General–to–Specific Model Selection Procedures for Structural Vector Autoregressions

Hans-Martin Krolzig *

Department of Economics and Nuffield College, Oxford University.

hans-martin.krolzig@nuf.ox.ac.uk

January 30, 2001

Abstract

Structural vector autoregressive (SVAR) models have emerged as a dominant research strategy in empirical macroeconomics. Despite their advantages, just-identified SVAR models suffer from (i) the great number of parameters ("curse of dimensionality"), (ii) the resulting uncertainty associated with impulse responses, (iii) the existence of alternative observationally-equivalent just-identified models and (iv) the lack of identification of the imposed causal ordering of the variables of the system. In this paper we propose general-to-specific reductions of just-identified SVAR models to overcome these limitations. We show that the computer-automated model selection algorithm embodied in *PcGets* (see Krolzig and Hendry, 2001) can be used for an efficient implementation of the underlying methodology. Since jointly selecting and diagnostic testing eludes theoretical analysis, we evaluate the proposed reduction and identification strategy by simulation. The application of the proposed reduction strategy to a US monetary system demonstrates the feasibility of *PcGets* for the analysis of large macroeconomic data sets.

JEL Classification: C51, C32, E52.

Keywords: Model selection; Vector autoregression; Subset model; Lag order determination; Data mining.

1 Introduction

Over the last two decades, vector autoregressive (VAR) models have emerged as an important research tool for the empirical analysis of macroeconomic time series, partly because of the critique in Sims (1980) of traditional macro-econometric modelling. VARs have widely exploited for the description of numerous macroeconomic data sets allowing fruitful insight on the interrelations between economic variables.

The popularity of VARs is due to various advantages of the approach: First, the flexibility of the VAR framework in producing econometric models with useful descriptive characteristics, within statistical tests of economically meaningful hypothesis can be executed. Secondly, the ease of the approach as econometric models can be formulated and data characterized without having to invoke economic theory to restrict the dynamic relations between variables. Thirdly, many completely specified economic models give rise to VAR representations as the reduced form of the variables of the system. Thus VARs do not only describe the data, but also allow the characterisation of macroeconomic models. Fourthly,

*Financial support from the UK Economic and Social Research Council under the grant L138251009 is gratefully acknowledged. The computations reported in this paper were carried out with PcGets1, see Hendry and Krolzig (2001), and Ox3, see Doornik (2001). Address of the author: Hans-Martin Krolzig, University of Oxford, Department of Economics, Manor Road Building, Oxford OX1 3UQ, UK.

the VAR model may be compatible with quite a wide variety of hypotheses regarding the formation of expectations, since in the reduced form of the system adaptive and rational expectations only differ in the weights applied to the variables representing dynamics (see, *inter alia*, the analysis of rational expectation models in Hansen and Sargent, 1981, and the real business cycle model in King, Plosser and Rebelo, 1988).

Reduced-form VARs have the drawback that in general the residuals are correlated across equations (as in the case of instantaneous causality), thus their innovations are not identified with the underlying structural errors. Therefore the impulse responses generated by a VAR do not possess a structural interpretation. There is no unique best way to deal with this problem. Since Sims (1980), a popular way of overcoming the problem is to transform the residuals to orthogonal form by triangulating the system which involves a causal ordering of the variables. This transformation of the VARs allows to interpret the evolution of the system as a function of orthogonalized innovations in the variables of the system.

A different approach to respond to the problem of interpreting VARs has been the development of structural vector autoregressions (SVARs) which introduces ‘theoretical’ restrictions to identify the underlying shocks. In this paper we focus on short-run restrictions specifying the causal ordering of the variables in the system (in contrast to long-run restrictions in the spirit of Blanchard and Quah, 1989).

Despite these advantages, SVAR models suffer from the fact that they are generally made recursive in contemporaneous variables and rarely impose any restrictions upon the dynamics in their implied structural equations. For example, the relatively small structural vector autoregressive (SVAR) models of Bernanke and Blinder (1992) and Sims (1992) for the US economy have the distinctive feature that each structural equation is saturated with lagged variables i.e. the dynamics are essentially unrestricted. This means that in a just-identified SVAR, the number of parameters grows with the square of the number of variables and quickly exhausts the degree of freedom. Having huge numbers of parameters, the structural equations of the SVAR are hard to interpret. This is unfortunate since it seems sensible to employ model reduction procedures which so far have been mainly used in the single-equation framework. Accordingly, these considerations point to the need for reductions of the systems which involve the utilization of exclusion restrictions upon the dynamics contained in each structural equation, so as to allow for easier interpretation of the system. This paper considers computer-automated simplification procedures which are based on the outcome of statistical tests.[1]

The SVAR approach tends to impose just enough restrictions to permit a coherent interpretation of the shocks to the system. While models which impose over-identifying restrictions can be tested, this is not the case for just-identifying restrictions. Thus other observationally-equivalent just-identified SVARs exist (see Pagan, 1994). This is related to the well-known fact that orthogonalized (structural) impulse-responses, which are based on a Choleski decomposition of the variance-covariance matrix of the reduced-form VAR, are not invariant against changes in the (causal) ordering of the variables.

It is also worth noting that the use of term ‘structural’ is not unproblematic. The types of restrictions used to identify VARs have been criticized as theoretically and empirically misleading: From a theoretical point of view, the identifying restrictions impose significant structure on a model without being firmly grounded in an economic model. Indeed, they can be incompatible with large classes of economic models (see, *inter alia*, Cooley and Dwyer, 1998). From an empirical point of view, the restrictions are imposed without establishing the congruence with the data (e.g. the constancy of the estimated coefficients (see, *inter alia*, Hendry and Mizon, 2000). Furthermore, the generality of the SVAR is not realized without assumptions regarding the absence of regime changes and the linear formation of expectations etc. For the economic interpretation of the model, it is therefore essential to ensure the

[1] An excellent empirical illustration of a modelling approach which combines simplifications based on statistical tests and economic considerations can be found in Dungey and Pagan (2001).

congruence of the assumptions made. It is a merit of the modelling approach proposed in this paper that it systematically checks for the presence of statistical misspecifications.

In this paper we propose general-to-specific reductions of just-identified SVAR models to overcome the limitations of just-identified SVAR models: (i) the great number of parameters ("curse of dimensionality"), (ii) the existence of alternative observationally-equivalent just-identified models and (iii) the lack of identification of the imposed causal ordering of the variables of the system. We will repeatedly argue that the reduction and identification of SVAR models is a natural area for the application of general–to–specific (*Gets*) reduction procedures. The *Gets* reduction process is designed to ensure that the reduced SVAR model will convey all the information embodied in the unrestricted SVAR. This is achieved by a joint selection and diagnostic testing process: starting from the unrestricted, congruent general model, standard testing procedures are used to eliminate statistically-insignificant variables, with diagnostic tests checking the validity of reductions, ensuring a congruent final selection. By reducing the complexity of the just-identified SVAR and checking the contained information, the selected simpler, more compact model provides an improved statistical description of the economic world (see Hendry, 1993, for an overview of the so-called 'LSE' methodology).

The existing literature on VAR model selection has mainly focused on the selection of lag order, p, of an otherwise unrestricted reduced-form VAR. In these selection procedures a model is usually selected by an information criterion which penalizes the likelihood function for the number of parameters.[2] Lütkepohl (1991) who discusses various strategies for selecting subset VAR models (i.e., VARs with zero constraints on the coefficients), which are based on the optimization of a specified model selection criterion. For a given maximal order p of VAR, a full search over all possible candidates is computationally unfeasible: in a K-dimensional VAR(p) without deterministic terms there are K^2p coefficients, any full search requires the estimation of a total of 2^{K^2p} subset models. Therefore various strategies have been proposed to overcome this problem (search over complete VAR matrices, top-down and bottom-up specification of the distributed lag lengths etc.). Brüggemann and Lütkepohl (2000) consider step-wise regression type single-equation reduction paths where the critical value is chosen such that an acceptance of the null hypothesis guarantees a marginal increase in a given information criterion. The recent developments in automatic model selection initiated by Hoover and Perez (1999) suggests that the operational characteristics of some computer-automated model selection algorithms are excellent across a wide range of states of nature. Naturally we will focus on the implementation of computer–automated model selection procedures using *PcGets* developed by Hendry and Krolzig (2001).

The structure of the paper is as follows: The following section (§2) introduce the VAR and the SVAR models. §3 discusses the 'General-to-specific' (*Gets*) approach to econometric modelling, the underlying theory of reduction and the computer automation of general-to-specific model selection procedures by *PcGets*. The application of the *Gets* approach to the reduction of reduced-form VAR models is considered in §4, which puts emphasis on the case of conditional independence. It will be shown that under this condition, single–equation model selection procedures such as *PcGets* are efficient. We then move forward to the reduction of SVAR models in §5 which will emerge as a very convenient framework for the application of *PcGets*. In §6 we discuss the use of the proposed modelling approach for select-

[2] The information criteria considered in the literature are defined as follows:

$$\begin{aligned} AIC &= -2L_T/T + 2n/T, \\ SC &= -2L_T/T + n\log(T)/T, \\ HQ &= -2L_T/T + 2n\log(\log(T))/T, \end{aligned}$$

where L_T is the maximized log-likelihood, n is the number of parameters and T is the sample size: see Akaike (1985), Schwarz (1978), and Hannan and Quinn (1979).

ing the causal ordering of the variables of the system. In §7 we investigate by simulation whether the *Gets* model-selection process works well or fails badly in the SVAR framework. Results are presented for a Monte Carlo experiment where the data generating process (DGP) is an over-identified trivariate SVAR(1) and the general unrestricted model (GUM) is an SVAR(5) or VAR(5). The focus will be on the selection properties as well as the precision and accuracy of the implied impulse-responses. The empirical illustration with a US monetary system presented in §8 evaluates the usefulness of proposed approach for the analysis of large macroeconomic data sets. In §9 we outline generalizations of the proposed reduction approach for cointegrated time series models and (just- or overidentified) simultaneous equation models. §10 concludes.

2 The vector autoregressive model

2.1 The reduced-form VAR

The basic model considered in the following is a vector autoregression possibly including deterministic terms and with independent Gaussian errors: the n-dimensional time series vector y_t is generated by a vector autoregressive process of order p, denoted VAR(p) model,

$$y_t = \nu + \sum_{i=1}^{p} A_i y_{t-i} + \varepsilon_t, \tag{1}$$

where $t = 1, \ldots, T$, the A_i and ν are coefficient matrices and the initial values of $Y_0 = (y_0, \ldots, y_{1-p})$ are fixed. The innovation process ε_t is an unobservable Gaussian zero-mean white noise process with a time-invariant positive-definite variance-covariance matrix $\mathsf{E}[\varepsilon_t \varepsilon_t'] = \Sigma$:

$$\varepsilon_t \sim \mathsf{NID}(\mathbf{0}, \Sigma). \tag{2}$$

The infinite-order vector moving-average representation of the VAR in (1) is

$$y_t = \mu + \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j}, \tag{3}$$

where $\mu = \left(\sum_{i=1}^{p} A_i\right)^{-1} \nu$ and where $\Psi_0 = I_K$. $\Psi(L) = \left(1 - \sum_{j=i}^{p} A_i L^i\right)^{-1}$, such that for VAR(1) processes, $\Psi_j = A^j$. The (k,l)-th element $\psi_{kl,j}$ of the MA matrix Ψ_j can be interpreted as the reaction of variable k in response to a unit shock in variable l, j periods ago.

A problematic assumption in this type of impulse response analysis is that the shock occurs only in one of the variables. Such an assumption might be reasonable in the case of conditional independence. If the residuals are correlated, the VAR can be readily transformed to interpret the evolution of the system as a function of orthogonalized innovations in any of the variables . Define $\eta_t^* = P^{-1}\varepsilon_t$ by decomposing Σ as $\Sigma = PP'$, where P is a lower triangular matrix, such that $\eta_t^* \sim \mathsf{NID}(0, I_K)$. Note that by selecting P, a causal ordering of the variables is implied.

Thus the orthogonalized vector moving average representation is given by:

$$y_t = \mu + \sum_{j=0}^{\infty} \Psi_j P P^{-1} \varepsilon_{t-j} = \mu + \sum_{j=0}^{\infty} \Phi_j^* \eta_{t-j}^*, \tag{4}$$

where $\Phi_0^* = P$ and $\Phi_j^* = \Psi_j P$. It is worth pointing out that the orthogonalized impulse responses depend on the choice of P.

A major objective of time series analysis is the creation of predictions. It is convenient to choose the optimal predictor as the minimizer of the mean squared prediction error (MSPE). Using the reduced-form VAR in (1), the MSPE-optimal prediction of y_{T+h} conditional on the information set $Y_T = (y_T, y_{T-1}, \ldots, y_{1-p})$ is given by the conditional expectation:

$$\mathsf{E}[y_{T+h}|Y_T] = \nu + \sum_{i=1}^{p} A_i \mathsf{E}[y_{T+h-i}|Y_T],$$

which can be solved recursively with the initial conditions $\mathsf{E}[y_{T+h-i}|Y_T] = y_{T+h-i}$ for $i \geq h$ (see, *inter alia*, Lütkepohl, 1991, for further details).

2.2 The structural VAR

The structural vector autoregressive process of order p, denoted SVAR(p) model, considered in this paper is given by

$$By_t = \delta + \sum_{i=1}^{p} \Gamma_i y_{t-i} + \eta_t, \tag{5}$$

where $t = 1, \ldots, T$, the Γ_i and δ are coefficient matrices and the initial values of $Y_0 = (y_0, \ldots, y_{1-p})$ are fixed. The innovation process η_t is an unobservable Gaussian zero-mean white noise process with a time-invariant diagonal variance-covariance matrix $\mathsf{E}[\eta_t \eta_t'] = \Omega$:

$$\eta_t \sim \mathsf{NID}(\mathbf{0}, \Omega). \tag{6}$$

The SVAR in equation 5 can be considered a particular simultaneous equation model in the spirit of the *Cowles approach*. Particularly, it is a recursive system of the sort proposed by Wold (1949) and Strotz and Wold (1960) and closely related to the concept of causal ordering introduced by Simon (1953).

To recover the structural parameters it is useful to consider the structural VAR in equation (5) in its reduced-form (1). The relation to the reduced form VAR is given by:

$$\begin{aligned} \Sigma &= B^{-1}\Omega B^{-1\prime}, \\ A_i &= B^{-1}\Gamma_i \qquad \text{for } i = 1, \ldots, p, \\ \nu &= B^{-1}\delta. \end{aligned}$$

The uniqueness of the model for the given structure, which guarantees the estimatability of the structural parameters, is ensured by the following just-identifying restrictions:

$$\begin{aligned} \Omega_{ij} &= 0 \qquad \text{for } i \neq j, \\ B_{ii} &= 1 \qquad \text{for } i = 1, \ldots, n, \\ B_{ij} &= 0 \qquad \text{for } i < j. \end{aligned} \tag{7}$$

The infinite-order structural vector moving-average representation results from (3) by $\varepsilon_{t-j} = B^{-1}\eta_{t-j}$ as:

$$y_t = \mu + \sum_{j=0}^{\infty} \Psi_j B^{-1} \eta_{t-j} = \mu + \sum_{j=0}^{\infty} \Phi_j \eta_{t-j}, \tag{8}$$

where $\Phi(L) = \left(B - \sum_{i=1}^{p} \Gamma_i L^i\right)^{-1}$ with $\Phi_0 = B^{-1}$. It is worth noting that for just-identified SVARs and unrestricted VARs with the same causal ordering, (8) differs from (4) only by the missing adjustment for the standard errors of the η_{kt}. In other words, the relation to (4) is given by $\eta_t^* = \Omega^{-\frac{1}{2}}\eta_t$.

3 Computer-automated general-to-specific model selection procedures

3.1 General-to-specific modelling

Empirical econometric modelling is an integral aspect of the attempt to make economics a quantitative science, although it raises many important methodological issues. Not least among these is how to select models. Economies are so high dimensional, non-stationary, and complicated that pure theory could never precisely specify the underlying process, and there are simply too many variables to rely solely on data evidence. Thus, model-selection methods must be used, and the methodology thereof deserves careful attention.

When the prior specification of a possible relationship is not known for certain, data evidence is essential to delineate the relevant from the irrelevant variables. Thus, model selection is inevitable in practice; and while that may be accomplished in many possible ways, we follow Hendry and Krolzig (2001) in arguing that simplification from a congruent GUM embodies the best features of a number of alternatives.

Unfortunately, there has been little agreement on which approaches should be adopted: see, *inter alia*, Leamer (1983b), Pagan (1987), Hendry, Leamer and Poirier (1990), Granger (1990) and Magnus and Morgan (1999). Hendry (2000) discusses the rather pessimistic perceptions extant in the literature, including difficulties deriving from ‘data-based model selection’ (see the attack by Keynes, 1939, 1940, on Tinbergen, 1940a, 1940b), ‘measurement without theory’ (Koopmans, 1947), ‘data mining’ (Lovell, 1983), pre-test biases’ (Judge and Bock, 1978), ‘ignoring selection effects’ (Leamer, 1978), ‘repeated testing’ (Leamer, 1983a), ‘arbitrary significance levels’ (Hendry *et al.*, 1990), ‘lack of identification’ (see Faust and Whiteman, 1997, for a recent reiteration), and the potential ‘path dependence of any selection’ (Pagan, 1987). Nevertheless, none of these problems is inherently insurmountable, most are based on theoretical arguments (rather than evidence), and most have counter criticisms. Instead, the sequence of developments in automatic model selection initiated by Hoover and Perez (1999) suggests the converse: the operational characteristics of some selection algorithms are excellent across a wide range of states of nature, as Hendry and Krolzig (2001) demonstrated.

General-to-specific approaches have a long pedigree: see *inter alia*, Anderson (1971), Sargan (1973, 1981), Mizon (1977b, 1977a), and Hendry (1979). White (1990) showed that with sufficiently-rigorous testing, the selected model will converge to the data generating process (DGP). Thus, any ‘overfitting’ and mis-specification problems are primarily finite sample. Moreover, Mayo (1981) emphasized the importance of diagnostic test information being effectively independent of the sufficient statistics from which parameter estimates are derived. Also, Hendry (1995) argued that congruent models are the appropriate class within which to search, that encompassing resolves many instances of ‘data mining’, and that in econometrics, theory dependence has as many drawbacks as sample dependence, so modeling procedures are essential.

Gets is the practical embodiment of the theory of reduction: see e.g., Florens and Mouchart (1980), Hendry and Richard (1983), Florens, Mouchart and Rolin (1990), Hendry (1987), and Hendry (1995, ch. 9). That theory explains how the data generation process (DGP) characterizing an economy is reduced to the ‘local’ DGP (LDGP), which is the joint distribution of the subset of variables under analysis. By operations of aggregating, marginalizing, conditioning, sequentially factorizing, and transforming, a potentially vast initial information set is reduced to the small transformed subset that is relevant for the problem in question. These reduction operations affect the parameters of the process, and thereby determine the properties of the LDGP.

3.2 The theory of reduction

The theory of reduction not only explains the origins of the LDGP, the possible losses of information from any given reduction can be measured relative to its ability to deliver the parameters of interest in the analysis. For example, inappropriate reductions, such as marginalizing with respect to relevant variables, can induce non-constant parameters, or autocorrelated residuals, or heteroscedasticity and so on. The resulting taxonomy of possible information losses highlights six main null hypotheses against which model evaluation can be conducted, relating to the past, present, and future of the data, measurement and theory information, and results obtained by rival models. A *congruent* model is one that matches the data evidence on all the measured attributes: the DGP can always be written as a congruent representation. Congruence is checked in practice by computing a set of mis-specification statistics to ascertain that the residuals are approximately homoscedastic, normal, white noise, that any conditioning is valid, and that the parameters are constant. Empirical congruence is shown by satisfactory performance on all these checks. A model that can explain the findings of other models of the same dependent variables is said to encompass them, see e.g., Cox (1961, 1962), Hendry and Richard (1982), Mizon (1984), Mizon and Richard (1986), and Hendry and Richard (1989). Since a model that explains the results of a more general model within which it is nested parsimoniously encompasses the latter (see Govaerts, Hendry and Richard, 1994, and Florens, Hendry and Richard, 1996), when no reductions lose relevant information, the LDGP would parsimoniously encompass the DGP with respect to the subset of variables under analysis.

To implement *Gets*, a general unrestricted model (GUM) is formulated to provide a congruent approximation to the LDGP, given the theoretical and empirical background knowledge. Importantly, Bontemps and Mizon (2001) show that an empirical model is congruent if it parsimoniously encompasses the LDGP. The empirical analysis commences from this GUM, after testing for mis-specifications, and if no problems are apparent, the GUM is simplified to a parsimonious, congruent model, each simplification step being checked by diagnostic testing. Simplification can be done in many ways: and although the goodness of a model is intrinsic to it, and not a property of the selection route, poor routes seem unlikely to deliver useful models.

3.3 The computer automation of *Gets*

While the joint issue of sequential variable selection and diagnostic testing using multiple criteria has eluded most attempts at theoretical analysis, an evaluation of the properties of the model-selection process can be achieved by simulation. Hoover and Perez (1999) reconsidered the Lovell (1983) experiments to evaluate the performance of *Gets*. Most important is their notion of commencing from the congruent general model by following a number of reduction search paths, terminated by either no further feasible reductions or significant diagnostic tests occurring. Hoover and Perez select among the surviving models the one which fits best. They show how much better a structured approach is than any method Lovell considered, suggesting that modeling *per se* need not be bad. Indeed, overall, the size of their selection procedure is close to that expected, and the power is reasonable. Moreover, re-running their experiments using *PcGets*, Hendry and Krolzig, 1999 found substantively better outcomes.

Naturally, we focus on *PcGets*: see Hendry and Krolzig (1999, 2001) and Krolzig and Hendry (2001). *PcGets* implements automatic general-to-specific (*Gets*) modelling for linear, dynamic, regression models based on the principles discussed in Hendry (1995). First, an initial general statistical model is tested for the absence of mis-specification (denoted congruence), which is then maintained throughout the selection process by diagnostic checks, thereby ensuring a congruent final model. The diagnostic tests require careful choice to ensure they characterize the salient attributes of congruency,

Table 1 The *PcGets* algorithm.

Stage 0. **Pre-search reductions**

(1) Estimation and test of the GUM;
(2) Outlier correction;
(3) Adjust significance level of diagnostics;
(4) Lag order pre-selection;
(5) Sort variables in order of their t^2 values

- Two-step top-down reduction;
- Bottom-up reduction;
- Encompassing.

Stage I. **Multiple model reduction paths:**

(1) Sequential estimation and test of reductions

(a) Remove insignificant variables.
(b) Model reductions are subjected to a wide range of diagnostic tests:

- Chow tests for structural stability;
- residual autocorrelation;
- ARCH effects in the residuals;
- normality;
- heteroscedasticity.

(2) Encompassing

Stage II. **Union testing**

(1) Estimation and test of the new GUM;
(2) Multiple model reduction paths;
(3) Encompassing and final model selection.

Stage III. **Sub-sample evaluation**

(1) Test the significance of every selected variable in two overlapping sub-samples;
(2) Penalize variable accordingly.

are correctly sized, and do not overly restrict reductions. Next statistically-insignificant variables are eliminated by selection tests, both in blocks and individually. Many reduction paths are searched, to prevent the algorithm from becoming stuck in a sequence that inadvertently eliminates a variable which matters, and thereby retains other variables as proxies. Path searches in PcGets terminate when no variable meets the pre-set criteria, or any diagnostic test becomes significant. Non-rejected models are tested by encompassing: if several remain acceptable, so are congruent, undominated, mutually-encompassing representations, the reduction process recommences from their union, providing that is a reduction of the GUM, till a unique outcome is obtained: otherwise, or if all selected simplifications re-appear, the search is terminated using the Schwarz (1978) information criterion. Lastly, sub-sample insignificance seeks to identify ‘spuriously significant’ regressors.

An overview of the algorithm is shown in Table 1, see Hendry and Krolzig (2001) for details. In the following we briefly discuss the econometrics of the different stages of the *PcGets* model-selection algorithm relevant for VAR modelling.

3.3.1 The GUM and pre-search tests (Stage 0)

The starting point for *Gets* model-selection is the GUM, so the key issues concern its specification and congruence. In the case of the VAR, the researcher has to specify the order and the dimension of the process. An overall F-test of all regressors checks that there is something to model, misspecification tests check the congruence of the model. *PcGets* then undertakes various ‘pre-search’ simplification F-tests to exclude variables from the GUM. Since variables found to be irrelevant on such tests are excluded from later analyses, this step uses a loose significance level (such as 50%). The lag order pre-selection consists of F-tests on the longest-lag blocks till the null is rejected. The next step consists of block (F) tests of groups of variables, ordered by their t^2-values in the GUM. In the *top-down* reduction sequence the t^2-test statistics are ordered from the smallest up, with cumulative F-tests on increasing block sizes till the null is rejected; the model size decreases until rejection. The *bottom-up* reduction sequence involves F-tests on decreasing block sizes from the largest t^2-tests down till the model is congruent. The model size increases until no misspecifications are found. According to the outcome of a block F test, *PcGets* will continue to work with one of the reductions.

3.3.2 Multi-stage multi-path search (Stages I/II)

The *PcGets* reduction path relies on a classical, sequential simplification and testing approach designed to reduce the complexity of the model by ensuring the congruency of the reduction. Many possible paths from that GUM are investigated: reduction paths considered include both multiple deletions as well as single, so t and/or F test statistics are used as simplification criteria. Along each path the least significant variable having a t-values less than the critical value is eliminated. If any diagnostic tests fail, that path is terminated, and the algorithm returns to the last accepted model of the search path: if the last accepted model cannot be further reduced, it becomes the terminal model of the particular search path; otherwise, the last removed variable is re-introduced, and the search path continues with a new reduction by removing the next least-insignificant variable of the last accepted model. If all tests are passed, but one or more variables are insignificant, the least significant variable of those is removed. If that specification has already been tested on a previous path, the current search path is terminated. Finally, if all diagnostic tests are passed, and all variables are significant, the model is the terminal model of that search path. Should multiple congruent contenders eventuate after a reduction round, encompassing can be used to test between them, with only the surviving non-nested specifications retained. If multiple models survive the ‘*testimation*’ process, their union forms a new general model, and selection path searches recommence. Such a process repeats till a unique contender emerges, or the previous union is reproduced, then stops. In the latter case a final selection is made using information criteria, otherwise a unique congruent and encompassing reduction has been located.

3.3.3 Sub-sample evaluation (Stage III)

As a check for potential over-selection in *Stage II*, *PcGets* exploits sub-sample information by investigating split samples for significance (as against constancy). This mimics the idea of recursive estimation: Since non-central ‘t’-values diverge with increasing sample size, whereas central ‘t’s fluctuate around zero, the latter have a low probability of exceeding any given critical value in two sub-samples, even when those sample overlap. Thus, adventitiously-significant variables may be revealed by their insignificance in one or both of the sub-samples. Consequently, a progressive research strategy can gradually eliminate ‘adventitiously-significant’ variables and tilt the size-power balance favorably. The sub-sample information is used to accord a ‘reliability’ score to variables, which investigators may use to guide their model choice.

3.3.4 Calibration

Balancing the objectives of small size and high power still involves a trade-off, but one that is dependent on the algorithm. The *'testimation'* process of *PcGets* depends on a number of decisions regarding the specification of the algorithm. Krolzig and Hendry (2001) investigate the calibration of *PcGets* with regard to the operational characteristics of the diagnostic tests, the selection probabilities of DGP variables, and the deletion probabilities of non-DGP variables. Based on intensive Monte-Carlo studies, Hendry and Krolzig (2001) propose a *'liberal'* and a *'conservative'* strategy which aim to provide maximum power at a controlled empirical size of 1% and 5%. These will be used in the following Monte Carlo experiments.

In the Monte Carlo experiments in Hendry and Krolzig (1999) and Krolzig and Hendry (2001), PcGets recovers the data generation process (DGP) with an accuracy close to what one would expect if the DGP specification were known, but nevertheless coefficient tests were conducted. Empirically, on the data sets analyzed by Davidson, Hendry, Srba and Yeo (1978) and Hendry and Ericsson (1991), PcGets selects (in seconds!) models at least as good as those developed over several years by their authors. Automatic model selection is in its infancy, yet already exceptional progress has been achieved, setting a high 'lower bound' on future performance. Moreover, there is a burgeoning symbiosis between the implementation and the theory – developments in each stimulate advances in the other.

So far section we described the model selection algorithm of *PcGets* as it has been developed by Hendry and Krolzig (2001) for linear single-equation models when the precise formulation of the economic system under analysis is not known. In the following we investigate the generalization of the algorithm for the reduction of highly parameterized SVAR models.

4 General-to-specific reductions in reduced-form VAR models

In the following we investigate *Gets* reductions of reduced-form VAR(p) processes as defined in equation (1). The GUM is an unrestricted VAR(p) model and the unknown DGP is a subset of the unrestricted VAR. Such systems can be analyzed one equation at a time, since every equation has the same set of regressors, but each variable is the regressand in turn. Thus, VAR modelling is a natural area for the application of *PcGets*. By reducing the complexity of the VAR and simultaneously ensuring that the parsimonious subset VAR will convey all the information embodied in the unrestricted VAR, the selected simpler model should provide an improved statistical description of the economic world. As *PcGets* has been developed for single-equation models, the critical issue is the potential loss in efficiency by analyzing the equations of a VAR once at a time using single-equation model selection algorithms rather than analyzing the system.

First, consider the case where the variance-covariance matrix of the system is diagonal, such that the equations of the VAR are unrelated to each other:

Proposition 1 (Reduction under conditional independence). *Suppose that in reduced-form VAR of (1), the variance-covariance matrix Σ is diagonal, i.e. all $\sigma_{ij} = 0$ for $i \neq j$.. Then, all possible reductions of the system can be efficiently estimated by OLS, and model-selection procedures can operate equation-by-equation without a loss in efficiency.*

Proof. Conditional independence of y_t conditional on its past Y_{t-1} allows the factorization of the probability density function of y_t in terms of its marginals:

$$f_y(y_t|Y_{t-1}) = \prod_{k=1}^{n} f_{y_k}(y_{kt}|Y_{t-1}).$$

This implies that the log-likelihood function $L_T(\theta)$ can be separated with regard to the parameters of interest, θ_k, of each equation $k = 1, \ldots, K$ of the system which can be varied freely:

$$L_T(\theta) = \sum_{k=1}^{K} \left\{ \sum_{t=1}^{T} \ln f_{y_k}(y_{kt}|Y_{t-1}; \theta_k) \right\}.$$

Consequently, all possible reductions of the system can be (asymptotically) efficiently estimated by single equation methods (OLS under normality), and reduction procedures can be applied equation-by-equation without a loss of (asymptotic) efficiency. ■

The great advantage of separability of the log-likelihood function is that the model space generated by the general unrestricted model (GUM), which is of dimension $2^{K(Kp+1)}$, can be searched in K subspaces of dimension 2^{Kp+1}. This is an econometrically and computationally attractive scenario.

This has strong implications for model selection procedures. Proposition 1 states that the efficiency of single-equation model selection algorithms depends on the absence of instantaneous causality. In other words, if the variance-covariance matrix of the system is diagonal, i.e., all $\sigma_{ij} = 0$ for $i \neq j$, *PcGets* can be used to model the system as in the single-equation framework it has been designed for.

Under the presence of instantaneous causality between the variables, i.e., some $\sigma_{ij} \neq 0$ for $i \neq j$, the separability property of the log-likelihood function is lost: the equations of the VAR are only seemingly unrelated to each other. Since eliminating a variable in one equation affects the others, single-equation model selection procedures are inefficient. This is directly related to the properties of the OLS estimation method: In contrast to full information maximum likelihood (FIML) and estimated generalized least squares (EGLS), OLS is an inefficient estimation method for subset VARs with non-diagonal variance-covariance matrices.

Krolzig (2001b) discusses generalizations of the *PcGets* algorithm for the analysis of VAR models summarized in table 2. System procedures involve (i) joint reductions of the system and (ii) reductions of the individual equations. In case of the reductions of the system we are interested in a system analysis of cross-equation restrictions whose acceptance would exclude a regressor from all equations of the system. Path reduction, for example, would search the system checking the coefficient with the lowest remaining t^2-value of the *system*:

$$(k^*, j^*, i^*) := \arg \min_{k=1,\ldots,K} \min_{j=1,\ldots,K} \min_{i=1,\ldots,p} \mathsf{t}^2_{\mathsf{kj,i}},$$

instead of checking the coefficient with the lowest remaining t-value of the k-th *equation*

$$(j^*, i^*) := \arg \min_{j=1,\ldots,K} \min_{i=1,\ldots,p} \mathsf{t}^2_{\mathsf{kj,i}}.$$

If the coefficient a_{kj^*,i^*} of regressor $y_{j,t-i}$ in equation k is insignificant, the coefficient is restricted to zero and the equation is re-estimated by OLS. Alternatively, EGLS could be used whereby the variance-covariance matrix is taken from the reduced, but otherwise unrestricted system.

Table 2 *Gets* Algorithm for stable VAR models under instantaneous causality.

(1) **Reductions of the system**

System analysis of joint restrictions (OLS): #state vector $= n$

- Presearch for the exclusion of blocks of variables from the system
- F-test search for the exclusion of single variables from the system
- Diagnostics for the vector of residuals

(2) **Reductions of the equations**

PcGets-style multi-stage multi-path model selection

- Pre-search tests
- Multiple model reduction paths:
  - Sequential estimation and test of reductions
  - Encompassing
- Union testing
- Sub-sample evaluation

Implementation

(a) Full-information procedure: System analysis (EGLS)
$\#$ state vector $= K(n-m)$

(b) Limited-information procedure: Single-equation analysis
$\# k^{th}$ state vector $= (n-m)$

5 General-to-specific reductions of SVAR models

In the case of contemporaneous correlation between the variables of the system, i.e. some $\sigma_{ij} \neq 0$ for $i \neq j$, proposition 1 does not hold. Hence a single-equation reduction procedure such as *PcGets* can not be an optimal implementation of the *Gets* methodology, Though the findings in Krolzig (2001a) suggest that they still might deliver reasonable results.

In this section we will argue that the recursive structure of SVARs as defined in (5) re-establish conditions for the (asymptotic) efficiency of single-equation *Gets* reduction procedures as implemented by *PcGets*. The crucial point is that the researcher does not aim to reduce the reduced-form VAR, but the recursive SVAR:

Proposition 2 (Reduction under causal ordering). *Suppose that the GUM is a just-identified SVAR of the form defined by (5). Then, all possible reductions of the SVAR can be efficiently estimated by OLS, and model-selection procedures can operate equation-by-equation without a loss in efficiency.*

Proof. Causal ordering of y_t allows the factorization of the probability density function of y_t in terms of marginals and conditionals:

$$f_y(y_t|Y_{t-1}) = f_{y_1}(y_{1t}|Y_{t-1}) \cdot f_{y_2|y_1}(y_{2t}|y_{1t}, Y_{t-1}) \cdot \ldots \cdot f_{y_K|y_1,\ldots,y_{K-1}}(y_{Kt}|y_{1t}, \ldots, y_{K-1t}, Y_{t-1}).$$

This implies that the log-likelihood function $L_T(\lambda)$ can be separated with regard to the parameters of interest, λ_k, of each equation $k = 1, \ldots, K$ of the system:

$$L_T(\lambda) = \sum_{t=1}^{T} \ln f_{y_1}(y_{1t}|Y_{t-1}, \lambda_1) + \sum_{k=2}^{K} \left\{ \sum_{t=1}^{T} \ln f_{y_k|y_1,\ldots,y_{k-1}}(y_{kt}|y_{1t}, \ldots, y_{k-1t}, Y_{t-1}, \lambda_k) \right\}.$$

In other words, $y_1, \ldots, y_{k-1}$ are weakly exogenous for the parameter vector $\lambda_k = (B_{k1}, \ldots, B_{kk-1}, \Gamma_{k1.1}, \ldots, \Gamma_{kk.p}, \omega_k^2)'$ of the k-th equation of the SVAR, which can be varied freely. Thus, reductions procedures of the SVAR can be implemented as single-equation techniques. ■

In §7 this case is studied in a Monte Carlo experiment. An alternative, asymptotically efficient single-equation *Gets* model selection procedure is given by proposition 3:

Proposition 3 (Two-stage reduction procedure). *Suppose that the SVAR is just identified by the restrictions in (7) such that the unrestricted B_{ij}'s are non-zero. Then a two-stage reduction consisting of (i) estimating B indirectly from the unrestricted reduced-form VAR as $P^{-1}\mathrm{dg}(P)$, where P is obtained from a Choleski decomposition of $\hat{\Sigma}$ and (ii)* Gets *reductions of the transformed system, is asymptotically equivalent to* Gets *reductions of the SVAR as stated by proposition 2.*

Proof. Depends on the consistency of the model selection procedure and that of the OLS estimate of Σ. ■

6 Selecting the causal structure

7 Monte Carlo results

Although the sequential nature of proposed model-selection process and its combination of variable-selection and diagnostic testing has eluded most attempts at theoretical analysis, an evaluation of the properties of the model-selection process can be achieved by simulation. In the Monte Carlo (MC) experiment considered here, the properties of the reduction of SVAR models with *PcGets*, foremost its 'size' and 'power', are studied under the conditions of proposition 2.

The DGP is a three-dimensional Gaussian SVAR(1) model with the causal ordering $y_{1t} \rightarrow y_{2t} \rightarrow y_{3t}$:

$$\begin{bmatrix} 1 & 0 & 0 \\ \beta_{21} & 1 & 0 \\ \beta_{31} & \beta_{32} & 1 \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} \gamma_{11} & 0 & \gamma_{13} \\ 0 & \gamma_{22} & \gamma_{23} \\ 0 & 0 & \gamma_{33} \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} + \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \\ \eta_{3t} \end{bmatrix} \quad (9)$$

$$\text{where } \eta_t \sim \mathsf{NID}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \omega_1^2 & 0 & 0 \\ 0 & \omega_2^2 & 0 \\ 0 & 0 & \omega_3^2 \end{bmatrix} \right).$$

The reduced-form VAR(1) model

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} \gamma_{11} & 0 & \gamma_{13} \\ -\beta_{21}\gamma_{11} & \gamma_{22} & \gamma_{23}-\beta_{21}\gamma_{13} \\ -\beta_{21}\gamma_{11} & -\beta_{32}\gamma_{22} & \gamma_{33}-\beta_{31}\gamma_{13}-\beta_{32}\gamma_{23} \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix} \quad (10)$$

$$\text{where } \varepsilon_t \sim \mathsf{NID}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \omega_1^2 & -\beta_{21}\omega_1^2 & -\beta_{31}\omega_1^2 \\ -\beta_{21}\omega_1^2 & \omega_2^2+\beta_{21}^2\omega_1^2 & -\beta_{32}\omega_2^2+\beta_{31}\beta_{21}\omega_1^2 \\ -\beta_{31}\omega_1^2 & -\beta_{32}\omega_2^2+\beta_{31}\beta_{21}\omega_1^2 & \omega_3^2+\beta_{31}^2\omega_1^2+\beta_{32}^2\omega_2^2 \end{bmatrix} \right).$$

We can think of the following ‘deep’ structure associated with recursive structure of the SVAR

$$\begin{aligned} \Delta y_{1t} &= -\alpha(y_{1t-1}-y_{3t-1})+\eta_1, \\ \Delta y_{2t} &= -\beta(y_{2t-1}-y_{3t-1}-y_{1t})+\eta_2, \\ \Delta y_{3t} &= \gamma(y_{2t}-y_{1t})-\rho y_{3t-1}+\eta_3, \end{aligned}$$

which results with $\alpha = 0.4$, $\beta = 0.4$, $\rho = 0.4$, and $\gamma = 0.5$ in the following parameterization of the SVAR in (9):

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.4 & 1 & 0 \\ 0.5 & -0.5 & 1 \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} 0.6 & 0 & 0.4 \\ 0 & 0.6 & 0.4 \\ 0 & 0 & 0.6 \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} + \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \\ \eta_{3t} \end{bmatrix} \quad (11)$$

$$\eta_t \sim \mathsf{NID}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)$$

Thus reduced form of the SVAR in (11) is given by the VAR(1) model:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} 0.60 & 0 & 0.40 \\ -0.24 & 0.60 & 0.24 \\ -0.42 & 0.30 & 0.52 \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix} \quad (12)$$

$$\varepsilon_t \sim \mathsf{NID}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1.00 & -0.40 & -0.50 \\ -0.40 & 1.16 & 0.78 \\ -0.50 & 0.78 & 1.74 \end{bmatrix} \right).$$

With eigenvalues of 0.672 ± 0.351 and 0.376 of A_1, the DGP is stationary with a zero mean.

In the Monte Carlo experiment we intend to evaluate the properties of the following modelling strategies

- (i) the true SVAR(1) in (11);
- (ii) a just-identified SVAR(5) with intercept, which nests (11);
- (iii) a *Gets* reduction of the SVAR(5) in (ii) using *PcGets*;
- (iv) the pseudo-true VAR(1) in (12);
- (v) an unrestricted VAR(5) with intercept, which nests (12) and itself is equivalent to the just-identified SVAR in (ii);
- (vi) a *Gets* reduction of the VAR(5) in (v) using *PcGets*, which in general will be different from the selected SVAR in (iii).

The sample size T is 100 and the number of replications M is 1000. In (iii) and (iv) the conservative strategy of PcGets is employed. In addition to the selection properties of (iii) and (vi), the analysis of the simulation experiment will also focus on the accuracy and precision of the resulting impulse-responses of (i)–(vi).

Table 3 Monte Carlo Results.

Model	SVAR			VAR		
Equation	$y_{1,t}$	$y_{2,t}$	$y_{3,t}$	$y_{1,t}$	$y_{2,t}$	$y_{3,t}$
DGP found when commencing from it	1.000	0.982	0.999	1.000	0.333	0.386
DGP found by *PcGets*	0.860	0.788	0.843	0.860	0.222	0.282
Non-deletion probability	0.140	0.189	0.155	0.140	0.325	0.264
Non-selection probability	0.020	0.056	0.021	0.020	0.738	0.675
DGP dominated by *PcGets*	0.099	0.153	0.113	0.099	0.554	0.550
PcGets dominated by DGP	0.026	0.042	0.031	0.026	0.033	0.038
Size	0.0157	0.0232	0.0169	0.0157	0.0376	0.0305
Size (*reliability based*)	0.9880	0.9770	0.9920	0.9880	0.7227	0.7420
Power	0.0116	0.0165	0.0129	0.0116	0.0315	0.0263
Power (*reliability based*)	0.9874	0.9727	0.9912	0.9874	0.7062	0.7250

Table 3 clarifies the 'success' and 'failure' of *PcGets*. Given the characteristics of the DGP, the probability to find the DGP by *PcGets* is in between 78.8% and 86% for the SVAR, and 22.2% to 86% for the VAR. . These figures depend on the design of DGP. For the VAR the figures appear to be small, but have to be compared to the probability of finding the DGP when starting the search from it which is in between 33.3% and 100%. As the overall probability to find or miss the DGP is not very informative, we check by an encompassing test whether the deviation of the model found by *PcGets* from the DGP results in a sound model that, based on statistically criteria, could not have been improved by knowing the truth. As long as *PcGets* is able to find a model that is not dominated by the DGP itself, the reduction process has been a success. If the specific model is dominated by the DGP, the search algorithm has failed. Our results indicate that the risk to find a model which is dominated by the DGP is relatively small (2.6% to 4.2%) when compared to the probability that the selection of *PcGets* dominates the truth which is 3 to 15 times higher. Note that by construction the outcome of *PcGets* always beats the unrestricted VAR(5) model.

The 'size' of *PcGets* (the average probability of selecting a nuisance regressor) is with 1.57% to 3.76% slightly higher than the nominal size of a t-test which is 1%. The size can be further improved by basing the inclusion decision on the reliability statistic of *Stage III*. In this case the 'size' shrinks to 1.16% – 3.15%. The 'power' of *PcGets* (the average probability of selecting a DGP variable) is in between 99.2% and 72.27%. Overall, *PcGets* works more than satisfactory despite the presence of collinearity among the regressors.

We now check how the good selection properties of the proposed selection strategy are translated into accuracy and precision of the impulse responses implied by the empirical model. Figures 1 to 5 display the (non-structural) impulse responses of the system for the six modelling approaches. Figures 2 to 4 the structural responses to a shock in the first two equations of the system in (8).[3] Note that the 90% confidence intervals for impulse-responses of the true and unrestricted VAR as well as SVAR only reflect estimation uncertainty, while the confidence intervals for impulse-responses of the models selected by *PcGets* also account for the specification uncertainty.

[3] Due to the causal order imposed, the responses to a structural shock in y_3 are identical to the corresponding non-structural one. So just one figure is required.

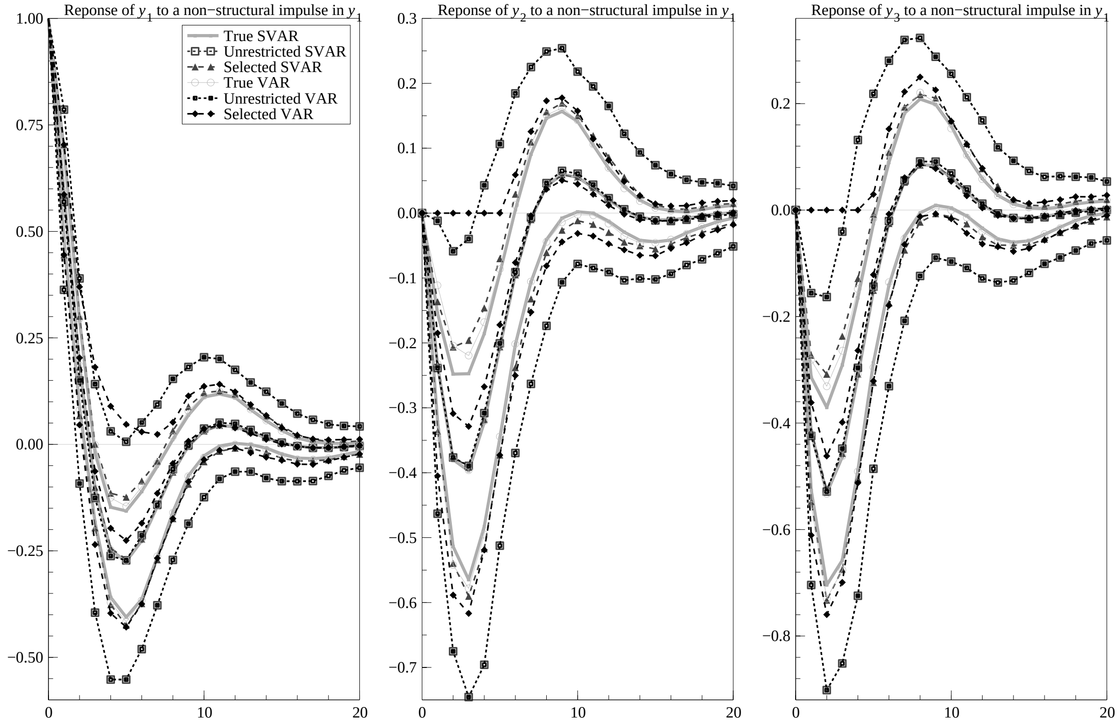

Figure 1 Responses to a non-structural shock in y_1.

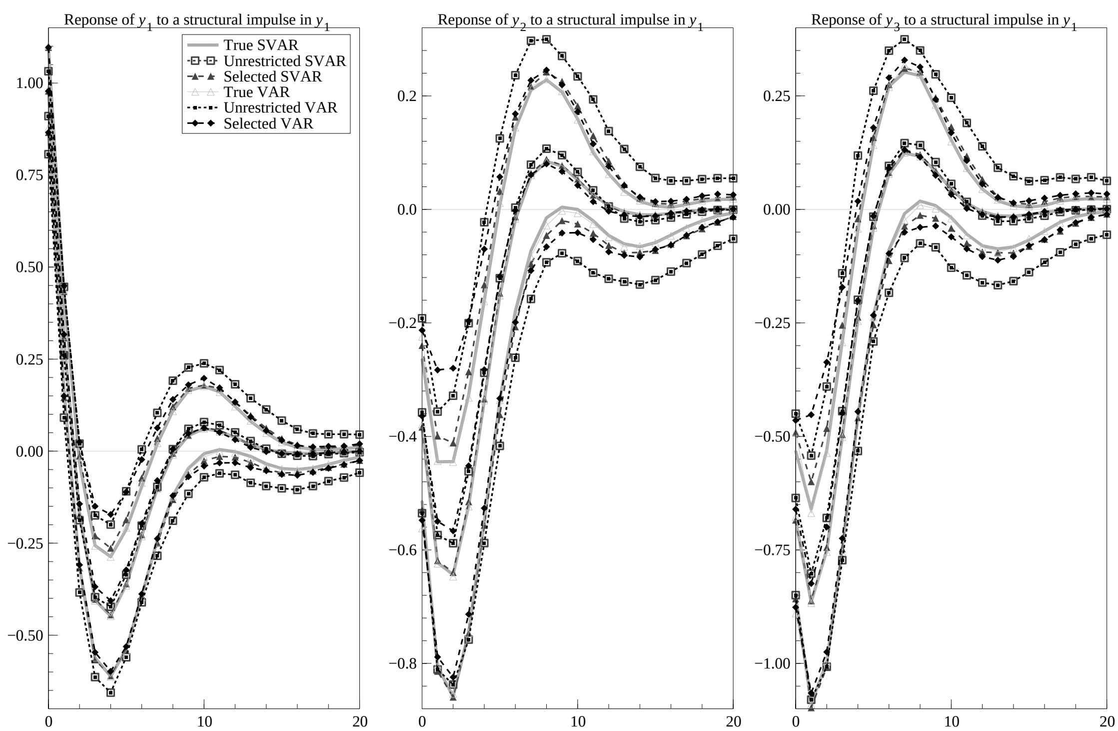

Figure 2 Responses to a structural shock in y_1.

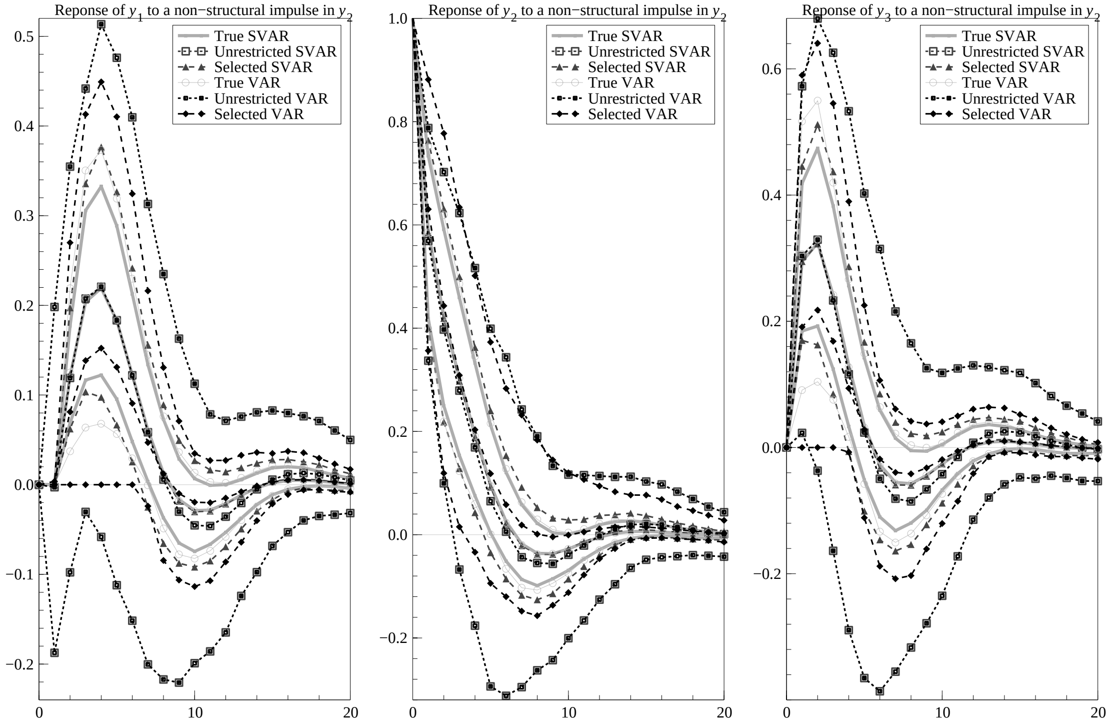

Figure 3 Responses to a non-structural shock in y_2.

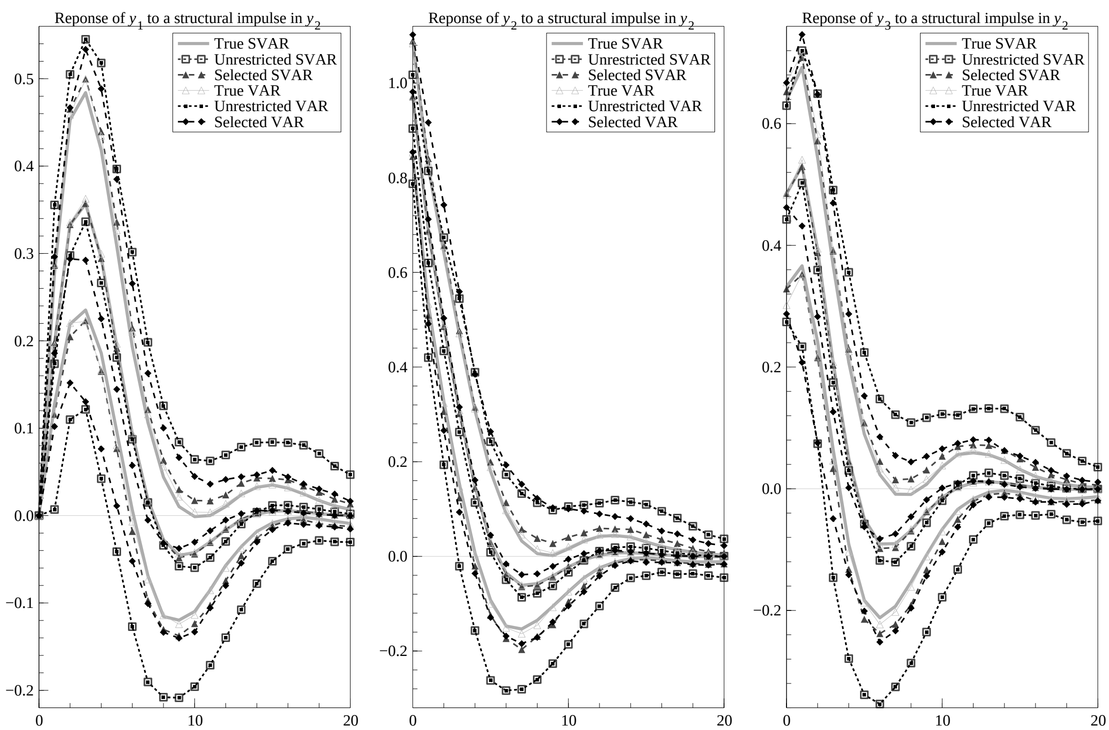

Figure 4 Responses to a structural shock in y_2.

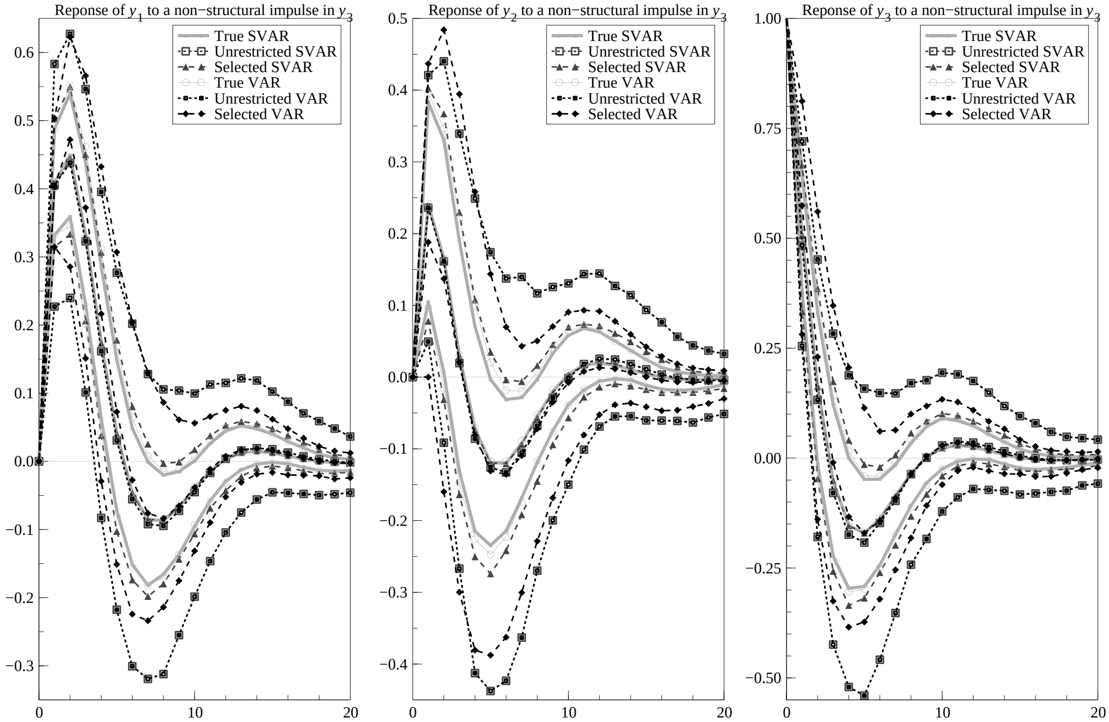

Figure 5 Responses to a shock y_3.

The figures suggest that, altogether, the precision of the model selection strategies can be ranked as follows:

(i) the true SVAR;
(i) the true VAR;
(iv) the selected SVAR;
(iv) the selected VAR;
(v) the just-identified SVAR/ unrestricted VAR.

Further research is required to demonstrate the robustness of these results.

8 Empirical illustration

To illustrate the proposed *Gets* procedures for SVARs, we will now use *PcGets* to analyze a small macro-econometric model for the US. The model is a sub-model of the monetary system considered by Christiano, Eichenbaum and Evans (1996) consisting of the log of real GDP, y, the log of the GDP deflator, p, the log of a commodity price index, p^{com}, the fed funds rate, i, the negative log of unborrowed reserves, r^U, the log of total reserves, r^T and the log of M1, m. Christiano *et al.* (1996) impose the following causal ordering

$$y_t \to p_t \to p_t^{com} \to i_t \to r_t^U \to r_t^T \to m_t$$

to analyze the effects of monetary policy shocks in an unrestricted reduced-form VAR(4) of the variables. Given the original interest in a recursive structure and the strong indication of instantaneous causality found by Krolzig (2001b), it seems appropriate to utilize the (potential) presence of a causal order during the reduction process in spirit of proposition 2.

In the following we will restrict our interest to a four-dimension system $(\Delta y, \Delta p, i, \Delta m)'$ while extending the data set to the period: 1962 (i) to 1999 (iii). Based on Johansen's cointegration analysis of the system $(y, p, i, m)'$ differencing of y, p and m does not affect a potential cointegrating relationships. The cointegration analysis of the system $(\Delta y, \Delta p, i, \Delta m)'$ still indicates the presence of stochastic trends even in the transformed system. This could potentially cause a problem as, to date, *PcGets* conducts all inferences as I(0). Fortunately, most selection tests remain valid even when the data are I(1), given the results in, say, Sims, Stock and Watson (1990). Only t- or F-tests for an effect that corresponds to a unit root require non-standard critical values. Similarly, Wooldridge (1999) shows that diagnostic tests on the GUM (and presumably simplifications thereof) remain valid even for integrated time series.

In spirit of proposition 4we consider *Gets* reductions of just identified SVARs and reduced-form VARs using the conservative strategy of *PcGets* 1. As it is a stylized fact that the US economy has been subject to structural change reducing the volatility of the economic process, the test for heteroscedasticity and ARCH of the residuals have removed from the test battery.

Starting point of the analysis is a just-identified SVAR(5) as defined in equation (5). All $4! = 24$ possible orderings of the four variables of the system are considered. Table 4 reports for each of the orderings the number of rejected misspecification tests in the just-identified SVAR (r) and the following properties of the SVAR selected by *PcGets*: the log-likelihood (L), the number of coefficients (n), the value of the Akaike information criterion (AIC, see Akaike, 1973), the Hannan-Quinn criterion (HQ, see Hannan and Quinn, 1979), the Schwarz criterion (SC, see Hannan and Quinn, 1979), and the reliability based criterion HK (see Krolzig and Hendry, 2001). The table is ordered with regard to the value of the HQ criterion. The marked models are of special interest and discussed in following.

Model A maximizes AIC, HQ and HK:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ - & + & 1 & 0 \\ + & - & - & 1 \end{bmatrix} \begin{bmatrix} \Delta p_t \\ \Delta m_t \\ i_t \\ \Delta y_t \end{bmatrix} = g^A(past) + \eta_t^A.$$

The conditional relationships are given by $i_t(\underset{(+}{\Delta p_t}, \underset{(-)}{\Delta m_t})$ and $\Delta y_t(\underset{(-)}{\Delta p_t}, \underset{(+)}{\Delta m_t}, \underset{(+)}{i_t})$. As it is counterintuitive, that economic activity is a positive function of the real interest rate, we might want to

Table 4 Comparison of the selected SVARs (ordered by HQ).

Causal Order	r	L	n	AIC	HQ	SC	HK	Model
$\Delta p_t \rightarrow \Delta m_t \rightarrow i_t \rightarrow \Delta y_t$	3	3044.00	24	-40.0000	-39.8052	-39.5204	-39.8052	A
$\Delta p_t \rightarrow \Delta y_t \rightarrow i_t \rightarrow \Delta m_t$	3	3043.28	24	-39.9905	-39.7956	-39.5109	-39.7892	C
$\Delta m_t \rightarrow \Delta p_t \rightarrow i_t \rightarrow \Delta y_t$	3	3044.65	25	-39.9953	-39.7924	-39.4958	-39.7860	
$\Delta m_t \rightarrow i_t \rightarrow \Delta p_t \rightarrow \Delta y_t$	3	3044.00	25	-39.9867	-39.7837	-39.4871	-39.7709	
$\Delta p_t \rightarrow \Delta m_t \rightarrow \Delta y_t \rightarrow i_t$	2	3042.30	24	-39.9775	-39.7827	-39.4980	-39.7742	
$\Delta p_t \rightarrow i_t \rightarrow \Delta m_t \rightarrow \Delta y_t$	3	3038.94	22	-39.9594	-39.7808	-39.5198	-39.7808	
$i_t \rightarrow \Delta m_t \rightarrow \Delta p_t \rightarrow \Delta y_t$	3	3042.00	24	-39.9735	-39.7787	-39.4939	-39.7723	
$\Delta p_t \rightarrow \Delta y_t \rightarrow \Delta m_t \rightarrow i_t$	3	3040.26	23	-39.9636	-39.7769	-39.5041	-39.7705	
$i_t \rightarrow \Delta p_t \rightarrow \Delta m_t \rightarrow \Delta y_t$	3	3040.16	23	-39.9624	-39.7757	-39.5028	-39.7757	
$\Delta p_t \rightarrow i_t \rightarrow \Delta y_t \rightarrow \Delta m_t$	3	3036.73	21	-39.9434	-39.7730	-39.5238	-39.7730	B
$i_t \rightarrow \Delta y_t \rightarrow \Delta m_t \rightarrow \Delta p_t$	3	3039.79	23	-39.9575	-39.7708	-39.4980	-39.7644	
$\Delta y_t \rightarrow \Delta p_t \rightarrow i_t \rightarrow \Delta m_t$	2	3047.83	28	-39.9977	-39.7704	-39.4382	-39.7046	CEE
$\Delta m_t \rightarrow \Delta p_t \rightarrow \Delta y_t \rightarrow i_t$	2	3042.95	25	-39.9729	-39.7699	-39.4733	-39.7550	
$i_t \rightarrow \Delta p_t \rightarrow \Delta y_t \rightarrow \Delta m_t$	3	3037.96	22	-39.9465	-39.7679	-39.5069	-39.7679	
$i_t \rightarrow \Delta y_t \rightarrow \Delta p_t \rightarrow \Delta m_t$	3	3037.96	22	-39.9465	-39.7679	-39.5069	-39.7679	
$\Delta m_t \rightarrow \Delta y_t \rightarrow \Delta p_t \rightarrow i_t$	2	3040.89	24	-39.9588	-39.7639	-39.4792	-39.7554	
$i_t \rightarrow \Delta m_t \rightarrow \Delta y_t \rightarrow \Delta p_t$	3	3039.08	23	-39.9480	-39.7613	-39.4885	-39.7549	
$\Delta y_t \rightarrow i_t \rightarrow \Delta m_t \rightarrow \Delta p_t$	2	3048.55	29	-39.9940	-39.7586	-39.4145	-39.6864	
$\Delta y_t \rightarrow i_t \rightarrow \Delta p_t \rightarrow \Delta m_t$	2	3046.71	28	-39.9829	-39.7556	-39.4234	-39.6898	
$\Delta y_t \rightarrow \Delta p_t \rightarrow \Delta m_t \rightarrow i_t$	2	3044.81	27	-39.9709	-39.7517	-39.4314	-39.6859	
$\Delta m_t \rightarrow \Delta y_t \rightarrow i_t \rightarrow \Delta p_t$	3	3039.85	24	-39.9451	-39.7502	-39.4655	-39.7353	
$\Delta y_t \rightarrow \Delta m_t \rightarrow \Delta p_t \rightarrow i_t$	2	3043.39	27	-39.9522	-39.7330	-39.4126	-39.6672	
$\Delta y_t \rightarrow \Delta m_t \rightarrow i_t \rightarrow \Delta p_t$	3	3042.36	27	-39.9385	-39.7193	-39.3990	-39.6471	
Reduced VAR	2	3032.51	25	-39.8346	-39.6317	-39.3350	-39.5659	
Unrestricted VAR	2	3060.17	84	-39.4195	-38.7376	-37.7410	-37.4183	
Just-identified SVAR	-	3076.43	90	-39.5554	-38.8248	-37.7570	———	

Notes: The reduction process used the *conservative* strategy of *PcGets* 1 with the tests for ARCH effects and heteroskedasticity switched off. The log-likelihood of the reduced-form VAR is calculated as the sum of the likelihood of the submodels. All just-identified SVARs are identical with regard to their likelihood, AIC, HQ and SC. The number of rejected misspecification tests and HK differ for different causal ordering.
r : number of rejected misspecification tests in the just-identified SVAR(5).
n : number of coefficients in the SVAR

consider Model B which maximizes SC:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ - & 1 & 0 & 0 \\ 0 & - & 1 & 0 \\ 0 & + & - & 1 \end{bmatrix} \begin{bmatrix} \Delta p_t \\ i_t \\ \Delta y_t \\ \Delta m_t \end{bmatrix} = g^B(past) + \eta_t^B.$$

But the parameters associated with the contemporaneous conditioning are again counterintuitive with $i_t(\underset{+}{\Delta p_t})$, $\Delta y_t(\underset{+}{i_t})$, and $\Delta m_t(\underset{-}{i_t}, \underset{+}{\Delta y_t})$.

Finally, we found with Model C an SVAR which is supported by the data and economically interpretable:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ - & - & 1 & 0 \\ 0 & - & + & 1 \end{bmatrix} \begin{bmatrix} \Delta p_t \\ \Delta y_t \\ i_t \\ \Delta m_t \end{bmatrix} = \begin{bmatrix} 0 \\ + \\ - \\ 0 \end{bmatrix} + \begin{bmatrix} + & 0 & - & + \\ 0 & 0 & 0 & 0 \\ 0 & + & + & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta p_{t-1} \\ \Delta y_{t-1} \\ i_{t-1} \\ \Delta m_{t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & + & - & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & + & + \end{bmatrix} \begin{bmatrix} \Delta p_{t-2} \\ \Delta y_{t-2} \\ i_{t-2} \\ \Delta m_{t-2} \end{bmatrix}$$

$$+ \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & + & 0 \\ 0 & 0 & + & + \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta p_{t-3} \\ \Delta y_{t-3} \\ i_{t-3} \\ \Delta m_{t-3} \end{bmatrix} + \begin{bmatrix} + & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & - & - \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta p_{t-4} \\ \Delta y_{t-4} \\ i_{t-4} \\ \Delta m_{t-4} \end{bmatrix} + \begin{bmatrix} 0 & 0 & - & 0 \\ 0 & 0 & 0 & 0 \\ 0 & + & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta p_{t-5} \\ \Delta y_{t-5} \\ i_{t-5} \\ \Delta m_{t-5} \end{bmatrix} + \begin{bmatrix} \eta_{1t} \\ \eta_{2t} \\ \eta_{3t} \\ \eta_{4t} \end{bmatrix}$$

where the third equation describes the nominal interest as a positive function of the contemporaneous growth and the inflation rate, $i_t(\underset{(+)}{\Delta p_t}, \underset{(+)}{\Delta y_t})$, and money (demand) growth as a positive function of the growth of nominal GDP and a negative function of the interest rate: $\Delta m_t(\underset{(+)}{\Delta p_t}, \underset{(+)}{\Delta y_t}, \underset{(-)}{i_t})$.

Interestingly the model is consistent with a *Gets* reduction of an SVAR with the causal ordering proposed by Christiano *et al.* (1996), denoted CEE, whose recursive structure is given by:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ - & - & 1 & 0 \\ - & 0 & + & 1 \end{bmatrix} \begin{bmatrix} \Delta y_t \\ \Delta p_t \\ i_t \\ \Delta m_t \end{bmatrix} \hat{=} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ - & - & 1 & 0 \\ 0 & - & + & 1 \end{bmatrix} \begin{bmatrix} \Delta p_t \\ \Delta y_t \\ i_t \\ \Delta m_t \end{bmatrix}$$

After reordering the equations, the same causal order as in Model C emerges. But Model C is clearly more parsimonious and, hence, preferred by all four information criteria. Altogether, we conclude that the support for the selected model is strong: Starting with the unrestricted SVAR(5) as the GUM, *PcGets* sets 66 zero restrictions, it finds 23 coefficients that are significant at the 1% percent level and one significant at the 5% percent level. The details are reported in the appendix.

For comparison, we also considered single-equation *Gets* reductions of the reduced-form using the reduction strategy of Krolzig (2001a). The results (reported in the appendix) confirm the findings of the Monte Carlo: There are clear advantages of applying *PcGets* to the SVAR when compared to reductions of the reduced-form VAR by *PcGets*. 59 null restrictions are set, 3 coefficients are insignificant, one coefficient is significant at the 5% percent level and 21 coefficients are significant at the 1% percent level. Interestingly, the reduction of the VAR is unanimously and clearly dominated by the reductions of the SVAR discussed earlier. It would be worth to compare the outcome of *PcGets* to other reduction procedures such as Brüggemann and Lütkepohl (2000) (see also Brüggemann, Krolzig and Lütkepohl, 2001). Further research on the pros and cons of the proposed procedure is required.

Figures 6 and 7 show the impulse responses for the unrestricted VAR, the selected SVAR (Model C) and reduced-form VAR selected by *PcGets*. In 7 the orthogonalized responses are plotted, where in case

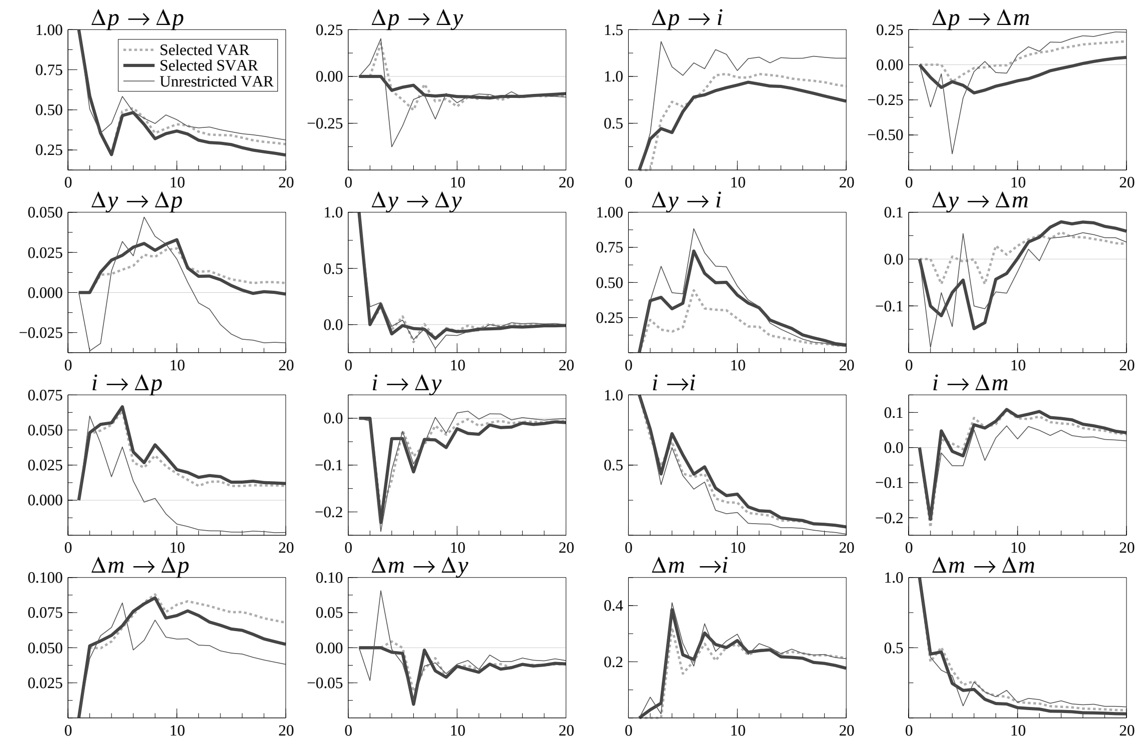

Figure 6 Impulse-responses.

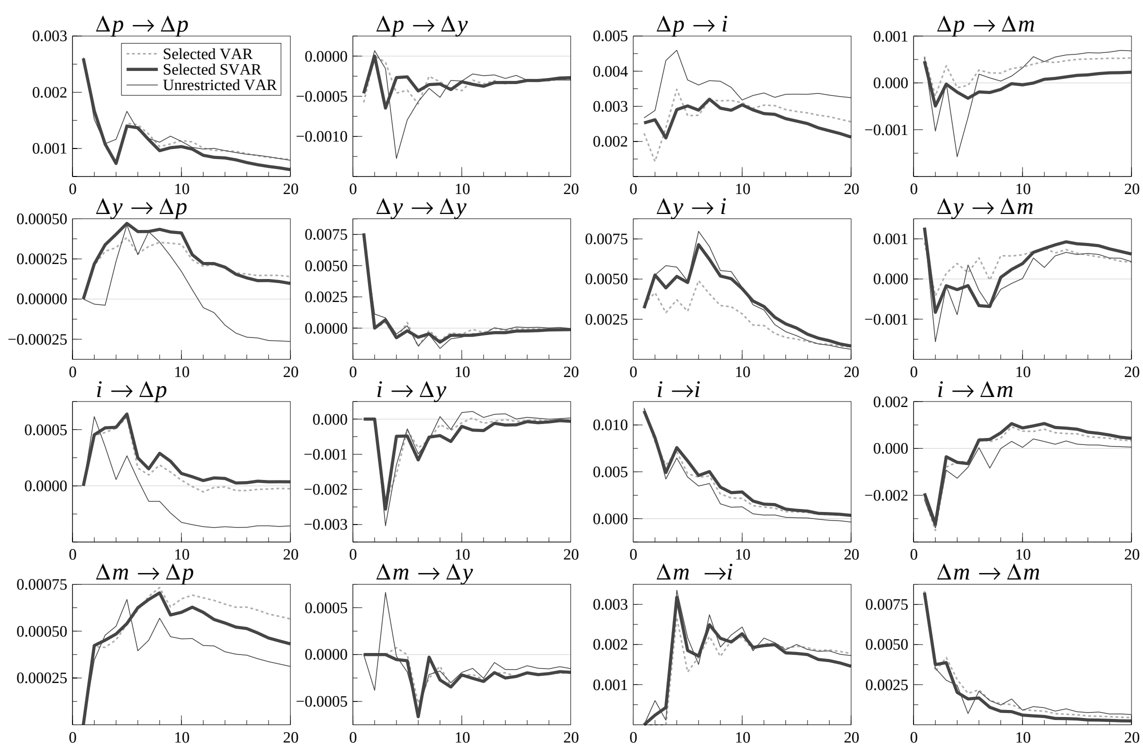

Figure 7 Structural (orthogonalized) impulse-responses.

of the reduced system the variance-covariance matrix as estimated by FIML has been used. The solid line represents the response in the selected SVAR, the dashed line the response of the unrestricted VAR (which is identical to that of a just-identified SVAR) and the dashed line the response in the selected reduced-form VAR. Interestingly, the responses of the VAR and its reduction show a very similar pattern. There is no indication of a bias caused by the reduction. Indeed, from the Monte Carlo study in §7 we expect that the responses in the reduced SVAR model estimated more precisely than in the alternative models. Overall, *PcGets* seems to be useful for the specification of SVAR and the analysis of their impulse responses.

9 Extensions

The approach considered so far is limited in two important respects: (i) the stationarity of the data-generating process and (ii) the recursive structure of the SVAR. In the following we outline how these limitations can be overcome in a more general setting which still allows the application of the reduction procedure discussed here.

First, we assumed that the VAR is stable. Most economic data show stochastic trends, so that we should allow for integrated and possibly co-integrated processes. Note that the reduced-form VAR(p) model in (1) can always be represented as a $p-$th order vector equilibrium correction model (VECM)

$$\Delta y_t = \nu + \Pi y_{t-1} + \sum_{i=1}^{p-1} \Upsilon_i \Delta y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathsf{NID}(\mathbf{0}, \Sigma). \tag{13}$$

If Π has full rank, the variables y_t are $I(0)$, and without further restriction, the VECM is just- identified. An identification issue arises when the $0 <$rank$(\Pi) = r < K$, in which case $y_t \sim I(1)$ and there exit r cointegration vectors $\beta' y_t \sim I(0)$. Various methods for the cointegration analysis of multiple time series have been proposed in the literature. As the *Gets* procedures discussed here are likelihood based, Johansen's concentrated-likelihood-function approach (see Johansen, 1995) is apparently the natural choice and suggests a three-stage reduction approach

(i) Johansen's reduced rank procedures are based on the unrestricted reduced-form VAR. Therefore a lag selection procedure to determine the order of the VAR precedes the cointegration analysis. This step could involve a liberal sequential F-test procedure of block restrictions $A_i = 0$ for $i = h, h-1, \ldots$ or AIC model comparisons.

(ii) The Johansen procedure for empirically determining the cointegration rank r, produces unique estimates of α and β as a result of requiring β to be orthogonal and normalized. This estimate provides a value for the unrestricted log-likelihood function to be compared to the value for the log-likelihood function under overidentifying restrictions for α and β which have an economic interpretation.

(iii) Given the outcome of the cointegration analysis (the cointegration rank r, cointegration matrix β and the structure of α), the analysis then focuses either on reduction of the short-run dynamics Υ_i in the corresponding (stationary) reduced-form VECM($p-1$),

$$\Delta y_t = \nu + \alpha_r \left(\beta_r' y_{t-1}\right) + \sum_{i=1}^{p-1} \Upsilon_i \Delta y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathsf{NID}(\mathbf{0}, \Sigma), \tag{14}$$

or, for a given recursive structure, on Υ_i^* in the structural VECM

$$B \Delta y_t = \delta + \alpha_r^* \left(\beta_r' y_{t-1}\right) + \sum_{i=1}^{p-1} \Upsilon_i^* \Delta y_{t-1} + \eta_t, \quad \eta_t \sim \mathsf{NID}(\mathbf{0}, \Omega). \tag{15}$$

The (just) identification of (15) follows from the restrictions formulated in (7) and the equivalence with (13) is given by:

$$\begin{aligned} \Sigma &= B^{-1}\Omega B^{-1\prime}, \\ \alpha_r &= B^{-1}\alpha_r^*, \\ \Upsilon_i^* &= B^{-1}\Upsilon_i \qquad \text{for } i = 1, \ldots, p-1, \\ \nu &= B^{-1}\delta. \end{aligned}$$

Equation (15) becomes the new GUM to which the *general-to-specific* reductions discussed above can then be applied.

Alternatively, when the cointegration vectors are imposed by economic theory, the *Gets* reduction approach might be limited to the short-run dynamics.

Secondly, the limitations of the recursive structure and closeness of the SVAR in (5) which is problematic if the underlying economic theory predicts the interdependence of the endogenous variables in the temporary macroeconomic equilibrium. Fortunately it is possible to relax the assumptions made in §4 without risking the basic properties of the *Gets* reduction approach. Though, in general, single-equation reduction procedures will be become inefficient.

In principle, the reduction approach proposed here can be applied to any simultaneous equation model

$$\begin{aligned} By_t &= \Gamma z_t + \eta_t, \\ \eta_t | z_t &\sim \mathsf{NID}(\mathbf{0}, \Omega), \end{aligned} \tag{16}$$

where restrictions on (B, Γ) garniture that the model is identified, i.e. in the sense of famous Cowles' Commission rank condition. If the system is the recursive (i.e. B triangular), proposition 2 holds and implies the (asymptotic) efficiency of single-equation reduction procedure.

Many SEMs used in practice suffer from sparsely formulated dynamics which can affect the structural stability and interpretability of the model. A useful application of the reduction approach proposed in this paper is the enrichment of the SEM in (16) with additional variables collected in the vector w_t which do not interfere with the identification restrictions of the original (i.e. they do not include elements of y_t and z_t) and enter each equation unrestrictedly:

$$\begin{aligned} B^* y_t &= \Gamma^* z_t + C w_t + \eta_t^*. \\ \eta_t^* | z_t, w_t &\sim \mathsf{NID}(\mathbf{0}, \Omega^*), \end{aligned} \tag{17}$$

The reduction process would look for simplifications of C conditional on the fixed structure of B and Γ. Note that the first step of *PcGets* is the test of $C = \mathbf{0}$ (currently implemented for each equation separately) which helps to control the size of the procedure. Apparently, this approach should help to test for omitted variables and dynamic misspecifications. But the approach might be useful for researchers which derive (16) from a theoretical model, but want to ensure the congruence (i.e. the absence of misspecifications) of the empirical model by introducing variables or dynamics which are of potential importance for the analyzed data but are not subject of the theoretical analysis (e.g. the use of monetary time series given a non-monetary theoretical model).

10 Conclusions

The aim of the paper was to propose and evaluate computerized model-selection strategies for structural VARs, to see if they worked well, indifferently, or failed badly. The results come much closer to the first.

In the Monte Carlo experiments *PcGets* recovered the DGP specification from a large just-identified VAR with anticipated size, and power close to commencing from the DGP itself. The accuracy and precision of the impulse responses of the SVAR selected by *PcGets* dominate those of unrestricted models even if the potential selection error is taken into account. The feasibility of *PcGets* for the SVAR analysis of macroeconomic data sets has been demonstrated. Possible extensions could involve cointegrated and identified interdependent simultaneous equation models.

References

Akaike, H. (1973). Information theory and the extension of the maximum likelihood principle. In Petrov, B., and Csaki, F. (eds.), *Second international symposium on Information Theory*, pp. 267–281. Budapest: Akademiai Kiado.

Akaike, H. (1985). Prediction and entropy. In Atkinson, A. C., and Fienberg, S. E. (eds.), *A Celebration of Statistics*, pp. 1–24. New York: Springer-Verlag.

Anderson, T. W. (1971). *The Statistical Analysis of Time Series.* New York: John Wiley & Sons.

Bernanke, B., and Blinder, A. (1992). The federal funds rate and the channels of monetary transmission. *American Economic Review*, **82**, 901–21.

Blanchard, O., and Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances. *American Economic Review*, **79**, 655–673.

Bontemps, C., and Mizon, G. E. (2001). Congruence and encompassing. in Stigum (2001).

Brüggemann, R., Krolzig, H.-M., and Lütkepohl, H. (2001). Comparison of model specification methods. Mimeo, Humboldt–University, Berlin.

Brüggemann, R., and Lütkepohl, H. (2000). Lag selection in subset VAR models with an application to a U.S. monetary system. in Friedmann *et al.* (2001), pp. 107–128.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (1996). The effects of monetary policy shocks: Evidence from the flow of funds. *Review of Economics and Statistics*, **78**, 16–34.

Cooley, T., and Dwyer, M. (1998). Business cycle analysis without much theory. a look at structural VARs. *Journal of Econometrics*, **83**, 57–88.

Cox, D. R. (1961). Tests of separate families of hypotheses. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 105–123 Berkeley: University of California Press.

Cox, D. R. (1962). Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society B*, **24**, 406–424.

Davidson, J. E. H., Hendry, D. F., Srba, F., and Yeo, J. S. (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal*, **88**, 661–692. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers, 1993, and Oxford University Press, 2000.

Doornik, J. A. (2001). *Object-Oriented Matrix Programming using Ox* 4th end. London: Timberlake Consultants Press.

Dungey, M., and Pagan, A. (2001). A structural var model of the australian economy. *Economic Record*, forthcoming.

Faust, J., and Whiteman, C. H. (1997). General-to-specific procedures for fitting a data-admissible, theory-inspired, congruent, parsimonious, encompassing, weakly-exogenous, identified, structural model of the DGP: A translation and critique. *Carnegie–Rochester Conference Series on Public Policy*, **47**, 121–161.

Florens, J.-P., Hendry, D. F., and Richard, J.-F. (1996). Encompassing and specificity. *Econometric Theory*, **12**, 620–656.

Florens, J.-P., and Mouchart, M. (1980). Initial and sequential reduction of Bayesian experiments. Discussion paper 8015, CORE, Louvain-La-Neuve, Belgium.

Florens, J.-P., Mouchart, M., and Rolin, J.-M. (1990). *Elements of Bayesian Statistics*. New York: Marcel Dekker.

Friedmann, R., Knüppel, L., and Lütkepohl, H. (eds.)(2001). *Econometric Studies - A Festschrift in Honour of Joachim Frohn*. Münster: LIT Verlag.

Govaerts, B., Hendry, D. F., and Richard, J.-F. (1994). Encompassing in stationary linear dynamic models. *Journal of Econometrics*, **63**, 245–270.

Granger, C. W. J. (ed.)(1990). *Modelling Economic Series*. Oxford: Clarendon Press.

Hannan, E. J., and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society*, **B, 41**, 190–195.

Hansen, L., and Sargent, T. (1981). Formulating and estimating dynamic linear rational expectations models. In *Rational Expectations and Econometric Practice*. Minneapolis: University of Minnesota Press.

Hendry, D. F. (1979). Predictive failure and econometric modelling in macro-economics: The transactions demand for money. In Ormerod, P. (ed.), *Economic Modelling*, pp. 217–242. London: Heinemann. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers, 1993, and Oxford University Press, 2000.

Hendry, D. F. (1987). Econometric methodology: A personal perspective. In Bewley, T. F. (ed.), *Advances in Econometrics*, C.H. 10. Cambridge: Cambridge University Press.

Hendry, D. F. (1993). *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.

Hendry, D. F. (2000). Epilogue: The success of general-to-specific model selection. In *Econometrics: Alchemy or Science?*, pp. 467–490. Oxford: Oxford University Press. New Edition.

Hendry, D. F., and Ericsson, N. R. (1991). Modeling the demand for narrow money in the United Kingdom and the United States. *European Economic Review*, **35**, 833–886.

Hendry, D. F., and Krolzig, H.-M. (1999). Improving on 'Data mining reconsidered' by K.D. Hoover and S.J. Perez. *Econometrics Journal*, **2**, 202–219.

Hendry, D. F., and Krolzig, H.-M. (2001). *Automatic Econometric Model Selection with PcGets*. London: Timberlake Consultants Press.

Hendry, D. F., Leamer, E. E., and Poirier, D. J. (1990). A conversation on econometric methodology. *Econometric Theory*, **6**, 171–261.

Hendry, D. F., and Mizon, G. E. (2000). Reformulating empirical macro-econometric modelling. *Oxford Review of Economic Policy*, **16**, 138–159.

Hendry, D. F., and Richard, J.-F. (1982). On the formulation of empirical models in dynamic econometrics. *Journal of Econometrics*, **20**, 3–33. Reprinted in Granger, C. W. J. (ed.) (1990), *Modelling Economic Series*. Oxford: Clarendon Press and in Hendry D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers 1993, and Oxford University Press, 2000.

Hendry, D. F., and Richard, J.-F. (1983). The econometric analysis of economic time series (with discussion). *International Statistical Review*, **51**, 111–163. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers, 1993, and Oxford University Press, 2000.

Hendry, D. F., and Richard, J.-F. (1989). Recent developments in the theory of encompassing. In Cornet, B., and Tulkens, H. (eds.), *Contributions to Operations Research and Economics. The XXth Anniversary of CORE*, pp. 393–440. Cambridge, MA: MIT Press.

Hendry, D. F., and Krolzig, H.-M. (2001). New developments in automatic general-to-specific modelling. in Stigum (2001). Forthcoming.

Hoover, K. D., and Perez, S. J. (1999). Data mining reconsidered: Encompassing and the general-to-specific approach to specification search. *Econometrics Journal*, **2**, 167–91.

Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

Judge, G. G., and Bock, M. E. (1978). *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*. Amsterdam: North Holland Publishing Company.

Keynes, J. M. (1939). Professor Tinbergen's method. *Economic Journal*, **44**, 558–568.

Keynes, J. M. (1940). Comment. *Economic Journal*, **50**, 154–156.

King, R. G., Plosser, C. I., and Rebelo, S. T. (1988). Production, growth and business cycles: I. The basic neoclassical model. *Journal of Monetary Economics*, *21*(2/3), 195–232.

Koopmans, T. C. (1947). Measurement without theory. *Review of Economics and Statistics*, **29**, 161–179.

Krolzig, H.-M. (2001a). Computer-automated single-equation general-to-specific reduction procedures for vector autoregressions. In Neck, R. (ed.), *IFAC SME 2001 Proceedings*. Oxford: Elsevier Science.

Krolzig, H.-M. (2001b). General-to-specific reductions in vector autoregressive processes. in Friedmann *et al.* (2001), pp. 129–157.

Krolzig, H.-M., and Hendry, D. F. (2001). Computer automation of general-to-specific model selection procedures. *Journal of Economic Dynamics and Control*, **25**, 831–866.

Leamer, E. E. (1978). *Specification Searches. Ad-Hoc Inference with Non-Experimental Data*. New York: John Wiley.

Leamer, E. E. (1983a). Let's take the con out of econometrics. *American Economic Review*, **73**, 31–43. Reprinted in Granger, C. W. J. (ed.) (1990), *Modelling Economic Series*. Oxford: Clarendon Press.

Leamer, E. E. (1983b). Model choice and specification analysis. In Griliches, Z., and Intriligator, M. D. (eds.), *Handbook of Econometrics*, Vol. 1, C.H. 5. Amsterdam: North-Holland.

Lovell, M. C. (1983). Data mining. *Review of Economics and Statistics*, **65**, 1–12.

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Berlin: Springer.

Magnus, J. R., and Morgan, M. S. (eds.)(1999). *Methodology and Tacit Knowledge: Two Experiments in Econometrics*. Chichester: John Wiley and Sons.

Mayo, D. (1981). Testing statistical testing. In Pitt, J. C. (ed.), *Philosophy in Economics*, pp. 175–230: D. Reidel Publishing Co. Reprinted as pp. 45–73 in Caldwell B. J. (1993), *The Philosophy and Methodology of Economics*, Vol. 2, Aldershot: Edward Elgar.

Mizon, G. E. (1977a). Inferential procedures in nonlinear models: An application in a UK industrial cross section study of factor substitution and returns to scale. *Econometrica*, **45**, 1221–1242.

Mizon, G. E. (1977b). Model selection procedures. In Artis, M. J., and Nobay, A. R. (eds.), *Studies in Modern Economic Analysis*, pp. P97–120. Oxford: Basil Blackwell.

Mizon, G. E. (1984). The encompassing approach in econometrics. In Hendry, D. F., and Wallis, K. F. (eds.), *Econometrics and Quantitative Economics*, pp. 135–172. Oxford: Basil Blackwell.

Mizon, G. E., and Richard, J.-F. (1986). The encompassing principle and its application to non-nested hypothesis tests. *Econometrica*, **54**, 657–678.

Pagan, A. R. (1987). Three econometric methodologies: A critical appraisal. *Journal of Economic Surveys*, **1**, 3–24. Reprinted in Granger, C. W. J. (ed.) (1990), *Modelling Economic Series*. Oxford: Clarendon Press.

Pagan, A. R. (1994). Structural VAR's: A review. mimeo, Australian National University.

Sargan, J. D. (1973). Model building and data mining. Discussion paper, London School of Economics. Presented to the Association of University Teachers of Economics, Meeting, Manchester, April 1973.

Sargan, J. D. (1981). The choice between sets of regressors. Mimeo, Economics Department, London School of Economics.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.

Simon, H. A. (1953). Causal ordering and identifiability. In Hood, W. C., and Koopmans, T. C. (eds.), *Studies in Econometric Method*, No. 14 in Cowles Commission Monograph, C.H. 3. New York: John Wiley & Sons.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, **48**, 1–48. Reprinted in Granger, C. W. J. (ed.) (1990), *Modelling Economic Series*. Oxford: Clarendon Press.

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

Sims, C. (1992). Interpreting the macroeconomic time series facts: The effects of monetary policy. *European Economic Review*, **36**, 975–1000.

Stigum, B. (ed.)(2001). *Studies in Economic Methodology*. Cambridge, Mass.: MIT Press.

Strotz, R. H., and Wold, H. O. A. (1960). Recursive versus non-recursive systems: An attempt at a synthesis. *Econometrica*, **28**, 417–421.

Tinbergen, J. (1940a). *Statistical Testing of Business-Cycle Theories*. Geneva: League of Nations. Vol. I: A Method and its application to Investment Activity.

Tinbergen, J. (1940b). *Statistical Testing of Business-Cycle Theories*. Geneva: League of Nations. Vol. II: Business Cycles in the United States of America, 1919–1932.

White, H. (1990). A consistent model selection. in Granger (1990), pp. 369–383.

Wold, H. O. A. (1949). Statistical estimation of economic relationships. *Econometrica*, **17**, 1–21. Supplement.

Wooldridge, J. M. (1999). Asymptotic properties of some specification tests in linear models with integrated processes. In Engle, R. F., and White, H. (eds.), *Cointegration, Causality and Forecasting*, pp. 366–384. Oxford: Oxford University Press.

Appendix

We report here the estimated SVARs A, B, and C and the reduced-form VAR of §8.

SVAR model A

Estimation results:

$$\widehat{\Delta p_t} = \underset{(0.0656)}{0.594}\ \Delta p_{t-1} + \underset{(0.0678)}{0.324}\ \Delta p_{t-4} + \underset{(0.0185)}{0.0509}\ \Delta m_{t-1} + \underset{(0.0107)}{0.047}\ i_{t-1} - \underset{(0.0107)}{0.0466}\ i_{t-5}$$

$$\widehat{\Delta m_t} = \underset{(0.0732)}{0.384}\ \Delta m_{t-1} + \underset{(0.0752)}{0.363}\ \Delta m_{t-2} - \underset{(0.0508)}{0.232}\ i_{t-1} + \underset{(0.0501)}{0.277}\ i_{t-2}$$

$$\widehat{i_t} = -\underset{(0.00391)}{0.0117} + \underset{(0.211)}{0.861}\ \Delta p_t - \underset{(0.0987)}{0.294}\ \Delta m_t + \underset{(0.0981)}{0.291}\ \Delta m_{t-3} + \underset{(0.0662)}{0.509}\ i_{t-1} + \underset{(0.0669)}{0.397}\ i_{t-3}$$

$$+\underset{(0.134)}{0.394}\ \Delta y_{t-1} + \underset{(0.127)}{0.348}\ \Delta y_{t-2} + \underset{(0.112)}{0.403}\ \Delta y_{t-5}$$

$$\widehat{\Delta y_t} = \underset{(0.00165)}{0.016} - \underset{(0.125)}{0.298}\ \Delta p_t + \underset{(0.0519)}{0.153}\ \Delta m_t + \underset{(0.0366)}{0.173}\ i_t - \underset{(0.0337)}{0.247}\ i_{t-2} - \underset{(0.0674)}{0.191}\ \Delta y_{t-5}$$

Estimation statistics:

		RSS	$\hat{\sigma}$	R^2	$\bar{R}^2$
Δp_t	:	0.0010	0.0026	0.8266	0.8210
Δm_t	:	0.0109	0.0086	0.4852	0.4747
i_t	:	0.0205	0.0120	0.8673	0.8598
Δy_t	:	0.0074	0.0071	0.3463	0.3237

SVAR model B

Estimation results:

$$\widehat{\Delta p_t} = \underset{(0.0656)}{0.594}\ \Delta p_{t-1} + \underset{(0.0678)}{0.324}\ \Delta p_{t-4} + \underset{(0.0107)}{0.047}\ i_{t-1} - \underset{(0.0107)}{0.0466}\ i_{t-5} + \underset{(0.0185)}{0.0509}\ \Delta m_{t-1}$$

$$\widehat{i_t} = -\underset{(0.00395)}{0.0123} + \underset{(0.213)}{0.857}\ \Delta p_t + \underset{(0.0659)}{0.566}\ i_{t-1} + \underset{(0.0666)}{0.341}\ i_{t-3} + \underset{(0.137)}{0.449}\ \Delta y_{t-1}$$

$$+\underset{(0.131)}{0.349}\ \Delta y_{t-2} + \underset{(0.115)}{0.413}\ \Delta y_{t-5}$$

$$\widehat{\Delta y_t} = \underset{(0.00159)}{0.0171} + \underset{(0.0341)}{0.122}\ i_t - \underset{(0.0338)}{0.226}\ i_{t-2} - \underset{(0.0691)}{0.19}\ \Delta y_{t-5}$$

$$\widehat{\Delta m_t} = -\underset{(0.0392)}{0.218}\ i_t + \underset{(0.0741)}{0.209}\ \Delta y_t + \underset{(0.0374)}{0.25}\ i_{t-2} + \underset{(0.0706)}{0.446}\ \Delta m_{t-1} + \underset{(0.0723)}{0.268}\ \Delta m_{t-2}$$

Estimation statistics:

		RSS	$\hat{\sigma}$	R^2	$\bar{R}^2$
Δp_t	:	0.0010	0.0026	0.8266	0.8219
i_t	:	0.0222	0.0124	0.8560	0.8500
Δy_t	:	0.0080	0.0074	0.2903	0.2758
Δm_t	:	0.0102	0.0084	0.5188	0.5056

SVAR model C

Estimation results:

$$\widehat{\Delta p_t} = \underset{(0.0656)}{0.594}\ \Delta p_{t-1} + \underset{(0.0678)}{0.324}\ \Delta p_{t-4} + \underset{(0.0107)}{0.047}\ i_{t-1} - \underset{(0.0107)}{0.0466}\ i_{t-5} + \underset{(0.0185)}{0.0509}\ \Delta m_{t-1}$$

$$\widehat{\Delta y_t} = \underset{(0.00186)}{0.0136} + \underset{(0.0781)}{0.176}\ \Delta y_{t-2} - \underset{(0.0445)}{0.222}\ i_{t-2} + \underset{(0.0461)}{0.125}\ i_{t-3}$$

$$\widehat{i_t} = -\underset{(0.00385)}{0.013} + \underset{(0.203)}{0.721}\ \Delta p_t + \underset{(0.124)}{0.45}\ \Delta y_t + \underset{(0.134)}{0.359}\ \Delta y_{t-1} + \underset{(0.112)}{0.413}\ \Delta y_{t-5}$$

$$+ \underset{(0.0629)}{0.691}\ i_{t-1} + \underset{(0.0828)}{0.473}\ i_{t-3} - \underset{(0.0729)}{0.245}\ i_{t-4} + \underset{(0.105)}{0.321}\ \Delta m_{t-3} - \underset{(0.107)}{0.246}\ \Delta m_{t-4}$$

$$\widehat{\Delta m_t} = \underset{(0.0741)}{0.209}\ \Delta y_t - \underset{(0.0392)}{0.218}\ i_t + \underset{(0.0374)}{0.25}\ i_{t-2} + \underset{(0.0706)}{0.446}\ \Delta m_{t-1} + \underset{(0.0723)}{0.268}\ \Delta m_{t-2}$$

Estimation statistics:

		RSS	$\hat{\sigma}$	R^2	$\bar{R}^2$
Δp_t	:	0.0010	0.0026	0.8266	0.8219
Δy_t	:	0.0084	0.0076	0.2550	0.2398
i_t	:	0.0194	0.0117	0.8742	0.8662
Δm_t	:	0.0102	0.0083	0.5188	0.5056

Causal Ordering as in Christiano et.al. (1996)

Estimation results:

$$\widehat{\Delta y_t} = \underset{(0.00226)}{0.0128} + \underset{(0.0822)}{0.216}\ \Delta y_{t-1} + \underset{(0.0761)}{0.096}\ \Delta y_{t-2} - \underset{(0.0722)}{0.144}\ \Delta y_{t-4} + \underset{(0.224)}{0.292}\ \Delta y_{t-5}$$

$$-\underset{(0.229)}{0.294}\ \Delta p_{t-2} - \underset{(0.0383)}{0.215}\ \Delta p_{t-3} + \underset{(0.0395)}{0.13}\ i_{t-2}$$

$$\widehat{\Delta p_t} = \underset{(0.0656)}{0.594}\ \Delta p_{t-1} + \underset{(0.0678)}{0.324}\ \Delta p_{t-4} + \underset{(0.0107)}{0.047}\ i_{t-1} - \underset{(0.0107)}{0.0466}\ i_{t-5} + \underset{(0.0185)}{0.0509}\ \Delta m_{t-1}$$

$$\widehat{i_t} = -\underset{(0.00385)}{0.013} + \underset{(0.124)}{0.45}\ \Delta y_t + \underset{(0.203)}{0.721}\ \Delta p_t + \underset{(0.134)}{0.359}\ \Delta y_{t-1} + \underset{(0.112)}{0.413}\ \Delta y_{t-5}$$

$$+ \underset{(0.0629)}{0.691}\ i_{t-1} + \underset{(0.0828)}{0.473}\ i_{t-3} - \underset{(0.0729)}{0.245}\ i_{t-4} + \underset{(0.105)}{0.321}\ \Delta m_{t-3} - \underset{(0.107)}{0.246}\ \Delta m_{t-4}$$

$$\widehat{\Delta m_t} = \underset{(0.0741)}{0.209}\ \Delta y_t - \underset{(0.0392)}{0.218}\ i_t + \underset{(0.0374)}{0.25}\ i_{t-2} + \underset{(0.0706)}{0.446}\ \Delta m_{t-1} + \underset{(0.0723)}{0.268}\ \Delta m_{t-2}$$

Estimation statistics:

		RSS	$\hat{\sigma}$	R^2	$\bar{R}^2$
Δy_t	:	0.0079	0.0074	0.2986	0.2642
Δp_t	:	0.0010	0.0027	0.8266	0.8219
i_t	:	0.0194	0.0117	0.8742	0.8662
Δm_t	:	0.0102	0.0083	0.5188	0.5056

Reduced-form VAR model

Estimation results:

$$\widehat{\Delta p_t} = \underset{(0.0656)}{0.594}\ \Delta p_{t-1} + \underset{(0.0678)}{0.324}\ \Delta p_{t-4} + \underset{(0.0107)}{0.047}\ i_{t-1} - \underset{(0.0107)}{0.0466}\ i_{t-5} + \underset{(0.0185)}{0.0509}\ \Delta m_{t-1}$$

$$\widehat{\Delta y_t} = \underset{(0.00226)}{0.0128} + \underset{(0.224)}{0.292}\ \Delta p_{t-2} - \underset{(0.229)}{0.294}\ \Delta p_{t-3} + \underset{(0.0822)}{0.216}\ \Delta y_{t-1}$$

$$+ \underset{(0.0761)}{0.096}\ \Delta y_{t-2} - \underset{(0.0722)}{0.144}\ \Delta y_{t-4} - \underset{(0.0383)}{0.215}\ \Delta y_{t-5} + \underset{(0.0395)}{0.13}\ i_{t-2}$$

$$\widehat{i_t} = \underset{(0.223)}{0.534}\ \Delta p_{t-2} + \underset{(0.116)}{0.296}\ \Delta y_{t-1} + \underset{(0.11)}{0.297}\ \Delta y_{t-5} + \underset{(0.0624)}{0.698}\ i_{t-1}$$

$$+ \underset{(0.087)}{0.39}\ i_{t-3} - \underset{(0.0769)}{0.241}\ i_{t-4} + \underset{(0.111)}{0.29}\ \Delta m_{t-3} - \underset{(0.111)}{0.24}\ \Delta m_{t-4}$$

$$\widehat{\Delta m_t} = - \underset{(0.0508)}{0.232}\ i_{t-1} + \underset{(0.0501)}{0.277}\ i_{t-2} + \underset{(0.0732)}{0.384}\ \Delta m_{t-1} + \underset{(0.0752)}{0.363}\ \Delta m_{t-2}$$

Estimation statistics:

		RSS	$\hat{\sigma}$	R^2	$\bar{R}^2$
Δp_t	:	0.0010	0.0026	0.8266	0.8219
Δy_t	:	0.0079	0.0074	0.2985	0.2642
i_t	:	0.0222	0.0125	0.8560	0.8489
Δm_t	:	0.0109	0.0086	0.4852	0.4747