Comparison of Model Reduction Methods for VAR Processes[1]

Ralf Brüggemann

Institut für Statistik und Ökonometrie, Humboldt–Universität zu Berlin, Spandauer Str. 1, D-10178 Berlin, GERMANY, Tel.: +49-30-2093-5603 Fax.: +49-30-2093-5712, email: brueggem@wiwi.hu-berlin.de

Hans-Martin Krolzig

Department of Economics, University of Oxford, Manor Road Building, Oxford OX1 3UQ, UK, and Nuffield College, Oxford, Tel: +44-1865-271085 Fax: +44-1865-271094, email: hans-martin.krolzig@nuf.ox.ac.uk

Helmut Lütkepohl

Humboldt–Universität zu Berlin and European University Institute, Florence, ITALY, Tel.: +39-055-4685271 Fax.: +39-055-4685202, email: helmut.luetkepohl@iue.it

Abstract

The objective of this study is to compare alternative computerized model-selection strategies in the context of the vector autoregressive (VAR) modeling framework. The focus is on a comparison of subset modeling strategies with the general-to-specific reduction approach automated by *PcGets*. Different measures of the possible gains of model selection are considered: (i) the chances of finding the 'correct' model, that is, a model which contains all necessary right-hand side variables and is as parsimonious as possible, (ii) the accuracy of the implied impulse-responses and (iii) the forecast performance of the models obtained with different specification algorithms. In the Monte Carlo experiments, the procedures recover the DGP specification from a large VAR with anticipated size and power close to commencing from the DGP itself when evaluated at the empirical size. We find that subset strategies and *PcGets* are close competitors in many respects, with the forecast comparison indicating a clear advantage of the *PcGets* algorithm.

JEL classification: C32, C51.

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

---

1 Introduction

Specifying a parsimonious model for the data generation process (DGP) of a set of variables is a crucial step in econometric analysis. Typically a general model or model class is chosen at the initial stage of the modeling exercise. If the model is found to be general enough to describe the relevant features of the DGP well, i.e., the model is congruent, usually restrictions are imposed to avoid overparameterization which in turn may result in inefficient estimates and forecasts as well as poor structural analysis. Therefore many tools have been suggested which can be used in this important part of the modeling task. Examples of such tools are statistical tests and model selection criteria. Moreover, different algorithms have been proposed which specify the sequence in which the tools should be used, in order to arrive at a useful model. In the context of econometric time series models so-called subset procedures are available, for example. One of their features is that they reduce the model along a specific path which is determined by a variable selection criterion or a sequence of statistical tests. As an extension of these procedures, expert system type algorithms have been developed recently for single equation models. They take advantage of the currently available substantial computing power and consider various reduction paths using a range of alternative model selection and model checking procedures.

The objective of this study is to compare straightforward subset modeling strategies with the more sophisticated reduction mechanism *PcGets* for the vector autoregressive (VAR) modeling framework. The question of primary interest is to determine whether there is any gain from and, if so, how much can be gained by using the more sophisticated apparatus incorporated in *PcGets* rather than just some standard statistical procedures. In this comparison different measures of the possible gains will be used. Specifically, we investigate the chances of finding the 'correct' model, that is, a model which contains all necessary right-hand side variables and is as parsimonious as possible. In addition, we check the accuracy of the implied estimated impulse responses and compare the forecast performance of the models obtained with different specification algorithms. Our comparison is based on Monte Carlo simulations and thereby it is in particular relevant for gaining information of the relative performance of the considered procedures in small samples.

The procedures to be compared are simple model selection strategies as discussed in Brüggemann & Lütkepohl (2001) on the one hand and the *PcGets* system on the other hand, as mentioned previously. The subset modeling procedures are based on sequential t-tests or model selection criteria such as Akaike's (1974) AIC, Hannan & Quinn's (1979) HQ or Schwarz's (1978) SC. Starting from a sufficiently large VAR model, zero restrictions are placed on the coefficients sequentially where a search for the best restriction is performed in each step. The subset procedures differ in the way they perform the sequential search for suitable restrictions. Because Brüggemann & Lütkepohl (2001) found little gains from considering the full system in each step we will focus on single equation search procedures in this study. In contrast, *PcGets* is a more sophisticated computer-automated approach to econometric modeling which achieves the model reduction by a joint selection and diagnostic testing process comparing different model reduction paths. Detailed

descriptions of all procedures included in the comparison will be given in the following section. Section 3 presents the Monte Carlo setup and discusses the results. Finally, conclusions are drawn in Section 4.

2 Model Selection Procedures

The comparison will be performed in the framework of standard finite order VAR(p) processes,

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{2.1}$$

where $y_t = (y_{1t}, \ldots, y_{Kt})'$ is a K-dimensional vector of observable time series variables, $A_1, \ldots, A_p$ are $(K \times K)$ VAR parameter matrices, ν is a $(K \times 1)$ intercept vector and u_t is a white noise process with zero mean and nonsingular, time invariant covariance matrix $E(u_t u_t') = \Sigma_u$.

The subset procedures as well as *PcGets* work on the individual equations of the model to determine possible zero restrictions, i.e., to eliminate lags of variables. If the restricted model is then used for impulse response analysis or prediction, a feasible GLS procedure is applied for estimating the full system at once. In the following we first present the subset procedures and then provide a brief description of the main features of the *PcGets* approach.

2.1 Subset Modeling Procedures

Individual parameter restrictions are placed on each system equation separately. Therefore, we write the k-th equation of (2.1) as

$$y_{kt} = x_{1t}\theta_1 + \cdots + x_{Nt}\theta_N + u_t, \quad t = 1, \ldots, T. \tag{2.2}$$

All right-hand side variables are denoted by x_{nt} including the constant. The variable selection criteria are of the form

$$CR(i_1, \ldots, i_n) = \log(SSE(i_1, \ldots, i_n)/T) + c_T n/T, \tag{2.3}$$

where $SSE(i_1, \ldots, i_n)$ is the sum of squared errors obtained by including $x_{i_1 t}, \ldots, x_{i_n t}$ as regressors in the regression model (2.2) and c_T is a sequence indexed by the sample size T. The following variable elimination strategies have, for instance, been considered in the literature (see, e.g., Lütkepohl (1991) and Brüggemann & Lütkepohl (2001)).

Full Search (FS)
Choose the regressors which minimize $CR(i_1, \ldots, i_n)$ for all subsets $\{i_1, \ldots, i_n\} \subset \{1, \ldots, N\}$ and $n = 0, \ldots, N$. □

This procedure requires a substantial computational effort if N is large because the set $\{1, \ldots, N\}$ has 2^N subsets. Consequently, 2^N models have to be compared. A computationally more efficient

procedure sequentially eliminates one variable in each step as follows. For simplicity we assume that the remaining variables are renumbered after each step such that in step j, $N - j + 1$ regressors are under consideration.

Sequential Elimination of Regressors (SER)

Sequentially delete those regressors which lead to the largest reduction of the given criterion until no further reduction is possible. Formally:

Step j: Delete x_{nt} if

$$CR(1, \ldots, n - 1, n + 1, \ldots, N - j + 1)$$
$$= \min_{l=1,\ldots,N-j+1} CR(1, \ldots, l - 1, l + 1, \ldots, N - j + 1)$$

and

$$CR(1, \ldots, n - 1, n + 1, \ldots, N - j + 1) \leq CR(1, \ldots, N - j + 1). \qquad \square$$

It is also possible, of course, to choose individual zero coefficients on the basis of the t-ratios of the parameter estimators. In that case, the regressors with the smallest absolute values of t-ratios are eliminated until they are all greater than some threshold value γ. A formal description of this procedure follows.

Testing Procedure (TP)

Let $t_n^{(j)}$ be the t-ratio from an OLS estimation associated with θ_n in the j-th step of the procedure. *Step j*: Delete x_{nt} if $|t_n^{(j)}| = \min_{i=1,\ldots,N-j+1} |t_i^{(j)}|$ and $|t_n^{(j)}| \leq \gamma$. Stop if all $|t_n^{(j)}| > \gamma$. $\qquad \square$

In each step of this procedure a single regressor is eliminated. Then new t-ratios are computed for the reduced model. It is shown by Brüggemann & Lütkepohl (2001) that this strategy is equivalent to the sequential elimination based on model selection criteria for a suitably chosen threshold value γ. More precisely, if $\gamma = \{[\exp(c_T/T) - 1](T - N + j - 1)\}^{1/2}$ in the jth step of the elimination procedure, the same final model is obtained that also results by sequentially minimizing the selection criterion defined by the penalty term c_T. Therefore it is apparent that the threshold value depends on the sample size, the number of regressors in the model and the selection criterion through c_T. The threshold values for the t-ratios correspond to the critical values of the tests. The well-known criteria AIC, HQ and SC may be used in these procedures. Their c_T values are $c_T = 2$, $c_T = 2 \log \log T$ and $c_T = \log T$, respectively. Figure 1 shows the implied significance level of TP as a function of the sample size T for a model with $N = 10$ regressors. For example, choosing a model by TP-AIC, TP-HQ or TP-SC with $T = 100$ observations roughly corresponds to eliminating all regressors with t-values which are not significant at the 18%, 10% or 5% level, respectively. Because SER and TP are equivalent, we only use TP in the Monte Carlo experiments with the critical values implied by the three selection criteria AIC, HQ and SC.

An alternative subset modeling strategy proceeds by checking the variables sequentially from

the last to the first. Formally it may be described in the following way.

Top Down Procedure (TD)

For $n = N, N-1, \ldots, 1$, delete x_{nt} if

$$CR(1, \ldots, n-1, n+1, j_n, \ldots, N_n) \leq CR(1, \ldots, j_n, \ldots, N_n),$$

where N_n is the largest lag remaining in the equation when x_{nt} is checked and $x_{j_n t}$ is the last previously checked regressor which remains in the equation. $\square$

Notice that the final outcome of this procedure may depend on the order in which the regressors are included in the model. Clearly, this introduces some arbitrariness into the procedure. In the present time series context it makes sense to check the higher order lags first before proceeding to smaller lags. Of course, there may be good arguments for alternative orderings in specific situations. For example, one may check the lags of one of the variables first and then turn to another variable and so on. In fact, this type of procedure was proposed by Hsiao (1979, 1982) for causality analysis. All subset procedures used in the Monte Carlo study have been implemented using Gauss and a convenient graphical user interface is provided by the software JMulTi.[2]

2.2 *PcGets*

PcGets developed by Hendry & Krolzig (2001) implements automatic general-to-specific (*Gets*) modeling for linear, dynamic, single-equation models based on the theory of reduction (see, *inter alia*, 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, 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 getting 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 and so are congruent, undominated, mutually encompassing representations, the reduction process recommences from their union, providing that is a reduction of the general unrestricted model (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 (for a detailed description see Hendry & Krolzig (1999), Krolzig & Hendry (2001) and Hendry & Krolzig (2001)). In the following we

---

[2]The software package JMulTi can be downloaded at: www.jmulti.de.

briefly discuss the econometrics of the different stages of the *PcGets* model-selection algorithm relevant for VAR modeling.

2.2.1 Statistical analysis of the GUM (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 only has to specify the variables of interest. An overall F-test of all regressors checks that there is something to model, misspecification tests check the congruence of the model. The initial maximum lag length can be given by the researcher or set automatically. The algorithm then determines the lag order of the process; the lag order pre-selection consists of F-tests on the longest-lag blocks till the null is rejected.

2.2.2 Pre-search reduction tests (Stage I)

PcGets then undertakes various 'pre-search' simplification F-tests to exclude variables from the GUM. Since variables found to be irrelevant by such tests are excluded from later analyses, this step uses a loose significance level (such as 50%). The first step analyses whether a block of regressors associated with a particular lag or variable can be eliminated from the system without losing relevant information. The next step consists of block (F) tests of groups of regressors, 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.

2.2.3 Recursive multiple-path encompassing search (Stage 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 variables having 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 is repeated till a unique contender emerges, or the previous union is reproduced, then it stops. In the latter case a final selection is made using information criteria, otherwise a unique congruent and encompassing reduction has been located.

2.2.4 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 samples 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.

2.2.5 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 & 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 & Krolzig (2001) propose a *'liberal'* and a *'conservative'* strategy which aim to provide maximum power at a controlled empirical size. Figure 1 illustrates the significance level of the strategies relative to AIC, SC and HQ (for 10 variables). The liberal strategy seeks to minimize the chances of omitting variables that matter, so uses a relatively loose significance level (with HQ as its upper and SC as its lower bound), whereas the Conservative uses a very stringent significance level (much tighter than SC) to minimize the chances of retaining variables that do not matter. The built-in strategies of *PcGets 1.1* will be used in the following Monte Carlo experiments.

3 Simulation Comparison

3.1 Monte Carlo Design

We have considered the following DGPs:

$$
\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix} + \begin{bmatrix} \alpha_{11,1} & \alpha_{12,1} \\ \alpha_{21,1} & \alpha_{22,1} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \alpha_{21,2} & 0 \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \tag{3.1}
$$

with identity white noise covariance matrix, $\Sigma_u = I_2$, and

$$
\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \\ 0 \end{bmatrix} + \begin{bmatrix} \alpha_{11,1} & 0 & 0 \\ 0 & \alpha_{22,1} & 0 \\ 0 & 0 & \alpha_{33,1} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} \alpha_{11,2} & \alpha_{12,2} & 0 \\ \alpha_{21,2} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \\ y_{3,t-2} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix} \tag{3.2}
$$

with white noise covariance matrix

$$
\Sigma_u = \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
$$

The ν and α coefficients are nonzero and assume values to be discussed shortly. A process similar to (3.1) was also used in simulations by Lütkepohl (1991), Brüggemann & Lütkepohl (2001) and Krolzig (2001). It is very simple and may in fact be too simple for our purposes. It is expected to give an advantage to the TD strategy because all nonzero coefficients are concentrated at the lowest lags. Whenever a nonzero coefficient appears, there is no zero coefficient at a corresponding lower lag. This feature contrasts with DGP (3.2), where nonzero coefficients appear, e.g., at lag 2 when the corresponding lag 1 coefficient is zero (see $\alpha_{12,i}$ and $\alpha_{21,i}$). The process may also be more difficult to specify correctly due to its higher dimension.

We generated $M = 1000$ sets of time series and applied the single equation strategies from Section 3 to the generated time series. To be more precise, we fitted VAR(4) models and then applied the model specification strategies. Of course, the coefficient matrices A_3 and A_4 of the true DGPs contain zeros only. We are interested in determining the ability of different procedures to detect the zero coefficients correctly without replacing too many parameters by zero. In other words, we are interested in the size and power of the procedures. Therefore, controlling the t-ratios is useful in our simulation comparison because a coefficient with a small population t-ratio is more likely to be restricted to zero than one with a larger t-ratio, given that the choice of restrictions in all model selection procedures to be compared is based to some extent on these quantities. For this reason we have chosen parameter values with preassigned absolute t-ratios.

Notice that in a regression model $y = X\theta + u$, the t-ratio of the kth coefficient is

$$
\mathsf{t}_k = \frac{\hat{\theta}_k}{\hat{\sigma}_{\theta_k}} = \frac{\sqrt{T}\hat{\theta}_k}{\hat{\sigma}_u \sqrt{(T^{-1}X'X)_{kk}^{-1}}}.
$$

The corresponding population statistic is

$$\mathsf{t}_k^* = \frac{\theta_k}{\sigma_{\theta_k}} = \frac{\sqrt{T}\theta_k}{\sigma_u \sqrt{Q_{kk}^{-1}}},$$

where $Q = \operatorname{plim} T^{-1}X'X$ and Q_{kk}^{-1} is the kth diagonal element of the inverse of Q. For our VAR(2) processes it can be shown that

$$Q = \begin{bmatrix} 1 & \tilde{\mu}' \\ \tilde{\mu} & \Gamma + \tilde{\mu}\tilde{\mu}' \end{bmatrix},$$

where $\tilde{\mu} = (1,1)' \otimes \mu$. Here $\mu = (I_K - A_1 - \cdots - A_p)^{-1}\nu$ is the mean of the VAR and Γ is the covariance matrix of (y_t, y_{t-1}). Hence, it is easy to work out the population t-ratios of the parameters in our DGPs. We have chosen the parameter values such that prespecified t-ratios of the nonzero coefficients are ensured. The actual parameter values for the DGPs are given in Tables 2 and 3. We also give the corresponding moduli of the nonzero reverse characteristic roots of the autoregressive polynomials $\det(I_K - A_1 z - \cdots - A_p z^p)$ in the tables. For example, for each of the DGPs (3.1) there are roots with two different nonzero moduli and there are also roots which are zero. Obviously, the larger the t-ratios, the closer the roots are to the unit circle. As these processes are closer to the nonstationarity region, we cannot expect that the selection properties of the reduction approaches will be monotonous functions of the underlying t-values. These and other characteristics of the DGPs will be of importance for the outcome of our experiment.

3.2 Evaluation of the Monte Carlo Results

3.2.1 Selection properties

The results for the DGPs (3.1) and (3.2) are summarized in Figures 2 - 7. In Figures 2 and 3 the average size and power of the different procedures are depicted. Here the average size is determined as the average relative frequency of including a zero coefficient. That is, the average rejection frequency of all coefficients which are actually zero is determined. The average power is the corresponding probability of including the nonzero coefficients. The average size is plotted on the horizontal axis whereas the power is shown on the vertical axis. For instance, in the upper left-hand panel of Figure 2, which shows results for the first equation of the DGP, the FS procedure combined with the SC criterion is seen to have a size of about 4% and a power of about 40%. For each of the procedures using model selection criteria, SC results in the smallest and AIC in the largest sizes whereas HQ is in between. To be able to assess the results, the theoretical power of a t-test of a coefficient with theoretical t-value corresponding to the one of the underlying DGP is shown by a solid line in the figures. That line is a power envelop because we cannot hope to do better than the nonsequential t-tests with our procedures. It is obvious that all the procedures are doing quite well in this respect. There are, however, some notable differences with respect to the empirical size. For example, subset procedures using the AIC criterion typically have a size around 20%, the size of HQ based algorithms is around 10% while the size of SC based strategies

is between 4 and 8%. The difference between FS, TP and TD is very small when the same selection criterion is used. The size of the conservative *PcGets* (*PcGets*-CON) algorithm is smallest of all selection methods, while the size of the liberal variant (*PcGets*-LIB) is typically very close to the SC based subset procedures, which is in line with the predictions of Figure 1.

In Figures 4 and 5 the overall size and power of the procedures are shown. The horizontal axis shows the estimated probability that at least one of the zero coefficients is included whereas the vertical axis shows the probability that all nonzero coefficients are included. The solid line is based on selecting the nonzero coefficients by multiple t-tests. Thus, it may again be regarded as a power envelop. In this case all procedures are a bit away from this envelop for DGPs with t-ratios less than 5. For the two-dimensional DGPs the TD strategy is slightly superior to the other subset modeling procedures which may be a consequence of the specific type of DGP, as mentioned earlier. Consequently, the ordering reverses in some cases when the three-dimensional DGPs are considered. In addition, for t-ratios less than 5 there is a clear trade-off between the probability of retaining irrelevant variables and selecting DGP variables. The liberal strategies based on AIC and HQ have the highest probability of including DGP variables but at the same time the probability of including non-DGP variables is also very high. In contrast, for $|t| > 5$ moving from AIC to SC based algorithms is associated with almost no loss in power (selection probability) while the probability of including non-DGP variables decreases drastically.

Finally, Figures 6 and 7 show the relative frequencies of finding precisely the true DGP, that is, all zero coefficients are deleted and all nonzero coefficients are included. In this respect the subset procedures are doing quite well relative to *PcGets* for processes with t-values smaller than 4 for the process types considered here. In fact, there are many cases when the simple subset strategies have a higher probability of finding the true DGP, however, the chance to find the DGP is very limited in any case. This has to be expected given the probability to find the DGP if one starts with it. However, for t-values larger than 4, the conservative variant of *PcGets* consistently selects the true DGP with highest probability, although some subset procedures, especially the parsimonious SC based methods, are close competitors.

3.2.2 Impulse response analysis

We now check how the selection properties of the proposed selection strategies are translated into the accuracy of the impulse responses implied by the empirical model.

The infinite-order vector moving-average representation of the VAR model in (2.1) is

$$y_t = \mu + \sum_{j=0}^{\infty} \Psi_j u_{t-j}, \tag{3.3}$$

where $\mu = (I_K - \sum_{i=1}^p A_i)^{-1} \nu$ and $\Psi_0 = I_K$. The matrix power series $\Psi(L) = \sum_{j=0}^{\infty} \Psi_j L^j$ is the inverse lag polynomial to $A(L) = I_K - \sum_{j=i}^p A_i L^i$. 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.

Figure 8 displays the mean squared error (MSE) of the impulse responses of the system for the 11 reduction approaches and the true VAR relative to the MSE for the full VAR(4). Let $\psi_h = \mathrm{vec}(\Psi_h)$ collect the response of the system to shocks h periods ago, then the graphs report for the various DGPs, each associated with a certain t-value, the relative MSE:

$$\frac{\mathsf{MSE}_{\mathcal{M}}(h)}{\mathsf{MSE}_{(4)}(h)}$$

for $h = 1$ and 5, where $\mathcal{M}$ is the reduction method, ψ_h is the theoretical response of the system given the specification of the DGP, and $\hat{\psi}_h$ is the estimated response of the system given the estimated VAR: $\hat{\Psi}(L) = \hat{A}(L)^{-1}$. The MSE is defined as

$$\mathsf{MSE}(h) = \frac{1}{M} \sum_{m=1}^{M} \left(\psi_h - \hat{\psi}_{h,m}\right)' \Sigma_{\psi_h}^{-1} \left(\psi_h - \hat{\psi}_{h,m}\right)$$

where Σ_{ψ_h} has been estimated using the simulated impulse-responses for the full VAR(4) and M is the number of replications in our simulation experiments.

Not surprisingly, using the true model always results in the smallest MSE. Obviously, using the reduction procedures may lead to substantial distortions for the initial impulse responses ($h = 1$, DGP (3.1)). Clearly they may impose false zero restrictions which in turn leads to a substantial bias in the corresponding estimated impulse responses towards zero and may result in relatively large MSEs. The problem diminishes when the coefficients have larger population t-ratios because in that case the reduction procedures are less likely to restrict them to zero. Also for larger lags the relative MSEs improve (see $h = 5$). For stationary processes the impulse responses taper off to zero as the lag goes to infinity. Therefore, imposing a false zero restriction and thereby forcing some estimated impulse responses to be zero at large lags may result in a more precise estimator with smaller MSE than an unconstrained full VAR model. This feature is clearly reflected in Figure 8, where the normalized MSEs of impulse responses based on restricted models are smaller than 1 for $h = 5$. Generally, when shrinking to zero helps improving the estimators, the more restrictive methods have an advantage over more liberal procedures and criteria.

3.2.3 Forecasting

To assess the forecasting power of the selected models relative to the full VAR(4) model we compare mean squared prediction errors (MSPEs) of 1- and 5-step ahead out-of-sample forecasts. More precisely, we compute the MSPE as

$$\mathsf{MSPE}(h) = \frac{1}{M} \sum_{m=1}^{M} (y_{T+h,m} - \hat{y}_T(h)_m)' \Sigma_h^{-1} (y_{T+h,m} - \hat{y}_T(h)_m) \tag{3.4}$$

where $y_{T+h,m}$ is the generated vector of the m-th K-variate time series for which the forecast is made, $\hat{y}_T(h)_m$ denotes the corresponding h-step ahead forecast at origin T from the model specification under consideration and Σ_h is the covariance matrix of the corresponding theoretical h-step

prediction error obtained when the forecast is based on the true DGP with known parameters. Here M is the number of replications of our simulation experiments. Now forecast precision is measured relative to the full VAR(4) as

$$\frac{\mathrm{MSPE}_{\mathcal{M}}(h)}{\mathrm{MSPE}_{(4)}(h)}, \tag{3.5}$$

where $\mathrm{MSPE}_{(4)}$ denotes the MSPE of the full VAR(4) model and $\mathrm{MSPE}_{\mathcal{M}}$ is the MSPE of the selected model.

In Figure 9 we give the 1- and 5-step ahead forecast MSPEs relative to the full VAR(4) model according to equation (3.5). The forecasts based on the 'true model' are obtained by estimating all nonzero coefficients of the DGP. In other words, the forecasts are affected by estimation uncertainty whereas the model selection uncertainty is removed. Therefore their precision is expected to be at the lower end of the forecasts based on model selection procedures. For DGPs (3.1) with population t-values 2 and 3 the subset procedures FS, TP and TD are associated with a loss in forecasting performance of up to 2%, while the *PcGets* selection gives a gain of 3.9-6.5% relative to a full VAR(4) model and is about as good as the precision of the 'true model' forecasts. With increasing population t-values the gain in forecasting performance increases only slightly but for $|\mathrm{t}| > 3$ all model selection methods lead to improved forecasts relative to the full VAR(4) model. A similar pattern can be observed for the 5-step forecasts. For $|\mathrm{t}| = 2$ all subset procedures and the conservative *PcGets* method are associated with a small loss ($< 1\%$) in forecasting power. Only the liberal variant of *PcGets* improves the forecasting power by 1.5%. Again, for $|\mathrm{t}| > 3$ model selection pays off, although gains are typically much smaller than for the 1-step ahead forecast ($< 3.7\%$). In all considered cases, *PcGets* selected models are better at forecasting than the subset procedures, although in some cases the differences are not very large. The difference between the subset strategies FS, TP and TD is usually fairly small. In fact, we often observe that for a given information criterion, FS, TP and TD lead to improvements with the same order of magnitude. Moreover, we typically observe that using the most parsimonious criterion (i.e. SC) leads to the largest improvement, especially for large population t-values.

In Figure 9 we also give results for the forecasting experiment for the 3-dimensional DGP (3.2). Similar to the results in the bivariate case, we find that *PcGets* consistently leads to the largest improvements in forecasting performance and again provides forecasts of similar precision as the 'true model'. Relative to a full VAR(4) model the improvement is 14.2% for $h = 1$ and 7.5% for $h = 5$, which compares to 10% ($h = 1$) and 5.8% ($h = 5$) for the best subset method available. From all subset methods the most parsimonious ones (with SC as an information criterion) work best. Consistent with the size and power results, TD often seems to have the edge when compared to FS and TP. As before, the gain for 1-step ahead forecasts is larger than for 5-step forecasts. In contrast to the bivariate case, all selected models now lead to improved forecast MSPEs, which might be an indication that model selection is especially useful in models with larger dimensions. Overall the forecast comparison indicates a clear advantage of the *PcGets* algorithm with the possible improvements being clearly DGP dependent.

11

4 Conclusions

In this paper we have evaluated alternative computerized model selection strategies for VAR models. In particular, we compared single path subset strategies based on information criteria (or sequential testing procedures) with the more sophisticated *PcGets* algorithm. The latter provides a computer automation of general-to-specific (*Gets*) modeling of linear, dynamic, regression models using multiple search paths along which the reduction of the model is conducted.

To assess the performance of alternative reduction methods we compared different measures of possible gains associated with model selection. Using Monte Carlo experiments, we first investigated the size and power properties. Moreover, we analyzed the chances of finding the 'correct' model, that is, a model which contains all right-hand side variables of the DGP and is as parsimonious as possible. Then the accuracy of the implied impulse-responses and the forecast performance of the models obtained with different specification algorithms were compared.

One aim of the paper was to evaluate alternative reduction strategies, to see if they worked well, indifferently, or failed badly. The results come much closer to the first. In fact, results from our Monte Carlo experiments show that the procedures recover the DGP specification from a large VAR with anticipated size, and power close to commencing from the DGP itself when evaluated at the empirical size. Results for the impulse response accuracy indicate many situations where model selection can successfully increase the precision relative to the full VAR model. However, there are also some cases (when the population t-value is small), where model selection strategies delete too many relevant variables and hence, adversely affect the impulse response accuracy. We find a similar picture for the forecasting exercise, where the largest gain from model selection is found in the trivariate systems.

The comparison of subset methods with *PcGets* revealed some interesting results: The simple subset methods are close competitors to the more sophisticated *PcGets* algorithm. In fact, in some cases a simple top down procedure has the best power-size trade-off. Moreover, for $|t| < 4$ there are some cases when subset strategies have the highest probability of finding the true DGP. On the other hand, for absolute t-values larger than 4, the conservative variant of *PcGets* consistently selects the true DGP with highest probability. Overall, for the DGPs considered here, the differences between *PcGets* and SC based subset methods are surprisingly small given the fact that *PcGets* is based on multiple search paths. But searching over multiple search paths obviously pays in some situations: The forecast comparison indicates a clear advantage of the *PcGets* algorithm for the DGPs considered in our Monte Carlo experiment.

In this study we only considered stationary DGPs. A natural extension will therefore include the investigation of different model selection methods in the presence of integrated and cointegrated time series.

References

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control* **AC-19**(6): 716–723.

Brüggemann, R. & Lütkepohl, H. (2001). Lag selection in Subset VAR models with an application to a U.S. monetary system, *in* R. Friedmann, L. Knüppel & H. Lütkepohl (eds), *Econometric Studies - A Festschrift in Honour of Joachim Frohn*, LIT: Münster, pp. 107–128.

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

Hannan, E. J. & Quinn, B. G. (1979). The determination of the order of an autoregression, *Journal of the Royal Statistical Society* **B41**: 190–195.

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

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

Hsiao, C. (1979). Autoregressive modeling of Canadian money and income data, *Journal of the American Statistical Association* **74**: 533–560.

Hsiao, C. (1982). Time series modelling and casual ordering of canadian money, income and interest rates, *in* O. Anderson (ed.), *Time Series Analysis: Theory and Practice I*, Amsterdam: North-Holland, pp. 671–699.

Krolzig, H.-M. (2001). General-to-specific reductions of vector autoregressive processes, *in* R. Friedmann, L. Knüppel & H. Lütkepohl (eds), *Econometric Studies - A Festschrift in Honour of Joachim Frohn*, LIT: Münster, pp. 129–157.

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

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

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

Table 1: The *PcGets* algorithm

Stage 0. **Statistical analysis of the GUM**

1. Estimation and test of the GUM;

2. Outlier correction;

3. Adjust significance level of diagnostics;

4. Lag order pre-selection.

Stage I. **Pre-search reductions**

1. Lags and variables pre-selection;

2. Sequential block reduction

   - Sort regressor in order of their t^2 values

   - Two-step top-down reduction and bottom-up reduction pathes;

   - Encompassing.

Stage II. **Recursive multiple-path encompassing search**

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

   (a) If a unique valid model survives, it becomes the selected model;

   (b) If there is a new set of non-dominated valid reductions, their union becomes the new general model and the search recommences;

   (c) If the algorithm converges to a particular set, no further reduction is feasible.

Stage III. **Sub-sample evaluation**

1. Test the significance of every selected variable in two overlapping sub-samples;

2. Penalize variables accordingly.

Table 2: Parameters for the 2-dimensional DGP (3.1) and $T = 100$

	ν_1	ν_2	$a_{11,1}$	$a_{21,1}$	$a_{12,1}$	$a_{22,1}$	$a_{21,2}$	Σ_u	modulus[a]
$|t| = 2$	0.213	0.214	0.192	0.196	-0.185	0.191	0.202	I_K	.43; .20
$|t| = 3$	0.352	0.353	0.276	0.289	-0.250	0.267	0.307	I_K	.57; .23
$|t| = 4$	0.531	0.538	0.348	0.379	-0.286	0.323	0.419	I_K	.68; .26
$|t| = 5$	0.748	0.787	0.407	0.467	-0.298	0.358	0.538	I_K	.76; .28

[a]Modulus of nonzero reverse characteristic roots of VAR process.

Table 3: Parameters for the 3-dimensional DGP (3.2) and $T = 100$

	ν_1	ν_2	$a_{11,1}$	$a_{22,1}$	$a_{33,1}$	$a_{11,2}$	$a_{21,2}$	$a_{12,2}$	σ_1^2	modulus[a]
$|t| = 2$	0.212	0.214	0.193	0.192	0.196	0.193	0.192	-0.187	0.951	.53; .36; .20
$|t| = 3$	0.320	0.360	0.279	0.275	0.287	0.280	0.299	-0.237	0.745	.65; .41; .29
$|t| = 4$	0.502	0.575	0.357	0.346	0.371	0.363	0.366	-0.287	0.706	.75; .43; .37
$|t| = 5$	0.598	0.930	0.429	0.403	0.447	0.448	0.535	-0.244	0.389	.82; .45; .44

[a]Modulus of nonzero reverse characteristic roots of VAR process.
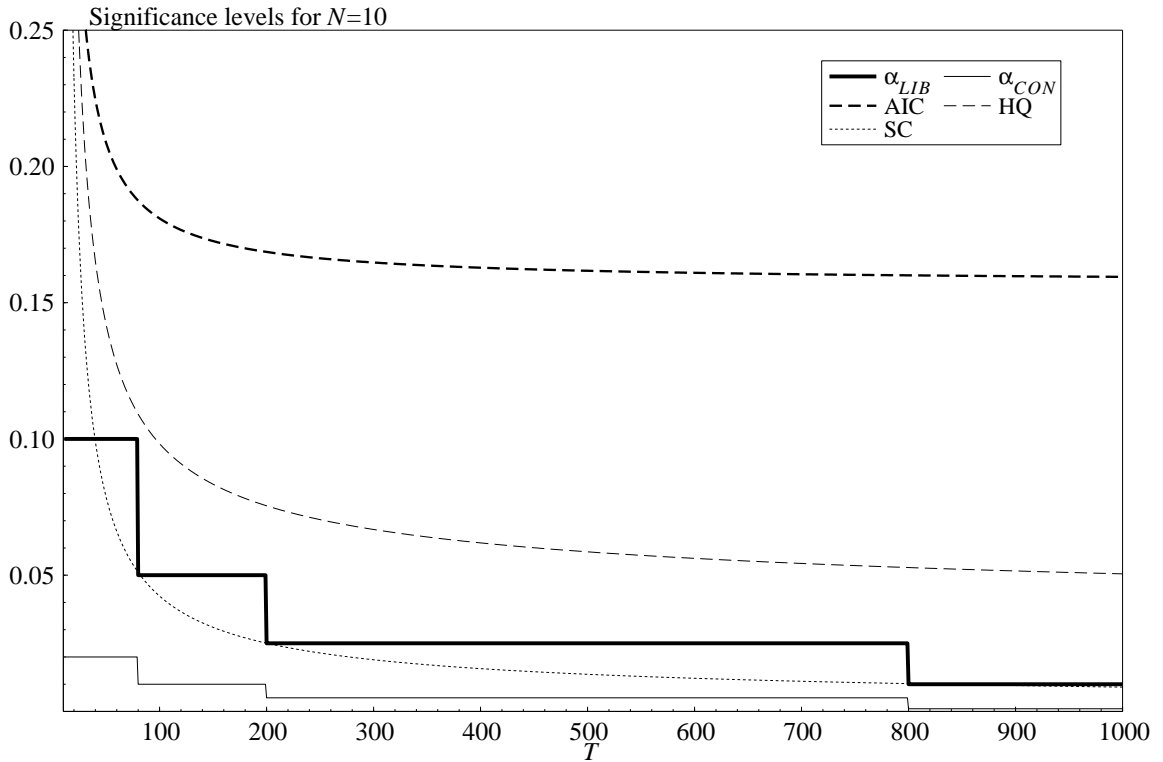
Figure 1: Significance level of t-tests used by the liberal (α_{LIB}) and conservative (α_{CON}) strategy of *PcGets* and implied by the information criteria AIC, HQ and SC
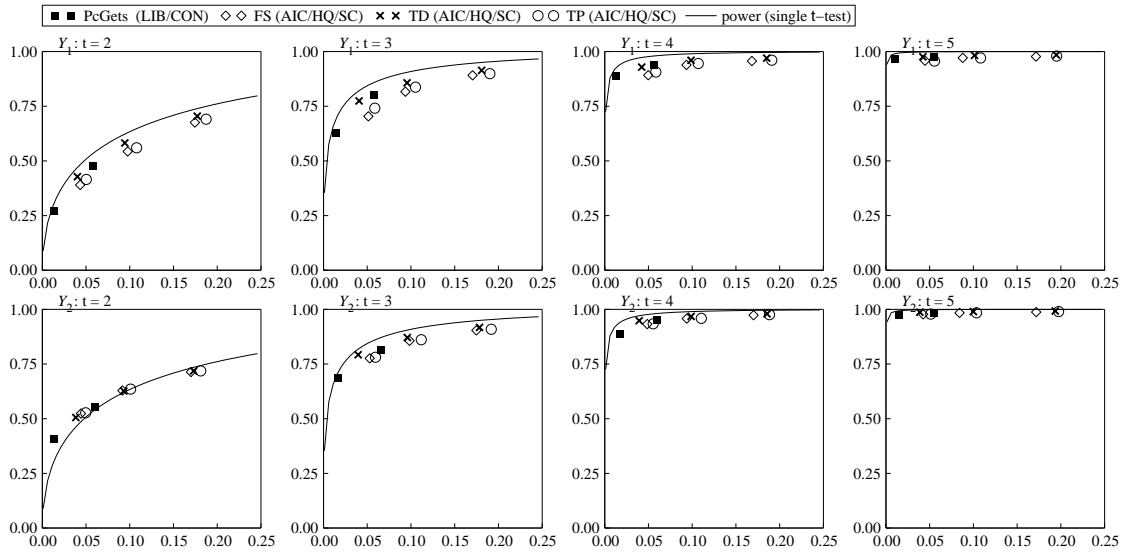
15

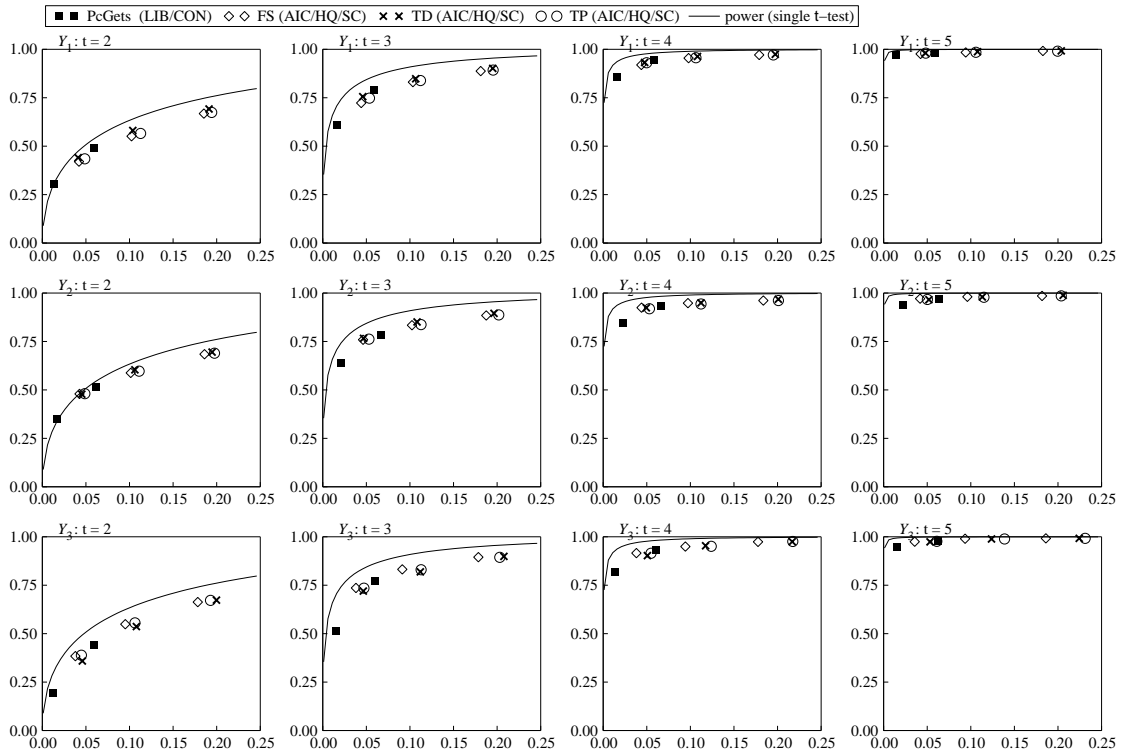Figure 2: DGP (3.1): Power – size trade-off

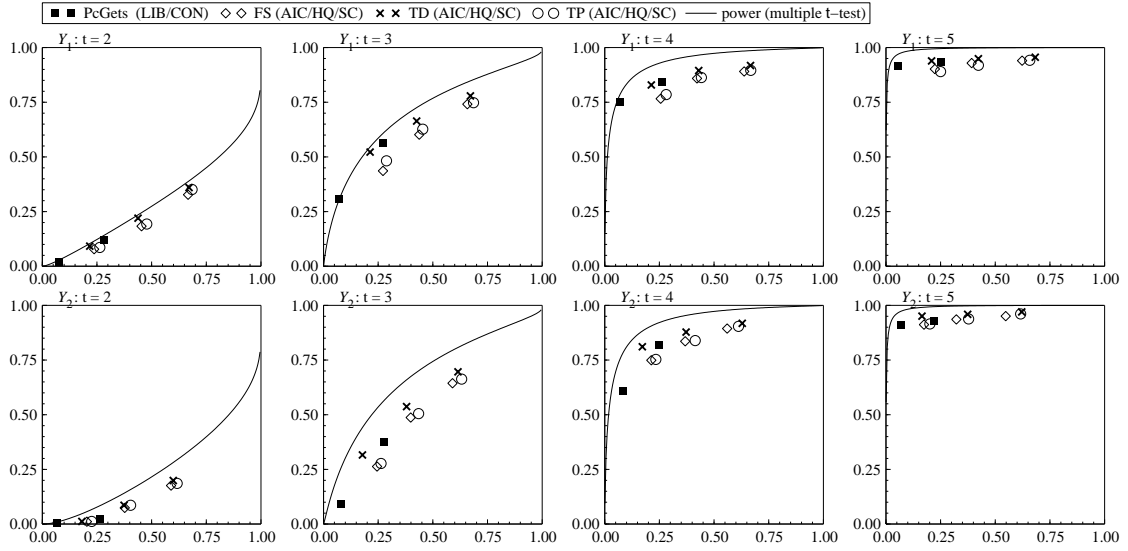Figure 3: DGP (3.2): Power – size trade-off

16

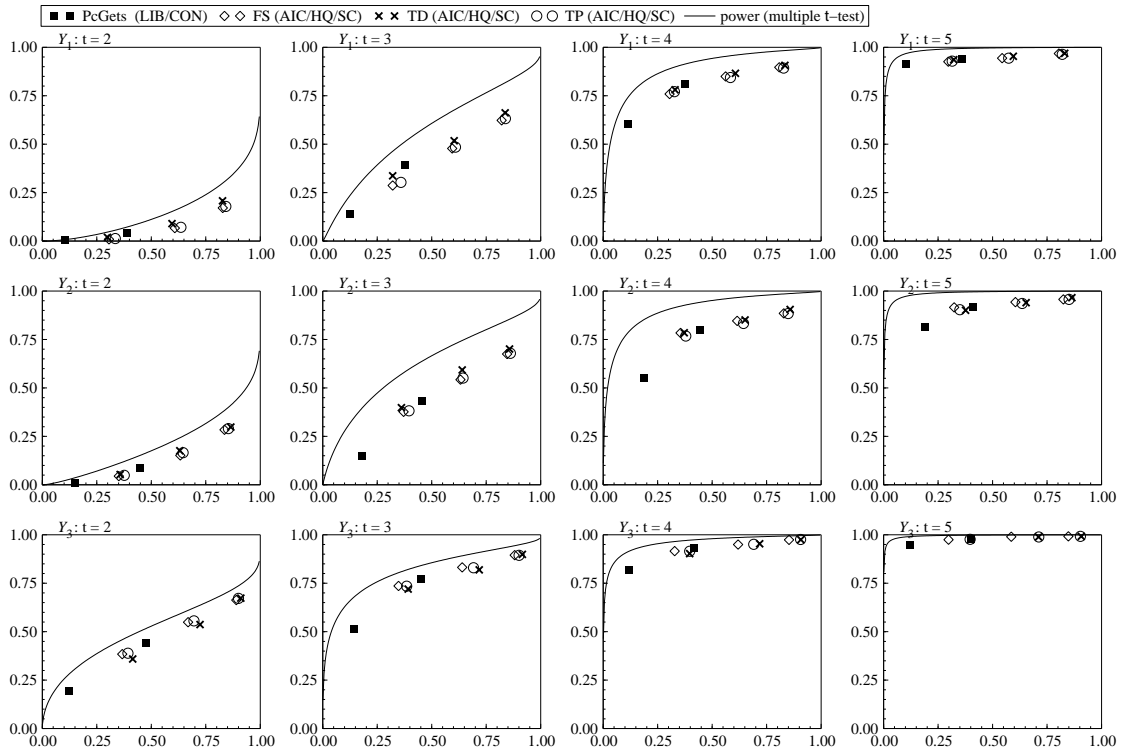Figure 4: DGP (3.1): Selection versus non-deletion probabilities

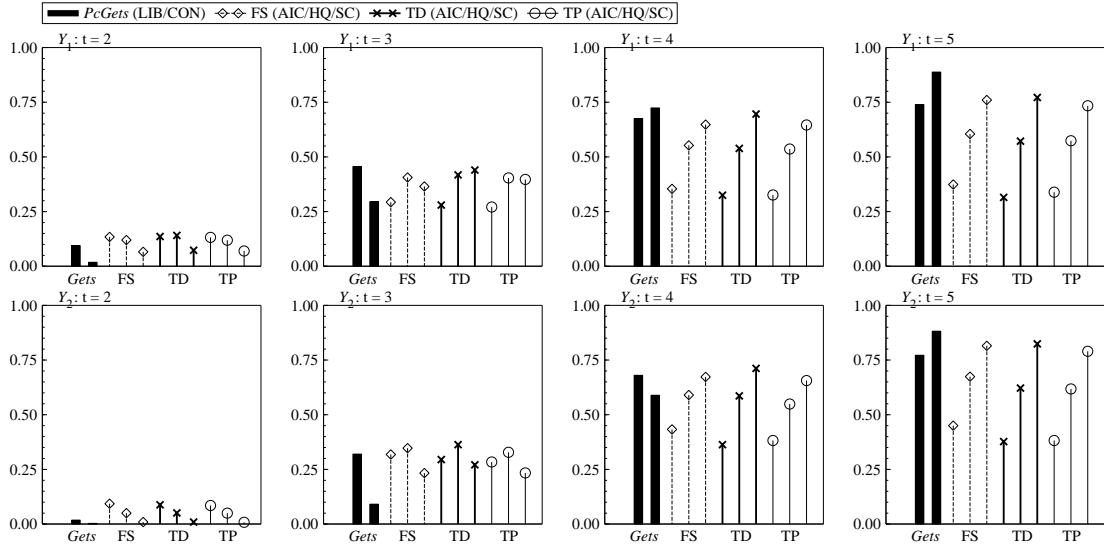Figure 5: DGP (3.2): Selection versus non-deletion probabilities

17

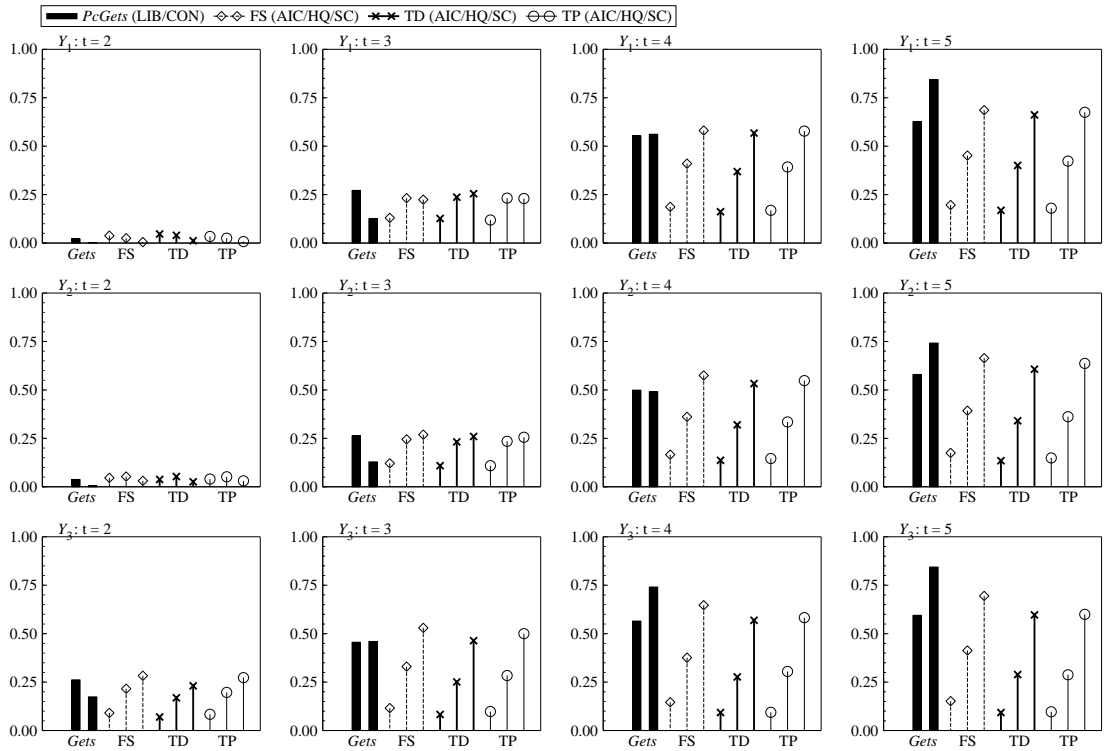Figure 6: Probability of finding the DGP (3.1)
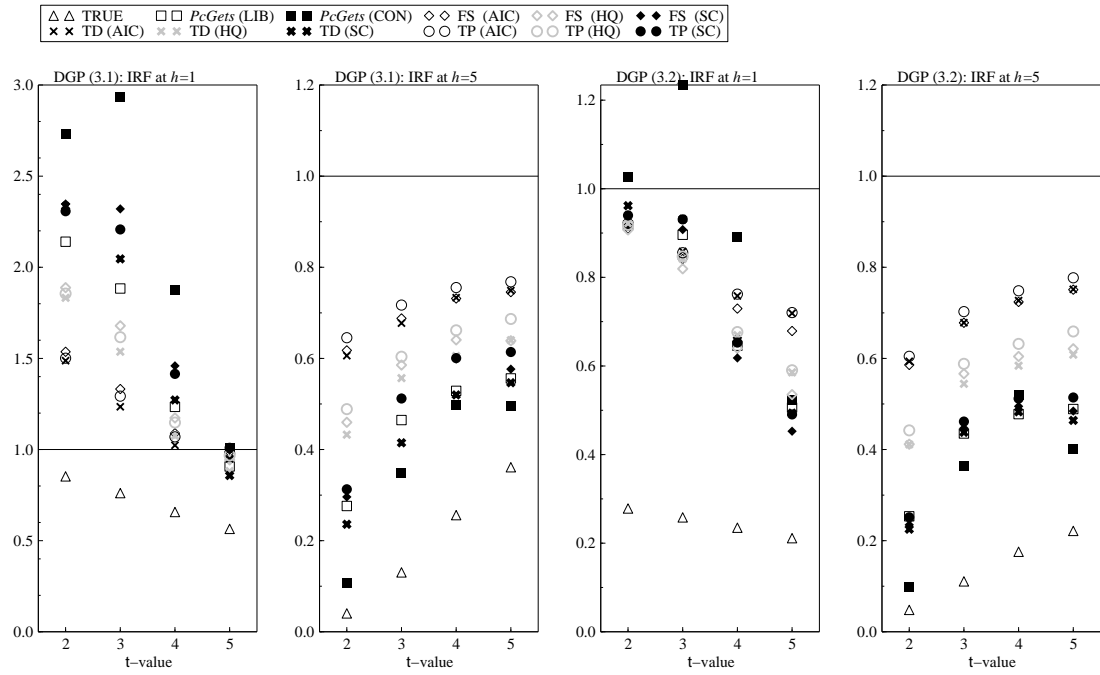

Figure 7: Probability of finding the DGP (3.2)

Figure 8: Normalized MSE of impulse-responses relative to the unrestricted VAR(4)
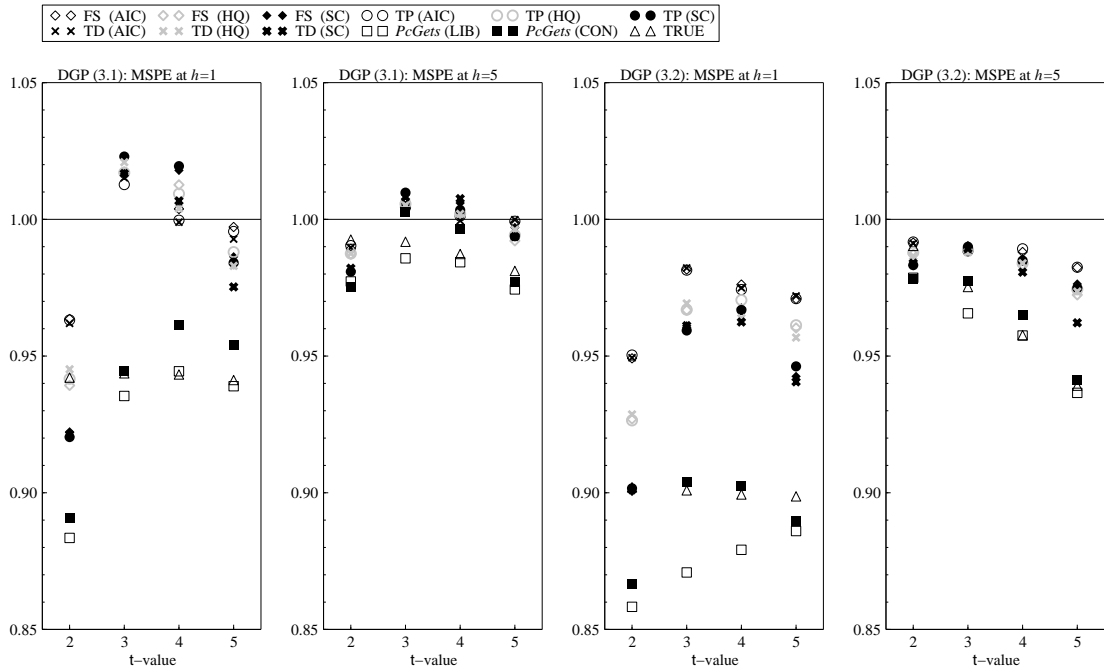

Figure 9: Normalized MSE of predictions relative to the unrestricted VAR(4)