January 27, 2000

Preliminary and incomplete!

Lag Selection in Subset VAR Models with an Application to a U.S. Monetary System[1]

Ralf Brüggemann and Helmut Lütkepohl

Institut für Statistik und Ökonometrie
Humboldt–Universität zu Berlin
Spandauer Str. 1
10178 Berlin
GERMANY
Tel.: +49-30-2093-5603/2093-5718
Fax.: +49-30-2093-5712
email: brueggem@wiwi.hu-berlin.de / luetke@wiwi.hu-berlin.de

Abstract

In this paper we consider alternative modeling strategies for specification of subset VAR models. We present four strategies and show that under certain conditions a testing procedure based on t-ratios is equivalent to eliminating sequentially lags that lead to the largest improvement in a prespecified model selection criterion. One finding from our Monte Carlo study is that differences between alternative strategies are small. Moreover, all strategies often fail to discover the true model. We argue that finding the correct model is not always the final modeling objective and find that using subset strategies results in models with improved forecast precision. To illustrate how these subset strategies can improve results from impulse response analysis, we use a VAR model of monetary policy shocks for the U.S. economy. While the response patterns from full and subset VARs are qualitatively identical, confidence bands from the unrestricted model are considerably wider. We conclude that subset strategies can be useful modeling tools when forecasting or impulse response analysis is the main objective.

Keywords: Model selection, monetary policy shocks, subset models, vector autoregressions

[1]We thank the Deutsche Forschungsgemeinschaft, SFB 373, for financial support.

1 Introduction

Following Sims' (1980) critique of classical econometric modeling, empirical macroeconomic studies are often based on vector autoregressive (VAR) models. In these models the relations between the variables are usually investigated within an impulse response analysis or innovation accounting. A criticisms that has been raised against this modeling strategy is that the number of parameters quickly becomes large if a moderate number of variables is considered and no or only a few restrictions are placed on the parameter matrices. In that case the sampling uncertainty in the estimated models makes it difficult to discriminate between different theories. Moreover, a theoretical problem related to the inference on impulse responses was pointed out by Benkwitz, Lütkepohl & Neumann (2000) and Benkwitz, Lütkepohl & Wolters (2000). These authors argue that standard bootstrap procedures which are often used for setting up confidence intervals for impulse responses may be grossly distorted if zero coefficients are estimated unrestrictedly. They advocate subset VAR models where zero restrictions are placed on some of the coefficients.

One possible approach is to decide on the restrictions on the basis of sample information and exclude, for example, insignificant lags of some of the variables. Using a statistical procedure for deciding on possible constraints may be advantageous compared to a procedure which is based on a priori economic theory if one desires to avoid biasing the results towards a particular theory at an early stage of an analysis. Therefore, in this study we will compare alternative statistical procedures that have been proposed and used for lag length selection in multivariate time series models. Typically applied researchers use testing procedures or model selection criteria in placing restrictions on a given VAR. We will compare both types of procedures in the following.

One testing procedure which is used occasionally is based on the t-ratios of the variables and eliminates the variables with lowest t-ratios sequentially until all remaining variables have t-ratios greater than some threshold value, say 2. We will discuss under what conditions such a procedure is equivalent to eliminating sequentially those lags of variables which lead to the largest improvement when the usual model selection criteria are applied instead of statistical tests. Moreover, we will compare these strategies with a full search procedure which chooses the restrictions that lead to the best overall model for a given model selection criterion.

The structure of this study is as follows. In the next section the model framework is presented and some alternative model selection strategies are considered in Sec. 3. In that section we distinguish between procedures which are based on single equation procedures, that is, procedures which treat the individual equations of a system separately and procedures which consider the full system at once. In Sec. 4, the results of a small sample comparison based on a Monte Carlo study are reported. In Sec. 5 we use U.S. macroeconomic data to illustrate the use of selection procedures and the effects of restrictions on impulse response analysis. Conclusions follow in Sec. 6.

2 Vector Autoregressive Models

The characteristics of the variables involved determine to some extent which model is a suitable representation of the data generation process (DGP). For instance, trending properties and seasonal fluctuations are of importance in setting up a suitable model. In the following we will focus on systems which contain potentially $I(0)$ and $I(1)$ variables.

Given a set of m time series variables $y_t = (y_{1t}, \ldots, y_{mt})'$, the basic VAR model considered in the following has the form

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

where ν is a constant $(m \times 1)$ vector, $A = [A_1 : \cdots : A_p]$, the A_i are $(m \times m)$ coefficient matrices, $Y_{t-1}^{t-p} = (y_{t-1}', \ldots, y_{t-p}')'$ and $u_t = (u_{1t}, \ldots, u_{mt})'$ is an unobservable zero mean white noise process with time invariant positive definite covariance matrix Σ_u. That is, the u_t are serially uncorrelated or independent. The model (2.1) is briefly referred to as a VAR(p) process because the number of lags is p. It is straightforward to introduce further deterministic terms such as seasonal dummy variables or polynomial trend terms in the model or include further exogenous variables. We use the simple model form (2.1) mainly for convenience in the following.

A VAR(p) process is *stable* if

$$\det(I_K - A_1 z - \cdots - A_p z^p) \neq 0 \quad \text{for} \quad |z| \leq 1. \tag{2.2}$$

Assuming that it has been initiated in the infinite past, it generates stationary time series which have time invariant means, variances and autocovariance structure. If the determinantal polynomial in (2.2) has roots for $z = 1$ (i.e., unit roots), then some or all of the

variables are $I(1)$ and they may also be cointegrated. Thus, the present model is general enough to accommodate variables with stochastic trends.

Clearly, the model (2.1) is in reduced form because all right-hand side variables are predetermined or deterministic and no instantaneous relations are modeled. Sometimes it is of interest to model also the instantaneous relations. In that case it may be useful to consider a structural form model,

$$A_0 y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t. \tag{2.3}$$

Of course, restrictions have to be imposed to identify the parameters of this model. For example, A_0 may be triangular and Σ_u diagonal so that the system is recursive.

3 Lag Order Selection Strategies

3.1 Single Equation Approaches

We begin with strategies which are based on the individual equations of the above system. Therefore we consider an equation of the form

$$y_t = \nu + x_{1t}\theta_1 + \cdots + x_{Kt}\theta_K + u_t, \quad t = 1, \ldots, T. \tag{3.1}$$

Here the right-hand side variables denoted by x_{kt} may include deterministic variables or unlagged endogenous variables if the equation belongs to a structural form. We wish to compare the following variable elimination strategies.

Full Search (FS)
Consider a criterion of the form

$$CR(i_1, \ldots, i_n) = \log(SSE(i_1, \ldots, i_n)/T) + c_T n/T, \tag{3.2}$$

where $SSE(i_1, \ldots, i_n)$ is the sum of squared errors obtained by including $x_{i_1 t}, \ldots, x_{i_n t}$ in the regression model (3.1) and c_T is a sequence indexed by the sample size. Choose the regressors which minimize $CR(i_1, \ldots, i_n)$ for all subsets $\{i_1, \ldots, i_n\} \subset \{1, \ldots, K\}$ and $n = 0, \ldots, K$. □

Clearly, this procedure involves a substantial computational effort if K is large. More precisely, the set $\{1, \ldots, K\}$ has 2^K subsets and, hence, there are as many models that have to be compared. The following elimination procedures proceed sequentially and are computationally less demanding. One variable only is eliminated in each step. For simplicity we assume that the remaining variables are renumbered after each step so that in Step j, $K - j + 1$ regressors are under consideration.

Sequential Elimination of Regressors (SER)

Sequentially delete those regressors which lead to the largest reduction in a given criterion of the type (3.2) until no further reduction is possible. Formally:

Step j: Delete x_{kt} if

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

and $CR(1, \ldots, k-1, k+1, \ldots, K-j+1) \leq CR(1, \ldots, K-j+1)$. □

Testing Procedure (TP)

Delete sequentially those regressors with the smallest absolute values of t-ratios until all t-ratios (in absolute value) are greater than some threshold value γ. Note that a single regressor is eliminated in each step only. Then new t-ratios are computed for the reduced model. Formally the procedure may be described as follows:

Let $t_k^{(j)}$ be the t-ratio associated with θ_k in the jth step of the procedure.

Step j: Delete x_{kt} if $|t_k^{(j)}| = \min_{i=1,\ldots,K-j+1} |t_i^{(j)}|$ and $|t_k^{(j)}| \leq \gamma$. Stop if all $|t_k^{(j)}| > \gamma$. □

The following proposition gives conditions under which these two lag selection procedures are equivalent.

Proposition 1

For given K and T, TP and SER lead to the same final model if the threshold value γ in TP is chosen as a function of the step j as follows: $\gamma = \gamma_j = \{[\exp(c_T/T) - 1](T - K + j - 1)\}^{1/2}$.

Note that the result in the proposition is purely algebraic and does not involve distributional assumptions. Of course, the t-ratios are not assumed to have actually a t- or standard normal

distribution. At this stage we do not even assume that all parameters are identified if (3.1) is a structural form equation. Also the proposition remains true if the search is refined to a subset of the regressors in the model (3.1). The proposition implies a computationally efficient way to determine the regressor whose elimination will lead to the greatest reduction in any one of the usual model selection criteria. We do not have to estimate all possible models with one coefficient restricted to zero but we just have to check the t-ratios of the coefficients.

Proof: For simplicity we assume that $K - j + 1 = n$ so that $x_{1t}, \ldots, x_{nt}$ are the regressors included before the jth step is performed. We show that both strategies eliminate the same regressor in that step. The squared t-ratio of the kth regressor is

$$\begin{aligned} t_k^2 &= (T-n)\tfrac{SSE(1,\ldots,k-1,k+1,\ldots,n)-SSE(1,\ldots,n)}{SSE(1,\ldots,n)} \\ &= (T-n)\left(\tfrac{SSE(1,\ldots,k-1,k+1,\ldots,n)}{SSE(1,\ldots,n)} - 1\right) \end{aligned}$$

(see Judge et al. (1988), Sec. 6.4). Obviously, t_k^2 is minimal if $(SSE(1,\ldots,k-1,k+1,\ldots,n)/T)/(SSE(1,\ldots,n)/T)$ and, hence,

$$\log(SSE(1,\ldots,k-1,k+1,\ldots,n)/T) - \log(SSE(1,\ldots,n)/T)$$

is minimal. Therefore the two strategies eliminate the same regressor if a regressor is deleted at all which happens if $t_k^2 \leq \gamma_j^2$. Choosing γ_j as in the proposition, this is equivalent to

$$\begin{aligned} &(T-n)\left(\tfrac{SSE(1,\ldots,k-1,k+1,\ldots,n)}{SSE(1,\ldots,n)} - 1\right) \leq \left[\exp\left(\tfrac{c_T}{T}\right) - 1\right](T-n) \\ &\iff \tfrac{SSE(1,\ldots,k-1,k+1,\ldots,n)/T}{SSE(1,\ldots,n)/T} \leq \exp\left(\tfrac{c_T}{T}\right) \\ &\iff \log(SSE(1,\ldots,k-1,k+1,\ldots,n)/T) - \log(SSE(1,\ldots,n)/T) \leq c_T/T \\ &\iff \log(SSE(1,\ldots,k-1,k+1,\ldots,n)/T) + \tfrac{c_T}{T}(n-1) \leq \log(SSE(1,\ldots,n)/T) + \tfrac{c_T}{T}n \\ &\iff CR(1,\ldots,k-1,k+1,\ldots,n) \leq CR(1,\ldots,n) \end{aligned}$$

which proves the proposition. □

In the simulation comparison in Sec. 4 we will also consider the following so-called top-down strategy which checks the importance of the regressors in reverse order of there subscripts (see Lütkepohl (1991, Ch. 5)). Thus, the procedure depends to some extent on the numbering of the regressors.

Top-Down Strategy (TD)

Delete the last regressor x_{Kt} from the equation if its removal does not increase a prespecified model selection criterion, CR say, keep it otherwise. Repeat the procedure for $x_{K-1,t}, \ldots, x_{1t}$. Formally:

Step j: Delete $x_{K-j+1,t}$ if removal does not increase CR. Keep it otherwise. □

Typical criteria used for time series model selection are

$$AIC(i_1, \ldots, i_n) = \log(SSE(i_1, \ldots, i_n)/T) + 2n/T,$$

(see Akaike (1974)),

$$HQ(i_1, \ldots, i_n) = \log(SSE(i_1, \ldots, i_n)/T) + \frac{2 \log \log T}{T} n$$

(see Hannan & Quinn (1979)) and

$$SC(i_1, \ldots, i_n) = \log(SSE(i_1, \ldots, i_n)/T) + \frac{\log T}{T} n$$

(see Schwarz (1978)). Because a time series length of about $T = 100$ is not untypical in macroeconomic studies with quarterly data we give some relevant γ_j for these three model selection criteria in Table 1. Obviously, the γ_j are in the range starting roughly at the 0.70 quantile of the standard normal distribution and reaching far out in its upper tail. With T=100 and K=12, for example, choosing a model by HQ roughly corresponds to eliminating all regressors with t-values which are not significant at 10% level. Because the t-ratios in the presently considered models with possibly integrated variables can be far from standard normal, the proposition also shows that using the SER procedure with standard model selection criteria may be problematic. All three lag selection procedures can be applied directly for choosing the right-hand side variables in a specific equation of the levels VAR model (2.1) or the structural form (2.3).

3.2 Systems Approaches

It is also possible to consider the complete system at once in selecting the lags and variables to be included. In that case a multivariate model of the type

$$y_t = X_t \theta + u_t, \qquad t = 1, \ldots, T, \tag{3.3}$$

is considered. Here X_t is a $(m \times J)$ regression matrix and θ is a $(J \times 1)$ parameter vector. Note that X_t is typically block-diagonal,

$$X_t = \begin{bmatrix} \mathbf{x}'_{1t} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}'_{2t} & & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{x}'_{mt} \end{bmatrix},$$

where $\mathbf{x}_{kt}$ is the $(J_k \times 1)$ vector of regressors in the kth equation and $J = J_1 + \cdots + J_m$. For example, if the full model (2.1) is considered, $X_t = I_m \otimes [1 : Y_{t-1}^{t-p'}]$, so that $J_k = mp + 1$ $(k = 1, \ldots, m)$. If lags are eliminated in some or all of the equations, the structure of X_t will usually become slightly more complicated with different numbers of regressors in the different equations. In that case, estimation is usually done by a GLS or SURE procedure, that is, the estimator of θ is

$$\hat{\theta} = \left(\sum_{t=1}^{T} X'_t \hat{\Sigma}_u^{-1} X_t \right)^{-1} \sum_{t=1}^{T} X'_t \hat{\Sigma}_u^{-1} y_t, \tag{3.4}$$

where $\hat{\Sigma}_u$ is a consistent estimator of Σ_u. For instance, $\hat{\Sigma}_u$ may be based on the residuals from an unrestricted or restricted OLS estimation of each individual equation.

Lag selection may again be based on sequential elimination of coefficients with smallest t-ratios. If the procedure is based on the full system we abbreviate it as STP. Alternatively, model selection criteria may be used in deciding on variables and lags to be eliminated. In the systems context, criteria of the type

$$VCR(\mathbf{j}) = \log \det(\tilde{\Sigma}_u(\mathbf{j})) + c_T J/T \tag{3.5}$$

are used, where c_T is as in (3.2), $\mathbf{j}$ indicates the restrictions placed on the model and $\tilde{\Sigma}_u(\mathbf{j})$ is a corresponding estimator of Σ_u. For example, starting from a full model such as (2.1), $\mathbf{j}$ may be an $((m^2p + m) \times 1)$ vector of zeros and ones where a one stands for a right-hand side variable which is included and a zero indicates a variable which is excluded. Ideally the ML estimator is used for $\tilde{\Sigma}_u(\mathbf{j})$. However, for restricted systems computing the exact ML estimator involves computationally demanding iterative methods in general. Therefore, using

$$\tilde{\Sigma}_u(\mathbf{j}) = T^{-1} \sum_{t=1}^{T} \hat{u}_t \hat{u}'_t \quad \text{with} \quad \hat{u}_t = y_t - X_t \hat{\theta} \quad (t = 1, \ldots, T) \tag{3.6}$$

is a feasible alternative. We will use this estimator in the following in combination with two types of selection strategies considered in the single equation context, FS and SER. In the

systems context they will be abbreviated as SFS and SSER. We do not use the Top-Down strategy in the system framework because it may result in a somewhat arbitrary model specification.

Note that the simple relation between STP and SSER given in Proposition 1 for the single equation case is no longer available in the presently considered systems case. The reason is that the Wald and the LR versions of tests for linear restrictions differ in the multivariate case. The usual t-test may be viewed as a Wald test whereas the LR version has the direct link to the lag selection criteria of the type (3.5) (see Lütkepohl (1991), Ch. 4). Given that the test versions are closely related, it is of course possible that both strategies lead to the same models in relevant small sample situations.

4 Simulation Comparison of Lag Selection Strategies

We have considered processes of different orders, dimensions and correlation characteristics to study the small sample properties of the selection strategies. In the following we will highlight some important findings and illustrate them with results from one example process (to be completed).

4.1 Single Equation Strategies

Many of the most important findings of our simulation study can be illustrated with the following stable bivariate VAR process from Lütkepohl (1991):

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} .02 \\ .03 \end{bmatrix} + \begin{bmatrix} .5 & .1 \\ .4 & .5 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ .25 & 0 \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \tag{4.1}$$

with covariance matrix

$$\Sigma_u = \begin{bmatrix} .09 & 0 \\ 0 & .04 \end{bmatrix}.$$

We simulated 1000 sets of time series. Then we applied the single equation strategies from Section 3 to the generated time series. Specifically, we applied the subset strategies to a VAR(3) process and thus the last coefficient matrix A_3 contains zeros only. For every

realization in the experiment we recorded whether the strategies decided correctly on the zero restrictions of individual coefficients and list the resulting relative frequencies of correct decisions.

We present results for $T = 30$ in Table 2. As there is very little sample information in this case, none of the criteria and strategies detects the zero elements with high probability. Especially, the small nonzero upper right element of the matrix A_1, denoted as $a_{12,1}$, is set to zero fairly often.

Comparing the results from SER/TP with those from FS it is interesting to note that the latter performs slightly better for coefficients which are actually zero. This can be seen from the values of A_3, for example. In contrast, SER/TP performs better for the nonzero coefficients.

For the TD strategy we find in all cases higher individual frequencies of correct decisions than for the SER/TP procedure, indicating that the TD procedure performs somewhat better than SER/TP for this particular process. With a few exceptions for small nonzero coefficients, TD performs better than FS when the actual coefficient is not equal to zero and vice versa.

On average the SC criterion selects models with more zero restrictions than HQ and AIC. This result is in line with the theoretical properties of these criteria (see Lütkepohl (1991, Ch. 4)). In particular, the SC criterion is less successful when the true coefficient is not equal but close to zero. On the other hand, if the true coefficient is zero, the SC criterion performs better than AIC and HQ. With $T = 30$, the difference between HQ and AIC is small in all subset strategies.

To get a more complete picture of the quality of the models selected, we also computed normalized forecast mean squared errors (MSEs) by adjusting for the theoretical forecast error covariance matrix. To be more precise, we denote the h-step forecast at time T of the n-th generated times series as $\hat{y}_T(h)_n$. Moreover, let $\Sigma_y(h)$ be the forecast error covariance matrix (see, e.g., Lütkepohl (1991) for precise expressions). Then the normalized forecast MSE in Tables 2 and 3 is

$$\frac{1}{m}\sum_{n=1}^{1000}(y_{T+h,n} - \hat{y}_T(h)_n)'\Sigma_y(h)^{-1}(y_{T+h,n} - \hat{y}_T(h)_n)/1000, \tag{4.2}$$

where $y_{T+h,n}$ is the generated value of the n-th time series corresponding to the forecast. The normalized MSE relates the forecast MSE of the estimated model based on the subset

strategies to the forecast MSE of the true model and should ideally be 1 because we divide by the dimension of the process m. Note that impulse responses are forecasts conditional on a specific assumed history of the process. Therefore the forecasting performance of the models is also an important characteristic if impulse response analysis is the objective of the analysis.

For $T = 30$ we show the 1- and 5-step MSEs in the last two columns of Table 2. The differences in forecast performance between the models chosen from alternative subset strategies are small. The corresponding 1- and 5-step MSE for the unrestricted model turn out to be 1.307 and 1.352, respectively. Thus, to improve the forecast precision compared to the unrestricted VAR, we can use any of the proposed strategies in conjunction with either AIC or HQ. The 5-step forecast MSE is largest in AIC models, followed by HQ and SC models. In contrast, the 1-step MSE is largest in models chosen by the SC criterion. This effect possibly results from incorrect restrictions set by the SC criterion when the sample size T is small.

In addition to what has been said, we are interested in the number of models with correctly identified restrictions, i.e. the number of models that have the same zero restrictions as the coefficient matrices of the true DGP. We call these models *'fully correct'* models. Moreover, we want to know in how many cases the strategies find a specification without incorrect zero restrictions. Therefore, we counted all models where at least the nonzero coefficients from the true model are unrestricted. Such a model is classified as *'not overly restricted'*. This implies, of course, that 'not overly restricted' models may include specifications where the true coefficient is zero but is not restricted by the selection procedure.

The first part of Table 4 shows the frequencies of 'fully correct' and 'not overly restricted' models for $T = 30$. For all strategies the number of data sets for which all zero coefficients are found and all restrictions are correct ('fully correct' models) is disappointingly small. When the SC criterion is used, the number of 'fully correct' models is zero and only 1 or 2 out of 1000 models have no incorrect zero restrictions. At best, when AIC is used in combination with TD the number of 'fully correct' models is 13, while nearly 86% of all chosen models show incorrect restrictions. Clearly, none of the strategies is very successful. Overall the Top-Down and SER/TP procedures with AIC and HQ do the best job. Somewhat surprisingly, the very computer intensive full search procedure does not do better and usually performs worse than these procedures.

We repeated the Monte Carlo study for $T = 100$ and list the results in Table 3. Naturally, the relative frequency of correct decisions increases with the sample size for all coefficients. However, the basic pattern is the same as observed before: In all strategy/criterion combinations, we find the lowest frequency for the small nonzero coefficient $a_{12,1}$. According to the normalized MSEs there is not much of a difference in forecasting accuracy between models chosen from different strategies. All strategies find models with MSEs close to one.[2]

The second part of Table 4 provides information on the number of 'fully correct' and 'not overly restricted' subset models for $T = 100$. As expected all subset strategies find the correct model more often than with $T = 30$. The Top-Down/HQ combination, for example, results in 111 'fully correct' models. Again, the Top-Down strategy seems to be the most successful. Also note the increasing number of 'not overly restricted' models: At best, around 37% of the models have no incorrect restrictions. In the majority of cases, however, the statistical procedures end up with incorrectly restricted models.

Overall our Monte Carlo experiments show that none of three subset procedures is very successful in finding the true underlying restricted model in samples of the size typically available in macroeconometric studies. The overall performance strongly depends on the structure of the underlying DGP. It is particularly interesting that the computationally efficient Top-Down and SER/TP strategies perform as well or even better than a Full Search procedure. Despite the fact that the true model is identified poorly, forecast precision tends to increase in the subset models relative to full models. Generally the subset strategies perform better in conjunction with the less parsimonious criteria HQ and AIC than with SC.

A test procedure that removes variables with lowest t-ratios until all remaining variables have t-ratios greater than 2 is often used in applied work. Note that such a strategy would be more parsimonious than a SER/TP procedure based on AIC or HQ. Even with these criteria the SER/TP procedure often fails to discover all zero restrictions correctly. In general, subset strategies may end up with misspecified models but still forecast well if the sample information is reasonably rich.

[2]MSE values smaller than one are due to the Monte Carlo variability.

4.2 Systems Approaches

to be completed

5 Empirical Application

The results from Section 4 suggest that the subset strategies presented are not very successful in identifying the true underlying model. Nevertheless, similar strategies are frequently used in applied work. In this section we illustrate how the use of subset strategies can in fact improve results of the final modeling objective, e.g. impulse responses or forecasts. To be more precise, we investigate the effects of a monetary policy shock in the U.S. measured as a shock to the federal funds rate. In the literature, unrestricted VAR models have been extensively used to analyze the effects of a monetary policy shock (see for example Christiano, Eichenbaum & Evans (1996), henceforth CEE). We compare impulse responses from an unrestricted VAR with results from a subset VAR as specified from the TP/SER procedure. Since impulse responses are special forecasts, as mentioned earlier, the results of the previous section indicate that our subset strategy may be useful for specifying suitable models for this kind of analysis as well.

As in the previous literature we identify monetary policy shocks as disturbances of a central bank's reaction function. To be more precise, the monetary authority is assumed to set its policy instrument according to a linear feedback rule that can be written in VAR(p) form as in (2.1), where the vector of variables y_t includes the monetary policy instrument and variables the central bank is looking at when setting the policy instrument. We distinguish between policy and non-policy variables. Policy variables are variables that are influenced immediately by central bank actions such as the federal funds rate, reserves or monetary aggregates. Non-policy variables that can only be influenced with a lag by the monetary authority, include the real GDP and a measures of the price level among others. With this setup, orthogonalized responses (see Lütkepohl (1991) for precise formulas) to an impulse in the monetary policy instrument show the effects of a monetary policy shock.

Following CEE we include the following variables in our analysis for the U.S.:

$$y_t = (gdp_t, p_t, pcom_t, ff_t, nbrd_t, tr_t, m1_t)' \tag{5.1}$$

where *gdp* is the log of real GDP, p the log of the GDP deflator, *pcom* the log of a commodity price index, ff the federal funds rate, *nbrd* the negative log of nonborrowed reserves, tr the log of total reserves and $m1$ the log of M1. Shocks to ff are regarded as a measure for monetary policy shocks. We use quarterly seasonally adjusted U.S. data over the period 1960q1–1992q4 ($T = 132$). The time series are depicted in Figure 1. Clearly, all time series used show trending behavior and given the choice of variables, cointegration between these variables might be possible. Since we are also interested in a comparison of our results to the ones of CEE, we ignore possible cointegration and proceed by estimating the model in levels representation.

We start out with an unrestricted VAR(4) model, i.e. we initially include four lags as in CEE. In this model, we estimate 203 parameters including the intercept terms. Since we wish to compare this unrestricted model to the restricted counterpart, we apply the SER/TP procedure presented in Section 3. We use the AIC criterion, which performed relatively well in the simulations. The resulting model has 115 zero restrictions leaving 88 parameters to be estimated. This subset model is then estimated with feasible GLS. We compute orthogonalized responses to a shock in the federal funds rate from both, the unrestricted and restricted system. In addition to the point estimates we compute confidence intervals to account for the fact that impulse responses are nonlinear functions of estimated coefficients and hence, estimates. We use a bootstrap procedure proposed by Hall (1992). For details on this bootstrap method see Benkwitz, Lütkepohl & Wolters (2000). They argue that Hall's percentile intervals are advantageous compared to the standard bootstrap intervals.

Figures 2 and 3 show the responses for all system variables to a monetary shock. The solid lines represent the point estimates, while the dashed lines are approximate 95% confidence bands computed from a bootstrap with 1000 draws. The left column shows responses from the unrestricted VAR(4) model, the right column is based on the subset model.

The results from our impulse response analysis are largely in line with the results of CEE. While CEE compute confidence bands that approximately correspond to the 90% confidence level, we draw our conclusions based on results from the subset VAR (restricted model) and on a 95% level. A positive impulse or shock in the federal funds rate corresponds to a contractionary monetary policy shock. This shock is associated with a persistent fall of real GDP and a delayed decline in the GDP price deflator. In contrast to CEE, we only find a

small delayed decline in commodity prices.[3] Moreover, we find that the shock in ff leads to a rise in the federal funds rate and a decline of nonborrowed and total reserves. In addition, the contractionary policy shock leads to a persistent decline of nominal money M1. In our example, the same conclusions can be drawn from the unrestricted model, however, using 90% level confidence intervals for the impulse responses.

The comparison of the unrestricted and the restricted model shows two interesting results: First, even though the restricted model includes a substantial number of zero coefficients, the pattern of estimated impulse responses remains basically unchanged. Overall, this indicates that restrictions from the SER/TP procedure with the AIC criterion seem to be reasonable.[4] There is no indication of any bias induced by these restrictions. Thus, the potential inference problems reported by Benkwitz, Lütkepohl & Neumann (2000) may not be present here.

Second and perhaps more important, the subset specification yields confidence bands that are substantially narrower than in the unrestricted model. Provided the restrictions are correct, impulse responses can be estimated more precisely and hence allow for easier interpretation.

As mentioned before, impulse responses are conditional forecasts. Therefore, it is especially interesting to evaluate the forecast performance of the specified subset model. We do so with the simulation technique presented in Section 4. To begin with, we assume that the subset specification used above is the true underlying model, i.e. we use the restricted EGLS estimates for the parameter matrices A_i and the covariance matrix Σ_u. We use this DGP and real observations as presample values to generate 1000 time series with length $T = 137$. Then, we apply SER/TP and the Top-Down procedure to specify subset VARs.[5] Table 5 shows normalized mean squared errors computed according to (4.2) from both strategies and from the unrestricted model. When using AIC and HQ models, both strategies perform better in forecasting than the unrestricted model. In contrast to simulation results from the previous section, MSEs from the SER/TP procedure are now slightly smaller than from Top-Down. However, differences between SER/TP and TD are only small. Given this result, it is not surprising that impulse responses from both strategies are very similar. Obviously,

[3]CEE find a sharp, immediate decline in commodity prices. The different pattern may be due to a different measure of commodity prices used in our study.

[4]We have also computed impulse responses from a subset model specified with the Top-Down procedure. These impulse responses show only minor differences to the ones in the right columns of Figures 2 and 3.

[5]The Full Search strategy is clearly infeasible in a large model. With 7 variables and 4 lags FS would have to compare 2^{28} models in each equation.

the SC criterion leads to models with too many incorrect zero restrictions and consequently to suboptimal forecasting properties.

The results from our empirical example can be summarized as follows: Both SER/TP and Top-Down lead to very similar subset VAR models. Patterns of impulse responses are nearly identical to those of the unrestricted VAR. The corresponding bootstrap confidence intervals are narrower indicating that responses are estimated more precisely. This reduces the uncertainty when interpreting the results, which are in line with results from CEE. For the present system with 7 variables the comparison of normalized MSEs shows that SER/TP has a slight advantage compared to the Top-Down strategy. We also conclude that using the SC criterion leads to models with many incorrect zero restrictions that spoil forecasts and possibly impulse responses. Thus, for impulse response analysis AIC and HQ are the better choice, at least for the present example model.

6 Conclusions

The present study considers alternative lag selection strategies within the VAR modeling framework. We present four different model selection procedures: Full Search (FS), Sequential Elimination of Regressors (SER), a Testing Procedure (TP), and a Top-Down (TD) procedure. We show that using the Test Procedure with threshold values as a function of the elimination step is equivalent to the SER strategy.

We compare the small sample properties of single equation strategies with Monte Carlo experiments. One finding for a small bivariate DGP is that none of the strategies specifies the correct model with high probability. It is particularly interesting that the computationally demanding Full Search procedure offers no advantage. The overall performance of the presented strategies strongly depends on the underlying DGP. We find that our results are sensitive to the absolute size of the DGP parameters. Therefore, it seems risky to generalize our results. A comparison of the forecast precision shows that in many situations the subset VAR models perform better than the corresponding unrestricted VAR model. From a forecasting point of view, using either subset strategy in combination with AIC or HQ is advantageous relative to full VAR modeling.

Although our results indicate that the presented procedures often fail to discover the true model, a testing procedure similar to SER/TP is frequently used in applied work. Finding

the zero coefficients is not necessarily the final modeling objective. If the researcher is interested in forecasts or impulse response analysis, the presented subset modeling strategies may help to improve the results. In our empirical example, a subset VAR identified from SER/TP results in impulse response patterns that are very similar to the ones from the full VAR. The confidence bands from the subset VAR are narrower, however, indicating that responses are estimated more precisely.

We conclude that subset strategies can be useful for forecasting purposes and impulse response analysis. Since they do not find all zero restrictions with high probability, we recommend to use subset strategies as additional modeling tools only. To avoid misspecification a comparison to the full VAR and the application of diagnostic tests is also advisable.

References

Akaike, H. (1974), A New Look at the Statistical Model Identification, *IEEE Transactions on Automatic Control*, AC-19, 716–723.

Benkwitz, A., H. Lütkepohl & M. H. Neumann (2000), Problems related to confidence intervals for impulse responses of autoregressive processes, *Econometric Reviews*, forthcoming.

Benkwitz, A., Lütkepohl, H. & Wolters, J. (2000), Comparison of Bootstrap Confidence Intervals for Impulse Responses of German Monetary Systems, *Macroeconomic Dynamics*, forthcoming.

Christiano, L. J., Eichenbaum, M. & Evans, C. L. (1996), The Effects of Monetary Policy Shocks: Evidence From the Flow of Funds, *Review of Economics and Statistics* **78** (1), 16–34.

Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*, New York: Springer.

Hannan, E.J. & B.G. Quinn (1979), The Determination of the Order of an Autoregression, *Journal of the Royal Statistical Society*, B41, 190-195.

Judge, G.G. et al. (1988), *Introduction to the Theory and Practice of Econometrics*, New York: John Wiley.

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.

Sims, C. A. (1980). Macroeconomics and reality, *Econometrica* **48**, 1–48.

Table 1. Threshold Values Corresponding to Model Selection Criteria

K	T	Criterion	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9	γ_{10}
12	50	AIC	1.25	1.26	1.28	1.29	1.31	1.32	1.34	1.36	1.37	1.38
		HQ	1.46	1.48	1.50	1.52	1.53	1.55	1.57	1.59	1.61	1.62
		SC	2.54	2.57	2.60	2.64	2.67	2.70	2.73	2.76	2.79	2.82
12	100	AIC	1.33	1.34	1.35	1.36	1.36	1.37	1.38	1.39	1.39	1.40
		HQ	1.65	1.66	1.67	1.68	1.69	1.70	1.71	1.72	1.73	1.73
		SC	2.91	2.93	2.95	2.96	2.98	3.00	3.01	3.03	3.04	3.06
12	200	AIC	1.37	1.38	1.38	1.39	1.39	1.39	1.40	1.40	1.40	1.41
		HQ	1.78	1.78	1.79	1.79	1.80	1.80	1.81	1.81	1.82	1.82
		SC	3.20	3.21	3.22	3.22	3.23	3.24	3.25	3.26	3.27	3.27
20	50	AIC	1.11	1.12	1.14	1.16	1.18	1.20	1.21	1.23	1.25	1.26
		HQ	1.30	1.32	1.34	1.36	1.38	1.40	1.42	1.44	1.46	1.48
		SC	2.25	2.29	2.33	2.36	2.40	2.43	2.47	2.50	2.54	2.57
20	100	AIC	1.27	1.28	1.29	1.29	1.30	1.31	1.32	1.33	1.33	1.34
		HQ	1.58	1.58	1.59	1.60	1.61	1.62	1.63	1.64	1.65	1.66
		SC	2.78	2.80	2.81	2.83	2.85	2.86	2.88	2.90	2.91	2.93
20	200	AIC	1.35	1.35	1.35	1.36	1.36	1.36	1.37	1.37	1.37	1.38
		HQ	1.74	1.74	1.75	1.75	1.76	1.76	1.77	1.77	1.78	1.78
		SC	3.13	3.14	3.15	3.16	3.16	3.17	3.18	3.19	3.20	3.21

Table 2. Relative frequency of correct decisions obtained from 1000 realizations of length $T = 30$ of the VAR(3) process (4.1)

subset strategy	model selection criterion	relative frequency of correct decisions A_1	A_2	A_3	normalized forecast MSE 1-step	5-step
Full Search (FS)	AIC	$\begin{bmatrix} .772 & .231 \\ .929 & .828 \end{bmatrix}$	$\begin{bmatrix} .809 & .797 \\ .701 & .768 \end{bmatrix}$	$\begin{bmatrix} .802 & .789 \\ .778 & .752 \end{bmatrix}$	1.293	1.315
	HQ	$\begin{bmatrix} .736 & .187 \\ .901 & .808 \end{bmatrix}$	$\begin{bmatrix} .858 & .839 \\ .646 & .824 \end{bmatrix}$	$\begin{bmatrix} .853 & .833 \\ .813 & .803 \end{bmatrix}$	1.299	1.311
	SC	$\begin{bmatrix} .443 & .035 \\ .703 & .723 \end{bmatrix}$	$\begin{bmatrix} .987 & .986 \\ .380 & .956 \end{bmatrix}$	$\begin{bmatrix} .989 & .973 \\ .940 & .965 \end{bmatrix}$	1.342	1.239
SER/TP	AIC	$\begin{bmatrix} .787 & .299 \\ .936 & .848 \end{bmatrix}$	$\begin{bmatrix} .770 & .735 \\ .698 & .741 \end{bmatrix}$	$\begin{bmatrix} .763 & .754 \\ .727 & .739 \end{bmatrix}$	1.296	1.327
	HQ	$\begin{bmatrix} .748 & .253 \\ .912 & .819 \end{bmatrix}$	$\begin{bmatrix} .822 & .787 \\ .650 & .791 \end{bmatrix}$	$\begin{bmatrix} .815 & .801 \\ .777 & .789 \end{bmatrix}$	1.306	1.320
	SC	$\begin{bmatrix} .440 & .043 \\ .717 & .694 \end{bmatrix}$	$\begin{bmatrix} .979 & .972 \\ .402 & .940 \end{bmatrix}$	$\begin{bmatrix} .979 & .965 \\ .919 & .946 \end{bmatrix}$	1.355	1.234
top-down	AIC	$\begin{bmatrix} .800 & .336 \\ .940 & .876 \end{bmatrix}$	$\begin{bmatrix} .777 & .757 \\ .702 & .745 \end{bmatrix}$	$\begin{bmatrix} .785 & .788 \\ .757 & .769 \end{bmatrix}$	1.292	1.313
	HQ	$\begin{bmatrix} .769 & .278 \\ .916 & .864 \end{bmatrix}$	$\begin{bmatrix} .831 & .795 \\ .644 & .803 \end{bmatrix}$	$\begin{bmatrix} .836 & .832 \\ .804 & .821 \end{bmatrix}$	1.281	1.296
	SC	$\begin{bmatrix} .471 & .031 \\ .761 & .849 \end{bmatrix}$	$\begin{bmatrix} .979 & .968 \\ .250 & .974 \end{bmatrix}$	$\begin{bmatrix} .980 & .973 \\ .979 & .978 \end{bmatrix}$	1.318	1.235

Table 3. Relative frequency of correct decisions obtained from 1000 realizations of length $T = 100$ of the VAR(3) process (4.1)

subset strategy	model selection criterion	relative frequency of correct decisions A_1	A_2	A_3	normalized forecast MSE 1-step	5-step
Full Search (FS)	AIC	$\begin{bmatrix} .999 & .273 \\ 1.00 & 1.00 \end{bmatrix}$	$\begin{bmatrix} .826 & .815 \\ .940 & .798 \end{bmatrix}$	$\begin{bmatrix} .826 & .831 \\ .892 & .808 \end{bmatrix}$	1.006	1.008
	HQ	$\begin{bmatrix} .998 & .169 \\ 1.00 & .999 \end{bmatrix}$	$\begin{bmatrix} .902 & .906 \\ .896 & .884 \end{bmatrix}$	$\begin{bmatrix} .909 & .924 \\ .953 & .912 \end{bmatrix}$	.996	1.008
	SC	$\begin{bmatrix} .978 & .009 \\ .995 & .995 \end{bmatrix}$	$\begin{bmatrix} .996 & .995 \\ .596 & .991 \end{bmatrix}$	$\begin{bmatrix} .998 & .998 \\ .996 & 1.00 \end{bmatrix}$	.995	1.006
SER/TP	AIC	$\begin{bmatrix} .999 & .320 \\ 1.00 & .999 \end{bmatrix}$	$\begin{bmatrix} .800 & .752 \\ .942 & .799 \end{bmatrix}$	$\begin{bmatrix} .805 & .797 \\ .813 & .811 \end{bmatrix}$	1.010	1.008
	HQ	$\begin{bmatrix} .998 & .218 \\ 1.00 & .999 \end{bmatrix}$	$\begin{bmatrix} .884 & .864 \\ .897 & .875 \end{bmatrix}$	$\begin{bmatrix} .893 & .894 \\ .911 & .898 \end{bmatrix}$	1.001	1.007
	SC	$\begin{bmatrix} .978 & .013 \\ .995 & .995 \end{bmatrix}$	$\begin{bmatrix} .993 & .993 \\ .570 & .991 \end{bmatrix}$	$\begin{bmatrix} .998 & .996 \\ .993 & .999 \end{bmatrix}$	.996	1.006
top-down	AIC	$\begin{bmatrix} .999 & .395 \\ 1.00 & 1.00 \end{bmatrix}$	$\begin{bmatrix} .816 & .787 \\ .948 & .782 \end{bmatrix}$	$\begin{bmatrix} .835 & .834 \\ .821 & .831 \end{bmatrix}$	1.002	1.008
	HQ	$\begin{bmatrix} .999 & .278 \\ 1.00 & 1.00 \end{bmatrix}$	$\begin{bmatrix} .897 & .890 \\ .904 & .874 \end{bmatrix}$	$\begin{bmatrix} .922 & .914 \\ .913 & .908 \end{bmatrix}$	.997	1.004
	SC	$\begin{bmatrix} .981 & .015 \\ .995 & .996 \end{bmatrix}$	$\begin{bmatrix} .997 & .997 \\ .570 & .994 \end{bmatrix}$	$\begin{bmatrix} .999 & .998 \\ .994 & .999 \end{bmatrix}$	.993	1.007

Table 4. Frequency of 'fully correct' and 'not overly restricted' models obtained from 1000 realizations of DGP (4.1)

Selection Procedure	T	Criterion	fully correct models	not overly restricted
Full Search (FS)	30	AIC	11	75
		HQ	8	48
		SC	0	1
SER/TP	30	AIC	10	111
		HQ	8	72
		SC	0	2
top-down	30	AIC	13	142
		HQ	11	94
		SC	0	1
Full Search (FS)	100	AIC	72	255
		HQ	80	150
		SC	5	5
SER/TP	100	AIC	59	297
		HQ	77	194
		SC	5	9
top-down	100	AIC	87	373
		HQ	111	249
		SC	5	9

Table 5. Normalized MSEs from 7-dimensional VAR

Selection Procedure	Criterion	normalized forecast MSE 1-step	5-step
SER/TP	AIC	1.35	1.54
	HQ	1.34	1.54
	SC	1.43	1.79
top-down	AIC	1.35	1.55
	HQ	1.36	1.56
	SC	1.50	2.07
unrestricted	–	1.37	1.56

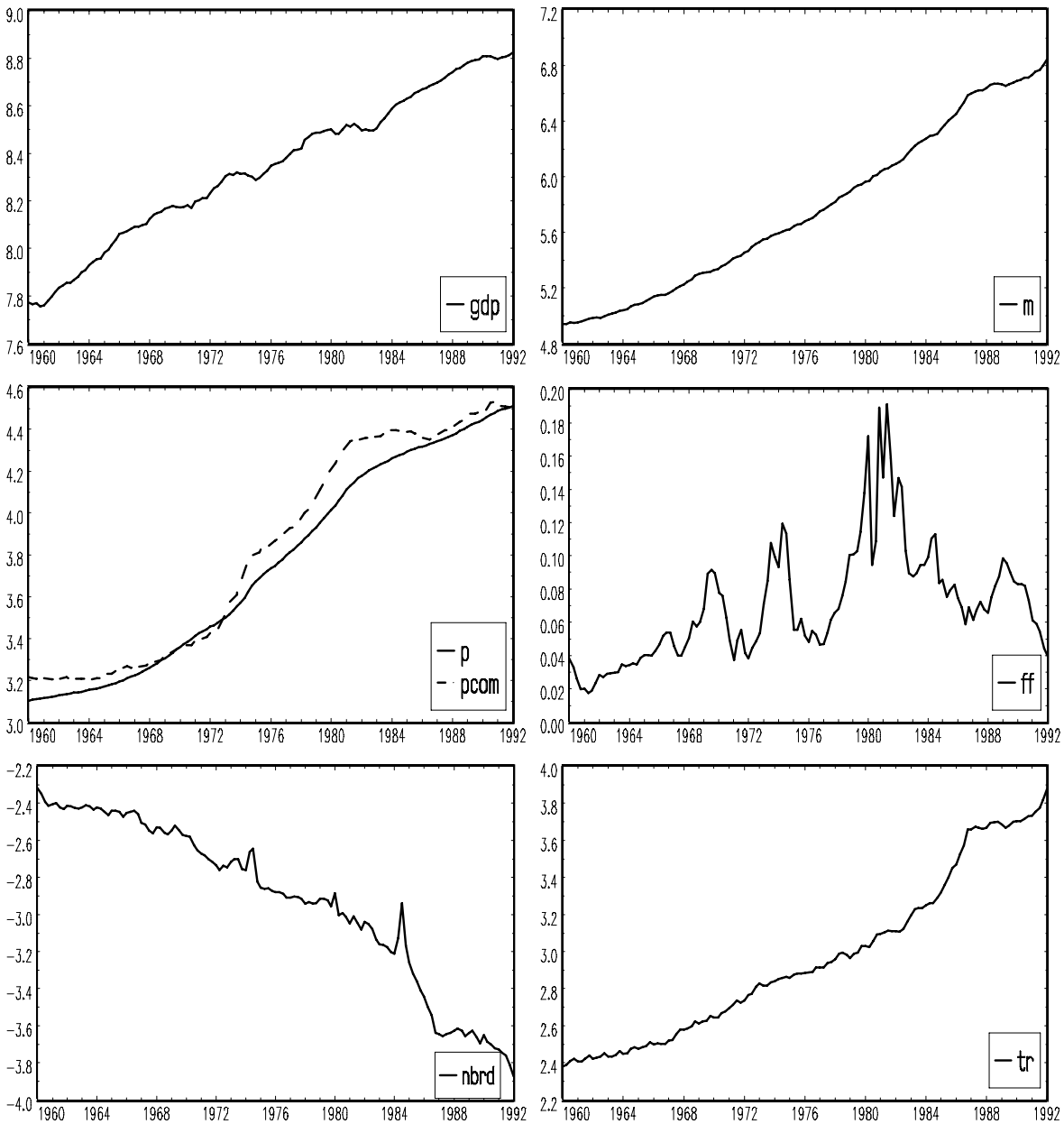

Figure 1. Time series analyzed

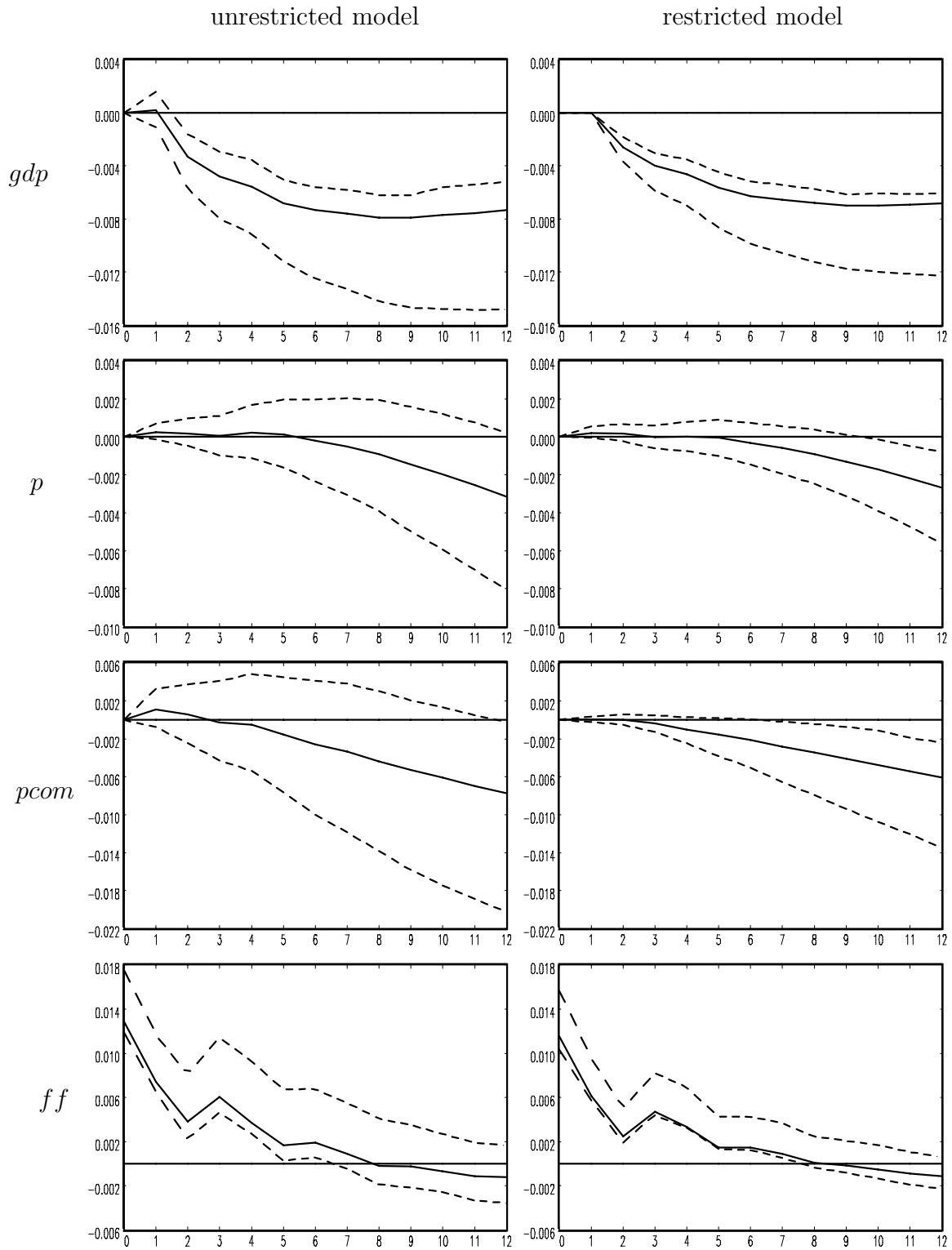

Figure 2. Responses of *gdp*, *p*, *pcom* and *ff* to a unit shock in *ff* computed from unrestricted (left) and restricted model (right)

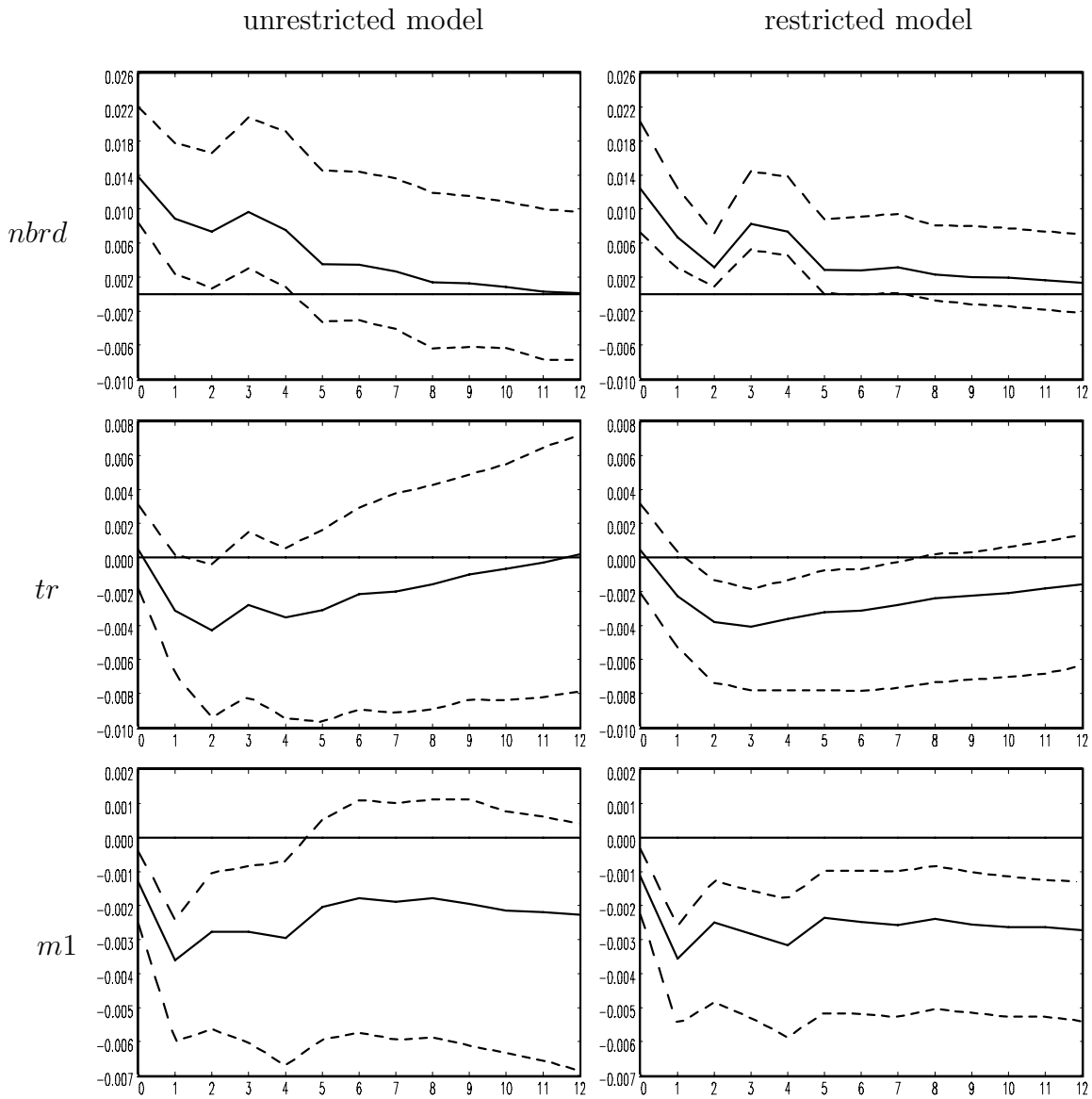

Figure 3. Responses of *nbrd*, *tr* and $m1$ to impulse in ff computed from unrestricted (left) and restricted model (right)