OPTIMAL LAG STRUCTURE SELECTION IN VEC–MODELS [1]

Peter Winker * **Dietmar Maringer** *

* *Department of Economics, Law and Social Sciences,*
University of Erfurt, Germany
{Peter.Winker,Dietmar.Maringer}@uni-erfurt.de

Abstract: For the modelling of time series, multivariate linear and nonlinear systems of equations became a standard tool. These models are also applied for non–stationary processes. However, estimation results in finite samples might depend on the specification of the model dynamics.
We propose a method for automatic identification of the dynamic part of VEC–models. Model selection is based on a modified information criterion. The resulting complex discrete optimization problem is tackled using a hybrid heuristic. We present the algorithm and results of a simulation study indicating the performance both with regard to the dynamic structure and the rank selection in the VEC–model.

Keywords: Time series analysis; variable structure; error correction; estimation algorithm; heuristic searches.

1. INTRODUCTION

For the modelling of time series, multivariate linear and nonlinear systems of equations became a standard tool. These models exhibit interesting features, e.g. dealing with non–stationary processes and cointegration. However, the issue of finite sample performance is relevant for these models which typically require the simultaneous estimation of a large set of parameters. In particular, vector autoregressive (VAR) models have been introduced as multivariate analog to the Box–Jenkins methodology. These models are used as reduced form approximations to model interdependent economic systems, e.g. for the analysis of transmission mechanisms. Vector error correction models (VECM) impose specific restrictions on the coefficients of a VAR in order to take into account non–stationarity and long–run comovement of the involved time series.

Both the results for the more traditional VAR– and the VEC–models depend on the specification of their dynamics. For example, Bewley and Yang (1998) compare the performance of different system tests for cointegration, i.e. the restrictions imposed in a VECM, when the lag length is selected by means of a standard information criterion. Both under– and overspecification of the lag length appear to have a negative impact. Ho and Sørensen (1996) considered higher dimensional systems and found that the negative impact of overspecification increases with the dimension.

In order to avoid or reduce this unwelcome specification effect, we propose an alternative approach. First, we employ a modified information criterion discussed by Chao and Phillips (1999) for the case of partially nonstationary VAR–models. Second,

[1] We are indebted to D. Hendry, S. Johansen, K. Juselius, H. Lütkepohl, and M. Meyer for helpful comments on an earlier draft of this paper. All remaining shortcomings lie solely with the authors.

we allow for “holes” in the lag structures, i.e. lag structures are not constrained to sequences of lags up to lag k. Using this approach, different lag structures can be used for different variables and in different equations of the system. This feature has to be taken into account in the estimation procedure for a given dynamic structure. For this purpose, we use a SUR–like modification of the two step reduced rank estimator proposed by Ahn and Reinsel (1990).

Consequently, the problem of model specification becomes an integer optimization problem on the huge set of all possible lag structures. In the context of VAR–models several methods have been proposed to tackle this problem of high computational complexity. In particular, heuristic optimization techniques which have already been applied to the linear VAR (Winker, 2000) can be extended to structural VAR– and VEC–models. However, the numerical methods used in estimating the model for a given dynamic structure become more involved in a VEC setting.

Section 2 introduces the model selection problem in the context of VEC–models. Section 3 describes the implementation of the optimization heuristic used to tackle the model selection problem in a VEC–setting. In section 4, we present some Monte Carlo evidence on the performance of the method. Section 5 summarizes the findings and provides an outlook to further steps of our analysis.

2. THE MODEL SELECTION PROCEDURE

The standard procedure for model selection in a VEC–model setting is sequential. First, information criteria are used to choose a lag length k for the unrestricted VAR–model. Then, it is assumed that the correct specification of the lag structure is given. However, for the sample size typically observed in economics, this assumption does not has to be fulfilled. In particular, by imposing the additional restriction that all lags up to k of all variables are included in all equations a correct lag specification becomes even less likely. The second step of the analysis consists in performing a sequence of cointegration tests. The statistical properties of this sequential procedure are difficult to assess. Consequently, it cannot be guaranteed that the final estimate of the cointegration rank obtained by this procedure is a consistent estimate of the true model structure (Johansen, 1992; Jacobson, 1995; Chao and Phillips, 1999).

Chao and Phillips (1999) propose to approach the task of rank determination as a model selection problem. They introduce a modification of the BIC and a posterior information criterion (PIC) for VEC–models. Using these criteria exhibits three advantages: First, lag structure and the cointegration rank can be selected in a single step. Second, the penalty function of both criteria reacts to under– and over–parameterization, which both might have a detrimental effect on the estimation of the cointegration rank. Third, the method can easily be extended to cover the case of different lag structures across equations including “holes”.

For our MC simulations, we consider both the modified BIC (BICm) and the modified posterior information criterion (PICm) presented in Chao and Phillips (1999, p. 236). However, the PICm is considered solely for a comparison of different criteria applied to models containing all lags up to a certain order.

We consider the d–dimensional VAR–model of order $k+1$

$$Y_t = \sum_{i=1}^{k+1} \Pi_i Y_{t-i} + \varepsilon_t \qquad (1)$$

with initial values $\{Y_0, Y_{-1}, ..., Y_{-k}\}$. Thereby, the error terms ε_t are assumed to be iid $N(0, \Omega)$. Furthermore, it is assumed that the characteristic polynomial of the VAR may have roots on the unit circle, but no explosive components. The VAR–model (1) can also be expressed in vector error–correction notation as

$$\Delta Y_t = \sum_{i=1}^{k} \Gamma_i \Delta Y_{t-i} + \Pi Y_{t-k-1} + \varepsilon_t \, . \qquad (2)$$

The matrix Π represents the parameters of the error correction term of the model. Consequently, the cointegration rank of the system is given by the rank of Π. For each $0 \leq r \leq d$, there exist $d \cdot r$ matrices α and β of full rank such that $\Pi = \alpha\beta'$.[2] Finally, we require that ΔY_t is a stationary process allowing for a Wold representation.

In the standard approach, only values for k and r have to be determined. If a maximum lag length k_{max} is given, the number of models to be considered amounts to k_{max}. A complete enumeration is feasible and will serve as a benchmark in our simulation analysis. However, a priori there is no reason to expect that the dynamic structure is of this standard type. Therefore, we allow for different lag structures across equations and for “holes” in the lag structure. Consequently, we have to choose a lag structure out of $2^{d^2 \cdot k_{max}}$ possible sets. A simple enumeration approach will fail in this case except for very small instances. As in Winker (2000) we use an optimization heuristic to tackle this problem.

For given lag structure and cointegration rank, the information criterion BICm is calculated using a modification of the iterative estimation procedure

[2] For $r = 0$, we choose $\alpha = \beta = 0$, for $r = d$, $\alpha = \Pi$ and $\beta = I$ is a solution.

proposed by Ahn and Reinsel (1990) for the reduced rank case. The modification takes into account that different lags might appear in different equation, i.e. allow for a SUR–like model.

3. THE ALGORITHM

The algorithm for finding the optimal lag structure is a hybrid heuristic combining ideas of the Threshold Accepting (TA) algorithm as described in Winker (2001) and of "Memetic Algorithms" (Moscato, 1999; Maringer and Winker, 2003). For a given cointegration rank r, a random initial lag structure is chosen, the parameters are estimated and the value for the information criterion BICm is computed. During the following iteration steps, a local search strategy is employed where the structure is modified by either including one additional or excluding one hitherto included lagged variable in one of the equations. If the information criterion is improved or if the impairment does not exceed a given threshold, i.e. if $\Delta BICm \leq T_i$, the modified lag structure is accepted. Otherwise, the modification is undone and the previous lag structure is restored. In the course of the iterations, the threshold is lowered, so that hardly any impairment is accepted in the last iterations. Consequently, the algorithm is well apt to overcome local optima and to fine–tune the solution once the "core structure" has been identified.

Whereas in TA a single agent is representing one solution per iteration, we enhanced the original TA concept much in the sense of Memetic Algorithms by replacing the single agent by a population of agents each of which follows the TA search strategy. In addition to their independent local search, the agents "compete" with each other and can combine parts of their solutions using a cross–over operator.

The heuristic optimization is repeated for all possible values of the rank r, i.e. $0 \leq r \leq d-1$. Let BICm_r denote the minimum value of the information criterion obtained by the optimization heuristic for a rank of r. Let $r^{opt} = \mathrm{argmin}_{0 \leq r \leq d-1} \mathrm{BICm}_r$, then the finally selected model is the one with rank r^{opt} and the corresponding dynamic lag structure. The selection of rank and lag length for the standard "take all up to the k–th lag" approach is performed in a similar way.

4. MONTE CARLO SIMULATION

4.1 Motivation

We use the iterative algorithm proposed by (Ahn and Reinsel, 1990) to estimate the parameters of the reduced rank models. Unfortunately, this procedure is quite time consuming even if good starting values are provided. Therefore, the high overall computational complexity of automatic lag order selection in the VEC–models limits the number of different settings which can be analyzed by means of MC simulation. Consequently, we tried to assess the relative performance of the method by considering a few typical cases. Given the page constraint of this contribution, we report only results for two artificial DGPs.

4.2 Simulation Setup

The results presented in this section are based on the simulation of two different DGPs with different rank and lag structure. The details of these DGPs are introduced below. The first DGP (DGP_1) is taken from Chao and Phillips (1999, pp. 242f, Experiment 5). The second DGP (DGP_2) adds a second cointegration vector and extends the dynamic structure.

For each replication of the two DGPs 300 observations have been generated from which the first 145 are eliminated, leaving a sample length of $T = 155$. We ran 100 and 200 replications, respectively, and for each replication the rank was estimated by the methods "all up to the k-th lag" (labelled "all") and our optimization heuristic allowing for structures with "holes" in the k_{max} lags (labelled "holes") with $k_{max} = 5$ for both methods.

DGP_1 Experiment 5 in Chao and Phillips (1999) is a three dimensional VECM with one cointegration vector entering a single equation of the system and a lag length of one. Thereby, lagged differences of the endogenous variables enter only the equation for the respective variables. The error correction term is described by the matrix Π, Γ_1 provides the coefficients of the dynamic part and Ω_ε the variance–covariance matrix of the normally distributed error terms:

$$\Pi = \begin{pmatrix} 0 \\ -0.01 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0.25 & 0.8 \end{pmatrix}$$

$$\Gamma_1 = \begin{pmatrix} 0.99 & 0 & 0 \\ 0 & 0.9025 & 0 \\ 0 & 0 & 0.99 \end{pmatrix} \Omega_\varepsilon = \begin{pmatrix} 2.25 & 2.55 & 1.95 \\ 2.55 & 3.25 & 2.81 \\ 1.95 & 2.81 & 2.78 \end{pmatrix} .$$

The moduli of nonzero reverse characteristic roots of the process are 1, 1, 0.99, 0.99, 0.95, 0.95.

DGP_2 Modifying the above DGP by adding a second cointegration vector and lags of order 2 and 3 in the dynamic part, we obtain DGP_2 with an actual rank of 2 and the following parameters:

$$\Pi = \begin{pmatrix} 0 & -0.005 \\ -0.005 & 0 \\ -0.002 & 0.003 \end{pmatrix} \begin{pmatrix} 0.8 & 0.25 & 0.5 \\ 0.4 & 0.10 & -0.3 \end{pmatrix}$$

$$\Gamma_1 = \begin{pmatrix} 0.59 & 0 & 0 \\ 0 & 0.725 & 0 \\ 0 & 0 & 0.84 \end{pmatrix} \Gamma_2 = \begin{pmatrix} 0.25 & 0 & 0 \\ 0.02 & 0.10 & 0 \\ -0.05 & 0 & 0.05 \end{pmatrix}$$

$$\Gamma_3 = \begin{pmatrix} 0 & 0.05 & -0.1 \\ 0 & 0 & 0 \\ 0.1 & -0.1 & 0.05 \end{pmatrix} \Omega_\varepsilon = \begin{pmatrix} 4.5 & 5.1 & 3.9 \\ 5.1 & 6.5 & 5.62 \\ 3.9 & 5.62 & 5.56 \end{pmatrix} .$$

The moduli of nonzero reverse characteristic roots of the process are 1, 0.99755, 0.96160, 0.96160, 0.88443, 0.88443, 0.35230, 0.35230, 0.30986, 0.30986, 0.13375. Obviously, the second root is very close to one. Thus, although the "true" rank of this model is two, it is close to a process with rank one.

4.3 Results

The evaluation of the Monte Carlo results focuses on the estimated cointegration rank. For the models allowing for "holes", we also present information on the average size and the average power. As a measure of possible overfitting, we report mean values for the quotient q_Σ of the determinant of the residual covariance matrix for the selected models and for the true DGP.

Results of the "take all up to the k–th lag" approach First, we present findings for the "take all up to the k–th lag" approach comparing different methods. Table 1 summarizes the findings for 1 000 replications of DGP_1. For the modified BIC and PIC criterion, the table entries indicate the number of times the corresponding rank and lag length has been selected by the criteria. For the Johansen testing procedure, a two–step approach is used. First, the lag length of the unrestricted VAR is selected according to the BIC. Then, the trace test for the cointegration rank is conducted using this lag length. The table entries indicate the number of times the corresponding rank and lag length is found by this two–step approach using a 1%– and a 5%–critical value for the trace test, respectively.

Obviously, for DGP_1 all four methods identify the actual lag length of one for all replications. Although the lag structure of DGP_1 is sparse since only the diagonal elements are different from zero, the high numerical values of these diagonal elements force all methods to choose a lag length of one. Nevertheless, the four methods differ markedly in their ability to identify the actual cointegration rank of the model. While the modified PIC points to the correct rank of one in 999 out of 1 000 replications, the share of correct identifications of the cointegration rank

Table 1. Results for the "take all up to the k–th lag" approach (DGP_1)

Modified BIC							Modified PIC						
	Lags							Lags					
Rank	0	1	2	3	4	5	Rank	0	1	2	3	4	5
0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	805	0	0	0	0	1	0	999	0	0	0	0
2	0	160	0	0	0	0	2	0	1	0	0	0	0
3	0	35	0	0	0	0	3	0	0	0	0	0	0
Johansen (1%)							Johansen (5%)						
	Lags							Lags					
Rank	0	1	2	3	4	5	Rank	0	1	2	3	4	5
0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	750	0	0	0	0	1	0	509	0	0	0	0
2	0	210	0	0	0	0	2	0	343	0	0	0	0
3	0	40	0	0	0	0	3	0	148	0	0	0	0

shrinks to 80.5% for the modified BIC and to 75% and 50.9%, respectively, when using Johansen's procedure with a nominal significance level of 1% and 5%, respectively.

Table 2. Results for the "take all up to the k–th lag" approach (DGP_2)

Modified BIC							Modified PIC						
	Lags							Lags					
Rank	0	1	2	3	4	5	Rank	0	1	2	3	4	5
0	0	81	376	1	0	0	0	0	18	879	4	0	0
1	0	60	357	17	0	0	1	0	7	65	27	0	0
2	0	14	78	1	0	0	2	0	0	0	0	0	0
3	0	0	15	0	0	0	3	0	0	0	0	0	0
Johansen (1%)							Johansen (5%)						
	Lags							Lags					
Rank	0	1	2	3	4	5	Rank	0	1	2	3	4	5
0	0	41	23	0	0	0	0	0	11	3	0	0	0
1	0	95	623	25	0	0	1	0	89	451	18	0	0
2	0	19	149	4	0	0	2	0	44	249	9	0	0
3	0	3	17	1	0	0	3	0	14	109	3	0	0

Table 2 exhibits the corresponding results for DGP_2. In contrast to the simpler dynamic structure of DGP_1, all four methods fail to identify the correct lag length for most replications. However, given that our main interest is in the long–run structure of the model, we might concentrate on the identification of the cointegration rank. The actual rank of two is found in 17.2% and 30.2% of the replications when using Johansen's procedure with level 1% and 5%, respectively. The modified BIC results in 9.3% correct estimates of the cointegration rank, while the modified PIC never results in a cointegration rank of two. These results confirm the findings by Gonzalo and Pitarakis (1999) that the relative performance of different methods might depend strongly on the DGP under consideration. Obviously, the second near unit root leads to the high rejection rates of the models with rank 2.

Summarizing the findings for the different criteria, at least for the two DGPs under consideration, the modified BIC criterion appears to be a sensible choice. Thus, the following results of the optimization approach concentrate on this criterion. Nevertheless, it is left to future research to also provide results for the PICm and Johansen's procedure.

Results of the Optimization Heuristic In the following, we present results of the implementation of the optimization heuristic described in section 3. In order to obtain a concise description of the results, we concentrate on the identification of the cointegration rank when using the modified BIC. Furthermore, we restrict our analysis to a cointegration rank between 0 and $d-1$. This implies the assumption that the case of a stationary VAR could be excluded by a standard unit root pretest. For DGP_1 and DGP_2 we analyze 200 different realizations with 150 observations. For each realization, three different methods have been used to determine the cointegration rank:

"known" The model is estimated for a cointegration rank of $p = 0, \ldots, d-1$ assuming that the actual lag structure is known, i.e. only the non–zero elements of the matrices Γ_i are included in the estimation.

"all" The model is estimated for a cointegration rank of $p = 0, \ldots, d-1$ and using all lags up to a given order $k = 1, \ldots, k_{max}$. For our application $d = 3$ and $k_{max} = 5$. Consequently, 15 different model specifications are estimated.

"holes" For each possible cointegration rank of $p = 0, \ldots, d-1$ a heuristic optimization is performed on the lags to be included in the dynamic part of the model.

For all three methods, the reported cointegration is defined by the minimum value of BICm obtained for the different rank conditions. Although the method "known" cannot be used in practical applications, it is used as a benchmark for our optimization approach ("holes"). By contrast, the method "all" represents the state of the art in criterion based model selection. Consequently, it is of interest to evaluate the relative performance of the last two methods.

Table 3 summarizes the results for the three methods applied to the two DGPs based on 200 replications. For all methods and DGPs the maximum lag length k_{max} has been fixed to five. The numbers in the table indicate the percentage share of replications for which the methods identify a cointegration rank of $p = 0, \ldots, 2$ based on the modified BIC.

For the first DGP with its quite simple dynamic structure, all three methods appear to work reasonably well. Nevertheless, the chance of identifying the actual cointegration rank $p = 1$ based on 150 observations is best if the true dynamics are known. The optimization procedure increases the frequency of finding the right cointegration rank

Table 3. Cointegration rank estimates

	Method		
Rank	"known"	"all"	"holes"
DGP_1 (200 replications)			
0	0.0%	0.0%	0.0%
1	95.0%	80.5%	88.0%
2	5.0%	19.5%	12.0%
DGP_2 (200 replications)			
0	13.0%	50.5%	13.4%
1	73.8%	35.8%	74.8%
2	13.2%	13.7%	11.8%

from 80% to 89% as compared to the standard method.

For DGP_2 with its quite complex lag structure and the second near unit root (0.99755), even when assuming that the true lag structure is known, only in 13% of all replications the actual cointegration rank of 2 is found. Further analysis is required to identify the reasons for this outcome. The two methods which can be used in applications, i.e. "all" and "holes", report the correct cointegration rank with frequency 13.7% and 11.8%, respectively. Although, "all" appears to have a slight advantage in finding the correct cointegration rank, it also results in a more than 50% chance of finding no cointegration at all, whereas the "holes" approach provides results quite similar to the ones obtained when the true DGP was known.

Although our main interest is in a correct specification of the cointegration part of our models, we finish by a short look on the dynamic structures selected by the three methods. Of course, this choice appears to be crucial for the determination of the cointegration rank. Table 4 reports on the dynamic structure for the three DPGs. The rows labelled ν shows the mean number of non zero elements estimated in the dynamic part of the model, i.e. the number of non zero entries in $\hat{\Gamma}_1, \ldots, \hat{\Gamma}_5$. In the rows with label "cl" we provide the share of lags present in the DGP which are included in the estimated models ("average power"), while "wl" provides the share of lags included in the estimated model, but not present in the DGPs ("average size"). Finally, q_Σ indicates the quotient of the determinant of the residual covariance matrices for the model under consideration as compared to the true DGP.

For a simple dynamic structure like DGP_1, the optimization method appears to work extremely well by finding the relevant lags (on the diagonal of $\hat{\Gamma}_1$) for all replications and including only a small number of additional lags. The standard method has to include all nine first order lags in order to capture the relevant lags. Consequently, the share of non relevant lags increases as the mean number of lags included ($\nu = 9$). Only for this rather simple DGP, q_Σ indicates a slight tendency

Table 4. Reported lag structure for different selection methods

	Method		
	"known"	"all"	"holes"
DGP_1 (200 replications)			
ν	3	9	4.47
cl	100%	100%	100%
wl	0%	14.3%	3.5%
q_Σ	1.00	0.92	0.94
DGP_2 (200 replications)			
ν	13	19.89	17.15
cl	100%	69.2%	52.0%
wl	0%	30.7%	27.1%
q_Σ	1.00	1.16	0.99

of overfitting for the "take all up" approach and – to a smaller extent – for the "holes" method. For the other DGP no overfitting is indicated by this measure, but the "holes" approach results in better fitting models with a determinant of the residual covariance matrix close to that of the true DGP. The share of relevant lags identified by the optimization heuristic is smaller than for the "all" heuristic, which is surprising at first sight given the larger search space. This result deserves further attention. Nevertheless, it is remarkable that the optimization heuristic seems to identify those lags allowing for a correct estimation of the cointegration rank more often than the "all" heuristic (see Table 3). Our preliminary results support earlier findings that the performance of model selection procedures in the context of cointegration depends heavily on the specific DGP. In particular, as Gredenhoff and Karlsson (1999, p. 184) we might conclude that "choosing the lag–length in VAR–models is not an easy task".

5. CONCLUSION

We compare different methods for model selection in VEC–models including methods based on information criteria and a two–step procedure employing Johansen's testing strategy. We introduce a discrete optimization heuristic allowing for the selection of lag structures with "holes". By means of a MC simulation, we find that the modelling of the dynamic part of VEC–models is crucial for a correct rank identification. Already our small set of DGPs indicates that this effect differs markedly for different DGPs. In particular, the practical guideline rather to include too many lags is not supported for all DGPs by our findings. However, we find that the optimization heuristic approach in combination with a modified BIC performs relatively well.

Future research will apply our method to a much larger set of different DGPs in order to find out how robust our results are and which factors are responsible for differences in the (relative) performance. Second, we want to include other procedures in our approach, in particular the modified PIC suggested by (Chao and Phillips, 1999) and Johansen's procedure.

Finally, the algorithm for model selection presented in this paper can also be applied for other model selection problems arising in economic applications and beyond.

REFERENCES

Ahn, S. K. and G. C. Reinsel (1990). Estimation for partially nonstationary multivariate autoregressive models. *Journal of the American Statistical Association* **85**(411), 813–823.

Bewley, R. and M. Yang (1998). On the size and power of system tests for cointegration. *The Review of Economics and Statistics* **80**(4), 675–679.

Chao, J. C. and P. C. B. Phillips (1999). Model selection in partially nonstationary vector autoregressive processes with reduced rank structure. *Journal of Econometrics* **91**, 227–271.

Gonzalo, J. and J.-Y. Pitarakis (1999). Lag length estimation in large dimensional systems. *Journal of Time Series Analysis* **23**(4), 401–423.

Gredenhoff, M. and S. Karlsson (1999). Lag–length selection in VAR–models using equal and unequal lag–length procedures. *Computational Statistics* **14**, 171–187.

Ho, M. S. and B. E. Sørensen (1996). Finding cointegration rank in high dimesional systems using the Johansen test: An illustration using data based Monte Carlo simulations. *The Review of Economics and Statistics* **78**(4), 726–732.

Jacobson, T. (1995). On the determination of lag order in vector autoregressions of cointegrated systems. *Computational Statistics* **10**(2), 177–192.

Johansen, S. (1992). Determination of cointegration rank in the presence of a linear trend. *Oxford Bulletin of Economics and Statistics* **54**(3), 383–397.

Maringer, D. and P. Winker (2003). Portfolio optimization under different risk constraints with memetic algorithms. Technical Report 2003–005E. Staatswissenschaftliche Fakultät. Universität Erfurt.

Moscato, P. (1999). Memetic algorithms: A short introduction. In: *New Ideas in Optimization* (D. Corne, M. Dorigo and F. Glover, Eds.). pp. 219–234. MacGraw-Hill. London.

Winker, P. (2000). Optimized multivariate lag structure selection. *Computational Economics* **16**, 87–103.

Winker, P. (2001). *Optimization Heuristics in Econometrics: Applications of Threshold Accepting*. Wiley. Chichester.