vector autoregressive models can be represented in a number of ways depending on the application and the preferences of the user. Here are some of the guises in which they occur most frequently:

The *VAR(p) representation* $\mathbf{x}_t = \nu + \mathbf{A}_1\mathbf{x}_{t-1} + \cdots + \mathbf{A}_p\mathbf{x}_{t-p} + \varepsilon_t$ forms the basis for all other representations. Here $\mathbf{x}_t$ is a $K\times 1$ vector of endogenous variables, where K is the dimension of the system; ν is a $K\times 1$ vector of constants; $\mathbf{A}_j$ is the jth $K\times K$ matrix of parameters; p is the autoregressive order; and ε_t is a $K\times 1$ vector of errors (also known as disturbances or shocks). The variance-covariance matrix of ε_t is $K\times K$ and will be referred to as $\mathbf{G}$ later on. Estimation of the model requires data for the time periods $t = 1,...,T$, where T is the sample size, and start-up values for $t = -p+1,..,0$.

If the disturbance vector ε_t is also allowed to have a lag structure of order q the model will not be VAR(p) but VARMA(p,q), and if in addition there are exogenous variables $\mathbf{Z}_t$ affecting the endogenous variable vector $\mathbf{x}_t$, the result will be a VARMAX model. Though only VAR(p) models will be studied here, there is every reason to expect the results to extend to VARMAX(p,0) as well (i.e. to VAR models with exogenous variables), but probably not as far as to VARMA or VARMAX models in general.

Every VAR(p) model can also be expressed as a VAR(1) model $\mathbf{x}_t^* = \nu^* + \mathbf{A}^*\mathbf{x}_{t-1}^* + \varepsilon_t^*$ by stacking $\mathbf{x}_t$ on $\mathbf{x}_{t-1}$, $\mathbf{x}_{t-2}$,...,$\mathbf{x}_{t-p+1}$ and $\mathbf{x}_{t-1}$ on $\mathbf{x}_{t-2}$, $\mathbf{x}_{t-3}$,...,$\mathbf{x}_{t-p}$ to form $\mathbf{x}_t^*$ and $\mathbf{x}_{t-1}^*$, respectively; stacking ν and ε_t on zero vectors to obtain ν^* and ε_t^*; and collecting all the $\mathbf{A}$ matrices in $\mathbf{A}^*$ together with an identity matrix and a zero matrix in the following manner:

$$\begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \\ \vdots \\ \mathbf{x}_{t-p+1} \end{bmatrix} = \begin{bmatrix} \nu \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} + \left[\begin{array}{ccc:c} \mathbf{A}_1 & \cdots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \hdashline & & & \mathbf{0} \\ & \mathbf{I} & & \vdots \\ & & & \mathbf{0} \end{array}\right] \begin{bmatrix} \mathbf{x}_{t-1} \\ \mathbf{x}_{t-2} \\ \vdots \\ \mathbf{x}_{t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$$

This VAR(1) representation will be used extensively in the chapters to follow.

Closely related to the VAR(p) representation is the *lag polynomial representation* $(\mathbf{I}-\mathbf{A}(L))\mathbf{x}_t = \nu + \varepsilon_t$ where L is the lag operator and $\mathbf{A}(L)$ is a $K\times K$ matrix polynomial of order p. Provided that $\mathbf{I}-\mathbf{A}(L)$ has an inverse, a *vector moving average representation* also exists. In its lag polynomial form it will be $\mathbf{x}_t = (\mathbf{I}-\mathbf{A}(L))^{-1}(\nu + \varepsilon_t) = \nu^* + \mathbf{B}(L)\varepsilon_t$ where $\mathbf{B}(L)$, the inverse of $\mathbf{I}-\mathbf{A}(L)$, is a $K\times K$ matrix polynomial of possibly infinite order and ν^* is the new constant vector $\mathbf{B}(L)\nu$. These two representations will not be used in the chapters to follow.

The *error correction representation* (ECM) is frequently used for cointegration analysis. It is obtained by taking differences throughout the standard VAR(p) model introduced above; $\Delta \mathbf{x}_t = \nu + \mathbf{D}_1 \Delta \mathbf{x}_{t-1} + \cdots + \mathbf{D}_{p-1} \Delta \mathbf{x}_{t-p+1} + \Pi \mathbf{x}_{t-1} + \varepsilon_t$ for instance for a first-order integrated process, where $\Delta \mathbf{x}_t = \mathbf{x}_t - \mathbf{x}_{t-1}$ and the new parameter matrix Π is the product of the loading matrix and the matrix of cointegrating vectors. Unit-root and cointegrated systems will be briefly discussed in Section 3.3, however without use of the error correction representation. Simulation studies by James LeSage (LeSage 1990) suggest that ECM models outperform VAR models in the presence of cointegration.

Occasionally it is desirable to write the VAR model on a form in which the errors are uncorrelated, i.e. in which the variance-covariance matrix of the errors is diagonal. Cholesky decomposition of the variance-covariance matrix $\mathbf{G}$ into $\mathbf{PP}'$, where $\mathbf{P}$ is a triangular matrix, is but one of many possible ways to achieve this. Premultiplying the original VAR(p) model by $\mathbf{P}^{-1}$ produces a new VAR(p) model with orthogonal errors. Even though this *orthogonal representation* forms the basis for the computed impulses and responses and for the forecast error decomposition in Chapter 5, it will not be explicitly referred to anywhere in the remainder.

2.3 Criticisms of VAR Methodology

It would be unfair to present vector autoregressive models without also presenting at least some of the points on which they, particularly in their unrestricted form, have been criticised. First of all, they are after all only a set of related linear difference equations, and as such they cannot be more than mere approximations of economic relationships which are far more complicated – nonlinear rather than linear, for instance.

The second point is the overparameterization described in the previous section, frequently leading to much better in-sample than out-of-sample fit, which in turn may seduce analysts to have more faith in their results than they ought to.

A third criticism concerns the interpretation of results, especially with regard to estimated responses to impulses and forecast error variance decompositions. Even though no analyst would seriously propose that a VAR model of fairly low dimension can correctly represent an entire economy in which everything depends on everything else, its estimates are frequently regarded as complete and ultimate.

Whereas this may not be a very big issue for parameter estimates and short-term forecasts, it may seriously distort the interpretation of impulse-response analyses and variance decompositions, both of which are statistical techniques with no connection to underlying economic factors. Variables which are omitted from the model for practical reasons will obviously have no impact on these results even if there is a causal relationship but, as Runkle (1987) demonstrates, marginally altering the specification may lead to radically changed estimated responses and decompositions. Thus the impulse-response functions and variance decompositions should be thought of as instruments for very partial analysis, limited to the variables included in the model and to the chosen model specification.

Another important inadequacy of vector autoregressive models, which incidentally applies to structural models as well, is that no distinction can be made between long-run and short-run dynamics. All results concerning dynamics are based on the estimated variance-covariance matrix, which is of course an average over the entire sample and therefore the modelling of short-term effects in particular will be hazardous. In fact, using the average variance-covariance matrix for such analysis implies an assumption of constancy with respect to the state of the system. Irrespective of how far the system is from its equilibrium and of where it is headed, the effects of shocking one of the variables are assumed to be the same at every instant – a very naïve idea at best.

Another important point, stressed by Tjalling Koopmans and others, is that even if a good fit has been obtained in a VAR model, there is no way to decide whether that fit is due to the suitability of the chosen model or to the regularity of data (cf. Koopmans 1947). This may seem like a matter of little importance, some may even question the difference between the two, but it is a vital point provided it is in the interest of the model user to make statements about the future. As an extreme example, whilst it is possible to model a series perfectly by estimating a polynomial of the same order as the length of the series, such an exercise would reveal nothing about the nature of the series and the analyst would be unable to say anything about its future. Equivalently, if the good fit of an estimated VAR model can be ascribed solely to the flexibility of the model and not at all to data properties, the model is useless for projections into the future. In part, this applies to structural models as well, but since they are (normally) based on some sort of economic theory and assumptions, the policy makers will at least be able to make conditional statements based on the results.

Even though the critique against imposing "unrealistic" restrictions in order to identify the structural model may be justified in many cases, it has the benefit of revealing which assumptions are unrealistic. When confronted with data, some restrictions and assumptions will appear untenable, and in that respect this procedure reduces the ignorance of the analyst, urging him or her to replace the least realistic restrictions by new and more realistic restrictions and assumptions. VAR models, on the other hand, offer no such guidance. As suggested above, they will most often fit well to data, and even when they do not there is no way of knowing how to improve the model. The user will have to respecify the model intuitively or, lacking intuition, at random.

Furthermore, Thomas Cooley and Stephen LeRoy question the use of VAR models for anything else than forecasting, tests for Granger-causality and searching for hypotheses, arguing that VAR results strictly speaking cannot be extended to other models than (observationally equivalent versions of) the estimated model. In their view, estimated VAR models must not be used to test economic theories, nor for innovation analysis, for instance.

To summarise the above points, VAR models have become an all-purpose tool for econometric analysis in spite of their shortcomings. This would not be a major cause for concern provided users took these limitations into account when interpreting their results, but this is rarely the case. In addition, VAR modelling can in certain respects be said to represent an undesirable step away from a deeper understanding of economic realities since it blurs rather than emphasises some aspects of those realities in order to explain others. As Epstein (1987) writes, "the challenge for econometrics will continue to be to devise accurate and powerful statistical tests to test the hypotheses that will be suggested in [more detailed] research. As a return to the special case which minimises the extent to which the more general structural system is able to model such detail, the VAR is a research strategy that is particularly ill-suited for this task of macroeconomics."

Another shortcoming of VAR models highlighted here is the uncertain properties of VAR parameter estimates, more precisely their bias and low efficiency. There are at least two reasons for this neglect; firstly analysts do not want to question their results on such grounds; and secondly, the critique applies to structural models as well. The reduced form of a system of structural equations will, in general, also contain autoregressive elements, causing biased parameter estimates. Here however, at least the bias will be treated as a real problem, moreover one that requires a remedy.

3. Monte Carlo Experiments

In this chapter the results of extensive Monte Carlo experiments will be reproduced. These experiments were carried out for two reasons; firstly, to assess the magnitude of the bias and variance and to investigate their dependence on eigenvalues, sample sizes, etc.; and secondly, to evaluate the performance of some asymptotic approximations to the bias. Whereas the evaluation of the bias approximations can be found in Chapter 4, the present chapter describes the experiments and the main results of the simulations.

3.1 Experiment Design

Two bivariate VAR models, $\mathbf{x}_t = \nu + \mathbf{A}_1\mathbf{x}_{t-1} + \varepsilon_t$ and $\mathbf{x}_t = \nu + \mathbf{A}_1\mathbf{x}_{t-1} + \mathbf{A}_2\mathbf{x}_{t-2} + \varepsilon_t$, where $\mathbf{x}_t$ is a 2×1 vector of variables; ν is a 2×1 vector of constants; $\mathbf{A}_1$ and $\mathbf{A}_2$ are 2×2 parameter matrices; and ε_t is a 2×1 vector of disturbances, have been Monte Carlo simulated for different parameter combinations and different sample sizes. For each of the 10,000 replications the parameter matrices were OLS estimated, and after the final replication means, variances and covariances were computed. The resulting average parameter estimates are taken to represent each estimate's expected value since the number of replicates is so large.

The VAR(1) representations of the two models are,

$$\text{for } p = 1, \quad \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix} + \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix} \tag{1}$$

and,

$$\text{for } p = 2, \quad \begin{bmatrix} x_t \\ y_t \\ x_{t-1} \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ x_{t-2} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ 0 \\ 0 \end{bmatrix} \tag{2}$$

where the elements of the **x** vectors have been defined as x and y rather than x_1 and x_2 in order to simplify the notation. The elements of ν and of ε, however, are indexed by 1 and 2 for the two equations of each system. The vectors **x**, ν and ε should also have indices distinguishing those of (1) from those of (2) since the two systems are different and therefore have different parameters, but since that distinction will never be crucial in the experiments, it is never made. The distinction between the elements of different **A** matrices, on the other hand, will be of interest in the remainder, and therefore the four parameters of **A** of the first-order system (1) are assigned single-digit indices and the eight parameters of **A** of the second-order system (2) double-digit indices. Consequently, a_1 and a_4 are first-order autoregressive parameters, as are a_{11} and a_{22} of (2); a_2 and a_3 of (1) and a_{12} and a_{21} of (2) are first-order cross-term parameters; a_{13} and a_{24} are second-order AR terms, and a_{14} and a_{23} are second-order cross terms.

Apparently, four parameters in the first-order case and eight for the second-order model can vary in the experiments. But these numbers can be reduced to two and four, respectively, by expressing the **A** matrices of (1) and (2) in terms of their eigenvalues (the solutions to the equations $|\lambda\mathbf{I}-\mathbf{A}| = 0$), by first generating data according to one of the following pairs of autoregressive processes:

$$\begin{aligned} z_t &= \rho_1 z_{t-1} + \varepsilon_{Z,t} \\ w_t &= \rho_2 w_{t-1} + \varepsilon_{W,t} \end{aligned} \tag{3}$$

$$\begin{aligned} z_t &= (\rho_{11} + \rho_{12}) z_{t-1} - \rho_{11}\rho_{12} z_{t-2} + \varepsilon_{Z,t} \\ w_t &= (\rho_{13} + \rho_{14}) w_{t-1} - \rho_{13}\rho_{14} w_{t-2} + \varepsilon_{W,t} \end{aligned} \tag{4}$$

where $\varepsilon_{Z,t}$ and $\varepsilon_{W,t}$ are i.i.d. with zero means. The generated z and w series are then transformed as $x_t = \alpha y_t - w_t$ and $y_t = x_t + z_t$ where α is a constant set by the experimenter ($0<\alpha<1$). This way, the relation between (3) and (1) is that ρ_1 and ρ_2 are the eigenvalues of **A** in (1), and likewise the eigenvalues of **A** in (2) are ρ_{11}, ρ_{12}, ρ_{13} and ρ_{14} which can be seen by first substituting y_t into the x_t equation:

$$x_t = \alpha y_t - w_t = \alpha(x_t + z_t) - w_t = \alpha x_t + \alpha z_t - w_t \tag{5}$$

which means that $(1-\alpha)x_t = \alpha z_t - w_t$ so that x_t will be a linear combination of the two generated series. That this holds for y_t as well can be seen by first multiplying the y_t equation by $(1-\alpha)$ and then inserting the expression for $(1-\alpha)x_t$ just obtained:

$$(1-\alpha)y_t = (1-\alpha)x_t + (1-\alpha)z_t = \alpha z_t - w_t + (1-\alpha)z_t = z_t - w_t \tag{6}$$

and so y_t is also just a linear combination of the two generated series.

The implications for the first-order system are that if (3) is inserted into (5) and (6), the following system obtains:

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \frac{1}{1-\alpha}\begin{bmatrix} \rho_2 - \alpha\rho_1 & \alpha\rho_1 - \alpha\rho_2 \\ \rho_2 - \rho_1 & \rho_1 - \alpha\rho_2 \end{bmatrix}\begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \frac{1}{1-\alpha}\begin{bmatrix} \alpha\varepsilon_{Z,t} - \varepsilon_{W,t} \\ \varepsilon_{Z,t} - \varepsilon_{W,t} \end{bmatrix}$$

$$\text{with } \varepsilon_t = \frac{1}{1-\alpha}\begin{bmatrix} \alpha\varepsilon_{Z,t} - \varepsilon_{W,t} \\ \varepsilon_{Z,t} - \varepsilon_{W,t} \end{bmatrix} \sim \left(\mathbf{0}, \frac{1}{(1-\alpha)^2}\begin{bmatrix} \alpha^2\sigma_Z^2 + \sigma_W^2 & \\ \alpha\sigma_Z^2 + \sigma_W^2 & \sigma_Z^2 + \sigma_W^2 \end{bmatrix}\right)$$

where σ_Z^2 and σ_W^2 are the variances of $\varepsilon_{Z,t}$ and $\varepsilon_{W,t}$, respectively. As can be seen, a first-order VAR model formulated in terms of the constant α, the two eigenvalues and the two error variances of the generated processes obtains. It should also be noted that it is a model without constants ν, which must not be the case but is desirable here. If the ν vector in (1) or (2) were to hold non-zero elements, this could easily have been achieved by generating the *z* and *w* series with non-zero constant terms.

For the second-order system, inserting (4) into (5) and (6) yields

$$\begin{bmatrix} x_t \\ y_t \\ x_{t-1} \\ y_{t-1} \end{bmatrix} = \frac{1}{1-\alpha}\begin{bmatrix} \rho_{13}+\rho_{14}-\alpha(\rho_{11}+\rho_{12}) & \alpha(\rho_{11}+\rho_{12}-\rho_{13}-\rho_{14}) & \alpha\rho_{11}\rho_{12}-\rho_{13}\rho_{14} & \alpha(\rho_{13}\rho_{14}-\rho_{11}\rho_{12}) \\ \rho_{13}+\rho_{14}-\rho_{11}-\rho_{12} & \rho_{11}+\rho_{12}-\alpha(\rho_{13}+\rho_{14}) & \rho_{11}\rho_{12}-\rho_{13}\rho_{14} & \alpha\rho_{13}\rho_{14}-\rho_{11}\rho_{12} \\ 1-\alpha & 0 & 0 & 0 \\ 0 & 1-\alpha & 0 & 0 \end{bmatrix}\begin{bmatrix} x_{t-1} \\ y_{t-1} \\ x_{t-2} \\ y_{t-2} \end{bmatrix} + \frac{1}{1-\alpha}\begin{bmatrix} \alpha\varepsilon_{Z,t} - \varepsilon_{W,t} \\ \varepsilon_{Z,t} - \varepsilon_{W,t} \\ 0 \\ 0 \end{bmatrix}$$

$$\text{with } \varepsilon_t = \frac{1}{1-\alpha}\begin{bmatrix} \alpha\varepsilon_{Z,t} - \varepsilon_{W,t} \\ \varepsilon_{Z,t} - \varepsilon_{W,t} \\ 0 \\ 0 \end{bmatrix} \sim \left(\mathbf{0}, \frac{1}{(1-\alpha)^2}\begin{bmatrix} \alpha^2\sigma_Z^2 + \sigma_W^2 & & & \\ \alpha\sigma_Z^2 + \sigma_W^2 & \sigma_Z^2 + \sigma_W^2 & & \\ 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 \end{bmatrix}\right)$$

The resulting model is similar to the first-order system (the zero constant vector is not reproduced here), but whereas the first-order parameters are linear combinations of all eigenvalues (as they are in the first-order system as well), the second-order parameters are linear combinations of the eigenvalue products $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$.

These transformations, while reducing the number of parameters to be varied by half, also have the more important effect of highlighting the eigenvalues of **A** rather than its parameters. This is important because the time-series properties of a system depend on its characteristic roots (or, equivalently, its eigenvalues), and since

different **A** matrices may have identical eigenvalues it is more efficient and more relevant to use the eigenvalues rather than the elements of **A** as the experiment variables. Ideally in a Monte Carlo study such as this, one would also like to allow for different α values and error variances σ_Z^2 and σ_W^2. Due to the scope of such an exercise, however, the possibility of a varying α, σ_Z^2 or σ_W^2 is ruled out by the following setting:

$$\alpha \equiv 0.2$$
$$\begin{bmatrix} \varepsilon_{Z,t} \\ \varepsilon_{W,t} \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} .01 & \\ 0 & .01 \end{bmatrix}\right)$$

The effects of assigning the value .2 to α are to impose a relation 1 to –5 on each pair of cross terms, and, in combination with the value .01 of the two variances σ_Z^2 and σ_W^2, to fix the variance-covariance matrices of the error vectors (and hence the correlation). Specifically, it means that the first-order system will be

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} 1.25(\rho_2 - 0.2\rho_1) & 0.25(\rho_1 - \rho_2) \\ 1.25(\rho_2 - \rho_1) & 1.25(\rho_1 - 0.2\rho_2) \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} 1.25(0.2\varepsilon_{Z,t} - \varepsilon_{W,t}) \\ 1.25(\varepsilon_{Z,t} - \varepsilon_{W,t}) \end{bmatrix} \quad (7)$$

with $\varepsilon_t \sim N\left(\mathbf{0}, \begin{bmatrix} .01625 & \\ .01875 & .03125 \end{bmatrix}\right)$, and it is clear that subject to these restrictions the error correlation will always be $\sqrt{9/13} \approx .83$ in this bivariate system. So no matter how many different eigenvalues and sample sizes are tried, the result will be a partial analysis only, strictly speaking applicable only to bivariate systems in which the correlation is around .83.

For the second-order system the following parameters obtain:

$$\begin{bmatrix} x_t \\ y_t \\ x_{t-1} \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} 1.25(\rho_{13}+\rho_{14}) - 0.25(\rho_{11}+\rho_{12}) & 0.25(\rho_{11}+\rho_{12}-\rho_{13}-\rho_{14}) & 1.25(0.2\rho_{11}\rho_{12}-\rho_{13}\rho_{14}) & 0.25(\rho_{13}\rho_{14}-\rho_{11}\rho_{12}) \\ 1.25(\rho_{13}+\rho_{14}-\rho_{11}-\rho_{12}) & 1.25(\rho_{11}+\rho_{12}) - 0.25(\rho_{13}+\rho_{14}) & 1.25(\rho_{11}\rho_{12}-\rho_{13}\rho_{14}) & 1.25(0.2\rho_{13}\rho_{14}-\rho_{11}\rho_{12}) \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ x_{t-2} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} 1.25(0.2\varepsilon_{Z,t} - \varepsilon_{W,t}) \\ 1.25(\varepsilon_{Z,t} - \varepsilon_{W,t}) \\ 0 \\ 0 \end{bmatrix} \quad (8)$$

with $\varepsilon_t \sim N\left(\mathbf{0}, \begin{bmatrix} .01625 & & & \\ .01875 & .03125 & & \\ 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 \end{bmatrix}\right)$ and again the correlation is $\sqrt{9/13}$.

Each eigenvalue is assigned the values {1.0, .9, .5, .3 .0, –.3, –.5, –.9}, where of course ρ=1.0 is a unit root, causing first- or second-order integration and, possibly, cointegration. Unit roots will be treated expressly in Section 3.3 and complex eigenvalues are not considered at all. The remaining seven eigenvalues are all less than one in absolute terms, hence they lead to stationary systems (but special attention will be paid to the "near-integrated" case of ρ=.9). As a result there are $8^2 = 64$ distinct eigenvalue combinations to simulate in the case of (7), 15 of which include one or two unit roots and 49 of which are stationary. For (8), though, there are not $8^4 = 4096$ distinct eigenvalue combinations since ρ_{11} and ρ_{12} on the one hand, and ρ_{13} and ρ_{14} on the other, always appear in pairs and are interchangeable. Instead, there are $\frac{1}{2} \cdot 8 \cdot 9 = 36$ ways to combine ρ_{11} and ρ_{12} and 36 ways to combine ρ_{13} and ρ_{14}, hence the total number of eigenvalue combinations to simulate will be $36^2 = 1296$ for (8), 512 of which include one or more unit roots and 784 of which are stationary. All 64 first-order combinations have been simulated as well as all 784 stationary second-order combinations, but only 98 of the unit-root processes since they turned out to be not very interesting (and because one of the three bias approximations to be introduced in Chapter 4 turned out to be sensitive to unit roots).

For each eigenvalue combination, z_t and w_t series of lengths 125, 150, 200 and 300 were generated according to (3) or (4). The first hundred observations were then discarded in order to alleviate the effects of start-up values, so the sample sizes are in effect 25, 50, 100 and 200. Transformations (5) and (6) (with $\alpha \equiv .2$) produce series x_t and y_t, the two variables of either (7) or (8), depending on whether data are generated by (3) or (4). When estimating a model, its structure is then assumed to be known, in particular that its dimension K is 2; its order p is 1 or 2; x_t and y_t are the two endogenous variables; and no exogenous variable enters. The constant vector ν is also estimated even though it is a zero vector in (7) and (8). Estimation is done by multivariate least squares, a technique known from econometric theory to produce biased and inefficient but consistent and asymptotically normally distributed estimates in autoregressive systems such as these. Here, these estimates coincide with maximum likelihood estimates (conditional on initial values) because the shocks to (7) and (8) are normally distributed, and consequently the following results apply to ML as well as to least-squares estimates.

For each eigenvalue combination and for each of the four sample sizes, the parameters of $\mathbf{A}$ and ν are estimated in each of the 10,000 replications. In addition, the variance-covariance matrix $\mathbf{G}$ and the vector of eigenvalues ρ are also estimated each time since they enter the three bias approximations to be introduced in Chapter 4.

The three approximations are also computed for each generated data set, and then, after the last of the 10,000 replications, means and variances are computed for the parameter estimates and for the approximations. These are taken to represent expected values since the number of replications is so large. It should also be observed that the variances are computed over the 10,000 estimates rather than as the average of the 10,000 estimated variances, since the latter would be biased.

3.2 Bias and Variance: Stationary Cases

In this section the results from the stationary Monte Carlo experiments will be discussed, mainly by modelling the observed biases and variances as functions of the eigenvalues and fitting surfaces by distance-weighted least squares. To prove that the observed bias is significant in most cases, relative biases will also be computed and plotted. All figures and tables in this section are based on 49 eigenvalue combinations for (7) and on 784 combinations for (8), all of which were simulated 10,000 times.

To start with the estimate of the autoregressive parameter a_1 of the first-order system (1), Figures 1 through 4 display fitted bias surfaces and Figures 5 through 8 fitted variance surfaces.

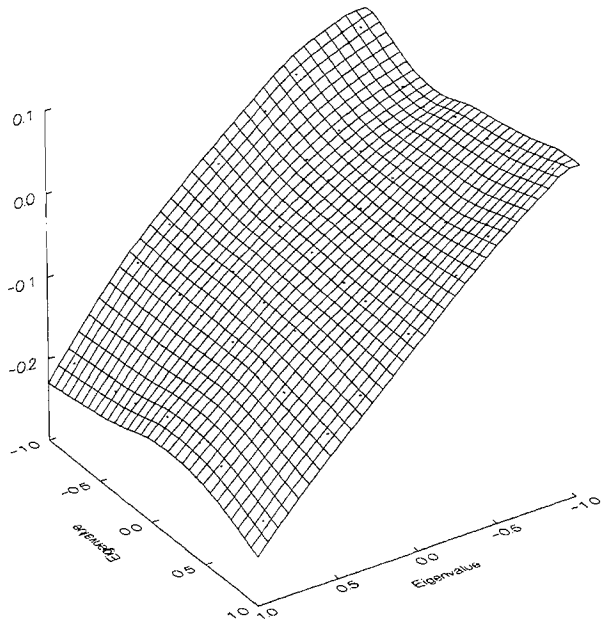

Figure 1: Bias of $\hat{a}_1$; sample size 25

As Figure 1 shows, the bias of $\hat{a}_1$ forms an almost perfect plane when plotted against the two eigenvalues ρ_1 and ρ_2. Only in the (.9,.9) corner in the foreground and in the

back (–.9,–.9) corner does the plane bend a little. The bias is negative for most eigenvalue combinations and ranges from about –.2 (when the eigenvalue on the right-hand scale is .9) to about .1 (when it is –.9), crossing zero around –.3. The other eigenvalue appears to have only marginal impact; smaller in fact than the 5-to-1 ratio between the eigenvalues prescribed by (7). But at least the linearity is there: a_1 is a linear function of ρ_1 and ρ_2 in (7), and here the bias $E(\hat{a}_1 - a_1)$ appears to be almost linear in ρ_1 and ρ_2 as well. (As will be seen in Chapter 4 however, the approximations to $E(\hat{a}_1 - a_1)$ are non-linear functions of the eigenvalues.)

Next, let us see what happens to the bias of $\hat{a}_1$ as the sample size increases to 50 (Figure 2), 100 (Figure 3) and 200 (Figure 4).

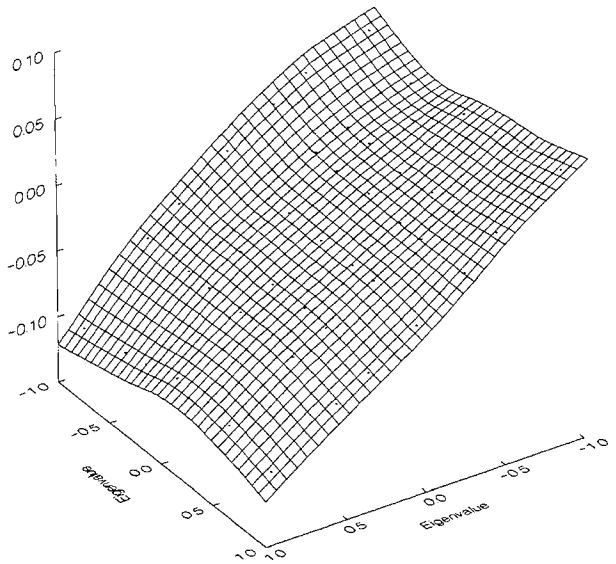

Figure 2: Bias of $\hat{a}_1$; sample size 50

As expected, the bias is reduced as the sample size increases. For T=50 it ranges from about –.1 to about .05, for T=100 from –.05 to .03, and for T=200 approximately from –.025 to .014. It appears to decrease more or less by the same rate as the sample size increases, as it should according to theory. Furthermore, for all four sample sizes it appears to be a linear function of the two eigenvalues (mainly of the right-hand eigenvalue), since the plane shape stays the same as the sample size increases.

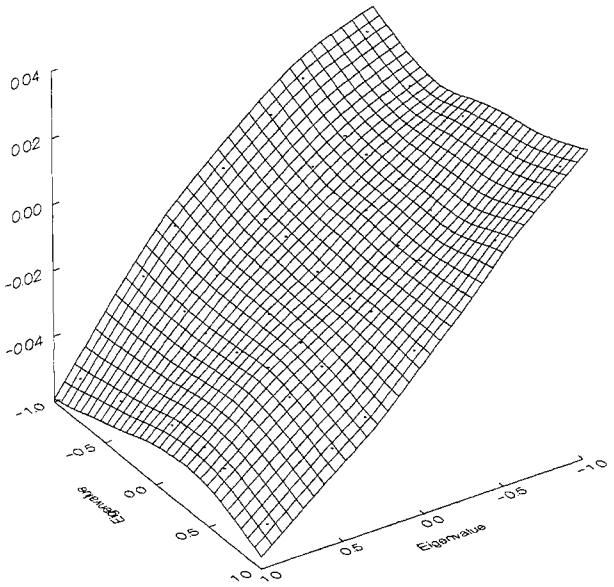

Figure 3: Bias of $\hat{a}_1$; sample size 100

Furthermore, it comes as no surprise that by each doubling of the sample size, the bias is reduced approximately by half. This is consistent with asymptotic theory for stationary processes, according to which $\hat{\mathbf{A}}$ should converge to $\mathbf{A}$ at the rate $1/T$, i.e. bias should be inversely proportional to sample size, as it appears to be for $\hat{a}_1$.

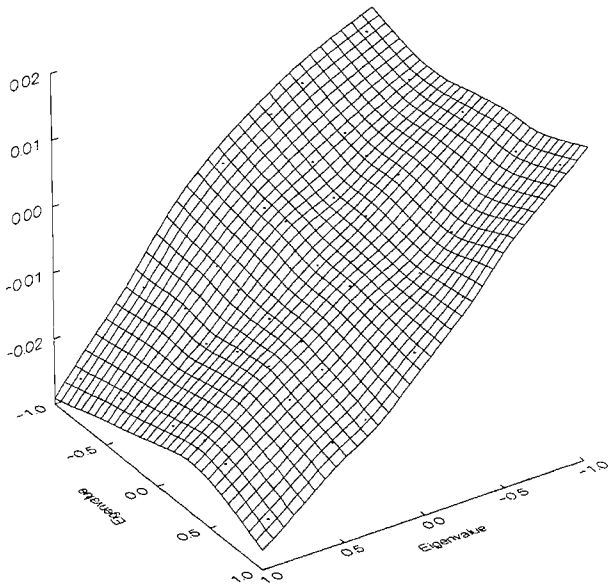

Figure 4: Bias of $\hat{a}_1$; sample size 200

When it comes to the variance of $\hat{a}_1$, the reader should be reminded once again that what is referred to here is the sample variance of all 10,000 estimates and not the mean variance.

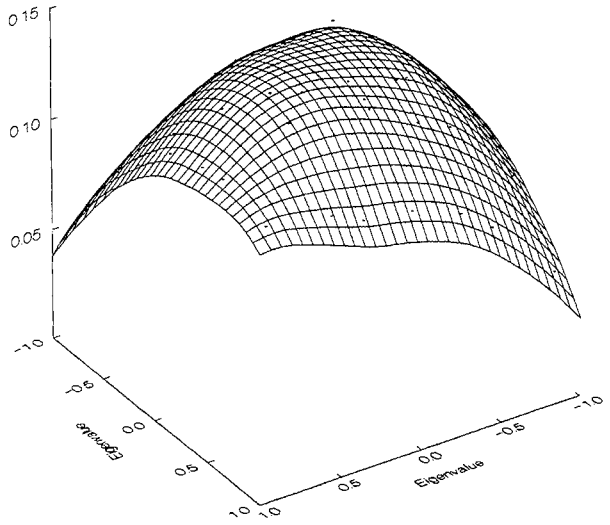

Figure 5: Variance of $\hat{a}_1$; sample size 25

As Figures 5 through 8 demonstrate, the sample variance of $\hat{a}_1$ is a quadratic function of both eigenvalues. For the smallest samples (Figures 5 and 6) it is not completely symmetric though; in the (.9,.9) corner the variance is higher than in the other three corners, but for sample sizes 100 and 200 it is almost perfectly symmetric around its maximum.

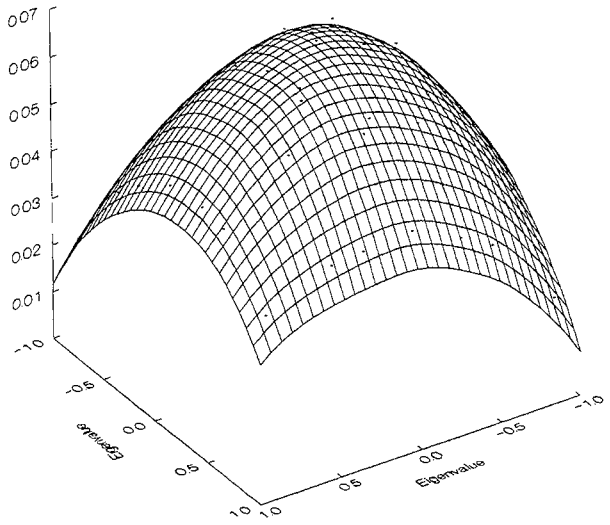

Figure 6: Variance of $\hat{a}_1$; sample size 50

The variance ranges from .06 to .15 for the smallest sample size, from .02 to .07 for T=50, from .01 to .035 for T=100, and from .004 to .017 for T=200. As in the case of bias, the rate of convergence should theoretically be T, but in practice it appears to be higher than that in many cases, in particular as one or both eigenvalues approach unity (in absolute terms). Another interesting feature about the variance figures is that the smallest variance is obtained in the four corners (±.9,±.9) and along the lines connecting them. Since the previous figures showed how the (absolute) bias reaches its

maxima in the same regions, this is where the most serious bias problems will arise in practice, since users are likely to obtain not only very biased estimates, but estimates they will treat as very reliable since standard errors will be small.

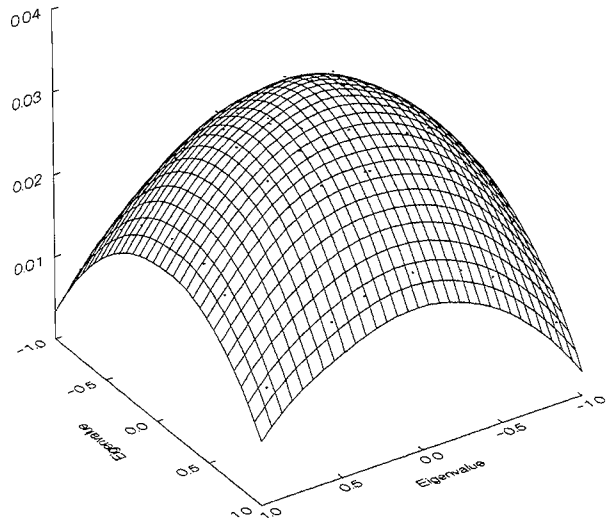

Figure 7: Variance of $\hat{a}_1$; sample size 100

On the other hand, analysts are less likely to be misled in the central region of the figures, within the (±.3,±.3) corners, say. At the same time as the bias is close to zero, the variance reaches its peak, and consequently we can expect the bias to be insignificant for at least some eigenvalue combinations. Without preceding the computation of relative bias at the end of this section, it can be said that based on these figures, it is reasonable to expect more significant biases in regions far from the (0,0) point than around it.

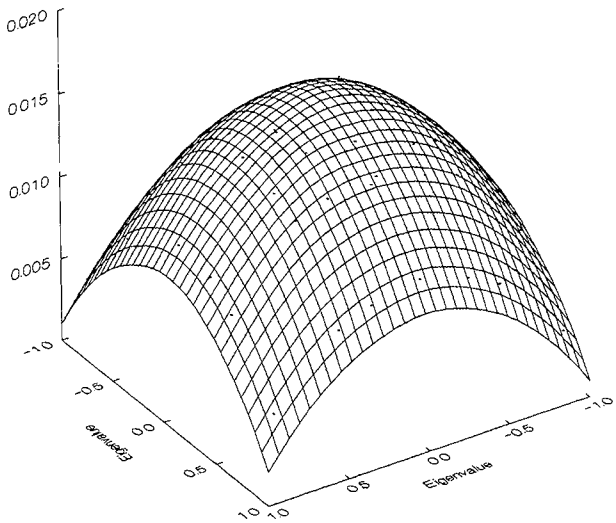

Figure 8: Variance of $\hat{a}_1$; sample size 200

As for the bias of the cross-term estimates $\hat{a}_2$ and $\hat{a}_3$, the most interesting comparison to make is between the two for each sample size. One of the consequences of the way the experiments were designed is to impose the relation $a_3 \equiv -5\, a_2$ between the parameters, and therefore it is of interest to see whether some sort of parallel relation can be detected between $\hat{a}_2$ and $\hat{a}_3$ as well. If so, it should also show in the bias plots, and therefore they are displayed pairwise below.

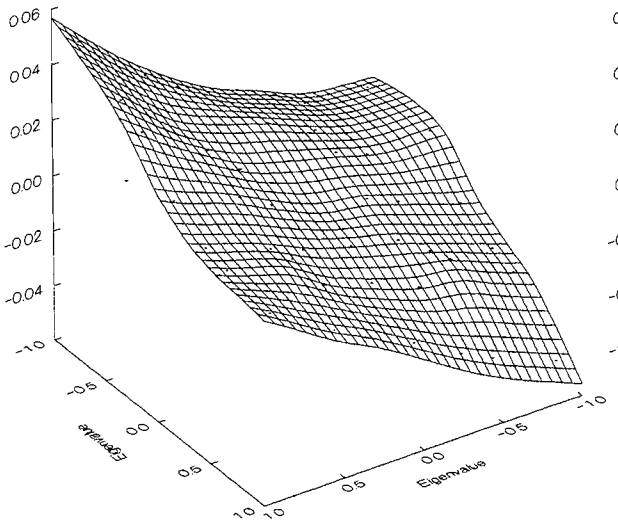

Figure 9: Bias of $\hat{a}_2$; sample size 25

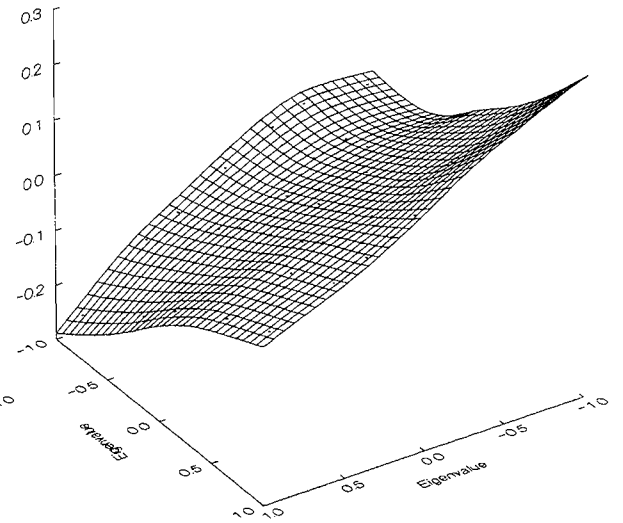

Figure 10: Bias of $\hat{a}_3$; sample size 25

Figures 9 through 16 display fitted bias surfaces for $\hat{a}_2$ and $\hat{a}_3$; Figures 17 through 24 give the corresponding variances. Like the bias of $\hat{a}_1$, the bias of $\hat{a}_2$ as well as the bias of $\hat{a}_3$ are relatively linear and thus form planes when plotted against the eigenvalues. Here however, they appear to be functions of both eigenvalues rather than mainly of one, as the bias of $\hat{a}_1$ (and of $\hat{a}_4$ below). This is no surprise since a_2 and a_3 are defined as equally weighted linear functions of ρ_1 and ρ_2 in (1), a property which apparently is carried over to the estimates and their bias.

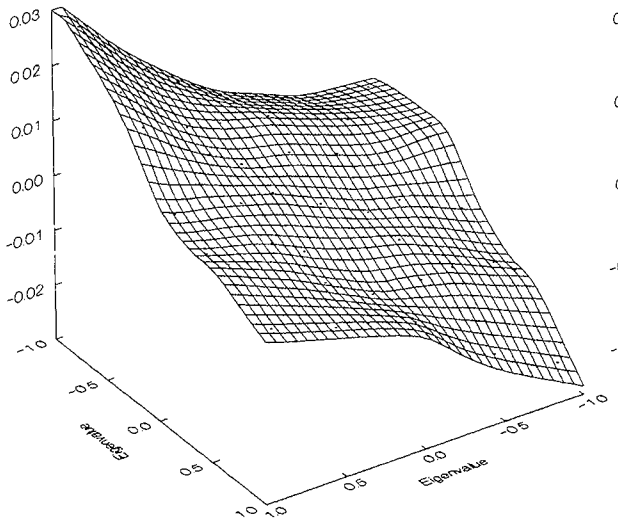

Figure 11: Bias of $\hat{a}_2$; sample size 50

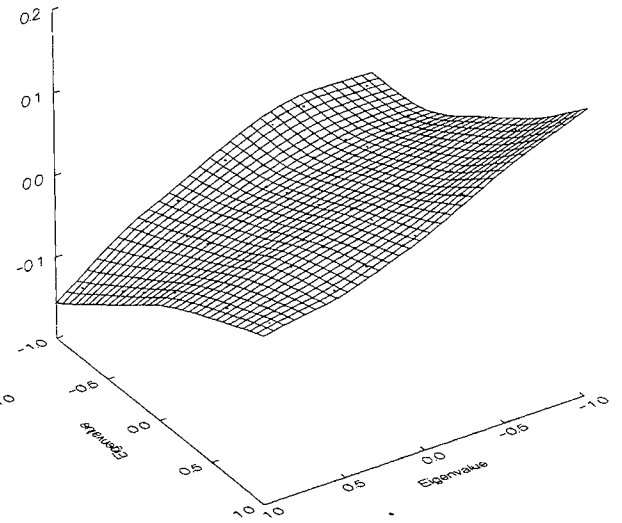

Figure 12: Bias of $\hat{a}_3$; sample size 50

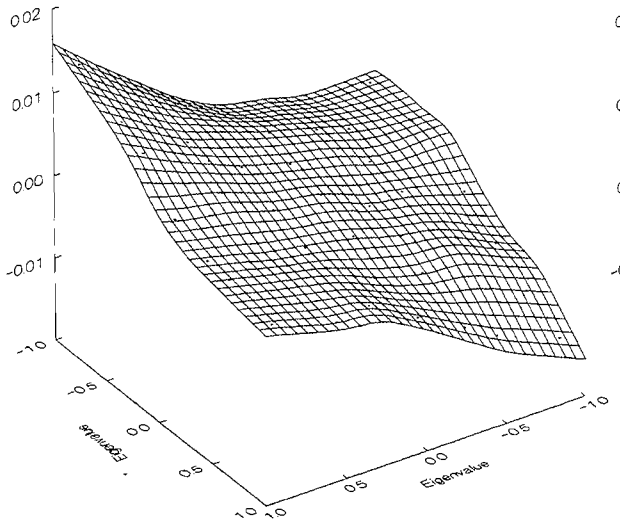

Figure 13: Bias of $\hat{a}_2$; sample size 100

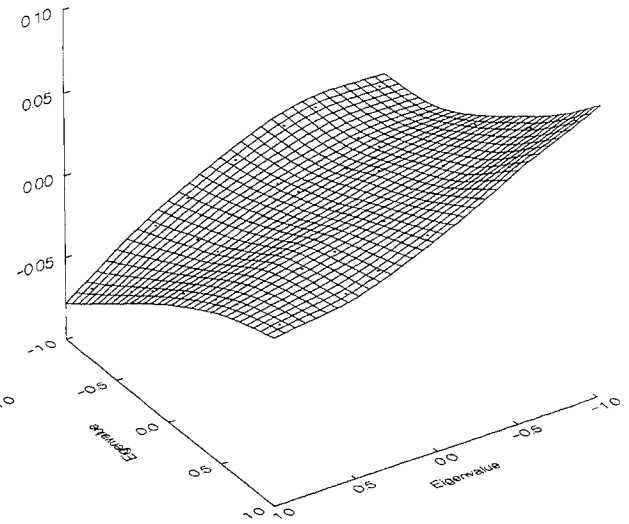

Figure 14: Bias of $\hat{a}_3$; sample size 100

As for the comparison between the bias of $\hat{a}_2$ and the bias of $\hat{a}_3$, the 5-to-1 relation with opposite signs between a_3 and a_2 also seems to apply to their estimates; the $\hat{a}_3$ bias surfaces are reciprocal to the $\hat{a}_2$ bias planes, and always about five times their size. The bias of $\hat{a}_2$ ranges from –.05 to .05 for T=25, from –.025 to .027 for T=50, from –.013 to .014 for T=100, and from –.007 to .007 for T=200. The corresponding ranges for the bias of $\hat{a}_3$ are from –.25 to .25 (T=25), from –.13 to .13 (T=50), from –.07 to .07 (T=100), and from –.03 to .03 (T=200).

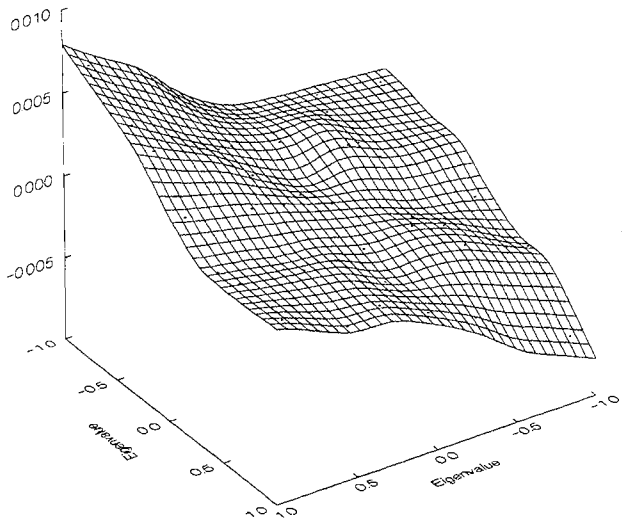

Figure 15: Bias of $\hat{a}_2$; sample size 200

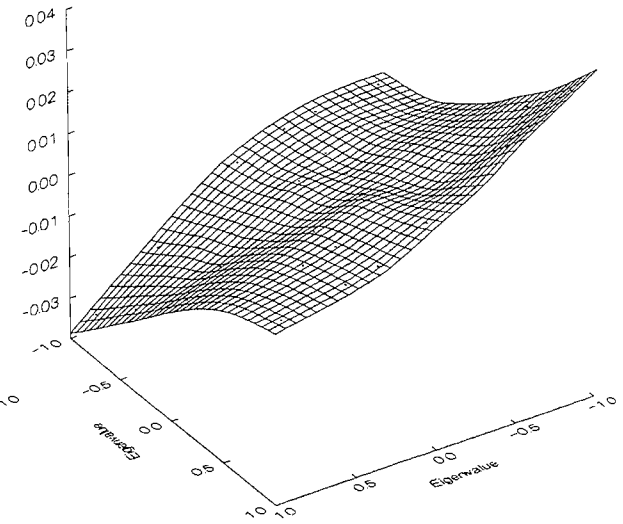

Figure 16: Bias of $\hat{a}_3$; sample size 200

Finally, two differences between these biases and the bias of $\hat{a}_1$ should be noted. Firstly, whereas the bias of $\hat{a}_1$ is zero when one of the eigenvalues is around –.3, a_2 and a_3 are estimated without bias when the two eigenvalues coincide. Since it is clearly the case that in such cases a_2 and a_3 in (7) cancel out, the comforting consequence is that if a_2 and a_3 do not enter the model their estimates will be zero or close to zero (whether

they are also insignificant will be explored later in this section), a conclusion which should hold generally and not only for this specific experiment design. The special case of coinciding eigenvalues will be returned to in Section 3.4.

Secondly, $\hat{a}_2$ and $\hat{a}_3$ are positively and negatively biased in more or less equal proportions, not predominantly negatively biased as $\hat{a}_1$ (and $\hat{a}_4$ below). These results also hold for more general settings than the present; autoregressive parameter estimates on the whole tend to be negatively biased, cross-term estimates do not.

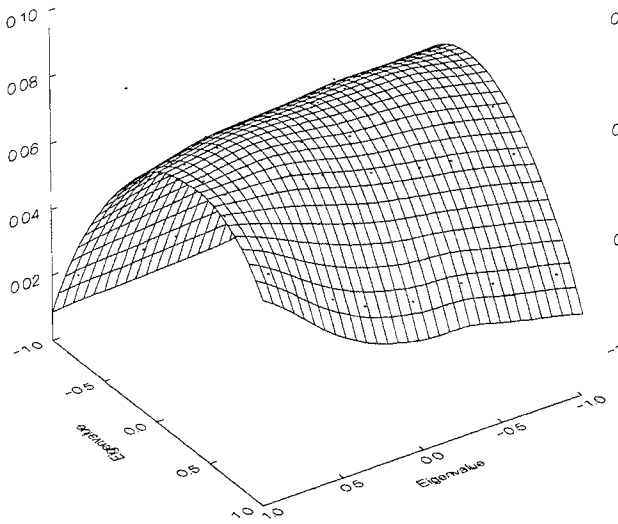

Figure 17: Variance of $\hat{a}_2$; sample size 25

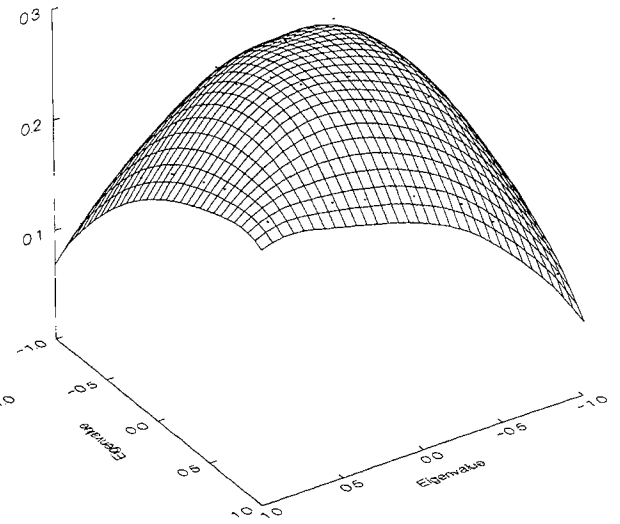

Figure 18: Variance of $\hat{a}_3$; sample size 25

When it comes to the variances of $\hat{a}_2$ and $\hat{a}_3$, the 5-to-1 relation imposed on a_3 and a_2 should have some effect, but probably not the same as on biases. And indeed, as Figures 17 to 24 show, it does make the variance of $\hat{a}_3$ greater than the variance of $\hat{a}_2$, but apart from that it is difficult to discern any pattern within the pairs of variances. In fact, whereas the surfaces fitted to the variances of $\hat{a}_2$ form quadratic functions of only the eigenvalue of the left-hand scale, the sample variance of $\hat{a}_3$ appears to be a quadratic function of both eigenvalues, like the variance of $\hat{a}_1$ above. That $\hat{a}_1$ and $\hat{a}_3$ have similar variance functions (except for a scale factor) is probably specific to this Monte Carlo study and may be related to the fact that a_1 and a_3 are parameters for x_{t-1} in (1) and that with the chosen value of the constant α, eigenvalues are given equal weights in the y_t equation (6) but not in the x_t equation (5). Later on it will be shown that $\hat{a}_2$ and $\hat{a}_4$ also have similar variances (again, except for a scale factor). These results are not likely to hold more generally though.

Like in previous variance plots, each surface is slightly bent around the (.9,.9) corner for the smaller sample sizes, but with increasing sample size this irregularity vanishes.

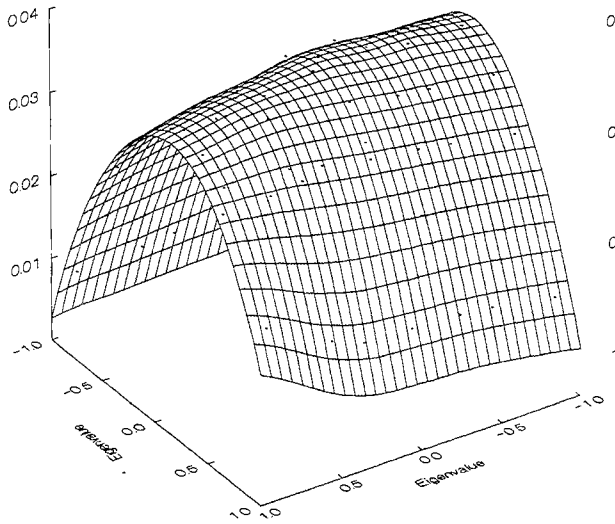

Figure 19: Variance of $\hat{a}_2$; sample size 50

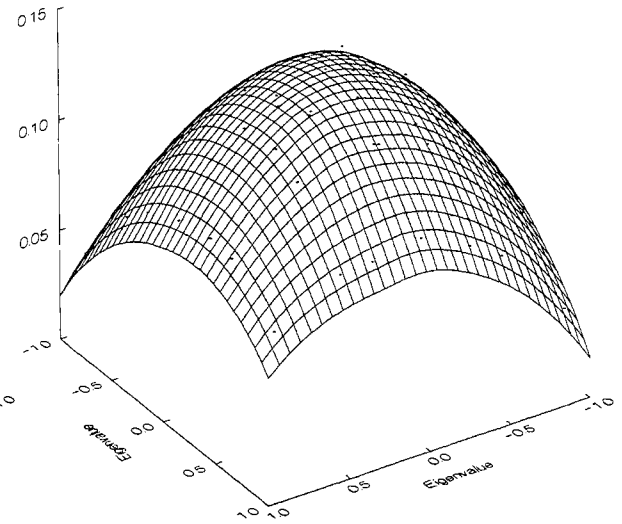

Figure 20: Variance of $\hat{a}_3$; sample size 50

Variances range from .02 to .08 for $\hat{a}_2$ and from .11 to .29 for $\hat{a}_3$ when T is 25; from .01 to .04 for $\hat{a}_2$ and from .04 to .14 for $\hat{a}_3$ when T=50; from .004 to .018 for $\hat{a}_2$ and from .017 to .066 for $\hat{a}_3$ when T=100; from .002 to .009 for $\hat{a}_2$ and from .007 to .032 for $\hat{a}_3$ when T=200. Like in the case of $\hat{a}_1$, in several cases variances appear to converge at a faster rate than by T.

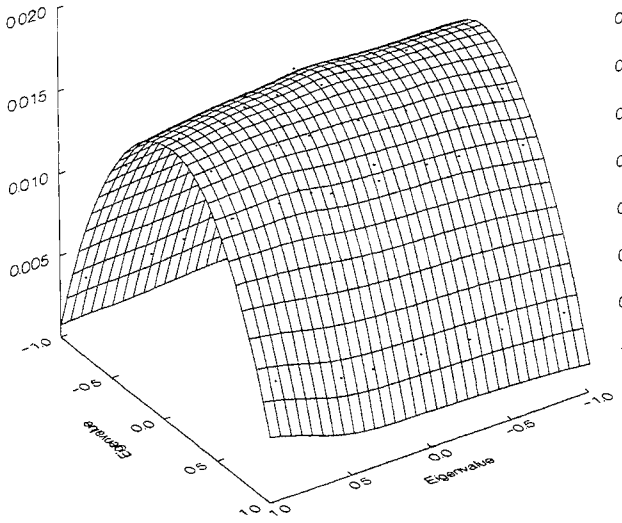

Figure 21: Variance of $\hat{a}_2$; sample size 100

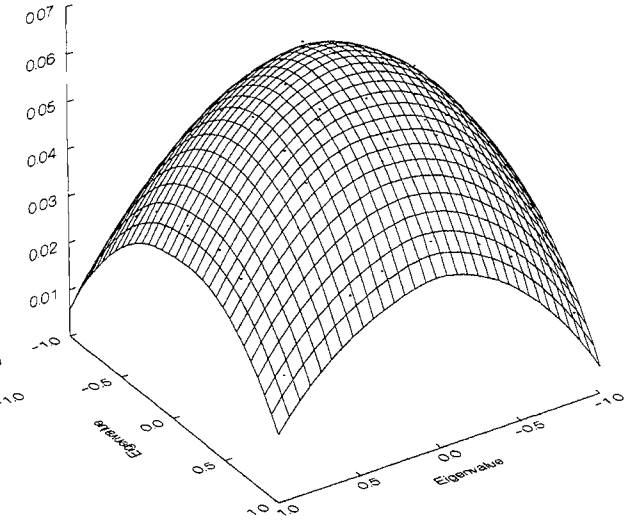

Figure 22: Variance of $\hat{a}_3$; sample size 100

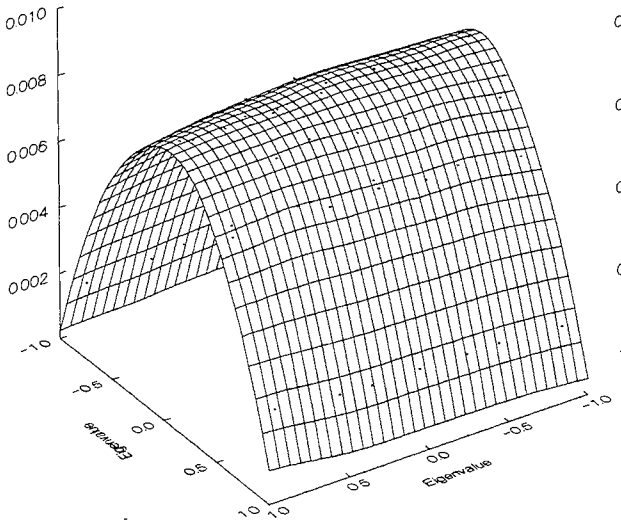

Figure 23: Variance of $\hat{a}_2$; sample size 200

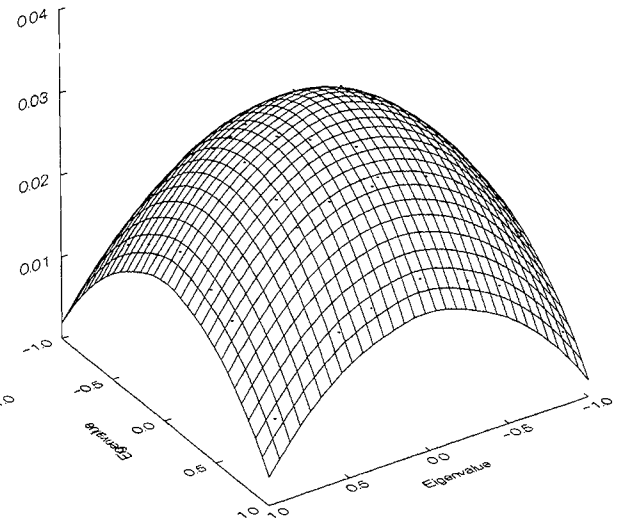

Figure 24: Variance of $\hat{a}_3$; sample size 200

When it comes to $\hat{a}_4$, the same sort of bias planes as for $\hat{a}_1$ can be expected since they are both autoregressive estimates. The weights of $\hat{a}_4$ are reversed though, 1.25 for ρ_1 and –.25 for ρ_2, and therefore the bias of $\hat{a}_4$ should be a linear function mainly of the left-hand eigenvalue ρ_1.

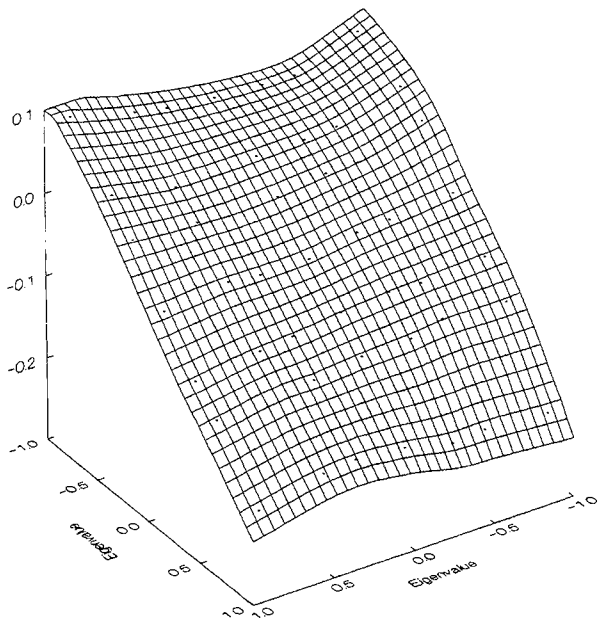

Figure 25: Bias of $\hat{a}_4$; sample size 25

And, as Figures 25 through 28 show, the bias of $\hat{a}_4$ is indeed an almost linear function of the left-hand eigenvalue, whereas the other eigenvalue appears to have virtually no influence (smaller than the 1:5 influence it has on a_4 according to (7)). As in the case of $\hat{a}_1$, the bias ranges approximately from –.2 to .1 for T=25, from –.1 to .06 for T=50, from –.05 to .03 for T=100, and from –.025 to .015 for T=200.

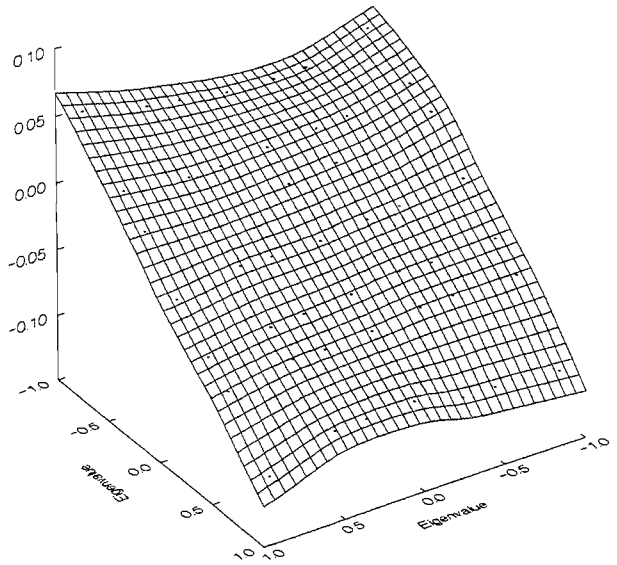

Figure 26: Bias of $\hat{a}_4$; sample size 50

Also parallel to the case of $\hat{a}_1$, there is virtually no bias around the value –.3 of the left-hand eigenvalue. Yet another parallel is the extra-ordinary behaviour in the (.9,.9) and (–.9,–.9) corners. In the (.9,.9) corner $\hat{a}_1$ and $\hat{a}_4$ are more negatively biased than in its vicinity, causing the fitted plane to bend downwards, and in the (–.9,–.9) corner they are more positively biased, causing it to bend up.

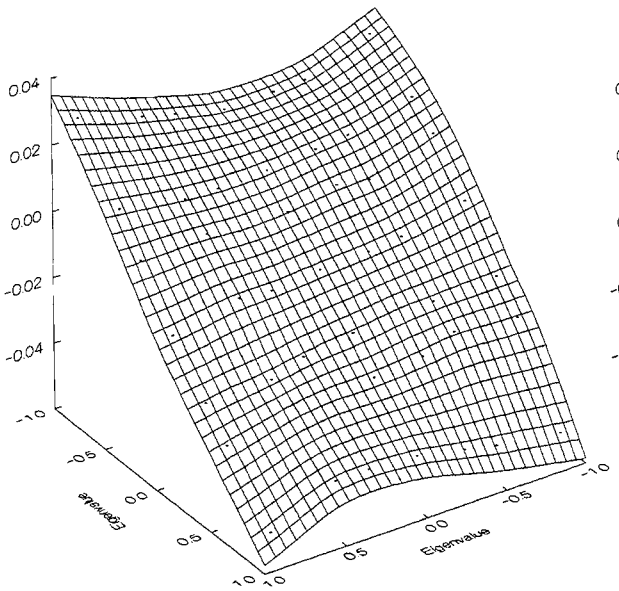

Figure 27: Bias of $\hat{a}_4$; sample size 100

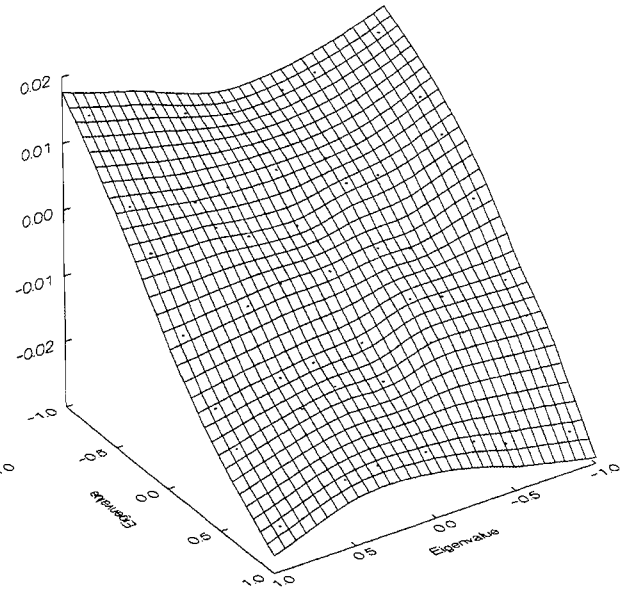

Figure 28: Bias of $\hat{a}_4$; sample size 200

Since a_4 is the parameter for y_{t-1} in the y_t equation, the same sort of variance shape as for $\hat{a}_2$, i.e. a quadratic function of one of the eigenvalues, can be expected. Figures 29 to 32 confirm this; the variance of $\hat{a}_4$ is identical to that of $\hat{a}_2$ except for the fact that the latter is about half the size. But exactly as for the variance of $\hat{a}_1$ versus the variance of $\hat{a}_3$, this is merely an effect of the experiment design and thus not to base general conclusions upon.

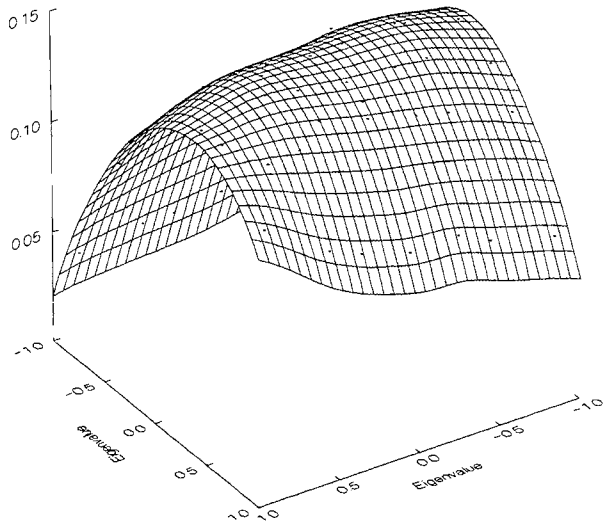

Figure 29: Variance of $\hat{a}_4$; sample size 25

For $\hat{a}_4$ sample variances range from .04 to .15 for T=25, from .02 to .07 for T=50, from .008 to .03 for T=100, and from .004 to .016 for T=200. An important point to make here is that the smallest variances obtain as the left-hand eigenvalue is ±.9, i.e. in exactly the same cases as $\hat{a}_4$ is the most biased (negatively or positively). Therefore the same conclusion as for $\hat{a}_1$ holds; analysts are most likely to be seriously misled in the two regions where the (absolute) bias reaches its maximum at the same time as the variance is small.

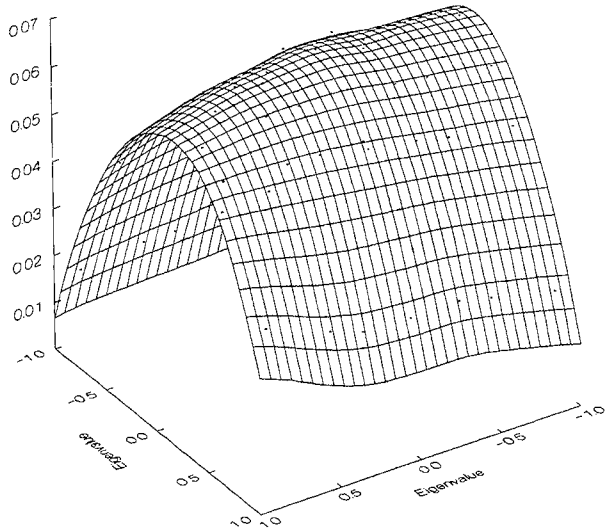

Figure 30: Variance of $\hat{a}_4$; sample size 50

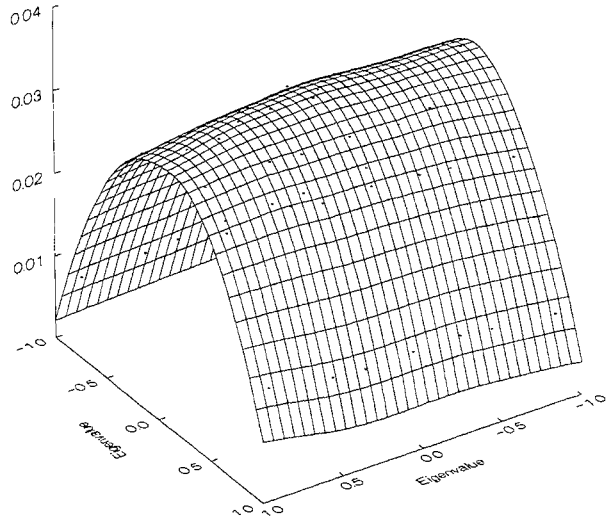

Figure 31: Variance of $\hat{a}_4$; sample size 100

Just like for the three previous parameter estimates, sample variances tend to be more than inversely proportional to T, and as always an irregularity can be detected around the (.9,.9) corner, which disappears however as the sample size increases.

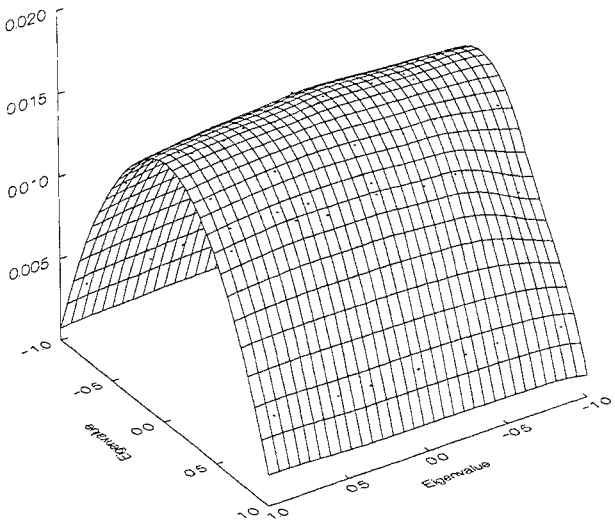

Figure 32: Variance of $\hat{a}_4$; sample size 200

Let us now leave (7) for a moment in order to investigate if the same sort of simple relationships can be established for the bias and variance of the parameter estimates of (8). When describing these, the first problem to be solved is of course that now there are four rather than two eigenvalues, and therefore it is not possible to plot biases and variances against all eigenvalues as for (7). But since the first-order parameters in (8) are really functions of the eigenvalue sums $\rho_{11}+\rho_{12}$ and $\rho_{13}+\rho_{14}$ rather than of each individual eigenvalue, it should be possible to plot biases and variances against the sums instead. Similarly, three-dimensional plots of the bias

and variance of the second-order estimates against the eigenvalue products $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$ should prove useful. Therefore the scales represent $\rho_{11}+\rho_{12}$ and $\rho_{13}+\rho_{14}$ in Figures 32 through 40 ($\hat{a}_{11}$) and $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$ in Figures 41 through 48 ($\hat{a}_{14}$). Thus only two of the eight estimates are shown here, but like in the first-order case a symmetry exists between the parameter estimates of the two equations of (8), so that $\hat{a}_{11}$ and $\hat{a}_{22}$ have similar bias functions and $\hat{a}_{11}$ and $\hat{a}_{21}$ have similar variance functions.

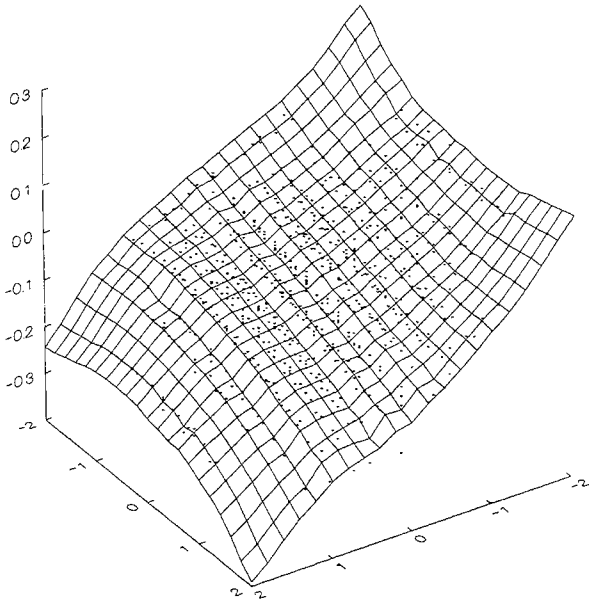

Figure 33: Bias of $\hat{a}_{11}$; sample size 25. Scales represent eigenvalue sums $\rho_{11}+\rho_{12}$ and $\rho_{13}+\rho_{14}$

Furthermore, the bias of $\hat{a}_{12}$ and the bias of $\hat{a}_{21}$ are similar, as are the variances of $\hat{a}_{12}$ and $\hat{a}_{22}$. The same holds for the second-order estimates; whereas $\hat{a}_{13}$ has a similar bias function as $\hat{a}_{24}$ and a similar variance function as $\hat{a}_{23}$, $\hat{a}_{14}$ has a similar bias function as $\hat{a}_{23}$ and a similar variance function as $\hat{a}_{24}$. Results for more estimates in (8) than $\hat{a}_{11}$ and $\hat{a}_{14}$ can be found in Brännström and Karlsson (1993).

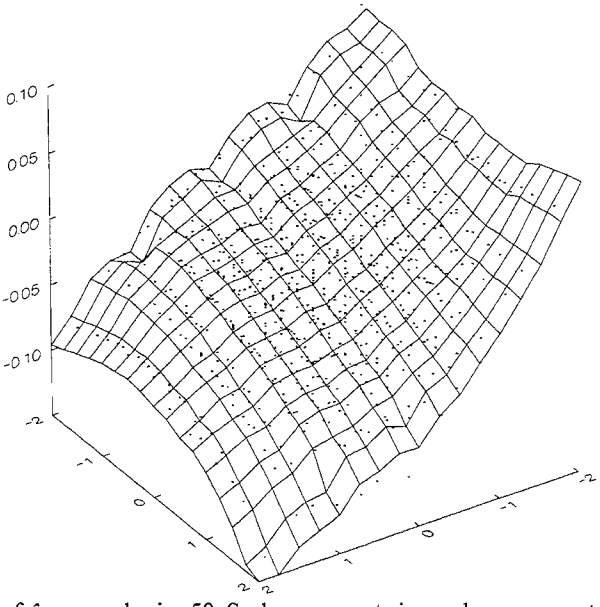

Figure 34: Bias of $\hat{a}_{11}$; sample size 50. Scales represent eigenvalue sums $\rho_{11}+\rho_{12}$ and $\rho_{13}+\rho_{14}$

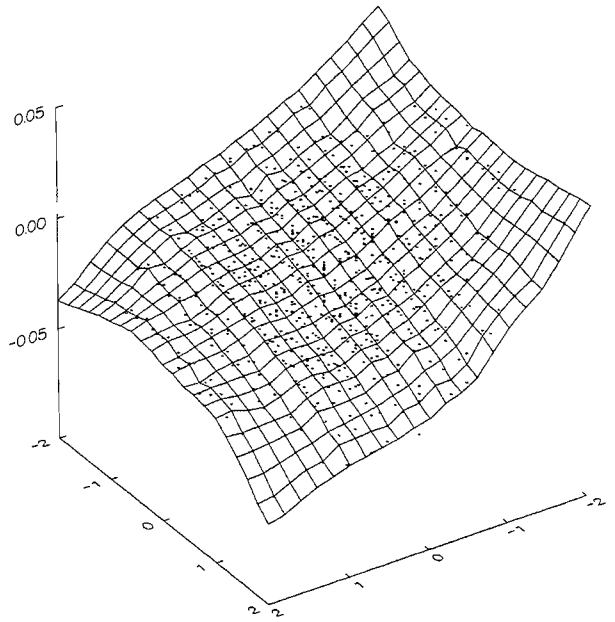

Figure 35: Bias of $\hat{a}_{11}$; sample size 100. Scales represent eigenvalue sums $\rho_{11}+\rho_{12}$ and $\rho_{13}+\rho_{14}$

The plane-like bias surfaces in Figures 33 through 36 are similar to those of $\hat{a}_1$ and $\hat{a}_4$, only less regular here. The bias of $\hat{a}_{11}$ appears to be more or less linear in $(\rho_{11}+\rho_{12})$ and $(\rho_{13}+\rho_{14})$, the latter sum (right-hand scale) having the greater impact (which is in line with (8)), but contrary to the first-order case the first eigenvalue sum (left-hand scale) also has some influence. In the corners, the same effect as in the first-order case arises: In the (1.8,1.8) corner in the foreground, all four eigenvalues being .9 leads to additional negative bias. In the opposite corner all four eigenvalues are –.9 and as a consequence $\hat{a}_{11}$ is more positively biased in than around that corner.

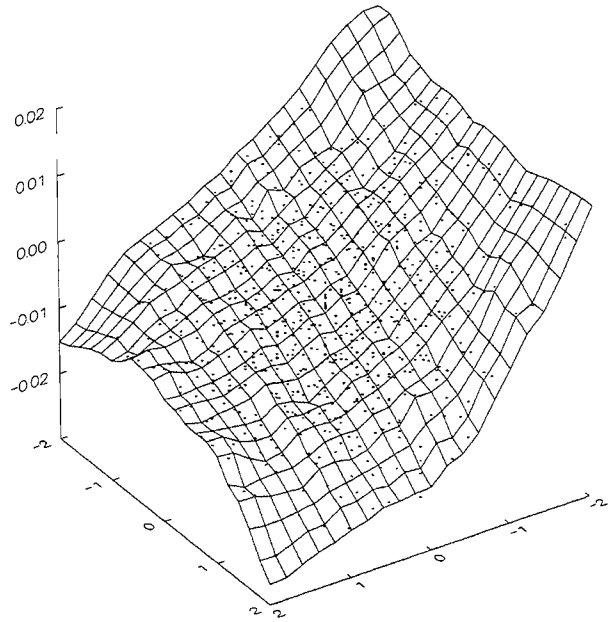

Figure 36: Bias of $\hat{a}_{11}$; sample size 200. Scales represent eigenvalue sums $\rho_{11}+\rho_{12}$ and $\rho_{13}+\rho_{14}$

But contrary to $\hat{a}_1$ and $\hat{a}_4$, the same effects occur in the remaining two corners as well, which can be seen more clearly as the sample size increases. In the right-hand corner the eigenvalue vector is (.9,.9,–.9,–.9), and since the sum of ρ_{13} an ρ_{14} dominates, this again adds to the already positive bias of $\hat{a}_{11}$. In the left-hand corner the eigenvalue vector is (–.9,–.9,.9,.9), and this time the stronger influence of ρ_{13} an ρ_{14} causes extra negative bias.

In the following four figures surfaces are fitted to the variance plots of $\hat{a}_{11}$ against the eigenvalue sums $(\rho_{11}+\rho_{12})$ and $(\rho_{13}+\rho_{14})$. If the parallel with the first-order system were to hold, the resulting surfaces would be shaped like quadratic functions of the two eigenvalue sums, but, as can be seen, the result is far from quadratic, even for the largest sample size. In fact, when the sample size is 25 the variance surface does not bend at all at the (1.8,1.8) and (–1.8,–1.8) corners, only around the (1.8,–1.8) and (–1.8,1.8) corners, forming a clearly non-quadratic function of the eigenvalues. For T=50 the surface bends in all four directions, although still not as much at the (1.8,1.8) and (–1.8,–1.8) corners as at the other two, and in its interior the graph is nearly flat and not quadratic at all. This could be due to the "trenches" appearing for the three largest sample sizes along the lines $\rho_{11}+\rho_{12}=0$ and $\rho_{13}+\rho_{14}=0$ in which the variance is actually lower than around them. The reason the variance drops there is that $\rho_{11}+\rho_{12}=0$ and $\rho_{13}+\rho_{14}=0$ obtain for a number of eigenvalue combinations satisfying $\rho_{11}=-\rho_{12}$ or $\rho_{13}=-\rho_{14}$, for instance by the combination (.9,–.9), which should lead to a very low variance, and by (0,0), which should lead to considerably higher variance. The fitted surfaces represent a weighted average along these lines, and therefore projecting the variance on the eigenvalue sums might not be such a good idea after all.

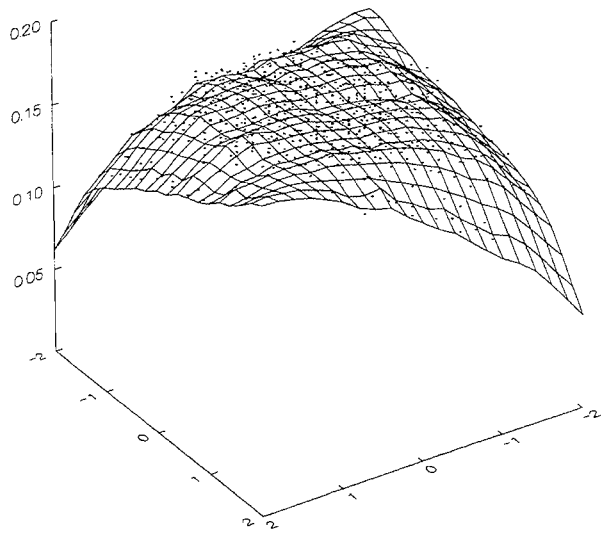

Figure 37: Variance of $\hat{a}_{11}$; sample size 25. Scales represent eigenvalue sums $\rho_{11}+\rho_{12}$ and $\rho_{13}+\rho_{14}$

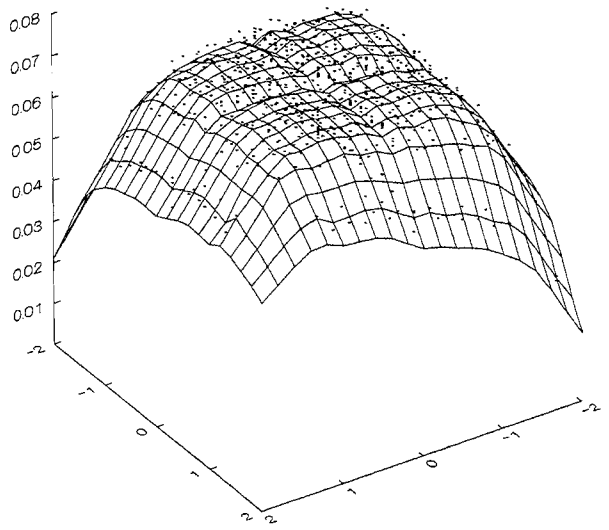

Figure 38: Variance of $\hat{a}_{11}$; sample size 50. Scales represent eigenvalue sums $\rho_{11}+\rho_{12}$ and $\rho_{13}+\rho_{14}$

As the sample size increases to 100 and 200, the variance surface assumes an increasingly smoother shape around and between the four corners, but in its interior it is still very flat, with "trenches" along the zero lines. This in fact makes it look like four overlapping functions rather than one. Whereas the trenches may be due to poor representation in three dimensions of a five-dimensional function, the flatness is more difficult to reconcile with. However, that does not alter one of the principal conclusions of this section, that the eigenvalue combinations leading to the most biased estimates (basically combinations involving ±.9) also produce the least variable estimates, and that therefore they are the most treacherous to VAR analysts. It should also be noted that the variance function, as well as the bias function, of $\hat{a}_{11}$ always assumes higher values than the corresponding functions of $\hat{a}_1$. As the following table shows though, this is not the case for every parameter estimate in (8).

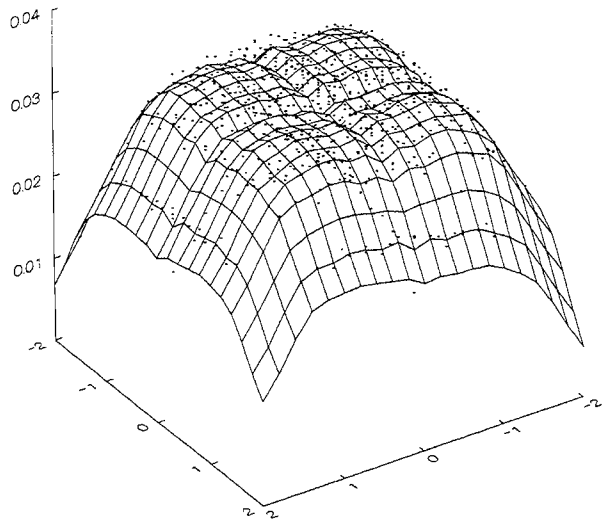

Figure 39: Variance of $\hat{a}_{11}$; sample size 100. Scales represent eigenvalue sums $\rho_{11}+\rho_{12}$ and $\rho_{13}+\rho_{14}$

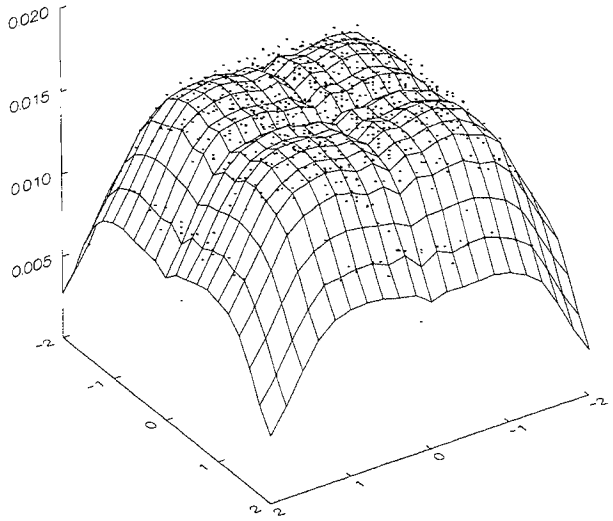

Figure 40: Variance of $\hat{a}_{11}$; sample size 200. Scales represent eigenvalue sums $\rho_{11}+\rho_{12}$ and $\rho_{13}+\rho_{14}$

Before proceeding to the bias and variance of $\hat{a}_{14}$, the following table gives the bias and variance ranges for all eight parameters. The same sort of bias symmetry as in the first-order case can be found; whereas autoregressive estimates ($\hat{a}_{11}$ vs. $\hat{a}_{22}$ and $\hat{a}_{13}$ vs. $\hat{a}_{24}$) are roughly as biased in the x_t as in the y_t equation; cross-term estimates ($\hat{a}_{12}$ vs. $\hat{a}_{21}$ and $\hat{a}_{14}$ vs. $\hat{a}_{23}$) are about five times as biased in the y_t than in the x_t equation.

Ranges of	T=25	T=50	T=100	T=200
Bias($\hat{a}_{11}$)	–.34 to .22	–.14 to .09	–.057 to. 041	–.025 to. 020
Var($\hat{a}_{11}$)	.080 to .188	.029 to .077	.012 to .036	.006 to .017
Bias($\hat{a}_{12}$)	–.053 to .053	–.022 to .021	–.009 to .009	–.004 to .004
Var($\hat{a}_{12}$)	.034 to .104	.014 to .039	.006 to .018	.003 to .009
Bias($\hat{a}_{13}$)	–.45 to .27	–.22 to .13	–.105 to .052	–.049 to .024
Var($\hat{a}_{13}$)	.079 to .179	.028 to .072	.012 to .034	.006 to .016
Bias($\hat{a}_{14}$)	–.12 to .12	–.056 to .058	–.026 to .027	–.012 to .012
Var($\hat{a}_{14}$)	.034 to .103	.014 to .040	.006 to .019	.003 to .009
Bias($\hat{a}_{21}$)	–.26 to .26	–.11 to .11	–.042 to .046	–.018 to .020
Var($\hat{a}_{21}$)	.162 to .335	.058 to .140	.024 to .067	.011 to .033
Bias($\hat{a}_{22}$)	–.35 to .22	–.14 to .09	–.054 to .039	–.024 to .017
Var($\hat{a}_{22}$)	.070 to .194	.028 to .076	.012 to .035	.006 to .017
Bias($\hat{a}_{23}$)	–.61 to .60	–.27 to .28	–.13 to .13	–.062 to .061
Var($\hat{a}_{23}$)	.151 to .352	.055 to .140	.024 to .066	.011 to .032
Bias($\hat{a}_{24}$)	–.44 to .35	–.22 to .13	–.103 to .055	–.049 to .024
Var($\hat{a}_{24}$)	.068 to .182	.027 to .073	.012 to .034	.006 to 016

Let us now focus on the bias and variance of $\hat{a}_{14}$. Since it is the estimated parameter for y_{t-2} in the x_t equation, it can be expected to have similar bias and variance functions as $\hat{a}_2$ in (7), the only difference being that bias (variance) should now be a linear (quadratic) function not of the eigenvalues but of the eigenvalue products $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$. Therefore scales represent these two products in Figures 41 through 48.

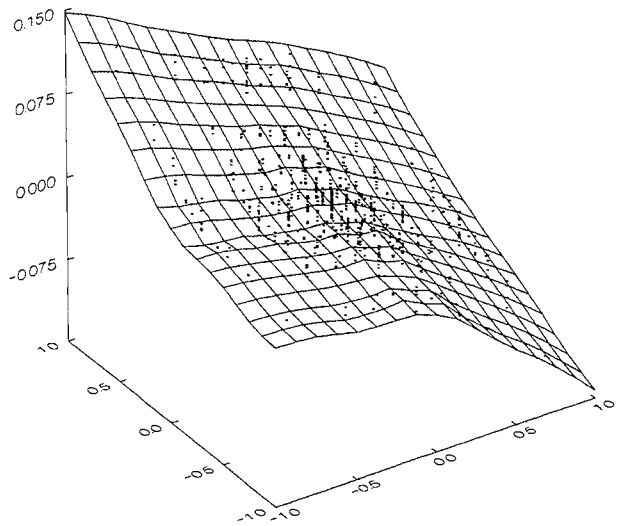

Figure 41: Bias of $\hat{a}_{14}$; sample size 25. Scales represent eigenvalue products $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$

As can be seen in Figures 41 through 44, the bias of $\hat{a}_{14}$ really forms a plane when plotted against $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$. For all four sample sizes though, in the foreground quarter the plane is somewhat distorted. This is where all four eigenvalues are relatively large and of opposite signs within pairs. Nevertheless, the linearity of bias cannot be disputed.

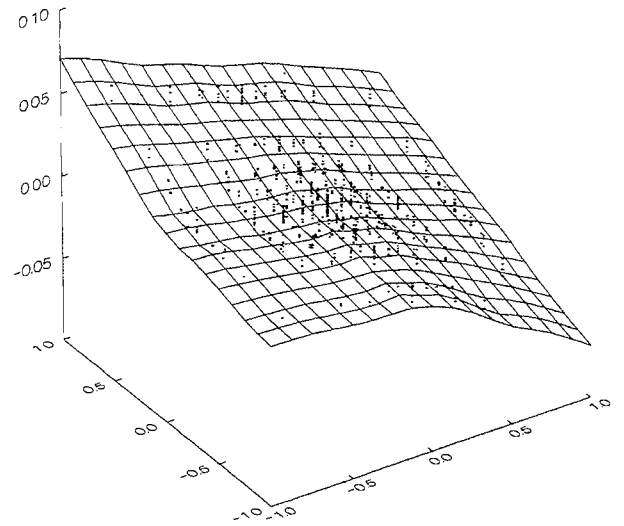

Figure 42: Bias of $\hat{a}_{14}$; sample size 50. Scales represent eigenvalue products $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$

The bias of $\hat{a}_{14}$ is around zero when a_{14} is zero, which it will be when $\rho_{11}\rho_{12}=\rho_{13}\rho_{14}$ (this holds for all four cross-term estimates), which is again a parallel with $\hat{a}_2$ and $\hat{a}_3$. However, the discussion of the significance of $\hat{a}_{14}$ ($\hat{a}_{12}$, $\hat{a}_{21}$, $\hat{a}_{23}$) will have to be postponed until the end of this section.

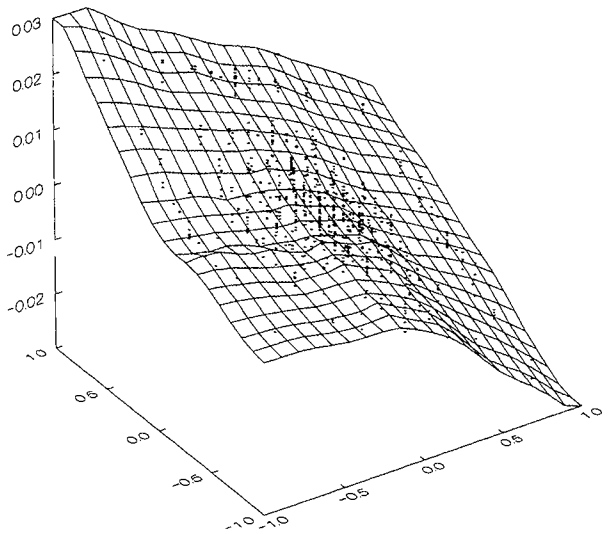

Figure 43: Bias of $\hat{a}_{14}$; sample size 100. Scales represent eigenvalue products $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$

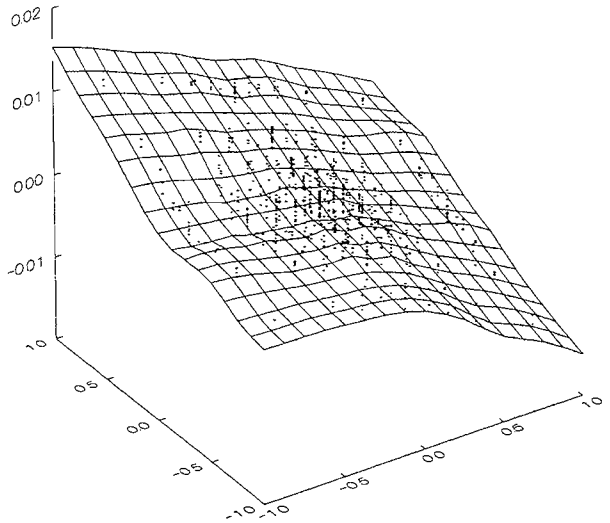

Figure 44: Bias of $\hat{a}_{14}$; sample size 200. Scales represent eigenvalue products $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$

The sample variance of $\hat{a}_{14}$ should be a quadratic function of either $\rho_{11}\rho_{12}$ or $\rho_{13}\rho_{14}$ if the parallel with $\hat{a}_2$ were to hold, but none of that can be traced in Figure 45. For the smallest sample size the variance is a highly irregular function, not at all supporting the idea of variance projection upon eigenvalue products. In fact, even though the sample variance of $\hat{a}_{11}$ also appeared completely different based on 25 observation than when longer series were used, at least then there was some sort of dependence on the eigenvalues, but here a statistical test would probably reject such a dependence (upon the eigenvalue products, at any rate). But, just like in the $\hat{a}_{11}$ case, a considerably brighter picture emerges once a sample of twice that size is used, i.e. T=50. In Figure 46 the same pattern as for $\hat{a}_2$ can be detected, in this case a quadratic function of $\rho_{11}\rho_{12}$ (left-hand scale), whereas the value of $\rho_{13}\rho_{14}$ (right-hand scale) has very little effect. Increasing the sample size to 100 (Figure 47) and 200 (Figure 48) leads to increasingly smoother quadratic functions.

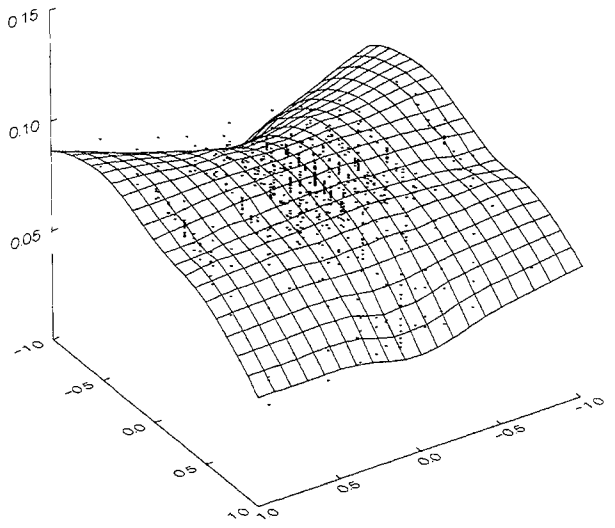

Figure 45: Variance of $\hat{a}_{14}$; sample size 25. Scales represent eigenvalue products $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$

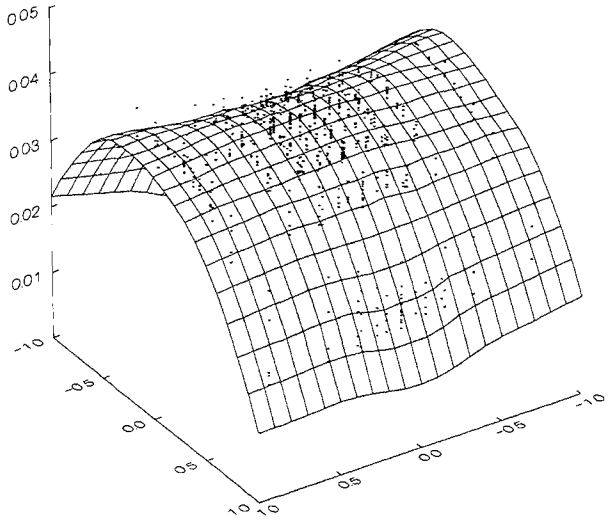

Figure 46: Variance of $\hat{a}_{14}$; sample size 50. Scales represent eigenvalue products $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$

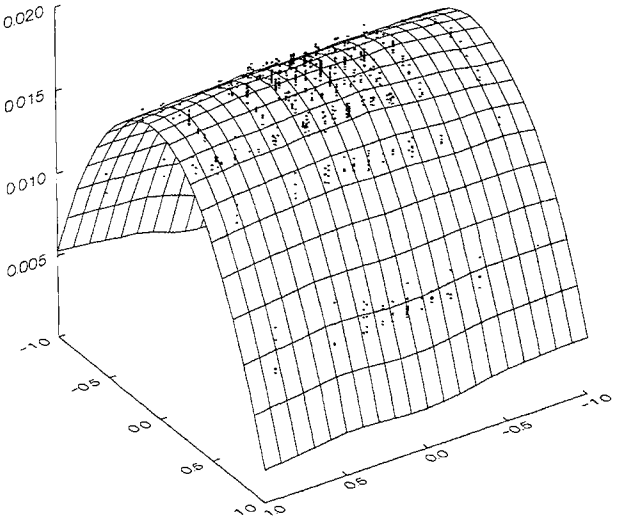

Figure 47: Variance of $\hat{a}_{14}$; sample size 100. Scales represent eigenvalue products $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$

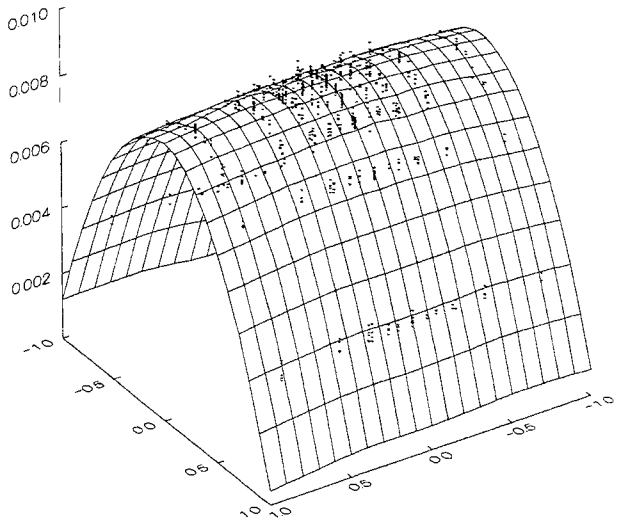

Figure 48: Variance of $\hat{a}_{14}$; sample size 200. Scales represent eigenvalue products $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$

Since all biases (variances) have been found to form fairly regular linear (quadratic) functions of the eigenvalues (at least for sample sizes 100 and 200), it is possible to estimate such functions very accurately. The estimated functions can be found below (with t ratios beneath each estimate), where biases and variances have been aggregated over sample sizes since both should be inversely proportional to sample size. Recall that the simulation results support this in the case of bias, whereas they suggest that variances be more than inversely proportional to sample size. Thus the aggregation of variances over sample sizes is not completely justified by data.

$$\text{Bias}(\hat{a}_1) = \frac{1}{T}\left(\underset{(-41.7)}{-1.07} - \underset{(-5.40)}{.241}\,\rho_1 - \underset{(-84.5)}{3.77}\,\rho_2\right) \quad \bar{R}^2 \approx .97$$

$$\text{Bias}(\hat{a}_2) = \frac{1}{T}\left(\underset{(.52)}{.0055} - \underset{(-32.5)}{.602}\,\rho_1 + \underset{(30.7)}{.569}\,\rho_2\right) \quad \bar{R}^2 \approx .91$$

$$\text{Bias}(\hat{a}_3) = \frac{1}{T}\left(\underset{(.11)}{.0041} + \underset{(44.2)}{2.99}\,\rho_1 - \underset{(-43.5)}{2.94}\,\rho_2\right) \quad \bar{R}^2 \approx .95$$

$$\text{Bias}(\hat{a}_4) = \frac{1}{T}\left(\underset{(-42.6)}{-1.07} - \underset{(-87.2)}{3.82}\,\rho_1 - \underset{(-5.92)}{.259}\,\rho_2\right) \quad \bar{R}^2 \approx .97$$

$$\text{Var}(\hat{a}_1) = \frac{1}{T}\left(\underset{(162)}{3.33} - \underset{(-42.6)}{1.56}\,\rho_1^2 - \underset{(-37.1)}{1.36}\,\rho_2^2 + \underset{(5.03)}{.177}\,\rho_1\rho_2\right) \quad \bar{R}^2 \approx .96$$

$$\text{Var}(\hat{a}_2) = \frac{1}{T}\left(\underset{(136)}{1.76} - \underset{(-66.9)}{1.54}\,\rho_1^2 - \underset{(-1.56)}{.036}\,\rho_2^2 + \underset{(5.17)}{.114}\,\rho_1\rho_2\right) \quad \bar{R}^2 \approx .97$$

$$\text{Var}(\hat{a}_3) = \frac{1}{T}\left(\underset{(149)}{6.51} - \underset{(-37.2)}{2.89}\,\rho_1^2 - \underset{(-36.9)}{2.86}\,\rho_2^2 + \underset{(4.65)}{.347}\,\rho_1\rho_2\right) \quad \bar{R}^2 \approx .95$$

$$\text{Var}(\hat{a}_4) = \frac{1}{T}\left(\underset{(142)}{3.33} - \underset{(-67.3)}{2.81}\,\rho_1^2 - \underset{(-2.67)}{.111}\,\rho_2^2 + \underset{(4.64)}{.186}\,\rho_1\rho_2\right) \quad \bar{R}^2 \approx .97$$

The estimated bias equations are based on $4 \cdot 7^2 = 196$ observations for $\hat{a}_1$, $\hat{a}_2$, $\hat{a}_3$ and $\hat{a}_4$ and on $4 \cdot 28^2 = 3136$ observations for $\hat{a}_{11}$ and $\hat{a}_{14}$, while the corresponding figures for the estimated variance equations are $3 \cdot 7^2 = 147$ and $3 \cdot 28^2 = 2352$, respectively, since the

variance in a sample of 25 observations proved too irregular to allow any equation to be fitted with reasonable accuracy.

$$\text{Bias}(\hat{a}_{11}) = \frac{1}{T}\left(\underset{(-103)}{-1.18} - \underset{(-75.4)}{1.00}(\rho_{11}+\rho_{12}) - \underset{(-113)}{1.50}(\rho_{13}+\rho_{14}) \right) \quad \overline{R}^2 \approx .85$$

$$\text{Bias}(\hat{a}_{14}) = \frac{1}{T}\left(\underset{(-.62)}{-.0029} + \underset{(99.3)}{1.32}\rho_{11}\rho_{12} - \underset{(-97.2)}{1.29}\rho_{13}\rho_{14} \right) \quad \overline{R}^2 \approx .86$$

$$\text{Var}(\hat{a}_{11}) = \frac{1}{T}\left(\underset{(634)}{3.49} - \underset{(-52.6)}{1.47}\rho_{11}^2\rho_{12}^2 - \underset{(-51.3)}{1.43}\rho_{13}^2\rho_{14}^2 + \underset{(3.59)}{.255}\rho_{11}\rho_{12}\rho_{13}\rho_{14} \right) \quad \overline{R}^2 \approx .79$$

$$\text{Var}(\hat{a}_{14}) = \frac{1}{T}\left(\underset{(610)}{1.81} - \underset{(-126)}{1.44}\rho_{11}^2\rho_{12}^2 - \underset{(-3.05)}{.035}\rho_{13}^2\rho_{14}^2 - \underset{(-2.94)}{.054}\rho_{11}\rho_{12}\rho_{13}\rho_{14} \right) \quad \overline{R}^2 \approx .87$$

Some of these estimates merit comments. First of all, the symmetry between the bias functions of $\hat{a}_1$ and $\hat{a}_4$ is striking; their estimated equations are virtually identical except for the roles played by ρ_1 and ρ_2. Between the bias functions of $\hat{a}_2$ and $\hat{a}_3$ a different type of symmetry can be found; each (significant) estimate in the $\hat{a}_3$ equation is of the opposite sign and about five times greater than the same estimate in the $\hat{a}_2$ equation, completely in line with the consequences on (7) of assigning the value .2 to the constant α. The same kind of symmetries can be found in the eight bias equations of (8) as well, though six of them are not shown here.

The symmetry between variances is less apparent since they are also affected by a scale factor, viz. the ratio between the variance of $\varepsilon_{X,t}$ (.01625) and the variance of $\varepsilon_{Y,t}$ (.03125). But except for that scale factor (.52) each estimate in the $\hat{a}_1$ variance equation is very close to the same estimate in the $\text{Var}(\hat{a}_3)$ equation, as are the estimates in the $\hat{a}_2$ and $\hat{a}_4$ variance equations. Thus the variance symmetry is between $\hat{a}_1$ and $\hat{a}_3$ and between $\hat{a}_2$ and $\hat{a}_4$ while the bias symmetry is between $\hat{a}_1$ and $\hat{a}_4$ and between $\hat{a}_2$ and $\hat{a}_3$. As already stated in connection with the variance figures, this is probably because $\hat{a}_1$ and $\hat{a}_3$ are estimates for x_{t-1} in the two equations while $\hat{a}_2$ and $\hat{a}_4$ are the estimates for y_{t-1}.

Unfortunately, despite the apparent appropriateness and accuracy of these equations they are completely useless for practical purposes since they require inaccessible information about the values of each eigenvalue. Even though it is of course possible to estimate eigenvalues, it is impossible to decide which eigenvalue is which. This is no vital matter for the bias of cross-term estimates since the eigenvalues have approximately equal weights there, but for the bias of autoregressive estimates it is crucial not to insert the wrong eigenvalue. Thus the above equations cannot be used for bias (or variance) approximation or reduction. Instead, that challenge will be faced by the three bias approximations to be introduced in Chapter 4.

At this stage one might question how serious the bias problem is and whether it is at all worthwhile to devote time and energy trying to reduce it. These questions will be answered below by the same kind of three-dimensional plots as before, but this time showing the relative bias, i.e. the ratio between bias and its standard error. In other words, each obtained bias $E(\hat{a}-a)$ in the bias figures is divided by the square root of the corresponding sample variance of $\hat{a}$ divided by 10,000, after which DWLS surfaces are fitted as before. The surface is actually less interesting than each relative bias here because the latter can be used to test whether or not there really is a bias for that eigenvalue combination. On the 5 per cent significance level, for instance, the possibility of no bias can be ruled out whenever the relative bias exceeds 1.96 in absolute terms, whereas for ratios between –1.96 and 1.96 it is not possible to rule that out.

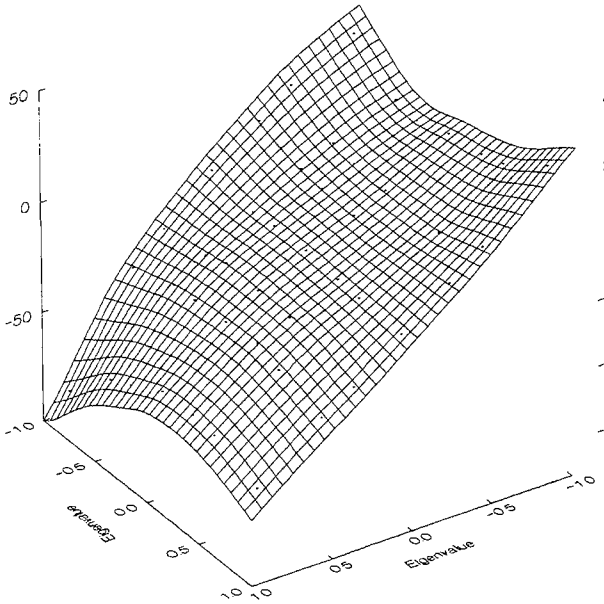

Figure 49: Relative bias of $\hat{a}_1$; sample size 25

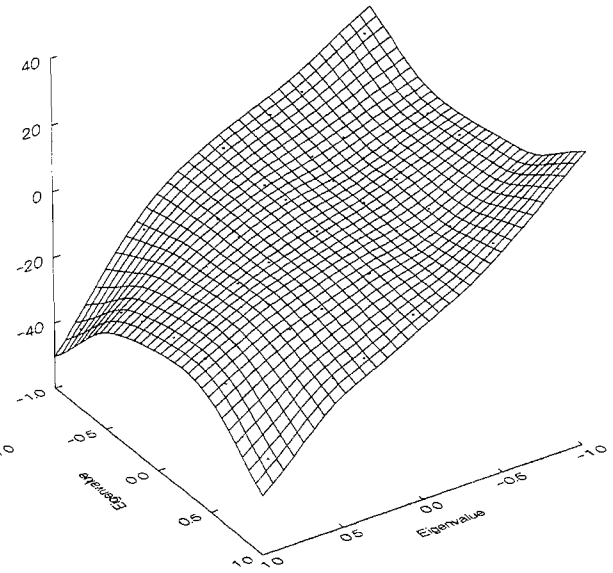

Figure 50: Relative bias of $\hat{a}_1$; sample size 200

There is very little to be gained from displaying the relative bias for all four sample sizes since for each estimated parameter it remains relatively unchanged as the sample size changes. Thus only the relative bias functions based on 25 and 200 observations will be taken to represent all four sample sizes. In Figure 49, for instance, the relative bias of $\hat{a}_1$ when T is 25 is shown, and in Figure 50 when it is 200. The same sort of planes as in Figures 1 and 4 obtain except for in the corners, where the extremely small standard errors (cf. Figures 5 and 8) add to the already sizeable relative bias, ranging from –85 to 40 for T=25 and from –40 to 25 for T=200. It is 5 per cent insignificant (i.e. between –1.96 and 1.96) only when the right-hand eigenvalue ρ_2 is –.3 (also for some combinations involving ρ_2=0 when T=200), which the reader may recall to be when the bias planes in Figures 1 to 4 cross the zero plane.

Turning to $\hat{a}_2$ (Figures 51 and 52), relative bias can again be seen to form basically the same planes as in Figures 9 and 15. It ranges from –25 to 35 for T=25 and from –15 to 16 for T=200, but a wide range of small relative biases along the line connecting the (.9,.9) corner with the (–.9,–.9) corner make $\hat{a}_2$ the most frequently unbiased estimate of all (at the 5 % level). It is clear from (7) that whenever the two eigenvalues are equal, a_2 will cancel out (as will a_3), and in all these cases $\hat{a}_2$ ($\hat{a}_3$) is insignificant for all four sample sizes. But even when ρ_1 and ρ_2 are as wide apart as –.3 and –.9 (as well as for numerous other combinations) the two estimates cannot be said to be significantly biased.

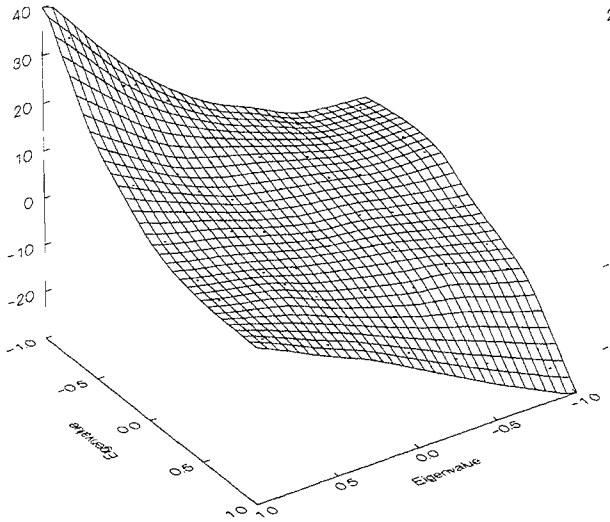

Figure 51: Relative bias of $\hat{a}_2$; sample size 25

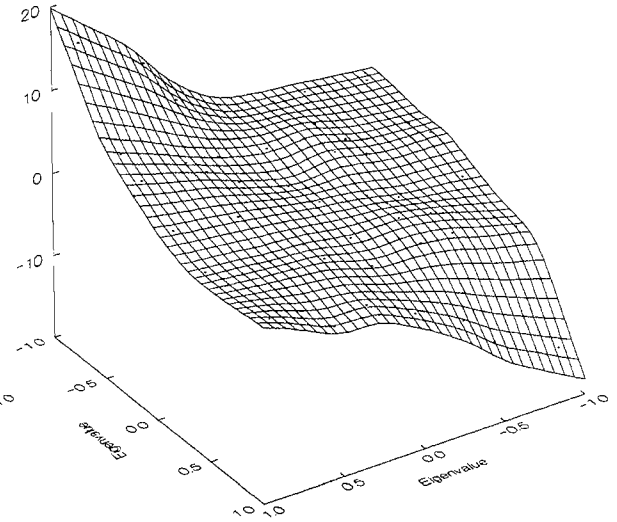

Figure 52: Relative bias of $\hat{a}_2$; sample size 200

Furthermore, as the sample size is increased to 200, a number of combinations are added for which $\hat{a}_2$ (but not $\hat{a}_3$) appears to be unbiased. In fact, the differences between the relative biases of $\hat{a}_2$ and $\hat{a}_3$ are surprising given the relation between a_2 and a_3 and the close resemblance between the bias functions of $\hat{a}_2$ and $\hat{a}_3$ in Figures 9 through 16

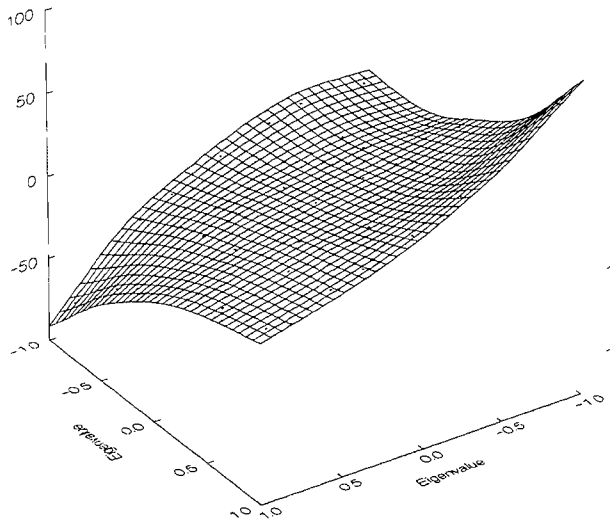

Figure 53: Relative bias of $\hat{a}_3$; sample size 25

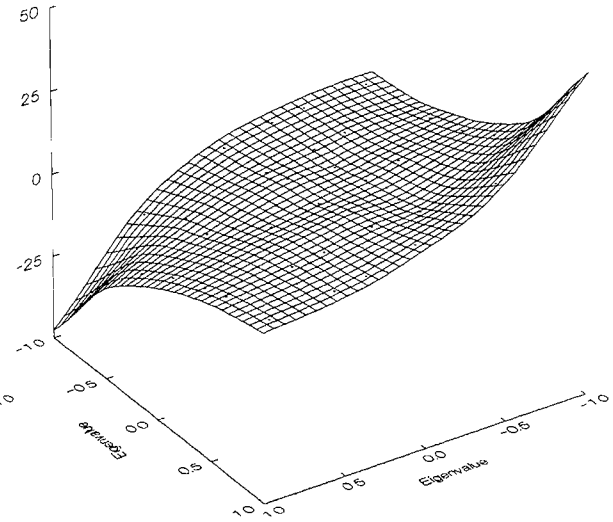

Figure 54: Relative bias of $\hat{a}_3$; sample size 200

(but then of course the variance functions are radically different). Just like $\hat{a}_2$, $\hat{a}_3$ is insignificantly biased whenever the two eigenvalues coincide and for some eigenvalue combinations in which they are close, but there are several such combinations for which $\hat{a}_3$ is significantly biased but $\hat{a}_2$ is not, and in the case of $\hat{a}_3$, increasing the sample size adds no new eigenvalue combinations to the list of combinations for which there is no significant bias.

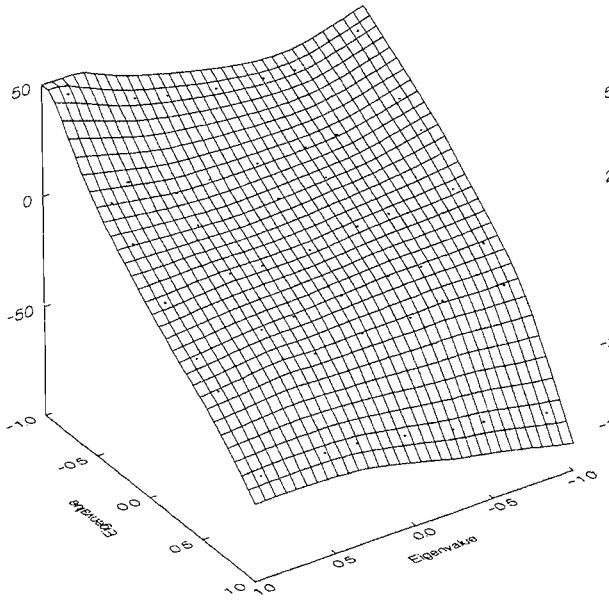

Figure 55: Relative bias of $\hat{a}_4$; sample size 25

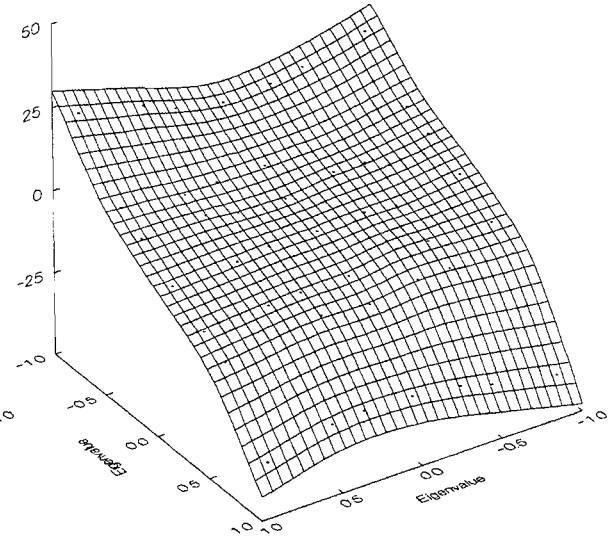

Figure 56: Relative bias of $\hat{a}_4$; sample size 200

Finally, the relative bias of $\hat{a}_4$ (Figures 55 and 56) is parallel to the relative bias of $\hat{a}_1$ in the same way as the bias of $\hat{a}_4$ is the reciprocal of the bias of $\hat{a}_1$, i.e. this time the left-hand eigenvalue ρ_1 plays the role ρ_2 does in Figures 49 and 50. Thus $\hat{a}_4$ will be significantly biased except for a few cases in which ρ_1 is –.3. Only for four eigenvalue combinations is this the case for T=25 though (in three more cases when T is 200), so the two autoregressive estimates are not completely parallel. Another difference is that whereas $\hat{a}_1$ is insignificantly biased for some eigenvalue combinations involving ρ_2=0 as well, this is not the case for $\hat{a}_4$ when ρ_1=0.

At this point it may be interesting to learn that it appears to be possible to estimate all four parameters of (7) without bias, but only under very special circumstances. If both eigenvalues are –.3 $\hat{a}_1$, $\hat{a}_2$ and $\hat{a}_3$ appear to be unbiased for all four sample sizes, but since $\hat{a}_4$ is biased for sample sizes 25 through 100 it will probably take long series and eigenvalues close to –.3 (the farther from –.3, the longer the series have to be) for all four estimates to be unbiased. (But as the sample size tends to infinity, the bias will of course disappear, since consistency is a property of least-squares estimates.)

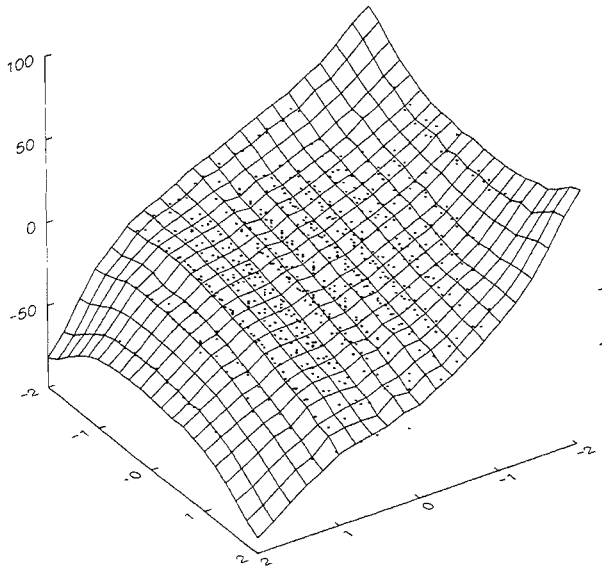

Figure 57: Relative bias of $\hat{a}_{11}$; sample size 25

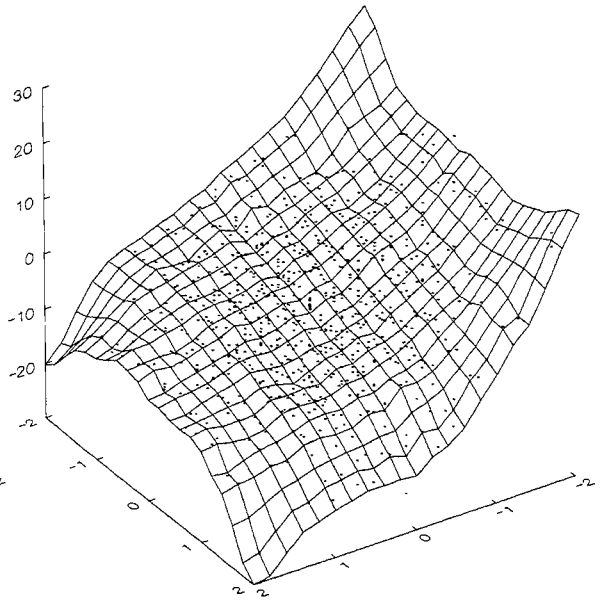

Figure 58: Relative bias of $\hat{a}_{11}$; sample size 200

The relative bias of $\hat{a}_{11}$ in Figures 57 and 58 is essentially the same as the planes in Figures 33 and 36, but with slight kinks in the four corners due to very small variances (cf. Figs 37 and 40). As for the relative bias of $\hat{a}_{14}$ (Figures 59 and 60), no such kinks can be found; the relative bias planes in fact look exactly like the corresponding bias planes in Figures 41 and 44 (except for a scale factor).

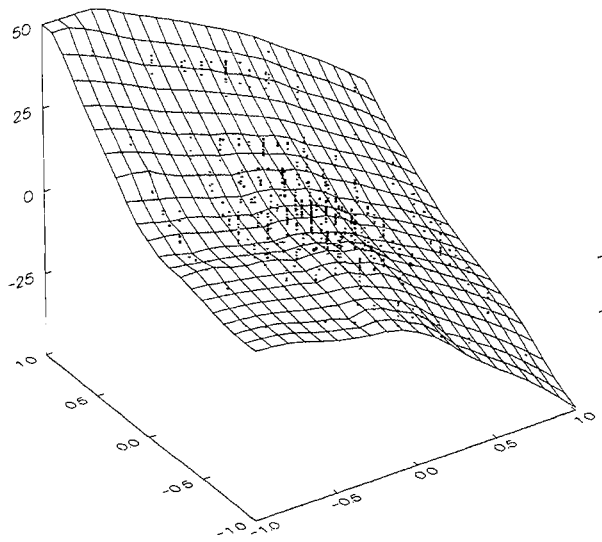

Figure 59: Relative bias of $\hat{a}_{14}$; sample size 25

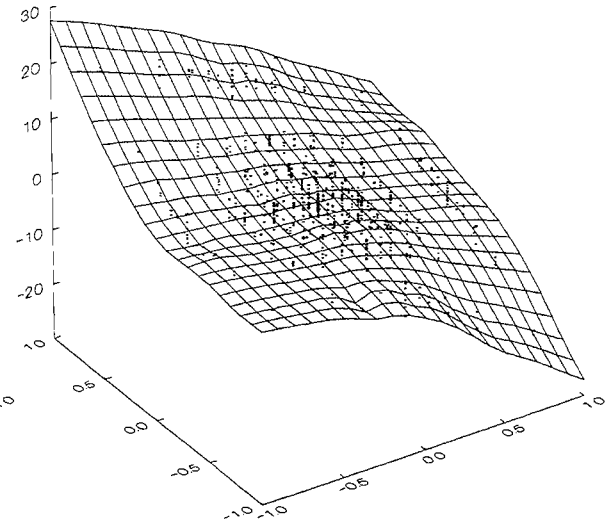

Figure 60: Relative bias of $\hat{a}_{14}$; sample size 200

The relative bias of $\hat{a}_{11}$ ranges from –80 to 60 when the sample size is 25 and from –30 to 25 when it is 200. It appears to be unbiased (smaller absolute relative bias than 1.96) whenever $\rho_{11}+\rho_{12}$ or $\rho_{13}+\rho_{14}$ is slightly negative. Corresponding ranges for $\hat{a}_{14}$ are from –40 to 40 when the sample size is 25 and from –20 to 20 when it is 200. $\hat{a}_{14}$ appears to be unbiased whenever $\rho_{11}\,\rho_{12}$ equals or is close to $\rho_{13}\,\rho_{14}$.

Both parameter estimates appear to be unbiased in quite a few additional cases as the sample size increases to 200.

To summarise the computation of relative biases:

(*i*) Autoregressive estimates ($\hat{a}_1$, $\hat{a}_4$, $\hat{a}_{11}$, $\hat{a}_{13}$, $\hat{a}_{22}$ and $\hat{a}_{24}$) are negatively biased for the majority of eigenvalue combinations. Only by exception are they unbiased (notably for intermediate eigenvalues).

(*ii*) Cross-term estimates ($\hat{a}_2$, $\hat{a}_3$, $\hat{a}_{12}$, $\hat{a}_{14}$, $\hat{a}_{21}$ and $\hat{a}_{23}$) are more or less as likely to be positively biased as they are of being negatively biased. They are unbiased for a number of eigenvalue combinations causing the underlying parameter (a_2, a_3, a_{12}, a_{14}, a_{21} and a_{23}) to be zero or close to zero. Thus if a cross term does not enter the model it seems very unlikely that its parameter should be significantly estimated, however the opposite case (a non-zero parameter being insignificantly estimated) is plausible.

(*iii*) As the sample size increases (in particular from 100 to 200 observations) the number of eigenvalue combinations yielding unbiased estimates grows. This is the case to a greater extent for cross-term estimates than for autoregressive estimates.

(*iv*) For both types of estimates the *t* ratios of the most significant biases decrease very slowly as the sample size increases, thus series of considerably longer length than 200 are needed for $\hat{\mathbf{A}}$ to be unbiased (at least at the 5 % significance level).

Finally, in Brännström and Karlsson (1993) mean square errors of $\hat{\mathbf{A}}$ are computed and plotted, but since they turn out to be completely dominated by the bias of $\hat{\mathbf{A}}$ (i.e. its MSE is, to a close approximation, equal to the square of the bias) and therefore not very interesting they are not reproduced here. In Chapter 4, however, MSE comparisons will be used to evaluate the bias-reduced estimates.

3.3 Bias and Variance: Integrated Cases

Theoretically at least, integrated cases (where at least one eigenvalue is at unity) represent an interesting class of special cases. Integrated and – more recently – cointegrated processes have gained massive attention in theoretical as well as applied work, and several important advances have been made. Therefore it is reasonable to expect integration and cointegration to affect estimates in a different way than in the previous section.

First of all five important propositions regarding the theoretical effects of unit roots in these Monte Carlo experiments are presented:

① If there are no unit roots in (7) or (8), x_t and y_t will be stationary, which follows from the definition of stationarity.

② If there is exactly one unit root, x_t and y_t will be first-order cointegrated, $CI(1,1)$, because if both variables are $I(1)$, the system must contain a cointegrating vector in order for there to be only one unit root.

③ If there are exactly two unit roots, x_t and y_t will be $I(1)$ but not cointegrated since in that case both equations are $I(1)$ and there is no cointegrating vector. However, in the second-order system two unit roots may also correspond to x_t and y_t being $CI(2,2)$ if the two unit roots pertain to the same equation. Thus if ρ_{11} and ρ_{12} are at unity but not ρ_{13} and ρ_{14}, or if ρ_{13} and ρ_{14} but not ρ_{11} and ρ_{12} are unit roots, (8) will be second-order cointegrated for the same reason as under ②. For every other allocation of the two unit roots, x_t and y_t will be $I(1)$ but not cointegrated in (8) as well.

④ If there are exactly three unit roots in (8), x_t and y_t will be $CI(2,1)$ for the same reason as before: In order for both variables to be $I(2)$ at the same time as there are only three unit roots, the system must contain a cointegrating vector.

⑤ If there are exactly four unit roots in (8), x_t and y_t will be $I(2)$ but not cointegrated. Both equations are $I(2)$ and there is no cointegrating vector.

Bearing these properties in mind, it is reasonable to expect different bias functions when unit roots are involved than for stationary processes, and in addition, cointegrated systems may lead to different biases than merely integrated systems. Whether or not this is the case will show in the bias plots below. For the first-order system they are based on sixty-four eigenvalue combinations; the same forty-nine stationary combinations as in the previous section plus fifteen unit-root combinations, fourteen of which leading to first-order cointegration and one ($\rho_1=\rho_2=1$) to first-order integration only.

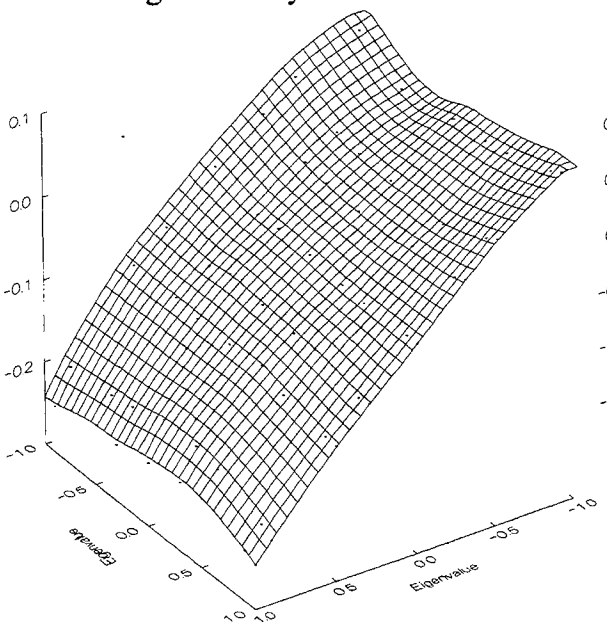

Figure 61: Bias of $\hat{a}_1$; sample size 25

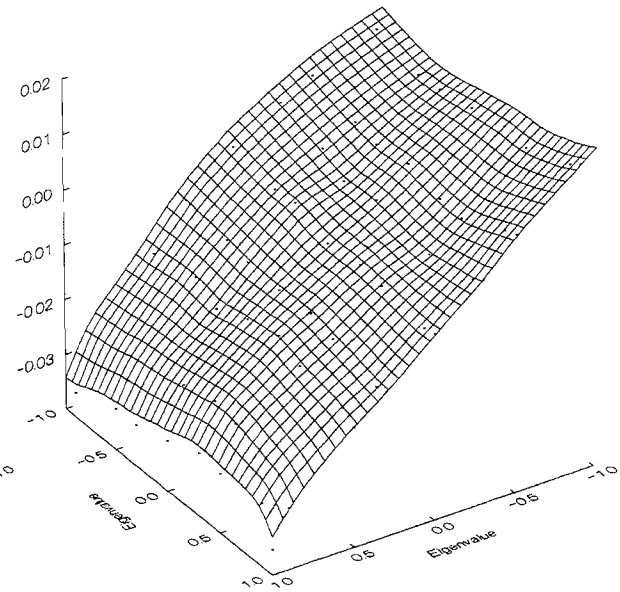

Figure 62: Bias of $\hat{a}_1$; sample size 200

Here again, results are shown only for sample sizes 25 and 200. In the case of $\hat{a}_1$ the bias of the new eigenvalue combinations (along the sides closest to the reader) correspond well with the forty-nine stationary combinations. When the influential right-hand eigenvalue ρ_2 is at unity, this only adds to the already negative bias of $\hat{a}_1$ in

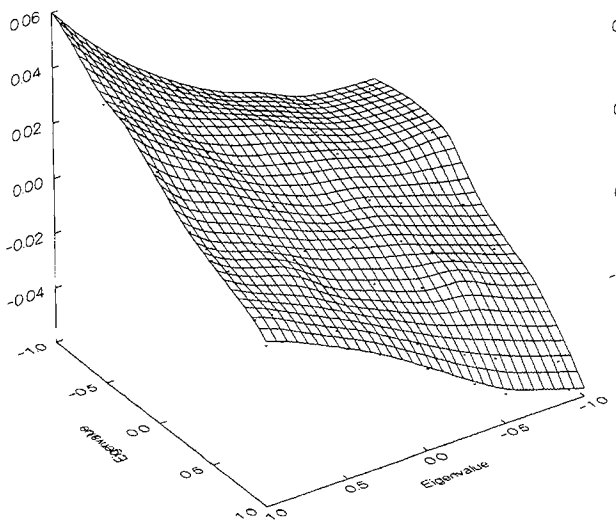

Figure 63: Bias of $\hat{a}_2$; sample size 25

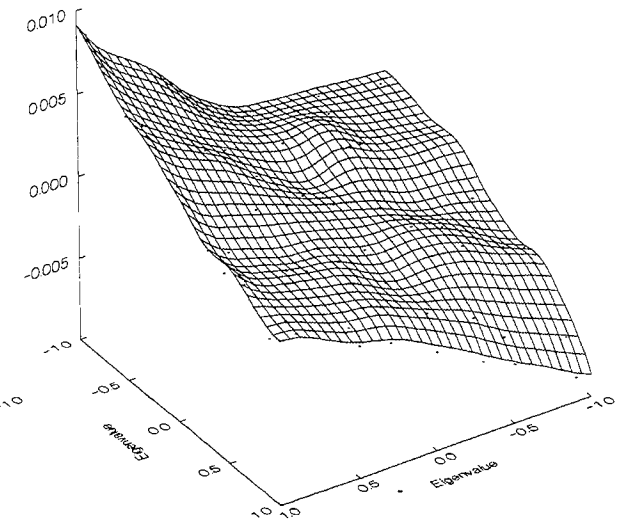

Figure 64: Bias of $\hat{a}_2$; sample size 200

Figures 61 and 62, but not in such a way as to distort the plane shape. Nor is the plane affected at all by the less important left-hand eigenvalue ρ_1 assuming a value of one.

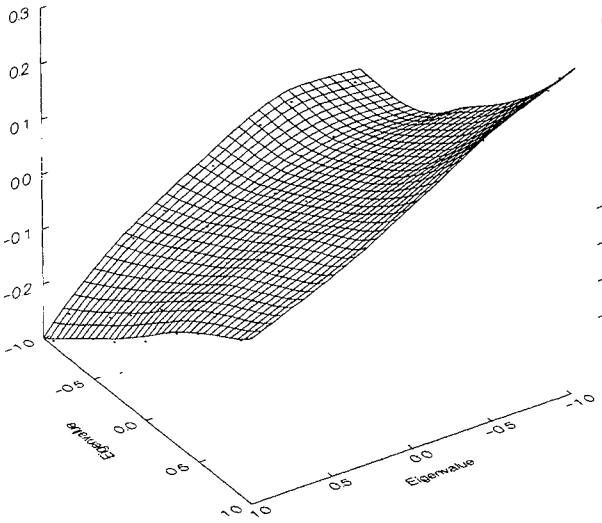

Figure 65: Bias of $\hat{a}_3$; sample size 25

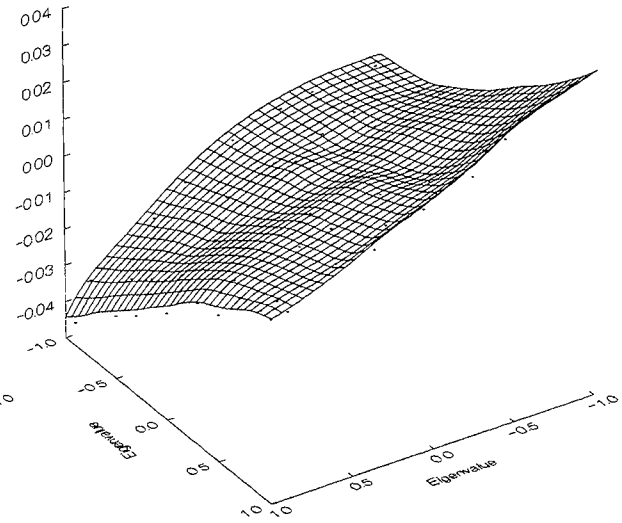

Figure 66: Bias of $\hat{a}_3$; sample size 200

That this holds for the cross-term estimates $\hat{a}_2$ and $\hat{a}_3$ as well should be clear from inspection of Figures 63 through 66, which again are each others' counterparts. More importantly, the same plane shapes as in the previous section can be found in these figures as well, thus the bias in estimated integrated and cointegrated systems is in no way different from that of stationary systems. In particular, when both eigenvalues are at unity (the $I(1)$ but not cointegrated case) and $a_2 = a_3 = 0$, then their estimates are also very close to zero. This special case of "spurious regression" (cf. Granger and Newbold 1974) will receive more attention in the following section.

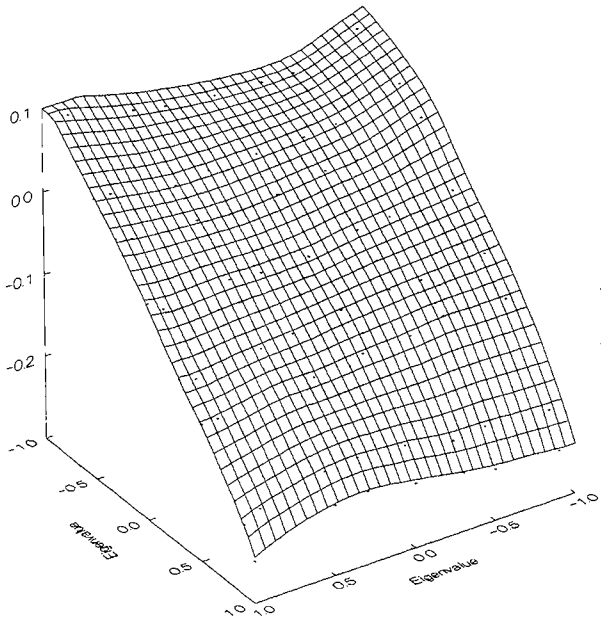

Figure 67: Bias of $\hat{a}_4$; sample size 25

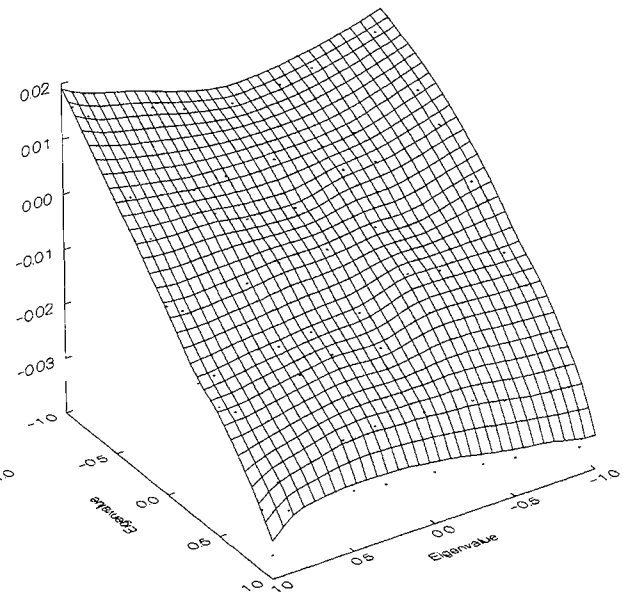

Figure 68: Bias of $\hat{a}_4$; sample size 200

Finally, the bias functions of $\hat{a}_4$ in Figures 67 and 68 also suggest that unit roots affect the estimates no differently than stationary systems. When the left-hand eigenvalue, which completely dominates this bias, is at unity (ρ_1=1), $\hat{a}_4$ is more negatively biased than when it is .9, but still in line with the plane. The right-hand eigenvalue ρ_2 has almost no effect on this bias, thus increasing it to unity does not alter the plane.

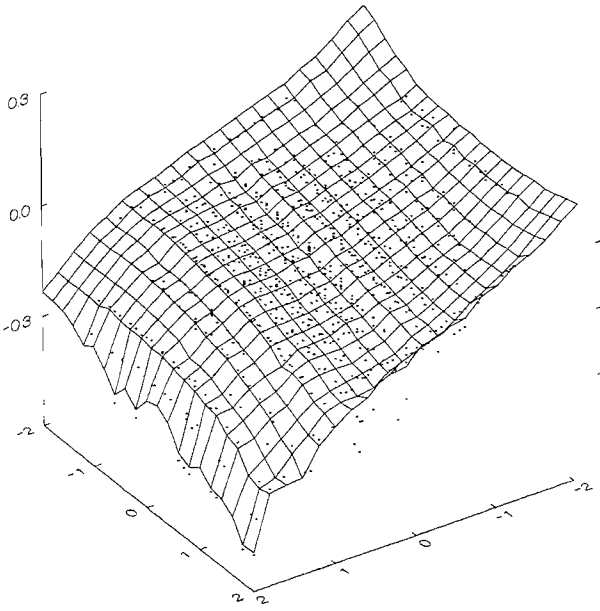

Figure 69: Bias of $\hat{a}_{11}$; sample size 25

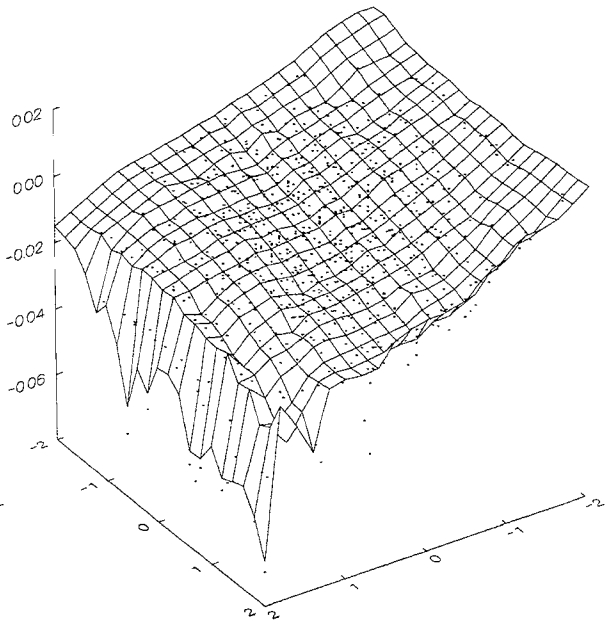

Figure 70: Bias of $\hat{a}_{11}$; sample size 200

More interesting effects occur when the parameters of the second-order system (8) are estimated subject to unit roots. The most spectacular effect is on $\hat{a}_{11}$ in $I(2)$ cases with or without cointegration, i.e. along the two edges closest to the reader in Figures 69 and 70. To the right of the (2,2) corner, cointegration has a positive effect on bias; to the left the effect is sharply negative.

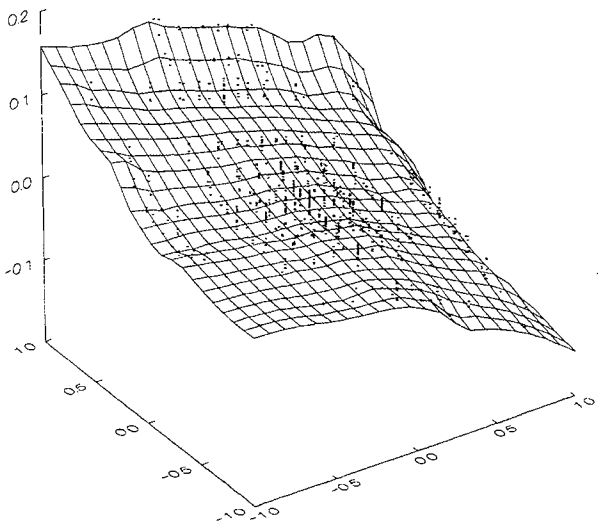

Figure 71: Bias of $\hat{a}_{14}$; sample size 25

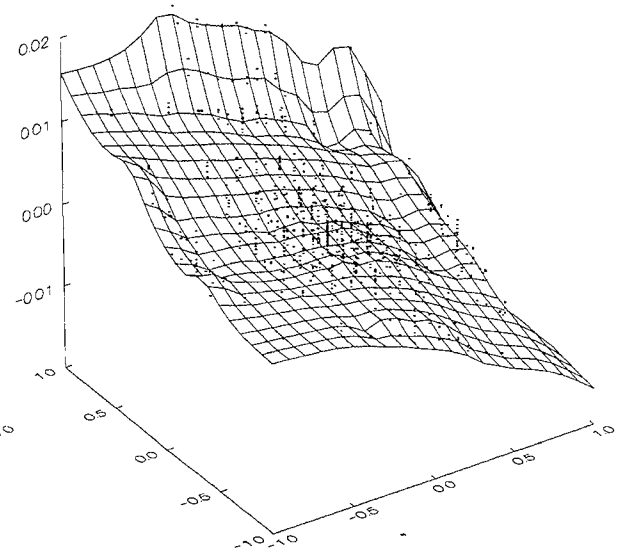

Figure 72: Bias of $\hat{a}_{14}$; sample size 200

The same effects can be seen (slightly less clearly) in Figures 71 and 72 as well, where the bias of $\hat{a}_{14}$ is much higher when $\rho_{11}\rho_{12}=1$ on the left-hand scale (recall that scales represent $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$ in Figures 71 and 72) than when $\rho_{11}\rho_{12}$ is less than unity, and slightly smaller when $\rho_{13}\rho_{14}$ on the right-hand scale is at unity than when $\rho_{13}\rho_{14}$ is smaller. It should also be noted that no such effects could be detected in the $CI(1,1)$ cases along the edges of Figures 61 to 68, nor for first-order cointegrated cases in the second-order system.

To summarise the effects on bias of introducing unit roots, it appears that if (7) or (8) is only $I(1)$ (with or without cointegration) the bias of their estimates is not dramatically affected, whereas profound effects can be observed on (8) if it is $I(2)$, with or without cointegration. If there is second-order (co-)integration ($\rho_{11}=\rho_{12}=1$ and/or $\rho_{13}=\rho_{14}=1$) the resulting bias is far more negative or positive than in the absence of such (co-)integration.

One important effect of unit roots not described until now and not easily detected in the figures above is that if (7) and (8) are $I(1)$ or $I(2)$, the bias of their estimates appears to be more persistent. Judging by these four sample sizes at least, it diminishes at a somewhat slower rate than T (in conflict with asymptotic theory), whereas the rate of convergence is slightly higher than by T for "near-integrated" combinations. An effect of this can be found in Figures 70 and 72, where the bias deflects much more sharply along the edges than in Figures 69 and 71. The reason is that as the sample grows eight times larger (from 25 to 200), most biases shrink at least by the same factor, except for along the edges.

Sample variances are smaller when the system is (co-)integrated than when it is "near-integrated", however they still correspond well with each estimated parameter's quadratic function as introduced in the previous section. Since it is probably more useful to derive the asymptotic variances from the information matrix, no results are provided here though. Instead, interested readers will find more details in Brännström and Karlsson (1993).

3.4 Overparameterization

Another class of special cases falls under the heading of overparameterization, defined as the estimation of parameters which are not in the model. One such case was considered in connection with the discussion of relative bias in Section 3.2, where it was suggested that when cross terms do not enter the model, the risk of obtaining significant cross-term estimates is relatively small. Thus one (in that respect harmless) form of overparameterization is when cross terms cancel out, as they will in the first-order model (7) whenever eigenvalues are equal. In the second-order system (8) it will be the case whenever $\rho_{11}=\rho_{13}$ and $\rho_{12}=\rho_{14}$, because only then are the two requirements $\rho_{11}+\rho_{12}=\rho_{13}+\rho_{14}$ and $\rho_{11}\rho_{12}=\rho_{13}\rho_{14}$ fulfilled.

Here interest focuses on the remaining, autoregressive estimates, more specifically on their bias and variance, because if cross terms do not enter, the model should not be estimated as a VAR but as seemingly unrelated regressions (SUR), which the following simple example demonstrates for (7): If $\rho_1=\rho_2=\rho$, the cross terms cancel out and (7) reduces to $\begin{bmatrix} x_t \\ y_t \end{bmatrix}=\begin{bmatrix} \rho & 0 \\ 0 & \rho \end{bmatrix}\begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix}+\varepsilon_t$, where $\varepsilon_t \sim N\left(\mathbf{0},\begin{bmatrix} .01625 & \\ .01875 & .03125 \end{bmatrix}\right)$, i.e. two AR(1) equations with contemporaneous correlation.

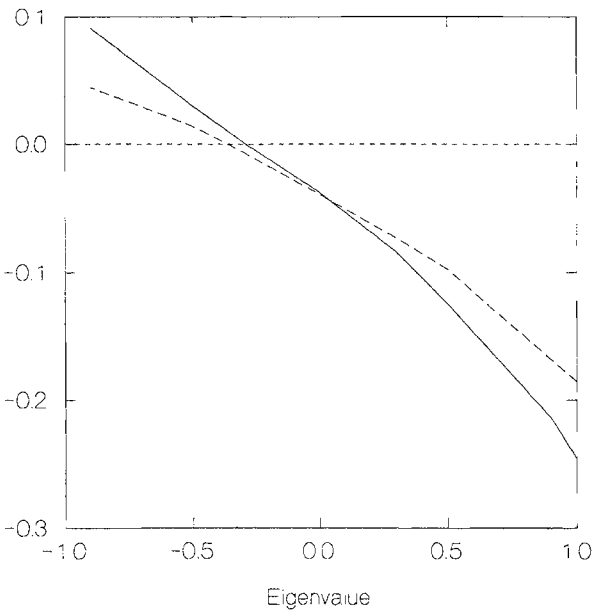

Figure 73: Bias of $\hat{a}_1$ when $\rho_1=\rho_2$; sample size 25
Solid line VAR estimate, dashed line SUR estimate

In Figures 73 through 76 the bias of $\hat{a}_1$ when estimated as a VAR parameter is compared with its bias when estimated as a SUR parameter (based on the same replications). Similar figures would obtain for $\hat{a}_4$ as well, therefore $\hat{a}_1$ is taken to represent both.

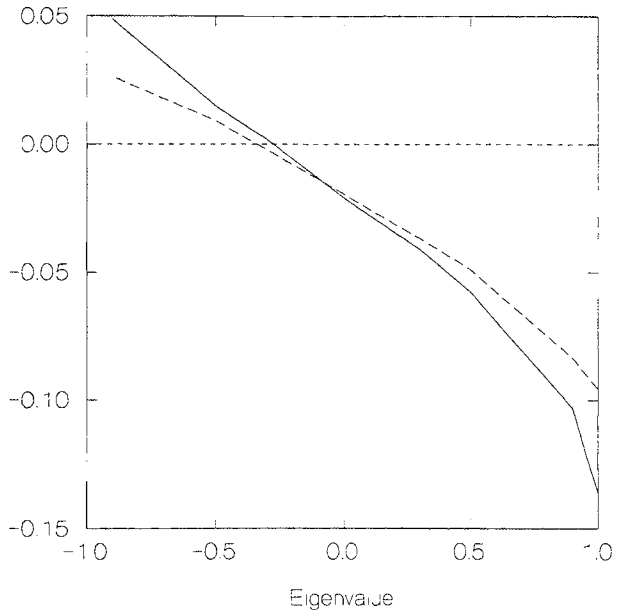

Figure 74: Bias of $\hat{a}_1$ when $\rho_1=\rho_2$; sample size 50
Solid line VAR estimate, dashed line SUR estimate

As the figures show, $\hat{a}_{VAR}$ is more biased than $\hat{a}_{SUR}$ for all values of ρ except for in a small region where both biases are small (and mainly insignificant). No doubt overparameterization is the reason for the inflated VAR bias; estimating two parameters (a_2 and a_3) which are not in the model reduces the degrees of freedom and blurs the estimation of a_1 and a_4. The general features are the same though; both estimates are positively biased for the most negative values of ρ, unbiased for intermediate negative values, and negatively biased for all remaining eigenvalues. Both biases are more or less linear, which is consistent with the figures in Section 3.2.

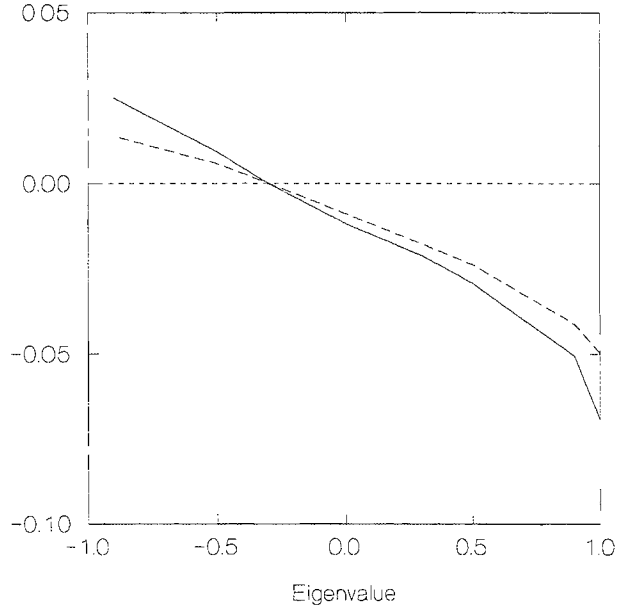

Figure 75: Bias of $\hat{a}_1$ when $\rho_1=\rho_2$; sample size 100
Solid line VAR estimate, dashed line SUR estimate

It should also be noted that $\hat{a}_{VAR}$ is far more biased when $\rho_1=\rho_2=1$, i.e. in the Granger and Newbold (1974) "spurious regression" case, than when $\rho_1=\rho_2=.9$. Spurious regression has a similar effect on $\hat{a}_{SUR}$ as well, but on a much smaller scale.

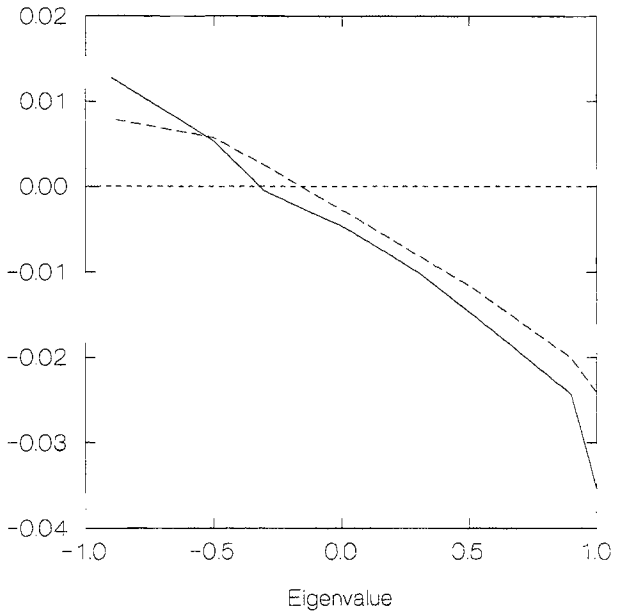

Figure 76: Bias of $\hat{a}_1$ when $\rho_1=\rho_2$; sample size 200
Solid line VAR estimate, dashed line SUR estimate

For variances, the same sort of comparisons as above can be found in Figures 77 through 80 below. Like previously in this chapter, the variance of $\hat{a}_1$ is more or less quadratic, irrespective of method of estimation. However, the variance of $\hat{a}_{VAR}$ is several times greater than the variance of $\hat{a}_{SUR}$. Again, this imprecision is probably due to overparameterization; estimating parameters which are not in the model makes remaining parameter estimates less accurate.

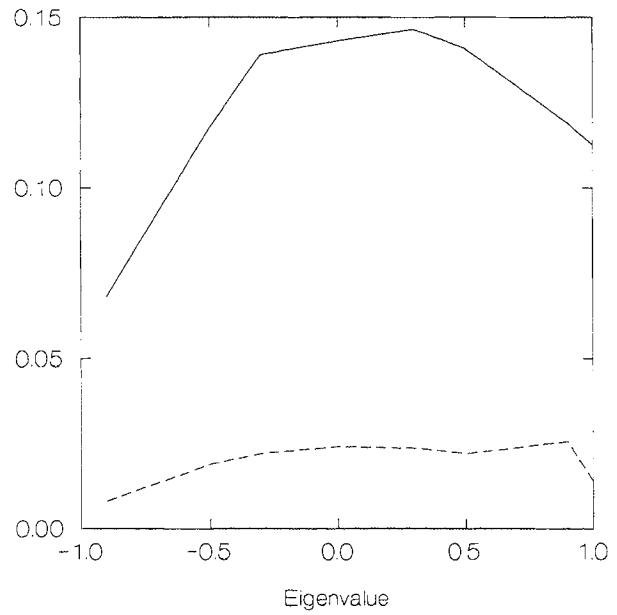

Figure 77: Sample variance of $\hat{a}_1$ when $\rho_1=\rho_2$; sample size 25
Solid line VAR estimate, dashed line SUR estimate

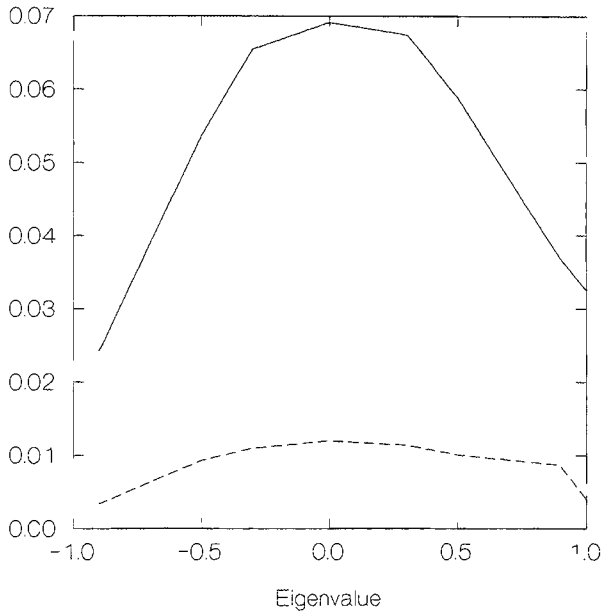

Figure 78: Sample variance of $\hat{a}_1$ when $\rho_1=\rho_2$; sample size 50
Solid line VAR estimate, dashed line SUR estimate

It should also be noted that the superior biases and variances are in no way limited to SUR estimates. Estimating without use of the covariance structure, i.e. as two separate AR(1) equations, leads to very similar results.

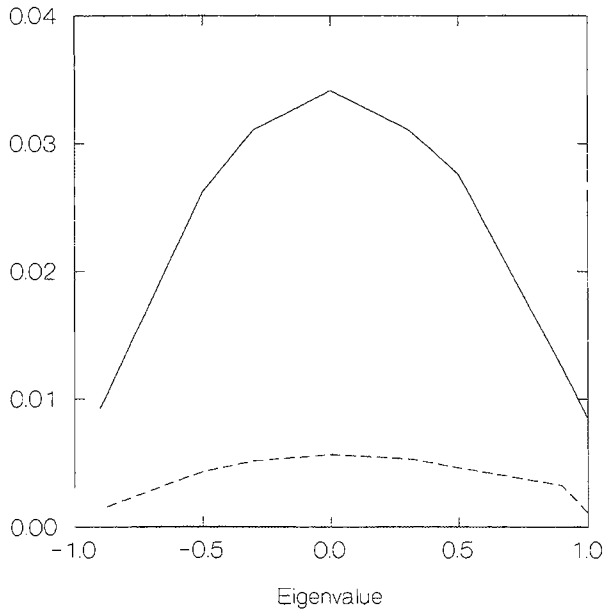

Figure 79: Sample variance of $\hat{a}_1$ when $\rho_1=\rho_2$; sample size 100
Solid line VAR estimate, dashed line SUR estimate

The same differences between VAR and SUR estimates obtain in the second-order model (8) as well, as seen in Brännström and Karlsson (1993). However, these results are not reproduced here.

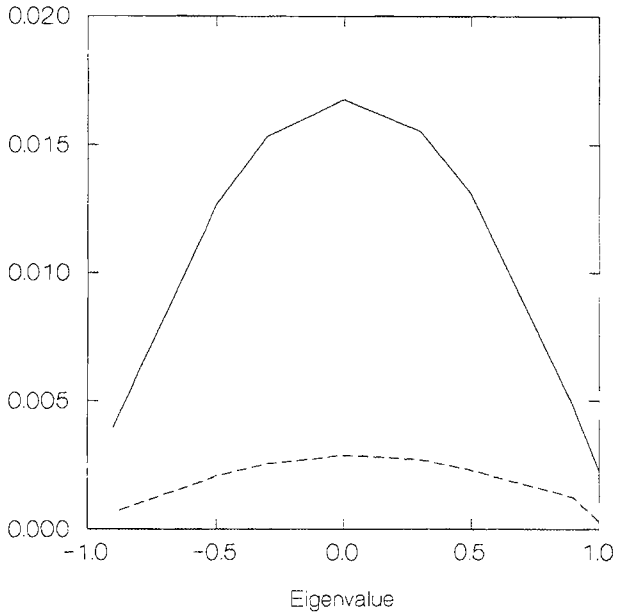

Figure 80: Sample variance of $\hat{a}_1$ when $\rho_1=\rho_2$; sample size 200
Solid line VAR estimate, dashed line SUR estimate

Now let us turn to another form of overparameterization, more precisely one which obtains if (8) is estimated when in fact (7) is the correct model. This will be the case whenever ρ_{11} or ρ_{12} and ρ_{13} or ρ_{14} are zero, because then the data-generating processes (4) will coincide with (3), and consequently (8) will reduce to (7). If a second-order system is estimated nevertheless in such a situation, then four parameters (a_{13}, a_{14}, a_{23} and a_{24}) are estimated which are not in the model, a redundancy which should affect the properties of all other estimates. As for the four incorrectly included parameters, it should be clear by now that the bias of the cross-term estimates $\hat{a}_{14}$ and $\hat{a}_{23}$ is highly likely to be insignificant since a_{14} and a_{23} are zero. On the other hand, $\hat{a}_{13}$ and $\hat{a}_{24}$ are autoregressive parameter estimates, and as such unbiased only by exception (for intermediate eigenvalues). Thus analysts relying too hard on t ratios risk being led to believe that the VAR model is of higher order than it is, since autoregressive parameters which are zero will be significantly estimated for most eigenvalue combinations.

Here however, interest will be concentrated on the first-order estimates in an incorrectly specified second-order model, i.e. on the bias and variance of $\hat{a}_{11}$, $\hat{a}_{12}$, $\hat{a}_{21}$ and $\hat{a}_{22}$, and in particular on how they compare to the bias and variance of the corresponding estimates in the (correct) first-order model. In particular, this form of overparameterization may also be expected to add to the bias and variance of the four first-order estimates since it is again a question of estimating parameters which should not have been included, just like previously in this section.

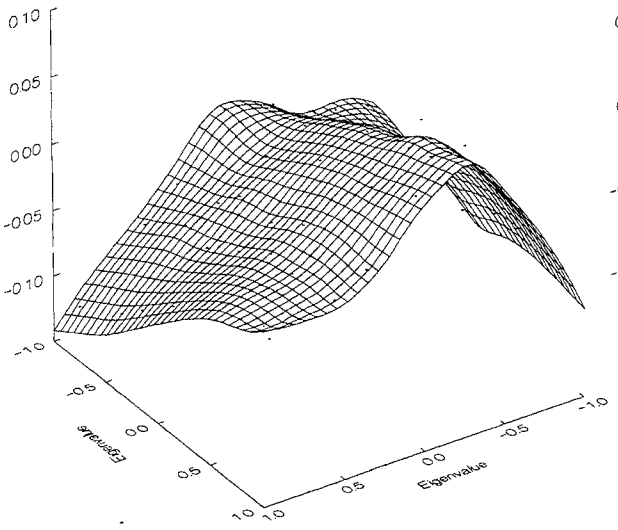

Figure 81: Abs. bias($\hat{a}_{11}$)–Abs. bias($\hat{a}_1$); T=25

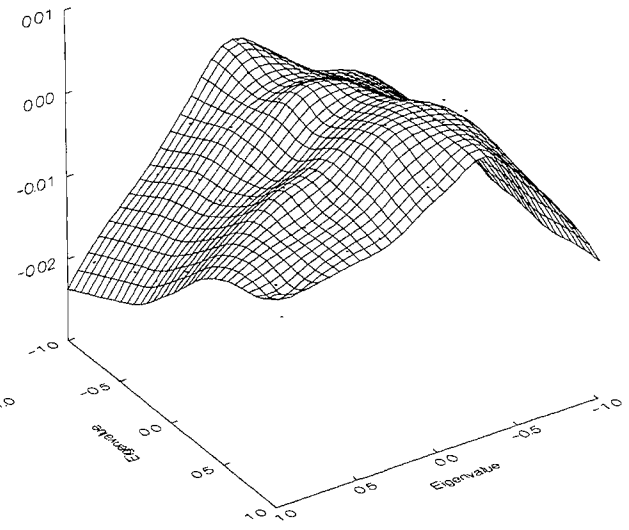

Figure 82: Abs. bias($\hat{a}_{11}$)–Abs. bias($\hat{a}_1$); T=200

But, as Figures 81 and 82 show for the first autoregressive estimate, the difference between the absolute bias of $\hat{a}_{11}$ in (8) and the absolute bias of $\hat{a}_1$ in (7) is negative for most eigenvalue combinations, which means that $\hat{a}_1$ is more biased than $\hat{a}_{11}$ in those regions. Only in the interior are the two estimates more or less equally biased or is $\hat{a}_{11}$ more biased than $\hat{a}_1$. (It should however be pointed out that the results are based on two different sets of generated data, one for (7) and another for (8), but this should affect the results only marginally since the number of replications is so large.)

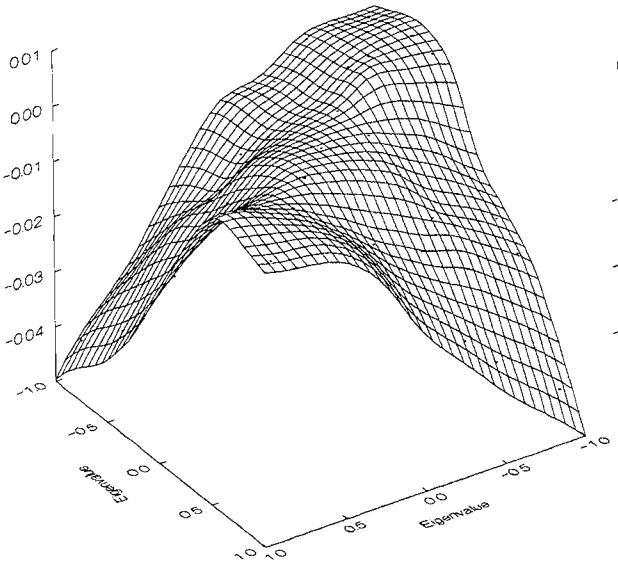

Figure 83: Abs. bias($\hat{a}_{12}$)–Abs. bias($\hat{a}_2$); T=25

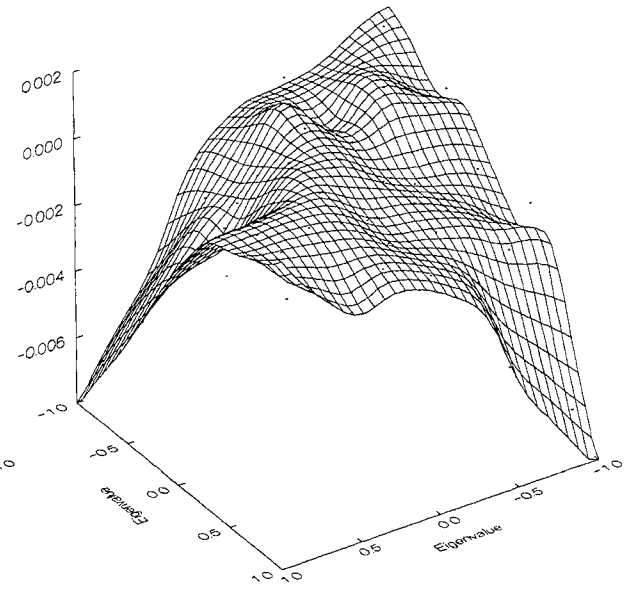

Figure 84: Abs. bias($\hat{a}_{12}$)–Abs. bias($\hat{a}_2$); T=200

The corresponding differences for the cross-term estimates $\hat{a}_2$ and $\hat{a}_{12}$ are given in Figures 83 and 84. Only along a line connecting the (1,1) with the (–.9,–.9) corner are the two estimates equally biased or does the bias of $\hat{a}_{12}$ exceed the bias of $\hat{a}_2$. Recalling Section 3.2, that comes as no surprise since both of them are more or less unbiased

along that line. On both sides of the line, however, $\hat{a}_2$ is apparently more biased than $\hat{a}_{12}$. It is also no surprise that Figures 85 and 86 below are very similar to Figures 83 and 84, since $\hat{a}_3$ and $\hat{a}_{21}$ are also cross-term estimates. Since $\hat{a}_3$ and $\hat{a}_{21}$ are approximately unbiased along the line from (1,1) to (–.9,–.9) they are also more or less equally biased there, whereas $\hat{a}_3$ is more biased than $\hat{a}_{21}$ off that line.

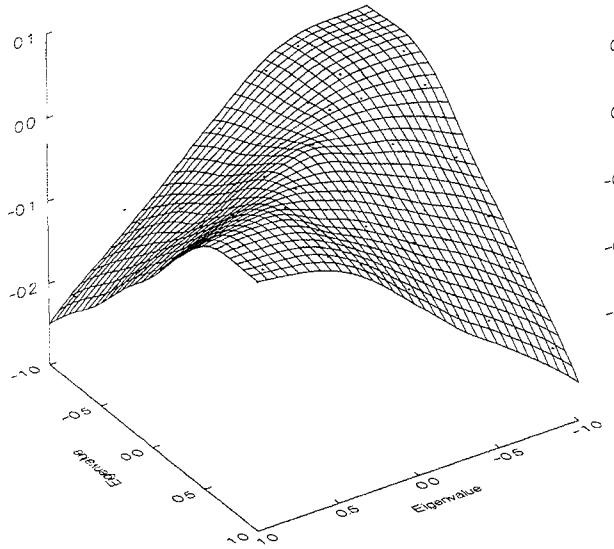

Figure 85: Abs. bias($\hat{a}_{21}$)–Abs. bias($\hat{a}_3$); T=25

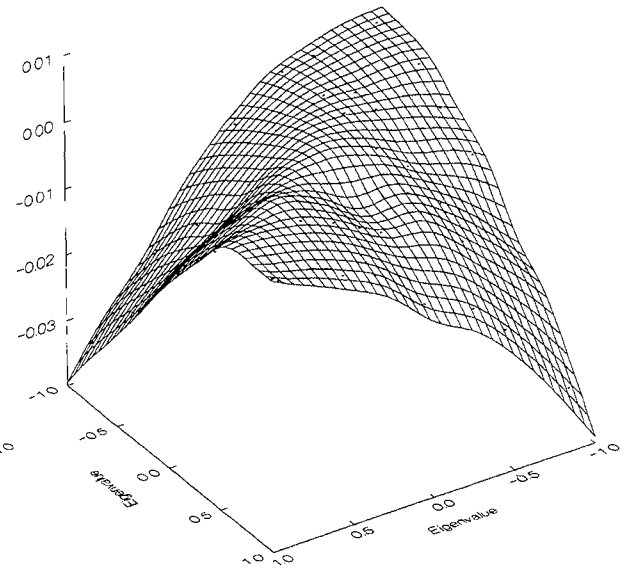

Figure 86: Abs. bias($\hat{a}_{21}$)–Abs. bias($\hat{a}_3$); T=200

Finally, the difference between the absolute bias of $\hat{a}_{22}$ and the absolute bias of $\hat{a}_4$ can be found in Figures 87 and 88. Like for the first pair of autoregressive estimates, $\hat{a}_4$ appears to be more biased than $\hat{a}_{22}$ except for in a small region in the interior, where the difference is positive and thus $\hat{a}_{22}$ is more biased than $\hat{a}_4$. The only difference between these figures and Figures 81 and 82 are the reversed roles played by the eigenvalues.

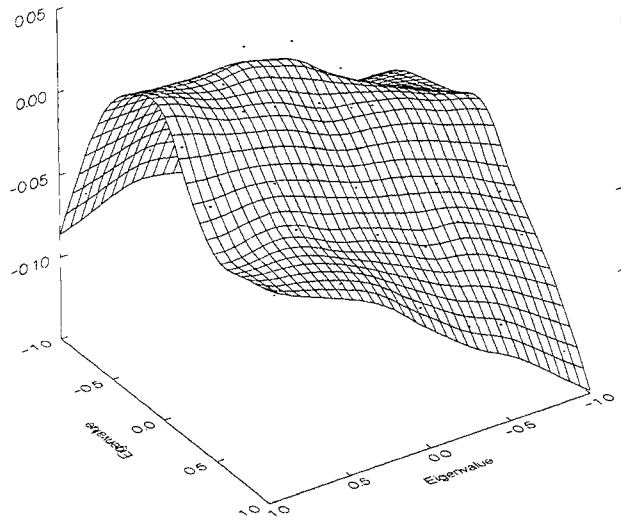

Figure 87: Abs. bias($\hat{a}_{22}$)–Abs. bias($\hat{a}_4$); T=25

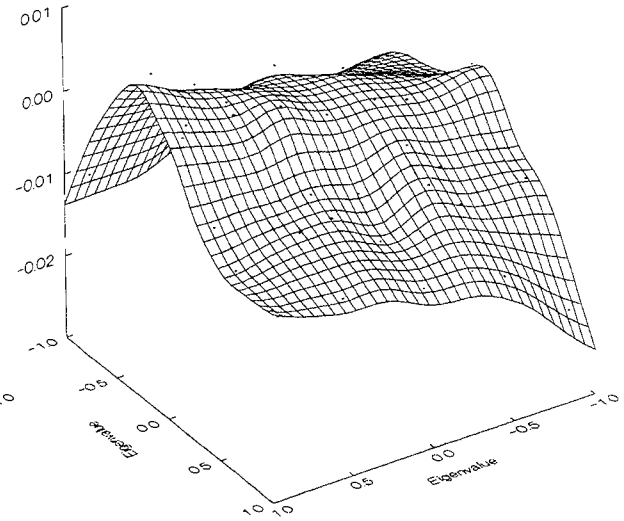

Figure 88: Abs. bias($\hat{a}_{22}$)–Abs. bias($\hat{a}_4$); T=200

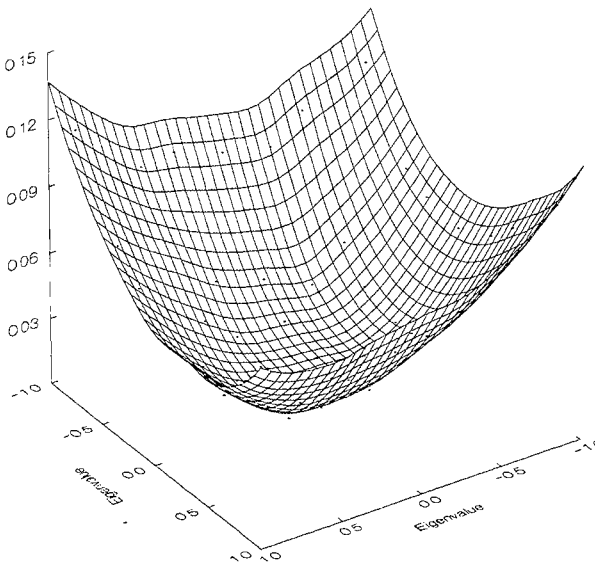

Figure 89: Variance($\hat{a}_{11}$)–Variance($\hat{a}_1$); sample size 25

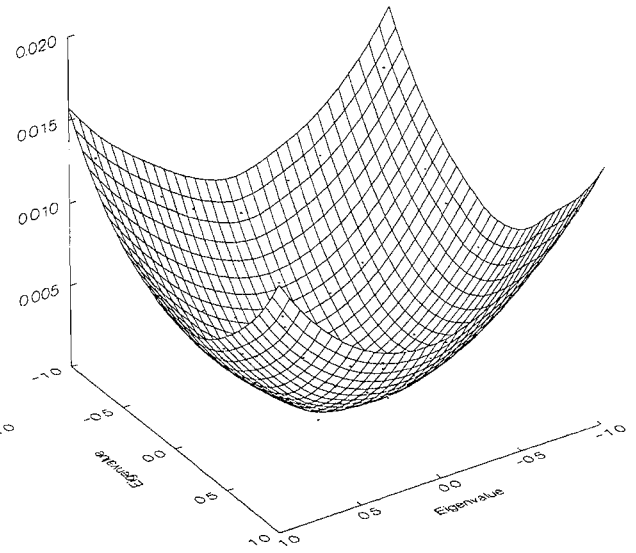

Figure 90: Variance($\hat{a}_{11}$)–Variance($\hat{a}_1$); sample size 200

Thus, estimating a VAR(2) model when the correct model is in fact VAR(1) has an effect on the bias of estimated first-order parameters, but only by exception the expected effect of increasing the bias. Instead it leads to less biased estimates for most eigenvalue combinations. But since estimating (8) when (7) is the correct model also produces four second-order estimates, two of which ($\hat{a}_{14}$ and $\hat{a}_{23}$) are likely to be unbiased and the other two ($\hat{a}_{13}$ and $\hat{a}_{24}$) probably biased, the total bias of an estimated second-order model may still exceed that of a first-order model. Nevertheless, if bias reduction is the only objective (as it will be in Chapters 4 and 5), estimating a model of higher order than the "true" order (or, in practice, than the order suggested by information criteria) and then discarding higher-order estimates appears to be a way to achieve this.

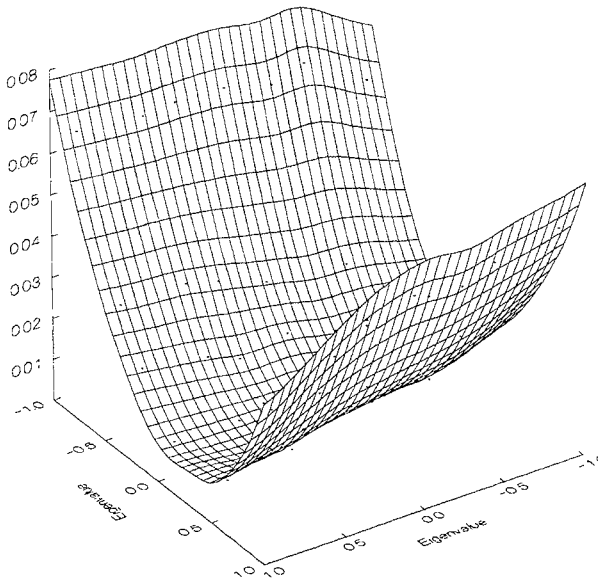

Figure 91: Variance($\hat{a}_{12}$)–Variance($\hat{a}_2$); sample size 25

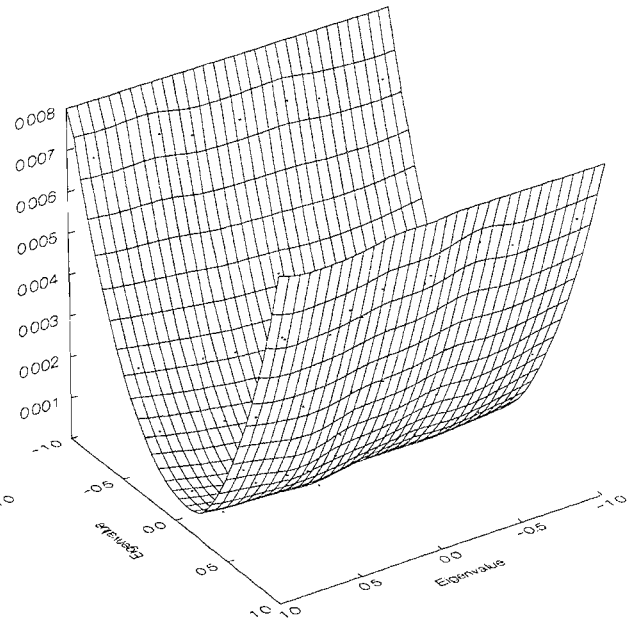

Figure 92: Variance($\hat{a}_{12}$)–Variance($\hat{a}_2$); sample size 200

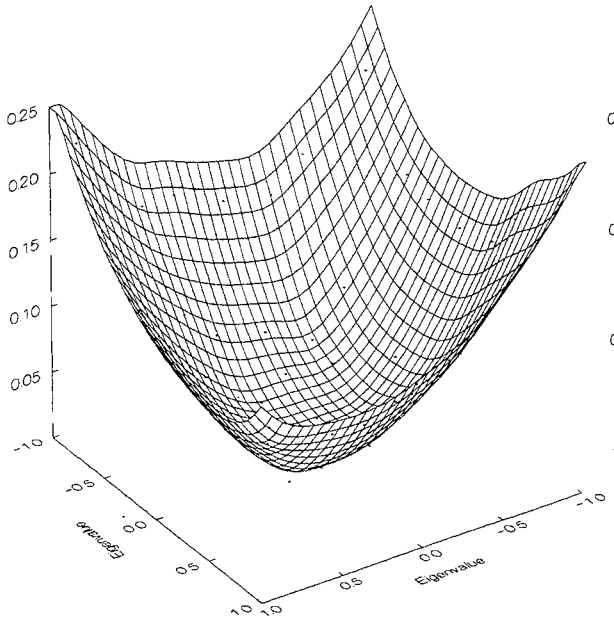

Figure 93: Variance($\hat{a}_{21}$)–Variance($\hat{a}_3$); sample size 25

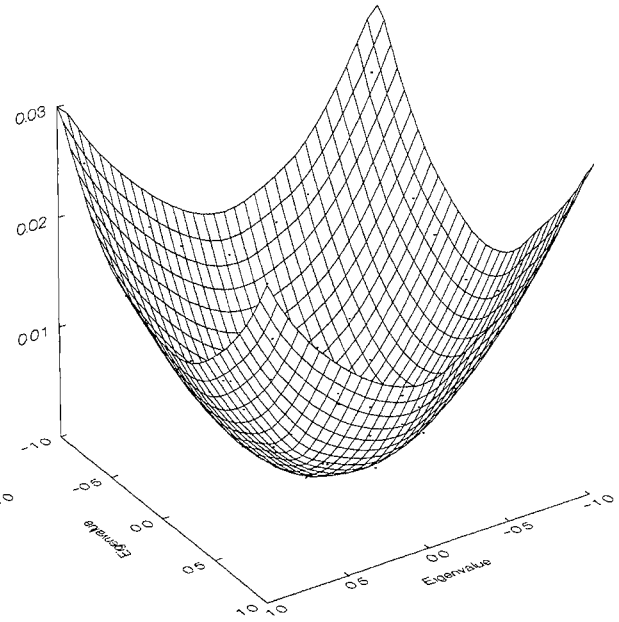

Figure 94: Variance($\hat{a}_{21}$)–Variance($\hat{a}_3$); sample size 200

Before trying to reduce bias by deliberately estimating models of too high an order, it is of course vital to know the effects on variance as well. These effects are illustrated in Figures 89 through 96, in which the sample variance of each first-order estimate in (7) is deducted from the variance of the corresponding first-order estimate in (8). The resulting functions are approximately quadratic functions of both eigenvalues (Figures 89, 90, 93 and 94) or of one eigenvalue (Figures 91, 92, 95 and 96), which is not surprising given that the underlying variance functions were found in Section 3.2 to depend on eigenvalues in that way as well. Since the quadratic curvature is stronger for estimates in (7) than for the estimates in (8), the difference will also be more or less quadratic. However, the most important property of the difference plots is

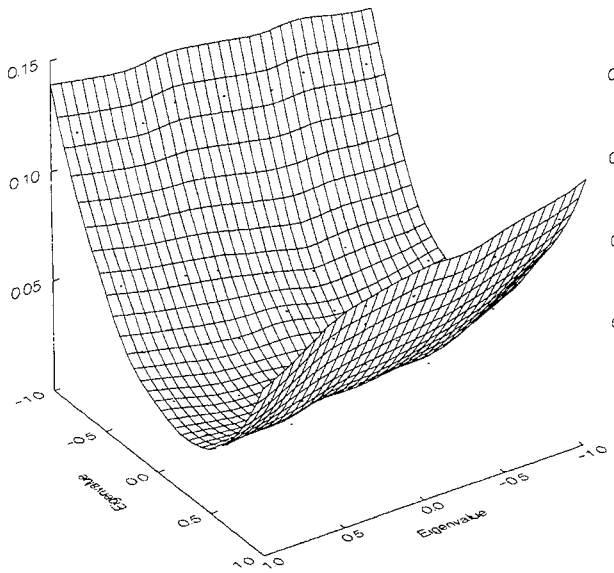

Figure 95: Variance($\hat{a}_{22}$)–Variance($\hat{a}_4$); sample size 25

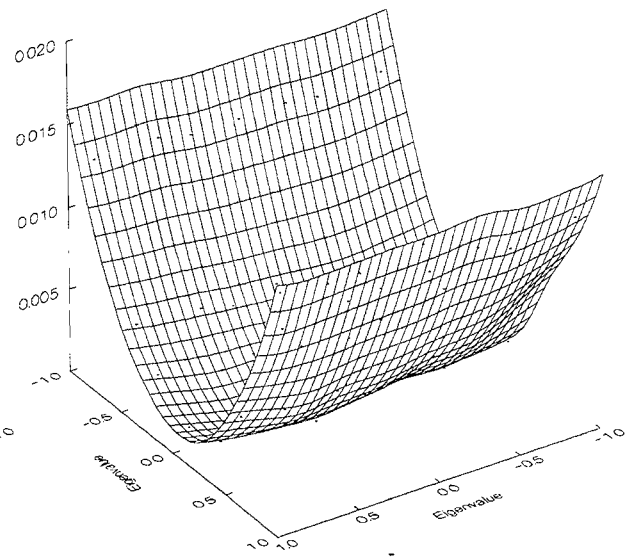

Figure 96: Variance($\hat{a}_{22}$)–Variance($\hat{a}_4$); sample size 200

that they are always positive, thus estimating a second-order model when the correct model is VAR(1) increases the variance of all four first-order estimates for all eigenvalue combinations, in particular in peripheral regions. Bearing in mind that the second-order estimates of (8) have sizeable variances as well, the total variance of (8) is likely to considerably exceed the variance of (7). So whereas this particular form of overparameterization may or may not increase the bias of first-order estimates, it appears to increase their variance almost surely.

3.5 Chapter Summary

The Monte Carlo results accounted for in this chapter are so ample and various that it is reasonable to summarise the most important of them before evaluating the bias approximations in Chapter 4.

(*i*) The bias of first-order estimates is a roughly linear function of eigenvalues; for second-order estimates it is roughly linear in the products of eigenvalues. In most cases it converges by the rate of T (the sample size); possibly faster for "near-integrated" systems and slower for unit root systems.

(*ii*) The sample variance of first-order estimates is a roughly quadratic function of eigenvalues and eigenvalue cross products; for second-order estimates it is a quadratic function of products of eigenvalues. For x_t parameter estimates the variance functions are quadratic in both horizontal dimensions; for y_t estimates only in one dimension. The rate of convergence is higher than T in most cases.

(*iii*) Biases appear to be highly significant for most eigenvalue combinations. Autoregressive estimates are unbiased only for a few intermediate combinations; cross-term estimates only if their parameters cancel out.

(*iv*) Unit roots not only affect the rate of convergence but also the size of bias; second-order (co-)integration in particular adds to the bias of all estimates in the second-order model. First order (co-)integration has no discernible such effects.

(*v*) Incorrectly estimating a VAR model when the true model is two single equations (without cross terms) adds to the bias as well as to the variance of estimates. Estimating a second-order VAR model when the true model is first order adds to the variance of estimates, but may or may not reduce their bias.

4. Bias Approximation and Bias Reduction

Bearing in mind the huge relative biases in Chapter 3, bias should be considered a real problem, and consequently bias reduction ought to lie in the interest of VAR analysts. In this chapter three similar bias approximations will be introduced, evaluated and used to construct bias-reduced estimates. All results in this chapter are based on the same generated data as in Chapter 3 in order to enhance comparison.

Bias approximations have a long history for univariate process estimates, e.g. Leo Hurwicz's asymptotic expansion for the bias of an estimated parameter in a univariate AR(1) process (Hurwicz 1950), or Maurice Kendall's parallel derivation for a univariate AR(1) with non-zero drift (Kendall 1954). John White extended Hurwicz's results even further by deriving expressions for the bias as well as for the variance of the parameter estimate subject to two different initial conditions for the process (White 1961), and L. Shenton and Whitney Johnson used moment generating functions in order to try to replicate Hurwicz's results (Shenton and Johnson 1965). They arrived at a different, basically inferior approximation though, but one which appears to perform better for "near-integrated" processes, and they managed to derive a variance approximation as well. Peter Phillips approximated the entire sample distribution in the AR(1) case (Phillips 1977). Katsuto Tanaka, like Kendall in 1954, examined the case of a univariate AR(1) process with a drift, and also derived the properties of some test statistics in that situation (Tanaka 1983), and finally Koichi Maekawa used Edgeworth expansions to derive approximate expressions for the expectations of estimates in a univariate AR(1) model with exogenous variables (Maekawa 1983, Maekawa 1985).

When it comes to multivariate processes, however, much less is known because only a few results have been published. Dag Tjøstheim and Jostein Paulsen bridged the gap between univariate and multivariate AR processes by deriving the small-sample properties of Yule-Walker estimates for univariate AR(1) and AR(2) processes as well as for multivariate AR(p) processes, and in their article (Tjøstheim and Paulsen 1983) they also suggest how to modify the Yule-Walker results in order to

apply them to least-squares estimates as well (see Section 4.1). Similar (but not identical) results were obtained by Des Nicholls and Adrian Pope even though they attacked the problem in a different way, focusing exclusively on least squares (Nicholls and Pope 1988). Also, Jan Kiviet and Garry Phillips have developed a technique to derive bias approximations in the multivariate case with exogenous variables (Kiviet and Phillips 1991, Kiviet and Phillips 1993) which can be applied to the same problem as Tjøstheim-Paulsen's and Nicholls-Pope's approximations address. In addition, they have recently derived a second-order bias approximation for the situation with exogenous variables (Kiviet and Phillips 1995).

4.1 The Bias Approximations

The three bias approximations (Tjøstheim and Paulsen 1983, Nicholls and Pope 1988, and Brännström 1995*a*) to be evaluated in Section 4.2 are very similar since they are all derived from the Yule-Walker relationship $\mathbf{A} = \mathbf{C}_{-1}\mathbf{C}_0^{-1}$, where $\mathbf{A}$ is the parameter matrix in the VAR(1) representation of a VAR(p) model, $\mathbf{C}_0$ is the autocovariance function of $\mathbf{x}_t$ and $\mathbf{C}_{-1}$ is the covariance function of $\mathbf{x}_t$ and $\mathbf{x}_{t-1}$, i.e. $\mathbf{C}_k = E(\mathbf{x}_t-\mu)(\mathbf{x}_{t+k}-\mu)'$ for k=0 or –1, where $E(\mathbf{x}_t)=E(\mathbf{x}_{t-1})=\mu$.

Tjøstheim and Paulsen (1983) started from the Yule-Walker estimate $\tilde{\mathbf{A}} = \hat{\mathbf{C}}_{-1}\hat{\mathbf{C}}_0^{-1}$, where $\hat{\mathbf{C}}_0$ and $\hat{\mathbf{C}}_{-1}$ are sample estimates of $\mathbf{C}_0$ and $\mathbf{C}_{-1}$. Following a number of Taylor expansions around $\mathbf{C}_0$ and $\mathbf{C}_1$ they managed to approximate $E(\hat{\mathbf{C}}_k - \mathbf{C}_k)$ and $E(\hat{\mathbf{C}}_k - \mathbf{C}_k)\mathbf{C}_0^{-1}(\hat{\mathbf{C}}_l - \mathbf{C}_l)$ (k,l=0,1) in order to eventually arrive at the following approximation for the bias of the Yule-Walker estimate $\tilde{\mathbf{A}}$:

$$E(\tilde{\mathbf{A}} - \mathbf{A}) \approx -\frac{1}{T}\left(2\mathbf{A} + (\mathbf{I}-\mathbf{A})\mathbf{C}_0(\mathbf{I}-\mathbf{A}')^{-1}\mathbf{C}_0^{-1} + \mathbf{G}\mathbf{A}'(\mathbf{I}-\mathbf{A}'\mathbf{A}')^{-1}\mathbf{C}_0^{-1} + \mathbf{G}\sum_{i=1}^{Kp}\rho_i(\mathbf{I}-\rho_i\mathbf{A}')^{-1}\mathbf{C}_0^{-1} \right)$$

where T is the sample size (25, 50, 100 or 200 in the Monte Carlo experiments); $\mathbf{I}$ is the identity matrix of order Kp, where K is the dimension of the process (2 in the experiments) and p is the autoregressive order (1 or 2 in the experiments); $\mathbf{G}$ is the variance-covariance matrix $E(\varepsilon_t\varepsilon_t')$ of the white noise vector ε_t; and ρ_i (i=1,...,Kp) are the eigenvalues of $\mathbf{A}$, i.e. the solutions to the characteristic equation $|\lambda\mathbf{I}-\mathbf{A}|=0$. (In the original paper though, the summation runs only from 1 to K, which is surely an error since a K-dimensional VAR(p) system will have Kp eigenvalues.) Like the two approximations to follow, it includes true parameters, matrices and vectors, which of

course limits its use since these will rarely be known by analysts (except of course for the sample size). However, when evaluating the performance of the approximations in Section 4.2, estimates will also be inserted into the expressions to assess their use under more realistic circumstances. Furthermore, they are all limited to stationary processes, i.e. to eigenvalues $|\rho|<1$, but that restriction will also be eased by using unit-root processes to generate data and then trying to approximate the resulting bias.

Even though it is valid only for Yule-Walker estimates, Tjøstheim and Paulsen suggested that their approximation could be extended to apply to least squares as well following a slight modification, more precisely by adding the first K rows of $(T-p+1)^{-1}\mathbf{A}$ to the above expression. The approximation to the bias of the least-squares estimate $\hat{\mathbf{A}}$ would thus be

$$E(\hat{\mathbf{A}}-\mathbf{A}) \approx -\frac{1}{T}\left(\mathbf{A}+(\mathbf{I}-\mathbf{A})\mathbf{C}_0(\mathbf{I}-\mathbf{A}')^{-1}\mathbf{C}_0^{-1}+\mathbf{G}\mathbf{A}'(\mathbf{I}-\mathbf{A}'\mathbf{A}')^{-1}\mathbf{C}_0^{-1}+\mathbf{G}\sum_{i=1}^{Kp}\rho_i(\mathbf{I}-\rho_i\mathbf{A}')^{-1}\mathbf{C}_0^{-1} \right) \tag{9a}$$

for first-order models (p=1), and

$$E(\hat{\mathbf{A}}-\mathbf{A}) \approx -\frac{1}{T}\left(\frac{T-2}{T-1}\mathbf{A}+(\mathbf{I}-\mathbf{A})\mathbf{C}_0(\mathbf{I}-\mathbf{A}')^{-1}\mathbf{C}_0^{-1}+\mathbf{G}\mathbf{A}'(\mathbf{I}-\mathbf{A}'\mathbf{A}')^{-1}\mathbf{C}_0^{-1}+\mathbf{G}\sum_{i=1}^{Kp}\rho_i(\mathbf{I}-\rho_i\mathbf{A}')^{-1}\mathbf{C}_0^{-1} \right) \tag{9b}$$

for second-order models (p=2). In general, the first term is $-\frac{1}{T}\left(\frac{T-2p+2}{T-p+1}\mathbf{A}+\cdots \right.$ for a VAR(p) model, and then the same consecutive terms as above follow. Whereas the first term of the approximation is $2\mathbf{A}$ for Yule-Walker estimates, it never exceeds $\mathbf{A}$ for least-squares estimates; in fact an increasingly smaller portion of $\mathbf{A}$ enters the expression as the autoregressive order p grows higher.

In their paper, Tjøstheim and Paulsen also performed relatively limited Monte Carlo experiments on univariate AR(2) processes in order to compare the bias of Yule-Walker and least-squares estimates and to assess the properties of (9*b*) in its univariate version. The only finding they reported, however, is that Yule-Walker estimates are more biased than least squares.

The scope of Nicholls and Pope (1988) was to replicate Tjøstheim and Paulsen's results in a mathematically more rigorous manner. They started by defining the two auxiliary matrices $\mathbf{P}$ and $\mathbf{Q}$ so that $\mathbf{P}=(\hat{\mathbf{C}}_1-\mathbf{C}_1)\mathbf{C}_0^{-1}$ and $\mathbf{Q}=(\hat{\mathbf{C}}_0-\mathbf{C}_0)\mathbf{C}_0^{-1}$. It is then easily verified that the least-squares estimate $\hat{\mathbf{A}}$ can be written as

$\hat{\mathbf{A}} = (\mathbf{A}+\mathbf{P})(\mathbf{I}+\mathbf{Q})^{-1}$ (the original paper incorrectly reports **I**+**P** instead of **I**+**Q**, though), and since $(\mathbf{I}+\mathbf{Q})^{-1} = \mathbf{I}-\mathbf{Q}+\mathbf{Q}^2-\mathbf{Q}^3+\cdots$, $\hat{\mathbf{A}}$ can be expressed as the infinite sum $(\mathbf{A}+\mathbf{P})\sum_{i=0}^{\infty}(-1)^i\mathbf{Q}^i$. This representation is very useful since it illustrates how the quality of the derived approximation can be improved by including more **Q** terms. Nicholls and Pope chose to include and derive expectations for all terms up to $\mathbf{AQ}^2$, neglecting $\mathbf{PQ}^2$ and all terms of higher order.

The resulting approximation was found to be

$$E(\hat{\mathbf{A}}-\mathbf{A}) \approx -\frac{1}{T}\mathbf{G}\left((\mathbf{I}-\mathbf{A}')^{-1}+\mathbf{A}'(\mathbf{I}-\mathbf{A}'\mathbf{A}')^{-1}+\sum_{i=1}^{Kp}\rho_i(\mathbf{I}-\rho_i\mathbf{A}')^{-1}\right)\mathbf{C}_0^{-1} \qquad (10)$$

which is similar, but apparently not identical, with (9). It can be shown that (10) is in fact identical with (9*a*) when p is 1 but not with (9*b*) when it is 2, nor for any other higher order p. The reason is the modification of the Yule-Walker bias approximation described above. For first-order systems (p=1) its denominator will be T, like it is for the 2**A** term of the original Yule-Walker approximation. Therefore only **A** remains in (9*a*), and then the sum of its first two terms equals the first term of (10), $\mathbf{G}(\mathbf{I}-\mathbf{A}')\mathbf{C}_0^{-1}$. But for systems of order 2 or higher the denominator in the modification term will be smaller than T, and consequently the sum of the first two terms of (9*b*) will not equal the first term of (10). One important effect of this difference is that Tjøstheim and Paulsen's approximation will diverge from Nicholls and Pope's approximation more the higher p is. Yet it is unclear why they should differ at all, if only marginally. Nicholls and Pope set out to solve the same problem as Tjøstheim and Paulsen did, and even though they used different techniques they should have reached the same result (provided of course that the derivations are correct). One possible explanation is the different sample mean corrections used in the computation of sample covariance functions $\hat{\mathbf{C}}_0$ and $\hat{\mathbf{C}}_{-1}$; where Tjøstheim and Paulsen used T–p in the denominator, Nicholls and Pope used T–1.

A common feature of these two bias approximations is that they include no $O(T^{-2})$ terms; in that sense they are first-order approximations only. This is a consequence of Tjøstheim and Paulsen using first-order Taylor expansions only when approximating $E(\hat{\mathbf{C}}_k-\mathbf{C}_k)\mathbf{C}_0^{-1}(\hat{\mathbf{C}}_l-\mathbf{C}_l)$ and $E(\hat{\mathbf{C}}_k-\mathbf{C}_k)$, and of Nicholls and Pope's decision to halt at the $\mathbf{PQ}^2$ term. But, as Sections 3.2 and 3.3 suggested and Section 4.3 will make clear, even though inverse proportionality to sample size appears to be reasonably correct for most eigenvalue combinations, it becomes increasingly inadequate as eigenvalues approach unity. These situations being too

for, presumably one involving second-order terms as well. Such a term was derived by Brännström (1995*a*) using the same technique as in Nicholls and Pope (1988) but extending the analysis to include the next two terms in their infinite sum, $\mathbf{PQ}^2$ and $\mathbf{AQ}^3$, as well. The expectation of the difference between these two terms turns out to be not too difficult to compute:

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

Nicholls and Pope (1988) show $\hat{\mathbf{C}}_k-E(\hat{\mathbf{C}}_k)$ (where k=0,–1) to be $O_p(T^{-1})$ and $\mathbf{Q}^2$–$E(\mathbf{Q}^2)$ to be $O_p(T^{-2})$, thus the second term on the last row will be $O_p(T^{-3})$ and can be ignored from this point. Decomposition of the first term into $E(\hat{\mathbf{C}}_{-1})\mathbf{C}_0^{-1}E(\mathbf{Q}^2)$–$\mathbf{A}E(\hat{\mathbf{C}}_0)\mathbf{C}_0^{-1}E(\mathbf{Q}^2)$ makes it clear that the expectations of $\hat{\mathbf{C}}_{-1}$, $\hat{\mathbf{C}}_0$ and $\mathbf{Q}^2$ must be approximated. Fortunately, these approximations can all be found in Nicholls and Pope (1988) since they were needed for the derivation of (10):

$$E(\hat{\mathbf{C}}_{-1})=\mathbf{AC}_0-\frac{1}{T}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{G}(\mathbf{I}-\mathbf{A}')^{-1}+O(T^{-2})$$
$$E(\hat{\mathbf{C}}_0)=\mathbf{C}_0-\frac{1}{T}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{G}(\mathbf{I}-\mathbf{A}')^{-1}+O(T^{-2})$$
$$(T-1)E(\mathbf{Q}^2)=(K+1)\mathbf{I}+\mathbf{A}^2(\mathbf{I}-\mathbf{A}^2)^{-1}+\mathbf{C}_0\mathbf{A}'\mathbf{A}'\mathbf{C}_0^{-1}(\mathbf{I}-\mathbf{C}_0\mathbf{A}'\mathbf{A}'\mathbf{C}_0^{-1})^{-1}+$$
$$+\sum_{i=1}^{Kp}\rho_i\left(\mathbf{A}(\mathbf{I}-\rho_i\mathbf{A})^{-1}+\mathbf{C}_0\mathbf{A}'\mathbf{C}_0^{-1}(\mathbf{I}-\rho_i\mathbf{C}_0\mathbf{A}'\mathbf{C}_0^{-1})^{-1}\right)+O(T^{-1})$$

After multiplication of these three expression (and some manipulations), the following second-order terms result:

$$-\frac{1}{T^2}\mathbf{G}(\mathbf{I}-\mathbf{A}')^{-1}\mathbf{C}_0^{-1}\left[(K+1)\mathbf{I}+\mathbf{C}_0\mathbf{A}'\mathbf{A}'\mathbf{C}_0^{-1}(\mathbf{I}-\mathbf{C}_0\mathbf{A}'\mathbf{A}'\mathbf{C}_0^{-1})^{-1}+\mathbf{A}^2(\mathbf{I}-\mathbf{A}^2)^{-1}+\sum\rho_i\left(\mathbf{A}(\mathbf{I}-\rho_i\mathbf{A})^{-1}+\mathbf{C}_0\mathbf{A}'\mathbf{C}_0^{-1}(\mathbf{I}-\rho_i\mathbf{C}_0\mathbf{A}'\mathbf{C}_0^{-1})^{-1}\right)\right]$$

The denominator should really be $T(T-1)$ rather than T^2 for some terms in this expression. Incidentally, this is the case for (10) as well (although Nicholls and Pope pretend not); (10) ought to be $-\mathbf{G}(\mathbf{I}-\mathbf{A}')^{-1}\frac{\hat{\mathbf{C}}_0^{-1}}{T}-\mathbf{G}\{\mathbf{A}'(\mathbf{I}-\mathbf{A}'\mathbf{A}')^{-1}+\Sigma\rho_i(\mathbf{I}-\rho_i\mathbf{A}')^{-1}\}\frac{\hat{\mathbf{C}}_0^{-1}}{T-1}$, but as Brännström (1994*b*) demonstrated, using T throughout in (10) has only marginal effects on the results, at least for samples of 25 observations or more. Therefore using T^2 where it would be more correct to use $T(T-1)$ would appear to be justified as well. It should also be emphasised that the second-order expression cannot be expected to

contain all second-order terms of the bias of $\hat{\mathbf{A}}$, but presumably it includes the leading terms. In fact, along the way towards the derivation of (10) Nicholls and Pope obtained a number of $O(T^{-2})$ remainders, at which they stopped since their objective was a first-order approximation, but which would have to be analysed for a full second-order approximation. Furthermore, it cannot be ruled out that at least some of the terms succeeding $\mathbf{PQ}^2$–$\mathbf{AQ}^3$ may also have second-order expectations, thus it would be necessary to compute expectations for all subsequent terms until finding only terms of lower order.

Even though the second-order term was derived from (10) and using the results and the technique of Nicholls and Pope (1988), it is probably applicable to Tjøstheim and Paulsen's first-order approximations as well. In the remainder however, we will not deal with second-order versions of (9*a*) or (9*b*), only of (10):

$$
\begin{aligned}
E(\hat{\mathbf{A}}-\mathbf{A}) \approx -\frac{1}{T}\mathbf{G}\Bigg(& (\mathbf{I}-\mathbf{A}')^{-1} + \mathbf{A}'(\mathbf{I}-\mathbf{A}'\mathbf{A}')^{-1} + \sum_{i=1}^{Kp}\rho_i(\mathbf{I}-\rho_i\mathbf{A}')^{-1}\Bigg)\mathbf{C}_0^{-1} \\
& -\frac{1}{T^2}\mathbf{G}(\mathbf{I}-\mathbf{A}')^{-1}\mathbf{C}_0^{-1}\Big[(K+1)\mathbf{I} + \mathbf{C}_0\mathbf{A}'\mathbf{A}'\mathbf{C}_0^{-1}\big(\mathbf{I}-\mathbf{C}_0\mathbf{A}'\mathbf{A}'\mathbf{C}_0^{-1}\big)^{-1} + \\
& \mathbf{A}^2\big(\mathbf{I}-\mathbf{A}^2\big)^{-1} + \sum_{i=1}^{Kp}\rho_i\Big(\mathbf{A}(\mathbf{I}-\rho_i\mathbf{A})^{-1} + \mathbf{C}_0\mathbf{A}'\mathbf{C}_0^{-1}\big(\mathbf{I}-\rho_i\mathbf{C}_0\mathbf{A}'\mathbf{C}_0^{-1}\big)^{-1}\Big)\Big]
\end{aligned}
\qquad (11)
$$

It is obviously a more complicated expression than (10), but more importantly, it requires exactly the same information as the two first-order approximations. In other words, if it is possible to compute (9) and (10), it will always be possible to compute (11) as well. The question of whether or not it is worth the extra computations will be answered in the next two sections.

4.2 Performance of Approximations

The most important distinction to make when evaluating (9), (10) and (11) is between their performance based on true parameters and based on estimates. All three bias approximations were derived and formulated in terms of the true parameter matrix $\mathbf{A}$, the true autocovariance function $\mathbf{C}_0$, the true variance-covariance matrix $\mathbf{G}$, the true eigenvalue vector ρ and the sample size T. Except for the sample size, none of these is very likely to be known in real situations, and if the performance of an approximation would turn out to hinge on these matrices and vectors, it would of course be useless for all practical purposes. Therefore, comparisons will be made between bias and

approximations based on true parameters as well as between bias and approximations based on estimates. Another important distinction is between stationary and (co-) integrated systems. As indicated in Section 4.1, although all three approximations are intended for stationary processes only, unit roots will be allowed when evaluating them. However, in unit-root cases there can be no evaluation based on true parameters, since (**I**–**A**) will then be non-invertible, causing (9), (10) and (11) to break down. Therefore only evaluations based on estimates will be possible.

In order to evaluate the performance of the bias approximations, simple linear regression equations of the type $Bias = \alpha + \beta \cdot Approximation + \varepsilon$ will be used. OLS estimates of α and β will be computed and reported together with the squared sample correlation R^2, the idea being that for an approximation to work well, not only should it correlate strongly with the bias, but the estimated intercept a should also be insignificantly close to zero and the estimated slope parameter b insignificantly close to unity. The standard OLS assumptions are invalidated by a correlation between error terms, and probably also between the approximation based on estimates and error terms, but lacking better alternatives evaluations will be based on a, b and R^2 anyway. Results will be reported for each sample size and will be based on 49 observations for (7) if only stationary eigenvalue combinations enter; if the 15 unit-root combinations are also taken into account the number of observations is 64. The corresponding figures for (8) are 784 and 882 observations, respectively.

Table 1 shows the results of regressing the average bias obtained when simulating (7) on the corresponding approximate bias. As indicated before, (9*a*) and (10) are equivalent in this case, therefore the table evaluates both approximations. When based on true parameters (left), the approximations perform excellently, explaining around 98 per cent of the variation in bias. Judging by the t ratios, the "zero intercept-unit slope" requirement is fulfilled (using 5 % significance, t ratios smaller than 1.96 in absolute terms) for all four sample sizes in the cases of a_2 and a_3 and for sample sizes greater than or equal to 50 in the cases of a_1 and a_4. Analysts applying (9*a*) or (10) to (7) can therefore expect to approximate its bias perfectly provided it is stationary, the sample is not too small, and **A**, **G**, ρ and $\mathbf{C}_0$ are all known. But since these are highly unrealistic assumptions, the excellent left-hand results are merely an academic exercise. Instead, more relevant results can be found in the right-hand part, where the average bias has been regressed on the average bias approximation based on estimates. For each of the 10,000 replications, not only was the least-squares estimate $\hat{\mathbf{A}}$ stored, but it was also used to compute $\hat{\rho}$ as the vector of solutions to $|\lambda\mathbf{I}-\hat{\mathbf{A}}|=0$ and, together with the least-squares estimate $\hat{\mathbf{G}}$, to compute $\hat{\mathbf{C}}_0$. All four were then inserted

into the bias approximation formulæ and the results were stored until the end of the simulations, upon which means, variances and covariances were computed. The great benefit of this rather tedious process is that approximations and biases are based on exactly the same generated data in order to make comparisons meaningful.

		Based on true parameters			Based on estimated parameters		
Parameter	Sample size	a	b	R^2	a	b	R^2
	25	−.0069	.9683	.985	−.0084	1.1328	.983
		(−4.04)	(−1.80)		(−4.78)	(6.18)	
a_1	50	−.0019	1.0090	.989	−.0021	1.0919	.989
		(−2.56)	(.59)		(−2.89)	(5.55)	
	100	−.0004	1.0174	.992	−.0004	1.0582	.991
		(−1.18)	(1.27)		(−1.27)	(4.03)	
	200	−.0002	1.0186	.989	−.0002	1.0385	.989
		(−1.23)	(1.20)		(−1.28)	(2.40)	
	25	.0002	.9580	.975	.0002	1.0936	.978
		(.38)	(−1.87)		(.41)	(3.94)	
a_2	50	.0000	1.0023	.986	.0000	1.0730	.988
		(.07)	(.13)		(.07)	(4.20)	
	100	.0000	1.0115	.987	.0000	1.0463	.987
		(.24)	(.69)		(.25)	(2.69)	
	200	.0000	1.0157	.987	.0000	1.0328	.986
		(.66)	(.91)		(.68)	(1.85)	
	25	−.0008	.9580	.975	−.0008	1.0936	.978
		(−.38)	(−1.87)		(−.41)	(3.94)	
a_3	50	−.0001	1.0023	.986	−.0001	1.0730	.988
		(−.07)	(.13)		(−.07)	(4.20)	
	100	−.0001	1.0115	.987	−.0001	1.0463	.987
		(−.24)	(.69)		(−.25)	(2.69)	
	200	−.0001	1.0159	.987	−.0002	1.0316	.986
		(−.59)	(.93)		(−.79)	(1.74)	
	25	−.0062	.9610	.984	−.0078	1.1222	.983
		(−3.57)	(−2.18)		(−4.43)	(5.73)	
a_4	50	−.0020	1.0008	.990	−.0022	1.0834	.991
		(−2.82)	(.06)		(−3.29)	(5.52)	
	100	−.0004	1.0081	.991	−.0004	1.0481	.991
		(−1.07)	(.58)		(−1.17)	(3.29)	
	200	−.0001	1.0055	.991	−.0001	1.0251	.991
		(−.76)	(.40)		(−.79)	(1.77)	

Table 1: OLS estimates of α and β in the model $Bias = \alpha + \beta \cdot Approx. + \varepsilon$, where the approximation to the bias is (9*a*) or (equivalently) (10), based on true (left) and estimated (right) parameters. t ratios in brackets are tests of zero intercept and unit slope, respectively. Based on the 49 stationary eigenvalue combinations of (7), each simulated 10,000 times.

When based on estimates, (9*a*) and (10) perform clearly worse, especially in the unit-slope sense. Sample correlations are still very high, in some cases actually higher than before, but the estimated slopes significantly exceed unity for all four parameters when T is smaller than 200. The interpretation of b being significantly greater than unity is that the approximation understates the bias; if it had been significantly smaller than unity the approximation would have exaggerated it. Thus analysts using (9*a*) or (10) based on estimates to assess the bias of (7) should beware that in samples of less than 200 observations, they will approximate the bias only in part.

Impressive as the results in the left-hand part of the table may be, they apply only to stationary processes (49 eigenvalue combinations). Next we will include the 15 unit-root combinations as well, since an approximation capable of handling integrated and cointegrated cases as well is much more interesting than one whose use is limited to stationary systems only. The results can be found in the following table, giving regression results for all 64 eigenvalue combinations based on estimates.

(No comparison based on true parameters can be made since all three approximations break down if unit roots are inserted, as pointed out earlier. Basing approximations on estimates is possible though, at least as long as no eigenvalue estimate is exactly at unity.)

		Based on 49 estimates			Based on 64 estimates		
Parameter	Sample size	a	b	R^2	a	b	R^2
	25	−.0084	1.1328	.983	−.0115	1.2708	.955
		(−4.78)	(6.18)		(−3.61)	(7.69)	
a_1	50	−.0021	1.0919	.991	−.0032	1.2163	.961
		(−2.89)	(5.55)		(−2.07)	(7.00)	
	100	−.0004	1.0582	.991	−.0008	1.1814	.988
		(−1.27)	(4.03)		(−1.02)	(6.02)	
	200	−.0002	1.0385	.989	−.0000	1.1610	.983
		(−1.28)	(2.40)		(−.00)	(5.03)	
	25	.0002	1.0936	.978	.0000	1.1742	.956
		(.41)	(3.94)		(.08)	(5.42)	
a_2	50	.0000	1.0730	.988	.0002	1.1588	.954
		(.07)	(4.20)		(.51)	(4.91)	
	100	.0000	1.0463	.987	.0001	1.1481	.950
		(.25)	(2.69)		(.47)	(4.20)	
	200	.0000	1.0328	.986	−.0000	1.1224	.912
		(.68)	(1.85)		(−.13)	(2.76)	
	25	−.0008	1.0936	.978	−.0009	1.1867	.970
		(−.41)	(3.94)		(−.36)	(7.06)	
a_3	50	−.0001	1.0730	.988	−.0002	1.1791	.970
		(−.07)	(4.20)		(−.12)	(6.86)	
	100	−.0001	1.0463	.987	−.0001	1.1636	.962
		(−.25)	(2.69)		(−.10)	(5.54)	
	200	−.0002	1.0316	.986	−.0001	1.1471	.954
		(−.79)	(1.74)		(−.27)	(4.62)	
	25	−.0078	1.1221	.983	−.0086	1.2187	.971
		(−4.43)	(5.73)		(−3.38)	(8.24)	
a_4	50	−.0022	1.0834	.991	−.0023	1.1873	.974
		(−3.29)	(5.52)		(−1.79)	(7.53)	
	100	−.0004	1.0481	.991	−.0005	1.1627	.967
		(−1.17)	(3.29)		(−.64)	(5.93)	
	200	−.0001	1.0251	.991	−.0002	1.1369	.963
		(−.79)	(1.77)		(−.54)	(4.82)	

Table 2: OLS estimates of α and β for (9a) and (10) based on estimates, without (49 observations; left) and with (64 observations; right) unit-root combinations. t ratios in brackets refer to tests of zero intercept and unit slope, respectively.

The left-hand half of the table holds the same results as before, i.e. OLS results for bias regressed on (9*a* or 10) based on estimates for 49 stationary eigenvalue combinations, and the right-hand half holds the same results for all 64 combinations. It is clear that the results deteriorate considerably when the 15 unit-root observations enter the regressions; sample correlations are smaller and all slope estimates significantly exceed unity, even for 200 observations. However, that does not mean that bias approximations cannot be used in such situations. The approximations are likely to perform reasonably but inadequately in unit-root cases since estimates will then be severely biased (the same effect can be seen on a smaller scale for eigenvalues of .9).

The mainly insignificant estimates of α indicate that the entire problem could be solved by simply inflating the bias approximations by some unknown factor in (co-)integrated cases, thereby placing them on the same unit-slope line as for stationary combinations.

		(9*a*) and (10)			(11)		
Parameter	Sample size	*a*	*b*	R^2	*a*	*b*	R^2
	25	−.0069 (−4.04)	.9683 (−1.80)	.985	−.0027 (−2.95)	.9641 (−3.95)	.996
a_1	50	−.0019 (−2.56)	1.0090 (.59)	.989	−.0008 (−1.43)	1.0077 (.69)	.994
	100	−.0004 (−1.18)	1.0174 (1.27)	.992	−.0001 (−.31)	1.0178 (1.45)	.993
	200	−.0002 (−1.23)	1.0186 (1.20)	.989	−.0001 (−.78)	1.0183 (1.17)	.989
	25	.0002 (.38)	.9580 (−1.87)	.975	.0002 (.75)	.9616 (−3.44)	.994
a_2	50	.0000 (.07)	1.0023 (.13)	.986	.0000 (.10)	1.0044 (.35)	.993
	100	.0000 (.24)	1.0115 (.69)	.987	.0000 (.27)	1.0124 (.82)	.990
	200	.0000 (.66)	1.0157 (.91)	.987	.0000 (.67)	1.0157 (.91)	.987
	25	−.0008 (−.38)	.9580 (−1.87)	.975	−.0003 (−.21)	.9414 (−4.12)	.989
a_3	50	−.0001 (−.07)	1.0023 (.13)	.986	−.0001 (−.11)	1.0007 (.07)	.994
	100	−.0001 (−.24)	1.0115 (.69)	.987	−.0001 (−.28)	1.0118 (.82)	.991
	200	−.0001 (−.59)	1.0159 (.93)	.987	−.0001 (−.59)	1.0156 (.90)	.987
	25	−.0062 (−3.57)	.9610 (−2.18)	.984	−.0015 (−1.52)	.9674 (−3.24)	.995
a_4	50	−.0020 (−2.82)	1.0008 (.06)	.990	−.0008 (−1.49)	1.0049 (.46)	.995
	100	−.0004 (−1.07)	1.0081 (.58)	.991	−.0001 (−.27)	1.0095 (.76)	.993
	200	−.0001 (−.76)	1.0055 (.40)	.991	−.0001 (−.32)	1.0056 (.42)	.992

Table 3: OLS estimates of α and β, where approximations (9*a*) and (10) (left) as well as (11) (right) are based on true parameters for 49 stationary eigenvalue combinations, each simulated 10,000 times. *t* ratios in brackets refer to tests of zero intercept and unit slope, respectively.

Bearing in mind the superb performance of (9*a*) and (10) based on true parameters as reflected in Table 1, there may not appear to be a need for a second-order approximation to model the bias of $\hat{\mathbf{A}}$ in (7). On the other hand, even though (11) was derived by Brännström (1995*a*) to handle irregularities in cases where first-order approximations would fail, it is vital that it also has no detrimental effect when first-order approximations are adequate, as they appear to be in Table 1. Furthermore, the performance of (9*a*) and (10) based on estimates could definitely be improved.

As expected, adding second-order terms to (10) has only small effects on estimates and sample correlations in Table 3 above. Correlations increase from already high levels, but in terms of inference about intercepts and slope parameters basically the same results as before emerge. Contrary to (9*a*) and (10) though, (11) significantly overstates the bias of $\hat{a}_2$ and $\hat{a}_3$ for T=25. On the other hand (11) appears to be the better bias approximation when the sample size is 50 since the intercept estimate is insignificant for all four parameters, which it is not for (9*a*) and (10).

		(9*a*) and (10)			(11)		
Parameter	Sample size	*a*	*b*	R^2	*a*	*b*	R^2
	25	−.0084	1.1328	.983	−.0054	1.0262	.973
		(−4.78)	(6.18)		(−2.39)	(1.05)	
a_1	50	−.0021	1.0919	.991	−.0013	1.0326	.991
		(−2.89)	(5.55)		(−1.83)	(2.23)	
	100	−.0004	1.0582	.991	−.0002	1.0161	.993
		(−1.27)	(4.03)		(−.59)	(1.33)	
	200	−.0002	1.0385	.989	.0003	1.0127	.989
		(−1.28)	(2.40)		(1.55)	(.82)	
	25	.0002	1.0936	.978	.0008	.8703	.961
		(.41)	(3.94)		(1.39)	(−5.07)	
a_2	50	.0000	1.0730	.988	.0003	.9663	.962
		(.07)	(4.20)		(1.11)	(−1.21)	
	100	.0000	1.0463	.987	.0001	.9765	.956
		(.25)	(2.69)		(.59)	(−.77)	
	200	.0000	1.0328	.986	−.0000	.9585	.959
		(.68)	(1.85)		(−.39)	(−1.41)	
	25	−.0008	1.0936	.978	−.0042	.8876	.983
		(−.41)	(3.94)		(−2.35)	(−6.67)	
a_3	50	−.0001	1.0730	.988	−.0003	.9857	.995
		(−.07)	(4.20)		(−.54)	(−1.35)	
	100	−.0001	1.0463	.987	−.0001	.9950	.991
		(−.25)	(2.69)		(−.29)	(−.36)	
	200	−.0002	1.0316	.986	−.0002	1.0040	.986
		(−.79)	(1.74)		(−.78)	(.22)	
	25	−.0078	1.1222	.983	−.0038	.9988	.989
		(−4.43)	(5.73)		(−2.64)	(−.08)	
a_4	50	−.0020	1.0834	.991	−.0010	1.0180	.996
		(−3.29)	(5.52)		(−2.12)	(1.92)	
	100	−.0004	1.0481	.991	−.0000	1.0054	.994
		(−1.17)	(3.29)		(−.14)	(.48)	
	200	−.0001	1.0251	.991	−.0000	1.0019	.992
		(−.79)	(1.77)		(−.26)	(.14)	

Table 4: OLS estimates of α and β, where approximations (9*a*) and (10) (left) as well as (11) (right) are based on estimates for 49 stationary eigenvalue combinations, each simulated 10,000 times. *t* ratios in brackets refer to tests of zero intercept and unit slope, respectively.

Next the performance of (11) based on estimates will be evaluated. As Table 4 above demonstrates, estimated intercepts and slopes are much improved by adding second-order terms to the approximation. Unfortunately (11) overadjusts in the case of a_2 and a_3 for T=25, leading to slope estimates significantly smaller than unity, but for series of more than 50 observations the second-order approximation based on estimates appears to model the bias perfectly.

Before looking at the same table but with the 15 unit-root cases added, a word of warning is due. It appears that (11) is much more sensitive to eigenvalues estimated close to unity than (9*a*) or (10), because occasionally the sample mean over the 10,000 replications turned out to be extremely large even though all other simulation results were normal. Most likely this is a consequence of inverting more complicated functions of **A** and ρ in (11) than in (9*a*) and (10). Such obviously implausible results have been discarded from Table 5 below (a total 12 out of 4·15=60 unit-root combinations, although there are probably less obvious cases which have not been discarded but may distort the results below) and for reasons of comparison the same observations were discarded in the reference left-hand part of the table as well. Therefore the reference results on the left are not identical with those presented in previous tables. The most important consequence though is for analysts wishing to approximate or reduce the bias of estimated unit-root systems. Apparently (11) is not altogether reliable in these cases, sometimes yielding strange results, and in a real situation it will of course be impossible to use only regular results, as has been done here.

Bearing these reservations in mind, (11) does seem to remedy most of the bias not approximated by the constantly understating approximations (9*a*) and (10). For $\hat{a}_2$ and $\hat{a}_3$ the estimated slopes in Table 5 differ significantly from unity for all four sample sizes when (9*a*) and (10) are applied, but only for the smallest sample size when (11) approximates the bias. (Once again (11) exaggerates the bias in the T=25 case, causing b to be significantly smaller than one.) For $\hat{a}_1$ and $\hat{a}_4$ slope estimates significantly exceed unity at least for the three smallest sample sizes (for all sample sizes in the case of $\hat{a}_1$), thus there is still some bias left after (11) has been applied but much less than in the case of (9*a*) and (10). Consequently the simulation results favour using (11) rather than (9*a*) and (10) to reduce the bias of (7) in stationary as well as in unit-root cases, at least as long as the sensitivity of (11) to underlying unit roots does not manifest itself in nonsensical results.

Parameter	Sample size	(9*a*) and (10) *a*	*b*	R^2	(11) *a*	*b*	R^2
	25	−.0117 (−3.58)	1.2564 (6.91)	.953	−.0060 (−2.25)	1.0362 (1.54)	.970
a_1	50	−.0026 (−1.76)	1.2000 (6.40)	.963	−.0016 (−1.45)	1.0666 (3.41)	.981
	100	−.0009 (−1.13)	1.1662 (5.24)	.960	−.0004 (−.70)	1.0504 (2.90)	.984
	200	.0001 (.23)	1.1151 (4.04)	.967	.0002 (1.15)	1.0335 (2.29)	.989
	25	.0004 (.73)	1.1425 (4.34)	.956	−.0010 (−.90)	.8505 (−3.29)	.954
a_2	50	.0000 (.02)	1.1448 (4.05)	.948	.0004 (1.22)	.9901 (−.39)	.963
	100	.0001 (.34)	1.1044 (2.68)	.934	.0000 (.01)	.9872 (−.38)	.935
	200	−.0000 (−.01)	1.1093 (2.06)	.917	−.0001 (−.80)	.9953 (−.13)	.931
	25	−.0023 (−.93)	1.1533 (5.66)	.970	−.0037 (−1.77)	.9304 (−4.14)	.983
a_3	50	.0009 (.65)	1.1552 (5.82)	.971	−.0009 (−1.01)	1.0023 (.16)	.988
	100	−.0002 (−.31)	1.1157 (3.70)	.957	−.0000 (−.02)	1.0159 (.84)	.981
	200	.0001 (.28)	1.1057 (3.35)	.959	−.0002 (−1.02)	1.0168 (1.21)	.990
	25	−.0089 (−3.41)	1.2099 (7.71)	.971	−.0043 (−2.16)	.9958 (−.25)	.984
a_4	50	−.0020 (−1.62)	1.1723 (6.93)	.975	−.0009 (−1.25)	1.0455 (3.68)	.992
	100	−.0007 (−.92)	1.1601 (5.83)	.969	−.0001 (−.25)	1.0342 (2.34)	.989
	200	−.0002 (−.54)	1.0899 (3.55)	.971	−.0001 (−.40)	1.0164 (1.26)	.991

Table 5: OLS estimates of α and β, where approximations (9*a*) and (10) (left) as well as (11) (right) are based on estimates for around 60 stationary eigenvalue combinations, each simulated 10,000 times. *t* ratios in brackets refer to tests of zero intercept and unit slope, respectively.

The following two tables concern the eight estimates of the second-order system (8). As already stated, Tjøstheim and Paulsen's approximation will differ from Nicholls and Pope's for every order p greater than one, but for p=2 the difference is smaller than for higher-order systems. In fact, comparing (9*a*), (9*b*) and (10) makes it clear that the difference between (9*b*) and (10) is just the difference between the $\frac{T-2}{T-1}\mathbf{A}$ term in (9*b*) and $\mathbf{A}$ in (9*a*) since (9*a*) is equivalent with (10). Thus (9*b*) and (10) will differ only by $(T-1)^{-1}\mathbf{A}$, which can be sizeable in very small samples but not really for the four sample sizes chosen here. Nevertheless, it is of course of interest to see what effects (if any) this difference has on the accuracy of the two approximations, and whether it is possible to say that one of them performs better than the other. Therefore there are three columns of regression results in Table 6 below, in which approximations are based on true parameters; one for (9*b*), one for (10) and one for (11). Table 7 holds corresponding results for approximations based on estimates.

Parameter	Sample size	(9b) a	(9b) b	(9b) R^2	(10) a	(10) b	(10) R^2	(11) a	(11) b	(11) R^2
	25	−.0010	1.3998	.886	−.0017	1.3798	.896	.0027	1.2092	.905
		(−.83)	(22.3)		(−1.57)	(22.7)		(2.54)	(15.0)	
a_{11}	50	−.0003	1.2489	.937	−.0005	1.2387	.940	.0008	1.1662	.951
		(−.77)	(21.5)		(−1.33)	(21.3)		(2.32)	(17.6)	
	100	−.0001	1.1309	.964	−.0002	1.1259	.965	.0001	1.0934	.971
		(−1.17)	(16.8)		(−1.57)	(16.4)		(.98)	(13.8)	
	200	−.0000	1.0684	.959	−.0000	1.0659	.959	.0000	1.0501	.961
		(−.40)	(8.71)		(−.59)	(8.42)		(.57)	(6.67)	
	25	−.0002	1.0944	.424	−.0002	1.1365	.478	.0001	.8943	.642
		(−.79)	(2.07)		(−.83)	(3.23)		(.50)	(−3.87)	
a_{12}	50	−.0001	1.1051	.599	−.0001	1.1088	.618	.0000	.9895	.733
		(−.62)	(3.26)		(−.63)	(3.50)		(.11)	(−.43)	
	100	−.0000	1.0628	.579	−.0000	1.0604	.584	.0000	1.0017	.670
		(−.26)	(1.94)		(−.25)	(1.89)		(.36)	(.06)	
	200	−.0001	1.0158	.432	−.0001	1.0131	.433	−.0000	.9736	.490
		(−1.68)	(.38)		(−1.68)	(.32)		(−.60)	(−.65)	
	25	.0252	1.1664	.971	.0243	1.1572	.971	.0262	1.0699	.952
		(28.1)	(23.1)		(27.2)	(22.0)		(22.5)	(8.16)	
a_{13}	50	.0075	1.1135	.984	.0073	1.1092	.984	.0080	1.0720	.982
		(23.7)	(22.6)		(23.1)	(21.8)		(23.5)	(13.7)	
	100	.0016	1.0534	.990	.0016	1.0514	.990	.0018	1.0349	.991
		(14.0)	(14.4)		(13.5)	(13.9)		(15.5)	(9.67)	
	200	.0003	1.0186	.986	.0003	1.0176	.986	.0004	1.0098	.986
		(4.54)	(4.28)		(4.37)	(4.06)		(5.21)	(2.28)	
	25	.0000	1.1820	.965	.0000	1.1728	.965	.0008	1.0879	.947
		(.21)	(22.6)		(.21)	(21.6)		(3.23)	(9.53)	
a_{14}	50	−.0001	1.1170	.967	−.0001	1.1128	.967	.0001	1.0783	.966
		(−.70)	(16.0)		(−.70)	(15.5)		(1.31)	(10.8)	
	100	.0000	1.0473	.957	.0000	1.0454	.957	.0000	1.0306	.958
		(.07)	(5.95)		(.07)	(5.72)		(.98)	(3.96)	
	200	−.0000	1.0102	.924	−.0000	1.0093	.924	−.0000	1.0024	.924
		(−.72)	(.99)		(−.72)	(.90)		(−.38)	(.24)	
	25	.0005	1.0877	.457	.0005	1.1293	.516	−.0025	.8852	.664
		(.38)	(2.08)		(.39)	(3.31)		(−2.57)	(−5.11)	
a_{21}	50	.0002	1.1015	.691	.0002	1.1064	.714	−.0007	.9828	.806
		(.41)	(3.86)		(.43)	(4.26)		(−2.16)	(−1.00)	
	100	.0000	1.0782	.821	.0000	1.0761	.828	−.0002	1.0122	.868
		(.14)	(4.35)		(.14)	(4.35)		(−1.70)	(.87)	
	200	.0000	1.0321	.796	.0000	1.0299	.797	−.0001	.9958	.809
		(.03)	(1.72)		(.03)	(1.61)		(−.75)	(−.24)	
	25	−.0009	1.4134	.887	−.0017	1.3931	.898	.0063	1.1931	.902
		(−.80)	(23.0)		(−1.56)	(23.4)		(5.58)	(13.7)	
a_{22}	50	−.0003	1.2569	.939	−.0005	1.2465	.942	.0017	1.1653	.954
		(−.93)	(22.5)		(−1.51)	(22.4)		(5.20)	(18.0)	
	100	−.0001	1.1370	.968	−.0002	1.1318	.969	.0004	1.0963	.975
		(−1.11)	(18.6)		(−1.54)	(18.2)		(3.50)	(15.3)	
	200	−.0001	1.0619	.964	−.0002	1.0594	.964	−.0000	1.0425	.966
		(−2.46)	(8.45)		(−2.67)	(8.15)		(−.44)	(6.08)	
	25	−.0002	1.1887	.976	−.0002	1.1794	.976	−.0040	1.0935	.957
		(−0.26)	(28.7)		(−.26)	(27.4)		(−3.64)	(11.3)	
a_{23}	50	.0001	1.1227	.986	.0001	1.1184	.986	−.0008	1.0833	.984
		(.44)	(25.9)		(.44)	(25.1)		(−2.54)	(16.8)	
	100	−.0000	1.0601	.990	−.0000	1.0581	.990	−.0002	1.0428	.991
		(−.08)	(16.1)		(−.08)	(15.6)		(−2.07)	(12.0)	
	200	.0000	1.0223	.986	.0000	1.0213	.986	−.0000	1.0141	.987
		(.70)	(5.19)		(.70)	(4.97)		(−.12)	(3.36)	
	25	.0258	1.1710	.967	.0248	1.1618	.967	.0285	1.0478	.929
		(26.9)	(22.2)		(26.5)	(21.2)		(19.7)	(4.61)	
a_{24}	50	.0073	1.1136	.984	.0070	1.1093	.984	.0085	1.0632	.976
		(22.9)	(22.4)		(22.4)	(21.7)		(21.8)	(10.7)	
	100	.0016	1.0550	.990	.0016	1.0530	.990	.0020	1.0333	.989
		(13.7)	(14.3)		(13.3)	(13.8)		(15.9)	(8.52)	
	200	.0003	1.0218	.986	.0003	1.0209	.986	.0004	1.0117	.986
		(4.29)	(4.97)		(4.11)	(4.76)		(5.63)	(2.69)	

Table 6: OLS estimates of α and β in the model *Bias* = $\alpha + \beta \cdot$*Approx.* + ε, where the approximations to the bias are, from left to right, (9*b*), (10) and (11), all based on true parameters. t ratios refer to tests of zero intercept and unit slope, respectively. Based on 784 stationary eigenvalue combinations, each simulated 10,000 times.

The regression results for the approximations based on true parameters applied to (8) are much less impressive than when applied to (7). Recall that (9*a*), (10) and (11) appeared to model the bias of (7) almost perfectly for sample sizes 50, 100 and 200, while overadjusting in a sample of 25 observations. Here however, all three approximations severely understate the bias of autoregressive estimates ($\hat{a}_{11}$, $\hat{a}_{13}$, $\hat{a}_{22}$ and $\hat{a}_{24}$) for all four sample sizes. Judging by the high t ratios even for T=200, it would appear that much longer series than 200 are required for the bias approximations to achieve the desired unit slopes and zero intercepts. Still, the strong correlation between approximation and bias for autoregressive estimates indicates that the approximations may nevertheless be apt, except for the fact that they need scaling up. In the case of the second-order autoregressive estimates $\hat{a}_{13}$ and $\hat{a}_{24}$ there is also the problem of highly significant intercept estimates a, meaning that the approximations will incorrectly indicate $\hat{a}_{13}$ and $\hat{a}_{24}$ as biased when in fact they are unbiased, and vice versa.

The performance of the approximations with regard to the second-order cross-term estimates $\hat{a}_{14}$ and $\hat{a}_{23}$ is very similar to their performance for $\hat{a}_{11}$ and $\hat{a}_{22}$, i.e. high correlations, insignificant intercept estimates and slope estimates significantly in excess of unity (although for $\hat{a}_{14}$, b is not significantly greater than one when T is 200). The most serious failure of the approximations in Table 6 is for the first-order cross-terms $\hat{a}_{12}$ and $\hat{a}_{21}$ though. In particular, squared correlations are alarmingly low, only between 40 and 80 per cent for (9*b*) and (10). Though they are considerably higher for (11), indicating that the failure may in part at least be due to second-order effects, they are still considerably lower than for the other six estimates. It is no comfort that the t ratios of some slope estimates do not differ significantly from unity, since that is mainly an effect of the poor explanatory power of the three approximations. Why they should perform so much worse for $\hat{a}_{12}$ and $\hat{a}_{21}$ than for $\hat{a}_{14}$ and $\hat{a}_{23}$ (or, for that matter, for $\hat{a}_2$ and $\hat{a}_3$ of the first-order system) remains a puzzle.

As for the relative performance of (9*b*), (10) and (11), the first two yield very similar regression results, which is consistent with the small differences between (9*b*) and (10) when p is only 2. Therefore it is difficult to decide which one is better, but since the high t ratios of b are slightly smaller for (10) than for (9*b*) and the correlations generally are stronger for (10) as well, (10) appears to be preferable to (9*b*) for second-order bivariate VAR models (to obtain stronger evidence, models of much higher order than 2 would have to be simulated). The comparison between (10) and (11) is more obvious, and basically in favour of the latter. Slope estimates as well as their t ratios are much smaller for (11), thus it understates the bias less on average, and correlations are also higher in most cases.

Parameter	Sample size	(9b) a	(9b) b	(9b) R^2	(10) a	(10) b	(10) R^2	(11) a	(11) b	(11) R^2
	25	−.0005	1.8254	.913	−.0024	1.7455	.936	.0039	1.5217	.931
		(−.44)	(38.2)		(−2.85)	(44.2)		(4.29)	(34.0)	
a_{11}	50	−.0002	1.4119	.942	−.0006	1.3875	.952	.0012	1.2932	.962
		(−.40)	(30.6)		(−1.80)	(34.2)		(4.06)	(31.3)	
	100	−.0001	1.2032	.962	−.0001	1.1950	.968	.0003	1.1516	.975
		(−.55)	(22.2)		(−1.19)	(24.9)		(2.93)	(22.7)	
	200	.0000	1.1049	.959	.0000	1.1045	.958	.0001	1.0836	.960
		(.02)	(11.1)		(.04)	(11.1)		(1.82)	(10.6)	
	25	−.0006	1.6468	.592	−.0004	1.6748	.689	−.0002	1.0226	.618
		(−2.40)	(12.5)		(−2.05)	(16.1)		(−.96)	(.75)	
a_{12}	50	−.0002	1.2846	.625	.0000	1.2958	.656	.0001	1.0646	.684
		(−1.91)	(7.55)		(.27)	(8.62)		(.80)	(2.44)	
	100	−.0001	1.1443	.564	−.0000	1.1633	.583	−.0000	1.0605	.596
		(−1.31)	(3.78)		(−.33)	(4.63)		(−.39)	(1.93)	
	200	−.0001	.9617	.375	−.0000	1.0788	.429	−.0000	1.0350	.434
		(−1.85)	(−.81)		(−1.48)	(1.77)		(−1.48)	(.83)	
	25	.0254	1.6272	.945	.0219	1.6026	.939	.0244	1.4794	.907
		(18.7)	(41.6)		(16.6)	(39.9)		(14.7)	(27.4)	
a_{13}	50	.0069	1.3080	.979	.0067	1.3143	.970	.0078	1.2672	.964
		(17.3)	(41.5)		(15.2)	(37.5)		(16.1)	(30.1)	
	100	.0014	1.1414	.989	.0016	1.1603	.980	.0020	1.1392	.979
		(10.8)	(30.9)		(9.78)	(27.0)		(11.6)	(23.4)	
	200	.0003	1.0607	.986	.0004	1.0802	.974	.0005	1.0698	.974
		(3.85)	(12.4)		(4.49)	(12.6)		(5.47)	(11.1)	
	25	.0003	1.6522	.935	.0001	1.5714	.937	.0003	1.3789	.881
		(1.05)	(38.4)		(.35)	(38.0)		(.76)	(20.3)	
a_{14}	50	−.0000	1.3146	.964	−.0001	1.2809	.962	−.0001	1.2072	.950
		(−.14)	(31.7)		(−1.60)	(30.2)		(−1.23)	(20.6)	
	100	−.0000	1.1386	.958	−.0000	1.1342	.956	−.0000	1.1134	.955
		(−.05)	(14.9)		(−.84)	(15.3)		(−.87)	(13.0)	
	200	−.0000	1.0524	.928	−.0001	1.0583	.924	−.0001	1.0499	.924
		(−1.24)	(4.59)		(−2.33)	(5.38)		(−2.32)	(4.64)	
	25	.0012	1.6204	.625	.0003	1.6843	.766	.0001	1.2864	.821
		(1.21)	(13.2)		(.34)	(19.6)		(.08)	(12.8)	
a_{21}	50	.0004	1.2746	.743	−.0002	1.3002	.798	−.0004	1.1366	.874
		(1.25)	(9.78)		(−.79)	(12.5)		(−1.52)	(8.63)	
	100	.0001	1.1513	.834	−.0000	1.1687	.850	−.0000	1.0873	.886
		(.86)	(7.88)		(−.21)	(9.56)		(−.28)	(6.23)	
	200	.0000	1.0702	.800	.0000	1.0714	.804	.0000	1.0301	.816
		(.49)	(3.51)		(.14)	(3.79)		(.16)	(1.73)	
	25	−.0016	1.8551	.908	−.0029	1.7335	.934	.0042	1.5148	.938
		(−1.33)	(37.3)		(−3.51)	(43.1)		(4.94)	(35.7)	
a_{22}	50	−.0007	1.4225	.941	−.0008	1.3649	.954	.0011	1.2749	.970
		(−1.59)	(30.4)		(−2.57)	(33.5)		(4.20)	(33.5)	
	100	−.0002	1.2065	.966	−.0003	1.1858	.969	.0002	1.1438	.967
		(−1.52)	(23.6)		(−2.29)	(24.5)		(1.87)	(22.4)	
	200	−.0002	1.0702	.960	−.0002	1.0852	.962	−.0001	1.0642	.964
		(−2.46)	(3.51)		(−2.78)	(11.1)		(−1.00)	(8.79)	
	25	−.0017	1.6437	.944	−.0002	1.5841	.948	−.0004	1.4305	.914
		(−1.27)	(41.6)		(−.26)	(42.6)		(−.29)	(26.6)	
a_{23}	50	−.0002	1.3153	.980	.0001	1.2966	.981	.0004	1.2478	.973
		(−.50)	(43.3)		(.44)	(45.1)		(1.03)	(33.1)	
	100	−.0000	1.1482	.990	−.0000	1.1420	.989	.0001	1.1235	.988
		(−.13)	(23.5)		(−.08)	(23.3)		(.63)	(20.5)	
	200	.0001	1.0646	.987	.0000	1.0627	.986	.0001	1.0542	.986
		(1.47)	(13.6)		(.70)	(13.6)		(1.92)	(12.1)	
	25	.0279	1.6396	.939	.0227	1.5821	.944	.0168	1.3583	.912
		(19.3)	(40.3)		(17.9)	(41.0)		(10.8)	(23.2)	
a_{24}	50	.0072	1.3120	.978	.0054	1.2776	.980	.0060	1.2230	.977
		(7.90)	(41.7)		(15.3)	(41.9)		(15.7)	(32.5)	
	100	.0016	1.1458	.989	.0013	1.1372	.989	.0016	1.1147	.988
		(11.6)	(31.4)		(10.4)	(31.5)		(12.4)	(26.4)	
	200	.0003	1.0671	.986	.0002	1.0652	.987	.0003	1.0555	.986
		(3.90)	(14.1)		(3.14)	(14.7)		(4.57)	(12.6)	

Table 7: OLS estimates of α and β in the model *Bias* = $\alpha + \beta \cdot$*Approx.* + ε, where the approximations to the bias are, from left to right, (9*b*), (10) and (11), all based on estimates. t ratios refer to tests of zero intercept and unit slope, respectively. Based on 784 stationary eigenvalue combinations, each simulated 10,000 times.

Table 7 holds the more interesting regression results for (9*b*), (10) and (11) based on estimates. Not surprisingly, the results are less impressive than in Table 6 because approximations tend to understate the bias more when based on estimates than when based on true parameters. Slope estimates clearly in excess of unity indicate that the problem of bias understatement is worse when estimates are used, as it was in the first-order case as well. The t ratios for T=200 are greater than 10 for most parameters, thus it would appear to take much longer series than that for the approximations to work well. Estimated intercepts are insignificant in many cases but are always significant for the second-order autoregressive estimates $\hat{a}_{13}$ and $\hat{a}_{24}$, as they are in Table 6 as well. Thus using estimates instead of true parameters does not affect intercepts. Squared sample correlations are also about the same as in Table 6, but generally somewhat higher for first-order estimates and slightly lower for second-order estimates. Though still very low, the correlations for $\hat{a}_{12}$ and $\hat{a}_{21}$ improve considerably when estimates are used.

4.3 Bias-Reduced Estimates

As a natural consequence of the relatively successful results in the previous section, it would appear possible to use the three approximations to reduce the bias of $\hat{\mathbf{A}}$. While such bias reduction ought to lie in the interest of every VAR model user, it is of course not obvious that the three approximations presented here should be used. In fact, to the extent that users have applied bias reduction, Bayesian methods have probably been employed (notably BVAR; cf. Section 2.1). Other possible means of reducing the bias include constrained estimation (imposing zero restrictions, for instance) and, as Section 3.4 indicated, estimating a model of too high an order and discarding parameter estimates of higher order may, in some cases at least, reduce bias as well. Ridge regression is another potential way to reduce bias, but one which has not been pursued and whose qualities are unknown.

Here however, the focus will be on the bias of $\hat{\mathbf{A}}$ before and after the approximate bias has been deducted. Since basing the approximations on true parameters has no meaning in terms of bias reduction, only approximations based on estimates will be used, and hence it is reasonable to expect some bias to remain even after the operation, bearing in mind the results presented in the previous section.

Starting with the first-order model (7), Figure 97 shows the remaining bias of $\hat{a}_1$ after (9*a*) (or (10), since they are equivalent here) based on estimates has been subtracted and Figure 98 the remaining bias after (11) based on estimates has been subtracted from the bias depicted in Figure 61 (since unit-root combinations have been included here, it is more appropriate to compare with Figure 61 than Figure 1).

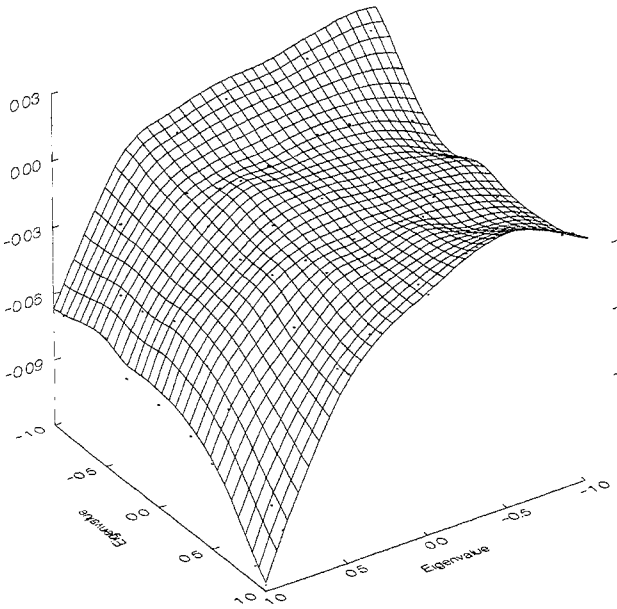

Figure 97: Bias of $\hat{a}_1$ less value of (9*a*); sample size 25

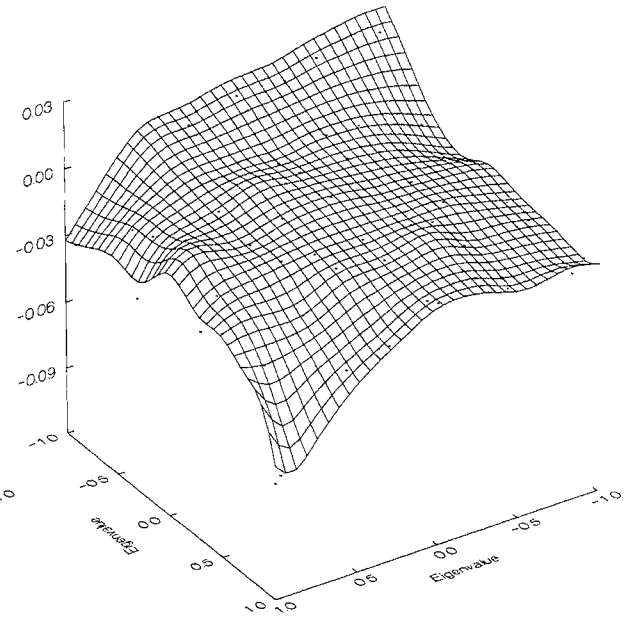

Figure 98: Bias of $\hat{a}_1$ less value of (11); sample size 25

Notice the change of scales here; whereas the bias of $\hat{a}_1$ in Figure 61 forms a plane approximately from –.2 to .1, the remaining bias in Figures 97 and 98 runs approximately from –.06 to .03, so bias reduction appears to work. Remaining bias is still mainly significant though; standard errors (now of course composed by the variance of $\hat{a}_1$, the variance of the approximation based on estimates, and the covariance between the two) range from .002 to .004 in this case, therefore only a small number of eigenvalue combinations in the interiors of Figures 97 and 98 have insignificant biases. (Later in this section (Figures 129 and 130) this issue will be addressed specifically.) For most eigenvalue combinations the ratio between bias and standard error is much smaller now than without bias reduction though (cf. Figure 49).

Unit-root combinations, and to some extent combinations involving eigenvalues of .9, apparently cause (9*a*) and (10) some trouble in Figure 97 above, the left wing of the figure being much more negative than the rest. It is obvious that as the right-hand eigenvalue approaches unity, it becomes increasingly difficult for (9*a*) and (10) to approximate the negative bias of $\hat{a}_1$ satisfactorily. Fortunately though, the second-order terms in (11) appear to handle that problem, almost completely adjusting

the left wing in Figure 98. The figure still bends though, indicating that a fuller second-order approximation than (11) or perhaps an approximation of even higher order is called for. It is nonetheless a great success that (11), which was developed precisely in order to better handle situations where first-order approximations would fail, has essentially the same effects as (9*a*) or (10) in regions where the first-order approximation performs well, but steps in to improve the performance in regions where it fails, in this case on the left.

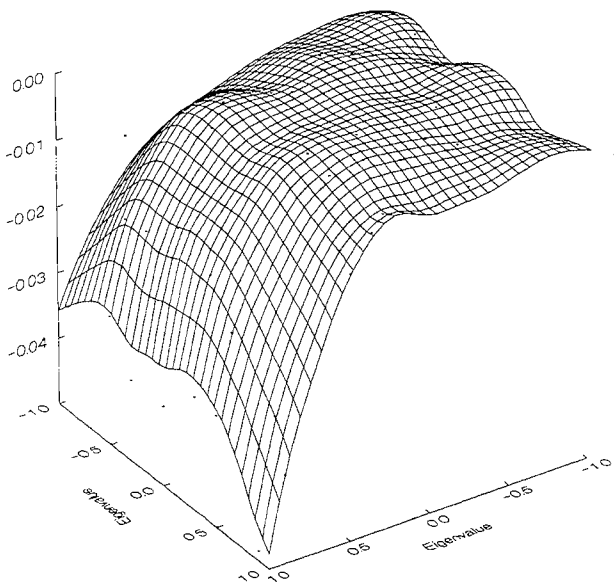

Figure 99: Bias of $\hat{a}_1$ less value of (9*a*); sample size 50

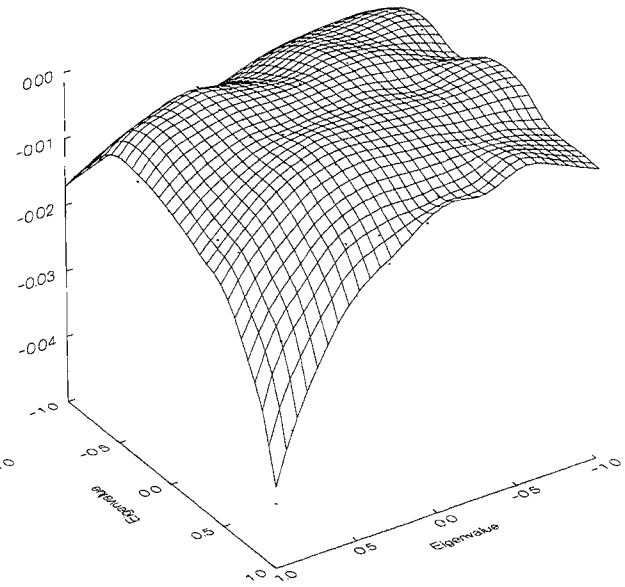

Figure 100: Bias of $\hat{a}_1$ less value of (11); sample size 50

As the sample size increases, the effects found in Figures 97 and 98 emerge even more clearly. For a sample of 50 observations, Figure 99 illustrates how the clearly positive bias of $\hat{a}_1$ on the right is reduced almost to zero by (9*a*) or (10), but also that the approximations are only partly successful in reducing the negative bias on the left, leaving a plane in Figure 99 which declines sharply from zero to –.04 as the right-hand eigenvalue approaches unity. This is greatly improved by (11) in Figure 100 though, except in the (1,1) corner (the integrated but not cointegrated case) where much of the bias still remains.

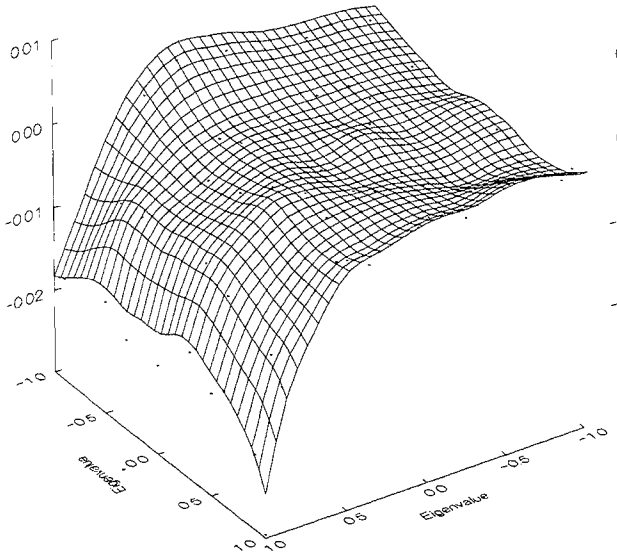

Fig. 101: Bias of $\hat{a}_1$ less value of (9*a*); sample size 100

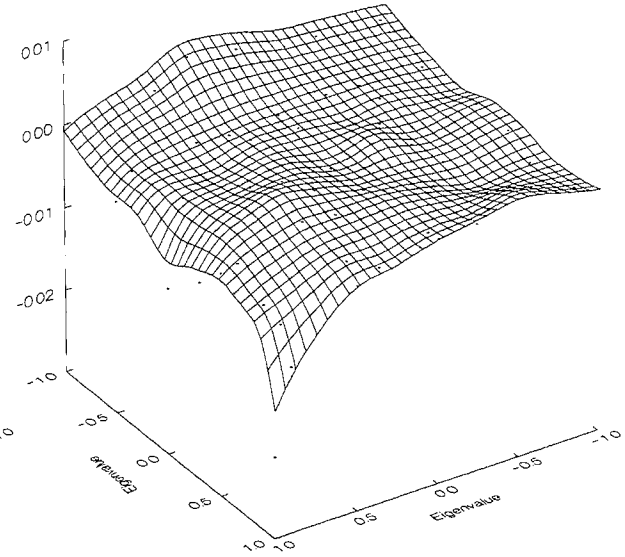

Fig. 102: Bias of $\hat{a}_1$ less value of (11); sample size 100

As the sample size increases to 100 (Figures 101 and 102) and 200 (Figures 103 and 104) the same effect can again be found: The positive bias on the right is approximated almost perfectly by the first-order approximation but the negative bias on the left is approximated increasingly worse as the right-hand eigenvalue approaches unity, causing the left-hand sections of the surfaces to bend down sharply. Using the second-order approximation (11) improves the situation in that region but still leaves considerable (and significant, as it will turn out) negative bias, in particular as both eigenvalues are close to or at unity (the case of first-order integration but no cointegration). Elsewhere, however, the second-order terms of (11) have little effect, as demonstrated by the ridge of remaining positive (and significant) bias in the back of Figure 103 below, which returns virtually unaffected in Figure 104.

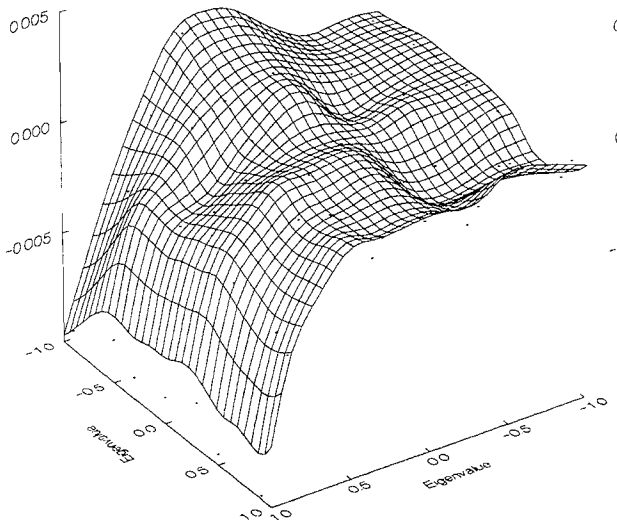

Fig. 103: Bias of $\hat{a}_1$ less value of (9*a*); sample size 200

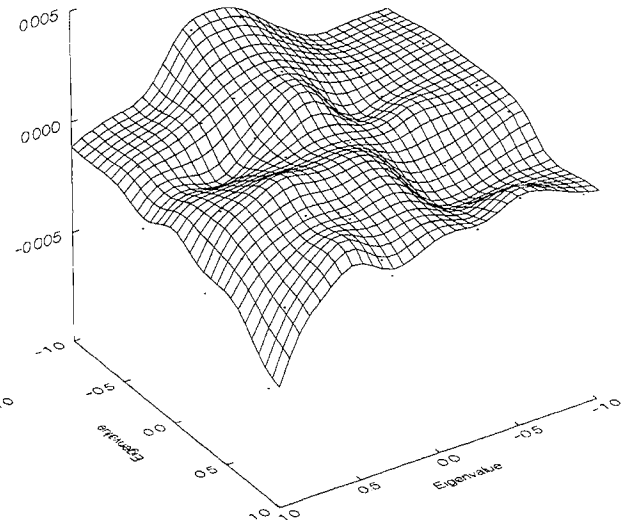

Fig. 104: Bias of $\hat{a}_1$ less value of (11); sample size 200

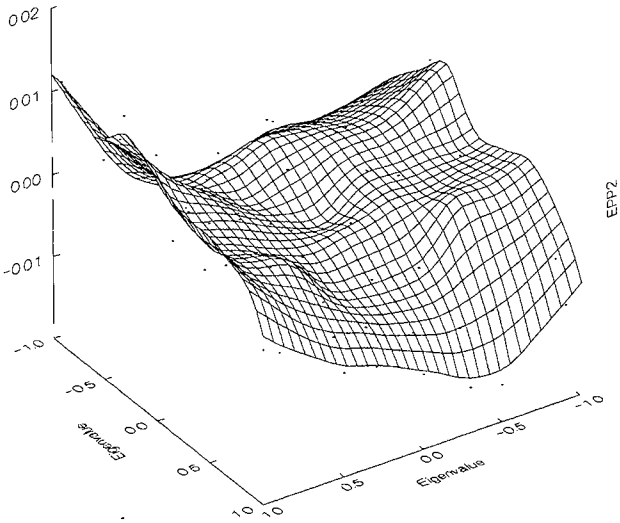

Figure 105: Bias of $\hat{a}_2$ less value of (9*a*); sample size 25

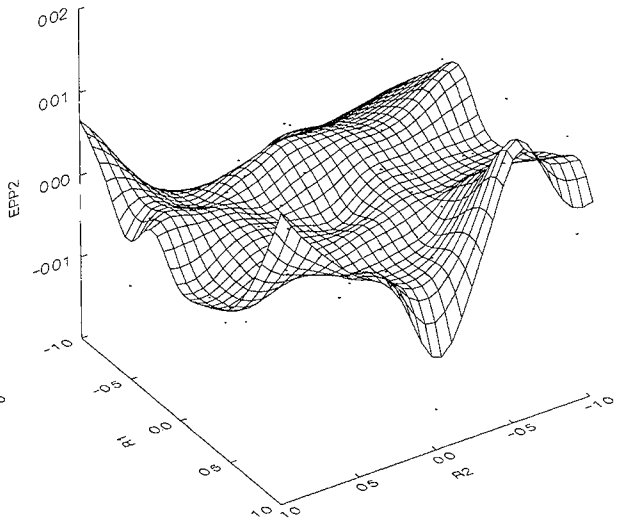

Figure 106: Bias of $\hat{a}_2$ less value of (11); sample size 25

A completely different picture obtains as the biases of the cross-term estimates $\hat{a}_2$ and $\hat{a}_3$ are approximated. Recall that the biases of $\hat{a}_2$ and of $\hat{a}_3$ have been found to form more or less linear functions of both eigenvalues rather than of only one, as in the case of $\hat{a}_1$ and $\hat{a}_4$. The approximations appear to be less able of handling the bias here than they proved to be for $\hat{a}_1$; in particular (11) does not prove as effective here as it did before. Figure 105 depicts the remaining bias from Figure 63 after the first-order approximation has been applied to reduce the bias. Most of the bias is gone, but some negative bias remains as the left-hand eigenvalue approaches unity and some positive bias as the right-hand eigenvalue tends to unity. The second-order approximation works in the right direction in both these areas, but to such an extent that in Figure 106 there are in fact two (significantly) positive bias peaks where negative bias prevailed in Figure 105. On the left as well, whereas there is significantly positive remaining bias in Figure 105, in Figure 106 the bias is slightly negative, so (11) appears to overadjust as eigenvalues approach unity.

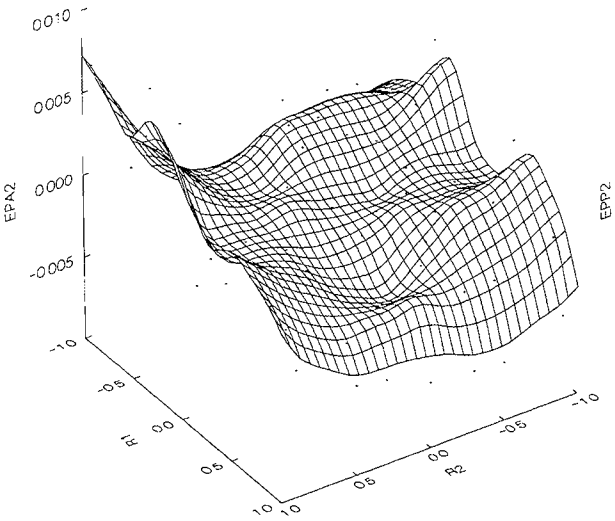

Figure 107: Bias of $\hat{a}_2$ less value of (9*a*); sample size 50

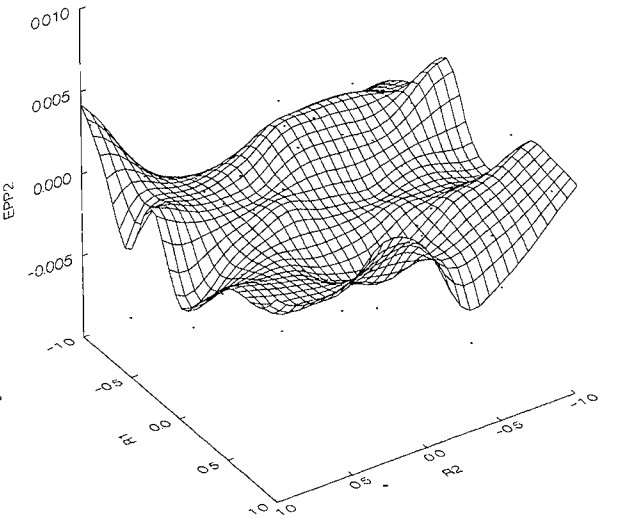

Figure 108: Bias of $\hat{a}_2$ less value of (11); sample size 50

Figures 107 and 108 illustrate the corresponding remaining biases when the sample size is 50. Using the first-order approximation to reduce the bias of $\hat{a}_2$ leaves the bias shown in Figure 107, which is similar to that in Figure 105, with negative remaining bias at the front and positive bias on the left. Again the second-order approximation appears to overadjust along the edges, leaving partly positive front bias and partly negative left-hand bias. The second-order terms appear to have little or no effect elsewhere.

However, in samples of 100 (Figures 109 and 110) or 200 (Figures 111 and 112), using (11) does not appear to cause overadjustment. The same sort of remaining bias shapes as before can be found in Figures 109 and 111 as well, significant biases occurring mainly along the front and left-hand edges. Applying (11) instead of (9*a*) or (10) reduces most of them to smaller (but in some cases still significant) levels in Figures 110 and 112, where apparently (11) works in the right direction without overadjusting the bias.

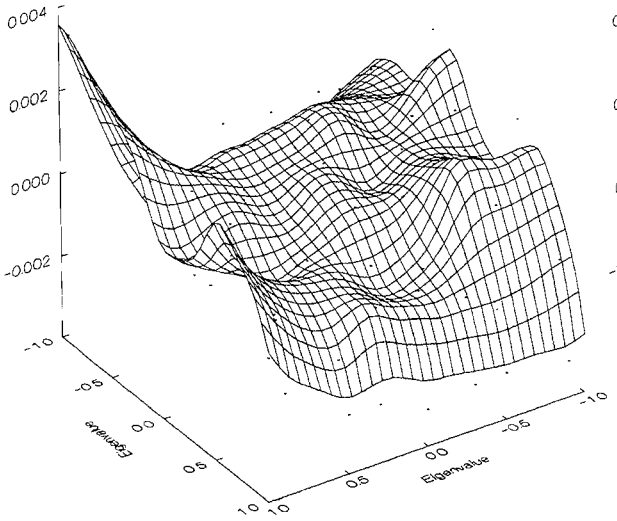

Fig. 109: Bias of $\hat{a}_2$ less value of (9*a*); sample size 100

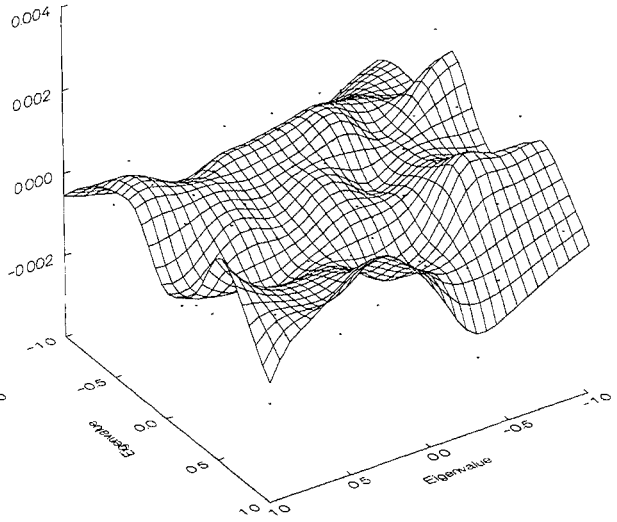

Fig. 110: Bias of $\hat{a}_2$ less value of (11); sample size 100

Turning to $\hat{a}_3$, the other cross-term estimate of (7), similar results are to be expected since in this particular setting a_3 will always be the same function of a_2 ($a_3 = -5a_2$). However, as Figures 113 through 120 indicate, the remaining bias of $\hat{a}_3$ is more regular and in a sense more systematic than the remaining bias of $\hat{a}_2$ presented above, which ought to improve the bias-reducing performance of (11) here.

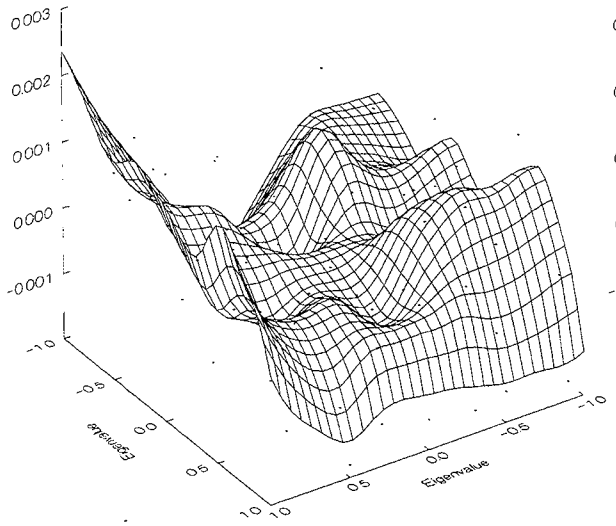

Fig. 111: Bias of $\hat{a}_2$ less value of (9*a*); sample size 200

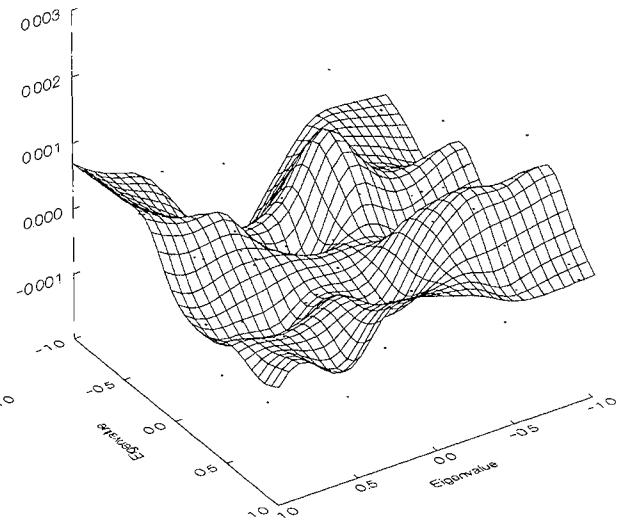

Fig. 112: Bias of $\hat{a}_2$ less value of (11); sample size 200

For the smallest sample size (T=25), Figure 113 below reveals that using a first-order bias approximation to reduce the bias of $\hat{a}_3$ (depicted in Figure 65) leaves positive biases when the left-hand eigenvalue approaches (positive or negative) unity, and negative biases when the right-hand eigenvalue approaches unity, together forming a hyperbolic parabola in three dimensions. The second-order approximation (Figure 114) has very little effect on the rear positive bias and the right-hand slightly negative bias, but very dramatic effects on the left-hand negative bias, which it reduces almost entirely, and on the positive bias in the foreground, where again (11) overadjusts, making it significantly negative in Figure 114. Nevertheless, in this case the second-order approximation must be said to work as intended, reducing the remaining bias in regions where one or both eigenvalues approach unity while virtually not affecting the remaining bias at all elsewhere.

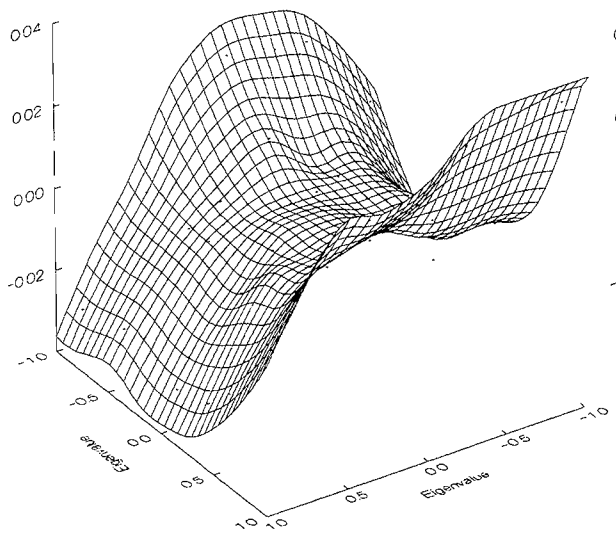

Figure 113: Bias of $\hat{a}_3$ less value of (9*a*); sample size 25

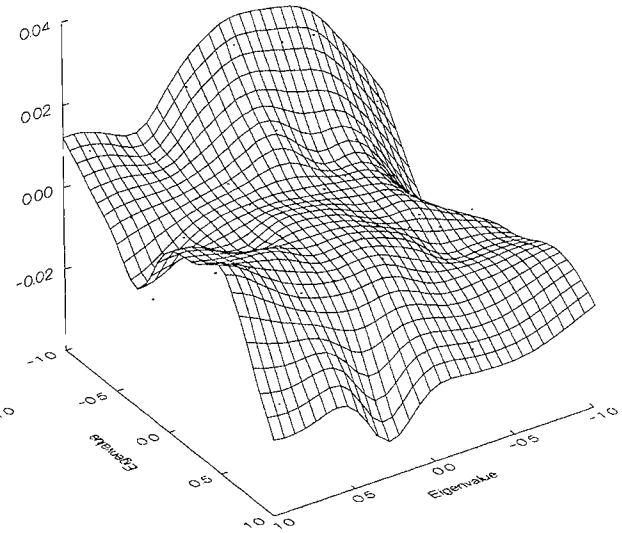

Figure 114: Bias of $\hat{a}_3$ less value of (11); sample size 25

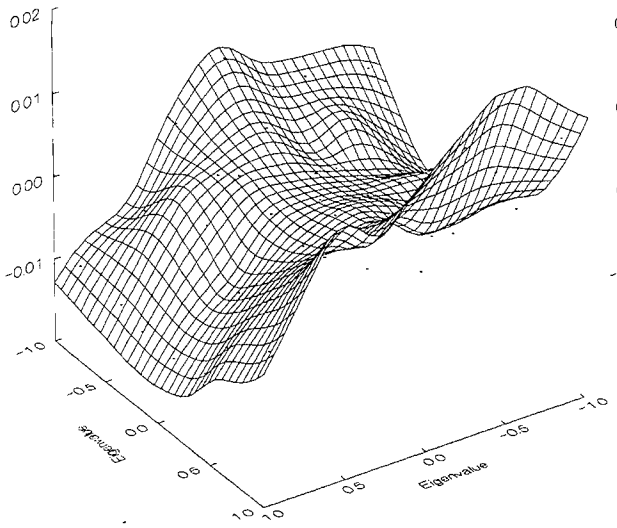

Figure 115: Bias of $\hat{a}_3$ less value of (9*a*); sample size 50

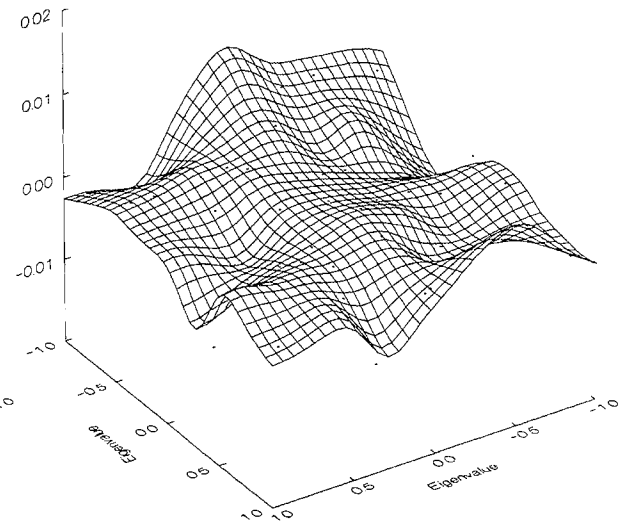

Figure 116: Bias of $\hat{a}_3$ less value of (11); sample size 50

Basically the same hyperbolic bias shape as before can be found in Figure 115 for T=50, but this time the second-order approximation does not overadjust as it did for 25 observations. As Figure 116 demonstrates, (11) now reduces the positive bias in front to more or less the same levels as the negative bias on the left. But again it has little or no effect on the rear positive and the right-hand negative biases.

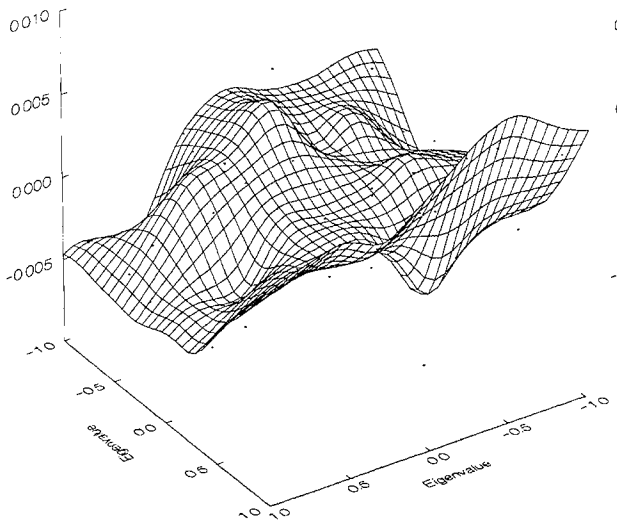

Fig. 117: Bias of $\hat{a}_3$ less value of (9*a*); sample size 100

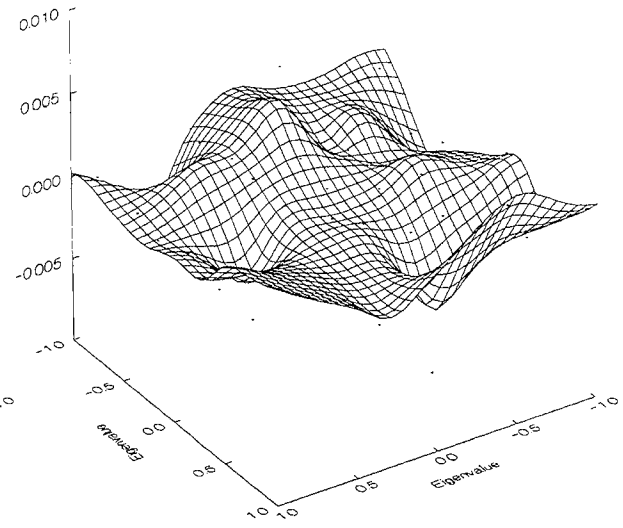

Fig. 118: Bias of $\hat{a}_3$ less value of (11); sample size 100

For T=100 the remaining bias in Figure 117 has a less regular shape than for the two smaller sample sizes, but regions of negative bias on the left and of positive bias in the back and in the foreground can still be found. Once again (11) handles the left-hand negative bias very well, but apparently the second-order bias reduction is too small in the region of positive bias closest to the reader, since there is still some positive bias left in Figure 118.

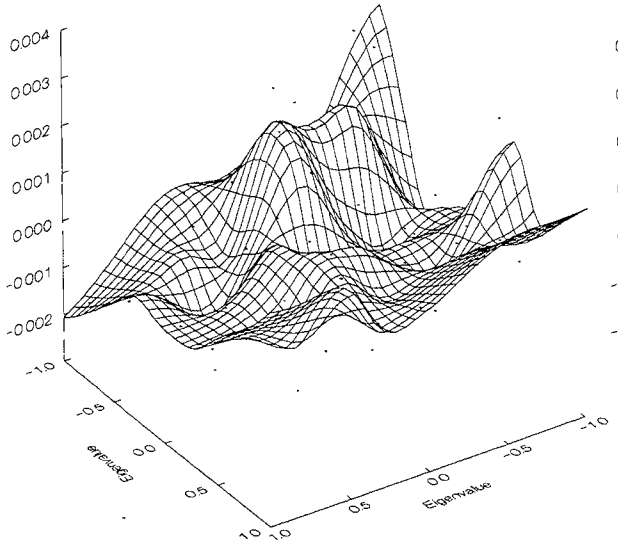

Fig. 119: Bias of $\hat{a}_3$ less value of (9*a*); sample size 200

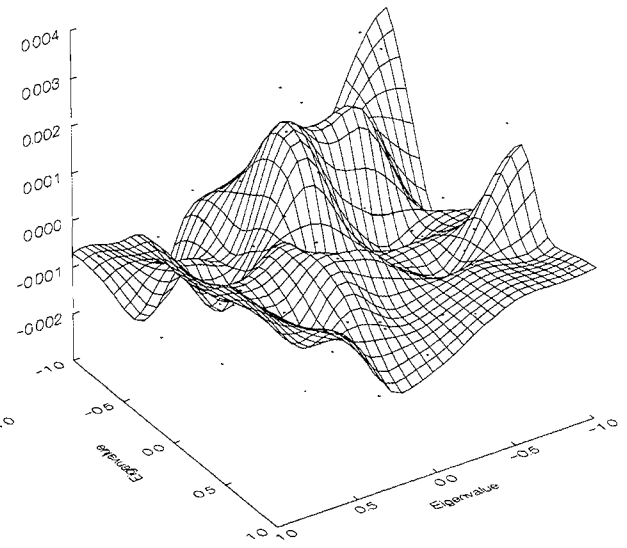

Fig. 120: Bias of $\hat{a}_3$ less value of (11); sample size 200

The results of using a first-order approximation to reduce the bias of $\hat{a}_3$ (depicted in Figure 66) when T is 200 can be found in Figure 119. It has a highly irregular shape with unexpected positive peaks on the right and in the (–.9,–.9) corner, but as before there is still a negative bias on the left and regions of positive bias in the background as well as in the foreground. The second-order approximation reduces the left-hand negative bias and the positive bias in front, but is of course completely unable to handle the two positive peaks and the rear positive bias.

Finally, the remaining bias of $\hat{a}_4$ is very similar to that of $\hat{a}_1$ except for the interchanged roles of the two eigenvalues. For T=25, reducing the bias of $\hat{a}_4$ (depicted in Figure 67) by a first-order approximation leaves the mainly negative bias found in Figure 121.

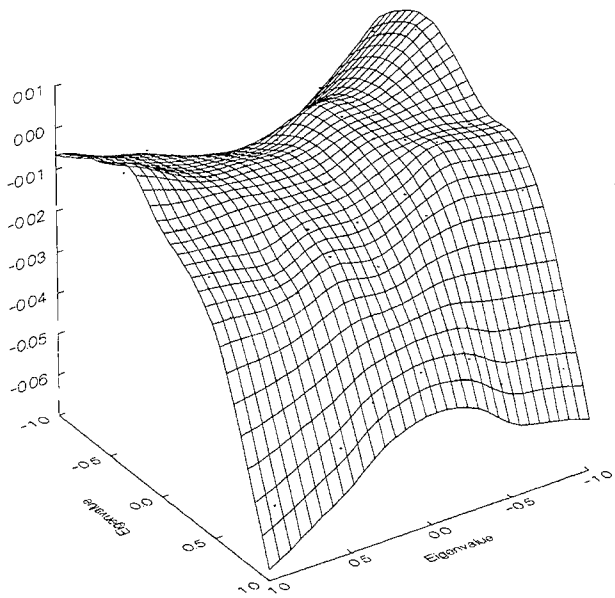

Figure 121: Bias of $\hat{a}_4$ less value of (9*a*); sample size 25

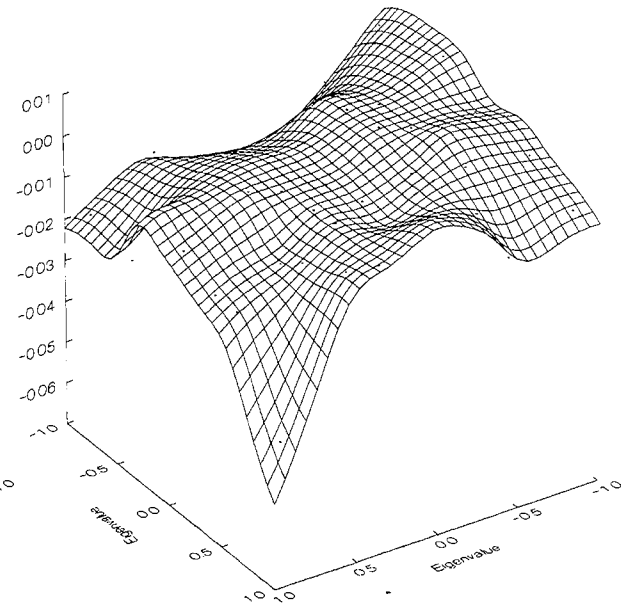

Figure 122: Bias of $\hat{a}_4$ less value of (11); sample size 25

It is positive only around the (–.9,–.9) corner and, parallel to the case of $\hat{a}_1$ in Figure 97, increasingly negative as the left-hand eigenvalue increases. Thus the potential for further bias reduction in that region is great, and as Figure 122 shows, using the second-order approximation indeed reduces the negative bias considerably except for in and around the (1,1) corner (the first-order integrated but not cointegrated case), which again is parallel to the case of $\hat{a}_1$. It should also be noted that (11) actually adds to the negative bias as the right-hand eigenvalue approaches unity in Figure 122, but as usual the second-order terms have negligible effects elsewhere.

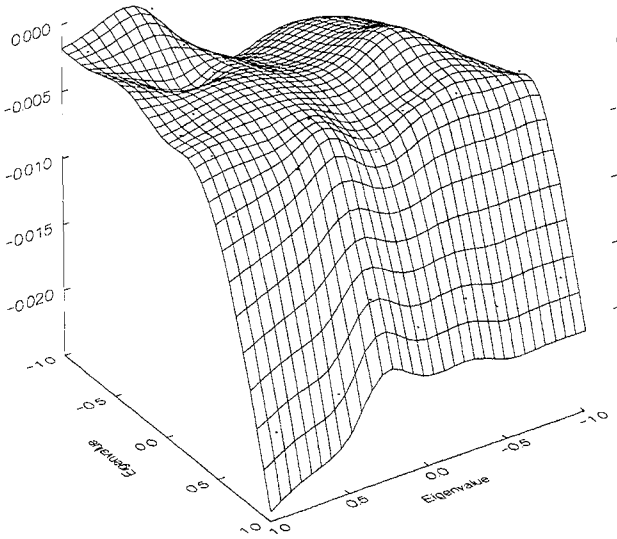

Figure 123: Bias of $\hat{a}_4$ less value of (9*a*); sample size 50

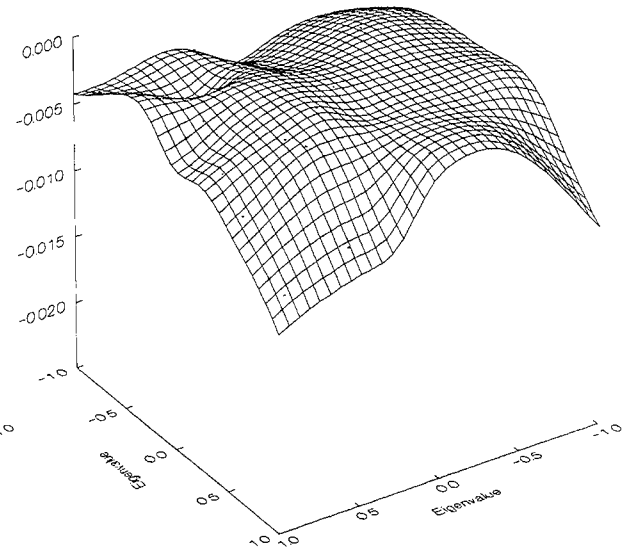

Figure 124: Bias of $\hat{a}_4$ less value of (11); sample size 50

As the sample size increases to 50 (Figures 123 and 124) and 100 (Figures 125 and 126), using the expanded bias approximation to reduce the bias finds even stronger support in the graphs. For instance, whereas Figure 123 indicates that (9*a*) and (10) reduce the positive bias well but leave considerable negative bias as the left-hand eigenvalue increases, Figure 124 shows that the second-order terms of (11) are highly capable of handling that negative bias, again except in the (1,1) corner (and to some extent around the right-hand corner). Here again (11) appears to add to the negative bias of $\hat{a}_4$ as the right-hand eigenvalue approaches unity but leaves the interior of Figure 123 largely unaffected.

In Figure 125, even though its interior is less regular than for 25 or 50 observations, the same pattern of sharply increasing negative bias can be found as the left-hand eigenvalue increases. And once again, the second-order terms of (11) handle this negative bias well, reducing it almost completely except in and around the (1,1) corner in Figure 126.

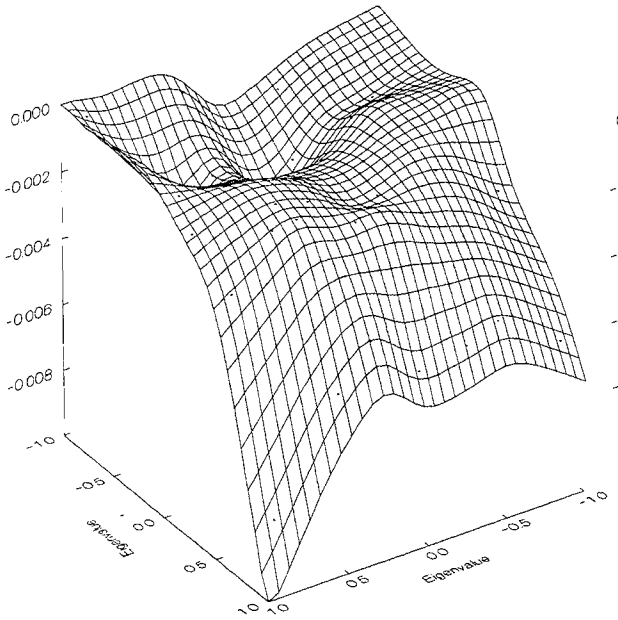

Fig. 125: Bias of $\hat{a}_4$ less value of (9*a*); sample size 100

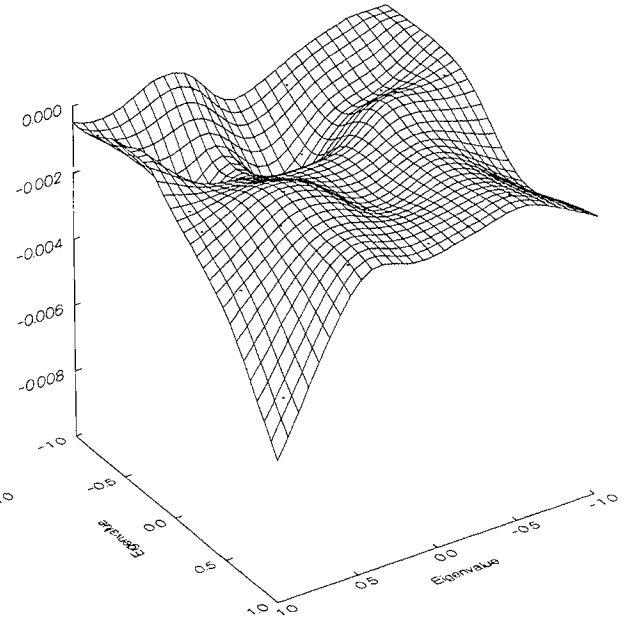

Fig. 126: Bias of $\hat{a}_4$ less value of (11); sample size 100

Finally, the remaining bias of $\hat{a}_4$ when then sample size is 200 is depicted in Figures 127 and 128 (cf. Figure 68 for the original bias). It is obviously much less regular than for the smaller sample sizes and holds a number of bias peaks which (11) is as unable to reduce as the first-order approximations are. As expected though, the second-order approximation manages to reduce the front negative bias in Figure 127, and actually does a decent job in the (1,1) corner as well.

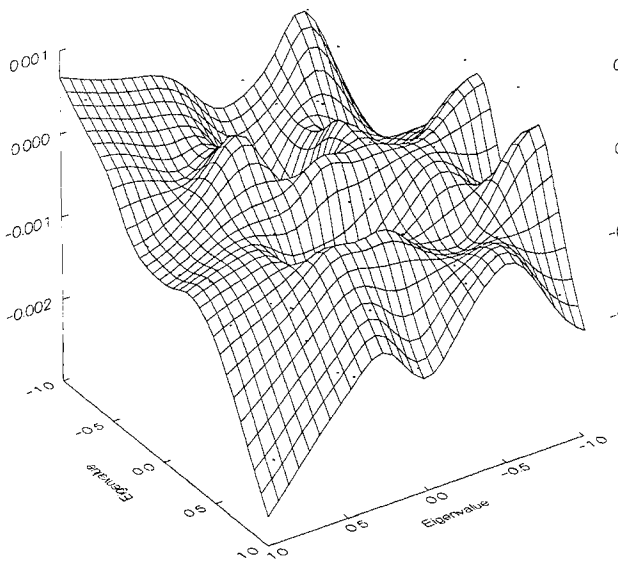

Fig. 127: Bias of $\hat{a}_4$ less value of (9*a*); sample size 200

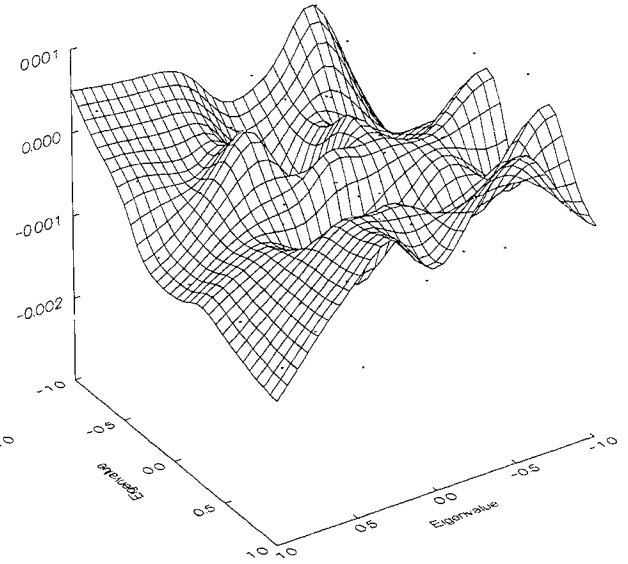

Fig. 128: Bias of $\hat{a}_4$ less value of (11); sample size 200

Before turning to the results for two of the estimates in the second-order model (8), some results concerning the significance of the remaining bias displayed in Figures 97 through 128 are due. As already indicated in connection with the bias plots, using (9*a*), (10) or (11) to reduce the bias normally does not leave unbiased estimates in the cases investigated here; significantly biased estimates are still to be expected, but in most cases the ratios between remaining bias and its standard error will be much smaller than in the corresponding Figures 49 to 56. Computing the standard errors is more complicated than in Section 3.2 though; since the definition of a bias-reduced estimate is the original estimate less a bias approximation based on that estimate, the standard error is no longer a function merely of the sample variance of the 10,000 estimates, but also of the variance of the 10,000 generated bias approximations and of the covariances between estimates and approximations based on estimates. Since the generated approximations based on estimates are less variable than their parameter estimates, the added variance will be of a smaller order of magnitude than the first variance. Estimated covariances can be substantial though, and since the approximations are often positively correlated with the parameter estimates, the resulting standard errors will sometimes be smaller than in Section 3.2. When the estimated correlation is negative though, as is often the case if eigenvalues are close to positive or negative unity, standard errors will of course increase.

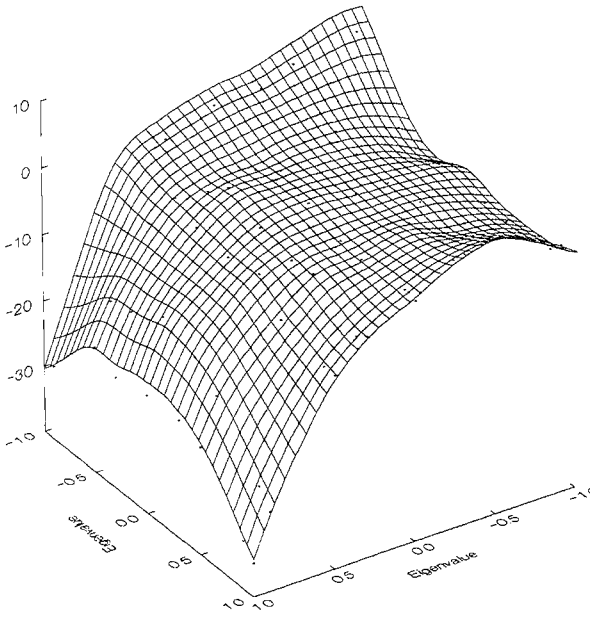

Figure 129: Relative of bias $\hat{a}_1$ less value of (9*a*); T=25

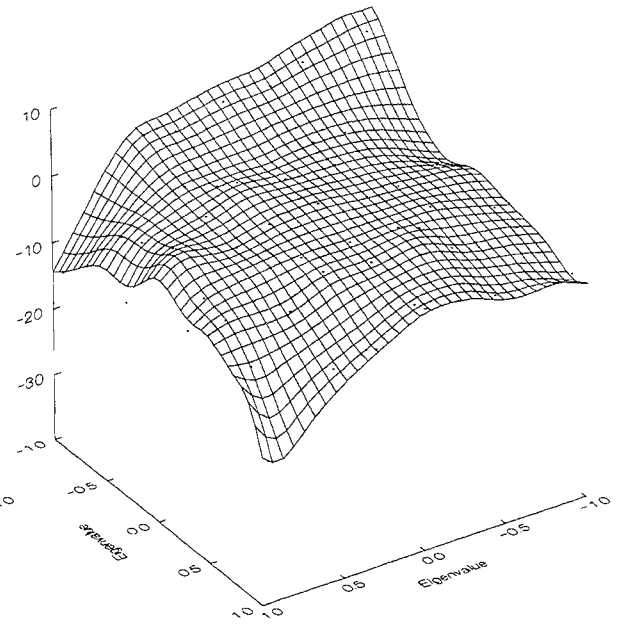

Figure 130: Relative bias of $\hat{a}_1$ less value of (11); T=25

The relative biases of the bias-reduced estimates of a_1 are graphed in Figures 129 and 130, Figure 129 showing the results of using a first-order approximation to reduce the bias (cf. Fig 97) and Figure 130 those of the second-order approximation (cf. Fig 98). Since the sample size is 25 here, both figures may be

compared with Figure 49, which holds the relative bias of $\hat{a}_1$ (but whereas Figure 49 is based only on stationary eigenvalue combinations, Figures 129 and 130 include unit-root cases as well). As the figures indicate, the bias-reduced estimate can still be expected to be significantly biased, but the relative biases are much smaller (between –30 to 8) than without bias reduction (recall that the corresponding ratios in Figure 49 ranged approximately from –85 to 40). Only for a small number of interior eigenvalue combinations are the bias-reduced estimates insignificantly biased (at the 5 per cent level, absolute relative bias not in excess of 1.96), but even if for most combinations a significant bias remains, it can be expected to be much smaller than without bias reduction. It should also be pointed out that the differences between Figure 129 and Figure 130 essentially reflect different remaining biases (cf. Figures 97 and 98). It is true that the denominators, i.e. the standard errors, will also be different since the second variance and the covariance will differ, but this effect is much smaller than the different numerators.

Like in Section 3.2, relative bias graphs are given for two sample sizes only, 25 and 200 observations. Figures 131 and 132 show the relative biases of the bias-reduced estimate of a_1 based on 200 observations, and therefore they relate to Figure 50. As before, the ranges of relative biases are much smaller than without bias reduction (–20 to 3 instead of –40 to 25), but more importantly, the majority of bias-reduced estimates turn out not to be 5 per cent significantly biased. Only along the edges of Figures 131 and 132 is significant bias still to be expected.

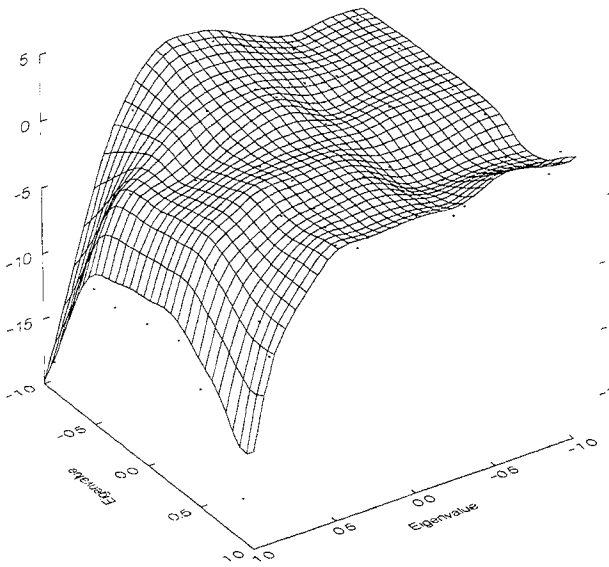

Figure 131: Relative bias of $\hat{a}_1$ less value of (9*a*); T=200

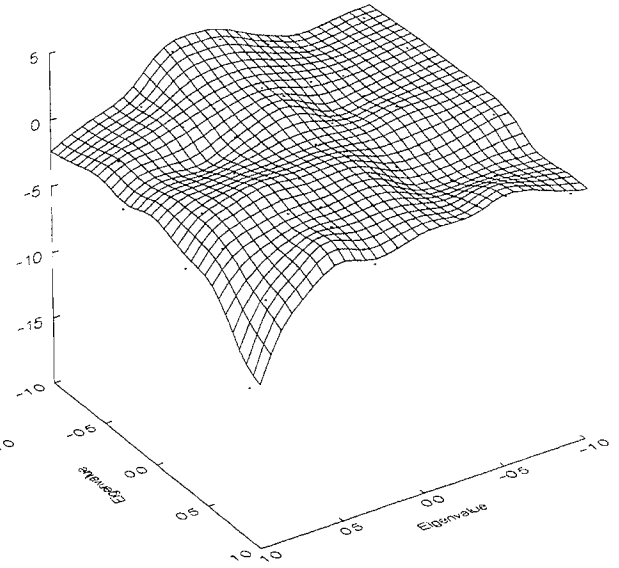

Figure 132: Relative bias of $\hat{a}_1$ less value of (11); T=200

Turning to the bias-reduced estimates of a_2, Figures 133 and 134 show their relative bias based on a sample of 25 observations. Without bias correction, the relative bias ranges from –25 to 35 and is insignificant only by exception (cf. Figure

51) but after bias reduction the range of relative biases is reduced approximately to –8 to 8, most of them being smaller than 1.96 in absolute terms. Like in the case of $\hat{a}_1$, adding second-order terms tends to reduce the relative bias even more.

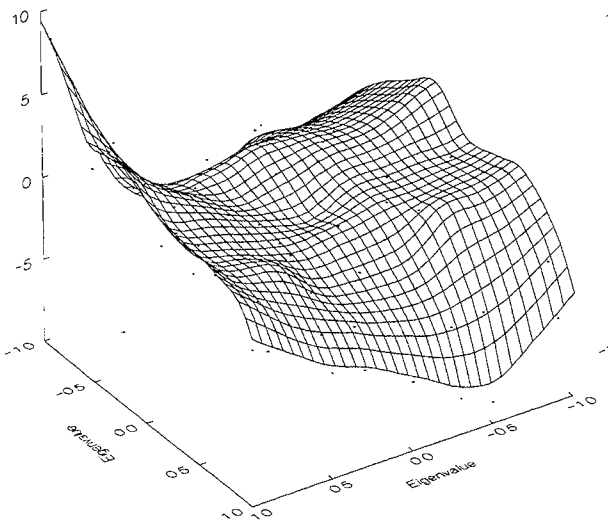

Figure 133: Relative bias of $\hat{a}_2$ less value of (9*a*); T=25

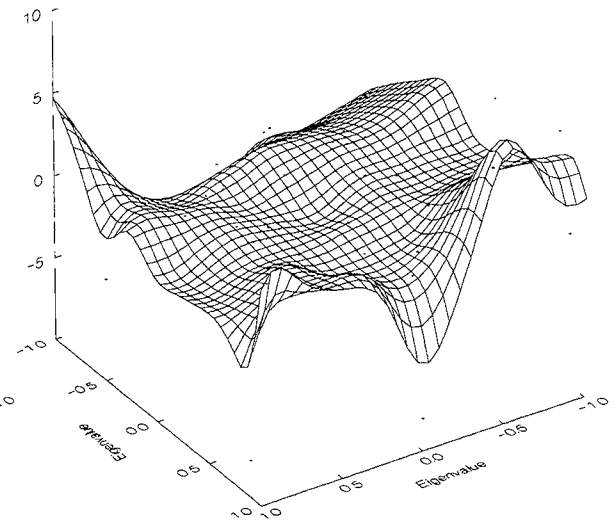

Figure 134: Relative bias of $\hat{a}_2$ less value of (11); T=25

Increasing the sample size to 200 of course reduces the relative bias in Figures 135 and 136, and more importantly, using the second-order approximation makes the estimate of a_2 insignificantly (at the 5 per cent level) biased in Figure 136 for most of the edge combinations for which a significant bias remains in Figure 135.

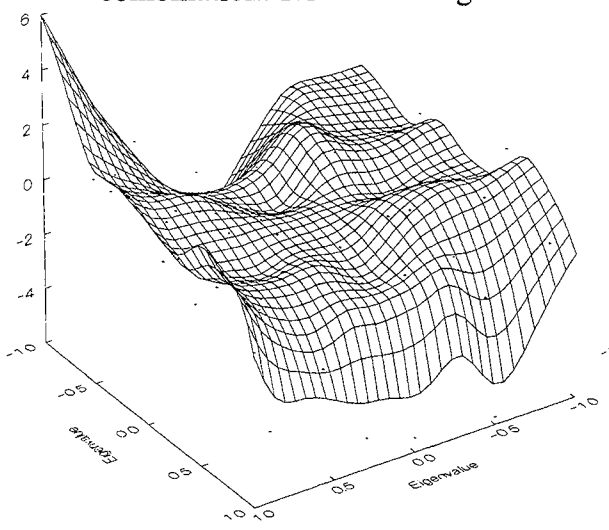

Figure 135: Relative bias of $\hat{a}_2$ less value of (9*a*); T=200

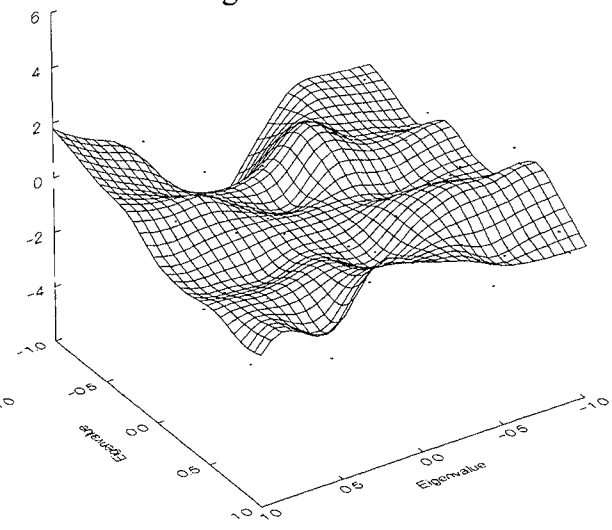

Figure 136: Relative bias of $\hat{a}_2$ less value of (11); T=200

As demonstrated earlier in this section, even though the 5-to-1 relation between a_3 and a_2 imposed by the experiment design can be found in their biases, after bias reduction it is more difficult to distinguish. As Figures 137 and 138 show, more regularly shaped relative bias plots than for $\hat{a}_2$ obtain, but with more significant biases. Nonetheless, the relative biases, ranging between –15 and 15, are of course smaller than without bias reduction (Figure 53).

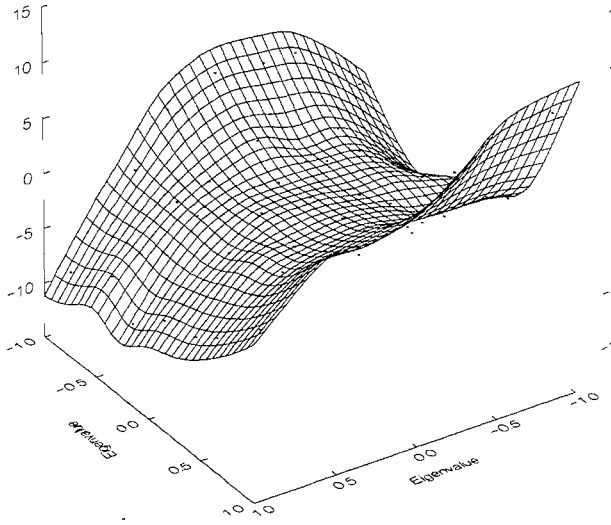

Figure 137: Relative bias of $\hat{a}_3$ less value of (9*a*); *T*=25

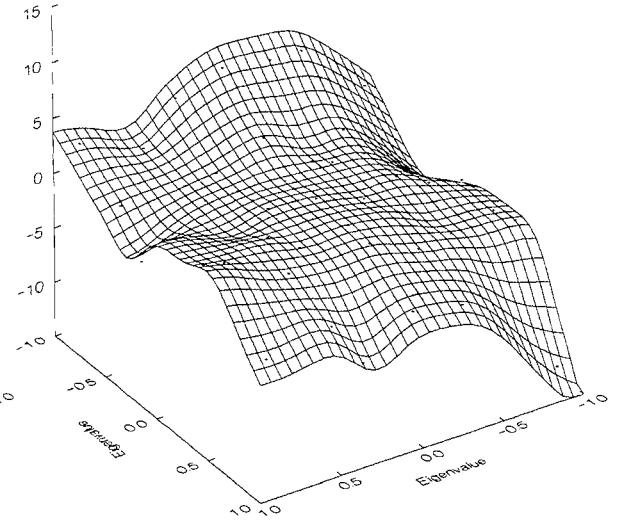

Figure 138: Relative bias of $\hat{a}_3$ less value of (11); *T*=25

As Figures 113 and 114 have demonstrated, (11) grossly overadjusts for bias in the foreground region, making the relative bias in Figure 138 more or less as large as in Figure 137, but negative instead of positive. On the other hand, the second-order terms appear to work satisfactorily along the left-hand edge, reducing the negative bias in Figure 137 to mostly insignificant levels in Figure 138.

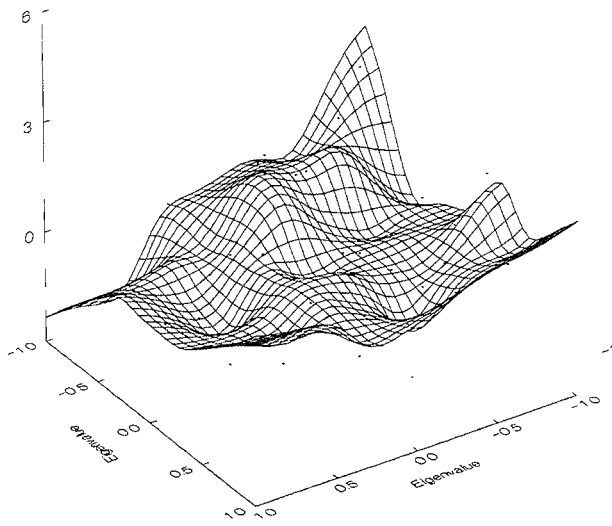

Figure 139: Relative bias of $\hat{a}_3$ less value of (9*a*); *T*=200

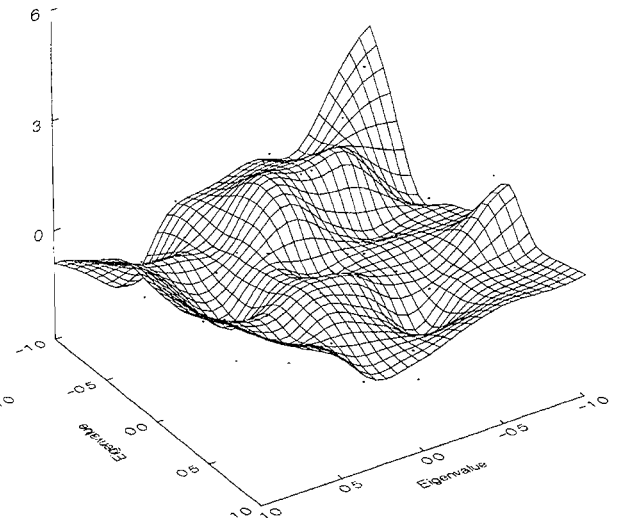

Figure 140: Relative bias of $\hat{a}_3$ less value of (11); *T*=200

Increasing the sample size to 200 leaves less regular relative bias shapes in Figures 139 and 140, with the same two distinguishing peaks as in Figures 119 and 120. For the majority of eigenvalue combinations the remaining bias appears to be insignificant though, and using (11) to reduce the bias of $\hat{a}_3$ dramatically improves the results along the front and left-hand edges. Its performance around the opposite (1,–.9) and (–.9,1) corners is particularly impressive; in both cases extreme biases in Figure 139 are reduced to more normal and insignificant levels in Figure 140.

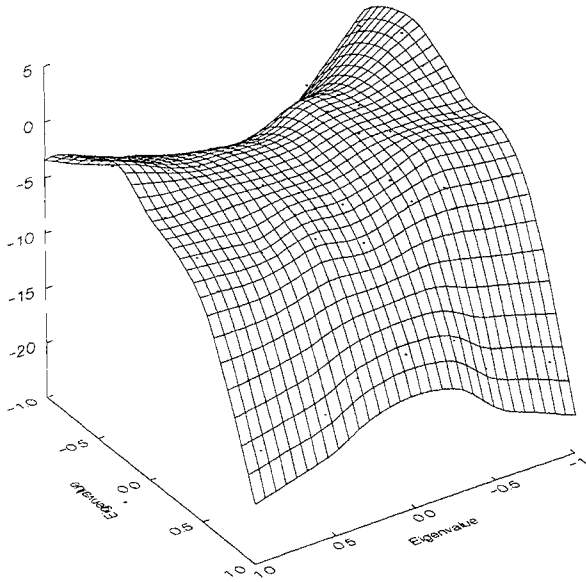

Figure 141: Relative bias of $\hat{a}_4$ less value of (9*a*); T=25

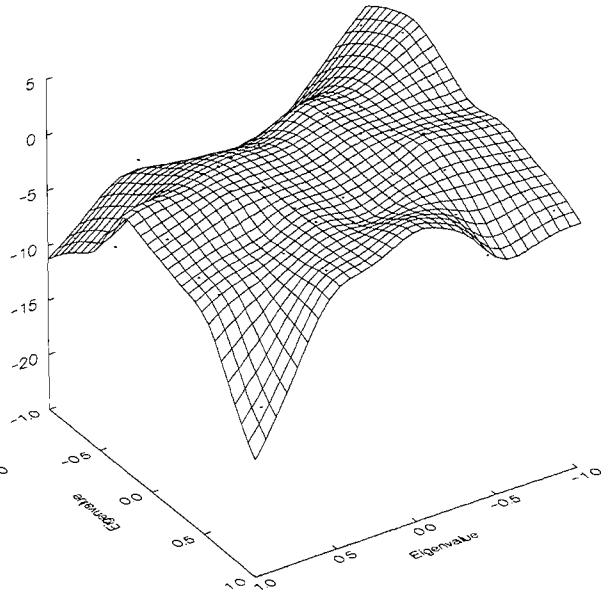

Figure 142: Relative bias of $\hat{a}_4$ less value of (11); T=25

Finally, the predominantly negative remaining bias of $\hat{a}_4$ when the sample size is 25 (cf. Figure 121) turns out in Figure 141 to be highly significant, with bias-to-standard error ratios close to –20 as the left-hand eigenvalue approaches unity. Whereas adding second-order terms improves the results considerably along the front, it actually makes relative biases even more negative along the left-hand edge, and has little effect elsewhere. It should also be noted that adding second-order terms has virtually no effect at all in the (1,1) corner, where a highly significant negative bias still prevails.

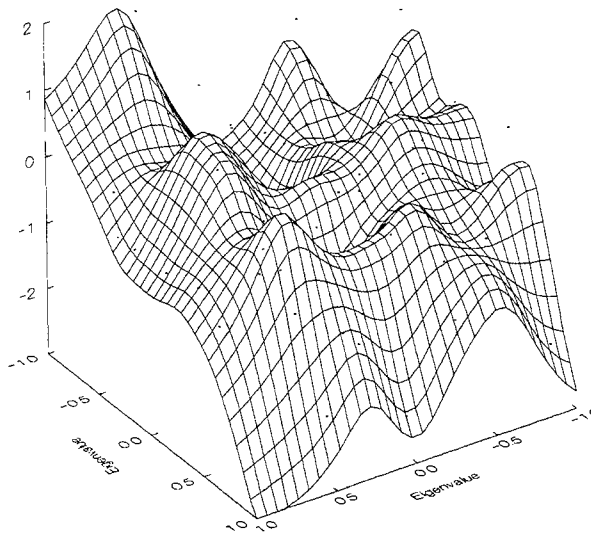

Figure 143: Relative bias of $\hat{a}_4$ less value of (9*a*); T=200

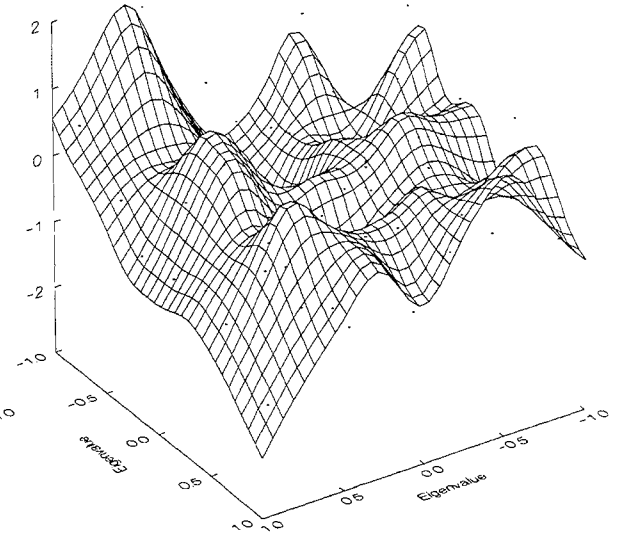

Figure 144: Relative bias of $\hat{a}_4$ less value of (11); T=200

Though much less regular, the bias plots for sample size 200 look far more encouraging. As Figure 143 indicates, reducing the bias of $\hat{a}_4$ by a first-order bias

approximation leaves 5 per cent significant bias only at a rear bias peak and for some eigenvalue combinations at the front. The second-order approximation handles most of them well though, including the (1,1) combination, but is incapable of reducing the positive bias peak in the rear.

Let us now focus on the corresponding results for $\hat{a}_{11}$ and $\hat{a}_{14}$ of the second-order model (8). As before, those two estimates are taken to represent all eight in order for the display of results to be relatively compact. Since the autoregressive order p is now 2, Tjøstheim-Paulsen's and Nicholls-Pope's approximations will no longer be equivalent, therefore using (9*b*) or (10) to reduce the bias of $\hat{\mathbf{A}}$ in (8) will yield different results, and so for each sample size below there will be three bias plots; one for (9*b*) based on estimates, one for (10) based on estimates, and one for (11) based on estimates.

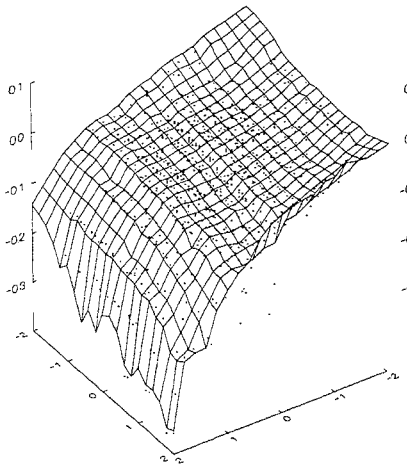

Figure 145: Bias of $\hat{a}_{11}$ less (9*b*)

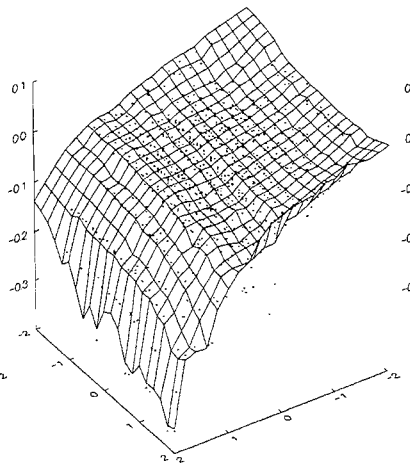

Figure 146: Bias of $\hat{a}_{11}$ less (10)

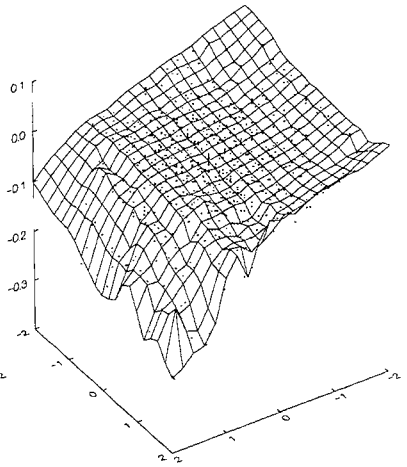

Figure 147: Bias of $\hat{a}_{11}$ less (11)

Based on estimates; sample size 25. Scales represent eigenvalue sums $\rho_{11}+\rho_{12}$ and $\rho_{13}+\rho_{14}$

Figures 145, 146 and 147 depict the remaining bias of $\hat{a}_{11}$ to be expected after bias reduction using (9*b*), (10) or (11), respectively, all based on estimates. Comparing these bias plots with Figure 69 (since unit-root combinations are included here, Figure 69 provides a better illustration than Figure 33 of the bias before bias reduction) verifies that all three approximations perform well in the sense that they reduce the bias of $\hat{a}_{11}$ in a sample of 25 observations. But their performance is less impressive than in the case of (7) because they significantly understate the bias of $\hat{a}_{11}$ (as demonstrated in Section 4.2), leaving considerable (and significant) unadjusted bias. This is particularly true in and around the (2,2) corner (where all four eigenvalues are at unity, so that the two series are $I(2)$ but not cointegrated) and along the $CI(2,2)$ edges, where on the left a negative bias remains, on the right a positive bias. Whereas (9*b*) and (10) perform almost identically, (11) handles second-order (co-)integration

better, reducing the left-hand negative as well as the right-hand positive bias. At the same time it is clear that much bias remains even after the second-order approximation has been applied, and this remaining bias will turn out to be significant in Figure 170.

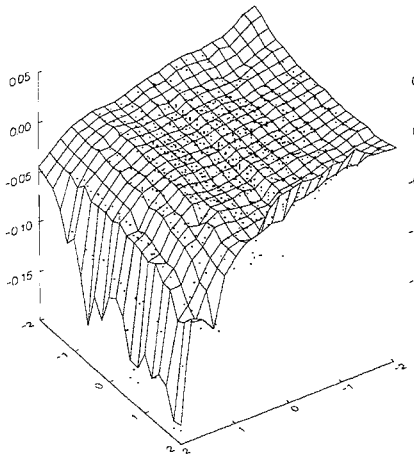

Figure 148: Bias of $\hat{a}_{11}$ less (9*b*)

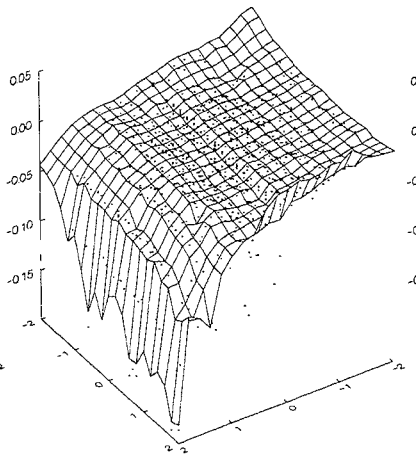

Figure 149: Bias of $\hat{a}_{11}$ less (10)

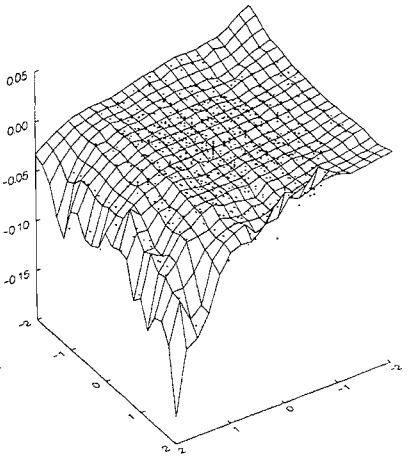

Figure 150: Bias of $\hat{a}_{11}$ less (11)

Based on estimates; sample size 50. Scales represent eigenvalue sums $\rho_{11}+\rho_{12}$ and $\rho_{13}+\rho_{14}$

The same results obtain even more clearly as the sample size increases to 50, 100 and 200 observations. Figures 148, 151 and 154 show the bias left unadjusted by (9*b*) for these three sample sizes, revealing its inability to handle the negative bias at and left of the (2,2) corner as well as the positive bias in the *CI*(2,2) situations on the right. Once again, the performance of (10) in Figures 149, 152 and 155 is very similar to that of (9*b*), whereas (11) appears to have less trouble with second-order cointegration, reducing the left-hand negative and the right-hand positive bias considerably. In the (2,2) corner, where there is no cointegration, (11) performs as badly as the first-order approximations. The three approximations also perform more or less equally well in the interior, in absence of any second-order (co-)integration. They are all relatively successful in reducing the bias, and as the sample size increases, more of the remaining bias turns out to be insignificant.

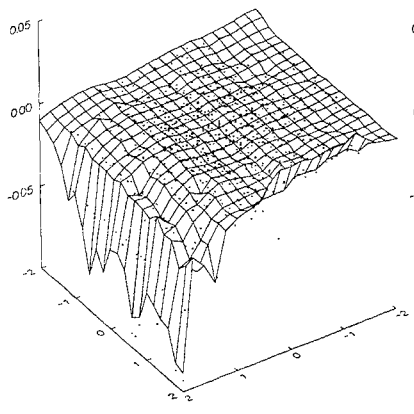

Figure 151: Bias of $\hat{a}_{11}$ less (9*b*)

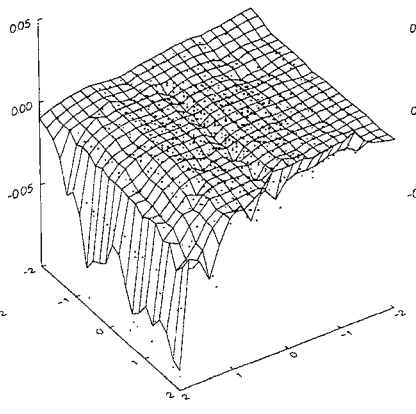

Figure 152: Bias of $\hat{a}_{11}$ less (10)

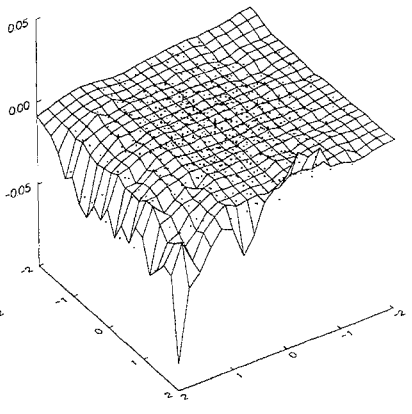

Figure 153: Bias of $\hat{a}_{11}$ less (11)

Based on estimates; sample size 100. Scales represent eigenvalue sums $\rho_{11}+\rho_{12}$ and $\rho_{13}+\rho_{14}$

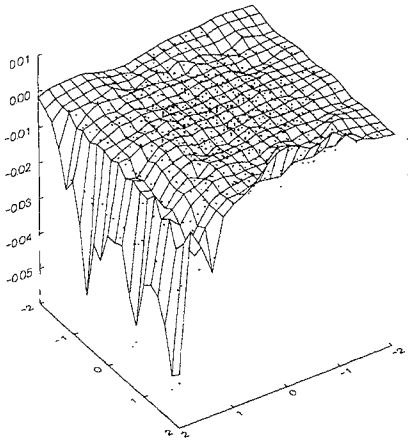

Figure 154: Bias of $\hat{a}_{11}$ less (9*b*)

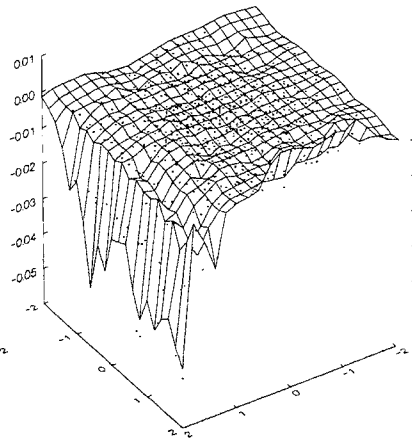

Figure 155: Bias of $\hat{a}_{11}$ less (10)

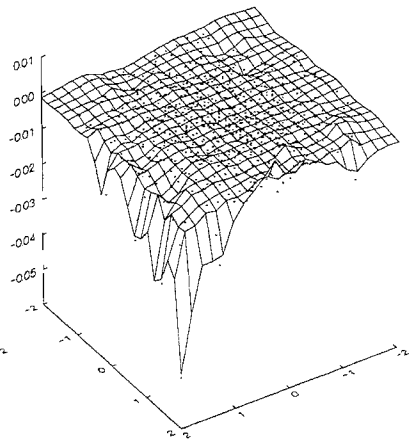

Figure 156: Bias of $\hat{a}_{11}$ less (11)

Based on estimates; sample size 200. Scales represent eigenvalue sums $\rho_{11}+\rho_{12}$ and $\rho_{13}+\rho_{14}$

At the end of this section it will be shown that for the larger sample sizes, the bias of $\hat{a}_{11}$ can in fact be expected to be insignificant for some eigenvalue combinations, but as the previous section demonstrated, all three approximations tend to severely understate the bias when based on estimates, thus their bias-reduced estimates are more likely to be significantly biased than the estimates of (7).

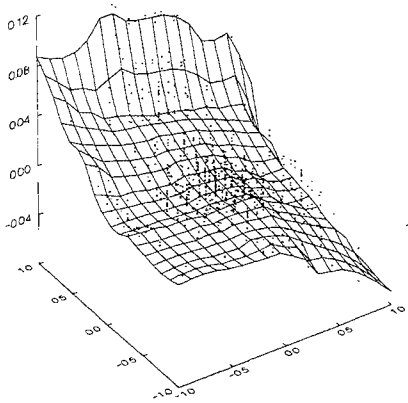

Figure 157: Bias of $\hat{a}_{14}$ less (9*b*)

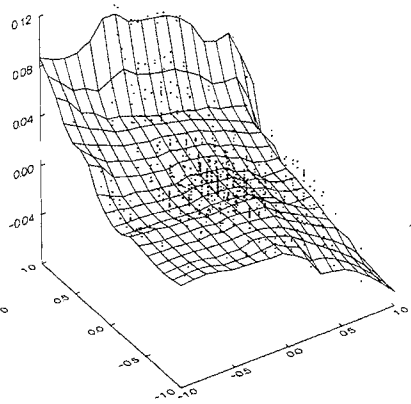

Figure 158: Bias of $\hat{a}_{14}$ less (10)

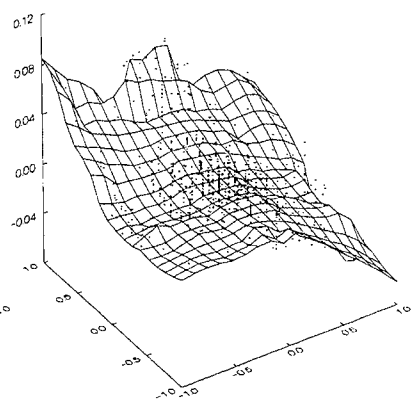

Figure 159: Bias of $\hat{a}_{14}$ less (11)

Based on estimates; sample size 25. Scales represent eigenvalue products $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$

In the case of $\hat{a}_{14}$, one of the second-order cross terms in (8), its bias has previously been found to form a fairly linear function of the eigenvalue products $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$, which is also the projection used in the bias plots to follow. The first three above illustrate the effects of bias reduction using approximations (9*b*), (10) and (11), respectively, in a sample of 25 observations. As in the case of $\hat{a}_{11}$, most of the bias depicted in Figure 71 is indeed reduced, however as the eigenvalue products tend to unity (signifying second-order cointegration) the bias approximations perform slightly worse, leaving unadjusted positive bias at the back as well as negative bias on the right.

Also parallel to the case of $\hat{a}_{11}$, the remaining bias plots of (9*b*) and (10) are almost identical, whereas (11) reduces more of the bias arising in cases of second-order cointegration. Even so, it will become clear that most of the remaining bias is highly significant for all three approximations.

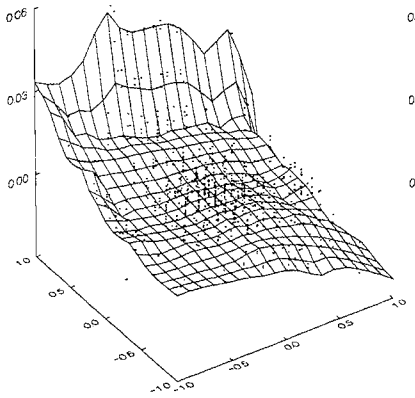

Figure 160: Bias of $\hat{a}_{14}$ less (9*b*)

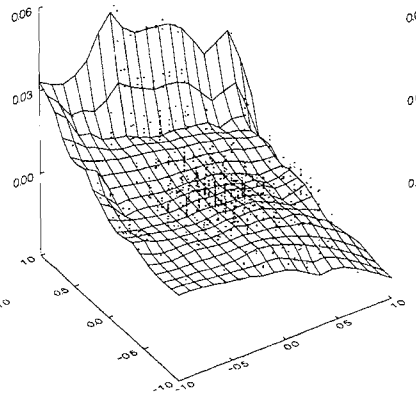

Figure 161: Bias of $\hat{a}_{14}$ less (10)

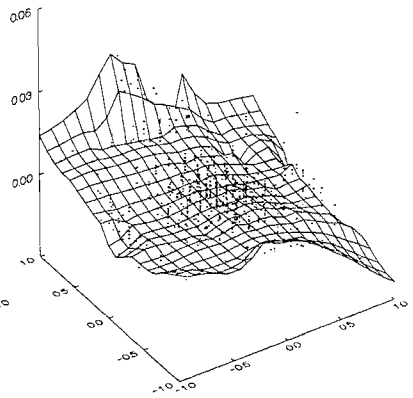

Figure 162: Bias of $\hat{a}_{14}$ less (11)

Based on estimates; sample size 50. Scales represent eigenvalue products $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$

Doubling the sample size to 50 and then again to 100 has little effect in terms of conclusions. Figures 160 through 165 indicate that (9*b*) and (10) again perform almost identically, understating the positive bias at the back as well as the negative bias on the right. As Figure 164 suggests though, (10) leaves more negative right-hand bias unadjusted than (9*b*). Figures 162 and 165 also confirm that the second-order approximation outperforms (9*b*) and (10) but still leaves considerable bias in these two regions.

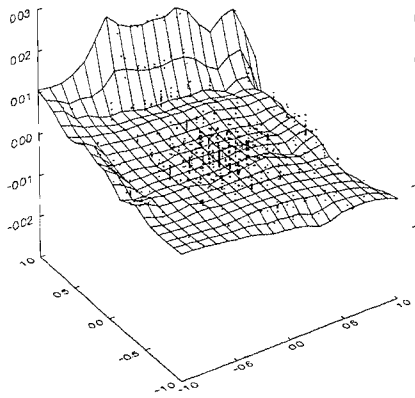

Figure 163: Bias of $\hat{a}_{14}$ less (9*b*)

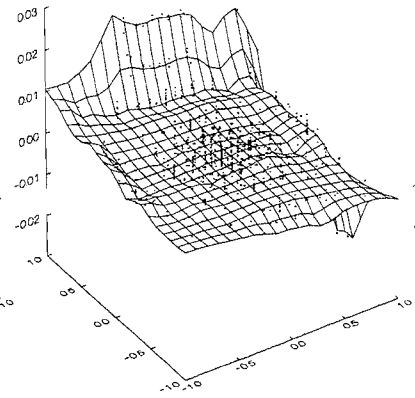

Figure 164: Bias of $\hat{a}_{14}$ less (10)

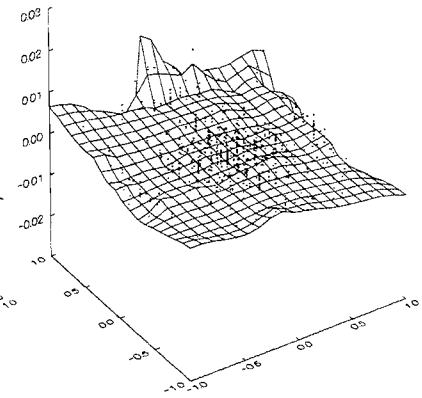

Figure 165: Bias of $\hat{a}_{14}$ less (11)

Based on estimates; sample size 100. Scales represent eigenvalue products $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$

When the sample size increases to 200 (Figures 166, 167 and 168) slightly different results obtain. The approximations again appear to perform more or less identically, but this time the remaining bias has a less regular shape in the interior than for the smaller sample sizes, no doubt because the approximations do not

understate the bias of $\hat{a}_{14}$ as severely as before. The two regions of second-order cointegration display the same sizeable positive and negative bias as before though, in particular for (9*b*) and (10) but to some extent for (11) as well. Figures 175 and 176 will reveal that the remaining bias is highly significant in these regions. They will also show whether or not any insignificant bias is likely to remain along the main diagonal, where a_{14} cancels out. So far, the figures have indicated that little bias remains along that line, in particular for T=200.

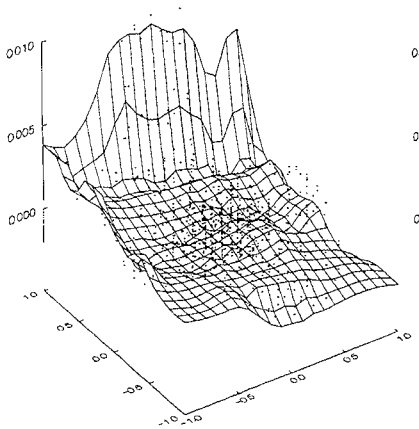

Figure 166: Bias of $\hat{a}_{14}$ less (9*b*)

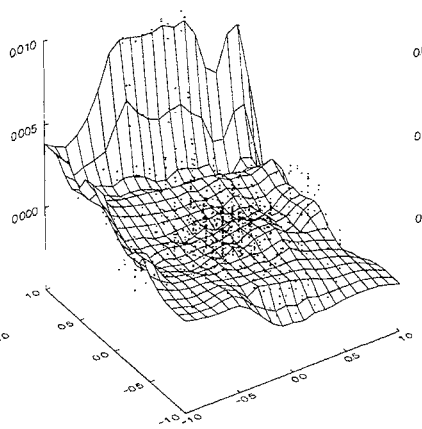

Figure 167: Bias of $\hat{a}_{14}$ less (10)

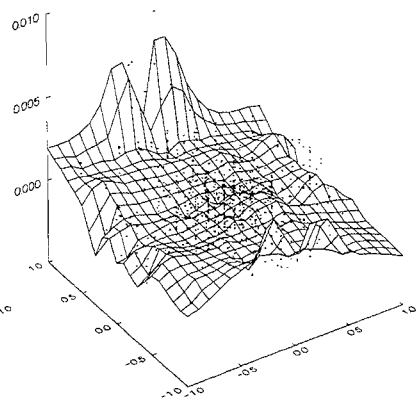

Figure 168: Bias of $\hat{a}_{14}$ less (11)

Based on estimates; sample size 200. Scales represent eigenvalue products $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$

But first Figures 169 and 170 depict the relative bias of $\hat{a}_{11}$ based on 25 observations, after bias reduction using (10) and (11), respectively (since it should be clear by now that (9*b*) and (10) perform almost identically, no relative bias plots are provided for the former approximation). As before, using different approximations to reduce the bias affects the numerator (the bias) as well as the denominator (its standard error) of the relative bias, but since standard errors will normally differ only marginally, differing relative biases will occur mainly as a result of bias differences.

The remaining bias of $\hat{a}_{11}$ after bias reduction turns out in Figures 169 and 170 to be highly significant for most eigenvalue combinations, relative biases ranging from –60 to 40 (cf. Figure 57, which includes only stationary eigenvalue combinations though). At the same time there are quite a few interior combinations for which the remaining bias is insignificant even when the sample size is only 25, so bias reduction appears to have at least some positive effect there. Since the second-order bias approximation was found to outperform the two first-order approximations along the right-hand and left-hand edges (i.e. in *CI*(2,1) and *CI*(2,2) situations) and the relative bias plots reflect mainly bias differences, Figure 170 generally holds smaller relative biases than Figure 169.

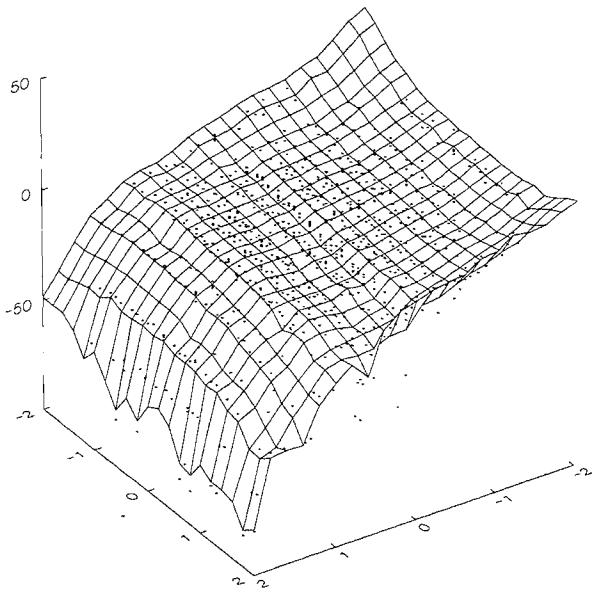

Figure 169: Relative bias of $\hat{a}_{11}$ less (10)

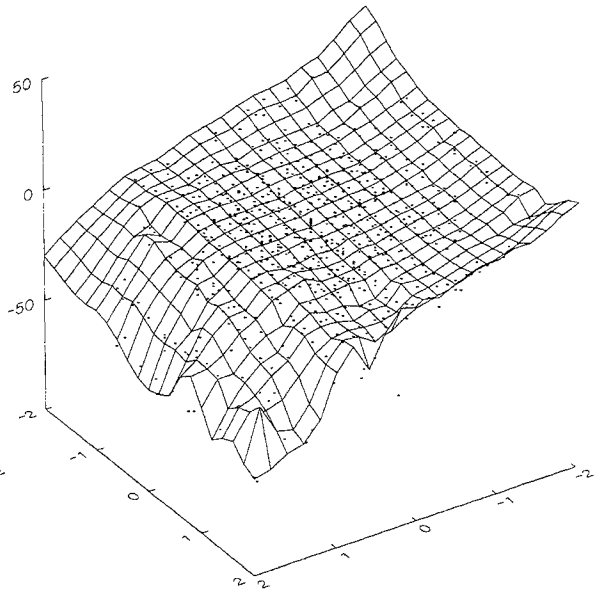

Figure 170: Relative bias of $\hat{a}_{11}$ less (11)

Based on estimates; sample size 25. Scales represent eigenvalue sums $\rho_{11}+\rho_{12}$ and $\rho_{13}+\rho_{14}$

Figures 171 and 172 below convey the benefits of bias reduction even more clearly, because when compared with Figure 58 (the relative bias of $\hat{a}_{11}$ without bias reduction in a sample of 200, excluding unit-root cases) it is clear that the relative biases are not only generally much smaller in Figures 171 and 172 (ranging from –25 to 10) but also mostly insignificant. Only along the four edges is the remaining bias significant, in particular when x_t and y_t are $I(2)$ (with or without cointegration) along the two front edges.

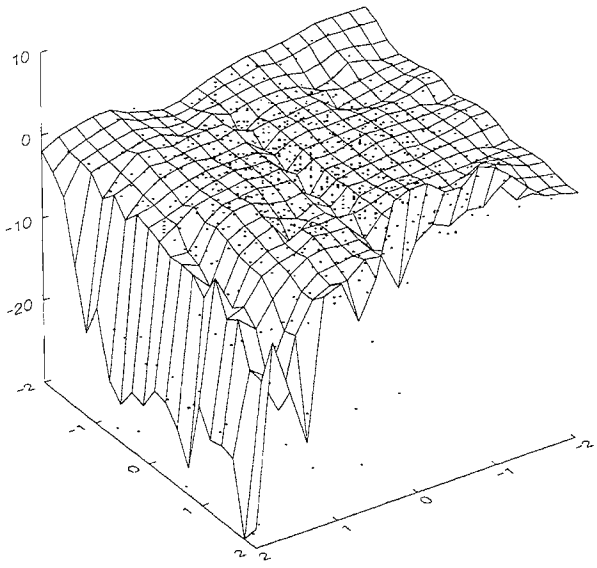

Figure 171: Relative bias of $\hat{a}_{11}$ less (10)

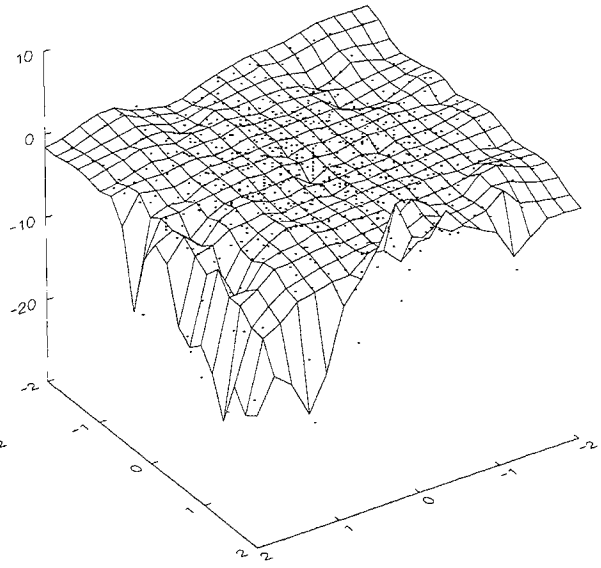

Figure 172: Relative bias of $\hat{a}_{11}$ less (11)

Based on estimates; sample size 200. Scales represent eigenvalue sums $\rho_{11}+\rho_{12}$ and $\rho_{13}+\rho_{14}$

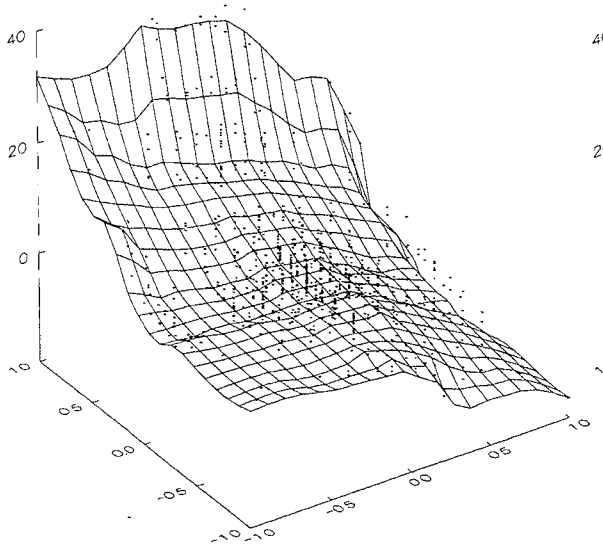

Figure 173: Relative bias of $\hat{a}_{14}$ less (10)

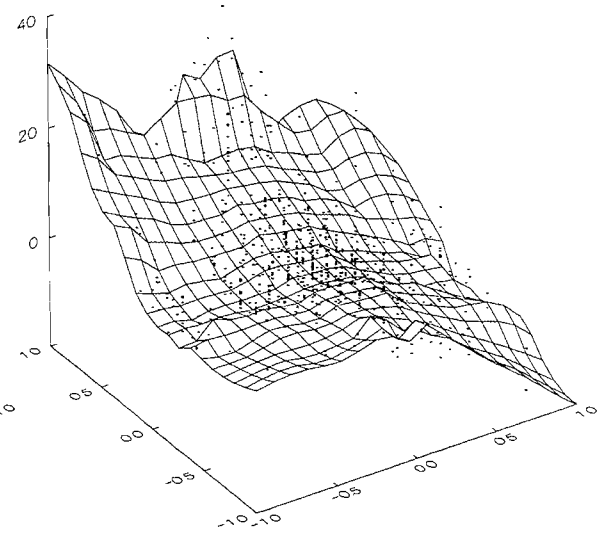

Figure 174: Relative bias of $\hat{a}_{14}$ less (11)

Based on estimates; sample size 25. Scales represent eigenvalue products $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$

The corresponding graphs for the remaining bias of $\hat{a}_{14}$ can be found in Figures 173 through 176. Not surprisingly, the remaining rear positive and right-hand negative bias of $\hat{a}_{14}$ turns out to be highly significant in the above figures (based on 25 observations). Once again a direct comparison with Figure 59 (the relative bias of $\hat{a}_{14}$ without bias reduction in a sample of 25) is invalidated by the fact that unit-root combinations have been included here, but it should be clear that the relative biases are smaller than without bias reduction.

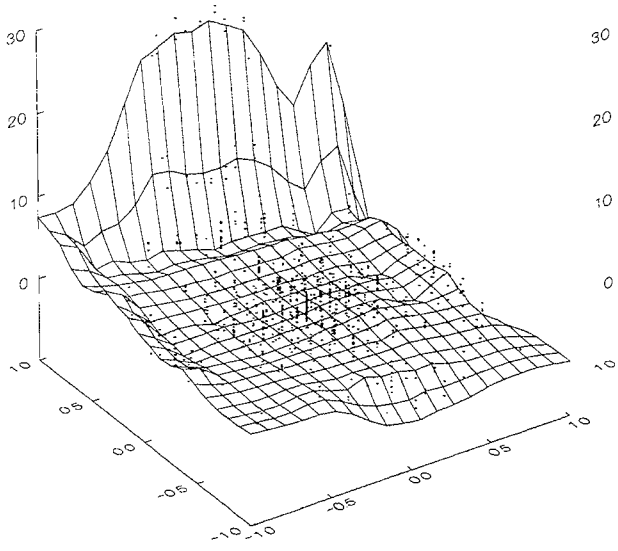

Figure 175: Relative bias of $\hat{a}_{14}$ less (10)

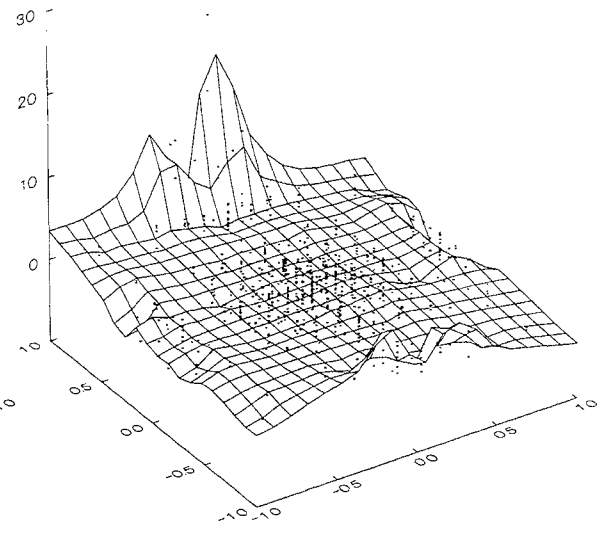

Figure 176: Relative bias of $\hat{a}_{14}$ less (11)

Based on estimates; sample size 200. Scales represent eigenvalue products $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$

In a sample of 200, most of the relative bias depicted in Figure 60 is reduced by (10) and (11). However, the failure of (10) to approximate the positive rear bias as x_t and y_t are $CI(2,2)$ results in enormous relative biases there. The second-order

approximation reduces most of the rear bias but still leaves significantly positive bias. In the interior of Figures 175 and 176, the remaining bias is mainly insignificant, notably along the main diagonal (where a_{14}=0). But at the same time the dispersion of biases at each eigenvalue point is so great that significant estimates of a_{14} along the main diagonal appear to be highly conceivable, as well as insignificant estimates quite far away from it.

4.4 Mean Square Error Effects

Some may argue that mean square error rather than bias is the important parameter by which to evaluate competing estimators. And indeed, little would be gained by bias reduction if it were to add to the variance so that MSE would increase. However, as the following figures (Figures 177 to 200) will demonstrate, this is not the case for any of the three bias reductions nor for any parameter estimate. Mean square errors in fact turn out to be completely dominated by the bias component for most eigenvalue combinations because although the variance of bias-reduced estimates is composed of the original variance, the variance of the approximation based on estimates and the covariance between the two, the latter two will typically be of a smaller order of magnitude than the original variance and therefore variances will not be greatly affected by bias reduction. Since the previous section demonstrated very clearly that the three bias approximations really do reduce the bias of $\hat{\mathbf{A}}$, the MSE comparisons will therefore add very little, mainly reflecting different biases.

The figures in this section depict differences between the MSE before and after bias reduction, so that positive values reflect MSE reductions resulting from bias reduction, while negative values indicate that bias reduction increases mean square errors. MSE difference surfaces will be displayed for sample sizes 25 and 200 only (Brännström (1994*a*) contains more comprehensive MSE tables). For the same reason as in Section 4.3, approximations are based on estimates rather than true values.

In the case of the autoregressive estimate $\hat{a}_1$ of (7), Figures 177 and 178 suggest that there is virtually no MSE difference in a region around ρ=–.3 on the right-hand scale. This is because $\hat{a}_1$ as well as the bias-reduced estimates are unbiased in that region, thus MSE differences there will reflect mainly differences between the variance of $\hat{a}_1$ and the variance of the bias-reduced estimate. As already pointed out, these differences are of a small order of magnitude. On both sides of that region

though, the surfaces are clearly positive, indicating sizeable MSE reductions in favour of the bias-reduced estimates. The slightly higher surface in Figure 178 than in Figure 177 suggests that (11) reduces mean square error more than (9*a*) and (10) as the right-hand eigenvalue tends to unity, but for all practical purposes the two figures below are identical.

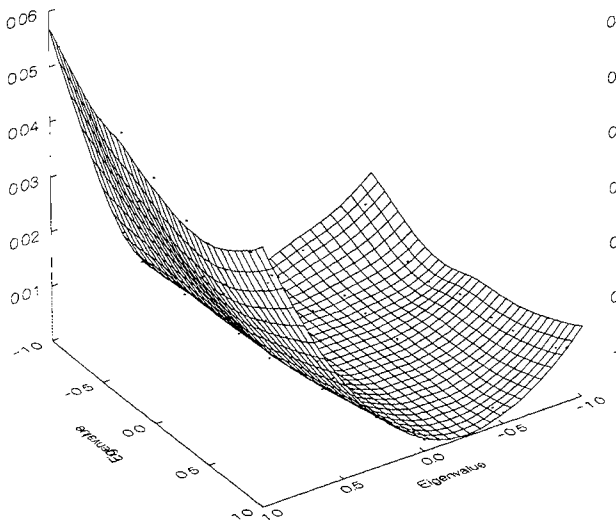

Figure 177: MSE($\hat{a}_1$) – MSE($\hat{a}_1$ less (9*a*) or (10)); T=25

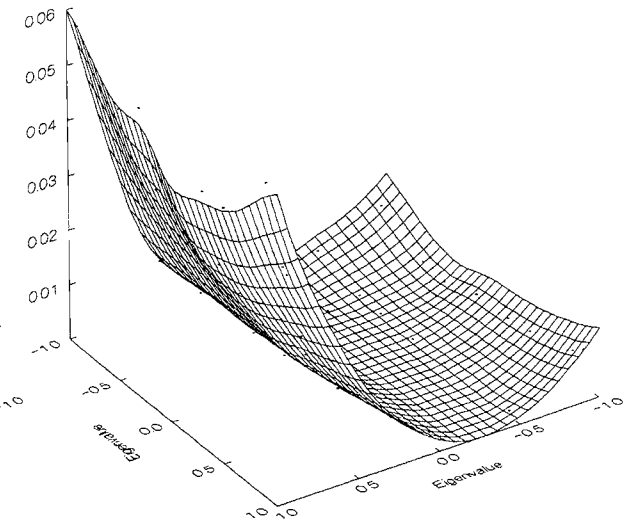

Figure 178: MSE($\hat{a}_1$) – MSE($\hat{a}_1$ less (11)); T=25

Increasing the sample size to 200 does not lead to any dramatically different results (Figures 179 and 180). The two surfaces are still virtually identical, displaying two regions of positive MSE differences and one region of virtually no difference at all as the right-hand eigenvalue is around –.3. But the important feature of all these figures is that there is no region of negative MSE differences, thus reducing the bias appears not to increase mean square error.

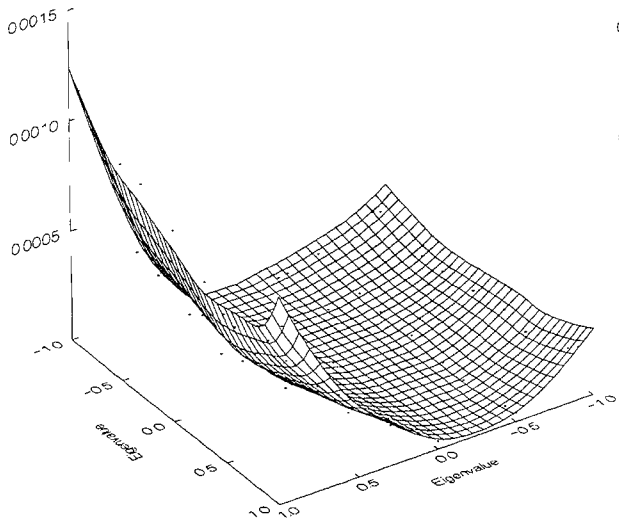

Figure 179: MSE($\hat{a}_1$) – MSE($\hat{a}_1$ less (9*a*) or (10)); T=200

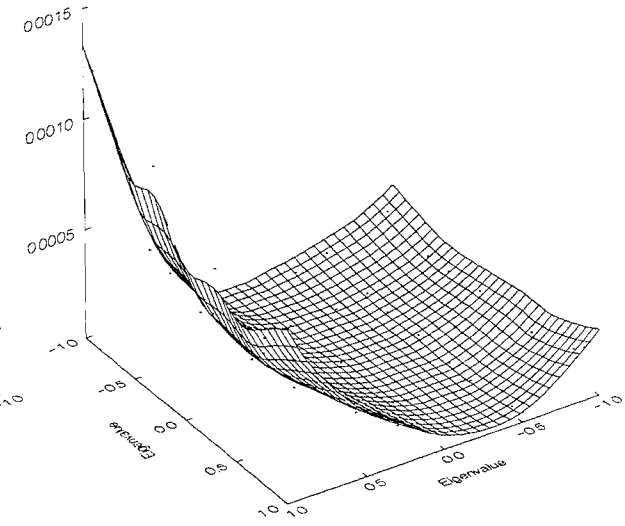

Figure 180: MSE($\hat{a}_1$) – MSE($\hat{a}_1$ less (11)); T=200

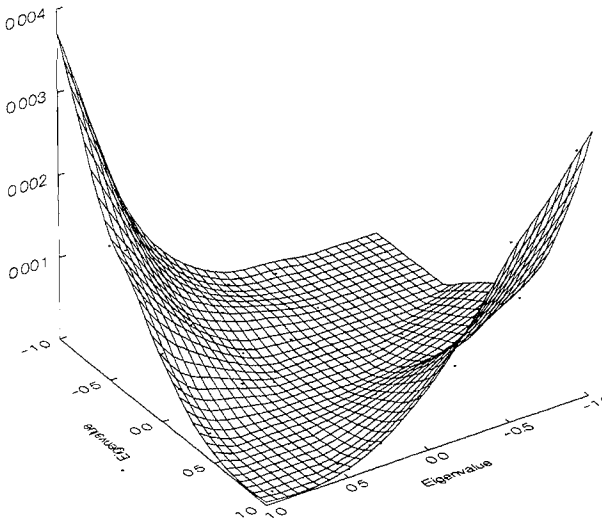

Figure 181: MSE($\hat{a}_2$) – MSE($\hat{a}_2$ less (9*a*) or (10)); T=25

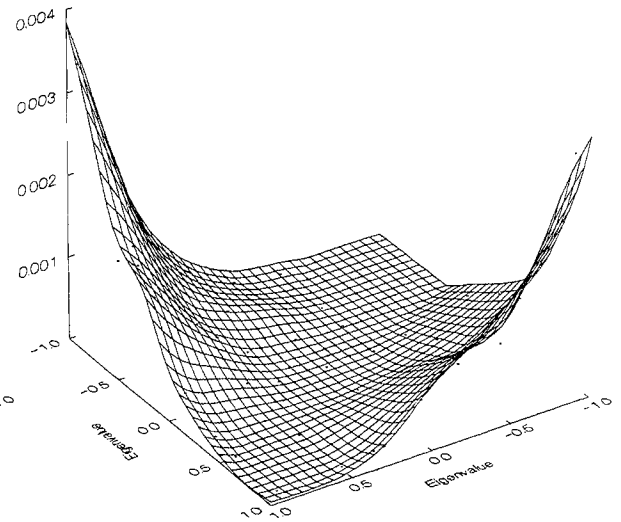

Figure 182: MSE($\hat{a}_2$) – MSE($\hat{a}_2$ less (11)); T=25

Turning to the MSE of $\hat{a}_2$ before and after bias reduction, its difference surfaces can be found in Figures 181 and 182 for T=25 and in Figures 183 and 184 for 200 observations. They differ from previous difference plots in that the regions of zero difference now follow the main diagonal, which is not surprising considering that a_2 is zero and $\hat{a}_2$ (before as well as after bias reduction) was found to be unbiased along that diagonal. On the other hand, $\hat{a}_2$ is severely biased around the two opposite (–1,1) and (1,–1) corners, which is also where the MSE difference plots are the most positive. Thus applying any of the three bias approximations reduces bias as well as mean square error around these corners.

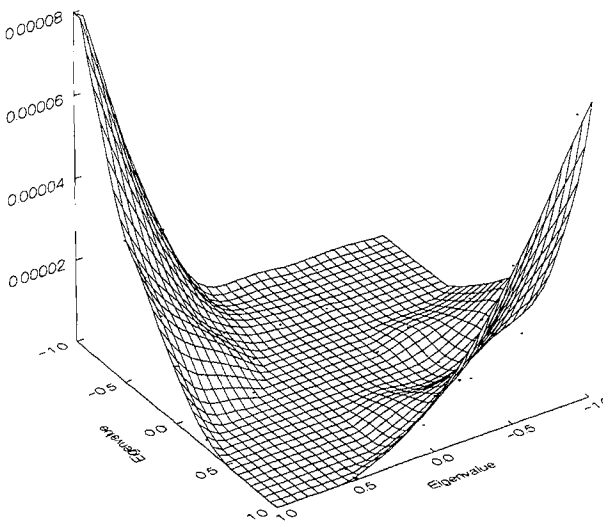

Figure 183: MSE($\hat{a}_2$) – MSE($\hat{a}_2$ less (9*a*) or (10)); T=200

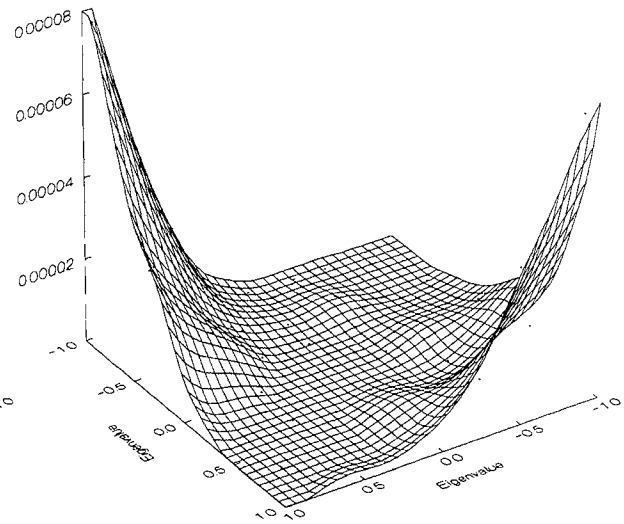

Figure 184: MSE($\hat{a}_2$) – MSE($\hat{a}_2$ less (11)); T=200

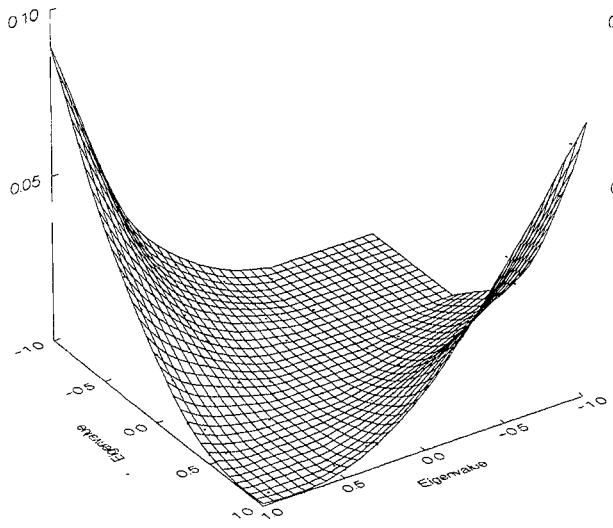

Figure 185: MSE($\hat{a}_3$) – MSE($\hat{a}_3$ less (9a) or (10)); T=25

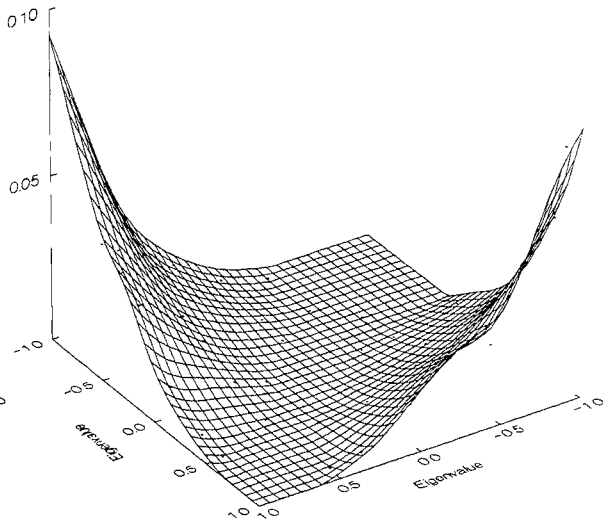

Figure 186: MSE($\hat{a}_3$) – MSE($\hat{a}_3$ less (11)); T=25

Not surprisingly, the MSE differences for $\hat{a}_3$ before and after bias reduction form surfaces similar to those of $\hat{a}_2$, with no MSE difference along the main diagonal where a_3 is zero and $\hat{a}_3$ can be expected to be unbiased. Around the (1,–1) and (–1,1) corners though, the mean square error of $\hat{a}_3$ before bias reduction exceeds that of the bias-reduced estimates, leaving positive MSE differences in Figures 185 through 188. Slightly greater MSE differences obtain when (11) is used to reduce the bias of $\hat{a}_3$ than for (9a) or (10), reflecting the advantage of the second-order approximation as eigenvalues tend to unity.

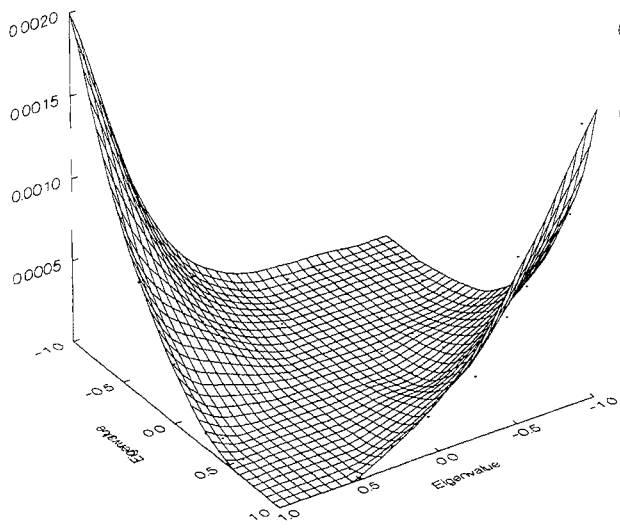

Figure 187: MSE($\hat{a}_3$) – MSE($\hat{a}_3$ less (9a) or (10)); T=200

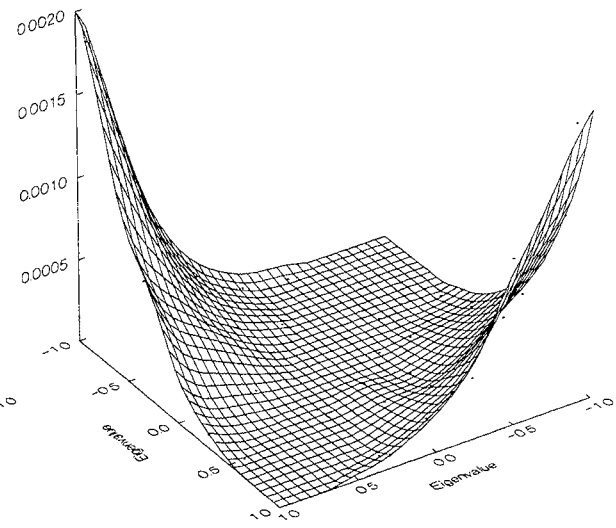

Figure 188: MSE($\hat{a}_3$) – MSE($\hat{a}_3$ less (11)); T=200

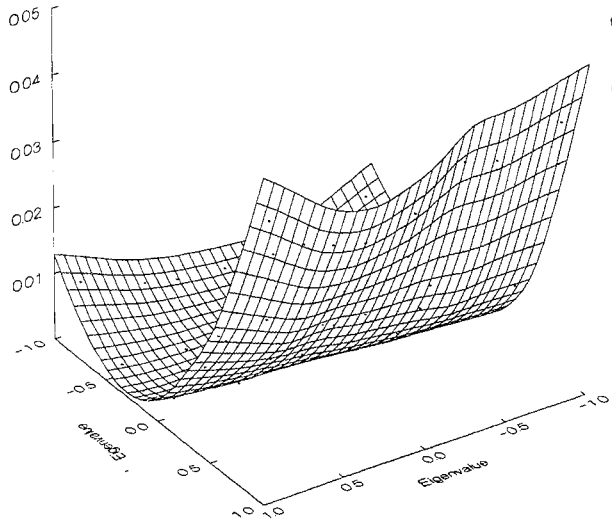

Figure 189: MSE($\hat{a}_4$) – MSE($\hat{a}_4$ less (9*a*) or (10)); T=25

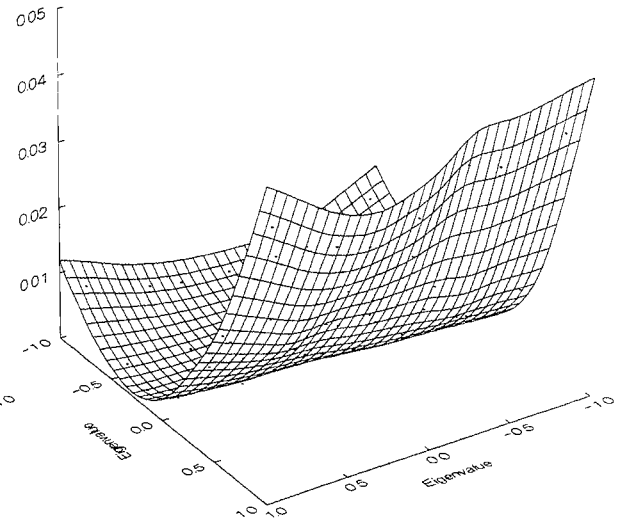

Figure 190: MSE($\hat{a}_4$) – MSE($\hat{a}_4$ less (11)); T=25

Finally, the MSE differences for $\hat{a}_4$ (Figures 189 through 192) are similar to those of $\hat{a}_1$ except for the roles played by the two eigenvalues. The difference is now close to zero as the left-hand eigenvalue is around –.3 and positive as it approaches ±1. The MSE difference plots based on first- and second-order approximations are virtually identical and always positive, again indicating that bias reduction does not lead to increasing MSE.

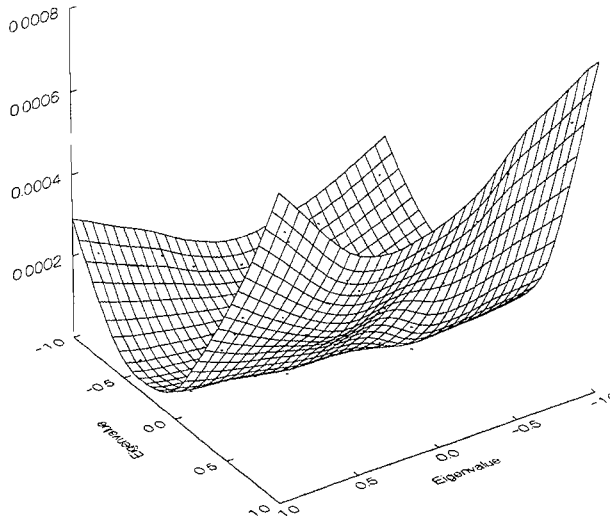

Figure 191: MSE($\hat{a}_4$) – MSE($\hat{a}_4$ less (9*a*) or (10)); T=200

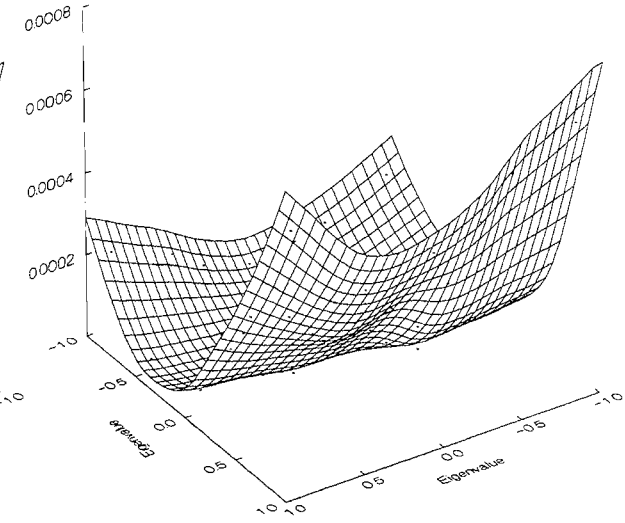

Figure 192: MSE($\hat{a}_4$) – MSE($\hat{a}_4$ less (11)); T=200

Turning to the corresponding difference plots for $\hat{a}_{11}$ and $\hat{a}_{14}$ of the second-order model, only figures based on (10) and (11) will be given here since it should be clear by now that (9*b*) and (10) differ only marginally when the order p is 2.

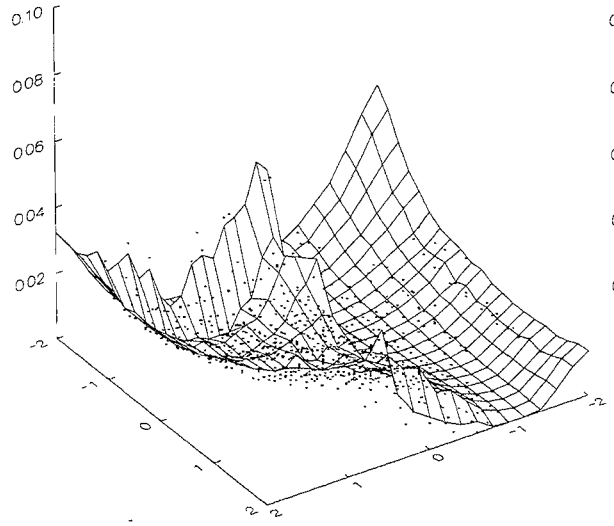

Figure 193: MSE($\hat{a}_{11}$) – MSE($\hat{a}_{11}$ less (10)); T=25

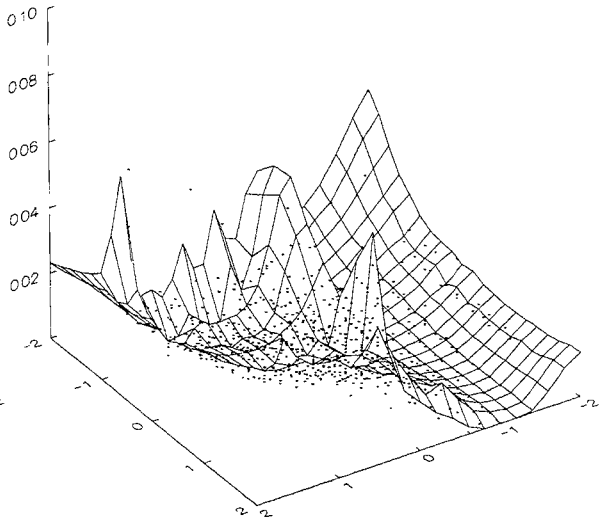

Figure 194: MSE($\hat{a}_{11}$) – MSE($\hat{a}_{11}$ less (11)); T=25

Scales represent eigenvalue sums $\rho_{11}+\rho_{12}$ and $\rho_{13}+\rho_{14}$

Using (10) or (11) to reduce the bias of $\hat{a}_{11}$ produces MSE difference plots similar to those of $\hat{a}_1$ (cf. Figures 177 to 180) only less regular. In a sample of 25, (10) and (11) appear to reduce mean square error in and around the four corners in Figures 193 and 194, the difference plots being positive there. At the (2,2) corner, for instance, (10) and (11) appear to reduce mean square error considerably, leaving very large positive differences in the figures above, despite their only partial success in reducing the bias (cf. Figures 146 and 147). On the other hand there is a vast region in the interior where there is virtually no difference between the MSE before and after bias reduction, typically for intermediate negative eigenvalue sums. That is also where $\hat{a}_{11}$ was found to be the least biased, before as well as after bias reduction. As always though, the important point to make is that no negative MSE differences occur.

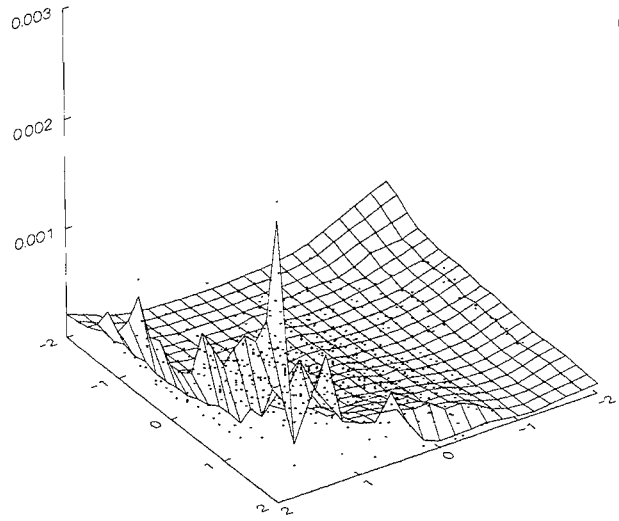

Figure 195: MSE($\hat{a}_{11}$) – MSE($\hat{a}_{11}$ less (10)); T=200

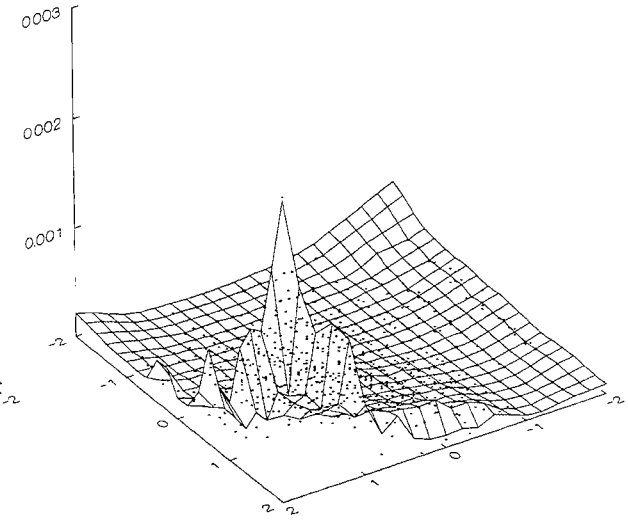

Figure 196: MSE($\hat{a}_{11}$) – MSE($\hat{a}_{11}$ less (11)); T=200

Scales represent eigenvalue sums $\rho_{11}+\rho_{12}$ and $\rho_{13}+\rho_{14}$

The corresponding figures for a sample size of 200 can be found in Figures 195 and 196. Again an interior region of little MSE difference is surrounded by regions of positive differences, in particular around the (2,2) corner, indicating that bias reduction reduces MSE as well.

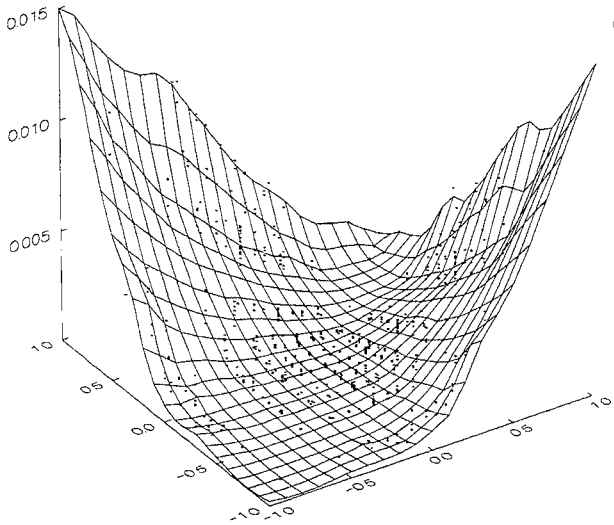

Figure 197: MSE($\hat{a}_{14}$) – MSE($\hat{a}_{14}$ less (10)); T=25

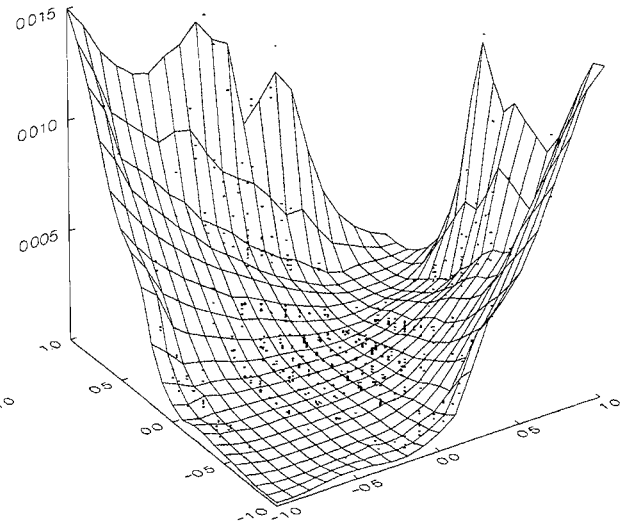

Figure 198: MSE($\hat{a}_{14}$) – MSE($\hat{a}_{14}$ less (11)); T=25

Scales represent eigenvalue products $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$

Finally, the difference plots for $\hat{a}_{14}$ (Figures 197 and 198 for T=25; Figures 199 and 200 for T=200) are very similar to those of the cross-term estimates $\hat{a}_2$ and $\hat{a}_3$ in that they display no MSE difference along the main diagonal, where a_{14} is zero and can be expected to be estimated without bias, but positive differences off that diagonal. The conclusion, that bias reduction does not increase mean square error but may in fact reduce it, was also reached by Sastry Pantula and Wayne Fuller for bias-reduced estimates in a univariate AR(2) setting (Pantula and Fuller 1985).

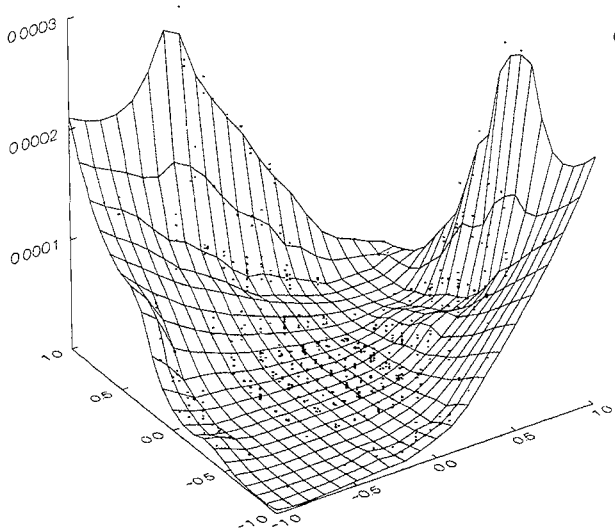

Figure 199: MSE($\hat{a}_{14}$) – MSE($\hat{a}_{14}$ less (10)); T=200

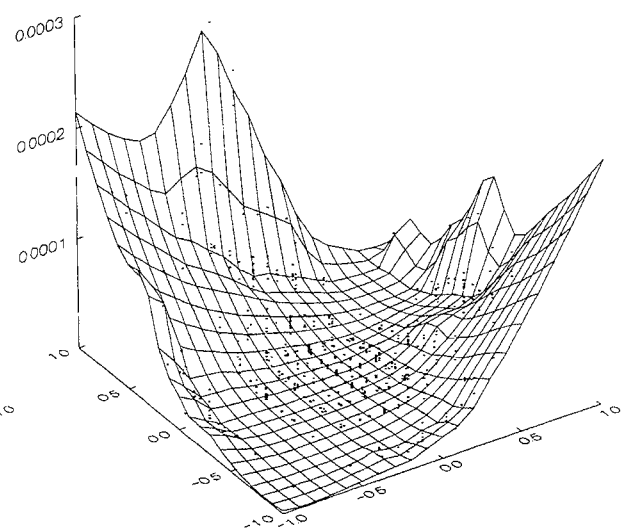

Figure 200: MSE($\hat{a}_{14}$) – MSE($\hat{a}_{14}$ less (11)); T=200

Scales represent eigenvalue products $\rho_{11}\rho_{12}$ and $\rho_{13}\rho_{14}$

4.5 Chapter Summary

By far the most important conclusion of the Monte Carlo-based evaluations in this chapter is that all three bias approximations are capable of bias approximation and reduction. Though approximating and reducing the bias perfectly only by exception (e.g. for sample sizes greater than 25 in a stationary bivariate first-order VAR model whose parameters **A**, **G**, $\mathbf{C}_0$ and ρ are known), the three approximations were found to work in the right direction and correlate strongly (except for $\hat{a}_{12}$ and $\hat{a}_{21}$) with the bias, thereby providing at least some bias reduction. One line of research not pursued here but open for future examination is the possibility of further reducing the bias in such cases by iterating the procedure used in this chapter until some form of convergence criteria are met, in each iteration deriving potentially less biased estimates of **A**, **G**, $\mathbf{C}_0$ and ρ.

To summarise this chapter's other important results, the approximations appear to perform better:

(*i*) When based on true parameters rather than estimates.

(*ii*) When the model is first-order (p=1) rather than second-order (p=2).

(*iii*) The larger the sample. Nonetheless, even series the length of 200 observations appear to be much too short for them to approximate the bias well based on estimates in a bivariate second-order model.

(*iv*) For stationary eigenvalue combinations rather than for integrated and co-integrated systems. Moreover they appear to work better in cointegrated than integrated systems.

(*v*) Based on (11) rather than (9) or (10). Whereas the latter two are equivalent if p is 1 and were found to be almost equivalent for p=2, (11) improves the performance considerably as eigenvalues approach unity, causing (9) and (10) to leave bias unadjusted. If the system is stationary though, (11) is practically equivalent with (9) and (10). In addition, the bias of $\hat{a}_{12}$ and $\hat{a}_{21}$ of the second-order model is better approximated (though still not well) by (11) than by the two first-order approximations.

5. Money Growth and Inflation: An Application

In this chapter the ideas put forward in previous chapters will be tested on a real data set consisting of quarterly observations on the US monetary base (M1) and consumer prices (CPI) from 1962 to 1994. First a bivariate VAR model will be estimated, then its bias will be reduced using the three bias approximations introduced in Chapter 3 and evaluated in Chapter 4. The bias-reduced estimates will then be evaluated by comparing the generated impulse-response functions, variance decompositions and forecasts for the four last years with those obtained without bias reduction.

Models of the demand for money and inflation appear very frequently in literature, most of the time with other variables, such as output and interest rates, included in the model as well. Suffice it to mention Sims (1980*b*), Litterman and Weiss (1985), Runkle (1987), Hallman *et al.* (1991), and Webb (1995). Christopher Sims (1980*b*) tried both a four-variable VAR (money, inflation, output and interest rates) and a model without the interest rate variable, and found the inclusion of interest rates to make a dramatic difference. For instance, without interest rates, innovations to the monetary base were found to explain some 37 per cent of the forecast error variance of output in the long run, but including interest rates that proportion dropped to a mere 4 per cent. Robert Litterman and Laurence Weiss (1985) tried to find support in data for the links between real and nominal variables suggested by real business cycle theory, but failing to do so they developed an alternative model in which output is unaffected by money supply changes, and money is supplied with the objective of stabilising prices in the short run. David Runkle (1987) estimated the same four-variable VAR model as Sims (1980*b*) in order to demonstrate that the results are very sensitive to model specification; different variable definitions and sampling periods may lead to very different results, for instance. John Hallman *et al.* (1991) advocated the use of their P^* model, linking the inflation rate to the level of money. Roy Webb (1995) explicitly focused on the often poor fit of the inflation equation of the model and the problems it leads to in terms of out-of-sample performance, etc. His results indicate that introducing variables to account for policy regime shifts and using different lag lengths in different equations may reduce the problem. This latter approach was also found to work satisfactorily in Edlund and Karlsson (1993).

The model presented in this chapter is far too simple to pretend to represent any real progress in this realm. It simply makes use of the standard quantity theory notion that a relative change in money supply will, eventually, lead to an equally large relative change in prices, and of Litterman and Weiss' observation that money supply is controlled so that short-run price stability be maintained.

5.1 Data

The data set used to exemplify the consequences of bias and of bias reduction involves two US series; quarterly M1 and CPI from 1962 to 1994 (from OECD's *Main Economic Indicators*). Both series were measured by the middle of the third month of the quarter and both series have been seasonally adjusted. Figure 201 shows the logarithms of the series:

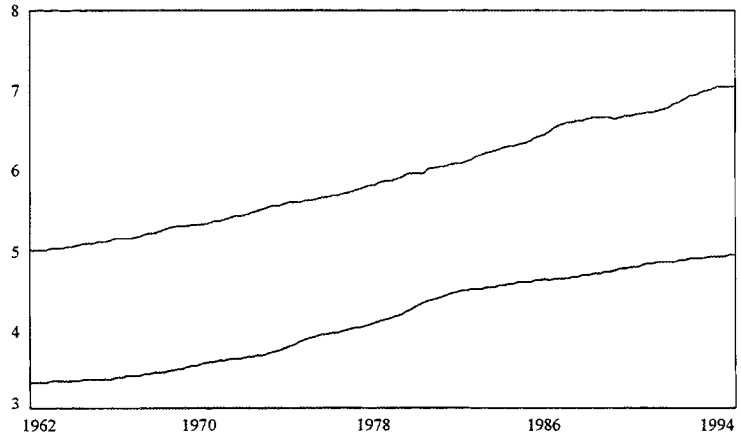

Figure 201: Logarithms of M1 (top) and CPI (bottom)

Both series appear to grow fairly steadily over this 33-year period, and in the case of M1 the growth rate does not change much over time (except for some fluctuations by the end of the period). The growth rate of prices changes though, more precisely there are three distinct inflation regimes in Figure 201: Until the middle of the 1960s US inflation was low, then it rose for a couple of years due to the Vietnam war build-up. The higher inflation rate stayed in effect even after the war was over, and not until Congress and the Federal Bank jointly decided to pursue a low-inflation policy in the early 1980s did it change to about the same rate as before the effects of the war build-up. Webb (1995) dates the first regime shift to the end of 1966 and the return to lower inflation rates to the middle of 1981, but that is an opinion not shared by everybody. For instance, Balke and Fomby (1991) identified the two regime shifts in 1968 and 1983 instead. Fortunately, that problem can be ignored completely here, even though Webb (1995) considered these regime shifts to be a major reason why estimated

inflation equations perform so badly. But since the objective here is not to estimate the best possible model of money and inflation, but a reasonably good benchmark model, the VAR model to be specified will not explicitly allow for regime shifts.

The first step towards a specified VAR model is a check for non-stationarity in the two series. To that end, the following two Dickey-Fuller equations were estimated over the full 33-year period (131 observations):

$$\widehat{\ln \mathrm{M1}_t} = .2017 + .0007t + .9603 \ln \mathrm{M1}_{t-1}$$

$$\widehat{\ln \mathrm{CPI}_t} = .0436 + .0002t + .9898 \ln \mathrm{CPI}_{t-1}$$

The hypothesis that M1 is integrated cannot be rejected at the 10 per cent significance level since $131(.9603-1) = -5.2$ (the 10 % critical values for $\hat{\rho}_\tau$ in Fuller (1976) are –2.62 and –20.7 for 100 observations) and the ratio of (.9603–1) to its standard error is –2.6 (10 % critical values –.90 and –3.45 for 100 observations). For prices the case is less clear since $131(.9898-1) = -1.3$ exceeds –2.62, whereas the ratio between (.9898–1) and its standard error is –1.1, which is not significant at the 10 per cent level. Since none of the series appears to be $I(2)$ (the $\hat{\rho}_\tau$ test statistics are –86 and –131, clearly significant at the 1 % level) both series will be treated as $I(1)$, which means that they may be modelled in differences rather than in levels, provided M1 and CPI are not cointegrated (as Holden (1995) points out, it is inappropriate to take differences of series which are cointegrated in levels). However, the hypothesis of no cointegrating equation cannot be rejected using Johansen's LR test. The test statistic is only 11.2, which is less than the 5 % critical value of 15.41 (Johansen 1991). Thus the conclusion must be that M1 and CPI are $I(1)$ but not $CI(1,1)$ (which is the conclusion in Webb (1995) as well) and therefore it is justified to use the following differences henceforth: $m_t = \ln \mathrm{M1}_t - \ln \mathrm{M1}_{t-1}$ and $p_t = \ln \mathrm{CPI}_t - \ln \mathrm{CPI}_{t-1}$. This operation, whilst shifting attention from the levels of money and prices to their relative changes, also makes both series stationary, which was seen in Chapter 4 to be a desirable property for the bias approximations to work well. The remainder of this chapter will deal exclusively with the 131 relative changes m_t and p_t depicted in Figure 202:

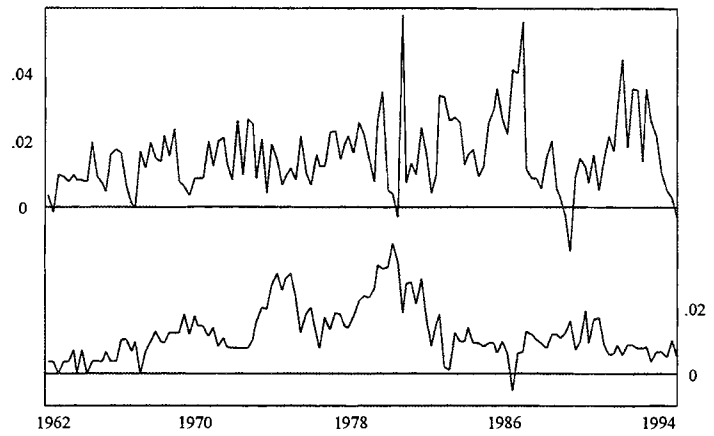

Figure 202: Quarterly relative M1 change m_t (left-hand scale) and quarterly relative CPI change p_t (right-hand scale)

As seen in Figure 202, both growth rates are mainly positive and fluctuate quite a lot, despite the regular appearance of $M1_t$ and CPI_t in Figure 201. The money supply growth rate m_t peaked in 1980Q3, when M1 grew by 5.7 per cent from the previous quarter, and in 1986Q4 (by 5.5 per cent). Decreasing money supply occurred in four quarters; by as much as 1.3 per cent in 1989Q2. Prices dropped in one quarter only, 1986Q2 (by half a per cent), and rose by a record 3.9 per cent in 1980Q1. The mean quarterly growth rates were about 1.6 per cent for money and about 1.2 per cent for prices and the standard deviations were about 1.1 per cent for m_t and .8 per cent for p_t.

If interest is centred on the dynamics between m_t and p_t, e.g. how shocks to one of the two affect the other over time, a bivariate VAR model is a suitable tool. Such a model will be specified and estimated in the following section.

5.2 The VAR Model

In order to examine the dynamics between m_t and p_t and to generate forecasts, the following bivariate VAR(p) model will be specified and estimated:

$$\begin{bmatrix} m_t \\ p_t \end{bmatrix} = \mu + \mathbf{A}_1 \begin{bmatrix} m_{t-1} \\ p_{t-1} \end{bmatrix} + \mathbf{A}_2 \begin{bmatrix} m_{t-2} \\ p_{t-2} \end{bmatrix} + \cdots + \mathbf{A}_p \begin{bmatrix} m_{t-p} \\ p_{t-p} \end{bmatrix} + \varepsilon_t$$

where μ and ε_t are 2×1 vectors and $\mathbf{A}_i$ (i=1,...,p) are 2×2 parameter matrices.

In order to first find the optimal value for the lag length p the following three information criteria were evaluated for the five lag lengths p = 0,...,4; Akaike's (AIC), Schwartz's (SC) and Hannan-Quinn's (HQ, also known as LIL):

$$\mathrm{AIC}(p) = \ln\left|\hat{\mathbf{G}}(p)\right| + \frac{2K^2 p}{T}$$

$$\mathrm{SC}(p) = \ln\left|\hat{\mathbf{G}}(p)\right| + \frac{K^2 p}{T}\ln T$$

$$\mathrm{HQ}(p) = \ln\left|\hat{\mathbf{G}}(p)\right| + \frac{2K^2 p}{T}\ln\ln T$$

where $\hat{\mathbf{G}}$ is the least-squares estimate of the variance-covariance matrix $\mathbf{G} = E(\varepsilon_t \varepsilon_t')$ of the vector ε_t of disturbances; K is the dimension of the process (K=2 in this case); and T is the sample size (T=131 here). The objective in all three cases is to minimise the value of the criterion and to select as the optimal lag length p that value which yields the minimum.

Estimating the bivariate model above by means of least squares with increasing lag lengths p results in the following values of AIC, SC and HQ:

p	AIC(p)	SC(p)	HQ(p)
0	−18.69	−18.69	−18.69
1	−19.84	−19.75	−19.79
2	−19.86	−19.69	−19.80
3	−19.90	−19.64	−19.79
4	−19.89	−19.55	−19.75

Unfortunately the three criteria are discordant; whereas Hannan-Quinn's criterion is minimised by p=2, Akaike's is minimised by p=3 and Schwartz's by p=1. As a compromise solution p=2 will be used in the remainder. (A first-order model was estimated as well, but due to its unsatisfactory diagnostics it was rejected.)

Thus the VAR(2) model to be estimated is

$$\begin{bmatrix} m_t \\ p_t \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} m_{t-1} \\ p_{t-1} \end{bmatrix} + \begin{bmatrix} a_{13} & a_{14} \\ a_{23} & a_{24} \end{bmatrix} \begin{bmatrix} m_{t-2} \\ p_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{m,t} \\ \varepsilon_{p,t} \end{bmatrix}$$

which can, as every VAR model, also be expressed as a VAR(1) process:

$$\begin{bmatrix} m_t \\ p_t \\ m_{t-1} \\ p_{t-1} \end{bmatrix} = \begin{bmatrix} \mu_m \\ \mu_p \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} m_{t-1} \\ p_{t-1} \\ m_{t-2} \\ p_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{m,t} \\ \varepsilon_{p,t} \\ 0 \\ 0 \end{bmatrix} \tag{12}$$

From this point we will deal with (12) only, bearing in mind that a_{11} and a_{22} are first-order autoregressive parameters, that a_{13} and a_{24} are second-order AR parameters, and that the remaining parameters of the parameter matrix **A** are cross-term parameters.

5.3 Estimated VAR Model

Estimating (12) by means of least squares over the full sample (1962Q2–1994Q4) yielded the following parameter estimates (with t ratios in parentheses below):

$$\hat{\boldsymbol{\mu}} = \begin{bmatrix} .010 \\ (3.81) \\ .000 \\ (.38) \\ 0 \\ 0 \end{bmatrix} \quad \hat{\mathbf{A}} = \begin{bmatrix} .310 & -.050 & .133 & -.003 \\ (3.41) & (-.26) & (1.45) & (-.02) \\ .031 & .617 & .058 & .235 \\ (.74) & (6.91) & (1.38) & (2.65) \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \hat{\mathbf{G}} = \frac{\mathbf{e}_t \mathbf{e}_t'}{T} = \begin{bmatrix} 10.2 & -.963 \\ -.963 & 2.14 \end{bmatrix} \cdot 10^{-5} \qquad (13)$$

Adjusted R^2 is only .12 for the money growth equation and .65 for the inflation equation. The estimated eigenvalues of $\hat{\mathbf{A}}$, i.e. the solutions to the equation $|\lambda \mathbf{I} - \hat{\mathbf{A}}| = 0$, are .56, –.23, .87, and –.28, which means that the estimated system is stationary since all eigenvalues are less than one (in absolute terms). Also, all four eigenvalues being real indicates that no seasonal or cyclical elements remain in the two series. Residuals are shown in Figure 203 (m_t residuals on left-hand scale; p_t residuals on right-hand scale).

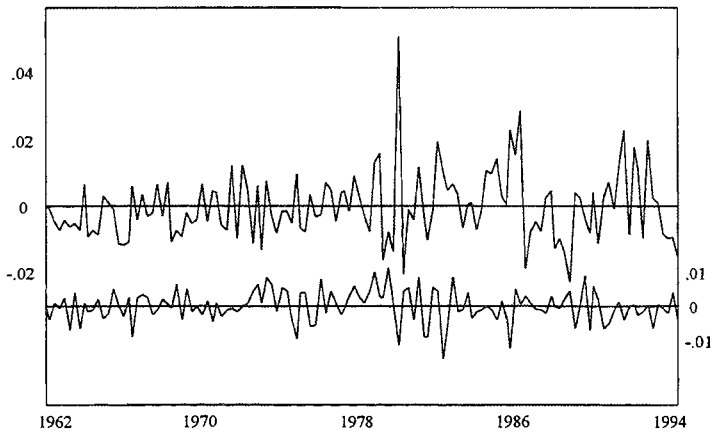

Figure 203: Full-sample residuals of m_t (left-hand scale) and of p_t (right-hand scale) based on (13)

Whereas the inflation residuals look fairly good, the money growth residuals show clear signs of heteroscedasticity and maybe autocorrelation as well. Nevertheless, the hypothesis that the two series of residuals are white noise cannot be rejected at the 5 % significance level in a nine-quarter portmanteau test (χ^2=40.4; critical value 41.3). Furthermore, augmented Dickey-Fuller tests reject the hypotheses that the residual series are integrated (t ratios 3.41 and 3.45, respectively, while the 5 % critical value reported in Engle and Yoo (1987) is 3.37 for 100 observations).

The two most apparent features about (13) are firstly, that m_t and p_t are negatively correlated ($r\approx-.21$), which not only contradicts standard quantity theory but is counterintuitive as well (it is consistent with the Litterman-Weiss model though) and secondly, that the t ratios of the four cross-term parameter estimates suggest they are all insignificant, at least at significance levels below 5 per cent. But a word of warning is due here: All inference in estimated VAR models is obscured by the fact that neither the parameter matrix **A** nor the variance-covariance matrix **G** can be estimated without bias, so it is not obvious that a_{23} (for instance) should be zero just because the t ratio of $\hat{a}_{23}$ is merely 1.38. Nevertheless, since the F statistic for the test of a_{12} and a_{14} being zero is only .10 (p-value .90), the proposition that inflation does not Granger-cause money growth appears to be reasonable even given the bias of $\hat{\mathbf{A}}$ and $\hat{\mathbf{G}}$. The test of the hypothesis $a_{21}=a_{23}=0$ has a p-value of only .16 though (F=1.88), which means that Granger-causality from money growth to inflation is not as easily dismissed, bearing in mind the bias of $\hat{\mathbf{A}}$ and $\hat{\mathbf{G}}$. Nevertheless, the following model of seemingly unrelated regressions was also estimated as an alternative:

$$\hat{\boldsymbol{\mu}} = \begin{bmatrix} .008 \\ (4.60) \\ .002 \\ (2.65) \\ 0 \\ 0 \end{bmatrix} \quad \hat{\mathbf{A}} = \begin{bmatrix} .329 & 0 & .157 & 0 \\ (3.84) & & (1.83) & \\ 0 & .619 & 0 & .216 \\ & (7.31) & & (2.57) \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \hat{\mathbf{G}} = \frac{\mathbf{e}_t \mathbf{e}_t'}{T} = \begin{bmatrix} 10.2 & -.995 \\ -.995 & 2.21 \end{bmatrix} \cdot 10^{-5} \qquad (14)$$

As can be seen, imposing the zero restrictions on (13) has relatively small effects on remaining estimates. Adjusted R^2 also remains largely unchanged, as do the estimated eigenvalues of $\hat{\mathbf{A}}$ and the residuals in Figure 204.

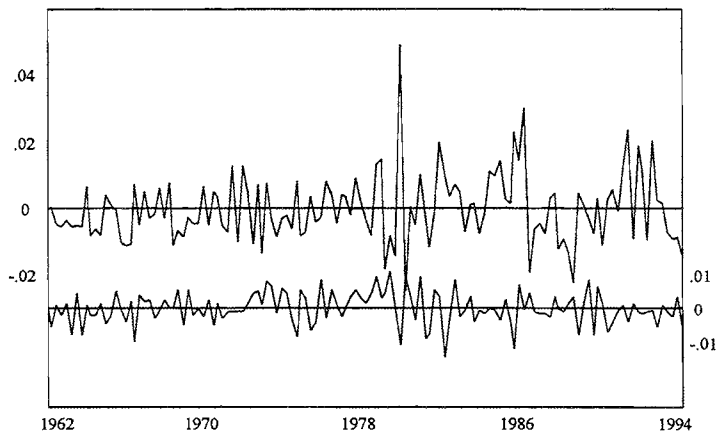

Figure 204: Full-sample residuals of m_t (left-hand scale) and of p_t (right-hand scale) based on (14)

Before turning to the different bias reductions, let us investigate the dynamics between m and p implied by the estimated full VAR model (13). The first main instrument with which to perform such an investigation is the impulse-response function, in which repercussions of shocks to the estimated system are traced through time. The second is variance decompositions, which depict to what variable the forecast error variances are attributable at different forecast horizons.

Figures 205 and 206 show the first 12 responses (three years) to shocks of one standard deviation to m_t and p_t respectively, based on the full-sample estimates in (13) after Cholesky decomposition of the estimated residual covariance matrix $\hat{\mathbf{G}}$.

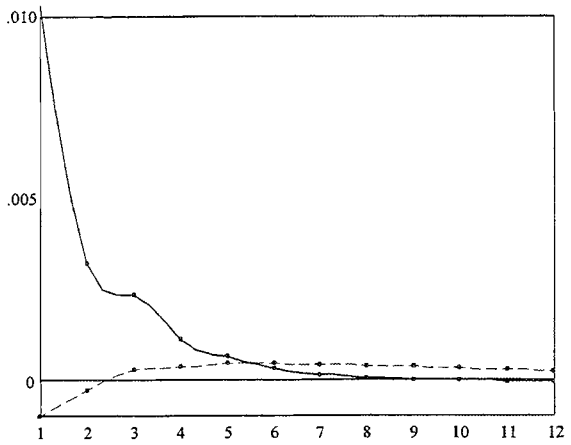

Figure 205: Responses to a shock to m of one standard deviation (.010) in period 1

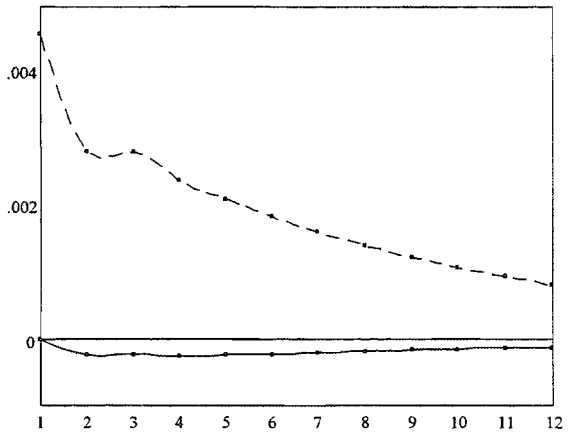

Figure 206: Responses to a shock to p of one standard deviation (.0047) in period 1

Solid lines represent m_t responses; dashed lines p_t responses

Money growth responses to a shock to m_t fade out rather rapidly; from the fifth quarter after the shock they are smaller than the (not very sizeable) inflation responses, but because both series are relative changes, the effects on the underlying variables M1 and CPI can be found in the areas covered by the graphs. These areas being positive for both variables in Figure 205 (except for inflation immediately after the shock to money growth) means that a positive shock to the money growth rate raises M1 as well as prices permanently. In Figure 206, the inflation responses fade out much more slowly than the money growth responses in Figure 205; after three years considerably higher inflation can still be found. Also contrary to Figure 205, an inflation shock aggregates into higher prices but lower money supply since the area under the m_t graph is negative. Recall that impulse-response functions can be criticised for a number of reasons, though; for example they do not represent a realistic situation but rather some sort of average responses to isolated shocks, and they can be very sensitive to the ordering of variables. The effects of ordering p_t over m_t can be found in Figures 207 and 208 below, but apparently this model is not very sensitive to the ordering, since

the responses remain largely unchanged. The most obvious effects are on the initial responses; p_t responds positively from the start in Figure 207 and the first responses of m_t in Figure 208 are clearly larger than when m is ordered over p.

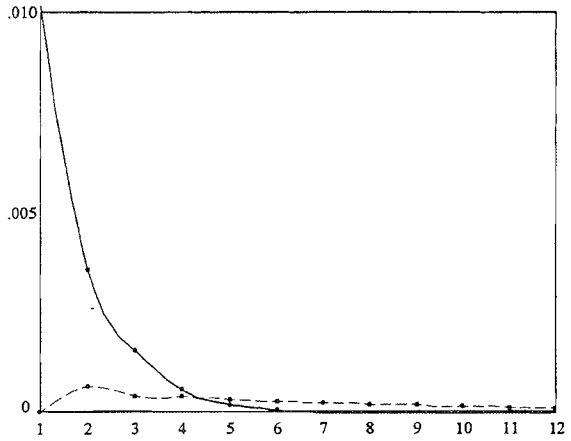

Figure 207: Responses to a shock to m of one standard deviation (.010) in period 1

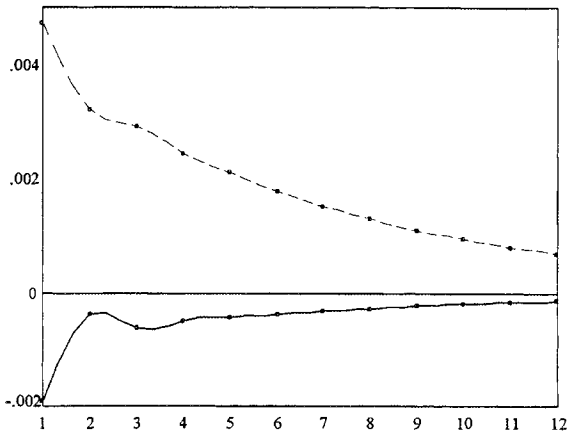

Figure 208: Responses to a shock to p of one standard deviation (.0047) in period 1

Solid lines represent m_t responses; dashed lines p_t responses

In the case of (14), because Granger-causality is ruled by the zero restrictions, impulse-response functions are not very interesting. They will not be zero though; depending on the ordering, shocks to p_t (or m_t) will carry over to m_t (p_t) in the same period through instant causality, and then decay.

Turning to the variance decompositions based on (13), Figure 209 shows how much of the variance of the m_t forecast errors is due to p_t variability and how much of the p_t forecast error variance is due to uncertainty about m_t, in both cases for the first twelve forecasts as well as in the long run (bars left unnumbered).

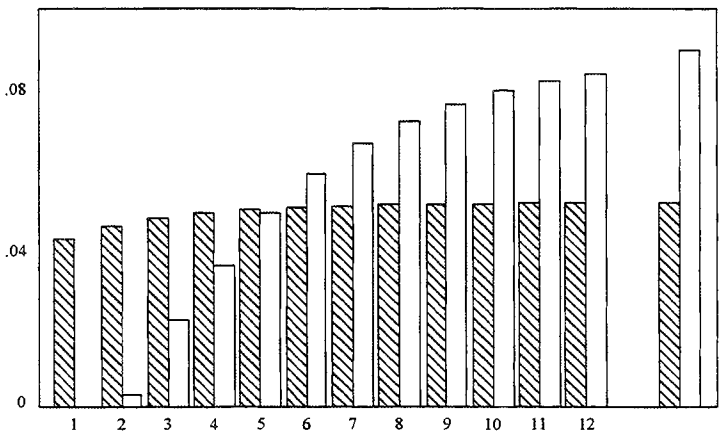

Figure 209: Decomposition of forecast error variances: Filled bars represent inflation's portion of the forecast error variance of m_t; unfilled bars represent money growth's portion of the forecast error variance of p_t

Not shown in Figure 209 are of course the proportion of the forecast error variance of m_t due to m_t and the proportion of the forecast error variance of p_t due to p_t, but since they are simply 100 per cent less the proportions in Figure 209, it is clear that both variables are responsible for more than 90 per cent of their forecast error variances in the short as well as in the long run. Apparently, the autoregressive part dominates the dynamics in (13), which is of course only another aspect of the questionable Granger-causality found previously.

5.4 Forecasts

In this section (13) and (14) will be re-estimated using the first 115 pairs of observations only, leaving the last sixteen observation pairs (1991Q1–1994Q4) to evaluate the generated forecasts. Long-term as well as one-step forecasts will be considered for reasons of comparison.

First, the least-squares estimates corresponding to (13) and (14) are computed for the sample period 1962Q2–1990Q4 and presented in (15) and (16), respectively.

$$\hat{\mu} = \begin{bmatrix} .010 \\ (3.91) \\ .001 \\ (.46) \\ 0 \\ 0 \end{bmatrix} \quad \hat{\mathbf{A}} = \begin{bmatrix} .288 & -.042 & .099 & -.003 \\ (2.96) & (-.21) & (1.02) & (-.02) \\ .045 & .610 & .061 & .233 \\ (.97) & (6.43) & (1.30) & (2.47) \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \hat{\mathbf{G}} = \begin{bmatrix} 9.79 & -.956 \\ -.956 & 2.28 \end{bmatrix} \cdot 10^{-5} \qquad (15)$$

As can be seen, leaving out the last sixteen observations has little effect on most estimates. The estimated parameters in the m_t equation are all reduced though, and particularly in the case of $\hat{a}_{11}$ and $\hat{a}_{13}$ the reductions are sizeable. Adjusted R^2, estimated eigenvalues and residuals are basically the same as for (13). The F statistic for the test of $a_{21}=a_{23}=0$ is now 2.04 (p-value .13), rather like the F=1.88 for (13), and for the test of $a_{12}=a_{14}=0$ the F statistic is even smaller than before, .07 (p-value .93), so the conclusion that p does not Granger-cause m still holds. As before, Granger-causality from m to p is more debatable, the p-value being only .13, still in (16) below the four cross-term zero restrictions have been imposed on (14). Again, $\hat{a}_{11}$ and $\hat{a}_{13}$ are more affected by the shorter sample than $\hat{a}_{22}$ and $\hat{a}_{24}$.

$$\hat{\mu} = \begin{bmatrix} .009 \\ (4.72) \\ .002 \\ (2.69) \\ 0 \\ 0 \end{bmatrix} \quad \hat{A} = \begin{bmatrix} .310 & 0 & .122 & 0 \\ (3.39) & & (1.33) & \\ 0 & .616 & 0 & .216 \\ & (6.84) & & (2.41) \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \hat{G} = \begin{bmatrix} 9.82 & -.993 \\ -.993 & 2.36 \end{bmatrix} \cdot 10^{-5} \qquad (16)$$

In terms of forecasts, it is of course unfortunate that the money growth rate is so poorly modelled. Recall that only about 13 per cent of its variation is explained by the model, which means that no predictor is likely to perform very well without some sort of extra information. But since all competing predictors in this section will face the same challenge, the comparison between them will not be distorted; the forecast errors will simply increase.

In the following table the h-step (h=1,2,3,...,16) money growth predictor based on all information up to 1990Q4 is denoted by $E_T(m_{T+h})$ and the optimal one-step predictor (including all information up to the previous quarter) by $E_{T+h-1}(m_{T+h})$. This predictor is presented in the table in order to assess the performance of $E_T(m_{T+h})$ since it is the optimal predictor given all available information. The values of the two predictors based on (15) are also displayed together with m_t and approximate 95 % prediction intervals in Figure 210.

		(15)		(16)	
	m_t	$E_T(m_{T+h})$	$E_{T+h-1}(m_{T+h})$	$E_T(m_{T+h})$	$E_{T+h-1}(m_{T+h})$
1991:1	.0146	.0123	.0123	.0123	.0123
2	.0215	.0133	.0143	.0131	.0139
3	.0171	.0144	.0175	.0143	.0173
4	.0286	.0148	.0168	.0148	.0167
1992:1	.0437	.0150	.0197	.0150	.0199
2	.0179	.0151	.0263	.0152	.0269
3	.0355	.0152	.0193	.0152	.0198
4	.0350	.0152	.0222	.0153	.0225
1993:1	.0139	.0152	.0247	.0153	.0254
2	.0356	.0152	.0179	.0153	.0180
3	.0260	.0152	.0219	.0153	.0219
4	.0214	.0152	.0224	.0153	.0226
1994:1	.0110	.0152	.0198	.0153	.0196
2	.0052	.0152	.0159	.0153	.0155
3	.0034	.0152	.0128	.0153	.0120
4	-.0027	.0152	.0112	.0153	.0104
MAD = $\sum_{h=1}^{16} \frac{\lvert E(m_{T+h}) - m_t \rvert}{16}$ =		.0113	.0100	.0113	.0098

The table as well as Figure 210 below depict how money growth is shocked in a major way a number of times during the period of interest. Long-term forecasts based on (15) or (16), i.e. $E_T(m_{T+h})$, are unable to predict such events since they stabilise around .015 after about a year. On the other hand, the one-step predictors $E_{T+h-1}(m_{T+h})$ also fail because using information up to and including t–1, they always lag one period behind.

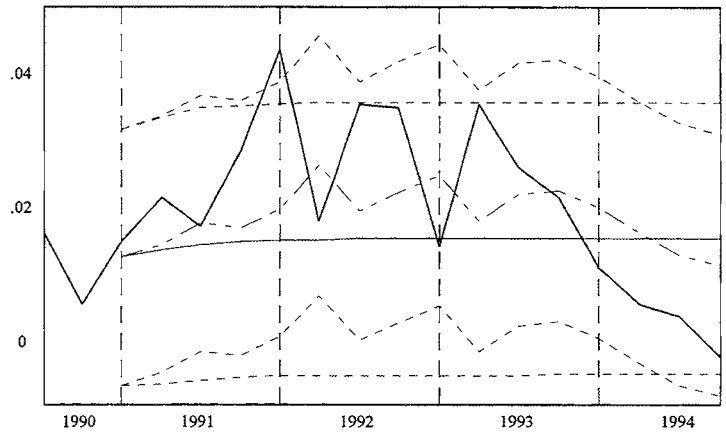

Figure 210: Long-term (solid line) and one-step (broken line) forecasts of m_t with approximate 95 % prediction intervals. Solid bold line represents m_t

This is why the fifth observation of the prediction period (1992Q1) lies clearly above the upper bound of the two prediction intervals. In addition there are two observations (1992Q3, 1993Q2) lying virtually on the upper bound of the more static long-term prediction, but within the bounds of the one-step predictor. But considering the poor fit for m_t in (15) an overall assessment of the predictors, long-term as well as one-step, must be that they are remarkably accurate over these four years. And, as the above table demonstrates, the mean absolute deviation (MAD) is .113 for long-term forecasts based on (15) or (16), only slightly higher than for one-step forecasts (around .10).

Turning to the inflation rate forecasts, the table below holds long-term and one-step forecasts based on (15) and (16). Long-term and one-step forecasts based on (15) are also graphed with p_t and approximate 95 % prediction intervals in Figure 211.

		(15)		(16)	
	p_t	$E_T(p_{T+h})$	$E_{T+h-1}(p_{T+h})$	$E_T(p_{T+h})$	$E_{T+h-1}(p_{T+h})$
1991:1	.0088	.0160	.0160	.0164	.0164
2	.0056	.0152	.0108	.0161	.0114
3	.0063	.0149	.0078	.0158	.0075
4	.0086	.0147	.0078	.0155	.0073
1992:1	.0054	.0146	.0097	.0153	.0090
2	.0085	.0145	.0095	.0151	.0074
3	.0084	.0144	.0104	.0149	.0087
4	.0076	.0143	.0102	.0148	.0092
1993:1	.0075	.0143	.0106	.0147	.0087
2	.0082	.0143	.0094	.0146	.0085
3	.0037	.0142	.0095	.0145	.0089
4	.0067	.0142	.0076	.0144	.0062
1994:1	.0066	.0142	.0076	.0143	.0071
2	.0051	.0142	.0077	.0143	.0077
3	.0102	.0141	.0059	.0142	.0067
4	.0050	.0141	.0084	.0142	.0096
$\text{MAD} = \sum_{h=1}^{16} \frac{\lvert E(p_{T+h}) - p_t \rvert}{16} =$		.0075	.0029	.0079	.0025

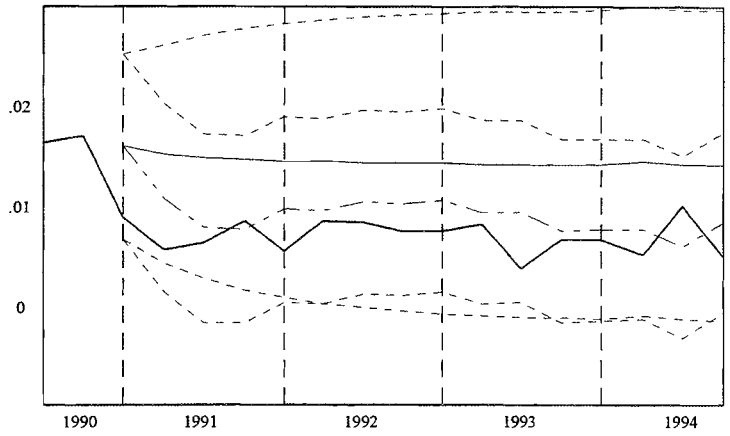

Figure 211: Long-term (solid line) and one-step (broken line) forecasts of p_t with approximate 95 % prediction intervals. Solid bold line represents p_t

It is clearly much easier to predict inflation than money growth during this period. Inflation declined slowly with minor perturbations, and so do the predictions, in the case of $E_T(p_{T+h})$ very slowly though. No extraordinary inflation rates occurred during the period and it was never outside the 95 per cent prediction intervals. The differences between mean absolute deviations of long-term and of one-step forecasts is greater for p_t than for m_t, no doubt due to the shocks to m_t and the more regular shape of p_t, but also because inflation was higher in 1990 than subsequently, leading all four predictors to forecast too high inflation for the first quarters of 1991. Whereas the one-step predictors manage to adjust to the lower inflation almost immediately, the long-term forecasts remained too high throughout the four years. It is also worth noting that the mean absolute deviation of long-term forecasts is smaller for (15) than for (16), whereas the opposite applies to the mean absolute deviation of one-step forecasts.

5.5 Reducing the Bias of $\hat{\mathbf{A}}$

The forecasts in the previous section and the impulse-response functions and variance decompositions in Section 5.3 are based on the biased estimates $\hat{\mathbf{A}}$ and $\hat{\mathbf{G}}$, and as such probably biased estimates of the corresponding "true" forecasts and functions. Moreover, since impulse-response functions and variance decompositions can be extremely sensitive to misspecification as pointed out by Runkle (1987), there is reason to believe they may be sensitive to bias as well. In this section the bias of $\hat{\mathbf{A}}$ will be reduced using the three bias approximations introduced in Chapter 4 after which new impulse-response functions, variance decompositions and forecasts based on the bias-reduced estimates will be computed and compared with their counterparts in the previous section.

The bias approximations used in this section are (9*b*), (10) and (11), all based on estimates. As the regressions in Chapter 4 demonstrate, approximations based on estimates are prone to severely understate the bias, thus when used to reduce the bias here they cannot be expected to do so more than in part. Therefore the resulting bias-reduced estimates will probably not be unbiased, only less biased than before.

In order to approximate the bias of $\hat{\mathbf{A}}$, the least-squares estimates of **A** and **G** were inserted into (9*b*), (10) and (11), together with $\hat{\mathbf{C}}_0$ derived from $\hat{\mathbf{A}}$ and $\hat{\mathbf{G}}$, and $\hat{\rho}$ derived as the solution to the characteristic equation $|\lambda\mathbf{I}-\hat{\mathbf{A}}| = 0$. The resulting bias approximations were then subtracted from $\hat{\mathbf{A}}$, after which the estimated residual variance-covariance matrix $\hat{\mathbf{G}}$ and estimated constant vector $\hat{\mu}$ were computed anew. The results corresponding to (13) based on (9*b*), (10) and (11) can be found in (17), (18) and (19), respectively. (13) is reproduced here as a reference.

$$\hat{\mu} = \begin{bmatrix} .010 \\ (4.40) \\ .001 \\ (.73) \\ 0 \\ 0 \end{bmatrix} \quad \hat{\mathbf{A}} = \begin{bmatrix} .293 & -.054 & .106 & -.013 \\ (3.32) & (-.28) & (1.19) & (-.07) \\ .031 & .603 & .059 & .216 \\ (.77) & (6.93) & (1.46) & (2.50) \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \hat{\mathbf{G}} = \frac{\mathbf{e}_t\mathbf{e}_t'}{T} = \begin{bmatrix} 10.2 & -.962 \\ -.962 & 2.15 \end{bmatrix} \cdot 10^{-5} \qquad (17)$$

$$\hat{\mu} = \begin{bmatrix} .010 \\ (4.40) \\ .001 \\ (.71) \\ 0 \\ 0 \end{bmatrix} \quad \hat{\mathbf{A}} = \begin{bmatrix} .293 & -.054 & .106 & -.013 \\ (3.32) & (-.28) & (1.19) & (-.07) \\ .031 & .604 & .060 & .218 \\ (.78) & (6.94) & (1.46) & (2.52) \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \hat{\mathbf{G}} = \frac{\mathbf{e}_t\mathbf{e}_t'}{T} = \begin{bmatrix} 10.2 & -.963 \\ -.963 & 2.15 \end{bmatrix} \cdot 10^{-5} \qquad (18)$$

$$\hat{\mu} = \begin{bmatrix} .010 \\ (4.42) \\ .001 \\ (.74) \\ 0 \\ 0 \end{bmatrix} \quad \hat{\mathbf{A}} = \begin{bmatrix} .292 & -.054 & .105 & -.014 \\ (3.31) & (-.28) & (1.18) & (-.07) \\ .031 & .603 & .059 & .216 \\ (.77) & (6.93) & (1.46) & (2.50) \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \hat{\mathbf{G}} = \frac{\mathbf{e}_t\mathbf{e}_t'}{T} = \begin{bmatrix} 10.2 & -.962 \\ -.962 & 2.15 \end{bmatrix} \cdot 10^{-5} \qquad (19)$$

$$\hat{\mu} = \begin{bmatrix} .010 \\ (3.81) \\ .000 \\ (.38) \\ 0 \\ 0 \end{bmatrix} \quad \hat{\mathbf{A}} = \begin{bmatrix} .310 & -.050 & .133 & -.003 \\ (3.41) & (-.26) & (1.45) & (-.02) \\ .031 & .617 & .058 & .235 \\ (.74) & (6.91) & (1.38) & (2.65) \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \hat{\mathbf{G}} = \frac{\mathbf{e}_t\mathbf{e}_t'}{T} = \begin{bmatrix} 10.2 & -.963 \\ -.963 & 2.14 \end{bmatrix} \cdot 10^{-5} \qquad (13)$$

Bias-reduced estimates corresponding to (14) based on (9*b*), (10) and (11) can be found in (20), (21) and (22), respectively. (14) is reproduced as a reference.

$$\hat{\mu} = \begin{bmatrix} .009 \\ (3.75) \\ .003 \\ (2.25) \\ 0 \\ 0 \end{bmatrix} \quad \hat{\mathbf{A}} = \begin{bmatrix} .312 & 0 & .131 & 0 \\ (3.54) & & (1.47) & \\ 0 & .603 & 0 & .194 \\ & (6.81) & & (2.21) \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \hat{\mathbf{G}} = \frac{\mathbf{e}_t \mathbf{e}_t'}{T} = \begin{bmatrix} 10.2 & -.976 \\ -.976 & 2.22 \end{bmatrix} \cdot 10^{-5} \tag{20}$$

$$\hat{\mu} = \begin{bmatrix} .009 \\ (3.75) \\ .003 \\ (2.29) \\ 0 \\ 0 \end{bmatrix} \quad \hat{\mathbf{A}} = \begin{bmatrix} .312 & 0 & .131 & 0 \\ (3.54) & & (1.47) & \\ 0 & .602 & 0 & .194 \\ & (6.81) & & (2.21) \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \hat{\mathbf{G}} = \frac{\mathbf{e}_t \mathbf{e}_t'}{T} = \begin{bmatrix} 10.2 & -.976 \\ -.976 & 2.22 \end{bmatrix} \cdot 10^{-5} \tag{21}$$

$$\hat{\mu} = \begin{bmatrix} .008 \\ (3.47) \\ .002 \\ (1.87) \\ 0 \\ 0 \end{bmatrix} \quad \hat{\mathbf{A}} = \begin{bmatrix} .311 & 0 & .130 & 0 \\ (3.53) & & (1.46) & \\ 0 & .601 & 0 & .192 \\ & (6.82) & & (2.19) \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \hat{\mathbf{G}} = \frac{\mathbf{e}_t \mathbf{e}_t'}{T} = \begin{bmatrix} 10.2 & -.976 \\ -.976 & 2.22 \end{bmatrix} \cdot 10^{-5} \tag{22}$$

$$\hat{\mu} = \begin{bmatrix} .008 \\ (4.60) \\ .002 \\ (2.65) \\ 0 \\ 0 \end{bmatrix} \quad \hat{\mathbf{A}} = \begin{bmatrix} .329 & 0 & .157 & 0 \\ (3.84) & & (1.83) & \\ 0 & .619 & 0 & .216 \\ & (7.31) & & (2.57) \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \hat{\mathbf{G}} = \frac{\mathbf{e}_t \mathbf{e}_t'}{T} = \begin{bmatrix} 10.2 & -.995 \\ -.995 & 2.21 \end{bmatrix} \cdot 10^{-5} \tag{14}$$

Apparently, the choice between (9*b*), (10) and (11) is not very crucial since they all lead to very similar bias reductions. This is not surprising since p is only 2 here and the series are relatively long (T=129). Compared to (13) and (14) though, the bias-reduced estimates have undergone major changes. The bias of most parameter estimates appears to be positive since the effect of the bias reduction is to reduce the estimates. Only in the case of $\hat{a}_{21}$ and $\hat{a}_{23}$ does the reduction work in the opposite direction. At this point it may also be of some interest to display the Monte Carlo results for the eigenvalue combination (.5, –.3, .9, –.3), since that is relatively close to the estimated eigenvalue combination of (13), (.56, –.23, .87, –.28). The table below demonstrates how (9*b*), (10) and (11) based on estimates approximate the average bias of the eight estimates in a sample of 100 observations.

Estimate	$\hat{a}_{11}$	$\hat{a}_{12}$	$\hat{a}_{13}$	$\hat{a}_{14}$	$\hat{a}_{21}$	$\hat{a}_{22}$	$\hat{a}_{23}$	$\hat{a}_{24}$
Bias	–.0284	.0043	–.0286	–.0017	–.0072	–.0169	.0072	–.0369
(9*b*) (based	–.0191	.0001	–.0282	–.0007	–.0003	–.0188	.0036	–.0326
(10) on	–.0192	.0001	–.0282	–.0007	–.0003	–.0188	.0036	–.0326
(11) estimates)	–.0218	.0005	–.0307	–.0003	–.0022	–.0191	.0017	–.0327

It should come as no surprise that (9*b*) and (10) perform almost identically in the table above, approximating the bias of $\hat{a}_{13}$ very well, the bias of $\hat{a}_{11}$, $\hat{a}_{22}$ and $\hat{a}_{24}$ less well, the bias of $\hat{a}_{14}$ and $\hat{a}_{23}$ poorly, and the bias of $\hat{a}_{12}$ and $\hat{a}_{21}$ very poorly. The second-order approximation performs better for $\hat{a}_{11}$, $\hat{a}_{12}$ and $\hat{a}_{21}$ but worse for $\hat{a}_{13}$, $\hat{a}_{14}$, $\hat{a}_{22}$ and $\hat{a}_{23}$. This eigenvalue combination is not very characteristic though; as the regressions in Section 4.2 demonstrate, the approximations on average perform well for all parameter estimates except $\hat{a}_{12}$ and $\hat{a}_{21}$.

It is of course interesting to see whether or not reducing the bias has any effects on impulse-response functions and variance decompositions. According to Runkle (1987) both are highly sensitive to specification, and there may be reason to believe they are sensitive to bias as well. On the other hand, since the estimated residual covariance matrix is only marginally affected by these operations and the parameter estimates are not dramatically altered, one would expect similar responses and forecast error variance decompositions as before. In the following figures, only the dynamics implied by (18) will be graphed since (18) was found to be virtually identical with (17) and (19). To facilitate comparisons with Figures 205 and 206, the same ordering and Cholesky decomposition have been used in Figures 212 and 213.

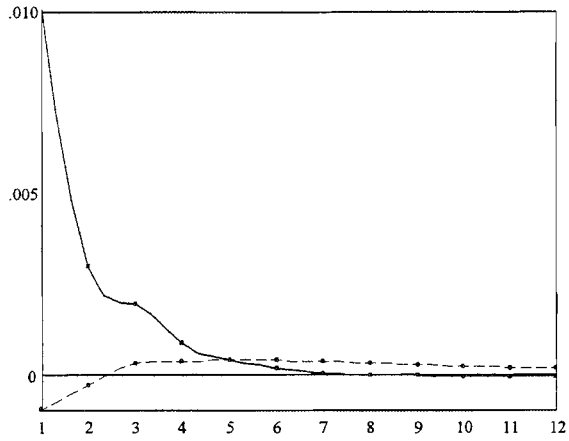

Figure 212: Responses to a shock to *m* of one standard deviation (.010) in period 1; based on (18)

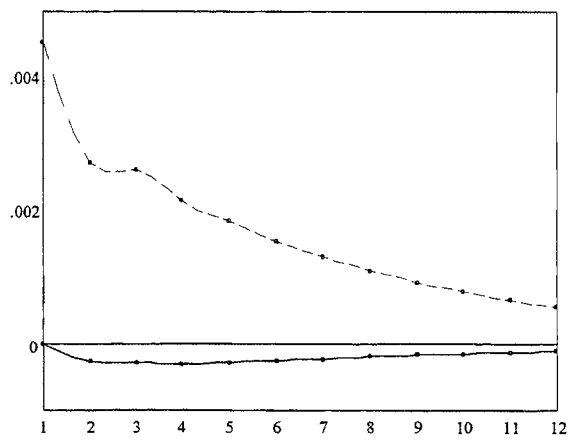

Figure 213: Responses to a shock to *p* of one standard deviation (.0047) in period 1; based on (18)

Solid lines represent m_t responses; dashed lines p_t responses

As Figures 212 and 213 clearly demonstrate, using (10) to reduce the bias of $\hat{\mathbf{A}}$ has essentially no effect on the responses. More or less identical responses obtain, and the same responses as in Figures 207 and 208 also obtain if p_t is ordered over m_t. As for the effects on the decomposition of forecast variances, that of (18) can be found in Figure 214. It too is very similar to Figure 209, but at least one difference can be observed, namely the lower values of the bars in Figure 214 than in Figure 209. It is only natural that such an effect is more easily detected in variance decompositions than

in impulse-response functions, since variance decompositions are simply ratios of sums of squared responses.

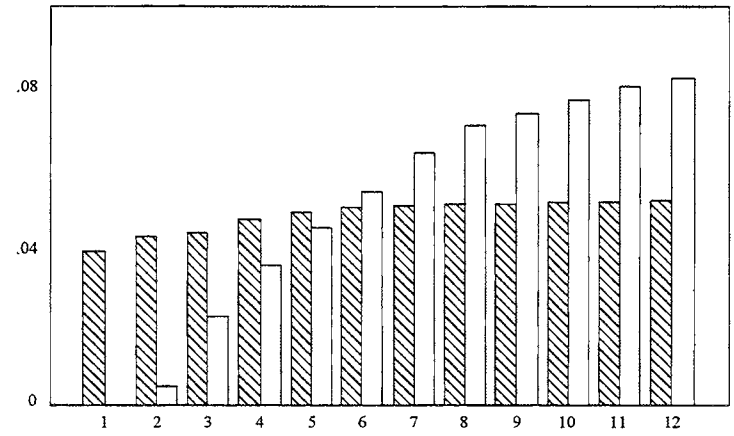

Figure 214: Decomposition of forecast error variances based on (18).
Filled bars represent inflation's portion of the forecast error variance of m_t;
unfilled bars represent money growth's portion of the forecast error variance of p_t

Thus, based on these results at least, there is little reason for going through the trouble to reduce the bias since the implied dynamics are hardly affected at all by such operations. On the other hand, this insensibility may be due to the fact that no strong dynamics were found between money growth and inflation in the first place; stronger results might be obtained in a model with stronger dynamics.

5.6 New Forecasts

In this section forecasts will again be generated for the last sixteen quarters, but this time based on the bias-reduced systems rather than on (15) and (16). Comparisons will then be made with actual money growth and inflation, respectively, as well as with the predictions in Section 5.4.

First of all (9*b*), (10) and (11) are used to compute bias-reduced estimates over the first 115 quarters (1962Q2–1990Q4) as alternatives to (15) and (16). Only the bias-reduced estimates based on (10) are reported in (23) below though; like in the previous section, the differences between the three bias reductions are negligible. As a reference, (15) is also reproduced and as before, t ratios are presented in parentheses.

$$\hat{\mu} = \begin{bmatrix} .011 \\ (4.47) \\ .001 \\ (.79) \end{bmatrix} \quad \hat{\mathbf{A}} = \begin{bmatrix} .269 & -.046 & .070 & -.015 \\ (2.85) & (-.24) & (.74) & (-.08) \\ .046 & .593 & .063 & .210 \\ (1.00) & (6.37) & (1.39) & (2.28) \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \hat{\mathbf{G}} = \begin{bmatrix} 9.81 & -.954 \\ -.954 & 2.29 \end{bmatrix} \cdot 10^{-5} \qquad (23)$$

$$\hat{\mu} = \begin{bmatrix} .010 \\ (3.91) \\ .001 \\ (.46) \\ 0 \\ 0 \end{bmatrix} \quad \hat{\mathbf{A}} = \begin{bmatrix} .288 & -.042 & .099 & -.003 \\ (2.96) & (-.21) & (1.02) & (-.02) \\ .045 & .610 & .061 & .233 \\ (.97) & (6.43) & (1.30) & (2.47) \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \hat{\mathbf{G}} = \begin{bmatrix} 9.79 & -.956 \\ -.956 & 2.28 \end{bmatrix} \cdot 10^{-5} \qquad (15)$$

Like for the full-sample estimates in the previous section, bias reduction affects all parameter estimates except $\hat{a}_{21}$ and $\hat{a}_{23}$ negatively, and again the effect is greater for the four autoregressive parameter estimates than for the cross-term estimates.

It is perhaps worth emphasising that it is not reasonable to expect the bias-reduced systems to predict the shocks to m_t any better than (15) does; they may in fact even perform worse if the shocks are thought of as external to the system and thus unpredictable. For more "normal" observations, however, less biased may be expected to perform better than more biased predictors.

In the following table, long-term forecasts $E_T(m_{T+h})$ based on all information up to 1990Q4 are presented with the actual money growth rate m_t. Forecasts have been generated by the original model (15) and by three bias-reduced models based on (9*b*), (10) and (11), respectively. Thus the column headings (9*b*), (10) and (11) do not mean that the forecasts below have been generated by (9*b*), (10) and (11), respectively, but by bias-reduced versions of (15), such as (23) above, using these three bias approximations to reduce the bias.

	m_t	(15)	(9*b*)	(10)	(11)
1991:1	.0146	.0123	.0124	.0124	.0124
2	.0215	.0133	.0136	.0136	.0136
3	.0171	.0144	.0145	.0145	.0145
4	.0286	.0148	.0149	.0149	.0148
1992:1	.0437	.0150	.0150	.0150	.0150
2	.0179	.0151	.0151	.0151	.0151
3	.0355	.0152	.0152	.0152	.0151
4	.0350	.0152	.0152	.0152	.0152
1993:1	.0139	.0152	.0152	.0152	.0152
2	.0356	.0152	.0152	.0152	.0152
3	.0260	.0152	.0152	.0152	.0152
4	.0214	.0152	.0152	.0152	.0152
1994:1	.0110	.0152	.0152	.0152	.0152
2	.0052	.0152	.0153	.0153	.0152
3	.0034	.0152	.0153	.0153	.0152
4	-.0027	.0152	.0153	.0153	.0152
MAD = $\sum_{h=1}^{16} \frac{\lvert E(m_{T+h}) - m_t \rvert}{16}$ =		.0113	.0113	.0113	.0113

Apparently, reducing the bias of (15) has almost no effect at all on predictions. The three bias-reduced models predict money growth more or less identically as (15) does; around .015 from the fourth quarter on. In view of the fluctuations money growth undergoes, all four predictors are very static and not at all impressive. Mean absolute

deviations (MAD) are also the same for all four predictors, thus the long-term money growth forecasts provide no argument in favour of bias reduction.

The picture looks considerably brighter when inflation predictions are compared, though. Probably due to the lack of extraordinary p_t values, reducing the bias also reduces the mean absolute deviation by some 15 per cent in the table below. The smaller coefficients of the bias-reduced systems yield smaller p_t predictions than (15), which turns out to be beneficial here since the lower inflation rates in 1991 than in 1990 cause each long-term predictor to overshoot the inflation rate during the last four years. Of course, $E_T(p_{T+h})$ based on biased-reduced models is also too high, only not quite as high as (15). But as Figure 215 also shows, the different predictors differ by only a fraction of the forecast errors, so even though bias reduction does improve long-term inflation forecasts, this model is still not very useful for predicting inflation.

	p_t	(15)	(9*b*)	(10)	(11)
1991:1	.0088	.0160	.0157	.0157	.0157
2	.0056	.0152	.0147	.0147	.0146
3	.0063	.0149	.0144	.0144	.0143
4	.0086	.0147	.0141	.0141	.0139
1992:1	.0054	.0146	.0139	.0139	.0137
2	.0085	.0145	.0137	.0137	.0136
3	.0084	.0144	.0136	.0136	.0134
4	.0076	.0143	.0135	.0135	.0133
1993:1	.0075	.0143	.0135	.0135	.0133
2	.0082	.0143	.0134	.0134	.0132
3	.0037	.0142	.0134	.0134	.0131
4	.0067	.0142	.0133	.0133	.0131
1994:1	.0066	.0142	.0133	.0133	.0131
2	.0051	.0142	.0133	.0133	.0130
3	.0102	.0141	.0132	.0132	.0130
4	.0050	.0141	.0132	.0132	.0130
MAD = $\sum_{h=1}^{16} \frac{\lvert E(p_{T+h}) - p_t \rvert}{16}$ =		.0075	.0067	.0067	.0066

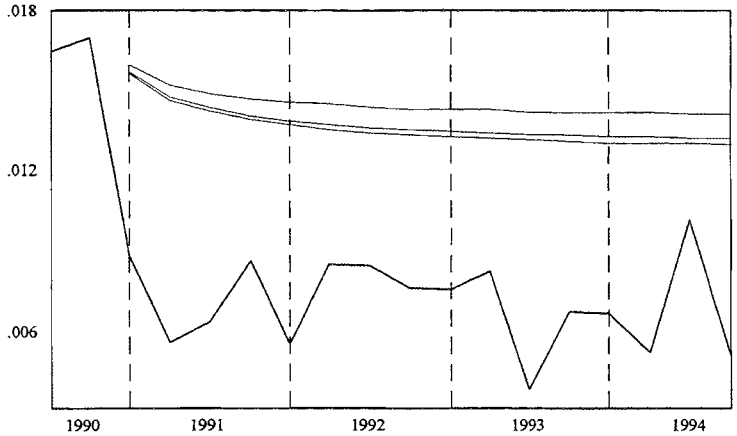

Figure 215: Quarterly inflation (bold line) with long-term forecasts; original (15) (top); bias-reduced using (10) (middle); bias-reduced using (11) (bottom). All observations are inside the 95 % prediction intervals (cf. Figure 211)

Next, the long-term forecast generated by bias-reduced versions of the seemingly unrelated regressions (16) will be evaluated. Using (16) to compute the three approximations (9*b*), (10) and (11) over the first 115 quarters (1962Q2–1990Q4) gives the following bias-reduced estimates (again, (10) is taken to represent all three bias-reduced models in (24), since very similar results obtain for the other two):

$$\hat{\boldsymbol{\mu}} = \begin{bmatrix} .009 \\ (3.80) \\ .003 \\ (2.34) \\ 0 \\ 0 \end{bmatrix} \quad \hat{\mathbf{A}} = \begin{bmatrix} .292 & 0 & .093 & 0 \\ (3.09) & & (.99) & \\ 0 & .597 & 0 & .190 \\ & (6.30) & & (2.03) \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \hat{\mathbf{G}} = \begin{bmatrix} 9.81 & -.966 \\ -.966 & 2.38 \end{bmatrix} \cdot 10^{-5} \tag{24}$$

$$\hat{\boldsymbol{\mu}} = \begin{bmatrix} .009 \\ (4.72) \\ .002 \\ (2.69) \\ 0 \\ 0 \end{bmatrix} \quad \hat{\mathbf{A}} = \begin{bmatrix} .310 & 0 & .122 & 0 \\ (3.39) & & (1.33) & \\ 0 & .616 & 0 & .216 \\ & (6.84) & & (2.41) \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \hat{\mathbf{G}} = \begin{bmatrix} 9.82 & -.993 \\ -.993 & 2.36 \end{bmatrix} \cdot 10^{-5} \tag{16}$$

Like for the full-sample estimates in Section 5.5, all four estimates are smaller in (24) than in (16), which may indicate a positive biased $\hat{\mathbf{A}}$ in (16). The implications for the long-term forecasts $E_T(m_{T+h})$ of using (9*b*), (10) or (11) to reduce the bias of (16) are illustrated in the following table. Predictions based on the biased-reduced versions of (16) can be found under (9*b*), (10) and (11), depending on which bias approximation was used. Like in the case of (23) versus (15), reducing the bias has essentially no effect on predictions; they all stabilise above .015 rather rapidly and fail completely to predict the shifting growth rate of money. Also parallel to the previous case, the mean absolute deviation is .0113 for all four predictors.

	m_t	(16)	(9*b*)	(10)	(11)
1991:1	.0146	.0123	.0123	.0123	.0124
2	.0215	.0131	.0134	.0134	.0135
3	.0171	.0143	.0143	.0143	.0145
4	.0286	.0148	.0147	.0147	.0149
1992:1	.0437	.0150	.0149	.0149	.0151
2	.0179	.0152	.0150	.0150	.0152
3	.0355	.0152	.0150	.0150	.0152
4	.0350	.0153	.0151	.0151	.0152
1993:1	.0139	.0153	.0151	.0151	.0153
2	.0356	.0153	.0151	.0151	.0153
3	.0260	.0153	.0151	.0151	.0153
4	.0214	.0153	.0151	.0151	.0153
1994:1	.0110	.0153	.0151	.0151	.0153
2	.0052	.0153	.0151	.0151	.0153
3	.0034	.0153	.0151	.0151	.0153
4	-.0027	.0153	.0151	.0151	.0153
MAD = $\sum_{h=1}^{16} \frac{\lvert E(m_{T+h}) - m_t \rvert}{16}$ =		.0113	.0113	.0113	.0113

More interesting are the results for $E_T(p_{T+h})$ generated by (16), shown in the table below. Like in the case of the unrestricted model (15), smaller bias-reduced parameter estimates lead to lower predicted inflation rates, and as a consequence the bias-reduced predictors come closer to p_t than (16) does. The effect appears to be smaller for (16) than for (15) though, since MAD is reduced by a mere 10 per cent here, whereas the reduction was close to 15 per cent in the case of (15). As Figure 216 shows, the difference between the bias reduction based on (10) and the bias reduction based on (11) is also smaller here than for (15). And again, it deserves pointing out that the improvement of forecasts is small relative to the remaining prediction errors.

	p_t	(16)	(9*b*)	(10)	(11)
1991:1	.0088	.0164	.0161	.0161	.0161
2	.0056	.0161	.0157	.0157	.0156
3	.0063	.0158	.0152	.0152	.0152
4	.0086	.0155	.0149	.0149	.0148
1992:1	.0054	.0153	.0146	.0146	.0145
2	.0085	.0151	.0144	.0144	.0143
3	.0084	.0149	.0142	.0142	.0141
4	.0076	.0148	.0140	.0140	.0139
1993:1	.0075	.0147	.0139	.0139	.0138
2	.0082	.0146	.0138	.0138	.0137
3	.0037	.0145	.0137	.0137	.0136
4	.0067	.0144	.0136	.0136	.0135
1994:1	.0066	.0143	.0135	.0135	.0134
2	.0051	.0143	.0135	.0135	.0134
3	.0102	.0142	.0134	.0134	.0133
4	.0050	.0142	.0134	.0134	.0133
MAD = $\sum_{h=1}^{16} \frac{\lvert E(p_{T+h})-p_t \rvert}{16}$ =		.0079	.0072	.0072	.0071

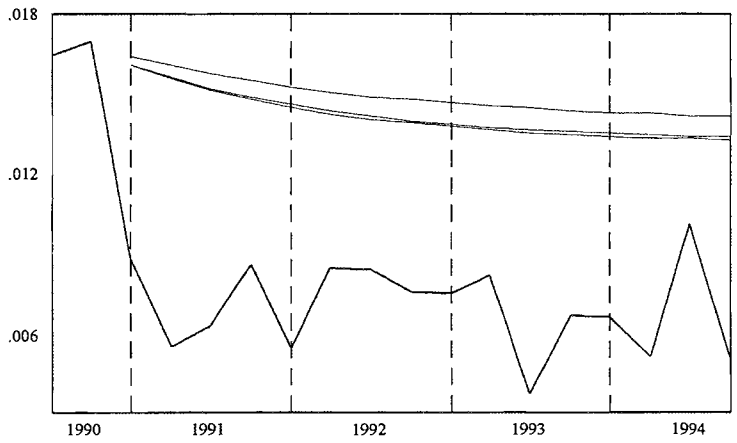

Figure 216: Quarterly inflation (bold line) with long-term forecasts; original (16) (top); bias-reduced using (10) (middle); bias-reduced using (11) (bottom) All observations are inside the 95 % prediction intervals (cf. Figure 211)

Computing bias-reduced one-step forecasts is rather a tedious occupation, since $\hat{\mathbf{A}}$ must first be estimated over T observations, then its bias is reduced using (9*b*), (10) and (11), after which predictions can be made for T+1. The same procedure is then repeated for observations up to T+1, T+2, etc., so that for this particular application the cycle is repeated sixteen times. Moreover, it is reasonable to expect the effects of bias reduction to be smaller for one-step forecasts than for longer-term forecasts since in the latter case, small initial effects can aggregate into more substantial differences over longer prediction periods. One-step forecasts, on the other hand, are constantly fed with lagged values of actual money growth and inflation rates, and therefore such aggregation of effects is impossible. All effects (if any) of bias reduction will be temporary only.

	m_t	(15)	(9*b*)	(10)	(11)
1991:1	.0146	.0123	.0124	.0124	.0124
2	.0215	.0143	.0145	.0145	.0145
3	.0171	.0175	.0174	.0174	.0174
4	.0286	.0168	.0166	.0166	.0166
1992:1	.0437	.0197	.0194	.0194	.0194
2	.0179	.0263	.0253	.0253	.0253
3	.0355	.0193	.0185	.0185	.0185
4	.0350	.0222	.0217	.0217	.0217
1993:1	.0139	.0247	.0238	.0238	.0237
2	.0356	.0179	.0174	.0174	.0174
3	.0260	.0219	.0216	.0215	.0215
4	.0214	.0224	.0217	.0217	.0217
1994:1	.0110	.0198	.0195	.0195	.0195
2	.0052	.0159	.0159	.0159	.0159
3	.0034	.0128	.0132	.0132	.0132
4	−.0027	.0112	.0118	.0118	.0118
MAD = $\sum_{h=1}^{16} \frac{\lvert E(m_{T+h})-m_t \rvert}{16}$ =		.0100	.0100	.0100	.0100

The above table holds one-step forecasts $E_{T+h-1}(m_{T+h})$ based on the unrestricted model (15) and its bias-reduced counterparts based on (9*b*), (10) and (11), respectively. Bias reduction actually appears to have a real effect on individual predictions since they differ quite a lot from the predictions generated by (15), but in aggregate it is nonetheless difficult to argue in favour of any of the three bias-reduced models, seeing that their predictions sum up to the same mean absolute deviation as the predictions generated by (15). Again, no support for bias reduction can be found.

Next, the same comparison is made between the restricted model (16) and its three bias-reduced versions using (9*b*), (10) and (11) to reduce the bias. Here as well, reducing the bias is seen in the following table to affect most one-step predictions, sometimes bringing them closer to p_t and sometimes removing them from p_t, but in aggregate all these effects sum up to zero, or rather somewhat greater than zero, since mean absolute deviations are actually higher than for (16).

	m_t	(16)	(9b)	(10)	(11)
1991:1	.0146	.0123	.0123	.0123	.0124
2	.0215	.0139	.0140	.0140	.0140
3	.0171	.0173	.0169	.0169	.0169
4	.0286	.0167	.0163	.0163	.0163
1992:1	.0437	.0199	.0194	.0194	.0193
2	.0179	.0269	.0257	.0257	.0257
3	.0355	.0198	.0188	.0188	.0188
4	.0350	.0225	.0218	.0218	.0218
1993:1	.0139	.0254	.0242	.0242	.0242
2	.0356	.0180	.0173	.0173	.0173
3	.0260	.0219	.0214	.0214	.0213
4	.0214	.0226	.0216	.0216	.0216
1994:1	.0110	.0196	.0190	.0190	.0190
2	.0052	.0155	.0153	.0153	.0153
3	.0034	.0120	.0121	.0121	.0121
4	−.0027	.0104	.0108	.0108	.0108
$\text{MAD} = \sum_{h=1}^{16} \frac{\lvert E(m_{T+h}) - m_t \rvert}{16} =$		.0098	.0099	.0099	.0099

Before turning to the one-step forecasts of inflation, Figure 217 depicts eight predictors of money growth: Long-term forecasts $E_T(m_{T+h})$ generated by the original full model (15), the original restricted model (16), the bias-reduced full model based on (15) and (10), and the bias-reduced restricted model based on (16) and (10); one-step forecasts $E_{T+h-1}(m_{T+h})$ generated by the original full model, the original restricted model, the bias-reduced full model using (10), and the bias-reduced restricted model using (10). Like in preceding tables, the predictions generated by bias-reduced models differ only marginally from those generated by the original models. That the differences are much smaller for long-term than for one-step forecasts is surprising; one would expect small differences to accumulate into more sizeable effects by the end of the four-year period, but the four long-term predictors all rapidly stabilise around .015 and stay there. One-step forecasts, on the other hand, react to the volatile growth rate of money, shifting the quarter after m_t shifts, and the four predictors appear to diverge more the more abruptly the money growth changes.

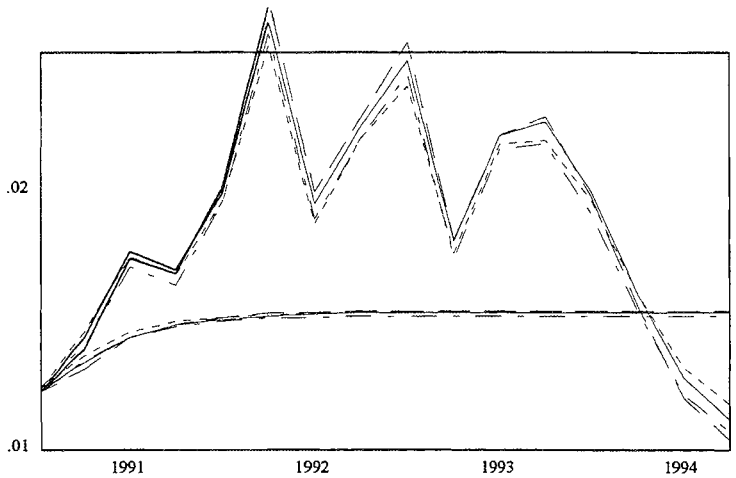

Figure 217: Long-term (smooth) and one-step (oscillating) forecasts of m_t based on (15) (solid); (16) (long dashes); (15) adjusted by (10) (short dashes); and (16) adjusted by (10) (broken) Money growth (not shown) is inside the 95 % prediction intervals except in 1992Q1

As for the one-step inflation forecasts $E_{T+h-1}(p_{T+h})$, the results of using (9*b*), (10) and (11) to reduce the bias of the full original model can be found in the following table. Recalling that the only positive effects of bias reduction was on the long-term inflation predictions generated by (15) and (16), one might expect to find a similar improvement in the short run, but that is not supported by the results in the table. As in the case of one-step money growth predictions, reducing the bias affects most predicted values (albeit due to scale effects not as dramatically as in the case of money growth), occasionally improving them, but most of the time making them worse by raising the original predictions somewhat. On the whole, the effect of using (9*b*), (10) or (11) to reduce the bias therefore appears to be negative, since these predictors have slightly higher mean absolute deviations than the original one-step predictor.

	p_t	(15)	(9*b*)	(10)	(11)
1991:1	.0088	.0160	.0157	.0157	.0157
2	.0056	.0108	.0107	.0107	.0107
3	.0063	.0078	.0080	.0080	.0080
4	.0086	.0078	.0080	.0080	.0080
1992:1	.0054	.0097	.0098	.0098	.0098
2	.0085	.0095	.0097	.0097	.0097
3	.0084	.0104	.0106	.0106	.0107
4	.0076	.0102	.0103	.0103	.0103
1993:1	.0075	.0106	.0108	.0108	.0108
2	.0082	.0094	.0096	.0096	.0096
3	.0037	.0095	.0096	.0096	.0096
4	.0067	.0076	.0078	.0078	.0078
1994:1	.0066	.0076	.0079	.0079	.0079
2	.0051	.0077	.0079	.0079	.0079
3	.0102	.0059	.0061	.0061	.0061
4	.0050	.0084	.0085	.0085	.0086
MAD = $\sum_{h=1}^{16} \frac{\lvert E(p_{T+h}) - p_t \rvert}{16}$ =		.0029	.0030	.0030	.0030

Similar results obtain if (9*b*), (10) and (11) are used to reduce the bias of the restricted model, i.e. of the two SUR equations. The bias-reduced models generate slightly higher one-step forecasts than the original model, which in most cases is tantamount to distancing them from the actual inflation rate. Consequently, the mean absolute deviations of three bias-reduced predictors are slightly in excess of that of the original predictor. Their MADs are still smaller than those in the above table, though, since the seemingly unrelated regressions model predicts inflation better than the unrestricted model in the short run.

	p_t	(16)	(9b)	(10)	(11)
1991:1	.0088	.0164	.0161	.0161	.0161
2	.0056	.0114	.0113	.0113	.0113
3	.0063	.0075	.0078	.0078	.0079
4	.0086	.0073	.0076	.0076	.0077
1992:1	.0054	.0090	.0092	.0092	.0092
2	.0085	.0074	.0076	.0076	.0076
3	.0084	.0087	.0089	.0089	.0089
4	.0076	.0092	.0094	.0094	.0094
1993:1	.0075	.0087	.0089	.0089	.0089
2	.0082	.0085	.0087	.0087	.0087
3	.0037	.0089	.0091	.0091	.0091
4	.0067	.0062	.0064	.0064	.0064
1994:1	.0066	.0071	.0074	.0074	.0074
2	.0051	.0077	.0079	.0079	.0079
3	.0102	.0067	.0069	.0069	.0069
4	.0050	.0096	.0097	.0097	.0097
MAD = $\sum_{h=1}^{16} \frac{\lvert E(p_{T+h}) - p_t \rvert}{16}$ =		.0025	.0026	.0026	.0026

Finally, the same eight-predictor comparison as for money growth is presented in Figure 218 below. For a number of reasons (clarity, for one) it is more interesting than Figure 217. First of all, the superiority of the one-step over the long-term predictors is even more evident here than in Figure 211. Furthermore, the differences between one-step forecasts based on the full model (solid and broken lines) and those based on the seemingly unrelated regressions model (dashed lines) is evident, the latter two closer to the rate of inflation most of the time. Thirdly, bias reduction has very little effect on one-step forecasts; the lines move closely together in pairs, the bias-reduced predictors (broken line and short dashes, respectively) most of time higher than the corresponding original predictor. Lastly, the effects of bias reduction are greater for long-term forecasts, which diverge (solid and broken line; dashed lines from each other) as expected. But then again, none of the four long-term predictors is very impressive.

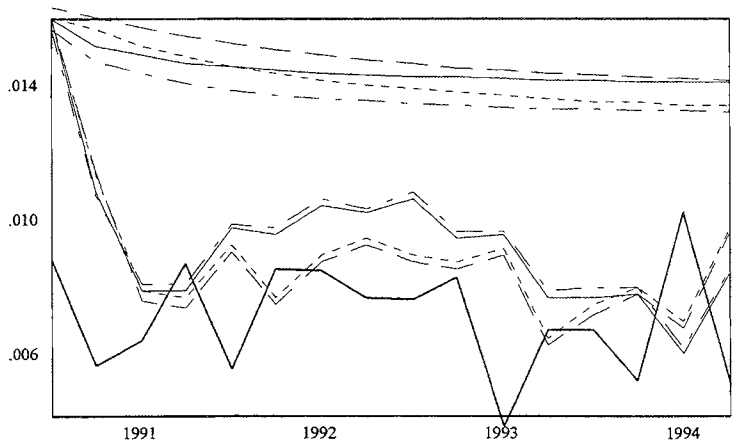

Figure 218: Long-term (top) and one-step (bottom) forecasts of p_t based on (15) (solid); (16) (long dashes); (15) adjusted by (10) (broken); and (16) adjusted by (10) (short dashes) Bold line, representing p_t, is inside the 95 % prediction intervals at all times

6. Concluding Remarks

This thesis, and the papers it is based on, have reached a number of important and potentially useful conclusions. On the most fundamental level, it can be concluded that the bias of parameter estimates will normally be a problem in estimated VAR models since it will often be huge both in absolute terms and relative to standard errors. This is particularly true when eigenvalues approach unity, i.e. in "near-integrated" and (co-) integrated and situations. While it should come as a surprise to nobody that estimated VAR parameters are biased, the size of that bias in these particular Monte Carlo experiments may astound some, especially since there is every reason to expect these results to hold more generally than for the two bivariate models simulated here.

There are two strategies to deal with this bias, the more popular being to ignore it. This thesis represents an argument in favour of the opposite approach of actively attempting to reduce the bias, although not necessarily by the methods evaluated here. As pointed out earlier, Bayesian methods, ridge regression, resampling methods or other bias approximations than the three presented here may in fact serve equally well or better. A principal conclusion of this thesis must be that doing something to reduce the bias is always better than doing nothing, though.

The extensive simulations were also used to evaluate the three bias approximations introduced in Chapter 4. For the bivariate first-order model, all three were found to approximate the bias well when based on true parameters but understated it when based on estimates. For the second-order bivariate model the approximations underachieved when based on true parameters and even more so when based on estimates, but since they tended to approximate in the right direction they can be expected to provide at least a partial bias reduction when deducted from the original estimates. Furthermore, the approximations had increasing difficulties dealing with the bias as roots approached the unit circle and finally, they performed better in larger than in smaller samples (which is not surprising considering that they were derived from asymptotic expansions).

As far as the performance of the three approximations relative to each other is concerned, it is very difficult to rank Tjøstheim-Paulsen's approximation ahead of Nicholls-Pope's or vice versa, their performances being identical for first-order models and almost identical for second-order models. However, since there is every reason to expect them to diverge as the order increases, determining which of the two is the best approximation in systems of much higher order than 2 remains an open and interesting question. When it comes to the modified approximation (including second-order terms), it seems to perform at least as well as the two first-order approximations and in many cases considerably better. More precisely, the second-order terms remain more or less inactive as long as they are not needed, i.e. as long as the first-order terms approximate the bias well, but step in to play an increasingly important role as eigenvalues approach unity. The second-order approximation is far from perfect though; it too will typically understate the bias in many situations, only less so than the two first-order approximations. In addition, the second-order approximation achieved better – but still far from good – results for $\hat{a}_{12}$ and $\hat{a}_{21}$, where the first-order approximations failed completely. Thus the choice between the three approximations would appear to be a simple one, except for the fact that the second-order approximation is more sensitive to near-unit roots and is more likely to break down in their presence. Therefore analysts may have to resort to first-order approximations in the very cases when a higher-order approximation would be really useful.

A number of other results have obtained as by-products along the way. They have already been commented upon in previous chapters, but deserve repeating here. First of all, whereas first-order integration was not found to add more bias than could be expected, second-order integration (i.e. two or more unit roots in the second-order model) with or without cointegrating vectors was found to cause extra bias, before as well as after bias reduction. Secondly, estimating a model with cross terms when in fact no cross terms enter appears to increase the bias and the variance of the autoregressive estimates. Estimating a model of too high an order, while severely increasing the variance of parameter estimates, may or may not increase their bias. Thus using a higher value for the order p than that suggested by data may in some situations be an alternative way to reduce bias than the approximations presented here.

Unfortunately, the bivariate application in Chapter 5 turned out to provide very little support for the idea of bias reduction. Even though the sample was not very large and the estimated eigenvalue vector seemed interesting, the effects of bias reduction were relatively small on parameter estimates and smaller still on

impulse-response functions, variance decompositions and predictions. Nonetheless, these results do not justify the conclusion that bias reduction does not matter and need never be applied. While this may be the case for this particular data set, the reason may very well be that there was little bias to reduce in the originally estimated VAR model. It is also conceivable that the very weak causality between the two variables makes bias reduction less useful.

Apart from the need to analyse biases and compare bias approximations in systems of higher order than 2 as pointed out earlier, it would also be interesting to increase the dimension K of the model, and to allow for exogenous variables. Another highly interesting research field outlined in this thesis is the possible consequences of iterating the bias-reduction process until some sort of convergence criteria are met. Questions to answer in this field include: Which criteria should be used? Under which circumstances will this process converge? Is convergence to something else than increasingly less biased estimates likely? Finally, as an alternative to using the first- or second-order approximations evaluated here, it should be possible to expand the approximation further, adding terms of higher order than 2. The fact that even a second-order approximation tends to reduce the bias only in part indicates the need for higher-order approximations.

References

Balke, S. and B. Fomby (1991) "Shifting Trends, Segmented Trends and Infrequent Permanent Shocks", *Journal of Monetary Economics*, **28**, 61-85

Boero, G. (1990) "Comparing *Ex-ante* Forecasts from a SEM and VAR Model: an Application to the Italian Economy", *Journal of Forecasting*, **9**, 13-24

Brännström, T. (1994*a*) "Reducing the Bias of OLS Estimates in Bivariate VAR Models", EFI Research Paper 6533. Stockholm: EFI

——— (1994*b*) "A Comparison Between Bias Approximations Applied to Bivariate VAR Models", Working Paper No. 22. Stockholm: Stockholm School of Economics Working Paper Series in Economics and Finance

——— (1995*a*) "A Second-Order Approximation to the Bias of OLS Estimates in Bivariate VAR Systems", Working Paper No. 81. Stockholm: Stockholm School of Economics Working Paper Series in Economics and Finance

——— (1995*b*) "Money Growth and Inflation: Implications of Reducing the Bias of VAR Estimates", Working Paper No. 82. Stockholm: Stockholm School of Economics Working Paper Series in Economics and Finance

——— and S. Karlsson (1993) "Least Squares Estimates in Bivariate VAR Systems", EFI Research Paper 6512. Stockholm: EFI

Cooley, T.F. and S.F. LeRoy (1985) "Atheoretical Macroeconometrics: A Critique", *Journal of Monetary Economics*, **16**, 283-308

Doan, T.A., R.B. Litterman and C.A. Sims (1984) "Forecasting and Conditional Projections using Realistic Prior Distributions" with discussion, *Econometric Reviews*, **3**, 1-100

Edlund, P-O. and S. Karlsson (1993) "Forecasting the Swedish Unemployment Rate. VAR vs. Transfer Function Modelling", *International Journal of Forecasting*, **9**, 61-76

Engle, R.F. and C.W.J. Granger (1987) "Co-Integration and Error Correction: Representation, Estimation and Testing", *Econometrica*, **55**, 251-276

Engle, R.F. and B.S. Yoo (1987) "Forecasting and Testing in Co-integrated Systems", *Journal of Econometrics*, **35**, 143-159

Epstein, R.J. (1987) *A History of Econometrics*. Amsterdam: North-Holland

Fuller, W.A. (1976) *Introduction to Statistical Time Series*. New York: Wiley

Granger, C.W.J. and P. Newbold (1974) "Spurious Regressions in Econometrics", *Journal of Econometrics*, **2**, 111-120

——— (1986) *Forecasting Economic Time Series*. San Diego: Academic Press

——— and A.A. Weiss (1983) "Time Series Analysis of Error-Correction Models", in *Studies in Econometrics, Time Series and Multivariate Statistics* (ed. S. Karlin), 255-278. New York: Academic Press

Hallman, J.J., R.D. Porter and D.H. Small (1991) "Is the Price Level Tied to the M2 Monetary Aggregate in the Long Run?", *American Economic Review* **81**, 841-858

Holden, K. (1995) "Vector Autoregression Modelling and Forecasting", *Journal of Forecasting* **14**, 159-166

Hsiao, C. (1979) "Autoregressive Modeling of Canadian Money and Income Data", *JASA* **74**, 553-560

Hurwicz, L. (1950) "Least-Squares Bias in Time Series", in *Statistic Inference in Dynamic Economic Models* (ed. T.C. Koopmans), 225-230. New York: Wiley

Johansen, S. (1991) "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models", *Econometrica* **51**, 1551-1580

Keating, J.W. (1990) "Identifying VAR Models under Rational Expectations", *Journal of Monetary Economics* **25**, 435-476

———— (1992) "Structural Approaches to Vector Autoregressions", *Federal Reserve Bank of St Louis Review* **74**, 37-57

Kendall, M.G. (1954) "Note on the Bias in the Estimation of Autocorrelation", *Biometrika* **41**, 403-404

Kiviet, J.F. and G.D.A. Phillips (1991) "Bias Reduction in Constrained and Transformed Autoregressive Distributed Lag Models", paper presented at the $(EC)^2$ conference, Rotterdam, December 1991

———— (1993) "Alternative Bias Approximations in Regressions with a Lagged-Dependent Variable", *Econometric Theory* **9**, 62-80

———— (1995) "Higher-Order Asymptotic Bias in First-Order Dynamic Regression Models", paper presented at the 7th Econometric Society Conference, Beijing, September 1995

Kohn, R. (1979) "Asymptotic Estimation and Hypothesis Testing Results for Vector Linear Time Series Models", *Econometrica* **47**, 1005-1030

Koopmans, T. (1947) unpubl. draft; in Epstein, R.J. (1987) *A History of Econometrics*

LeSage, J.P. (1990) "A Comparison of the Forecasting Ability of ECM and VAR Models", *Review of Economics and Statistics* **72**, 664-671

Litterman, R.B. (1986*a*) "A Statistical Approach to Economic Forecasting", *Journal of Business & Economic Statistics* **4**, 1-4

———— (1986*b*) "Forecasting with Bayesian Vector Autoregressions – Five Years of Experience", *Journal of Business & Economic Statistics* **4**, 25-38

———— and L. Weiss (1985) "Money, Real Interest Rates, and Output: A Re-interpretation of Postwar U.S. Data", *Econometrica* **53**, 129-155

Lütkepohl, H. (1984) "Linear Transformations of Vector ARMA Processes", *Journal of Econometrics* **26**, 283-293

———— (1993) *Introduction to Multiple Time Series Analysis*. Berlin: Springer

Maekawa, K. (1983) "An Approximation to the Distribution of the Least Squares Estimator in an Autoregressive Model with Exogenous Variables", *Econometrica*, **51**, 229-238

———— (1985) "Edgeworth Expansion for the OLS Estimator in a Time Series Regression Model", *Econometric Theory*, **1**, 223-239

Main Economic Indicators (Historical Statistics), various editions. Paris: OECD Publications

Mann, H.B. and A. Wald (1943) "On the Statistical Treatment of Linear Stochastic Difference Equations", *Econometrica*, **11**, 173-220

McNees, S.K. (1986) "Forecasting Accuracy of Alternative Techniques: A Comparison of U.S. Macroeconomic Forecasts" with discussion, *Journal of Business & Economic Statistics*, **4**, 5-23

Nicholls, D.F. and A.L. Pope (1988) "Bias in the Estimation of Multivariate Auto-regressions", *Australian Journal of Statistics*, **30A**, 296-309

Pantula, S.G. and W.A. Fuller (1985) "Mean Estimation Bias in Least Squares Estimation of Autoregressive Processes", *Journal of Econometrics*, **27**, 99-121

Phillips, A.W. (1978) "Estimation of Systems of Difference Equations with Moving Average Disturbances", in *Stability and Inflation* (ed. A.R. Bergstrom *et al.*), 181-199. New York: Wiley

Phillips, P.C.B. (1977) "Approximations to some Finite Sample Distributions associated with a First-Order Stochastic Difference Equation", *Econometrica*, **45**, 463-485

Robertson, D. and M. Wickens (1994) "VAR Modelling", in *Applied Economic Forecasting Techniques* (ed. S. Hall), 222-318. New York: Harvester Weatsheaf

Runkle, D.E. (1987) "Vector Autoregressions and Reality" with discussion, *Journal of Business & Economic Statistics*, **5**, 437-453

Sarantis, N. and C. Stewart (1995) "Structural, VAR and BVAR Models of Exchange Rate Determination: A Comparison of Their Forecasting Performance", *Journal of Forecasting*, **14**, 201-215

Sargent, T.J. (1979*a*) *Macroeconomic Theory*. New York: Academic Press

———— (1979*b*) "Estimating Vector Autoregressions using Methods not Based on Explicit Economic Theories", *Federal Reserve Bank of Minneapolis Quarterly Review*, Summer 1979, 8-15

———— and C.A. Sims (1977) "Business Cycle Modeling Without Pretending to Have Too Much *A Priori* Economic Theory", in *New methods in business cycle research: Proceedings from a conference* (ed. C.A. Sims), 45-109. Minneapolis: Federal Reserve Bank of Minneapolis

Shenton, L.R. and W.L. Johnson (1965) "Moments of a Serial Correlation Coefficient", *American Journal of Statistics*, **27**, 308-320

Sims, C.A. (1980*a*) "Macroeconomics and Reality", *Econometrica*, **48**, 1-45

———— (1980*b*) "Business Comparison of Interwar and Postwar Cycles: Monetarism Reconsidered", *American Economic Review*, **70**, 250-257

———— (1986) "Are Forecasting Models Usable for Policy Analysis?", *Federal Reserve Bank of Minneapolis Quarterly Review*, Winter 1986, 2-16

Stock, J.H. (1987) "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors", *Econometrica*, **55**, 1035-1056

———— and M.W. Watson (1988) "Testing for Common Trends", *JASA*, **83**, 1097-1107

Tanaka, K. (1983) "Asymptotic Expansions Associated with the AR(1) Model with Unknown Mean", *Econometrica*, **51**, 1221-1231

Tiao, G.C. and G.E.P. Box (1981) "Modelling Multiple Time Series with Applications", *JASA*, **76**, 802-816

Tjøstheim, D. and J. Paulsen (1983) "Bias of some Commonly-Used Time Series Estimates", *Biometrika*, **70**, 389-399

Warne, A. (1990) *Vector Autoregressions and Common Trends in Macro and Financial Economics*, Unpubl. Ph.D. thesis. Stockholm: EFI

Webb, R.H. (1995) "Forecasts of Inflation from VAR Models", *Journal of Forecasting*, **14**, 267-285

White, J.S. (1961) "Asymptotic Expansions for the Mean and Variance of the Serial Correlation Coefficient", *Biometrika*, **48**, 85-94