May 12, 2000

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

Alternative modeling strategies for specifying subset VAR models are considered. It is shown that under certain conditions a testing procedure based on t-ratios is equivalent to sequentially eliminating lags that lead to the largest improvement in a prespecified model selection criterion. A Monte Carlo study is used to illustrate the properties of different procedures. It is found that the differences between alternative strategies are small. In small samples, the strategies often fail to discover the true model. Nevertheless, using subset strategies results in models with improved forecast precision. To illustrate how these subset strategies can improve results from impulse response analysis, a VAR model is used to analyze the effects 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: Lag selection, model selection, monetary policy shocks, subset models, vector autoregressions

JEL classification: C32, E52

[1]Financial support by the Deutsche Forschungsgemeinschaft, SFB 373, and the European Commission under the Training and Mobility of Researchers Programme (contract No. ERBFMRXCT980213) is gratefully acknowledged. We thank Hannah Nielsen for useful comments.

1 Introduction

As a consequence of Sims' (1980) critique of classical econometric modeling, empirical macroeconomic studies are often based on vector autoregressive (VAR) models. In these models the relations between 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 the variables. Using a statistical procedure for deciding on possible constraints may be advantageous compared to a procedure which is based on assumed a priori knowledge 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 (see e.g. Lütkepohl (1991, Ch. 5) for a survey of some procedures). Typically applied researchers use testing procedures or model selection criteria in placing restrictions on a given VAR model. We will compare both types of procedures in the following.

A widely used testing procedure eliminates the variables with lowest t-ratios sequentially until all remaining coefficients 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. These lag selection procedures are discussed and compared within both a single equation and a systems framework. Their small sample properties are explored with a Monte Carlo study and their virtue is illustrated by applying

them to a U.S. monetary system which has been analyzed previously on the basis of an unrestricted VAR model.

A number of other subset methods have been proposed in the literature (see, e.g., Lütkepohl (1991, Ch. 5)). Some of these methods are not suitable for our purposes. For example, the methods proposed by Hsiao (1979, 1982) and Penm & Terrell (1984) are not designed for finding all zero entries in the VAR coefficient matrices but aim at detecting special zero coefficients only. Some of the methods presented in Lütkepohl (1991) may be viewed as direct competitors to the methods considered here, however. Since they are not very popular in practice, they will not be treated in detail in the following.

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 single equation 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. Furthermore, in Sec. 5, U.S. macroeconomic data are used to illustrate the usefulness of applying variable selection procedures and the effects of restrictions on impulse response analysis. Conclusions follow in Sec. 6.

2 Vector Autoregressive Models

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 fixed $(M \times 1)$ 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. 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. The process may have unit roots so that integrated and cointegrated variables are not excluded.

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,

$$A_0 y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t. \tag{2.2}$$

In the next section, strategies will be discussed for imposing zero restrictions on a model such as (2.1) or (2.2).

3 Lag Order Selection Strategies

Because the dimension of the parameter space tends to be large for VAR models if the number of variables, M, is moderate or large, it is desirable to impose restrictions to improve the precision of statistical inference. Several statistical procedures are in common use for this purpose (e.g. Lütkepohl (1991, Ch. 4)). For example, procedures for specifying a suitable lag order p exist. Moreover, a number of proposals have been made for eliminating individual lags of variables or complete variables from the equations of a system. These proposals are typically based on statistical tests or model selection criteria. Furthermore, they either consider the full system at once or they are based on individual equations. In this section some procedures will be reviewed and compared on the basis of their theoretical properties. We begin with methods which consider the equations individually.

3.1 Single Equation Approaches

Consider the m-th equation of the system (2.1) or (2.2) of the form

$$y_{mt} = \nu_m + x_{1t}\theta_1 + \cdots + x_{Kt}\theta_K + u_{mt}, \quad t = 1, \ldots, T. \tag{3.1}$$

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 variable elimination strategies for the single equation (3.1). The first procedure compares all possible sets of zero restrictions for the x_{kt} $(k = 1, \ldots, K)$ and chooses the model which is optimal according to a given model selection criterion.

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

Although less demanding than a full search, the SER procedure still requires a considerable computational effort because, in each step, as many models have to be compared as there are regressors. In contrast, the following elimination strategy requires one ordinary least squares (OLS) regression in each step only. It deletes the regressor with the smallest t-ratio in each step.

Testing Procedure (TP)
Estimate the model by OLS and delete sequentially those regressors with the smallest absolute values of the corresponding 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 from an OLS estimation associated with θ_k in the j-th 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 the last 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 t- or standard normal distributions. Thus, some or all of the variables may be integrated and the errors do not have to be white noise. 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 anyone 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 estimate the equation by OLS once and 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 j-th step is performed. We show that both strategies eliminate the same regressor in that step. The squared t-ratio of the k-th regressor is

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

(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 one 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

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

Thereby the proposition is established. □

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 between $T = 50$ and $T = 200$ is not unusual in macroeconomic studies with quarterly data we give some relevant γ_j obtained from Proposition 1 for these three model selection criteria in Table 1. Obviously, the γ_j are in a range starting roughly at the .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 a 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 if a procedure with a typical type I error is desired.

All the lag selection procedures considered so far 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 (2.2). In the next subsection procedures will be considered which are based on the full system.

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	1.76	1.78	1.80	1.83	1.85	1.87	1.89	1.91	1.93	1.96
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	2.04	2.05	2.06	2.07	2.08	2.09	2.10	2.12	2.13	2.14
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	2.25	2.25	2.26	2.26	2.27	2.28	2.28	2.29	2.29	2.30
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	1.56	1.59	1.61	1.64	1.66	1.69	1.71	1.74	1.76	1.78
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	1.94	1.95	1.97	1.98	1.99	2.00	2.01	2.02	2.04	2.05
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	2.20	2.20	2.21	2.22	2.22	2.23	2.23	2.24	2.25	2.25

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 usually 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 $\mathbf{x}_{mt}$ is the $(J_m \times 1)$ vector of regressors in the m-th 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 $J_m = Mp+1$ $(m = 1, \ldots, M)$. 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 each equation. In that case, estimation may be 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 the 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 a $((M^2p + M) \times 1)$ 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 maximum likelihood (ML) estimator is used for $\tilde{\Sigma}_u(\mathbf{j})$. However, for restricted systems computing the exact Gaussian 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 the FS and SER selection strategies considered in the single equation context. In the systems context they will be abbreviated as SFS and SSER.

Note that the simple relation between the testing procedure and the sequential elimination of regressors 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 likelihood ratio (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, correlation and integration 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 the following simple bivariate VAR(2) example process:

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

with covariance matrix

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

This process was also used in simulations by Lütkepohl (1991). We simulated 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 a VAR(4) and then applied the subset strategies. In this case the coefficient matrices A_3 and A_4 of the true data generating process (DGP) contain zeros only. For each realization of the DGP we recorded whether the strategies decided correctly on the zero restrictions of individual coefficients and list the resulting relative frequencies of correct decisions.

Obviously, the example process is relatively small compared to VAR models that are found in applied studies. It is still useful to consider such a simple DGP here because the results for this process can be presented in a lucid way. To obtain a realistic ratio of the number of parameters and the sample size we present results for a smaller than typical sample size. More precisely, we show 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 is set to zero fairly often. Generally there is not much to choose between the different strategies and the very computer intensive full search strategy, FS, does not perform better than the more efficient SER/TP strategy.

Table 2: Relative Frequency of Correct Decisions Obtained from 1000 Realizations of Length $T = 30$ of VAR(4) Process

Selection Procedure	Criterion	Relative Frequency of Correct Decisions A_1	A_2	A_3	A_4
Full Search (FS)	AIC	$\begin{bmatrix} .756 & .240 \\ .925 & .790 \end{bmatrix}$	$\begin{bmatrix} .788 & .783 \\ .709 & .754 \end{bmatrix}$	$\begin{bmatrix} .795 & .773 \\ .741 & .763 \end{bmatrix}$	$\begin{bmatrix} .783 & .772 \\ .738 & .731 \end{bmatrix}$
	HQ	$\begin{bmatrix} .721 & .199 \\ .897 & .762 \end{bmatrix}$	$\begin{bmatrix} .847 & .832 \\ .665 & .809 \end{bmatrix}$	$\begin{bmatrix} .845 & .823 \\ .781 & .808 \end{bmatrix}$	$\begin{bmatrix} .836 & .828 \\ .795 & .788 \end{bmatrix}$
	SC	$\begin{bmatrix} .638 & .122 \\ .860 & .747 \end{bmatrix}$	$\begin{bmatrix} .910 & .899 \\ .573 & .869 \end{bmatrix}$	$\begin{bmatrix} .919 & .889 \\ .854 & .892 \end{bmatrix}$	$\begin{bmatrix} .900 & .896 \\ .878 & .870 \end{bmatrix}$
SER/TP	AIC	$\begin{bmatrix} .774 & .298 \\ .926 & .806 \end{bmatrix}$	$\begin{bmatrix} .748 & .731 \\ .718 & .721 \end{bmatrix}$	$\begin{bmatrix} .755 & .731 \\ .689 & .738 \end{bmatrix}$	$\begin{bmatrix} .741 & .740 \\ .707 & .694 \end{bmatrix}$
	HQ	$\begin{bmatrix} .736 & .259 \\ .905 & .781 \end{bmatrix}$	$\begin{bmatrix} .799 & .774 \\ .669 & .771 \end{bmatrix}$	$\begin{bmatrix} .796 & .777 \\ .737 & .792 \end{bmatrix}$	$\begin{bmatrix} .798 & .796 \\ .763 & .757 \end{bmatrix}$
	SC	$\begin{bmatrix} .652 & .171 \\ .868 & .743 \end{bmatrix}$	$\begin{bmatrix} .881 & .848 \\ .589 & .833 \end{bmatrix}$	$\begin{bmatrix} .880 & .852 \\ .816 & .852 \end{bmatrix}$	$\begin{bmatrix} .876 & .866 \\ .833 & .824 \end{bmatrix}$

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.

In Table 3 the numbers of models with correctly identified zero restrictions, i.e., the numbers of models that have the same zero restrictions as the coefficient matrices of the true DGP are given. We call models *'fully correct'* if all zero coefficients are found and none of the nonzero coefficients is incorrectly restricted. A model is classified as *'not overly restricted'* if no false zero restrictions are imposed. This implies, of course, that 'not overly restricted models' may include unrestricted coefficients which are actually zero in the true DGP.

Table 3: Frequency of 'fully correct' and 'not overly restricted models' Obtained from 1000 Realizations of VAR(4) DGP Based on (4.1)

		$T = 30$		$T = 100$	
Selection Procedure	Criterion	fully correct	not overly restricted	fully correct	not overly restricted
Full Search (FS)	AIC	4	76	46	258
	HQ	2	51	60	154
	SC	0	14	53	82
SER/TP	AIC	2	114	35	297
	HQ	3	76	53	197
	SC	0	26	48	99

Columns 3 and 4 of Table 3 show 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') is disappointingly small. This number ranges from 0 to 4 out of 1000 models. Clearly, none of the strategies is very successful. Somewhat surprisingly, the very computer intensive full search procedure does not perform better than the other procedures. In fact, FS specifies more models with false restrictions.

We repeated the Monte Carlo study for $T = 100$. The last two columns of Table 3 provide information on the number of 'fully correct' and 'not overly restricted' subset models for this case. Naturally, the relative frequency of correct decisions increases with the sample size. Nevertheless, the success rate is still disappointing. At best, around 30% of the models have no incorrect restrictions. In the majority of cases, however, the statistical procedures end up with incorrectly restricted models. 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. Such a strategy would be more parsimonious than a SER/TP procedure based on AIC or HQ and hence will impose more incorrect restrictions.

In general our Monte Carlo experiments show that none of the 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 SER/TP strategy performs as well as or even better than a full search procedure.

One may wonder whether some of the other subset methods proposed in the literature

Table 4: Forecasting Properties of Subset Models

Selection Procedure	Criterion	$T = 30$ normalized forecast MSE 1-step	$T = 30$ normalized forecast MSE 5-step	$T = 100$ normalized forecast MSE 1-step	$T = 100$ normalized forecast MSE 5-step
Full Search (FS)	AIC	1.375	1.387	1.021	1.019
	HQ	1.360	1.366	1.004	1.020
	SC	1.334	1.314	.993	1.012
SER/TP	AIC	1.397	1.403	1.026	1.018
	HQ	1.379	1.375	1.009	1.020
	SC	1.351	1.331	1.001	1.015
Full VAR(4)		1.422	1.455	1.050	1.028

have more favorable properties in this respect. Therefore it may be worth noting that the simulation results reported by Lütkepohl (1991, Ch. 5) for some other methods are not unlike the results in our Table 2. We have also included the top-down strategy of Lütkepohl (1991) in the present comparison and found that the overall performance is similar to that of the procedures considered here.

To get a more comprehensive 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 by $\hat{y}_T(h)_n$. Moreover, let $\Sigma_y(h)$ be the forecast error covariance matrix (see, e.g., Lütkepohl (1991) for precise expressions). Then the normalized forecast MSEs in Table 4 are computed as

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

where $y_{T+h,n}$ is the generated vector of the n-th bivariate time series for which the forecast is made. 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.

We show the 1- and 5-step MSEs in the third and fourth column of Table 4. The differences in forecast performance between the models chosen from alternative subset strategies are

small. For $T = 30$, the corresponding 1- and 5-step MSEs for the unrestricted model turn out to be 1.422 and 1.455, respectively. Our results indicate that we can use any of the proposed strategies in conjunction with any of the criteria to improve the forecast precision compared to the unrestricted VAR. In other words, despite the fact that the true model is rarely found, forecast precision tends to increase in the subset models relative to full models.

For $T = 100$, according to the normalized MSEs there is very little difference in forecasting accuracy between models chosen from different strategies. All strategies find models with MSEs close to one.[2] Moreover, all subset strategies end up with slightly smaller MSEs than the unrestricted models, despite the fact that incorrect zero restrictions are imposed quite often.

We have also applied the systems approaches for choosing the zero restrictions. In that case the general picture is similar to what is seen in Tables 2, 3, and 4. In other words, the computationally more demanding system strategies SFS and SSER do not outperform their single equation counterparts. This result was also obtained for a model with nondiagonal residual covariance matrix Σ_u. Overall, the results from our Monte Carlo experiments indicate no evidence in favor of system strategies. The simple STP with a fixed threshold often performs worse than the related single equation testing procedure where the threshold value γ is chosen as a function of the reduction step and the sample size T. SFS and SSER perform very much like their single equation counterparts. Since system procedures involve feasible GLS estimation in each reduction step they are, of course, computationally more demanding than the single equation strategies. Taking this fact into account, we recommend using the single equation strategies in applied work.

In applied time series analysis, researchers often face nonstationary and cointegrated time series. Despite the fact that cointegration methods have become standard to analyze such data, it is still popular to estimate VAR models of nonstationary series in levels form. Since the empirical example in the next section involves nonstationary and possibly cointegrated time series, we were interested in the performance of the lag selection strategies when the underlying process is integrated but estimated as a VAR in levels. Therefore, we also conducted a Monte Carlo experiment based on a cointegrated VAR process. The general conclusion from these simulations is that the overall performance of subset strategies when the time series are nonstationary and cointegrated is very similar to the one from a stable process. Therefore we do not present specific results here in order to save space.

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

In summary, our simulations based on stationary and cointegrated processes indicate that a single equation strategy based on eliminating regressors with smallest t-ratios is the preferred procedure. The critical values or, equivalently, the choice of model selection criterion will have an impact on the final model. Forecasts and, hence, impulse responses will not be very sensitive to this choice if one of the usual values or criteria is used.

5 Empirical Application

The results from Section 4 suggest that our subset strategies 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 response analysis or forecasting. 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 for analyzing 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. We use single equation strategies because they are considerably faster and perform as well as systems strategies.

As in the previous literature we identify monetary policy shocks as disturbances of a central bank's reaction function. More precisely, 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 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 measure of the price level among others. Within this setup, orthogonalized responses (see Lütkepohl (1991) for precise formulas) are used to investigate the effects of an impulse in the monetary policy instrument.

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 ($T = 132$) as in CEE[3]. 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 is possible. Since we are also interested in a comparison of our results to those of CEE, we ignore possible cointegration and proceed by estimating the model in levels.

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 Hall's (1992) percentile bootstrap confidence intervals for the impulse responses, as recommended by Benkwitz, Lütkepohl & Wolters (2000). They argue that these 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

[3]Data sources: The original series have been obtained from the Federal Reserve Economic Data (FRED) database.

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.[4] 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 using 90% level confidence intervals for the impulse responses as in CEE.

The comparison of the unrestricted and the restricted models 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. 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 smaller than in the unrestricted model. Provided the restrictions are correct, impulse responses can be estimated more precisely. This reduces the uncertainty when interpreting the results

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 using the simulation technique presented in Section 4. To begin with, we assume that the subset specification used above is the true underlying DGP, 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 the SER/TP procedure to specify subset VARs.[5] Table 5 shows normalized MSEs computed according to (4.2) from both strategies and from the unrestricted model. When using AIC and HQ models, the subset strategy performs better in forecasting than the unrestricted model. However, the differences between SER/TP and the full model MSEs are only small. Given this result, it is not surprising that impulse responses from both models are very similar. In addition, the SC criterion sometimes leads to models with too many incorrect zero restrictions and consequently to suboptimal 5-step forecasting properties.

[4]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.

[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} regressions in each equation.

Table 5: Normalized MSEs for 7-dimensional VAR

Model	Criterion	normalized forecast MSE 1-step	5-step
Subset	AIC	1.35	1.54
SER/TP	HQ	1.34	1.54
	SC	1.36	1.58
unrestricted	–	1.37	1.56

We conclude that using the SC criterion can lead to models with many incorrect zero restrictions that spoil forecasts and possibly impulse responses. Thus, for impulse response analysis, AIC and HQ are likely to be 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 three different single equation model selection procedures: Full search (FS), sequential elimination of regressors (SER) and a testing procedure (TP) based on t-ratios from OLS regressions. It is shown that using the test procedure with critical values as a function of the elimination step is equivalent to the SER strategy. Furthermore, we propose systems strategies related to the single equation strategies above.

The small sample properties of single equation and systems strategies are compared with Monte Carlo experiments. It is found that single equation and systems strategies produce very similar results. Given that system strategies are computationally more demanding, we recommend using the single equation procedures.

Our results indicate 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 strategies presented strongly depends on the underlying DGP. We find that the results are sensitive to the absolute size of the DGP parameters. Therefore, it seems risky to generalize our findings. A comparison of the forecast precision shows that in many situations the subset VAR models perform better than the corresponding unrestricted VAR models even if the zero restrictions imposed in

the subset VAR are not correct. Thus, if a 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 smaller, however, indicating that responses are estimated more precisely.

We conclude that the single equation 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 other modeling tools in addition. To avoid misspecification a comparison to the full VAR and the application of diagnostic tests is 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*, 19, 69-103.

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.

Hsiao, C. (1979), Autoregressive modeling of Canadian money and income data, *Journal of the American Statistical Association*, 74, 553-560.

Hsiao, C. (1982), Time series modelling and causal ordering of Canadian money, income and interest rates, in O.D. Anderson (ed.), *Time Series Analysis: Theory and Practice 1*, Amsterdam: North-Holland, 671-699.

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.

Penm, J.H.W. & R.D. Terrell (1984), Multivariate subset autoregressive modelling with zero constraints for detecting 'overall causality', *Journal of Econometrics*, 24, 311-330.

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.

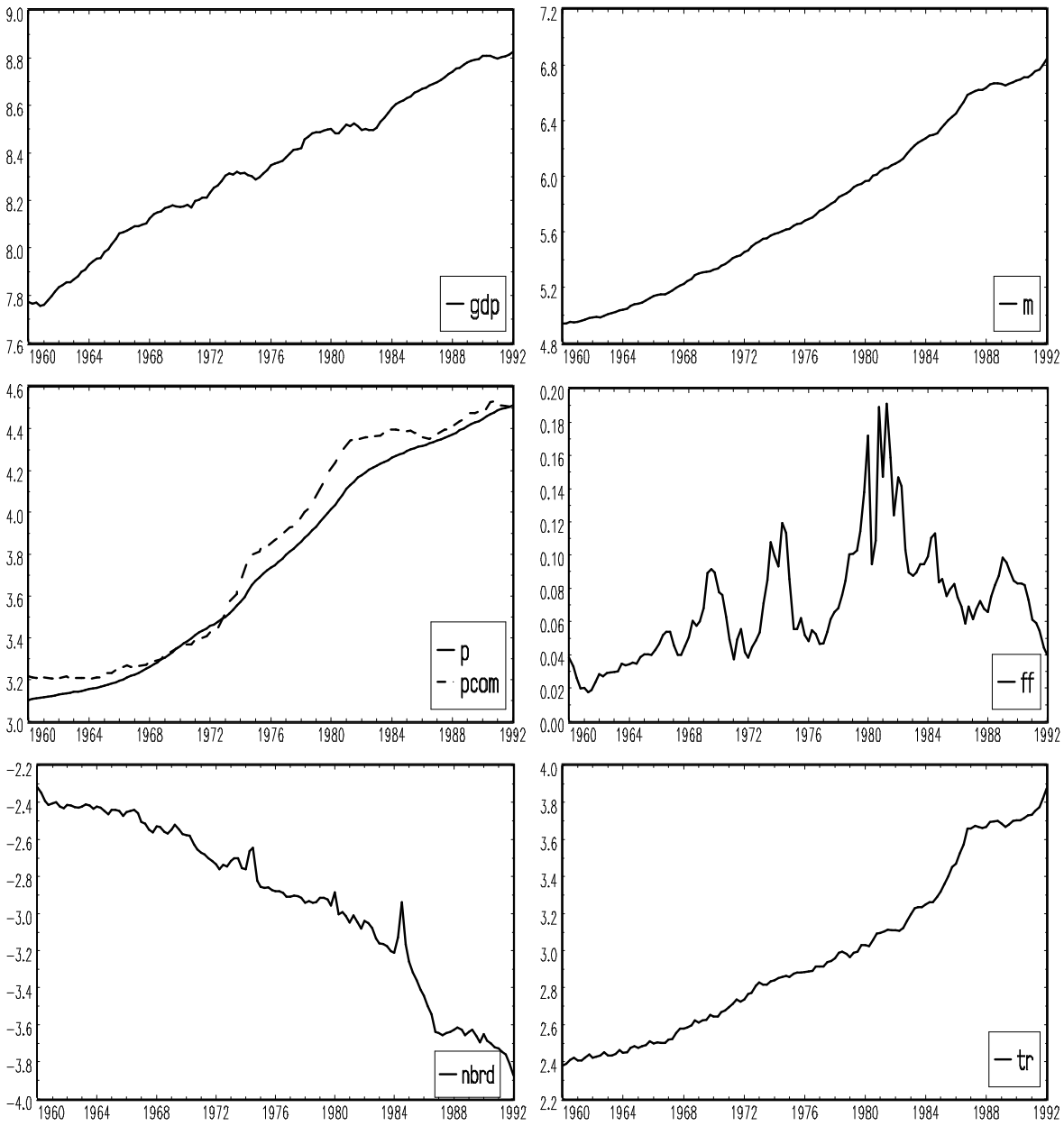

Figure 1: Time series analyzed

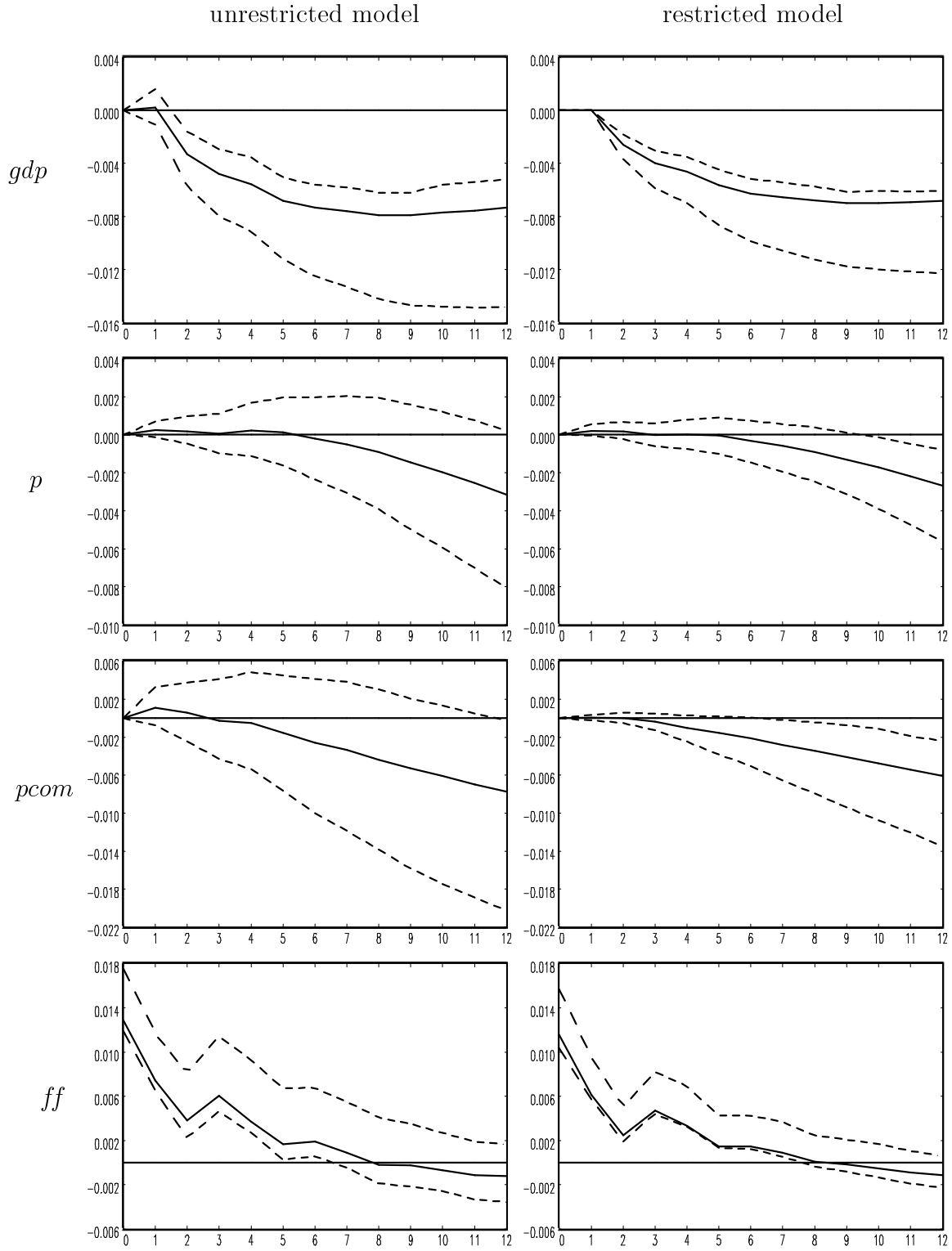

Figure 2: Responses of *gdp*, *p*, *pcom* and *ff* to a unit shock in *ff* computed from unrestricted (left) and restricted model (right)

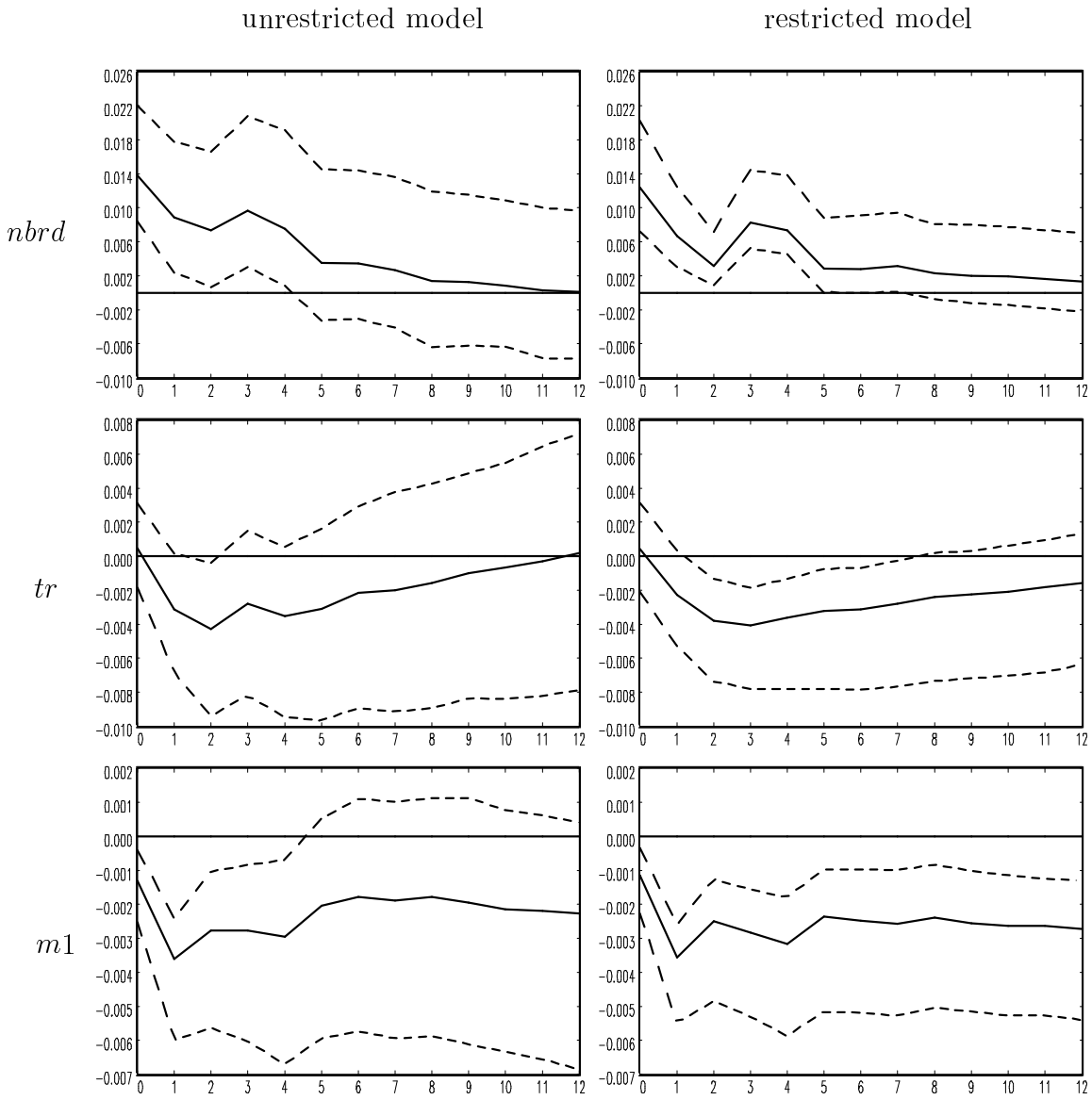

Figure 3: Responses of $nbrd$, tr and $m1$ to impulse in ff computed from unrestricted (left) and restricted model (right)