To difference or not to difference: a Monte Carlo investigation of inference in vector autoregression models

Richard A. Ashley*

Department of Economics
Virginia Tech (VPI), USA
E-mail: ashleyr@vt.edu
*Corresponding author

Randal J. Verbrugge

Division of Price and Index Number Research (3105)
Bureau of Labor Statistics, USA
E-mail: verbrugge_r@bls.gov

Abstract: It is often unclear whether time series displaying substantial persistence should be modelled as a vector autoregression in levels (perhaps with a trend term) or in differences. The impact of this decision on inference is examined here using Monte Carlo simulation. In particular, the size and power of variable inclusion (Granger causality) tests and the coverage of impulse response function confidence intervals are examined for simulated vector autoregression models using a variety of estimation techniques. We conclude that testing should be done using differenced regressors, but that overdifferencing a model yields poor impulse response function confidence interval coverage; modelling in Hodrick-Prescott filtered levels yields poor results in any case. We find that the lag-augmented vector autoregression method suggested by Toda and Yamamoto (1995) – which models the level of the series but allows for variable inclusion testing on changes in the series – performs well for both Granger causality testing and impulse response function estimation.

Keywords: vector autoregressions; VAR model; differencing; impulse response function; unit root.

Reference to this paper should be made as follows: Ashley, R.A. and Verbrugge, R.J. (2009) 'To difference or not to difference: a Monte Carlo investigation of inference in vector autoregression models', *Int. J. Data Analysis Techniques and Strategies*, Vol. 1, No. 3, pp.242–274.

Biographical notes: Richard A. Ashley received his PhD in Economics from the University of California, San Diego. He has taught at the University of Texas at Austin and is now a member of the Department of Economics at Virginia Tech.

Randal J. Verbrugge received his PhD in Economics from Stanford University. He has taught at Virginia Tech and is now on the research staff at the Bureau of Labor Statistics.

This paper is available at http://ashleymac.econ.vt.edu/working_papers/varsim.pdf. All views expressed in this paper are those of the authors and do not reflect the views or policies of the Bureau of Labor Statistics (BLS) or the views of other BLS staff members.

1 Introduction

1.1 Overview

Model mis-specification is a fact of life in time-series econometric modelling: the samples are too small, the underlying structures we seek to model too unstable for it to be otherwise. But some mis-specifications are more consequential than others; and some modelling choices are superior to others in that they yield models whose statistical properties are relatively insensitive to the specification errors which inevitably occur. In the context of highly persistent time series data – such as are often encountered in macroeconometric modelling – this paper investigates the risk of model mis-specification and attempts to give some guidance about which modelling choices involve the least amount of downside risk for the practitioner.

1.2 Review of the literature

Beginning with Sims (1980), empirical macroeconomics has increasingly involved the estimation and analysis of Vector Autoregressive (VAR) models and related multivariate times series models. In much of the earlier work in this area, VAR models were specified in terms of the *levels* of the economic variables. Much work – *e.g.*, Christiano *et al*. (1996) – continues this practice.

This specification choice may seem odd in view of the widespread agreement that it is difficult to reject the null hypothesis of {I(1)} nonstationarity for most macroeconomic time series and given the widely known “spurious regression” simulations of Granger and Newbold (1977) who show that estimated t ratios from regressions among independently generated random walks are routinely misleading.[1] More recently, Ohanian (1988; 1991) simulated a variation on Sims’ (1980) VAR model, adding an independently generated random walk to the model. He observed what amounts to spurious regressions, in that this additional variate appears to Granger cause the macroeconomic variables, and plays a significant role in variance decompositions. And, still more recently, Granger *et al*. (2001) showed that spurious regressions can easily occur between even stationary AR(1) processes if they are fairly persistent.

Proponents of VAR modelling in levels initially met these objections with two rebuttals. First, they noted that the diagonal element lag structures in the VAR-in-levels model are free to mimic a first difference in the data generating process for each series in the model. Second, they noted that VAR models in levels generally have higher R^2 values than do VAR models in differences.[2]

More recently, the levels-versus-differences issue has re-emerged on two levels. First, it has become common in the dynamic general equilibrium literature to apply the HP filter (Hodrick and Prescott, 1980) to the levels of the data prior to further analysis – *e.g.*, Bernanke *et al.* (1997). Consequently, some might view a VAR-in-HP-filtered-data model as the appropriate alternative to a VAR-in-levels model.[3]

Second, the explosion in research on cointegrated systems has generated new theoretical results which provide an asymptotic justification for running VARs in levels. For example, results in Sims *et al*. (1990) and Toda and Phillips (1993) indicate that cointegration implies a standard asymptotic distribution for any estimated regression coefficient in the VAR-in-levels model for which the explanatory variable appears in a cointegrating relationship. In fact, their work further indicates that, even if the data are not cointegrated, some hypothesis tests on parameters multiplying covariance stationary variates will have the usual asymptotic distribution.

These results have also motivated the development of several alternatives to the VAR-in-levels model in which the levels of the series are modelled using a mixture of lagged differences in the series and a lag in the levels of the series. The most prominent of these alternatives is the lag-augmented VAR approach suggested by Toda and Yamamoto (1995); this approach will be further described below.

In the latest strand of this debate some have suggested analysing persistent data using a VAR on fractionally differenced data. Recent work, however – *e.g*., Granger and Hyung (2004), Jensen and Liu (2006), and Ashley and Patterson (2008) – indicates that the apparent fractional integration observed in some time series is essentially an artefact of structural breaks and/or weak trends, so this suggestion is not analysed here.

1.3 Summary of the analysis

Practitioners appear to have strong prior beliefs regarding what ought to be the default specification choice in this context of modelling relatively persistent time series, but there is a dearth of evidence at modest sample lengths to validate these positions. Below we provide such evidence.

The data generation processes and estimation techniques considered here are described in Section 2; these consist of:

- pairs of (trended) random walks
- pairs of stationary (in some cases borderline stationary) processes
- a simple bivariate cointegrated system.

Data simulated from these processes are then analysed using model estimation techniques including:

- a VAR in levels
- a VAR in levels including a time trend variable
- a VAR in differences
- the "lag-augmented VAR" approach of Toda and Yamamoto (1995), (which is a VAR model in levels *and* differences)[4]
- an error-correction model (simply assuming the existence of cointegration)
- a VAR model in HP-filtered data.

Four issues are then examined in Section 3 using artificial data simulated from these three data generating processes, two related to Granger causality tests and two related to impulse response confidence intervals:

1 *Which estimation techniques yield model specification (Granger causality) tests that are correctly sized?*

This is important because incorrectly sized causality tests can easily cause one to conclude that a relationship exists when it does not: 'spurious regression'.

2 *Of the correctly sized techniques, does one technique have an advantage in terms of power to reject the null hypothesis of non-causality?*

This is important because using a test with unnecessarily low power can cause one to fail to detect a relationship that does exist.

3 *Which estimation techniques yield estimated impulse response function confidence intervals with correct coverage?*

The estimation of impulse response functions is often the principal motivation for VAR modelling. Since sampling error contaminates estimated impulse response functions, confidence intervals are the only meaningful way to summarise what an estimated model has to say about the actual impulse response function. If a particular method yields 95% confidence intervals whose coverage differs substantially from .95, then the impulse response function estimates from this method are of dubious utility.

4 *Which method of computing impulse response function confidence intervals is preferable: the asymptotic "delta" method of Lütkepohl (1990) or the bias-corrected bootstrap method of Kilian (1998)?*

A priori, one would think that the bootstrap results are likely to be superior, but more computationally intensive. On the other hand, perhaps the bootstrap results are more sensitive to making the 'wrong' choice as to whether or not to difference the data.

Our results on all four of these questions are summarised in Section 4. Overall, it is worth noting here that statistical inference using models estimated in levels can be problematic (even with substantial sample lengths) but that the estimation of impulse response functions using differenced models can be equally problematic. Notably, however, the lag-augmented VAR approach – in which the explanatory variables appear in both differenced and level form, but statistical inference is done only on the coefficients of the differenced variables – yields good results for both inference and for impulse response function estimation. In contrast, models based HP-filtered levels yield uniformly poor results.

2 Data generating processes and estimation models

We examine the size and power of model specification (Granger causality) tests and the coverage of the analogous impulse response function confidence intervals using samples of length N {x(t), y(t), t = 1 ... N} generated from each of the five Data Generating Processes (DGPs) given below. In each case, the innovations $\varepsilon(t)$ and $\eta(t)$ are independently generated NIID(0, 1) processes.

DGP I. Unit root processes $\{\rho = 0 \text{ or } .7;\ \psi = 0 \text{ or } .15\}$

$$x_t = 1.0 + x_{t-1} + \rho\{x_{t-1} - x_{t-2}\} + \varepsilon_t$$

$$y_t = 1.0 + y_{t-1} + \rho\{y_{t-1} - y_{t-2}\} + \psi\{x_{t-1} - x_{t-2}\} + \eta_t.$$

DGP II. Stationary processes – causality in partial difference

$\{\varphi = .99, .95, .50, .10; \rho = .0, .7; \psi = 0 \text{ or } .15\}$

$$x_t = 1.0 + \varphi x_{t-1} + \rho\{x_{t-1} - \varphi x_{t-2}\} + \varepsilon_t$$

$$y_t = 1.0 + \varphi y_{t-1} + \rho\{y_{t-1} - \varphi y_{t-2}\} + \psi\{x_{t-1} - \varphi x_{t-2}\} + \eta_t.$$

DGP III. Stationary processes – causality in level

$\{\varphi = .99, .95, .50, .10; \rho = .0, .7; \psi = 0 \text{ or } .15\}$

$$x_t = 1.0 + \varphi x_{t-1} + \rho\{x_{t-1} - \varphi x_{t-2}\} + \varepsilon_t$$

$$y_t = 1.0 + \varphi y_{t-1} + \rho\{y_{t-1} - \varphi y_{t-2}\} + \psi x_{t-1} + \eta_t.$$

DGP IV. Simple cointegrated system – x_t adjusts towards y_t

$$x_t = 1.0 + x_{t-1} + \rho\{x_{t-1} - x_{t-2}\} - .07\{x_{t-1} - y_{t-1}\} + \varepsilon_t$$

$$y_t = 1.0 + y_{t-1} + \rho\{y_{t-1} - y_{t-2}\} + \eta_t.$$

DGP V. Simple cointegrated system – y_t adjusts towards x_t

$$x_t = 1.0 + x_{t-1} + \rho\{x_{t-1} - x_{t-2}\} + \varepsilon_t$$

$$y_t = 1.0 + y_{t-1} + \rho\{y_{t-1} - y_{t-2}\} - .07\{y_{t-1} - x_{t-1}\} + \eta_t.$$

Simulation results are presented below for samples of length 50, 200, 400, and 2000. The coefficient of –.07 on the error correction term in the two cointegrated DGPs (IV and V) is chosen so as to be statistically significant on average across the simulations; the average t ratio estimate was 2.25 in the simulations at N = 50. The null hypothesis that x(t) does not Granger-cause y(t) corresponds to setting $\psi = 0$ in DGPs I, II, and III; in DGPs IV and V there is a contemporaneous long run relationship between x(t) and y(t).[5]

The following six regression models for y_t were estimated for each of 10 000 samples of length N generated from the models given above:

1 VAR in levels

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 x_{t-1} + \alpha_4 x_{t-2} + \eta_t.$$

2 VAR in levels, plus a trend term

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 x_{t-1} + \alpha_4 x_{t-2} + \alpha_5 t + \eta_t.$$

3 VAR in differences

$$\Delta y_t = \alpha_0 + \alpha_1 \Delta y_{t-1} + \alpha_2 \Delta x_{t-1} + \eta_t.$$

4 Lag-augmented VAR

$$y_t = \alpha_0 + \alpha_1 \Delta y_{t-1} + \alpha_2 \Delta x_{t-1} + \alpha_3 y_{t-2} + \alpha_4 x_{t-2} + \eta_t.$$

5 Error correction model[6]

$$\Delta y_t = \alpha_0 + \alpha_1 \Delta y_{t-1} + \alpha_2 \Delta x_{t-1} + \alpha_3 \hat{v}_{t-1} + \eta_t.$$

6 HP-filtered VAR in levels

$$HP(y_t) = \alpha_0 + \alpha_1 HP(y_{t-1}) + \alpha_2 HP(y_{t-2}) + \alpha_3 HP(x_{t-1}) + \alpha_4 HP(x_{t-2}) + \eta_t.$$

This sixth regression model is identical to Model A except that both x(t) and y(t) are first HP-filtered.[7]

For Models A, B, and F the null hypothesis that x_t does not Granger cause y_t corresponds to x_{t-1} and x_{t-2} not appearing in the regression equation; for Models C, D, and E, this corresponds to Δx_{t-1} not appearing in the regression equation.[8] The null hypothesis is in each case tested using the usual F or t statistic. Simulation results on the empirical sizes and powers of these tests are reported in Sections 3.1 and 3.2 below. In Section 3.3 we examine the coverage of estimated confidence intervals for the impulse response function for y_t based on these models, first using the usual large-sample estimates (Lütkepohl, 1990) and then using the bias-corrected bootstrap intervals suggested by Kilian (1998).

3 Simulation results

3.1 Empirical sizes of model-specification tests

Imposing the null hypothesis that x_t does *not* Granger-cause y_t – by setting the parameter ψ to zero – the five DGPs specified in Section 2 reduce to just three:

1 a pair of independent unit root processes (DGP I)

2 a pair of independent stationary processes (DGP II and DGP III)

3 a pair of cointegrated processes (DGP IV),

since DGP II and DGP III defined in Section 2 are equivalent when ψ is zero and since the Granger non-causality null hypothesis is actually false for the cointegrated system, DGP V.

Each of these processes is used to generate 10 000 samples of length N on $\{x_t , y_t\}$. For each N-sample, all six regression models specified in Section 2 are estimated and the appropriate t or F test of the relevant null hypothesis – that there is no relationship between y_t and past x_t – is performed.

The results of this procedure for sample lengths of N = 50, 200, 400, and 2000 are given in Tables 1 through 5; these tables are collected together in the Appendix. A range of values for the parameter φ (φ = .99, .95, .50, and .10) is used in specifying the pair of stationary processes. The larger values of φ correspond to processes which are close to having a unit root; the smaller values of φ correspond to processes which are clearly stationary. The top portion of each table gives the proportion of the 10 000 trials in which the null hypothesis is (incorrectly) rejected at the 5% level of significance – *i.e.*, the size of the test – for both simple (ρ = 0) and for somewhat more complex (ρ = .7) models. Since the test is at the 5% level, this proportion should be close to .05. Analogous results for tests at the 1% level are given in the bottom portion of each table.

Several conclusions emerge from these results:

- Consistent with theory (Sims *et al.*, 1990), the estimation methods in which the explanatory variables enter in level form yield model specification tests which are over-sized when the data is I(1) but not cointegrated, even at N = 2000.

- When the data generating process is stationary but fairly persistent – *e.g.*, φ = .99 or .95 – the estimation methods in which the explanatory variables enter in level form yield model specification tests which are still quite over-sized even at N = 400. Convergence to the limiting distributions is evidently quite slow in this case.
- When the data generating process is either I(1) *or* when it is stationary and even mildly persistent, the VAR in HP-filtered levels estimation method yields substantially over-sized tests.
- Differencing the dependent variable (*i.e.*, using the VAR in differences or error correction model estimation methods) when φ is well below one yields noticeable (but modest) size distortions.
- Except for the smallest sample length (N = 50), the lag-augmented VAR estimation method generally yields reasonably-sized tests for all of the values of φ and ρ considered. (See also, Swanson *et al.*, 2003 for similar results.)
- Where the data are actually cointegrated but Granger causality from x_t to y_t is not present (DGP IV), all three tests involving differenced explanatory variables (VAR in differences, lag-augmented VAR, and error correction model) are generally well-sized; by N = 400 the levels-based tests which are not HP-filtered are reasonably well-sized also.

We conclude that, for moderately-to-severely persistent data, it is not a good idea to estimate a VAR model in levels for purposes of investigating Granger causality relationships – *i.e.*, for determining which explanatory variables belong in the relationship – and that HP filtering one's data does not help matters. In contrast, all three estimation methods in which the explanatory variables appear in differenced form – *i.e.*, the VAR in differences, the lag-augmented VAR, and the error correction model methods – appear to yield reasonably sized hypothesis tests, with the lag-augmented VAR estimation method yielding the most robust results.[9]

Thus, for this kind of model specification analysis, unit root testing is unnecessary: if the data are persistent enough that unit root testing seems warranted, then one should not be using the levels-based estimation methods at all.[10] Rather, one should be using a method in which the explanatory variables enter in differenced form, such as a lag-augmented VAR.

3.2 Empirical power of model specification tests

Evidence was presented in Section 3.1 indicating that model specification tests in VAR models estimated on the levels (or HP-filtered levels) of the data are poorly sized, whereas all three estimation models using differenced explanatory variables (VAR in differences, lag-augmented VAR, and error correction model) yield reasonably well-sized tests. Correct size, however, is no guarantee that a test will have good power. It is widely believed that differencing greatly reduces the power of Granger causality tests.[11] And, in any case, it is of interest to examine whether the power of model specification tests based on one of these three estimation methods dominates that of the others.

Results on the empirical power of the Granger causality tests are presented in Table 5 for the three correctly sized tests; Table 6 is displayed in the Appendix. Each of these power estimates is obtained as the fraction of 10 000 trials for which the null hypothesis of non-causality from past x(t) to current y(t) is rejected at the 5% level when this hypothesis is false because the data were generated using a non-zero value (.15) for the parameter ψ.[12]

The results given in Table 5 indicate that causality tests based on each estimation model can have substantial power; this result does not support the conventional wisdom that differencing substantially reduces the power of causality tests.

The power of the causality tests based on the error correction estimation model is typically somewhat lower than that of the tests based on the other two estimation models. Where cointegration is not present, this is because the error correction model is actually the VAR in differences model plus an additional explanatory variable which enters with coefficient zero.

When the data are generated by DGP V, the cointegration implies Granger causality from x_t to y_t. Here the tests based on the VAR in differences and the lag-augmented VAR estimation models do reject the null hypothesis of no relationship, especially in the cases where ρ is .70. (This relationship is detected in the error correction estimation model also, but there it is appropriately attributed entirely to the cointegration vector term.) Thus, these tests are correctly indicating that past values of some form of x_t belong in the equation for y_t; they can hardly be recommended as tests for cointegration, however.

We also find that the VAR in differences estimation method yields causality tests with higher power than the lag-augmented VAR method for positive ρ and vice-versa for ρ equal to zero. The most reasonable interpretation of this result is that which of these two methods yields higher power depends on the details of the DGP.[13] This observation suggests that it would be useful to run such tests using either or both the VAR in differences and the lag-augmented VAR estimation methods.

3.3 *Impulse response function confidence intervals*

The dynamic implications of an estimated VAR model are commonly analysed by considering its implied impulse response functions. These quantify the impact of each innovation on future values of the dependent variables. For example, for a two dimensional VAR with a maximum lag length of two periods, such as:

$$A(B)\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} A_{yy}(B) & A_{yx}(B) \\ A_{xy}(B) & A_{xx}(B) \end{bmatrix}\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}$$

where:

$$A_{yy}(B) = 1 - A_{yy1}B - A_{yy2}B^2$$

$$A_{yx}(B) = 1 - A_{yx1}B - A_{yx2}B^2$$

$$A_{xy}(B) = 1 - A_{xy1}B - A_{xy2}B^2$$

$$A_{xx}(B) = 1 - A_{xx1}B - A_{xx2}B^2$$

the impulse response functions for y_t are $\Omega_{y\varepsilon}(B)$ and $\Omega_{y\eta}(B)$ in the MA(∞) representation:

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = A^{-1}(B)\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix} = \begin{bmatrix} \Omega_{y\varepsilon}(B) & \Omega_{y\eta}(B) \\ \Omega_{x\varepsilon}(B) & \Omega_{x\eta}(B) \end{bmatrix}\begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}$$

where:

$$\Omega_{y\varepsilon}(B) = 1 + \Omega_{y\varepsilon 1}B + \Omega_{y\varepsilon 2}B^2 + \Omega_{y\varepsilon 3}B^3 + \Omega_{y\varepsilon 4}B^4 + \ldots$$

$$\Omega_{y\eta}(B) = 1 + \Omega_{y\eta 1}B + \Omega_{y\eta 2}B^2 + \Omega_{y\eta 3}B^3 + \Omega_{y\eta 4}B^4 + \ldots$$

Since sampling error contaminates estimated impulse response functions (*i.e.*, estimates of $\Omega_{y\varepsilon 1}$, $\Omega_{y\varepsilon 2}$, $\Omega_{y\varepsilon 3}$, $\Omega_{y\varepsilon 4}$, $\Omega_{y\eta 1}$, $\Omega_{y\eta 2}$, $\Omega_{y\eta 3}$, and $\Omega_{y\eta 4}$), a sequence of *confidence intervals* – one for each of these terms – is the only meaningful way to summarise what an estimated model has to say about the actual impulse response function. If, using a particular estimation method, the actual coverage of these (95%) confidence intervals differs substantially from the nominal value (.95), then the impulse response function estimates from this method are of dubious utility.

Table 6 and Table 7 report estimates of the coverage of 95% confidence intervals obtained by applying each of the estimation methods described in Section 2 to data generated from each of the DGPs described there. Table 6 reports on the coverage estimates so obtained with confidence intervals based on asymptotic standard errors for the $\Omega_{y\varepsilon j}$ and $\Omega_{y\eta j}$ coefficients; Table 7 reports analogous coverage estimates for confidence intervals based on the bias-corrected bootstrap method of Kilian (1998).[14]

Asymptotic standard errors for the relevant coefficients in Ω(B) were obtained, using the usual 'delta' method – *e.g.*, Lütkepohl (1990) – by locally linearising the relationship between Ω(B) and the estimated coefficient matrix, A(B), using a first order Taylor expansion. Ordinarily the required partial derivatives for this expansion are obtained by numerical differentiation of each coefficient in Ω(B) with respect to each coefficient in A(B). Due to the large number of simulations, estimation models, and DGPs considered here, this was computationally infeasible.[15] Consequently, we instead derived analytical expressions for $\Omega_{y\varepsilon j}$ and $\Omega_{y\eta j}$ as functions of the elements of A(B) and obtained explicit expressions for the relevant partial derivatives. These results tremendously alleviated the computational burden but, because the algebra becomes overwhelming for j larger than four, the calculations are limited to the first four lags in each of the two impulse response functions. Fortunately, the impact of lag length on the confidence interval coverage for the various estimation models and DGPs is already quite apparent by lag four.

Kilian (1998) provides an exposition of the computational details of his bias-corrected bootstrap approach, but some comment seems appropriate here as to (1) why one might expect the bootstrap to provide better confidence intervals in this setting than those obtained using the asymptotic standard errors and (2) why one might expect the bias-corrected bootstrap to work better than the ordinary bootstrap.

First, since $\Omega_{y\varepsilon j}$ and $\Omega_{y\eta j}$ (the impulse response coefficients for y_t at lag j) depend on terms like $(A_{yy1})^j$, it is reasonable to expect that these coefficients will be substantially non-Gaussian for values of j greater than one. This, along with its generally excellent small sample properties, suggests that the bootstrap might provide superior confidence intervals for coefficients like $\Omega_{y\varepsilon j}$ and $\Omega_{y\eta j}$. However, the essential bootstrap approximation – replacement of the actual distribution of the innovations by the empirical distribution of the fitting errors – hinges on these fitting errors being identically and independently distributed. This is an appropriate assumption for large samples, assuming that sufficient lags are included in the VAR model. However, least squares parameter estimation of dynamic regression models is known to exhibit substantial bias in small samples when the data is fairly persistent, as is precisely the case when the issue is whether to estimate in levels or differences. Consequently, there is ample reason to suspect that it might be preferable to first use the bootstrap to estimate these finite sample parameter estimation biases and then re-sample out of residuals based on coefficient estimates corrected for these estimated biases.[16] We confirmed the result of Kilian (1998) that this is the case, so only the bias-corrected bootstrap coverage results are reported in Table 7.

Our implementation of the Kilian (1998) algorithm is essentially identical to his except in two details. First, we use our analytic expressions for $\Omega_{y\varepsilon j}$ and $\Omega_{y\eta j}$ as functions of the elements of A(B) instead of computing them numerically. Again, this limits us to a maximum lag length of four, but these results are comparable to those obtained using the asymptotic standard errors and, in any case, the patterns that are most relevant here – those relating to the relative performance of the various estimation techniques across the various DGPs – are all quite clear by lag four.[17] The other difference between our implementation and Kilian (1998) is that, as in Kilian and Chang (2000), we re-estimate the bias correction for each newly simulated data set whereas Kilian (1998) estimates these biases once and for all using the first simulated data set. As he points out, this makes no difference asymptotically, but it does make a noticeable (albeit not dramatic) difference in the coverage results for the estimation models and sample lengths we consider here.

Tables 6 and 7 summarise our estimated impulse response function confidence interval coverage results using these two approaches. The estimates are in each case based on 10 000 samples of length N generated from each of the five DGPs specified in Section 2 and in each case the impulse response function confidence intervals are obtained using each of the six estimation methods also specified in that Section. Since these are 95% confidence intervals, coverage would ideally be .95 for every interval.[18]

The results in these tables address the questions of which estimation method, and which confidence interval construction method, one should use so as to minimise the potential losses associated with mis-specifying the nature of the data generating process. Several conclusions are noteworthy:

- (Lag-augmented VAR ≈ VAR in levels.) Impulse response function confidence interval coverages resulting from estimation using a lag-augmented VAR model and a VAR in levels model are similar, irrespective of the data generating process and of which confidence interval construction method was used. This is not entirely surprising, since a lag-augmented VAR is simply a linear transformation of a levels VAR. This result implies that there is no reason based upon the impulse response function confidence interval coverage results to prefer one of these estimation models over the other, but it is worth noting again that the results in Section 3.1 strongly indicate that the lag augmented VAR estimation model is much superior to the VAR in levels estimation model in terms of providing correctly sized model specification inference results.

- (Include a trend term.) Upon comparing the coverage of a VAR in levels with and without a trend, the results presented in Tables 6 and 7 suggest that one is somewhat better off omitting a time trend. This should not be surprising, since in all the cases considered in these tables, the addition of a time trend is a mis-specification. However, results not reported in these tables completely overturn this conclusion: when the data generating process consists of stationary fluctuations around a linear trend, the penalty to omitting a trend term in a VAR in levels can be severe. For example, at N = 200, the coverage of the confidence intervals from the VAR in levels estimation method for this case is below .03 at all lags. The downside risk to omitting a trend is so large that we advise its inclusion unless one is nearly certain that the data generating process is I(1).
- (Do not HP filter.) Estimation using HP-filtered data gives rise to extremely poor confidence interval coverage, regardless of the method of confidence interval estimation, a result which should give pause to practitioners who commonly HP-filter data prior to analysis.
- (Bootstrap vs. delta method.) Bias-corrected bootstrapped impulse response function confidence intervals are generally somewhat superior to those obtained using asymptotic standard errors, but not markedly so. Comparing the results in Table 6 to those in Table 7, the evidence is mixed. At least for the first four lags, neither approach provides substantially better coverage than the other consistently across the variety of DGPs and (mostly mis-specified) estimation models considered here.[19]
- (Differencing is risky.) The differenced model yields excellent confidence interval coverage when the data is really I(1), but not cointegrated, as in DGP V. However, if the data is stationary or cointegrated, as in DGP V, then the fully differenced estimation method yields poor results except in very small samples. In fact, the results from the fully differenced estimation method actually worsen as the sample length increases. This is intuitively appealing (as larger data sets allow the estimation routine to react more and more strongly to the mis-specification involved in over-differencing the data) but the intensity of this effect on the confidence interval coverages is unexpected. Significantly, in this situation the confidence interval coverage from this estimation method is becoming extremely poor in the very circumstance (large sample length) where one would be most likely to give one's inference results a high degree of credence.[20]

Thus, unless one is quite confident that the data is I(1), impulse response functions should be estimated using a model whose dependent variable enters in level form – such as a VAR in levels or lag-augmented VAR. Results in Tables 6 and 7 indicate that, in contrast to differenced models, both the lag-augmented VAR and the VAR in levels models yield impulse response function confidence interval coverage which converge properly and which are relatively robust with respect to whether the data is stationary, nonstationary, or cointegrated.[21]

Overall, however, the coverage of the confidence intervals for impulses to y_t (*i.e.*, shocks t to ε_t) is not very good for persistent data, especially beyond lags one or two: though the sampling distributions of these impulse response coefficients do converge properly, their convergence is notably slower than one might hope for or expect. In contrast, the coverage of the confidence intervals for impulses to x_t (*i.e.*, shocks to η_t) is reasonably good.

To difference or not to difference? In contrast to our results on model specification testing in Sections 3.1 and 3.2, here we come to the opposite conclusion. While (unsurprisingly) it is always helpful to know whether or not the data actually is I(1), for the purposes of estimating impulse response function confidence intervals with good coverage, the downside risk of overdifferencing is substantially larger than that of underdifferencing.

4 Conclusions

Where one's principal concern is valid model specification (Granger causality) testing, our results strongly contraindicate the use of VAR in levels estimation models: For data which are persistent – precisely the kind where one might well be wondering whether to difference or not – the VAR in levels estimation models yield poorly sized tests, even when the DGP is actually stationary and even when the sample is fairly long, such as N = 400. HP-filtering the data and then estimating a VAR model in levels yields even worse results. In contrast, our results do support the use of the three estimation methods in which the explanatory variables appear in differenced form – *i.e.*, the VAR in differences, lag-augmented VAR, and error correction estimation models – as long as cointegration is not an issue.[22] When cointegration is present, we find that standard Granger causality tests may or may not detect this long run relationship, depending upon the form that the cointegration takes.

Conversely, where one's principal concern is in obtaining impulse response function confidence intervals with good coverage, we find that VAR in levels and lag-augmented VAR estimation methods perform adequately, at least for rather large samples, but that the differenced estimation models (*i.e.*, the VAR in differences and error correction models) are problematic. If the data are really I(1) but not cointegrated, then these estimation models yield excellent results. However, if the data are persistent but really I(0), then both of these differenced modelling choices yield extremely poor results as the sample length becomes large. This result is intuitively appealing: as the sample length increases, the mis-specification involved in differencing an I(0) series becomes increasingly consequential. Our results show that this effect becomes substantial at sample lengths as low as N = 200 for the impulse response coefficients at longer lags. HP-filtering the data gives rise to confidence intervals with extremely poor coverage regardless of the data generating process.

We came to this project with a sense that many economists (ourselves included) are reluctant to estimate a VAR in levels for any data which are persistent enough that a unit root seems a reasonable possibility – primarily because of the danger of spurious regressions, along the lines illustrated by Granger and Newbold (1974). The development of the asymptotic theory (Sims *et al*., 1990) validating the use of VARs in levels for some inference purposes was not entirely convincing, given the substantial small-sample downward bias in estimated coefficients on lagged dependent variables when the data are persistent.

Our results indicate that this reluctance is well-founded in the context of model specification hypothesis testing: consistent with asymptotic theory, VAR in levels models yield substantially incorrect test sizes with I(1) data, even in fairly large samples. Moreover, we find that these tests remain mis-sized for data which is I(0) but strongly persistent, even for samples of length 400. Thus, for persistent data, such testing should be done on explanatory variables in differenced form. Nevertheless, the results we have obtained here indicate that modelling in levels *is* appropriate in the context of impulse response function confidence interval estimation. In that context, the mis-specification danger from possible overdifferencing clearly outweighs the danger of spurious regression.

Notably, however, the lag-augmented VAR estimation approach – in which the level of the time series is modelled but the explanatory variables are transformed so that inference can be done on explanatory variables which are in differenced form – yields good results in either context.

What, then, is a non-Bayesian to do in practice?[23] Based on the results obtained here, we recommend the following:

- Do not HP-filter the data.[24]
- For model specification (short-term Granger causality) testing, use either a VAR in differences model or a lag-augmented VAR model; either has reasonably accurate size and comparable power in our simulations.
- Estimate the impulse response functions (and confidence intervals for same) using a model in which the dependent variable is in levels – using either the levels model or the lag-augmented VAR model – with a trend term included. The bias-corrected bootstrap confidence intervals appear to be preferable to using asymptotic standard errors, but not dramatically so. Unless the sample length is *quite* large, however, be aware that the actual coverage of nominally 95% confidence intervals may be *substantially* less than 95%, especially past lags one or two – the best available statistical inference machinery simply does not converge all that quickly with persistently dependent data.[25]

References

Ashley, R. and Patterson, D.M. (2008) 'Apparent long memory in time series as an artifact of a time-varying mean', Unpublished manuscript, http://ashleymac.econ.vt.edu/working_papers/long_memory.pdf.

Ashley, R. and Verbrugge, R. (2009) 'Frequency dependence in regression model coefficients: an alternative approach for modeling nonlinear dynamic relationships', *Econometric Reviews*, forthcoming, http://ashleymac.econ.vt.edu/working_papers/freq_depend.pdf.

Bernanke, B.S., Gertler, M. and Watson, M. (1997) 'Systematic monetary policy and the effects of oil price shocks', *Brookings Papers on Economic Activity*, pp.91–157.

Christiano, L.J., Eichenbaum, M. and Evans, C. (1996) 'The effects of monetary policy shocks: evidence from the flow of funds', *Review of Economics and Statistics*, Vol. 78, No. 1, pp.16–34.

Cogley, T. and Nason, J.M. (1995) 'Effects of the Hodrick-Prescott filter on trend and difference stationary time series: implications for business cycle research', *J. Economic Dynamics and Control*, Vol. 19, pp.253–278.

Diebold, F.X. and Kilian, L. (2000) 'Unit root tests are useful for selecting forecasting models', *Journal of Business and Economic Statistics*, Vol. 18, No. 3, pp.265–273.

Granger, C.W.J. and Hyung, N. (2004) 'Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns', *Journal of Empirical Finance*, Vol. 11, pp.399–421.

Granger, C.W.J., Hyung, N. and Jeon, Y. (2001) 'Spurious regressions with stationary series', *Applied Economics*, Vol. 33, pp.899–904.

Granger, C.W.J. and Newbold, P. (1977) *Forecasting Economic Time Series*, New York: Academic Press.

Harvey, A.C. and Jaeger, A. (1993) 'Detrending, stylized facts and the business cycle', *J. Applied Econometrics*, Vol. 8, pp.231–247.

Hodrick, R. and Prescott, E.C. (1980) 'Post-war U.S. business cycles: an empirical investigation', *Mimeo*, Carnegie-Mellon University, Pittsburgh, Pennsylvania.

Jensen, M.J. and Liu, M. (2006) 'Do long swings in the business cycle lead to strong persistence in output', *Journal of Monetary Economics*, Vol. 53, No. 3, pp.597–611.

Kilian, L. (1998) 'Small-sample confidence intervals for impulse response functions', *Review of Economics and Statistics*, Vol. 80, No. 2, pp.218–230.

Kilian, L. and Chang, P. (2000) 'How accurate are confidence intervals for impulse response functions in large VAR models?', *Economics Letters*, Vol. 69, pp.299–307.

King, R.G. and Rebelo, S.T. (1993) 'Low frequency filtering and real business cycles', *J. Economic Dynamics and Control*, Vol. 17, pp.207–231.

Kurozumi, E. and Yamamoto, T. (2000) 'Modified lag augmented vector autoregressions', *Econometric Reviews*, Vol. 19, No. 2, pp.207–231.

Lin, J.L. and Tsay, R. (1996) 'Cointegration constraints and forecasting: an empirical examination', *Journal of Applied Econometrics*, Vol. 11, pp.519–538.

Lütkepohl, H. (1990) 'Asymptotic distributions of impulse response functions and forecast error decompositions of vector autoregressive models', *Review of Economics and Statistics*, Vol. 72, pp.116–125.

Ohanian, L.E. (1988) 'The spurious effects of unit roots on vector autoregressions: a Monte Carlo study', *Journal of Econometrics*, Vol. 39, pp.251–256.

Ohanian, L.E. (1991) 'A note on spurious inference in a linearly detrended vector autoregression', *Review of Economics and Statistics*, Vol. 73, No. 3, pp.568–571.

Pesavento, E. and Rossi, B. (2006) 'Impulse response function confidence intervals for persistent data: what have we learned', *Journal of Economic Dynamics and Control*, Vol. 31, No. 7, pp.2398–2412.

Phillips, P.C.B. (1986) 'Understanding spurious regressions in econometrics', *J. Econometrics*, Vol. 33, pp.311–340.

Phillips, P.C.B. (1995) 'Fully modified least squares and vector autoregression', *Econometrica*, Vol. 63, pp.1023–1078.

Plosser, C.I. and Schwert, G.W. (1977) 'Estimation of a non-invertible moving average process', *Journal of Econometrics*, Vol. 6, pp.199–224.

Plosser, C.I. and Schwert, G.W. (1978) 'Money, income, and sunspots: measuring economic relationships and the effects of differencing', *Journal of Monetary Economics*, Vol. 4, pp.637–660.

Sims, C.A. (1980) 'Macroeconomics and reality', *Econometrica*, Vol. 48, No. 1, pp.1–48.

Sims, C.A., Stock, J.H. and Watson, M.W. (1990) 'Inference in linear times series models with some unit roots', *Econometrica*, Vol. 58, pp.113–144.

Sims, C.A. and Zha, T. (1999) 'Error bands for impulse responses', *Econometrica*, Vol. 67, pp.1113–1155.

Swanson, N.R., Ozyildrim, A. and Pisu, M. (2003) 'A comparison of alternative causality and predictive accuracy tests in the presence of integrated and co-integrated economic variables', in D.E.A. Giles (Ed.) *Computer-Aided Econometrics*, London: Routledge, pp.91–148.

Toda, H.Y. and Phillips, P.C.B. (1993) 'Vector autoregressions and causality', *Econometrica*, Vol. 61, pp.1367–1393.

Toda, H.Y. and Phillips, P.C.B. (1994) 'Vector autoregressions and causality: a theoretical overview and simulation study', *Econometric Reviews*, Vol. 13, No. 2, pp.259–285.

Toda, H.Y. and Yamamoto, T. (1995) 'Statistical inference in vector autoregressions with possibly integrated processes', *J. Econometrics*, Vol. 66, pp.225–250.

Yamada, H. and Toda, H.Y. (1998) 'Inference in possibly integrated vector autoregressive models: some finite sample evidence', *J. Econometrics*, Vol. 86, pp.55–95.

Notes

1 See also Plosser and Schwert (1977; 1978). Phillips (1986) provides a firm theoretical basis for Granger and Newbold's results by showing that limiting distributions do not exist for the estimated t ratio in such circumstances so even attempts to adjust the critical values are asymptotically unjustified in this context.

2 However, a higher R^2 does not even imply a better fit to the data since the dependent variables are different for the VAR-in-differences and VAR-in-levels models. Note also that R^2 does not even necessarily admit of an interpretation as proportion of variance explained in the VAR-in-levels model, since the variance is not finite for an I(1) level series.

3 This is, of course, despite the published work (*e.g.*, Harvey and Jaeger, 1993; King and Rebelo, 1993; Cogley and Nason, 1995) showing that HP filtering can severely distort estimated dynamic relationships.

4 See also Kurozumi and Yamamoto (2000). The related FM-VAR approach suggested by Phillips (1995) and the sequential ECM approach suggested by Toda and Phillips (1994) are not implemented here based on simulation results presented in Yamada and Toda (1998) indicating that the lag-augmented VAR ('LA-VAR') approach is better-sized in small samples.

5 Data generating processes with feedback are not explicitly considered because including such terms did not materially affect the results quoted below. Note that equations for generating both x(t) and y(t) are needed, but the focus below is entirely on the estimated equations for y(t).

6 Here $\hat{v}_t$ is the fitting error from estimating $x_t = \beta_0 + \beta_1 y_t + v_t$.

7 HP filtering software written by E. Prescott was obtained from the Quantitative Macroeconomics and Real Business Cycle website (http://ideas.uqam.ca/QM RBC); the smoothing parameter was set to the standard value of 1600.

8 Note that $\alpha_2 = 0$ in Model D corresponds to $\rho = 0$ (*i.e.*, to x_t not Granger causing y) in all five DGPs; thus, there is no need for Model D to include terms in Δx_{t-2} and x_{t-3}.

9 The modified lag-augmented VAR procedure proposed by Kurozumi and Yamamoto (2000) might provide still better results.

10 This is not to say that unit root testing may not be useful in other contexts. For example, Diebold and Kilian (2000) find such testing helpful in choosing forecasting models. Nor – as becomes apparent from the results of Section 3.3 below – does it imply that differenced-based models are necessarily preferable for producing confidence intervals on impulse response functions in models for which Granger causality is already known to be present.

11 Note, however, that the results in Section 3.1 above rationalise the existence of this belief without validating it, since the apparent power of an over-sized test will be artificially inflated.

12 This fraction is meaningless for the tests based on the models estimated in levels or in HP-filtered levels since, in these cases, the actual size of a nominally 5% test is substantially different from .05. The parameter ψ explicitly appears in DGP I, DGP II, and DGP III in Section 2.

13 All three estimation methods have approximately the same power (ca. .05) for the generating process where x_t adjusts to y_t (DGP IV of Section 2) since x_t does not Granger cause y_t in that case.

14 Bayesian interval estimation (as in Sims and Zha, 1999) is beyond the scope of the present paper. We also note that *both* methods are only asymptotically justified; the relevant issue is the relative and absolute effectiveness of each method using the finite sample lengths considered here.

15 Numerical precision issues also arise in calculating $\partial \Omega y\varepsilon j / \partial Ayy1$, say, for large j since terms like $\Omega y\varepsilon j$ involve $(Ayy1)^j$.

16 See Kilian (1998) for details on how this bias correction is performed; for example, the correction is explicitly adjusted so that it never makes an originally stationary model nonstationary.

17 In fairness, however, we should note that this lag limitation in our results may cause us to understate the superiority of the bootstrap confidence intervals over those obtained using the asymptotic standard errors.

18 Estimated models involving differenced variables were re-written (post-estimation) in terms of the levels of the variables. The estimated sampling distributions of $A_{yy1} \ldots A_{xx2}$ were obtained based on the estimated sampling distributions of the coefficients in the equations actually estimated. Tables 6 and 7 are based on parameter values of $\psi = .15$ and $\rho = 0$.

19 We cannot rule out the possibility that the bootstrap is substantially superior at larger lags. Also, results which we do not report in these tables overwhelmingly support the contention of Kilian (1998) that bias-corrected bootstrap intervals are definitely preferable to ordinary bootstrap intervals.

20 Over differencing is also consequential for long term forecasting – see Lin and Tsay (1996).

21 In results not reported in Tables 6 and 7, we find very similar coverage results for $\psi = 0$; thus, this result is also robust with respect to whether or not this explanatory variable was erroneously included in the model.

22 Note that the dependent variables in lag-augmented VAR model equations are in levels and that each explanatory variable also appears in level form at the largest lag, so it is not appropriate to think of the lag-augmented VAR as a 'differenced' model.

23 Bayesian interval estimation – *e.g.*, Sims and Zha (1999) – is beyond the scope of the present paper.

24 Users of HP filtering sometimes imply that their filtering choice is justified by an interest in relationships 'only at business cycle frequencies'. If one really believes that the relevant relationships are frequency-dependent, then regression methods which explicitly allow for this – *e.g.*, Ashley and Verbrugge (2009) – should be employed.

25 This conclusion is consistent with the results of Kilian and Chang (2000), who look at impulse response function confidence interval coverage for several higher dimensional VAR models (see also Pesavento and Rossi, 2006).

Appendix

Simulation results

Table 1a Test sizes at N = 50 (ρ = 0.00 case)

5% test results	*Generating process*					
	Unit root processes		*Stationary processes (II and III)*			
Estimated model	*Independent (I)*	*Cointegrated (IV)*	$\varphi = .99$	$\varphi = .95$	$\varphi = .50$	$\varphi = .10$
VAR in levels	.213	.131	.119	.107	.058	.052
VAR in levels with trend	.118	.123	.124	.114	.062	.053
VAR in differences	.051	.051	.052	.053	.063	.069
Lag-augmented VAR	.057	.056	.060	.056	.050	.049
Error correction model	.063	.061	.049	.046	.037	.031
VAR in HP-filtered levels	.089	.103	.093	.092	.068	.055

1% test results	*Generating process*					
	Unit root processes		*Stationary processes (II and III)*			
Estimated model	*Independent (I)*	*Cointegrated (IV)*	$\varphi = .99$	$\varphi = .95$	$\varphi = .50$	$\varphi = .10$
VAR in levels	.069	.036	.036	.028	.012	.011
VAR in levels with trend	.037	.037	.037	.035	.014	.011
VAR in differences	.009	.011	.010	.011	.015	.018
Lag-augmented VAR	.012	.012	.011	.012	.011	.011
Error correction model	.013	.013	.010	.010	.005	.006
VAR in HP-filtered levels	.023	.026	.026	.026	.015	.013

Table 1b Test sizes at N = 50 (ρ = 0.70 case)

5% test results	*Generating process*					
	Unit root processes		*Stationary processes (II and III)*			
Estimated model	*Independent (I)*	*Cointegrated (IV)*	$\varphi = .99$	$\varphi = .95$	$\varphi = .50$	$\varphi = .10$
VAR in levels	.275	.119	.232	.161	.074	.067
VAR in levels with trend	.197	.139	.194	.184	.087	.074
VAR in differences	.062	.061	.068	.067	.060	.057
Lag-augmented VAR	.070	.083	.077	.074	.052	.052
Error correction model	.057	.058	.063	.057	.042	.040
VAR in HP-filtered levels	.122	.137	.125	.125	.093	.079

Table 1b Test sizes at N = 50 (ρ = 0.70 case) (continued)

1% test results	*Generating process*					
	Unit root processes		*Stationary processes (II and III)*			
Estimated model	*Independent (I)*	*Cointegrated (IV)*	*φ = .99*	*φ = .95*	*φ = .50*	*φ = .10*
VAR in levels	.107	.035	.084	.051	.018	.014
VAR in levels with trend	.069	.043	.076	.069	.022	.016
VAR in differences	.014	.017	.017	.015	.013	.014
Lag-augmented VAR	.019	.020	.019	.020	.010	.011
Error correction model	.012	.016	.015	.012	.007	.007
VAR in HP-filtered levels	.036	.042	.038	.039	.025	.019

Table 2a Test sizes at N = 200 (ρ = 0.00 case)

5% test results	*Generating process*					
	Unit root processes		*Stationary processes (II and III)*			
Estimated model	*Independent (I)*	*Cointegrated (IV)*	*φ = .99*	*φ = .95*	*φ = .50*	*φ = .10*
VAR in levels	.233	.074	.113	.075	.052	.051
VAR in levels with trend	.129	.096	.126	.089	.053	.051
VAR in differences	.054	.046	.052	.052	.061	.068
Lag-augmented VAR	.050	.045	.055	.053	.051	.051
Error correction model	.051	.048	.051	.049	.037	.033
VAR in HP-filtered levels	.079	.146	.085	.085	.067	.060

1% test results	*Generating process*					
	Unit root processes		*Stationary processes (II and III)*			
Estimated model	*Independent (I)*	*Cointegrated (IV)*	*φ = .99*	*φ = .95*	*φ = .50*	*φ = .10*
VAR in levels	.079	.018	.034	.017	.013	.011
VAR in levels with trend	.041	.026	.039	.023	.013	.011
VAR in differences	.009	.010	.012	.012	.016	.019
Lag-augmented VAR	.009	.008	.012	.011	.012	.012
Error correction model	.011	.011	.011	.011	.007	.005
VAR in HP-filtered levels	.020	.053	.024	.023	.017	.015

Table 2b Test sizes at N = 200 (ρ = 0.70 case)

5% test results	*Generating process*					
	Unit root processes		*Stationary processes (II and III)*			
Estimated model	*Independent (I)*	*Cointegrated (IV)*	*φ = .99*	*φ = .95*	*φ = .50*	*φ = .10*
VAR in levels	.246	.065	.163	.084	.058	.056
VAR in levels with trend	.153	.063	.171	.106	.060	.058
VAR in differences	.047	.054	.060	.058	.060	.057
Lag-augmented VAR	.053	.057	.058	.056	.048	.049
Error correction model	.048	.052	.053	.053	.047	.044
VAR in HP-filtered levels	.092	.209	.098	.098	.085	.077

1% test results	*Generating process*					
	Unit root processes		*Stationary processes (II and III)*			
Estimated model	*Independent (I)*	*Cointegrated (IV)*	*φ = .99*	*φ = .95*	*φ = .50*	*φ = .10*
VAR in levels	.087	.016	.050	.022	.012	.013
VAR in levels with trend	.052	.013	.058	.030	.012	.013
VAR in differences	.009	.010	.014	.014	.011	.015
Lag-augmented VAR	.010	.014	.013	.011	.010	.011
Error correction model	.008	.011	.013	.012	.009	.008
VAR in HP-filtered levels	.025	.082	.028	.027	.022	.019

Table 3a Test sizes at N = 400 (ρ = 0.00 case)

5% test results	*Generating process*					
	Unit root processes		*Stationary processes (II and III)*			
Estimated model	*Independent (I)*	*Cointegrated (IV)*	*φ = .99*	*φ = .95*	*φ = .50*	*φ = .10*
VAR in levels	.240	.063	.097	.062	.045	.047
VAR in levels with trend	.130	.078	.113	.070	.045	.048
VAR in differences	.045	.049	.046	.048	.058	.066
Lag-augmented VAR	.045	.046	.048	.047	.044	.045
Error correction model	.051	.051	.047	.044	.036	.033
VAR in HP-filtered levels	.081	.218	.079	.079	.064	.059

Table 3a Test sizes at N = 400 (ρ = 0.00 case) (continued)

1% test results	*Generating process*					
	Unit root processes		*Stationary Processes (II and III)*			
Estimated model	*Independent (I)*	*Cointegrated (IV)*	$\varphi = .99$	$\varphi = .95$	$\varphi = .50$	$\varphi = .10$
VAR in levels	.080	.015	.030	.015	.009	.009
VAR in levels with trend	.045	.020	.035	.017	.009	.009
VAR in differences	.009	.009	.009	.009	.011	.016
Lag-augmented VAR	.009	.008	.008	.009	.009	.008
Error correction model	.009	.011	.008	.008	.006	.006
VAR in HP-filtered levels	.021	.093	.019	.020	.014	.012

Table 3b Test sizes at N = 400 (ρ = 0.70 case)

5% test results	*Generating process*					
	Unit root processes		*Stationary processes (II and III)*			
Estimated model	*Independent (I)*	*Cointegrated (IV)*	$\varphi = .99$	$\varphi = .95$	$\varphi = .50$	$\varphi = .10$
VAR in levels	.243	.055	.116	.068	.048	.048
VAR in levels with trend	.145	.052	.146	.081	.050	.047
VAR in differences	.047	.051	.068	.049	.058	.055
Lag-augmented VAR	.047	.050	.050	.047	.048	.045
Error correction model	.046	.050	.063	.047	.045	.041
VAR in HP-filtered levels	.090	.339	.090	.090	.077	.071

1% test results	*Generating process*					
	Unit root processes		*Stationary processes (II and III)*			
Estimated model	*Independent (I)*	*Cointegrated (IV)*	$\varphi = .99$	$\varphi = .95$	$\varphi = .50$	$\varphi = .10$
VAR in levels	.087	.013	.034	.016	.010	.009
VAR in levels with trend	.051	.012	.047	.019	.011	.010
VAR in differences	.010	.010	.016	.011	.011	.010
Lag-augmented VAR	.011	.011	.010	.010	.008	.008
Error correction model	.010	.010	.016	.010	.009	.007
VAR in HP-filtered levels	.025	.163	.024	.024	.020	.017

Table 4a Test sizes at N = 2000 (ρ = 0.00 case)

5% test results	*Generating process*					
	Unit root processes		*Stationary processes (II and III)*			
Estimated model	*Independent (I)*	*Cointegrated (IV)*	*φ = .99*	*φ = .95*	*φ = .50*	*φ = .10*
VAR in levels	.243	.056	.069	.051	.044	.043
VAR in levels with trend	.134	.056	.075	.052	.044	.044
VAR in differences	.049	.047	.046	.045	.054	.062
Lag-augmented VAR	.049	.042	.046	.046	.045	.045
Error correction model	.050	.048	.045	.043	.036	.030
VAR in HP-filtered levels	.089	.697	.085	.086	.062	.054

1% test results	*Generating process*					
	Unit root processes		*Stationary processes (II and III)*			
Estimated model	*Independent (I)*	*Cointegrated (IV)*	*φ = .99*	*φ = .95*	*φ = .50*	*φ = .10*
VAR in levels	.084	.013	.016	.011	.009	.008
VAR in levels with trend	.043	.012	.018	.011	.009	.008
VAR in differences	.009	.009	.008	.008	.012	.016
Lag-augmented VAR	.009	.008	.007	.007	.008	.008
Error correction model	.009	.009	.008	.008	.007	.005
VAR in HP-filtered levels	.023	.489	.024	.023	.014	.012

Table 4b Test sizes at N = 2000 (ρ = 0.70 case)

5% test results	*Generating process*					
	Unit root processes		*Stationary processes (II and III)*			
Estimated model	*Independent (I)*	*Cointegrated (IV)*	*φ = .99*	*φ = .95*	*φ = .50*	*φ = .10*
VAR in levels	.236	.053	.068	.056	.050	.049
VAR in levels with trend	.130	.051	.077	.057	.049	.047
VAR in differences	.039	.052	.056	.051	.059	.052
Lag-augmented VAR	.039	.051	.044	.044	.045	.044
Error correction model	.038	.052	.055	.048	.049	.041
VAR in HP-filtered levels	.096	.932	.100	.100	.086	.074

Table 4b Test sizes at N = 2000 (ρ = 0.70 case) (continued)

1% test results	*Generating process*					
	Unit root processes		*Stationary processes (II and III)*			
Estimated model	*Independent (I)*	*Cointegrated (IV)*	$\varphi = .99$	$\varphi = .95$	$\varphi = .50$	$\varphi = .10$
VAR in levels	.082	.011	.015	.013	.010	.009
VAR in levels with trend	.046	.010	.018	.012	.010	.009
VAR in differences	.006	.011	.012	.011	.014	.011
Lag-augmented VAR	.006	.008	.007	.008	.010	.008
Error correction model	.006	.010	.012	.009	.010	.006
VAR in HP-filtered levels	.027	.820	.028	.028	.021	.017

Table 5a Empirical power of the Granger causality tests (ρ = 0.00 case)[a]

		Generating process			
		Unit root processes		*Stationary processes ($\varphi = .95$)*	
Estimated model	*N*	*Independent (I)*	*Cointegrated (V)*	*Causality in partial difference (II)*	*Causality in level (III)*
VAR in differences	50	.169	.065	.170	.099
Lag-augmented VAR	50	.244	.091	.170	.157
Error correction model	50	.071	.057	.133	.048
VAR in differences	200	.540	.138	.541	.319
Lag-augmented VAR	200	.593	.167	.539	.532
Error correction model	200	.437	.050	.485	.048
VAR in differences	400	.840	.241	.837	.581
Lag-augmented VAR	400	.861	.275	.837	.834
Error correction model	400	.785	.052	.798	.047
VAR in differences	2000	1.000	.832	1.000	.998
Lag-augmented VAR	2000	1.000	.863	1.000	1.000
Error correction model	2000	1.000	.052	1.000	.056

Note: [a] $\psi = .15$ in each case.

Table 5b Empirical power of the Granger causality tests (ρ = 0.70 case)[a]

		Generating process			
		Unit root processes		*Stationary processes (φ = .95)*	
Estimated model	*N*	*Independent (I)*	*Cointegrated (V)*	*Causality in partial difference (II)*	*Causality in level (III)*
VAR in differences	50	.300	.259	.294	.487
Lag-augmented VAR	50	.278	.075	.243	.161
Error correction model	50	.161	.059	.204	.255
VAR in differences	200	.824	.820	.813	.990
Lag-augmented VAR	200	.806	.198	.761	.663
Error correction model	200	.742	.051	.756	.438
VAR in differences	400	.985	.986	.982	1.000
Lag-augmented VAR	400	.984	.383	.972	.945
Error correction model	400	.977	.050	.972	.411
VAR in differences	2000	1.000	1.000	1.000	1.000
Lag-augmented VAR	2000	1.000	.980	1.000	1.000
Error correction model	2000	1.000	.047	1.000	.437

Note: [a]ψ = .15 in each case.

Table 6 Impulse response function – empirical coverage of asymptotic 95% confidence interval

a Unit root generating processes (DGP I)

Estimated model	*Own innovation (ε_t)*				*X_t innovation (η_t)*			
	Lag 1	*Lag 2*	*Lag 3*	*Lag 4*	*Lag 1*	*Lag 2*	*Lag 3*	*Lag 4*
VAR in levels								
N = 50	.87	.74	.72	.71	.94	.92	.89	.86
N = 200	.93	.86	.82	.79	.95	.94	.90	.87
N = 400	.95	.90	.87	.83	.95	.94	.92	.89
N = 2000	.95	.94	.93	.92	.95	.95	.94	.94
VAR in levels with trend								
N = 50	.74	.45	.31	.24	.94	.87	.81	.76
N = 200	.90	.68	.53	.42	.95	.93	.90	.87
N = 400	.93	.79	.66	.55	.95	.94	.92	.90
N = 2000	.94	.91	.86	.81	.95	.95	.95	.94
VAR in differences								
N = 50	.94	.92	.92	.92	.94	.93	.95	.95
N = 200	.95	.94	.95	.95	.95	.95	.96	.96
N = 400	.95	.95	.95	.95	.95	.96	.96	.96
N = 2000	.95	.95	.95	.95	.95	.95	.95	.95
Lag-augmented VAR								
N = 50	.87	.74	.72	.71	.94	.92	.89	.86
N = 200	.93	.86	.82	.79	.95	.94	.90	.87
N = 400	.95	.90	.87	.83	.95	.94	.92	.89
N = 2000	.95	.94	.93	.92	.95	.95	.94	.94
Error correction model								
N = 50	.94	.92	.93	.94	.91	.90	.92	.92
N = 200	.94	.95	.95	.95	.94	.94	.95	.95
N = 400	.95	.95	.95	.96	.95	.95	.95	.95
N = 2000	.94	.95	.95	.95	.95	.95	.95	.95
VAR in HP-filtered levels								
N = 50	.50	.11	.02	.01	.93	.89	.86	.80
N = 200	.10	.00	.00	.00	.94	.90	.82	.71
N = 400	.01	.00	.00	.00	.94	.89	.74	.56
N = 2000	.00	.00	.00	.00	.94	.78	.29	.06

Table 6 Impulse response function – empirical coverage of asymptotic 95% confidence interval (continued)

b Borderline-stationary (φ = .95) Generating processes – causality in partial differences (DGP II)

Estimated model	*Own innovation (ε_t)*				*X_t innovation (η_t)*			
	Lag 1	*Lag 2*	*Lag 3*	*Lag 4*	*Lag 1*	*Lag 2*	*Lag 3*	*Lag 4*
VAR in levels								
N = 50	.88	.74	.68	.64	.94	.90	.86	.84
N = 200	.94	.88	.85	.83	.95	.94	.92	.91
N = 400	.95	.92	.90	.89	.95	.95	.94	.94
N = 2000	.95	.94	.93	.92	.95	.95	.95	.95
VAR in levels with trend								
N = 50	.80	.58	.47	.42	.93	.88	.83	.80
N = 200	.93	.82	.75	.72	.94	.93	.91	.90
N = 400	.95	.89	.85	.83	.95	.95	.94	.93
N = 2000	.95	.94	.93	.92	.95	.95	.95	.95
VAR in differences								
N = 50	.95	.96	.97	.95	.94	.94	.96	.96
N = 200	.93	.87	.62	.28	.95	.95	.95	.95
N = 400	.92	.69	.25	.02	.95	.95	.95	.95
N = 2000	.78	.04	.00	.00	.96	.96	.96	.96
Lag-augmented VAR								
N = 50	.88	.74	.68	.64	.94	.90	.86	.84
N = 200	.94	.88	.85	.83	.95	.94	.92	.91
N = 400	.95	.92	.90	.89	.95	.95	.94	.94
N = 2000	.95	.94	.93	.92	.95	.95	.95	.95
Error correction model								
N = 50	.95	.96	.98	.96	.94	.94	.94	.96
N = 200	.93	.87	.63	.28	.95	.95	.96	.95
N = 400	.92	.69	.25	.02	.95	.96	.96	.95
N = 2000	.78	.04	.00	.00	.96	.96	.96	.96
VAR in HP-filtered levels								
N = 50	.59	.19	.09	.04	.93	.88	.86	.83
N = 200	.22	.00	.00	.00	.94	.89	.83	.76
N = 400	.05	.00	.00	.00	.94	.90	.79	.66
N = 2000	.00	.00	.00	.00	.95	.83	.44	.16

Table 6 Impulse response function – empirical coverage of asymptotic 95% confidence interval (continued)

c Borderline-stationary (φ = .95) Generating processes – causality in level (DGP III)

Estimated model	*Own innovation (ε_t)*				*X_t innovation (η_t)*			
	Lag 1	*Lag 2*	*Lag 3*	*Lag 4*	*Lag 1*	*Lag 2*	*Lag 3*	*Lag 4*
VAR in levels								
N = 50	.93	.86	.86	.85	.94	.94	.92	.88
N = 200	.94	.91	.90	.89	.95	.95	.95	.95
N = 400	.95	.92	.91	.91	.95	.95	.95	.95
N = 2000	.94	.94	.93	.93	.95	.96	.96	.95
VAR in levels with trend								
N = 50	.89	.77	.73	.68	.93	.92	.92	.85
N = 200	.94	.89	.87	.85	.95	.94	.94	.94
N = 400	.94	.90	.88	.87	.95	.95	.95	.94
N = 2000	.94	.93	.92	.92	.95	.96	.95	.95
VAR in differences								
N = 50	.52	.61	.62	.53	.93	.62	.34	.18
N = 200	.57	.43	.17	.04	.93	.45	.09	.01
N = 400	.83	.52	.13	.01	.94	.60	.21	.04
N = 2000	.76	.03	.00	.00	.94	.77	.50	.25
Lag-augmented VAR								
N = 50	.93	.86	.86	.85	.94	.94	.92	.88
N = 200	.94	.91	.90	.89	.95	.95	.95	.95
N = 400	.95	.92	.91	.91	.95	.95	.95	.95
N = 2000	.94	.94	.93	.93	.95	.96	.96	.95
Error correction model								
N = 50	.92	.94	.94	.91	.53	.16	.05	.02
N = 200	.92	.82	.53	.19	.48	.05	.00	.00
N = 400	.89	.62	.18	.01	.75	.29	.06	.01
N = 2000	.76	.03	.00	.00	.87	.64	.35	.15
VAR in HP-filtered levels								
N = 50	.68	.29	.15	.09	.93	.89	.80	.57
N = 200	.34	.00	.00	.00	.90	.85	.63	.25
N = 400	.11	.00	.00	.00	.85	.78	.41	.06
N = 2000	.00	.00	.00	.00	.51	.29	.00	.00

Table 6 Impulse response function – empirical coverage of asymptotic 95% confidence interval (continued)

d Cointegrated processes – x_t adjusts to y_t (DGP IV)

Estimated model	*Own innovation (ε_t)*				*X_t innovation (η_t)*			
	Lag 1	*Lag 2*	*Lag 3*	*Lag 4*	*Lag 1*	*Lag 2*	*Lag 3*	*Lag 4*
VAR in levels								
N = 50	.92	.79	.74	.71	.94	.89	.84	.79
N = 200	.94	.90	.89	.88	.95	.94	.93	.91
N = 400	.95	93	.92	.91	.95	.95	.94	.94
N = 2000	.95	95	.95	.94	.96	.96	.96	.95
VAR in levels with trend								
N = 50	.76	.43	.30	.23	.94	.88	.85	.85
N = 200	.90	.73	.61	.53	.95	.94	.92	.91
N = 400	.93	.83	.75	.69	.95	.95	.94	.93
N = 2000	.95	.93	.91	.89	.96	.96	.96	.95
VAR in differences								
N = 50	.95	.92	.94	.93	.94	.96	.96	.96
N = 200	.95	.94	.95	.95	.95	.96	.96	.96
N = 400	.95	.95	.95	.95	.95	.95	.95	.95
N = 2000	.95	.95	.95	.95	.95	.95	.95	.95
Lag-augmented VAR								
N = 50	.92	.79	.74	.71	.94	.89	.84	.79
N = 200	.94	.90	.89	.88	.95	.94	.93	.91
N = 400	.95	.93	.92	.91	.95	.95	.94	.94
N = 2000	.95	.95	.95	.94	.96	.96	.96	.95
Error correction model								
N = 50	.95	.94	.96	.96	.93	.95	.95	.96
N = 200	.95	.95	.95	.95	.95	.95	.95	.96
N = 400	.95	.95	.96	.96	.95	.95	.95	.95
N = 2000	.95	.95	.95	.95	.95	.95	.95	.95
VAR in HP-filtered levels								
N = 50	.50	.10	.03	.01	.93	.88	.87	.91
N = 200	.09	.00	.00	.00	.94	.84	.82	.82
N = 400	.01	.00	.00	.00	.92	.78	.73	.73
N = 2000	.00	.00	.00	.00	.84	.34	.23	.23

Table 6 Impulse response function – empirical coverage of asymptotic 95% confidence interval (continued)

e Cointegrated processes – y_t adjusts to x_t (DGP V)

Estimated model	*Own innovation (ε_t)*				*X_t innovation (η_t)*			
	Lag 1	*Lag 2*	*Lag 3*	*Lag 4*	*Lag 1*	*Lag 2*	*Lag 3*	*Lag 4*
VAR in levels								
N = 50	.91	.79	.77	.76	.94	.91	.87	.86
N = 200	.91	.91	.89	.89	.95	.95	.93	.92
N = 400	.95	.93	.92	.92	.95	.95	.94	.94
N = 2000	.95	.94	.94	.94	.95	.96	.95	.95
VAR in levels with trend								
N = 50	.83	.61	.52	.46	.93	.89	.86	.84
N = 200	.94	.86	.82	.80	.95	.94	.93	.93
N = 400	.95	.91	.89	.88	.95	.95	.94	.94
N = 2000	.95	.94	.94	.93	.95	.96	.96	.96
VAR in differences								
N = 50	.94	.95	.93	.77	.95	.89	.82	.70
N = 200	.86	.57	.14	.01	.95	.82	.55	.29
N = 400	.74	.19	.00	.00	.95	.72	.30	.07
N = 2000	.14	.00	.00	.00	.95	.17	.00	.00
Lag-augmented VAR								
N = 50	.91	.79	.77	.76	.94	.91	.87	.86
N = 200	.95	.91	.89	.89	.95	.95	.93	.92
N = 400	.95	.93	.92	.92	.95	.95	.94	.94
N = 2000	.95	.94	.94	.94	.95	.96	.95	.95
Error correction model								
N = 50	.90	.90	.85	.63	.87	.77	.66	.52
N = 200	.80	.46	.08	.00	.79	.48	.21	.07
N = 400	.67	.12	.00	.00	.68	.23	.00	.00
N = 2000	.08	.00	.00	.00	.11	.00	.00	.00
VAR in HP-filtered levels								
N = 50	.65	.25	.12	.07	.93	.89	.87	.75
N = 200	.31	.01	.00	.00	.93	.90	.83	.63
N = 400	.09	.00	.00	.00	.92	.89	.77	.44
N = 2000	.00	.00	.00	.00	.82	.75	.35	.02

Table 7 Impulse response function – empirical coverage of bootstrapped 95% confidence interval

a Unit root generating processes (DGP I)

	Own innovation (ε_t)				X_t innovation (η_t)			
Estimated model	*Lag 1*	*Lag 2*	*Lag 3*	*Lag 4*	*Lag 1*	*Lag 2*	*Lag 3*	*Lag 4*
VAR in levels								
N = 50	.79	.65	.54	.50	.95	.91	.86	.86
N = 200	.91	.80	.67	.58	.93	.94	.93	.94
N = 400	.95	.89	.80	.72	.95	.96	.96	.95
N = 2000	.95	.94	.93	.90	.97	.96	.96	.96
VAR in levels with trend								
N = 50	.75	.56	.40	.30	.93	.83	.73	.65
N = 200	.93	.82	.68	.55	.93	.91	.85	.79
N = 400	.95	.88	.77	.66	.95	.94	.92	.88
N = 2000	.94	.93	.89	.84	.97	.96	.94	.93
VAR in differences								
N = 50	.95	.96	.96	.96	.96	.96	.96	.96
N = 200	.96	.95	.96	.96	.93	.94	.94	.94
N = 400	.97	.96	.96	.96	.95	.95	.95	.96
N = 2000	.94	.95	.95	.95	.97	.96	.96	.96
Lag-augmented VAR								
N = 50	.80	.65	.54	.50	.94	.91	.86	.85
N = 200	.91	.79	.67	.58	.93	.94	.94	.94
N = 400	.94	.88	.80	.72	.95	.96	.95	.95
N = 2000	.94	.94	.92	.90	.97	.96	.96	.96
Error correction model								
N = 50	.82	.74	.68	.65	.94	.88	.82	.79
N = 200	.89	.86	.83	.79	.90	.91	.89	.88
N = 400	.90	.88	.87	.86	.92	.92	.91	.90
N = 2000	.89	.90	.89	.89	.94	.93	.94	.94
VAR in HP-filtered levels								
N = 50	.60	.31	.18	.12	.95	.93	.91	.89
N = 200	.12	.00	.00	.00	.93	.90	.84	.75
N = 400	.00	.00	.00	.00	.94	.88	.75	.59
N = 2000	.00	.00	.00	.00	.95	.76	.29	.05

Table 7 Impulse response function – empirical coverage of bootstrapped 95% confidence interval (continued)

b Borderline-stationary (φ = .95) Generating processes – causality in partial differences (DGP II)

Estimated model	*Own innovation (ε_t)*				*X_t innovation (η_t)*			
	Lag 1	*Lag 2*	*Lag 3*	*Lag 4*	*Lag 1*	*Lag 2*	*Lag 3*	*Lag 4*
VAR in levels								
N = 50	.89	.82	.73	.69	.94	.91	.84	.79
N = 200	.96	.94	.91	.89	.96	.95	.92	.90
N = 400	.97	.95	.93	.91	.95	.95	.94	.92
N = 2000	.94	.94	.94	.94	.95	.96	.96	.96
VAR in levels with trend								
N = 50	.85	.72	.60	.55	.94	.87	.83	.78
N = 200	.95	.93	.89	.86	.96	.94	.92	.89
N = 400	.97	.94	.93	.91	.94	.95	.93	.92
N = 2000	.93	.94	.94	.94	.95	.96	.97	.96
VAR in differences								
N = 50	.95	.93	.86	.73	.95	.95	.96	.96
N = 200	.95	.78	.49	.18	.95	.96	.95	.95
N = 400	.94	.61	.18	.01	.94	.95	.95	.95
N = 2000	.76	.03	.00	.00	.95	.95	.95	.92
Lag-augmented VAR								
N = 50	.89	.82	.73	.68	.95	.91	.85	.79
N = 200	.96	.94	.91	.88	.96	.95	.93	.90
N = 400	.96	.95	.93	.90	.94	.95	.94	.92
N = 2000	.94	.94	.93	.93	.95	.96	.97	.96
Error correction model								
N = 50	.83	.76	.72	.71	.73	.54	.45	.41
N = 200	.95	.90	.86	.83	.80	.55	.33	.23
N = 400	.96	.93	.89	.87	.81	.45	.21	.09
N = 2000	.93	.94	.93	.90	.54	.04	.00	.00
VAR in HP-filtered levels								
N = 50	.70	.43	.33	.29	.94	.93	.92	.89
N = 200	.29	.00	.00	.00	.95	.92	.86	.81
N = 400	.05	.00	.00	.00	.93	.90	.80	.69
N = 2000	.00	.00	.00	.00	.94	.83	.43	.16

Table 7 Impulse response function – empirical coverage of bootstrapped 95% confidence interval (continued)

c Borderline-stationary (φ = .95) Generating processes – causality in level (DGP III)

Estimated model	*Own innovation (ε_t)*				*X_t innovation (η_t)*			
	Lag 1	*Lag 2*	*Lag 3*	*Lag 4*	*Lag 1*	*Lag 2*	*Lag 3*	*Lag 4*
VAR in levels								
N = 50	.91	.89	.83	.77	.93	.91	.86	.80
N = 200	.95	.95	.93	.91	.96	.95	.94	.93
N = 400	.95	.95	.93	.93	.95	.95	.95	.94
N = 2000	.94	.95	.95	.94	.96	.96	.96	.96
VAR in levels with trend								
N = 50	.87	.76	.65	.54	.91	.85	.77	.69
N = 200	.94	.94	.91	.88	.96	.95	.94	.92
N = 400	.95	.95	.94	.91	.96	.95	.94	.94
N = 2000	.94	.95	.95	.95	.96	.96	.96	.96
VAR in differences								
N = 50	.88	.82	.71	.55	.94	.74	.53	.36
N = 200	.55	.31	.10	.02	.93	.46	.11	.01
N = 400	.31	.07	.01	.00	.91	.20	.01	.00
N = 2000	.00	.00	.00	.00	.67	.00	.00	.00
Lag-augmented VAR								
N = 50	.91	.88	.83	.77	.93	.91	.86	.80
N = 200	.95	.95	.94	.91	.96	.95	.94	.94
N = 400	.95	.95	.94	.93	.95	.95	.95	.94
N = 2000	.95	.95	.95	.94	.97	.96	.96	.96
Error correction model								
N = 50	.65	.71	.71	.72	.65	.71	.72	.76
N = 200	.92	.93	.93	.93	.89	.71	.49	.34
N = 400	.85	.88	.89	.91	.84	.31	.07	.03
N = 2000	.30	.36	.44	.55	.24	.00	.00	.00
VAR in HP-filtered levels								
N = 50	.76	.54	.42	.35	.92	.91	.84	.71
N = 200	.40	.03	.00	.00	.90	.85	.64	.29
N = 400	.13	.00	.00	.00	.87	.79	.43	.06
N = 2000	.00	.00	.00	.00	.50	.28	.01	.00

Table 7 Impulse response function – empirical coverage of bootstrapped 95% confidence interval (continued)

d Cointegrated processes – x_t adjusts to y_t (DGP IV)

Estimated model	*Own innovation (ε_t)*				*X_t innovation (η_t)*			
	Lag 1	*Lag 2*	*Lag 3*	*Lag 4*	*Lag 1*	*Lag 2*	*Lag 3*	*Lag 4*
VAR in levels								
N = 50	.87	.77	.68	.63	.95	.91	.89	.87
N = 200	.93	.86	.79	.73	.95	.95	.95	.95
N = 400	.96	.92	.87	.82	.95	.96	.96	.96
N = 2000	.94	.93	.91	.90	.96	.97	.96	.96
VAR in levels with trend								
N = 50	.80	.60	.43	.33	.93	.82	.70	.63
N = 200	.94	.87	.76	.68	.95	.92	.87	.82
N = 400	.96	.92	.87	.82	.96	.95	.92	.89
N = 2000	.95	.93	.90	.89	.96	.96	.95	.95
VAR in differences								
N = 50	.95	.95	.95	.95	.95	.95	.95	.95
N = 200	.96	.96	.96	.96	.95	.95	.95	.95
N = 400	.97	.96	.96	.96	.95	.95	.95	.95
N = 2000	.94	.94	.94	.94	.96	.96	.96	.96
Lag-augmented VAR								
N = 50	.87	.77	.68	.63	.96	.91	.89	.87
N = 200	.94	.86	.79	.72	.95	.95	.95	.95
N = 400	.96	.92	.87	.82	.95	.96	.96	.96
N = 2000	.94	.92	.91	.90	.96	.97	.96	.96
Error correction model								
N = 50	.91	.88	.85	.84	.95	.92	.90	.88
N = 200	.96	.96	.95	.94	.95	.96	.96	.95
N = 400	.97	.97	.97	.96	.95	.96	.96	.96
N = 2000	.94	.95	.93	.93	.96	.97	.96	.96
VAR in HP-filtered levels								
N = 50	.60	.28	.18	.13	.94	.93	.92	.92
N = 200	.11	.00	.00	.00	.93	.86	.85	.84
N = 400	.01	.00	.00	.00	.92	.81	.75	.74
N = 2000	.00	.00	.00	.00	.83	.34	.22	.22

Table 7 Impulse response function – empirical coverage of bootstrapped 95% confidence interval (continued)

e Cointegrated processes – y_t adjusts to x_t (DGP V)

	Own innovation (ε_t)				*X_t innovation (η_t)*			
Estimated model	*Lag 1*	*Lag 2*	*Lag 3*	*Lag 4*	*Lag 1*	*Lag 2*	*Lag 3*	*Lag 4*
VAR in levels								
N = 50	.88	.79	.72	.70	.94	.90	.89	.91
N = 200	.94	.89	.84	.81	.95	.95	.95	.95
N = 400	.96	.93	.90	.88	.96	.96	.96	.96
N = 2000	.95	.95	.95	.94	.96	.96	.97	.97
VAR in levels with trend								
N = 50	.86	.73	.59	.51	.91	.83	.74	.65
N = 200	.94	.93	.90	.87	.95	.95	.92	.88
N = 400	.97	.95	.93	.91	.96	.95	.94	.93
N = 2000	.95	.95	.94	.94	.96	.96	.97	.96
VAR in differences								
N = 50	.96	.87	.67	.39	.94	.89	.84	.75
N = 200	.86	.42	.09	.00	.96	.81	.57	.31
N = 400	.73	.11	.00	.00	.97	.74	.32	.06
N = 2000	.14	.00	.00	.00	.95	.15	.00	.00
Lag-augmented VAR								
N = 50	.88	.79	.71	.69	.94	.90	.89	.90
N = 200	.94	.89	.84	.81	.95	.95	.95	.95
N = 400	.96	.93	.90	.89	.96	.95	.95	.96
N = 2000	.95	.95	.95	.94	.95	.96	.97	.97
Error correction model								
N = 50	.92	.88	.86	.86	.95	.92	.90	.90
N = 200	.95	.94	.93	.93	.96	.96	.96	.95
N = 400	.97	.96	.95	.95	.96	.95	.96	.96
N = 2000	.95	.96	.96	.96	.96	.96	.97	.97
VAR in HP-filtered levels								
N = 50	.74	.49	.39	.34	91	.91	.91	.85
N = 200	.36	.02	.00	.00	.93	.91	.86	.68
N = 400	.00	.00	.00	.00	.93	.90	.78	.48
N = 2000	.00	.00	.00	.00	.82	.76	.31	.01