Working Paper 95-9 / Document de travail 95-9

Selection of the Truncation Lag in Structural VARs (or VECMs) with Long-Run Restrictions

by
Alain DeSerres and Alain Guay

Bank of Canada Banque du Canada

October 1995

Selection of the Truncation Lag in Structural VARs (or VECMs) with Long-Run Restrictions

by

Alain DeSerres and Alain Guay

International Department
Bank of Canada
Ottawa, Ontario, Canada
K1A 0G9

Tel.: (613) 992-5555
Fax.: (613) 992-5773

Internet: alain.guay@fin.x400.gc.ca
alain.deserres@fin.x400.gc.ca

Acknowledgments

We would like to thank Marcel Kasumovich, Robert Lafrance, Pierre St-Amant and especially Simon van Norden for useful comments and suggestions. We also thank Jeffrey Gable and Robert Vigfusson for their excellent technical assistance and Helen Meubus for her thoughtful editorial comments. The responsibility for errors is ours.

ISSN 1192-5434
ISBN 0-662-23793-5

Printed in Canada on recycled paper

Abstract

The authors examine the issue of lag-length selection in the context of a structural vector autoregression (VAR) and a vector error-correction model with long-run restrictions. First, they show that imposing long-run restrictions implies, in general, a moving-average (MA) component in the stationary multivariate representation. Then they examine the sensitivity of estimates of the permanent and transitory components to the selection of the lag length required in a VAR system to approximate this MA component. In summary, they find that using a lag structure that is too short can lead to a significant estimation bias of the permanent and transitory components. In addition, in comparing four different lag-selection criteria, they find that the Schwarz information criterion systematically underperforms relative to the other tests. More generally, as the order of the VAR that best approximates the data-generating process increases, the sequence-based tests (Wald, likelihood ratio) tend to provide more reliable results than the information-based tests (Akaike, Schwarz).

Résumé

Dans la présente étude, les auteurs examinent la question du choix des retards dans un modèle structurel d'autorégression vectorielle ou modèle vectoriel de correction des erreurs avec contraintes de long terme. Ils montrent tout d'abord que l'imposition de telles contraintes implique généralement qu'une composante de moyennes mobiles (MA) intervient dans la représentation stationnaire du modèle à plusieurs variables. Puis ils examinent le degré de sensibilité que les estimations des composantes permanente et temporaire du modèle doivent afficher par rapport au choix du retard dans un système autorégressif pour s'approcher de cette composante MA. En somme, ils trouvent que l'utilisation d'une structure de retard de trop courte durée peut entraîner une distortion importante de l'estimation des composantes permanente et temporaire. En outre, une comparaison de quatre critères différents du choix des retards révèle que les tests axés sur le critère d'information de Schwarz produisent toujours des résultats moins probants que les autres. De façon générale, au fur et à mesure que s'accroît l'ordre du vecteur autorégressif qui fournit la meilleure approximation du processus de génération de données, les tests séquentiels (Wald, rapport des vraisemblances) tendent à produire des résultats plus fiables que ceux qui s'appuient sur les critères d'information (Akaike, Schwarz).

Contents

1 Introduction

In order to legitimize the use of the vector autoregression (VAR) methodology to conduct structural interpretation of impulse response functions, Sims (1986), Bernanke (1986) and Blanchard and Watson (1986) introduced "structural" VARs (or SVARs). Each proposed slightly different contemporaneous identifying restrictions that had the advantage of being more directly drawn from economic theory compared with the standard and commonly used alternative, which consisted of imposing a contemporaneous recursive structure on the vector of reduced-form errors, often without any theoretical justification.

Blanchard and Quah (1989) as well as Shapiro and Watson (1988) proposed to identify the structural shocks by imposing restrictions on their long-run effects. The main advantage was that long-run restrictions are often easier to justify on the basis of economic theory than contemporaneous restrictions. The methodology was also extended to representations with cointegration by King, Plosser, Stock and Watson (1991). Since publication of these papers, applications of the SVAR methodology, based either on contemporaneous or long-run restrictions (or a mixture of both), have gained popularity. So far, SVARs have been used mostly to evaluate the relative contribution of different "structural" shocks to fluctuations of output in various countries. In some cases, the results have contributed to the ongoing debate in the literature regarding the relative size of permanent and transitory shocks in output fluctuations. Given the failure of univariate methods to provide non-arbitrary decompositions, the multivariate (SVAR) approach can be seen as a useful alternative.

Even though one important class of SVARs relies on long-run restrictions, the method has been commonly applied to relatively short samples. This raises the difficulty of obtaining a good approximation of the matrix of long-run multipliers in small samples. While this type of econometric problem has been the object of several papers in the last few years in the context of univariate time series, it has not yet received as much attention in the multivariate context. Nevertheless, there have been some recent attempts to provide a general characterization of the properties of the SVAR methodology based on long-run restrictions. For instance, Faust and Leeper (1994) argue that to obtain a meaningful structural interpretation, long-run restrictions must be accompanied by restrictions on the finite-horizon dynamics, regardless of the sample size.[1]

The main purpose of this paper is to examine the issue of lag-length selection in the context of a structural VAR or vector error-correction model (VECM) with long-run restrictions. In particular, we focus on the estimation of the matrix of contemporaneous decomposition of reduced-form errors into their structural counterparts. In this regard, our paper can be seen as an

1. To a certain extent, their argument is similar to the point raised by Watson (1986) who showed in the univariate context that different ARIMA processes yielding almost identical short-run dynamics have very different long-run implications in terms of the size of the long-run effect of the permanent shock.

extension to the multivariate case of the work done by Hall (1995) and Ng and Perron (1994) in the context of univariate series. While both papers examine this issue in the context of unit root tests, Ng and Perron look at the implications of approximating an ARMA model with an AR specification. Extending this approach to the multivariate context, we look at the implications of using a VAR representation to approximate a VARMA model. The SVAR methodology relying on long-run restrictions is particularly exposed to this approximation problem. Indeed, we show that when some shocks have only temporary effects on one of the variables, the first difference of the multivariate representation will, in general, contain an MA component.

Even though a parallel can be drawn with the applications of some cointegration tests based on the estimation of the matrix of long-run multipliers, the examination of this issue is not one of the objectives of this paper. We concentrate instead on the implications for the structural interpretation of impulse response functions and variance decompositions derived from the application of the SVAR methodology with long-run restrictions.

To do so, we generate various bivariate VARMA models, all of which satisfy all identifying restrictions used in the estimation of SVARs. These models differ according to the size of the MA component. Given that the VAR representation of these models has an infinite lag structure, the models can only be approximated with a finite VAR specification.

Then, we estimate the VAR representation on each artificial data set for truncation lags that vary from 1 to 12. We also compare the performance of four data-based lag-length selection criteria. Two of these, Akaike (AIC) and Schwarz (SIC), are information-based criteria. The other two, the Wald and likelihood ratio (LR) criteria, are sequence-based tests and are applied according to a general-to-specific strategy for 5 and 10 per cent critical values.

To briefly summarize our results, we find that using a lag structure which is too parsimonious can lead to a significant estimation bias of the permanent and temporary components. In addition, we find that the Schwarz criterion systematically underperforms relative to the other tests and that more generally, as the order of the VAR that best approximates the data-generating process (DGP) increases, the sequence-based tests tend to provide more reliable estimates than the information-based tests.

The rest of the paper is organized as follows. Section 2 provides a description of the identification method based on long-run restrictions. The four lag-length selection criteria used are introduced in Section 3. In Section 4, we describe the basic DGP. In Section 5, we discuss the choice of parameter settings to achieve our objectives. Section 6 presents the results from the simulations. Conclusions follow.

2 Identification method based on long-run restrictions

In this section, we describe the identification method based on long-run restrictions for a general representation with integrated variables I(1) or with stationary variables I(0) that encompasses the cases with and without cointegration.

Consider the following Wold representation of a reduced-form model:

$$\Delta X_t = \mu + C(L)\varepsilon_t \tag{2.1}$$

where ΔX_t is a n x 1 vector of I(0) random variables, μ is a vector of constant, $C(0) = I$ and ε_t is a vector of i.i.d. reduced-form disturbances (Σ) with finite fourth moments.[2] The roots of $det\ C(L)$ must lie on or outside the unit circle.[3] We wish to identify the structural parameters and residuals from estimates of the reduced form, where the structural form is

$$\Delta X_t = \mu + A(L)\eta_t \tag{2.2}$$

and η_t is i.i.d. The variance-covariance matrix of η_t is usually normalized to the identity matrix, which implies that the structural disturbances are assumed to be uncorrelated. Therefore, it follows that

$$A(0)A(0)' = \Sigma \tag{2.3}$$

and

$$\varepsilon_t = A(0)\eta_t \tag{2.4}$$

The identification problem is given by these two relationships. From (2.4), we see that the n x n matrix $A(0)$, relates the reduced-form errors to their structural counterparts and has n^2 elements. Since the variance-covariance matrix Σ in (2.3) contains only $n(n+1)/2$ independent terms, $n(n-1)/2$ additional constraints are necessary to identify $A(0)$ and thereby recover the structural shocks from the reduced-form residuals.

The conventional approach to identifying the structural parameters has been to set restrictions on the matrix of contemporaneous effects between the variables included in the system by applying a Choleski decomposition on the covariance matrix of the reduced form. However,

2. ΔX_t may include stationary variables in level, as was the case in Blanchard and Quah (1989), where the unemployment rate in level is used in the VAR representation.

3. This condition rules out non-fundamental representation as defined by Lippi and Reichlin (1993).

Cooley and Leroy (1985) criticized this method on the grounds that recursive contemporaneous structure implied is often hard to reconcile with any economic theory. In order to address this line of criticism, Sims (1986), Bernanke (1986) and Blanchard and Watson (1986), proposed alternative methods whereby structural equations can be recovered from reduced-form estimates by imposing contemporaneous restrictions more directly drawn from economic theory.

Blanchard and Quah (1989) and Shapiro and Watson (1988) proposed an alternative method based on long-run restrictions. Instead of using the contemporaneous matrix, these authors imposed some restrictions on the long-run effect of the shocks. The advantage was that long-run restrictions are usually the object of a broader consensus across various economic theories. King, Plosser, Stock and Watson (1991) extended the Blanchard-Quah (BQ) methodology to a representation with cointegration vectors where the long-run restrictions correspond to common stochastic trends.

Given its reliance on long-run restrictions, the BQ method can be used in a multivariate context to identify the stochastic permanent and temporary components of a particular non-stationary time series. We can transform the reduced form in a way to make these two components explicit:

$$\Delta X_t = \mu + C(1)\,\varepsilon_t + C^*(L)\,\varepsilon_t \tag{2.5}$$

where $C(1)\,\varepsilon_t$ is permanent and $C^*(L)\,\varepsilon_t$ is transitory. The representation (2.5) corresponds to the multivariate Beveridge and Nelson (1981) decomposition. As a result, the long-run variance-covariance matrix of (2.5) is

$$C(1)\,\Sigma C(1)' \tag{2.6}$$

Given the assumption of zero correlation between the structural shocks and the normalization of their variance to unity, the long-run variance-covariance matrix of the structural form is

$$A(1)\,A(1)' \tag{2.7}$$

Therefore (2.3), (2.6) and (2.7) imply the following relation:

$$C(1)\,A(0) = A(1) \tag{2.8}$$

This relation suggests that by imposing restrictions on the long-run covariance matrix of the structural form $A(1)$, we can identify the matrix $A(0)$. For example, Blanchard and Quah estimate a VAR system for U.S. output growth and the unemployment rate. Their main objective is to identify the relative contribution of supply and demand shocks to the variance of output innovations. In their application, they assume that the two shocks driving the output and unemployment processes are mutually uncorrelated, and only one of the two (defined as the supply shock) has a permanent effect on output.

King, Plosser, Stock and Watson (henceforth, KPSW) consider the case where cointegration between the I(1) variables exists. In this case, the matrix of the reduced form $C(1)$ has the dimension n x (n-r), where r is the number of cointegration relations and n-r the number of common trends (Stock and Watson 1988). KPSW describe how cointegration relationships can be used as a set of long-run restrictions on the structural form. In their application, they try to evaluate the importance of productivity shocks on economic fluctuations. These shocks are identified as disturbances to the common trend in output, consumption and investment.

To obtain the moving-average representation of the reduced form (2.1) and the matrix of long-run multipliers $C(1)$, we need to estimate a VAR in the case without cointegration and a VECM in the case with cointegration.

In the case without cointegration, where the roots of $C(L)$ are outside the unit circle, (2.1) can be expressed as an infinite VAR. In the case with cointegration, the Granger Representation Theorem implies that any cointegrated system has an error-correction representation. Unless this representation is a finite autoregression, the error-correction model corresponds to a infinite VECM. The general infinite VECM representation is[4]

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{\infty} B_i \Delta X_{t-i} + \varepsilon_t \tag{2.9}$$

where $\Pi = 0$ if there is no cointegration. However, to estimate the system, we need to fit a finite autoregression of order k:

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k} B_i \Delta X_{t-i} + \varepsilon_{kt} \tag{2.10}$$

Now, we examine the link between the choice of truncation length k and the estimated long-run covariance matrix. In order to facilitate the discussion, we introduce some useful definitions:

$$\Gamma_x(k) = \sum_{t=k+1}^{T} (X'_{t-1}, \Delta X'_{t-1}, \ldots, \Delta X'_{t-k})' (X'_{t-1}, \Delta X'_{t-1}, \ldots, \Delta X'_{t-k}) \tag{2.11}$$

$$\hat{\Sigma}_k = \frac{1}{(T-k)} \sum_{t=k+1}^{T} \hat{\varepsilon}_{kt} \hat{\varepsilon}_{kt}' \tag{2.12}$$

4. For simplicity, we examine the case without a constant.

where $\Gamma_x(k)$ has dimension (nk+1) x (nk+1) and $\hat{\Sigma}_k$ is n x n. The vector of estimated parameters is the following:

$$\beta^k = vec(B^k) \tag{2.13}$$

β^k is a $(kn^2 \times 1)$ vector formed by stacking the columns of B_i (coefficients associated with the lagged variables in the first difference) underneath each other such that

$$B^k = (B_1, B_2, \ldots, B_k)$$

The estimation of the long-run covariance matrix (2.6) depends on the estimation of the long-run multiplier $C(1)$ and on the estimation of Σ. In the cointegration case, when a VECM is used to approximate (2.1), the long-run multiplier $C(1)$ is equal to

$$\alpha_{\perp}[I - \delta'_{\perp}B^k(1)\alpha_{\perp}]^{-1}\delta'_{\perp} \tag{2.14}$$

where α is the n x r matrix of cointegrating vectors, δ is the n x r matrix of adjustment coefficients such that $\Pi = \delta\alpha'$, $\alpha'_{\perp}\alpha = 0$ and $\delta'_{\perp}\delta = 0$. In the VAR case, $C(1)$ corresponds to

$$[I - B^k(1)]^{-1} \tag{2.15}$$

By (2.14) and (2.15), it is clear that the estimation of the long-run matrix $C(1)$ depends on the number of lags k.

In general, the problem of underestimating the appropriate number of lags will result in inconsistent estimates, regardless of whether the identification restrictions are applied to the matrix of short-run effects or to the matrix of long-run multipliers. For instance, suppose that the DGP is a VAR with k_0. If we select a VAR representation with $k < k_0$, the sum of the autoregressive process will differ from the true value. As a result,

$$[I - \hat{B}^k(1)]^{-1} \neq [I - B_0(1)]^{-1}$$

where $B_0(1)$ is the sum of AR coefficients in the DGP. The inconsistency in the estimates of the long-run covariance matrix may have an important impact on the decomposition between permanent and transitory shock, unless the inconsistency in the long-run multiplier exactly compensates for the inconsistency in the variance-covariance matrix Σ. The problem is exacerbated if the true DGP includes an MA component. In such a case, the number of lags required to approximate the MA component will increase with the sample size at a given rate. Even with a persistent, albeit temporary, component in the MA part, the truncation lag (k) must be large enough to approximate the true dynamic properly. In practice, the choice of an appropriate value for k is usually limited by the degrees of freedom available. In addition, if a higher number of lags helps to reduce the variance of the residuals, this usually comes at the expense of a higher variance of parameter estimates, which leads to a trade-off between the size of the bias in parameter estimates and the width of the confidence band around the estimates.

In the following lemme, we state that, in general, the presence of structural permanent and transitory components for at least one of the variables, which is needed for the identification based on long-run restrictions, implies a multivariate representation with an MA component.

Lemme: *For the structural form (2.2), if at least one variable is characterized by orthogonal transitory and permanent components, the multivariate representation of this structural form possesses an MA component unless the autoregressive component shares the same roots of this MA component.*

Proof: see Appendix.

By this lemme, we need, in general, to approximate an infinite VAR or VECM by a finite autoregression.

Lewis and Reinsel (1985) derived the asymptotic distribution of estimated autoregressive coefficients, obtained by fitting a VAR model of order k to a multivariate series of T observations from an infinite order autoregressive process, as k and $T \to \infty$. They developed two theorems showing the consistency of the OLS estimates when k is chosen as a function of T.

In the first theorem, they establish the consistency of $\hat{B}^k$ when (i) k is chosen as a function of T such that $k^2/T \to 0$ as k, $T \to \infty$ and ii) k is chosen as a function of T such that

$$k^{1/2} \sum_{j = k+1}^{\infty} \|B_j\| \to 0 \tag{2.16}$$

where $\|B_j\|^2 = tr(B'_j B_j)$. The first condition indicates the rate at which k must rise to control the variance of the estimators. The second condition prevents k from increasing too slowly for the goodness of fit of the approximation. As Ng and Perron (1994) point out, the second condition includes k increasing at a logarithmic rate. This condition is satisfied for any k increasing such that $k \to \infty$ when $T \to \infty$.

In the second theorem, Lewis and Reinsel show the consistency of any arbitrary linear combination of $\hat{B}^k$ when (i) k is chosen as a function of T such that $k^3/T \to 0$ as k, $T \to \infty$, and (ii) k is chosen as a function of T such that

$$T^{1/2} \sum_{j=k+1}^{\infty} \|B_j\| \to 0.$$

when k, $T \to \infty$. In this case, the second condition rules out a choice of k increasing at the logarithmic rate (see Ng and Perron 1994). More importantly, these conditions imply a rate of convergence at $\sqrt{T}$. This condition is stronger than the previous one, since k must, in this case, increase faster. The difference between the two theorems is the rate of convergence of the estimator. In finite sample, a faster rate can presumably lead to better properties.

These two theorems give the condition for the consistency of the estimator of the sum of the autoregressive parameters $\hat{B}(1)$. By virtue of the consistency of $\hat{B}^k$, we have the consistency of the estimator of the variance-covariance matrix $\hat{\Sigma}_k$. As a result, we obtain the consistency of the estimator of the long-run covariance matrix $C(1)\Sigma C(1)'$. Therefore, if we choose k correctly, we can get consistent structural decompositions. However, this leaves the question of how to choose k in finite samples. We examine this issue in the next section.

3 Criteria for lag-length selection

In this section, we describe four data-dependent rules for the choice of a truncation lag. First, we present two criteria based on the minimization of an objective function. These criteria are based on the Kullback-Leibler mean information. Following Hannan and Deistler (1988), the function to be minimized for the choice of a truncation lag is of the form

$$I_k = \log|\Sigma_k| + d(k)\, C_T/T \tag{3.1}$$

where d(k) is the number of parameters in the VAR representation. For instance, in the case of a representation with a constant, $d(k) = n(nk+1)$. While the first term is a function of the square of residuals $\hat{\Sigma}_k$, which diminish with the number of lags, the second term imposes a penalty for increasing the number of parameters. For both criteria considered below, it assumes that $C_T > 0$ and $C_T/T \to 0$. The difference between the two criteria is in the specification of C_T.

In our simulation exercise, we consider two criteria: Schwarz (SIC) and Akaike (AIC). In the case of AIC, the criterion is the following:

$$AIC_k = \log\left|\hat{\Sigma}_k\right| + d(k)\,2/T \tag{3.2}$$

and in the case of SIC:

$$SIC_k = \log\left|\hat{\Sigma}_k\right| + d(k)\log T/T \tag{3.3}$$

The Schwarz criterion is therefore more restrictive than the Akaike. Shibata (1976) has shown that, for a univariate finite order AR process, AIC asymptotically overestimates the order with positive probability, whereas Schwarz provides a consistent estimator asymptotically with probability 1.[5] These results are also valid in a VAR case. Even in a univariate Gaussian ARMA model, Shibata (1980) has shown that AIC with $C_T = 2$ chooses k in proportion to log(T). Hannan and Deistler have shown the same result in an invertible multivariate ARMA model. As Ng and Perron point out, both criteria satisfy the conditions of the first Lewis-Reinsel theorem but not the conditions of the second one.

In addition to these two criteria, we consider sequential tests for the significance of the coefficients on lags. Hall (1995) and Ng and Perron (1994) consider sequential tests for the choice of lags in unit root tests. The sequential test is applied following a general-to-specific strategy. Considering the same definition as the one used in these two papers, suppose that we want to compare a model with j lags with a model with p=j+m lags.

Definition: *The general to specific strategy chooses $\hat{k}$ as either i) j+1 where Q(j,m) is the first statistic in the sequence of $\{Q(i,m), i = J-1, \ldots, 0\}$, which is significantly different from zero at significance level α; or ii) 0 if Q(i,m) is not significantly different from zero for all $i = J-1, \ldots, 0$.*

The first of the two sequential tests that we consider is the Wald statistic, which under the hypothesis that the last m x n lags are jointly different from zero, has the following quadratic form:

$$WS_T(j,m) = \hat{\beta}(n^2(p-j))\left[\left(\hat{\Sigma}_k \otimes \Gamma_x^{-1}(k)_{(p-j)}\right)\right]^{-1}\hat{\beta}(n^2(p-j)) \sim \chi^2(n^2(p-j)) \tag{3.5}$$

where $(\,)_{(p-j)}$ is the lower right hand (p-j) x (p-j) block of $\Gamma_x^{-1}(k)$ and $\hat{\beta}(n^2(p-j))$ are the last $n^2(p-j)$ elements of the vector $\hat{\beta}$.

5. This is the main reason why we consider the Schwarz criterion in our simulations.

The second test we examine is based on the LR statistic, which has the following form:

$$LRS_T(j, m) = (T - c(k))\ (\log|\hat{\Sigma}_p| - \log|\hat{\Sigma}_j|) \sim \chi^2(n^2(p-j))$$

where c(k) is a correction factor equal to the number of variables in each unrestricted equation in the VAR, as suggested by Sims (1980).

In the next section, we use simulation methods to achieve two objectives. First, we illustrate the sensitivity of the distortion in structural parameter estimates to the underestimation of the appropriate lag length. Second, based on the results, we compare the performance of the different lag-selection criteria described in this section.

4 Description of the data-generating process

In this section, we present the general form of the DGP. Although the implications of selecting a truncation lag that is too short have been discussed in a general context, our general DGP will concentrate on the case without cointegration.

To generate the artificial data set we use a simple linear state-space modelling approach. This has several advantages. First, it is a general representation that allows for a transparent specification of the dynamics and relative size of each of the structural components included in the system. Second, it has become a familiar framework in modern macro modelling since the solution of a dynamic general equilibrium model is usually cast in a state-space model (e.g. King, Plosser and Rebelo 1988). Finally, a state-space representation makes it easier to specify a DGP that is general enough to generate results that have implications for a wide class of multivariate processes and yet simple enough that we maintain our ability to conduct controlled experiments with clear econometric interpretations.

Our state-space model has the following measurement and transition equations:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} & 0 \\ z_{21} & 0 & z_{22} \end{bmatrix} \cdot \begin{bmatrix} \alpha_{1t} \\ \alpha_{2t} \\ \alpha_{3t} \end{bmatrix} \tag{4.1}$$

$$\begin{bmatrix} \Delta\alpha_{1t} \\ \alpha_{2t} \\ \Delta\alpha_{3t} \end{bmatrix} = \begin{bmatrix} \rho_1 & 0 & 0 \\ 0 & \rho_2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \Delta\alpha_{1t-1} \\ \alpha_{2t-1} \\ \Delta\alpha_{3t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix} \tag{4.2}$$

The measurement equation (4.1) relates each of the two observable variables y_t to a pair of independent unobservable components α_{it}. The parameters z_{ii} measure the impact effect of each shock on the two variables of the system. According to the specification of the transition equation (4.2), α_1 and α_3 are non-stationary processes (α_3 is a random walk), whereas α_2 is an AR(1) stationary process that is driven by the same stochastic shock as α_3. While the shock η_{1t} has a permanent effect on both variables in y_t, the other disturbance η_{2t} has a permanent effect on y_{2t} and a temporary effect on y_{1t}. As a result, this simple structure generates two data series that satisfy all the conditions implicit in the application of a particular SVAR decomposition. First, both series are nonstationary and non-cointegrated. Second, the two shocks are mutually and serially uncorrelated. Third, one shock has a permanent effect on one variable and a temporary effect on the other.

Notice that in contrast to some univariate methods of trend-cycle decomposition, the permanent component in a multivariate decomposition method based on long-run restrictions need not follow a random walk.[6] Indeed, according to our formulation, the permanent and transitory components of y_1 can have different dynamics as indicated by the parameters ρ_1 and ρ_2. However, for the purpose of transparency, we chose to keep the dynamics of the DGP to simple low-order ARMA processes. Combining equations (4.1) and (4.2) yields the following system for the two series in first differences:

$$\begin{bmatrix} (1-\rho_1 L)\ (1-\rho_2 L) & 0 \\ 0 & 1-\rho_1 L \end{bmatrix} \begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} z_{11}(1-\rho_2 L) & z_{12}(1-\rho_1 L)\ (1-L) \\ z_{21} & z_{22}(1-\rho_1 L) \end{bmatrix} \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix} \tag{4.3}$$

In general form, this system can be expressed as

$$\Phi(L)\Delta y_t = \theta(L)\eta_t \tag{4.4}$$

If $\rho_1, \rho_2 < 1$, the MA part is invertible and the VAR representation can then be derived:

$$\theta^{-1}(L)\Phi(L)\Delta y_{t_t} = \eta_t \tag{4.5}$$

Introducing the following relationships:

$$\alpha = (z_{11}z_{22} - z_{21}z_{12}) \tag{4.6}$$

and

$$\lambda = (z_{11}z_{22}\rho_2 - z_{21}z_{12}/z_{11}z_{22} - z_{21}z_{12}) \tag{4.8}$$

6. This is the case for most multivariate methods, regardless of whether they use long-run restrictions or not.

we can rewrite (4.5) in terms of the parameters of the system:

$$\frac{\alpha}{(1-\lambda L)}\begin{bmatrix} z_{22}(1-\rho_1 L)\ (1-\rho_2 L) & -z_{12}(1-L)\ (1-\rho_1 L) \\ -z_{21}(1-\rho_2 L) & z_{11}(1-\rho_2 L)\end{bmatrix}\begin{bmatrix}\Delta y_{1t} \\ \Delta y_{2t}\end{bmatrix} = \begin{bmatrix}\eta_{1t} \\ \eta_{2t}\end{bmatrix} \tag{4.8}$$

It is clear from (4.8) that the corresponding VAR representation has an infinite lag structure and therefore can at best be approximated with a finite VAR specification. In addition, since the term λ represents the size of the moving-average component, it can be directly related to the parameters of the model by (4.7). Notice that it depends on all the structural parameters of the model except ρ_1. It is therefore independent of the dynamics of the permanent component. Moreover, we see that removing the persistence in the transitory component (i.e. setting ρ_2 to 0) is not sufficient to eliminate the MA component. The implications of this are discussed in more detail in the next section.

5 Calibration

To calibrate this state-space model, we must choose values for the parameters z_{ii}, ρ_i and the standard deviation of the two uncorrelated structural shocks. In order to motivate the choice of particular values for these parameters, we first need to state clearly the objective of the simulation exercise. The primary goal is to assess the extent to which the misspecification of the lag length in the VAR leads to distorted estimates of the structural parameters of the model and, thereby, of the relative size of the permanent and transitory components.

Given that the lag length that will provide the best approximation of the infinite VAR will vary with the size of the MA component (parameter λ in 4.8), we wish to choose values for the z_{ii} and ρ_i that will vary λ and then test the sensitivity of structural parameter estimates to these changes. In order to be more specific we need to determine which structural parameters we should focus on.

We know from Section 2 that identification of the elements of the matrix of contemporaneous decomposition A(0) is central to estimation of the entire structural model. As shown in equation 2.4, the matrix A(0) contains the elements necessary to uncover the structural residuals from the estimated reduced-form residuals. Looking at the contemporaneous relationship between the structural shocks and the two series implied by equations (4.1) and (4.2), we find that by setting the variance of the two structural shocks to unity, each element of the matrix A(0) corresponds in fact to one of the z_{ii}. Accordingly, in our discussion of the implications of the misspecification of the appropriate lag length, we will focus on the z_{ii} parameters.

More precisely, since we are essentially interested in the decomposition of y_1, we will concentrate mainly on z_{11} and z_{12}, given that the value of these two parameters is sufficient to compute the one-period-ahead variance decomposition of the prediction error on y_1.[7] This variance decomposition is often used in the application of structural VAR methodology as a measure of the relative size of each component.

In the simulation exercise that follows, we use DGPs based on three different sets of z_{ii}. In the simplest case, we set both z_{11} and z_{12} to 0.5 so that each shock accounts for 50 per cent of the one-period-ahead variance decomposition of the prediction error of y_1. z_{21} and z_{22} are then arbitrarily set to -0.5 and 0.5.[8] We then examine two alternative specifications, one where the one-period-ahead variance decomposition of the prediction error of y_1 is largely dominated by the transitory component and the other where it is dominated by the permanent component. The values for the z_{ii} used to generate these cases and the implied variance decompositions are shown in Table 1.

This calibration strategy allows us to fulfill part of our objective, which is to assess the extent to which the size of the distortion in the estimates of the structural parameters z_{11} and z_{12} is sensitive to variations in the size of the MA component (λ). In this case, we vary λ by selecting different values for the z_{ii} in the true model. Given that λ depends also on ρ_2 but not on ρ_1 (see equation 4.7) we verify how sensitive the estimated size of the structural parameters is to the degree of persistence in the dynamic adjustment of the permanent and transitory components. This is particularly relevant given that the decomposition of the variable y_1 into its permanent and transitory components is based on a restriction regarding the long-run effect of the two structural shocks.

As is the case for the z_{ii}, we consider three cases for the ρ_i. In the first case, we treat the permanent and the transitory components as being characterized by symmetric adjustment paths. To do so, we set ρ_1 equal to ρ_2 and examine the implications of varying the degree of persistence from 0 to 0.9. In Table 1, using equation (4.8), we show the size of the MA component corresponding to these various degrees of persistence for each set of values for the z_{ii}. Note that when we set $\rho_1 = \rho_2 = 0$, we obtain a value of 0.5 for the MA component (for values of the z_{ii} that correspond to the basic scenario) even though there are no dynamics in the transition equation (4.2). This shows that the existence of the MA component is associated with

---

7. The main reason for focussing on y_{1t} is that in the application of a bivariate structural VAR using long-run restrictions, usually one of the variables is the object of the decomposition, while the other is used to provide an extra error term and the information necessary to obtain a non-arbitrary decomposition of the first series. In our system, y_1 is the main variable of interest and y_2 represents the variable that provides extra information.

8. The reason for setting the parameter z_{21} to -0.5 is that using 0.5 would have implied a singular A(0) matrix.

the presence of a transitory component rather than with just the dynamics of the permanent or transitory shocks.

In the second case, we treat the permanent component as a random walk by setting ρ_1 to zero and we let ρ_2 vary over the same range used in the base case (i.e. from 0 to 0.9). This case is interesting, since in structural time-series models, the permanent component is often modelled as a random walk. Since λ is independent of ρ_1, we obtain the same values for the size of the MA component as in the previous case.

Conversely, in the final case considered, we set ρ_2 to zero and allow instead ρ_1 to take values from 0 to 0.9 as in the base case. Since the value of λ will in this case depend only on the z_{ii}, the size of the MA component is given by the first column of the right panel in Table 1 (under $\rho_2 = 0$), regardless of the value taken by ρ_1. This implies that in such a case, the size of the MA component will never be larger than 0.75, whereas in the previous two cases, it could be as high as 0.975 under certain parameter values.

To summarize, for each set of parameter values for the z_{ii} we have three sets of values for the ρ_i. In each of these sets, we use five different degrees of persistence (different values for either ρ_1, ρ_2 or for both of them, depending on the case considered) ranging from 0 to 0.9. This implies 13 different models for each set of z_{ii}. Given that we consider three different sets of z_{ii}, we have a total of 39 models or DGPs used to generate data and on which to apply the SVAR methodology based on long-run restrictions.

6 Results

In this section, we report the results of Monte Carlo experiments based on the models described above. Given the large number of experiments implied in total, we can only report a subset of the results. Therefore we will present in detail the results based on the 13 models corresponding to the case where $\rho_1 = \rho_2$. We then summarize how results change when this restriction is relaxed.

In reporting the results, we focus on the values of the parameter z_{11} and z_{12}, which measure the impact effect of each shock on y_1. The value of these two parameters is sufficient to compute the one-period-ahead variance decomposition of the prediction error on y_1.[9] Each experiment is conducted over 1000 replications and over a fixed sample of 200 observations.

9. We could, instead, have reported the variance decomposition directly, but then when a bias was found we would not have known whether it came from z_{11}, z_{12} or both.

The rest of the section is divided in two parts. In the first part, we verify how the distortion in the estimates of z_{11} and z_{12} evolve with the order of the VAR. In the second part, instead of imposing a lag length, we compare the performance of four optimal lag-selection criteria.

6.1 Sensitivity of the bias to different truncation lags

Since the 39 models presented above have different VAR representations, we first report the mean estimates of z_{11} and z_{12} obtained when we impose the same lag length for each trial. We conduct this experiment for each of the 39 models and for lags 1 to 12.

6.1.1 DGPs with ρ_1 equal to ρ_2

The results for the case where ρ_1 is equal to ρ_2 are shown in Tables 2-A to 2-C. Each table corresponds to a particular pair of values for z_{11} and z_{12} in the true model, and therefore imply a different size for the MA component. In Table 2-A, both parameters are set to 0.5, which implies an equal size for the permanent and transitory components in the one-period-ahead variance decomposition. This also implies a size of the MA component ranging from 0.5, when $\rho_1 = \rho_2 = 0$, to 0.95 when $\rho_1 = \rho_2 = 0.9$. In each table, the rows correspond to the number of lags imposed during an experiment, while each column corresponds to a different value for ρ_1 and ρ_2.

The results from Table 2-A provide some idea of the potential severity of the distortion when the truncation lag in the VAR is set too low. Even in the case where there is little or no dynamics in the transition equation, the presence of the MA component in the DGP implies that four to five lags are necessary to properly approximate the VARMA model, as shown in the first column of Table 2. In this case, using only one or two lags would lead to an overestimation of the permanent component in the variance decomposition, arising from both an overestimation of z_{11} and an underestimation of z_{12}.

Surprisingly, as the degree of persistence increases (higher value for the ρ_i), fewer lags are necessary to produce unbiased estimates of the z_{ii}, even though the size of the MA component gets significantly larger. In fact, in the case where $\rho_1 = \rho_2 = 0.9$, we find that one lag is sufficient to yield undistorted estimates of the z_{ii}. Intuitively, this can be explained by the fact that as we raise the size of the MA component via a joint increase in the value of ρ_1 and ρ_2, the difference in the dynamic persistence of the permanent and transitory components tends to diminish. As a result, even when the VAR is too parsimonious to capture the full dynamic adjustment, the complete effect of both components is more likely to be underestimated in a proportional amount, and therefore neither component will tend to be overestimated.

The results shown in Table 2-B are generated from the model where the values for z_{11} and z_{12} are set to 0.25 and 0.75 respectively, which implies a size of the permanent component equal to 10 per cent in the one-period-ahead variance decomposition (see Table 1). Generally speaking, we observe the same pattern as in Table 2-A, except that the extent of the bias in the parameter estimates is more pronounced. The distortion always goes in the direction of an overestimation of the permanent component. Again, as we increase the size of ρ_1 and ρ_2, the magnitude of the bias diminishes for low-order VARs. However, as we extend the order of the VAR we obtain a much faster convergence towards the true values when ρ_1 and ρ_2 are small.

Table 2-C reports the results for models generating a series for y_1 that is largely dominated by the permanent component. As expected, given the smaller size of the MA component, fewer lags are necessary to get an unbiased estimate of the relative size of each component in comparison to the previous cases.

6.1.2 DGPs with $\rho_1 = 0$

Next, we set ρ_1 to zero and let ρ_2 vary over the same range as before. Even though the same three sets of calibration for the z_{ii} were simulated, we only report the results for the case where z_{11} and z_{12} are both set to 0.5. These results appear in Table 3.[10]

We find that, as in the case where ρ_1 is equal to ρ_2, a parsimonious lag structure leads to a systematic overestimation of the permanent component (due to both an overestimation of z_{11} and an underestimation of z_{12}). However, comparing the results from Tables 3 and 2-A, we see that when $\rho_1 = 0$, the overestimation of the permanent component in low-order VAR systems becomes more rather than less significant as the value of ρ_2 increases (compare the last column and first row of Tables 3 and 2-A). Therefore, much longer lags are necessary to eliminate the bias for larger values of ρ_2. In fact, when ρ_2 equals 0.8 or 0.9, even with 12 lags we find some distortion in the estimates of z_{12} and therefore, in the variance decomposition.[11]

The explanation for this difference may lie in the measurement of the long-run effects of each shock. When ρ_1 is set to zero, the permanent component is a random walk, so the long-run effect of a permanent shock is likely to be estimated properly even when the number of lags included in the VAR is too small to capture the complete dynamic process of the system. In contrast, as ρ_2 increases, a longer lag structure is required to properly estimate the effect of a transitory shock. As a result, when the order of the VAR is too low, a portion of the transitory shocks may be interpreted as permanent shocks. However, for this explanation to hold, an overestimation of the transitory component should be observed in the case where the adjustment to the temporary shock is much faster than the adjustment to the permanent shock. We investigate this possibility next.

10. In the case where $\rho_2 = 0$ (first column), the results will not differ from those reported in Table 2-A.

11. Although the results are not reported, the same general pattern is observed in the case where z_{11} and z_{12} are set to 0.25 and 0.75, respectively. The main difference is that the extent of the overestimation of the permanent component is much more pronounced. This is consistent with the fact that for each value of ρ_2, the size of the corresponding MA component is larger than in the case reported in Table 3. In contrast, when z_{11} and z_{12} are set to 0.75 and 0.25, respectively, the relatively smaller values of the MA component contribute to limit the size of the distortion in parameter estimates.

6.1.3 DGPs with $\rho_2 = 0$

We perform a final set of experiments, where we set ρ_2 to zero and allow ρ_1 to take values from 0 to 0.9. The results for the case where z_{11} and z_{12} are both set to 0.5 appear in Table 4. As can be seen, for values of ρ_1 below 0.6, the permanent component will tend to be overestimated as a result of estimating a low-order VAR. However, for values of ρ_1 greater or equal to 0.6, we find, in contrast, that a parsimonious VAR will tend to produce an overestimation of the transitory component. In fact, the results obtained with $\rho_1 = 0.9$ are almost exactly the opposite of those obtained with $\rho_1 = 0$.

To better understand this result we can rewrite equation (4.8) for the case where ρ_2 is set to zero:

$$\frac{\alpha}{(1-\lambda' L)}\begin{bmatrix} z_{22}(1-\rho_1 L) & -z_{12}(1-\rho_1 L) \\ -z_{21} & z_{11} \end{bmatrix}\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

where

$$\lambda' = (-z_{21}z_{12}) / (z_{11}z_{22} - z_{21}z_{12})$$

In the special case where z_{11} and z_{12} are both set to 0.5, we find that the size of the MA component λ' is 0.5. Given that the MA component in our model is associated with the transitory component, we should expect to find the smallest bias in the relative size of each component in the case where ρ_1 is set at a value close to 0.5. In fact, we can see that with a value of 0.5 for ρ_1, the terms $(1-\lambda' L)$ and $(1-\rho_1 L)$ would cancel out in the equation for Δy_1. In other words, even when $\rho_2 = 0$, it is only for values of ρ_1 greater than the size of the MA component (0.5 in this case) that the dynamics of the permanent component dominate those of the transitory component. Moreover, the larger is the size of the transitory component in the DGP, the higher has to be the value of ρ_1 to obtain unbiased estimates when a parsimonious VAR is specified.[12]

12. Again, the same general pattern emerges in the case where z_{11} and z_{12} are set to 0.25 and 0.75, respectively. The main difference is that a value of ρ_1 equal to 0.75 rather than 0.5 is necessary to cancel the MA component and produce undistorted estimates of z_{11} and z_{12}. However, for values of ρ_1 equal to 0.8 or 0.9, the overestimation of the transitory component is much more pronounced. In contrast, when z_{11} and z_{12} are set to 0.75 and 0.25, overestimation of the transitory component is obtained for values of ρ_1 greater than 0.25, but given the relatively small size of the MA component in this case, the extent of the distortion in the estimates of z_{11} and z_{12} is fairly limited even for high values of ρ_1. These results are not reported here owing to space constraints.

To summarize, the results from Tables 2 to 4 indicate that structural parameter estimates can be highly sensitive to an underestimation of the appropriate lag length. As a consequence, if the two structural components are characterized by significantly different dynamic adjustment processes, choosing a truncation lag for the VAR that is too low can result in severe distortions in the estimation of the relative size of the permanent and transitory components. Moreover, we find that the extent of this bias depends not only on the dynamic profile of each component but also on their relative size in the DGP. In the context of our model, this is explained by the fact that the size of the MA component, and therefore the appropriate lag length in the VAR, increase with the relative size of the transitory component.

6.2 Performance of different optimal lag-selection criteria

In the previous section, we performed each simulation by imposing a specific lag length throughout the 1000 replications. This provided a background against which we can compare the performance of the four different data-dependent lag selection criteria introduced in Section 3.

As mentioned above, the Schwarz (SIC) and Akaike (AIC) criteria are based on the minimization of an objective function and are referred to as the information-based criteria. As for the Wald and LR tests, they correspond to a general-to-specific modelling strategy and will be referred to as sequence-based criteria. Since the latter two are based on critical values, we apply both of them at the 5 and 10 per cent significance levels.

The experiments were conducted over the same 39 models as before, and we again focus exclusively on the estimates of the parameter z_{11} and z_{12}. The results for the case where $\rho_1 = \rho_2$ are reported in Table 5-A. Tables 5-B and 5-C contain the results corresponding to the cases where $\rho_1 = 0$ and $\rho_2 = 0$, respectively. The three panels included in each table correspond to a pair of values for z_{11} and z_{12}.

The emphasis will be put on the cases where the selection of a short lag-length yields distorted estimates of the parameter z_{11} and z_{12}. For example, we know that in the case where ρ_1 is equal to ρ_2, the gap between parameter estimates and the true values is largest when z_{11} and z_{12} are set to 0.25 and 0.75, respectively. In particular, the results shown in the first two columns of Table 2-B suggest that under small values for ρ_1 and ρ_2 (0 or 0.2), a minimum of six or seven lags are required to produce parameter estimates that are not too far from the true value.

Against this background, we can evaluate the performance of the various lag-selection criteria by looking at the results shown in the top panel of Table 5-A. An examination of the first two columns indicates that a generally better performance is achieved when the sequence-based tests are applied. Among the information-based tests, the performance of the Akaike criterion is not substantially different from that of the sequence-based test evaluated at 5 per cent. The Schwarz criterion produces estimates with the largest distortion.

The difference in the performance between the information-based and the sequence-based tests is more evident in the case where the permanent component is a random walk ($\rho_1 = 0$). From the results of Table 3, we know that in such a case the number of lags necessary to produce unbiased results increases rapidly with higher values for ρ_2. As a result, for values of ρ_2 greater than 0.6, a parsimonious VAR produces highly distorted estimates of z_{11} and z_{12}.

The performance of the different lag-selection criteria in this case is shown in the middle panel of Table 5-B.[13] We find that for small values of ρ_2 all the criteria except the Schwarz criterion produce estimates of z_{11} and z_{12} that are close to their true values. However, as the value of ρ_2 increases, the information-based criteria tend to produce estimates that are much more distorted than the sequence-based tests. Moreover, in such a case, there is much less difference in the performance of the Akaike and Schwarz criteria since both perform just as poorly. Overall, when a long lag structure is necessary, the 10 per cent Wald and LR tests provide the least-biased estimates of the parameters z_{11} and z_{12}, with a slight advantage for the Wald test. Nevertheless, all the criteria produce large distortions.

In order to understand why the various criteria produce different results, Table 6 reports the frequency distribution of the number of lags selected for the results reported in the middle panel of Table 5-B. First, we observe that under the Schwarz criterion, the number of lags selected is always less than three for 99 per cent of the replications, which explains its poor performance. In addition, we find that although the Akaike criterion selects longer lags than the Schwarz, the maximum lag selected exceeds eight, less than 1 per cent of the time, and we find that 78 per cent of the distribution is concentrated in lags one to three.

The longest mean lag is usually found when the Wald test is used. Indeed, the 10 per cent Wald statistic is the only one systematically showing a higher proportion of lags selected in the range 10 to 12 than in the range 1 to 4. This is consistent with the fact that this statistic generally yields mean estimates for z_{11} and z_{12} that are the closest to the true values.

A final point is worth mentioning. As shown in Table 3, the higher is the degree of persistence in the transitory component (ρ_2), the longer is the lag length required to eliminate the bias in parameter estimates. However, the frequency distribution of selected lags shows the opposite phenomenon. That is, as the value of ρ_2 increases and approaches unity, an optimal lag length of order one is found much more frequently than with lower values of ρ_2. The problem is most severe in the case of the information-based criteria and is much more limited when the 10 per cent Wald or LR statistics are used.

13. This is the case where z_{11} and z_{12} are both set to 0.5, and it therefore corresponds to the case reported in Table 3.

This can be attributed to the common factor problem encountered as the value of ρ_2 approaches one. To better understand the nature of the problem, we can rewrite the VAR representation of the VARMA model (4.8) in the case where $\rho_1 = 0$:

$$\frac{\alpha}{(1-\lambda L)}\begin{bmatrix} z_{22}(1-\rho_2 L) & -z_{12}(1-L) \\ -z_{21}(1-\rho_2 L) & z_{11}(1-\rho_2 L) \end{bmatrix}\begin{bmatrix}\Delta y_{1t} \\ \Delta y_{2t}\end{bmatrix} = \begin{bmatrix}\varepsilon_{1t} \\ \varepsilon_{2t}\end{bmatrix}$$

where

$$\lambda = (z_{11}z_{22}\rho_2 - z_{21}z_{12}/z_{11}z_{22} - z_{21}z_{12})$$

We can see from that equation that in the limiting case where $\rho_2 = 1$, λ is also equal to one, in which case all the terms with a lag operator cancel out and the optimal lag in the VAR is zero. Ng and Perron (1994) obtained a similar result in the context of univariate data. Earlier, the problem had been documented by Schwert (1989), which found Phillips and Perron unit root tests to have low power for time series whose first difference had a large negative MA component.

6.3 Impulse response functions and variance decompositions

In the previous two sections, we focussed exclusively on the value of the parameters z_{11} and z_{12}, To give a more complete picture of the extent of the estimation bias and how it affects the dynamic structure, we show the impulse response functions and the variance decompositions with their respective confidence intervals for one of the models used above.

The impulse response functions we show are based on the model with values for z_{11} and z_{12} set to 0.5, and values for ρ_1 and ρ_2 set to 0 and 0.6, respectively. (The mean estimated values for z_{11} and z_{12} obtained from the imposition of a given lag length appear in the third column of results in Table 3). The mean estimated values obtained from the various lag-selection criteria appear in the middle panel of Table 5-B). We chose this particular case because it provides a good illustration of the relative performance of the different criteria. In particular, we know that even though a VAR with a short lag length will produce highly distorted estimates of z_{11} and z_{12}, a truncation lag of 12 is more than sufficient to eliminate most of the distortion.

Figures 1 to 3 show the mean impulse responses of y_1 to permanent and transitory shocks estimated from the application of the 10 per cent Wald statistic, as well as the Akaike and Schwarz criteria. The impulse responses corresponding to the Wald test are shown in Figures 1a and 1b. We know from the results of Table 5-B that the mean estimated impulses at period one are equal to 0.52 and 0.40 for the permanent and temporary shocks, respectively. As can be seen, the VAR in this case seems to approximate the VARMA adequately, since the estimates are not only unbiased but also fairly close to the true values.

Not surprisingly, a different picture emerges when the Akaike criterion is applied. As shown in Figure 2b, the true impulse response to the temporary shock lies just outside the upper bound, and as a result, the estimated impulse is significantly downward biased at a 10 per cent significance level. Even though the response to the permanent shock is not biased significantly (Figure 2a), the two results combined provide a large relative overestimation of the permanent component at short and medium horizons. Finally, when the Schwarz criterion is used, the result is a large and statistically significant bias in the estimated impulse response to both a permanent and a temporary shock (see figure 3).

The corresponding mean variance decompositions are shown in Table 7. Given that the variance decomposition is a non-linear transformation of the impulse response, even a relatively small bias in the impulse can lead to a completely misleading variance decomposition. As was the case with the impulse response, the mean estimated contribution of the temporary shock to the variance of the forecast error of y_1 varies widely across the different lag-selection criteria. Not surprisingly, the 10 per cent Wald statistic provides estimates that are much closer to the true values than the two alternatives. However, because the Wald statistic selects longer lags, the estimated variance decompositions based on the Wald statistic are associated with very large confidence intervals. In contrast, the estimates obtained from the application of both the Akaike and Schwarz criteria have narrower confidence intervals, but the relative contribution of the transitory shock to the variance decomposition is underestimated to the point of not being significantly different from zero even in the short run.

7 Conclusions

In this paper, we use simulation techniques to examine the consequences of the choice of truncation lags in the context of a structural VAR with long-run restrictions. In particular, we focus on the estimation bias found in the matrix of the contemporaneous decomposition of reduced-form errors into their structural counterparts resulting from the specification of a VAR that is too parsimonious. We do this by looking at the implications of approximating a VARMA model with a finite VAR structure.

The results obtained indicate that, under our DGP, using a lag structure that is too parsimonious can lead to an important estimation bias of the permanent and/or temporary components. In addition, we find that the Schwarz test systematically underperforms the other criteria and that more generally, as the order of the VAR that best approximates the DGP increases, the sequence-based tests tend to provide more reliable estimates than the information-based tests. These results tend to confirm those obtained by Hall (1995) and Ng and Perron (1994) in the univariate context of unit root tests.

Based on our findings, we conclude that application of the VAR methodology with long-run restrictions requires taking special care in selecting the lag length. To avoid biases, sequence-based methods (in particular the Wald test) should be preferred to information-based tests as selection criteria, even though they may imply a longer lag structure and therefore wider confidence intervals around the point estimates. In addition, it may be wise to examine the sensitivity of the key parameter estimates to the choice of truncation lags.

Appendix

First, we look at the case without cointegration. The lemme is shown in a bivariate case but could be easily generalized to the case with n-variables. Assume the following structural form:

$$\Delta X_t = A(L)\eta_t \tag{A.1}$$

where the roots of $det\,A(L)$ lie outside the unit circle. Imposing long-run restriction implies that at least one of the two variables possesses orthogonal permanent and transitory components. Therefore, we can always rewrite (A.1) in the following way:

$$\begin{bmatrix} \Delta X_{1t} \\ \Delta X_{2t} \end{bmatrix} = A^*(L)\begin{bmatrix} a_{11} & a_{12}(1-L) \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix} \tag{A.2}$$

where ΔX_{1t} has a permanent component related to the error term η_{1t} and a transitory component related to the error term η_{2t}. We see that the decomposition implied by the long-run restrictions involve an MA component corresponding to the matrix $a_{ij}(L)$, where $i, j = 1, 2$, regardless of whether the inverse of $A^*(L)$ has a finite vector autoregression representation. Taking the inverse of the moving-average representation yields the following autoregressive representation:

$$det(a(L))^{-1}adj(a(L))A^*(L)^{-1}\begin{bmatrix} \Delta X_{1t} \\ \Delta X_{2t} \end{bmatrix} = \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \end{bmatrix}$$

This representation is a finite VAR only if $A^*(L)^{-1}$ is finite and has a common root corresponding to the root of $det(a(L))^{-1}$ such that

$$A^*(L)^{-1} = det(a(L))\Phi(L)$$

where $\Phi(L)$ is of finite order. Otherwise, we are in the presence of an infinite autoregressive representation.

In the cointegration case, the number of permanent components will be smaller than the number of variables in the multivariate representation. According to the Granger Representation Theorem, there exists an error correction representation of the structural form which is

$$B(L)\Delta X_t = \Pi X_{t-1} + d(L)\eta_t$$

where $d(L)$ is a scalar lag polynomial. An finite vector autoregression exists only if $d(L) = 1$ or if $B(L)$ has roots equal to the roots of $det(d(L))$.

References

Bernanke, B. S., 1986. “Alternative explanations of the money income correlation,” *Carnegie-Rochester Conference Series on Public Policy.* pp. 49-100.

Beveridge, S. and C. R. Nelson 1981. “A new approach to the decomposition of economic time series into permanent and transitory components, with particular attention to the measurement of the business cycle.” *Journal of Monetary Economics.*

Blanchard, O. J. and D. Quah 1989. “The dynamic effects of aggregate demand and supply disturbances.” *American Economic Review 79*. pp. 655-673.

Blanchard, O. J. and M. W. Watson 1986. “Are Business Cycles all Alike?” in R. J. Gordon, ed., The American Business Cycle (University of Chicago Press).

Cooley, T. F. and S. F. Leroy. 1985. “Atheoretical Macroeconometrics: A Critique.” *Journal of Monetary Economics*. November: 283-308.

Faust, J. and E. Leeper. 1994. “When Do Long-Run Identifying Restrictions Give Reliable Results?” *Federal Reserve Bank of Atlanta*, Working Paper 94-2, April.

Hall, A. 1995. “Testing for a Unit Root in Times Series with Pretest Data Based Model Selection.” Forthcoming in *Journal of Business and Economic Statistics*.

Hannan, E. J. and M. Deistler. 1988. The Statistical Theory of Linear Systems, New-York: John Wiley and Sons.

King, R. G., C. I. Plosser and S. T. Rebelo. 1988. “Production, Growth and Business Cycles I: The Basic Mechanical Model.” *Journal of Monetary Economics*, 21, 195-232.

King, R. G., Plosser, C. I., Stock, J. H. and M. W. Watson. 1991. “Stochastic Trends and Economic Fluctuations,” *American Economic Review*, 81, September, 819-840.

Lewis, R. and G. C. Reinsel. 1985. “Prediction of Multivariate Time Series by Autoregressive Model Fitting,” *Journal of Multivariate Analysis*, 16, 393-411.

Lippi, M. and L. Reichlin. 1993. “The Dynamic Effects of Aggregate Demand and Supply Disturbances: A Comment.”, *American Economic Review*, 83, 644-52.

Ng, S. and P. Perron. 1994. “Unit Root Tests in ARMA Models with Data Dependent Methods for the Selection of the Truncation Lag,” Forthcoming in *Journal of the American Statistical Association.*

Schwert, G. W. 1989. “Tests for Unit Roots: A Monte Carlo Investigation.” *Journal of Business and Economic Statistics*, 7, 147-160

Shapiro, M. D. and M. Watson. 1988. “Sources of business cycle fluctuations.” NBER Working Paper No. 2589.

Shibata, R. 1976. “Selection of Order of an Autoregressive Model By Akaike’s Information Criterion,” *Biometrika*, 63, 117-126.

Shibata, R. 1980. “Asymptotically Efficient Solution of the Order of the Model for Estimating Parameters of a Linear Process.” *Annals of Statistics*, 8, 147-164.

Sims, C. A. 1980. “Macroeconomics and Reality.” *Econometrica*, January: 1-48.

Sims, C.A. 1986. “Are Forecasting Models Usable for Policy Analysis?” *Federal Reserve Bank of Minneapolis Quarterly Review* 10 (Winter): 2-16.

Watson, M. W. 1986. “Univariate Detrending Methods with Stochastic Trends.” *Journal of Monetary Economics*18: 49-75

Table 1

Main characteristics of the models in terms of the variance decomposition and the size of the MA component in the VAR representation									
					Value of ρ_2				
Model	Parameters of the impact effect of each shock on:		Percentage contribution of:[a]		0.0	0.2	0.6	0.8	0.9
	y_{1t}	y_{2t}	η_{1t}[b]	η_{2t}[c]	Size of the MA component				
Case 1	$z_{11} = 0.5$ $z_{12} = 0.5$	$z_{21} = -0.5$ $z_{22} = 0.5$	50	50	0.50	0.60	0.80	0.90	0.95
Case 2	$z_{11} = 0.25$ $z_{12} = 0.75$	$z_{21} = -0.25$ $z_{22} = 0.25$	10	90	0.75	0.80	0.90	0.95	0.975
Case 3	$z_{11} = 0.75$ $z_{12} = 0.25$	$z_{21} = -0.75$ $z_{22} = 0.75$	90	10	0.25	0.40	0.70	0.85	0.925

a. Variance decomposition of the one-period-ahead prediction error on the level of y_{1t}. Since we are mainly interested in the decomposition of y_{1t}, we do not report the variance decomposition of y_{2t},

b. Permanent shock to y_1.

c. Temporary shock to y_1.

Table 2-A: Mean estimated values for z_{11} and z_{12}

Case with dynamics in both components ($\rho_1 = \rho_2$)					
True values: $z_{11} = 0.50$ and $z_{12} = 0.50$					
	Value for ρ_1 and ρ_2				
	0.0	0.2	0.6	0.8	0.9
	Corresponding value of the MA parameter				
	0.5	0.6	0.8	0.9	0.95
No. of lags	Mean estimates of z_{11} and z_{12} (1000 draws)				
1	0.65/0.32	0.63/0.35	0.58/0.41	0.54/0.45	0.52/0.47
2	0.57/0.41	0.57/0.41	0.56/0.43	0.54/0.45	0.52/0.47
3	0.53/0.45	0.53/0.44	0.54/0.44	0.52/0.45	0.51/0.47
4	0.50/0.47	0.51/0.46	0.52/0.45	0.51/0.46	0.50/0.47
5	0.49/0.47	0.50/0.47	0.51/0.45	0.51/0.45	0.50/0.46
6	0.48/0.47	0.49/0.47	0.50/0.46	0.50/0.46	0.49/0.46
7	0.47/0.47	0.48/0.47	0.49/0.46	0.49/0.45	0.49/0.46
8	0.47/0.47	0.47/0.46	0.49/0.45	0.49/0.45	0.49/0.45
9	0.47/0.46	0.47/0.46	0.48/0.45	0.49/0.45	0.48/0.45
10	0.46/0.46	0.46/0.46	0.47/0.45	0.48/0.45	0.48/0.45
11	0.46/0.45	0.46/0.45	0.47/0.45	0.47/0.44	0.47/0.44
12	0.46/0.45	0.46/0.45	0.46/0.44	0.47/0.44	0.47/0.44

Table 2-B: Mean estimated values for z_{11} and z_{12}

Case with dynamics in both components ($\rho_1 = \rho_2$)					
True values: $z_{11} = 0.25$ and $z_{12} = 0.75$					
	Value for ρ_1 and ρ_2				
	0.0	0.2	0.6	0.8	0.9
	Corresponding value of the MA parameter				
	0.75	0.8	0.9	0.95	0.975
No. of lags	Mean estimates of z_{11} and z_{12} (1000 draws)				
1	0.77/0.41	0.74/0.43	0.64/0.51	0.52/0.61	0.41/0.67
2	0.62/0.54	0.62/0.53	0.59/0.55	0.49/0.62	0.40/0.67
3	0.51/0.60	0.54/0.58	0.54/0.58	0.47/0.63	0.39/0.67
4	0.44/0.64	0.47/0.62	0.49/0.61	0.44/0.64	0.38/0.67
5	0.39/0.66	0.42/0.64	0.46/0.62	0.42/0.64	0.37/0.67
6	0.35/0.67	0.38/0.66	0.42/0.63	0.41/0.65	0.36/0.67
7	0.31/0.68	0.34/0.67	0.40/0.64	0.39/0.65	0.35/0.66
8	0.29/0.68	0.32/0.67	0.38/0.65	0.38/0.65	0.35/0.66
9	0.28/0.68	0.30/0.67	0.36/0.65	0.37/0.65	0.34/0.66
10	0.27/0.68	0.29/0.67	0.34/0.65	0.36/0.65	0.33/0.66
11	0.25/0.68	0.27/0.67	0.33/0.65	0.35/0.64	0.33/0.65
12	0.25/0.67	0.27/0.67	0.32/0.65	0.34/0.64	0.32/0.65

Table 2-C: Mean estimated values for z_{11} and z_{12}

Case with dynamics in both components ($\rho_1 = \rho_2$)					
True values: $z_{11} = 0.75$ and $z_{12} = 0.25$					
Value for ρ_1 and ρ_2					
	0.0	0.2	0.6	0.8	0.9
	Corresponding value of the MA parameter				
	0.25	0.4	0.7	0.85	0.925
No. of lags	Mean estimates of z_{11} and z_{12} (1000 draws)				
1	0.76/0.19	0.76/0.20	0.75/0.22	0.75/0.23	0.74/0.24
2	0.74/0.23	0.74/0.23	0.74/0.23	0.74/0.23	0.74/0.24
3	0.73/0.24	0.73/0.24	0.74/0.23	0.73/0.23	0.73/0.24
4	0.73/0.24	0.73/0.24	0.73/0.24	0.73/0.24	0.73/0.24
5	0.72/0.24	0.72/0.24	0.72/0.23	0.72/0.23	0.72/0.23
6	0.71/0.24	0.71/0.24	0.72/0.24	0.72/0.23	0.72/0.23
7	0.70/0.24	0.71/0.23	0.71/0.23	0.71/0.23	0.71/0.23
8	0.70/0.23	0.70/0.23	0.70/0.23	0.71/0.23	0.70/0.23
9	0.70/0.23	0.70/0.23	0.70/0.23	0.70/0.23	0.70/0.22
10	0.69/0.22	0.69/0.22	0.69/0.23	0.69/0.22	0.69/0.22
11	0.68/0.22	0.68/0.22	0.69/0.22	0.69/0.22	0.69/0.22
12	0.68/0.22	0.68/0.22	0.68/0.22	0.68/0.22	0.68/0.22

Table 3: Mean estimated values for z_{11} and z_{12}

Case where permanent component is a random walk ($\rho_1 = 0$)					
True values: $z_{11} = 0.50$ and $z_{12} = 0.50$					
Value for ρ_2					
	0.0	0.2	0.6	0.8	0.9
	Corresponding value of the MA parameter				
	0.5	0.6	0.8	0.9	0.95
No. of lags	Mean estimates of z_{11} and z_{12} (1000 draws)				
1	0.65/0.32	0.68/0.27	0.71/0.14	0.71/0.07	0.71/0.04
2	0.57/0.41	0.60/0.37	0.67/0.23	0.70/0.13	0.70/0.07
3	0.53/0.45	0.55/0.42	0.64/0.29	0.68/0.18	0.69/0.10
4	0.50/0.47	0.52/0.45	0.60/0.34	0.66/0.22	0.68/0.12
5	0.49/0.47	0.50/0.46	0.57/0.37	0.64/0.25	0.67/0.15
6	0.48/0.47	0.49/0.46	0.55/0.39	0.62/0.28	0.66/0.17
7	0.47/0.47	0.48/0.47	0.53/0.41	0.60/0.30	0.64/0.19
8	0.47/0.47	0.48/0.46	0.52/0.42	0.59/0.32	0.63/0.20
9	0.47/0.46	0.47/0.46	0.50/0.42	0.57/0.33	0.62/0.22
10	0.46/0.46	0.47/0.46	0.49/0.43	0.56/0.34	0.61/0.23
11	0.46/0.45	0.46/0.45	0.48/0.43	0.54/0.35	0.60/0.24
12	0.46/0.45	0.46/0.45	0.48/0.43	0.53/0.36	0.59/0.25

Table 4: Mean estimated values for z_{11} and z_{12}

Case with no dynamics in transitory component ($\rho_2 = 0$)					
True values: $z_{11} = 0.50$ and $z_{12} = 0.50$					
	Value for ρ_1				
	0.0	0.2	0.6	0.8	0.9
	Corresponding value for the MA component				
	0.5	0.5	0.5	0.5	0.5
No. of lags	Mean estimates of z_{11} and z_{12} (1000 draws)				
1	0.65/0.32	0.60/0.39	0.45/0.54	0.36/0.60	0.31/0.63
2	0.57/0.41	0.54/0.44	0.48/0.51	0.43/0.54	0.41/0.56
3	0.53/0.45	0.51/0.47	0.48/0.50	0.46/0.51	0.45/0.52
4	0.50/0.47	0.50/0.47	0.48/0.49	0.47/0.50	0.47/0.50
5	0.49/0.47	0.49/0.48	0.48/0.48	0.48/0.49	0.47/0.49
6	0.48/0.47	0.47/0.48	0.48/0.48	0.47/0.48	0.47/0.48
7	0.47/0.47	0.47/0.47	0.47/0.47	0.47/0.47	0.47/0.47
8	0.47/0.47	0.47/0.47	0.47/0.47	0.47/0.47	0.47/0.47
9	0.47/0.46	0.47/0.46	0.47/0.46	0.47/0.46	0.47/0.46
10	0.46/0.46	0.46/0.46	0.46/0.46	0.46/0.46	0.46/0.46
11	0.46/0.45	0.46/0.45	0.46/0.45	0.46/0.45	0.46/0.45
12	0.46/0.45	0.46/0.45	0.46/0.45	0.46/0.45	0.46/0.45

Table 5-A: Mean estimated values for z_{11} and z_{12}

Case with dynamics in both components ($\rho_1 = \rho_2$)							
True Values	Lag-selection procedure		Value for ρ_1 and ρ_2				
			0.0	0.2	0.6	0.8	0.9
$z_{11} = 0.25$, $z_{12} = 0.75$	AIC		0.41/0.64	0.44/0.62	0.53/0.57	0.50/0.61	0.41/0.67
	SIC		0.59/0.54	0.65/0.49	0.64/0.51	0.52/0.60	0.41/0.67
	Wald	5%	0.35/0.65	0.36/0.65	0.41/0.62	0.41/0.63	0.37/0.66
		10%	0.30/0.67	0.32/0.66	0.37/0.64	0.38/0.64	0.35/0.66
	LR	5%	0.37/0.65	0.40/0.63	0.45/0.60	0.44/0.62	0.38/0.66
		10%	0.33/0.66	0.35/0.65	0.39/0.63	0.39/0.63	0.36/0.66
$z_{11} = 0.50$, $z_{12} = 0.50$	AIC		0.54/0.43	0.55/0.42	0.55/0.42	0.53/0.45	0.52/0.47
	SIC		0.62/0.35	0.62/0.35	0.58/0.41	0.54/0.45	0.52/0.47
	Wald	5%	0.51/0.44	0.51/0.43	0.51/0.44	0.50/0.45	0.50/0.46
		10%	0.48/0.45	0.48/0.45	0.49/0.45	0.49/0.45	0.49/0.45
	LR	5%	0.52/0.43	0.53/0.43	0.52/0.43	0.51/0.45	0.50/0.46
		10%	0.49/0.45	0.49/0.44	0.50/0.44	0.50/0.45	0.49/0.45
$z_{11} = 0.75$, $z_{12} = 0.25$	AIC		0.75/0.21	0.75/0.21	0.74/0.23	0.74/0.24	0.74/0.24
	SIC		0.76/0.19	0.76/0.20	0.75/0.22	0.75/0.23	0.74/0.24
	Wald	5%	0.72/0.22	0.72/0.22	0.72/0.23	0.72/0.23	0.72/0.23
		10%	0.70/0.23	0.70/0.23	0.70/0.23	0.70/0.23	0.70/0.23
	LR	5%	0.73/0.21	0.73/0.22	0.73/0.23	0.72/0.23	0.72/0.23
		10%	0.71/0.23	0.71/0.23	0.71/0.23	0.71/0.23	0.71/0.23

Table 5-B: Mean estimated values for z_{11} and z_{12}

Case where permanent component is a random walk ($\rho_1 = 0$)							
True Values	Lag-selection procedure		Value for ρ_2				
			0.0	0.2	0.6	0.8	0.9
$z_{11} = 0.25$ $z_{12} = 0.75$	AIC		0.41/0.64	0.46/0.61	0.63/0.43	0.76/0.19	0.79/0.08
	SIC		0.59/0.54	0.66/0.47	0.80/0.20	0.81/0.09	0.80/0.04
	Wald	5%	0.35/0.65	0.37/0.64	0.51/0.52	0.65/0.35	0.72/0.20
		10%	0.30/0.67	0.33/0.65	0.47/0.56	0.60/0.41	0.69/0.26
	LR	5%	0.37/0.65	0.41/0.62	0.55/0.49	0.68/0.31	0.74/0.16
		10%	0.33/0.66	0.35/0.64	0.49/0.54	0.63/0.38	0.71/0.22
$z_{11} = 0.50$ $z_{12} = 0.50$	AIC		0.54/0.43	0.56/0.40	0.63/0.27	0.69/0.12	0.70/0.06
	SIC		0.62/0.35	0.66/0.29	0.71/0.14	0.71/0.07	0.71/0.04
	Wald	5%	0.51/0.44	0.52/0.43	0.56/0.35	0.61/0.25	0.65/0.15
		10%	0.48/0.45	0.49/0.44	0.52/0.40	0.58/0.31	0.62/0.20
	LR	5%	0.52/0.43	0.53/0.41	0.58/0.32	0.63/0.21	0.66/0.12
		10%	0.49/0.45	0.50/0.44	0.54/0.38	0.59/0.28	0.64/0.17
$z_{11} = 0.75$ $z_{12} = 0.25$	AIC		0.75/0.21	0.75/0.19	0.77/0.11	0.78/0.06	0.78/0.03
	SIC		0.76/0.19	0.77/0.16	0.78/0.08	0.78/0.04	0.78/0.02
	Wald	5%	0.72/0.22	0.72/0.21	0.73/0.17	0.74/0.13	0.75/0.08
		10%	0.70/0.23	0.70/0.22	0.71/0.20	0.72/0.16	0.73/0.11
	LR	5%	0.73/0.21	0.73/0.20	0.74/0.15	0.75/0.11	0.76/0.07
		10%	0.71/0.23	0.71/0.22	0.72/0.19	0.73/0.14	0.74/0.09

Table 5-C: Mean estimated values for z_{11} and z_{12}

Case with no dynamics in transitory component ($\rho_2 = 0$)							
True Values	Lag-selection procedure		Value for ρ_1				
			0.0	0.2	0.6	0.8	0.9
$z_{11} = 0.25$ $z_{12} = 0.75$	AIC		0.40/0.64	0.39/0.65	0.31/0.71	0.22/0.74	0.17/0.75
	SIC		0.59/0.54	0.58/0.57	0.40/0.71	0.19/0.79	0.06/0.81
	Wald	5%	0.35/0.65	0.34/0.66	0.28/0.70	0.23/0.72	0.20/0.72
		10%	0.30/0.67	0.30/0.67	0.26/0.69	0.23/0.70	0.22/0.71
	LR	5%	0.37/0.65	0.36/0.66	0.29/0.70	0.23/0.72	0.19/0.73
		10%	0.33/0.66	0.32/0.67	0.27/0.70	0.23/0.71	0.21/0.72
$z_{11} = 0.50$ $z_{12} = 0.50$	AIC		0.54/0.43	0.53/0.45	0.47/0.51	0.44/0.53	0.43/0.54
	SIC		0.62/0.35	0.59/0.40	0.46/0.53	0.38/0.59	0.34/0.61
	Wald	5%	0.51/0.44	0.50/0.45	0.46/0.49	0.45/0.50	0.44/0.50
		10%	0.48/0.45	0.48/0.45	0.46/0.47	0.46/0.47	0.46/0.47
	LR	5%	0.52/0.43	0.51/0.45	0.47/0.49	0.44/0.51	0.43/0.51
		10%	0.49/0.45	0.49/0.45	0.46/0.48	0.45/0.49	0.45/0.49
$z_{11} = 0.75$ $z_{12} = 0.25$	AIC		0.75/0.21	0.74/0.24	0.72/0.29	0.71/0.31	0.71/0.31
	SIC		0.76/0.19	0.75/0.24	0.71/0.32	0.70/0.36	0.69/0.38
	Wald	5%	0.72/0.22	0.71/0.24	0.70/0.27	0.70/0.27	0.70/0.28
		10%	0.70/0.23	0.70/0.23	0.69/0.25	0.69/0.25	0.69/0.25
	LR	5%	0.73/0.21	0.72/0.24	0.71/0.28	0.70/0.29	0.70/0.29
		10%	0.71/0.23	0.71/0.24	0.70/0.25	0.70/0.26	0.71/0.23

Table 6: Frequency distribution of lags selected

Case with $\rho_1 = 0$ and $z_{11} = z_{12} = 0.5$														
			# of Times Lag chosen (out of 1000)											
ρ_2	procedure		1	2	3	4	5	6	7	8	9	10	11	12
0.0	AIC		70	559	244	80	26	8	6	5	1	0	0	1
	S		739	256	5	0	0	0	0	0	0	0	0	0
	Wald	5%	65	299	112	56	50	41	35	52	59	59	85	87
		10%	17	162	83	47	50	50	44	70	91	89	145	152
	LR	5%	86	402	132	51	42	38	30	41	45	45	46	42
		10%	25	220	121	66	60	57	38	72	68	67	105	101
0.2	AIC		68	458	288	114	39	18	8	5	1	0	0	1
	S		773	221	6	0	0	0	0	0	0	0	0	0
	Wald	5%	50	253	131	77	58	56	41	51	58	56	85	84
		10%	10	109	95	72	61	58	45	72	86	90	142	160
	LR	5%	78	338	153	71	56	49	36	47	41	40	48	43
		10%	24	183	127	84	67	63	45	66	69	71	99	102
0.6	AIC		304	300	176	108	57	24	11	13	5	1	0	1
	S		956	42	2	0	0	0	0	0	0	0	0	0
	Wald	5%	150	100	107	94	75	60	53	57	60	72	83	89
		10%	35	60	52	67	71	68	65	84	95	102	138	163
	LR	5%	221	134	116	102	74	64	40	52	41	48	56	52
		10%	75	82	87	98	87	69	63	79	76	87	96	101
0.8	AIC		645	181	75	44	26	10	9	4	4	1	0	1
	S		995	5	0	0	0	0	0	0	0	0	0	0
	Wald	5%	261	52	55	67	53	55	54	65	59	81	91	107
		10%	89	26	41	56	61	63	65	79	89	118	140	173
	LR	5%	386	67	64	67	52	63	41	46	43	53	57	61
		10%	150	51	55	72	72	63	67	83	72	94	101	120
0.9	AIC		807	110	41	19	8	5	3	3	2	1	0	1
	S		997	3	0	0	0	0	0	0	0	0	0	0
	Wald	5%	352	43	44	41	46	41	53	64	51	80	79	106
		10%	138	21	28	45	56	62	60	78	81	115	138	178
	LR	5%	500	43	44	44	41	42	38	42	33	56	53	64
		10%	243	39	42	58	57	53	59	75	69	96	94	115

Table 7: Variance decomposition

Percentage contribution of the temporary shock to the variance decomposition of the forecast error of y_1										
Case with $\rho_1 = 0$, $\rho_2 = 0.6$ and $z_{11} = z_{12} = 0.5$										
		Mean estimated percentage contribution with 10% confidence interval								
	True	10% Wald test			Akaike			Schwarz		
No. of periods		Low	Mean	Up	Low	Mean	Up	Low	Mean	Up
1	50.0	6.9	38.6	77.8	1.5	18.2	50.7	0.5	4.7	11.9
2	40.5	4.3	31.0	69.0	0.9	13.3	42.2	0.3	2.7	6.9
3	33.2	3.3	25.4	60.3	0.6	10.2	33.6	0.2	1.9	4.7
4	27.7	2.7	21.5	52.2	0.5	8.3	27.5	0.2	1.4	3.5
5	23.7	2.2	18.6	45.9	0.4	6.9	22.6	0.1	1.1	2.8
6	20.6	1.9	16.3	41.0	0.3	5.8	19.3	0.1	1.0	2.4
7	18.2	1.7	14.5	37.4	0.3	5.1	16.6	0.1	0.8	2.0
8	16.3	1.6	13.0	33.5	0.2	4.5	14.7	0.1	0.7	1.8
12	11.5	1.1	9.2	23.5	0.2	3.1	10.0	0.1	0.5	1.2
16	8.9	0.8	7.1	17.8	0.1	2.3	7.7	0.0	0.4	1.0
20	7.2	0.7	5.8	14.7	0.1	1.9	6.2	0.0	0.3	0.7
24	6.1	0.6	4.9	12.4	0.1	1.6	5.2	0.0	0.2	0.6

FIGURE 1a.

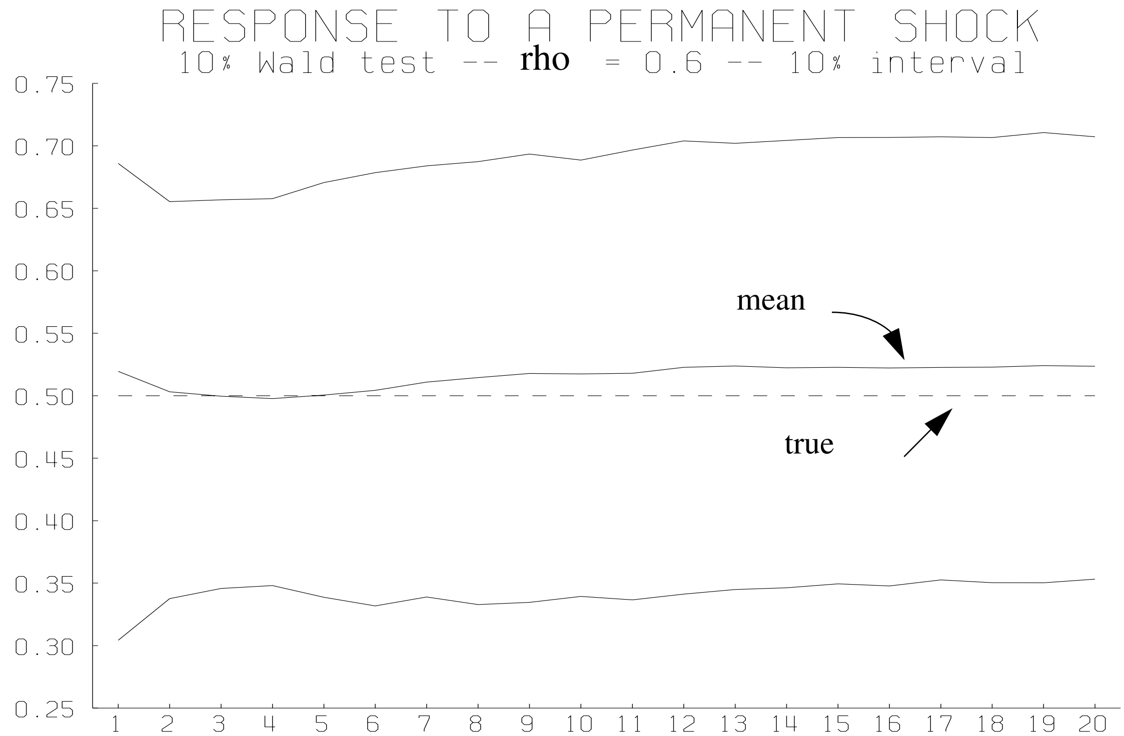


FIGURE 1b.

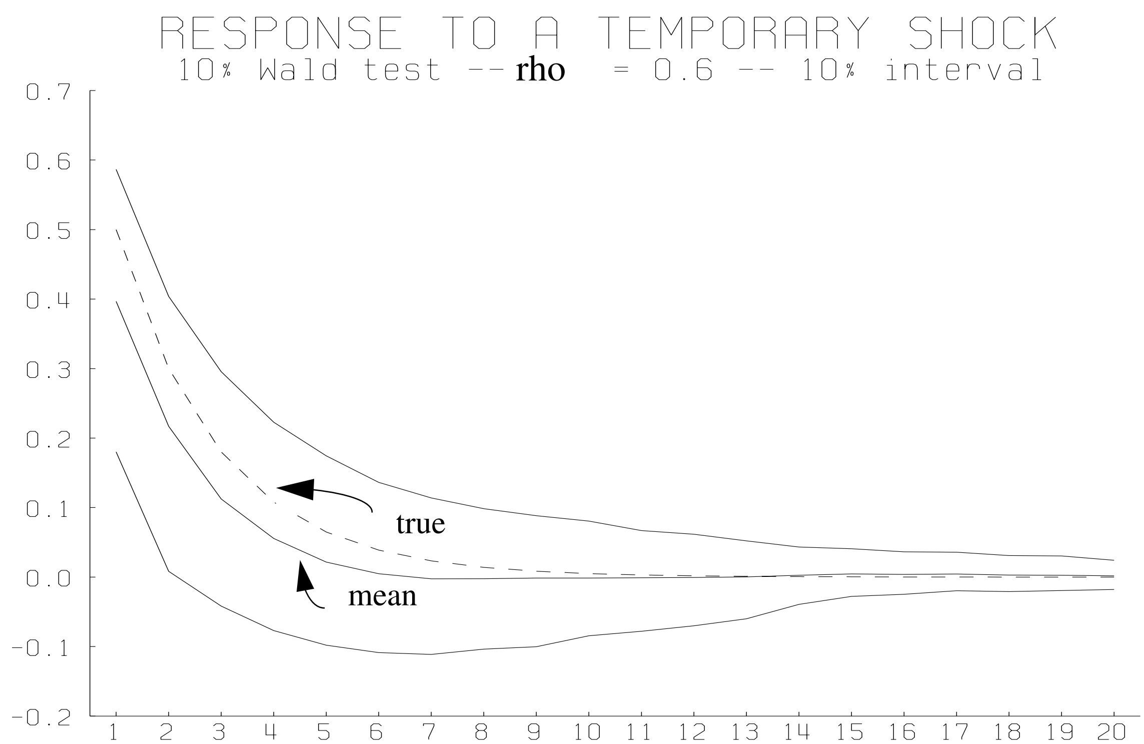


FIGURE 2a.

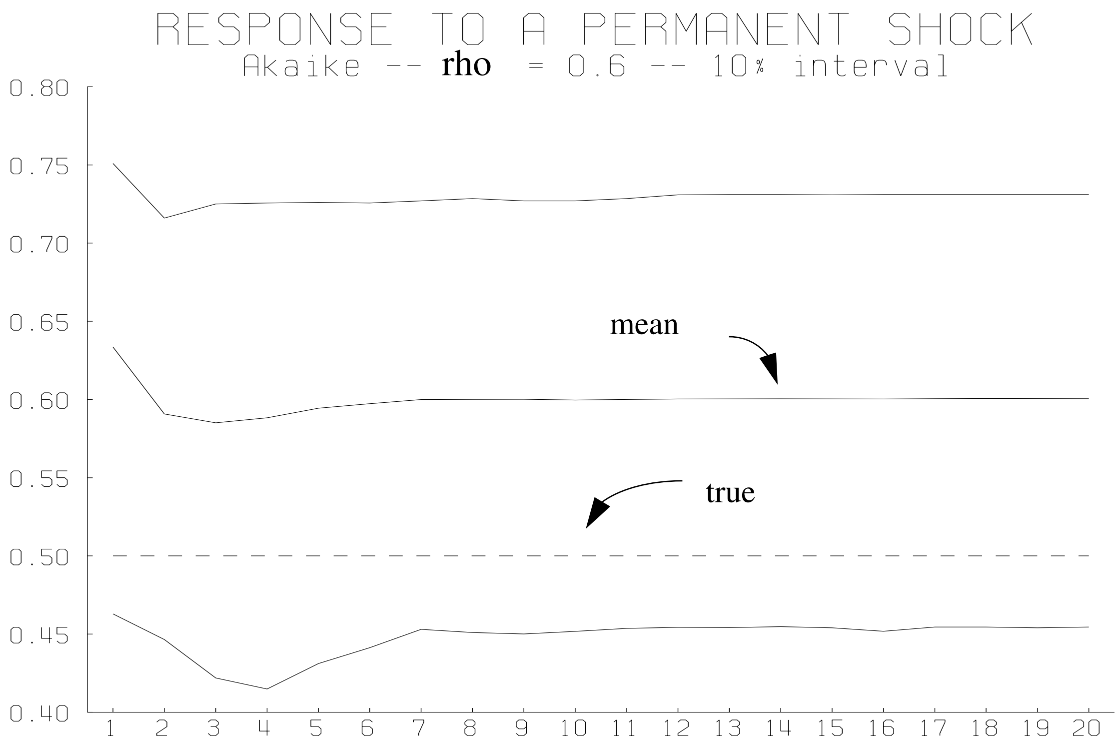


FIGURE 2b.

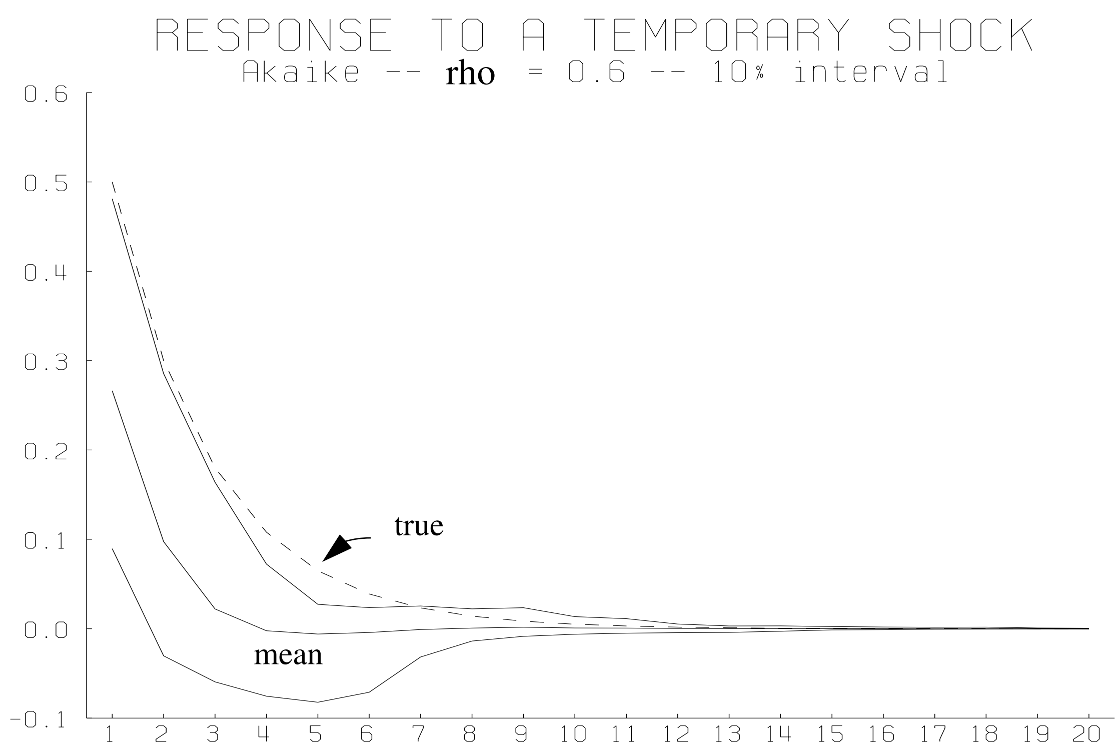


FIGURE 3a.

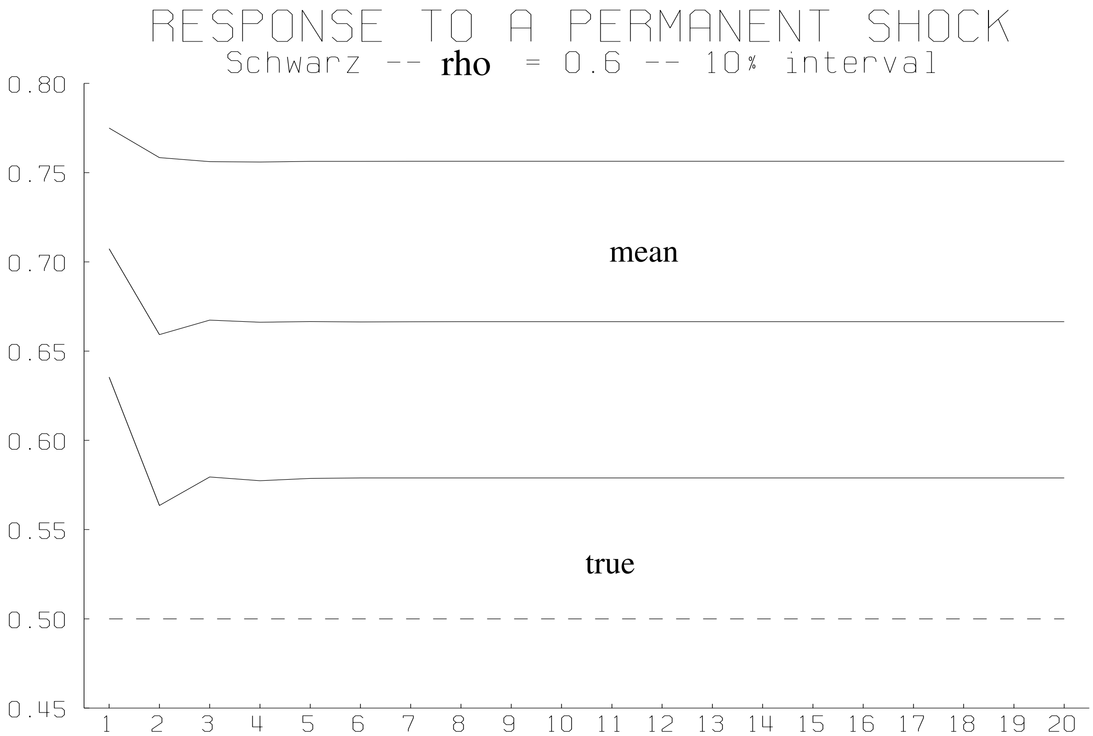
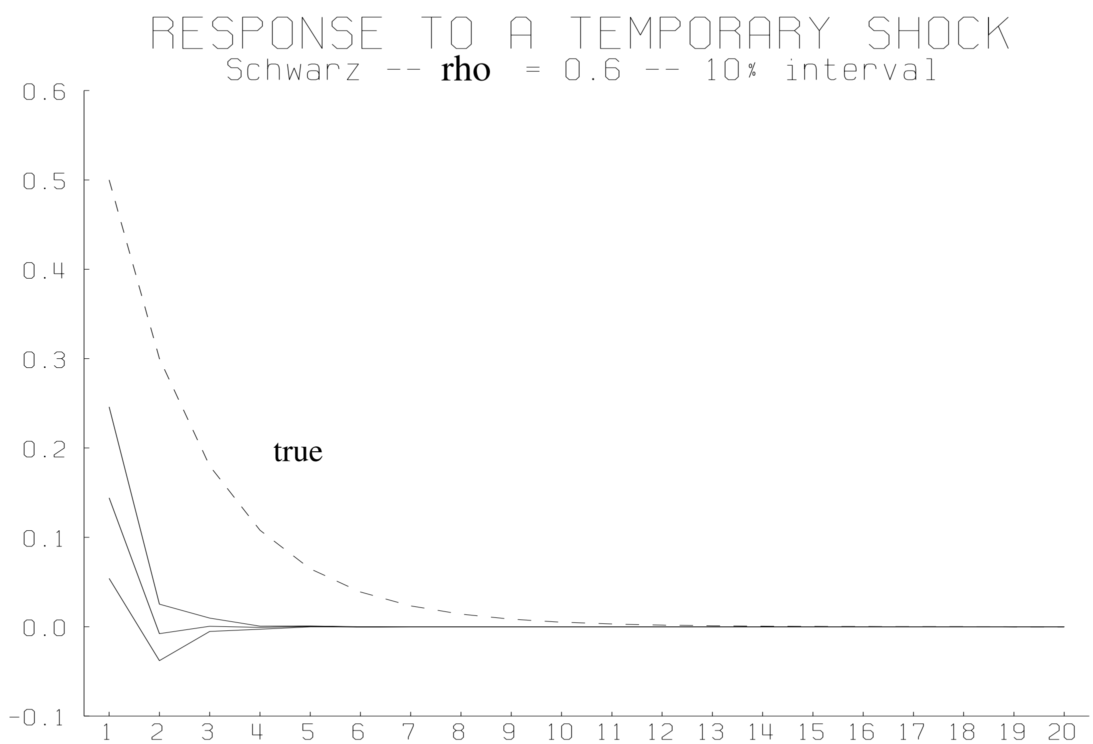


FIGURE 3b.



Bank of Canada Working Papers

1995

95-1	Deriving Agents' Inflation Forecasts from the Term Structure of Interest Rates	C. Ragan
95-2	Estimating and Projecting Potential Output Using Structural VAR Methodology: The Case of the Mexican Economy	A. DeSerres, A. Guay and P. St-Amant
95-3	Empirical Evidence on the Cost of Adjustment and Dynamic Labour Demand	R. A. Amano
95-4	Government Debt and Deficits in Canada: A Macro Simulation Analysis	T. Macklem, D. Rose and R. Tetlow
95-5	Changes in the Inflation Process in Canada: Evidence and Implications	D. Hostland
95-6	Inflation, Learning and Monetary Policy Regimes in the G-7 Economies	N. Ricketts and D. Rose
95-7	Analytical Derivatives for Markov-Switching Models	J. Gable, S. van Norden and R. Vigfusson
95-8	Exchange Rates and Oil Prices	R. A. Amano and S. van Norden
95-9	Selection of the Truncation Lag in Structural VARs (or VECMS) with Long-Run Restrictions	A. DeSerres and A. Guay

1994

(Earlier 1994 papers not listed here are also available.)

94-7	L'endettement du secteur privé au Canada : un examen macroéconomique	J.-F. Fillion
94-8	An Empirical Investigation into Government Spending and Private Sector Behaviour	R. A. Amano and T. S. Wirjanto
94-9	Symétrie des chocs touchant les régions canadiennes et choix d'un régime de change	A. DeSerres and R. Lalonde
94-10	Les provinces canadiennes et la convergence : une évaluation empirique	M. Lefebvre
94-11	The Causes of Unemployment in Canada: A Review of the Evidence	S. S. Poloz
94-12	Searching for the Liquidity Effect in Canada	B. Fung and R. Gupta

Single copies of Bank of Canada papers may be obtained from

Publications Distribution
Bank of Canada
234 Wellington Street
Ottawa, Ontario K1A 0G9

E-mail: publications@bank-banque-canada.ca

The papers are also available by anonymous FTP to the following address, in the subdirectory /pub/publications: ftp.bank-banque-canada.ca