Optimal Lag Structure Selection in VEC–Models*

Peter Winker[†]   Dietmar Maringer[‡]

Preliminary version: February 26, 2004

Abstract

For the modelling of economic and financial time series, multivariate linear and nonlinear systems of equations became a standard tool. These models might also be applied in the context of non–stationary processes. However, estimation results in finite samples might depend strongly 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 lag structure of the model is selected according to this objective function allowing for "holes". The resulting complex discrete optimization problem is tackled using a hybrid heuristic combining ideas from threshold accepting and memetic algorithms. 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. The results indicate that the selected rank might depend strongly on the dynamic specification of the VEC–model.

Keywords: Model selection; cointegration rank; information criterion; order selection; reduced rank regression; VECM.
JEL classification: C32; C61.

[†]Department of Economics, Law and Social Sciences, University of Erfurt
[‡]Department of Economics, Law and Social Sciences, University of Erfurt

1 Introduction

For the modelling of economic and financial time series, multivariate linear (VAR, SVAR, VECM) and nonlinear systems of equations (MS–VAR) became a standard tool over the last few years. As compared to univariate approaches, these models exhibit interesting features, e.g. dealing with non–stationary processes and cointegration. However, the issue of finite sample performance becomes even more relevant for these models which typically require the simultaneous estimation of a large set of parameters.[1] While economic theory might provide a guideline for long–run or equilibrium relationships,[2] the modelling of the dynamic part has to rely on different inputs.[3]

In order to avoid a simple ad hoc specification of the dynamic part, several statistical procedures have been proposed. For example, the lag structure in VAR–models is based on tests on residual autocorrelation (Jacobson, 1995) or information criteria like AIC and BIC (Winker, 2000).[4] However, these approaches did not take into account potential non–stationarity of the time series and the restrictions imposed by the rank conditions in a cointegration framework. In several contributions,[5] the effect of lag length selection on the outcomes of tests for cointegration,[6] in particular on the cointegration rank, has been demonstrated. Bewley and Yang (1998) compare the performance of different system tests for cointegration when the lag length is selected by means of a standard information criterion (AIC or BIC, respectively). Both under– and overspecification of the lag length appear to have a negative impact.[7] While the effect on size is small for tests for rank equal to zero, the

---

[1] For example, a small sample correction of Johansen's test is proposed by Johansen (2002).

[2] The issue of identification and restriction of long–run relationships based on statistical tests and prior information is discussed by Omtzigt (2002).

[3] In this paper, we neglect the specification of deterministic trend terms, which might have similar implications on the outcomes (Ahking, 2002).

[4] A different approach focusing on general–to–specific reductions, which eliminates statistically insignificant variables and uses diagnostic tests to check the validity of reductions is presented by Hendry and Krolzig (2001) and Krolzig and Hendry (2001). Brüggemann *et al.* (2003) provide a comparison of different methods.

[5] See also the references provided by Ho and Sørensen (1996) in their introduction and by Pötscher (1991).

[6] In general, the analysis is conducted in the framework of Johansen's testing procedure (Johansen, 1988; Johansen, 1991; Johansen, 1992; Johansen, 1995).

[7] Gonzalo and Pitarakis (1999) analyse the performance of model selection criteria in large dimensional VARs. They find that underfitting might become as important as over-

effect on power is more substantial for certain parameter combinations. This effect becomes even more pronounced for tests of the null of a cointegration rank of one. Ho and Sørensen (1996) considered higher dimensional systems and found that the negative impact of overspecification increases with the dimension. In particular, application of Johansen's test tends to underestimate the number of unit roots in the system, and, in due course, to overestimate the cointegration rank in this case.

The model selection procedure analyzed in this paper differs in two aspects from the methods mentioned above. First, we employ a modified information criterion discussed by Chao and Phillips (1999) for the case of partially nonstationary VAR–models.[8] Consequently, the dynamic model selection is performed taking the restrictions of reduced rank regressions into account. Second, we allow for "holes" in the lag structures, i.e. lag structures are not constrained to sequences of lags up to lag k, but might consist, e.g., of the first and fourth lag only in an application to quarterly data. 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.[9] For this purpose, we use a SURE–like modification of the two step reduced rank estimator proposed by Ahn and Reinsel (1990).[10]

Using this approach, 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. Exact algorithms are based on an intelligent enumeration of possible models avoiding the evaluation of all cases

---

fitting when the dimension of the process increases even for the AIC. Furthermore, the performance of information criteria depends critically on the specific DGP under consideration.

[8] Analyzing the effects of choosing alternative criteria, e.g. along the lines suggested by Campos *et al.* (2003), is left for future research. The same applies to the combination with a pre–selecting step also discussed by Campos *et al.* (2003).

[9] As pointed out by Gredenhoff and Karlsson (1999), in the literature on model selection in VAR–models, the possibility that the true model may have unequal lag–length or even holes in the lag structure has received little attention. Although they use the Hsiao procedure which does not investigate all combinations of lag–lengths for the different variables and does not allow for holes at all, their simulation results indicate that, in particular for more complex lag structures, their unequal lag length procedure appears improve results.

[10] An application in the ML setting of Johansen's procedure is left for future research.

(Gatu and Kontoghiorghes, 2003). Nevertheless, this approach is still of high computational complexity and appears to be limited in the current stage to linear models without further restrictions. In contrast, 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, i.e. the two step reduced rank estimator used in this paper or a ML estimator, become more involved in a VEC setting. In due course, the overall computational complexity increases. This high computational load sets a limit to the number of different data generating processes (DGP) and parameter settings which can be used for our MC simulation analysis. Nevertheless, our first results indicate that the method works well in practice and might be superior to the standard "take all up to the k–th lag" approach in specific settings.

The paper is organized as follows. Section 2 introduces the model selection problem in the context of VEC–models. We present the model, the information criteria and the resulting integer optimization problem. Section 3 describes the implementation of the heuristic used to solve this optimization problem. In section 4 we present some Monte Carlo evidence on the performance of the method applied to different data generating processes. The results are compared to the standard method of choosing all lags up to a certain order. 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 consists of a sequential procedure. First, information criteria like AIC or BIC are used to choose a lag length for the unrestricted VAR–model.[11] For the next steps of the analysis, it is assumed that the correct specification of the lag structure is given.[12] Then, for the determination of the cointegration rank, a sequence of cointegration tests is performed. The statistical properties of

---

[11]Ho and Sørensen (1996) find evidence in favour of using BIC when a cointegration analysis is intended. Winker (1995, 2000) generalizes this model selection step to allow for different lag structures across equations including "holes".

[12]However, several simulation studies have demonstrated that lag misspecification adversely affects the outcome of the cointegration tests conducted in the second step (Ho and Sørensen, 1996; Bewley and Yang, 1998).

this sequential procedure are difficult to assess. Consequently, it cannot be guaranteed that the final estimation of the cointegration rank obtained by this procedure is a consistent estimate (Johansen, 1992; Jacobson, 1995; Chao and Phillips, 1999).

In order to circumvent these shortcomings of the traditional approach, Chao and Phillips (1999) propose to reconsider the problem from the viewpoint of model selection.[13] They propose a modification of the BIC and a posterior information criterion (PIC) for the application to VAR processes with reduced rank cointegration structure. Using these criteria for model selection 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 (Bewley and Yang, 1998). Application of this criterion provides a consistent estimation of lag structure and cointegration rank (Chao and Phillips, 1999). Third, the method can easily be extended to cover the case of different lag structures across equations including "holes".

For the MC simulations presented in this paper, 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. Our goal is to assess the advantage of allowing for holes in the lag structure for the determination of the cointegration rank of a VECM as compared to the standard "take all up to the k–th lag" approach. For the optimization of lag structures allowing for holes, we restrict the analysis to the BICm for computational reasons. Inclusion of the PICm is part of our future research agenda.

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

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

with initial values $\{Y_0, Y_{-1}, ..., Y_{-k}\}$. Thereby, the error terms ε_t are assumed to be iid $\mathcal{N}(0, \Omega)$. Furthermore, it is assumed that the characteristic polynomial of the VAR may have unit roots on the unit circle, but no explosive

---

[13]Bahmani-Oskooee and Brooks (2003) also propose a global criterion based on the goodness of fit of the resulting long–run relationships. However, they do not provide MC evidence on the relative performance of this approach.

components. The VAR–model (1) can also be expressed in vector error–correction notation as[14]

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

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 $\Pi = \alpha\beta'$.[15] Finally, we require that ΔY_t is a stationary process allowing for a Wold representation (Chao and Phillips, 1999, p. 229).

In the standard approach, i.e. taking all lags up to a specified order in all equations, the model selection problem consists in determining values for k and r. If a maximum lag length k_{max} is assumed to be given, the number of models to be considered amounts to $d \cdot k_{max}$. A complete enumeration of these models 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 extend this approach to allow for different lag structures across equations and for "holes" in the lag structure of any equation. Consequently, we have to choose a lag structure out of $2^{d^2 \cdot k_{max}}$ possible sets. Obviously, a simple enumeration approach will fail in this case except for very small instances. As in Winker (1995, 2000), we employ a heuristic optimization technique to tackle this problem. The method will be described in section 3.

Before turning to the optimal selection of the dynamic lag structure, we have to provide more details on the calculation of the information criterion BICm for given lag structure and cointegration rank. We use a modification of the iterative estimation procedure proposed by Ahn and Reinsel (1990) for the reduced rank case. However, we have to modify this method to allow for different lag structures across equations. The parameter estimates are obtained iteratively by

$$\widehat{b}^{\iota+1} = \widehat{b}^{\iota} + \left(\sum_{t=1}^{T} U_t^* \widehat{\Omega}_\varepsilon^{-1} U_t^{*'} \right)^{(-1)} \left(\sum_{t=1}^{T} U_t^* \widehat{\Omega}_\varepsilon^{-1} \widehat{\varepsilon}_t \right) \tag{3}$$

---

[14]For the empirical application, we employ the asymptotically equivalent representation from, e.g., Ahn and Reinsel (1990, p. 817), of $\Delta Y_t = \sum_{i=1}^{k} \Gamma_i \Delta Y_{t-i} + \Pi Y_{t-1} + \varepsilon_t$.

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

where ι is the current iteration, $\widehat{\Omega}_\varepsilon$ is the covariance matrix of the residuals, $\hat{\varepsilon}$, under the current parameter estimates $\widehat{\beta}^\iota$ and

$$U_t^* = \left[(\alpha' \otimes [0, I_{d-r}]Y_{t-1})', I_d \otimes [(\beta Y_{t-1})', \Delta Y_{t-1}, ..., \Delta Y_{t-k}]' \right]' \qquad (4)$$

where β is normalized so that $\beta = [I_r, \beta_0]$. The matrices Π and Γ_i can then be determined by decomposing the parameter vector b with

$$b = [vec(\beta_0')', vec((\alpha, \Gamma_1, ..., \Gamma_k)')']'. \qquad (5)$$

The initial solution for b^0, can be found from a full rank SUR estimate which decreases the number of necessary iterations significantly and therefore increases the convergence speed.

To introduce the "holes" into the lag structure, i.e., setting some of the elements of the Γ_i's equal to 0, the respective columns in $U_t^{*'}$ are eliminated and the (de–)composition of b has to be adapted. The information criterion BICm can then be calculated according to

$$BICm = \ln |\widehat{\Omega_{\Upsilon,r}}| + \frac{\nu + r(d - r) + dr}{T} \cdot \ln(T) \qquad (6)$$

where Υ denotes the set of elements of the Γ_i's, and $\nu = \sharp \Upsilon$, i.e. is the number of elements of the Γ_i's that are not equal to zero. $\widehat{\Omega_{\Upsilon,r}}$ is computed following Chao and Phillips (1999) in two steps: first, the parameters α, β and Γ_i are estimated iteratively as described previously for given rank r and lag structure Υ. Next, the corrected values $\Delta Y_t^* = \Delta Y_t - \alpha \beta' Y_{t-1}$ are computed. Finally, the Γ_i's are re–estimated by running a SUR estimation of ΔY_t^* on the ΔY_{t-i}'s included in the given lag structure Υ. $\widehat{\Omega_{\Upsilon,r}}$ denotes the covariance matrix of the residuals of this last regression.

3 The Algorithm

The algorithm for finding the optimal lag structure allowing for holes is a hybrid heuristic combining ideas of the Threshold Accepting (TA) algorithm as described in Winker (2001) and of "Memetic Algorithms"[16]. 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

---

[16]Cf. Moscato (1999) and Maringer and Winker (2003).

along the lines described in the previous section. 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 is acceptable in the sense that it does not exceed a given threshold, i.e. if $\Delta BICm \leq T_i$, the modified lag structure is accepted. If, however, the modified lag structure degrades the information criterion more than tolerated by the current value of the threshold sequence, this modification is undone and the previous lag structure is restored. During the early iteration steps, the threshold is chosen rather generously and most of the modifications are actually accepted. In the course of the iterations, the threshold is persistently 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 on a regular basis where one agent challenges another and passes his (current) structure on to the challenged agent if the change in the challenged agent's information criterion does not violate the threshold criterion. Also, agents can combine parts of their solutions using a cross–over operator (Fogel, 2001) where an offspring will replace a parent if, again, the impairment in the information criterion does not exceed the threshold.

The heuristic optimization is repeated for all possible values of the rank r, i.e. $0 \leq r \leq d - 1$.[17] Let $BICm_r$ denote the minimum value of the information criterion obtained by the optimization heuristic for a rank of r. Let $r^{opt} = \operatorname{argmin}_{0 \leq r \leq d-1} BICm_r$, 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. For all possible values of the rank r and all k, $0 \leq k \leq k_{max}$, the value of the criterion $BICm$ is calculated. The pair (r, k)

---

[17]The case $r = d$ is not considered as it corresponds to a stationary VAR–model. A method for model selection in VAR–models by means of optimization heuristics is presented in Winker (2000).

resulting in the minimum value of $BICm$ describes the model identified by the "take all up to the k–th lag" approach.[18]

4 Monte Carlo Simulation

4.1 Motivation

The evaluation of the information criterion BICm used for model selection in this paper requires the estimation of the parameters of the reduced rank models. For this purpose we employ the iterative algorithm proposed by Ahn and Reinsel (1990). This procedure is quite time consuming even if good starting values are provided. Consequently, the number of iterations of our hybrid heuristic has to be limited in order to allow for at least some replications in a MC setting. Finally, this 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. Besides using artificial DGPs, we follow Ho and Sørensen (1996) for some of our simulations by using parameter values obtained from an estimation using actual data. Given that our simulations can only pick a small number of parameter settings out of a huge parameter space, this approach ascertains that we might select empirically relevant parameter settings.

4.2 Simulation Setup

The results presented in this section are based on the simulation of three 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, Experiment5). The second DGP (DGP_2) is based on this example, but adds a second cointegration vector and extends the dynamic structure. Finally, the third DGP (DGP_3) is based on the estimation of a simple money demand system.

For each replication of the first two DGPs 300 observations have been generated from which the first 145 are eliminated, leaving a sample length

---

[18]Depending on the assumed rank r, the complexity of the lag structure and the length of the data series, the CPU time ranged from 3 to 7 minutes per independent optimization run on a Pentium 4 with 2.8 GHz using Matlab R13.

of $T = 155$. For DGP$_3$, the process was initialized with the historical values of the variables and samples of length $T = 200$ have been simulated for each replication. 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 modulus of nonzero reverse characteristic roots of the process[19] 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

---

[19]See Lütkepohl (1993) for a description. The roots are calculated using a Maple implementation with 100 digit precision.

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 modulus 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 which would correspond to a cointegration rank of two. This finding should be taken into account when analyzing the results obtained for this process.

DGP$_3$

Finally, DGP$_3$ has been obtained by fitting a VECM to the logarithms of M3 (lm), the nominal GDP (ly) and the GDP–Deflator (lp) for the period 1973.1 to 1989.4, i.e. restricted to West German data. All series have been detrended and seasonally adjusted by regressing them on a constant, a linear trend and seasonal dummies.[20] The cointegration space is spanned by the two vectors corresponding to long–run neutrality of money $(0, 1, -1)$ and no money illusion in the long–run $(1, -1, 0)$. Imposing these restrictions and reducing the model dynamics by a general–to–specific approach, the following parameters resulted:

---

[20]From an economic point of view, it appears reasonable to include an indicator of the cost of holding money, too. However, extending the analysis to a four dimensional system is left for future analysis.

$$\Pi Y_{t-1} = \begin{pmatrix} 0 & 0 \\ 0.20237 & -0.20453 \\ -0.069713 & 0.16597 \end{pmatrix} \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} \mathtt{lm}_{t-1} \\ \mathtt{ly}_{t-1} \\ \mathtt{lp}_{t-1} \end{pmatrix}$$

$$\Gamma_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -0.35885 & -0.55845 \\ 0 & 0 & 0.29401 \end{pmatrix} \quad \Gamma_2 = \begin{pmatrix} 0 & 0 & 0.28511 \\ 0.42088 & -0.24879 & -0.28601 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\Gamma_3 = \begin{pmatrix} 0.15953 & 0 & 0 \\ 0.38674 & 0 & -0.32608 \\ 0.026677 & 0 & 0 \end{pmatrix} \quad \Gamma_4 = \begin{pmatrix} 0.13814 & -0.087458 & -0.20802 \\ 0.19475 & 0.24129 & 0 \\ 0 & -0.090326 & 0.062985 \end{pmatrix}$$

$$\Gamma_5 = \begin{pmatrix} 0 & 0 & 0.20721 \\ 0 & 0 & -0.46694 \\ 0 & -0.082374 & 0 \end{pmatrix}$$

$$\Omega_\varepsilon = \begin{pmatrix} 0.0075836 & 0 & 0 \\ 0 & 0.011097 & 0 \\ 0 & 0 & 0.0056604 \end{pmatrix}$$

The modulus of nonzero reverse characteristic roots for this pseudo empirical process are 1, 0.94975, 0.82505, 0.82505,, 0.78820, 0.78820, 0.68524, 0.68524, 0.66087, 0.66087, 0.62045, 0.62045, 0.61235, 0.61235, 0.57458.

4.3   Results

The evaluation of the Monte Carlo results could be based on different properties. However, given that our main interest is on the effects of model selection on rank estimation, we focus on the estimated cointegration rank. For the models allowing for "holes", we also present information on the average size, i.e. the average relative frequency of including a zero coefficient, and the average power, i.e. the probability of including the nonzero coefficients. Alternative criteria comprise, e.g. the relative frequency of finding exactly the true DGP, which is considered to be an uninteresting statistic for real applications, the accuracy of impulse response analysis based on the selected model as compared to the true model or the relative forecasting performance.[21] 'As

---

[21]Brüggemann *et al.* (2003) use these characteristics for a comparison of model selection procedures for stationary VAR processes. They find that "model selection is especially useful in models with larger dimensions."

a measure of possible overfitting, which might be relevant in a forecasting setting, we also 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. Hendry and Krolzig (2003) suggest that values of this quotient close to or above 1 indicate that overfitting does not occur.

Results of the "take all up to the k–th lag" approach

Before turning to the results of our optimization algorithm, we first 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.[22]

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,[23] the share of correct identifications of the cointegration rank shrinks to 80.5% for the modified BIC and to 75% or 50.9%, respectively, when using Johansen's procedure with a nominal significance level of 1% and 5%, respectively.

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. Obviously, both the absolute

---

[22]It should be noted that for the trace test, standard critical values have been used. Of course, by taking into account the known structure of the DGPs, exact critical values could be obtained by means of simulation.

[23]This corresponds to the findings presented by Chao and Phillips (1999, p. 248).

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

and the relative performance of the methods changes drastically. 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. In particular, our results do not support results of other simulation studies where simpler lag structures allowed for the tentative conclusion that over-fitting might be less distorting than underfitting in a cointegration context.[24] Obviously, the second near unit root leads to the high rejection rates of the models with rank 2.

For our third DGP (DGP_3), the comparison of the different criteria is based on only 100 replications, which all have been initialized with the historical values of the detrended series. The results are summarized in 3.

Again, the modified BIC and Johansen's procedure appear more suitable for selecting the actual cointegration rank of two. However, both methods

---

[24]See, e.g. Cheung and Lai (1993) and Jacobson (1995).

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

do so while missing the high order dynamic dependencies embedded in the sparse lag structure of the DGP. By contrast, using the modified PIC points more often to higher lag orders. However, the actual lag order of five is identified only in 4% of all cases as compared to 1% for the other criteria. This deficit might be attributed to the sparse lag structure which implies that in order to capture the actual high order dynamics, the "take all up to the k–th lag" approach has to provide estimates for all entries in the matrices $\Gamma_1, \ldots, \Gamma_5$. As demonstrated for stationary VAR–processes in Winker (2000) allowing for holes in the lag structure might sensibly reduce this effect and improve the model identification of the dynamic part.

Summarizing the findings for the different criteria, at least for the three 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.

Table 3: Results for DGP$_3$ and all up to the k–th lag approach

Modified BIC							Modified PIC						
	Lags							Lags					
Rank	0	1	2	3	4	5	Rank	0	1	2	3	4	5
0	0	0	4	5	0	0	0	0	0	5	11	5	0
1	0	3	10	0	1	0	1	0	0	21	3	13	0
2	33	21	16	0	1	1	2	2	10	20	3	3	4
3	0	2	3	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	12	8	0	0	0	1	0	4	2	0	0	0
2	36	18	17	0	1	1	2	34	21	21	0	1	1
3	3	2	2	0	0	0	3	5	7	4	0	0	0

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.[25] Furthermore, due to constraints by available computer resources, 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, while for DGP$_3$ only 100 realizations with 200 observations are considered. For each realization, three different methods have been used to obtain an estimate of 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. The

---

[25]Detailed results on the identification of the dynamic structure are available on request.

cointegration rank identified by this method is defined by the minimum value of BICm obtained for the different rank conditions.

"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. The model resulting in the minimum value of BICm provides the rank estimate of this method.

"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. Afterwards, out of the d resulting models the one resulting in the smallest value of the modified BIC is selected.

Obviously, the method "known" cannot be used in practical applications, as the true lag structure will not be known. Nevertheless, it is used as a benchmark for our optimization approach ("holes") in order to make sure that by employing the optimization heuristic the identified model has a value of the criterion BICm which is smaller than or equal to the value of BICm for the actual lag structure.[26] 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 4 summarizes the results for the three methods applied to the three DGPs based on 200, 200, and 100 replications, respectively. 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. For practical applications only a comparison of the standard method of taking all lags up to a certain order ("all") with our optimization procedure is relevant. For DGP_1 the optimization procedure increases the frequency of finding the right cointegration rank from 80 to 89%.

---

[26]In fact, this goal is achieved for almost all runs of the optimization algorithm despite a small number of iterations used.

Table 4: Cointegration rank estimates (%)

Rank	Method		
	"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%
DGP$_3$ (100 replications)			
0	0.0%	10.0%	3.0%
1	19.0%	23.0%	28.0%
2	81.0%	67.0%	69.0%

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, in particular, we have to check wether increasing the number of observations tends to improve the results. 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 to cointegration at all, whereas the "holes" approach provides results quite similar to the ones obtained when the true DGP was known.

Finally, for the example DGP$_3$ constructed from a real data example, the assumption of a known lag structure would result in the best chance to find the real cointegration rank, while the difference between "all" and "holes" is smallest for this example, but still in favour of the optimization approach. We assume that the relative performance of the optimization approach will improve when a larger number of iterations can be performed

in the optimization step. This has been impossible for these first results due to constraints in computational resources.

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 5 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 $\hat{\Gamma}_1$, ..., $\hat{\Gamma}_5$. 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.

Table 5: 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
DGP$_3$ (100 replications)			
ν	21	15.12	13.06
cl	100%	27.8%	57.7%
wl	0%	25.1%	2.8%
q_Σ	1.00	1.65	1.08

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 of overfitting for the "take all up" approach and – to a smaller extent - for the "holes" method. For the other DGPs 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. For DGP_2, 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 4). Finally, for DGP_3, the optimization heuristic is much more successful in selecting the relevant lags and avoiding to include non relevant ones. Nevertheless, this advantage in the modelling of the dynamic part does not show up in a dramatic improvement in selecting the right cointegration rank. Again our first 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". Consequently, more research is needed to identify the features of DGPs affecting the performance of different model selection methods.

5   Conclusion

In this paper, we discuss the model selection issue in the context of non–stationary VAR–models with stationary cointegration relationships. Our reading of the literature suggests that the modelling of the dynamic part of these VEC–models is crucial for a correct rank identification. We compare different methods for model selection in a MC simulation including methods based on information criteria and a two–step procedure employing Johansen's testing strategy. Furthermore, we introduce a discrete optimization heuristic allowing for the selection of lag structures out of a much larger search

space than the usual "take all up to the k–th lag" approach. Again, a MC simulation is used to assess the relative performance of this algorithm.

Our findings support the view that the results of methods aiming at identifying the cointegration rank of a VECM depend heavily on the modelling of the dynamic structure. In contrast to earlier studies and in accordance with more recent findings, already a very small set of DGPs indicates that this effect might differ 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 as compared to the "take all up to the k–th lag" approach.

Given the small set of DGPs considered in this paper and the restriction to a single model selection procedure in the optimization context points directly towards future research. First, we have to apply our method to a much larger set of different DGPs in order to find out how robust our results are and which factors might be 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, we have to improve the performance of the estimation step in the optimization heuristic in order to allow for a larger number of iterations for a given problem in order to have more reliable results from the optimization method.

Despite our limited and preliminary results the tentative conclusion seems admitted that employing a more refined method for identifying the dynamic structure of a VECM might improve the performance in terms of rank order identification.

References

Ahking, F. W. (2002). Model mis–specification and Johansen's co–integration analysis: An application to the US money demand. *Journal of Macroeconomics* **24**, 51–66.

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.

Bahmani-Oskooee, M. and T. J. Brooks (2003). A new criteria for selecting the optimum lags in Johansen's cointegration technique. *Applied Economics* **35**, 875–880.

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.

Brüggemann, R., H.-M. Krolzig and H. Lütkepohl (2003). Comparison of model reduction methods for VAR processes. Technical Report 2003–W13. Economics Group, Nuffield College. University of Oxford.

Campos, J., D. F. Hendry and H.-M. Krolzig (2003). Consistent model selection by an automatic gets approach. *Oxford Bulletin of Economics & Statistics* **65**(s1), 803–820.

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

Cheung, Y.-.W and K. S. Lai (1993). Finite–sample sizes of Johansen's likelihood ratio test for cointegration. *Oxford Bulletin of Economics and Statistics* **55**, 313–328.

Fogel, D. (2001). *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*. 2nd ed.. IEEE Press. New York, NY.

Gatu, Cristian and Erricos J. Kontoghiorghes (2003). Parallel algorithms for computing all possible subset regression models using the QR decomposition. *Parallel Computing* p. forthcoming.

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.

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

Hendry, D. F. and H.-M. Krolzig (2003). Automatic model selection: A new instrument for social science. Technical report. Economics Department. Oxford University.

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. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* **12**, 231–254.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica* **59**(6), 1551–1580.

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

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

Johansen, S. (2002). A small sample correction for the test of cointegration rank in the vector autoregressive model. *Econometrica* **70**(5), 1929–1961.

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

Lütkepohl, H. (1993). *Introduction to Multiple Time Series Analysis*. 2 ed.. Springer. Berlin.

Maringer, Dietmar and Peter 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* (M. Doriga D. Corne and F. Glover, Eds.). pp. 219–234. MacGraw-Hill. London.

Omtzigt, Pieter (2002). Automatic identification and restriction of the cointegration space. Technical Report 2002/25. Faculty of Economics. University of Insubria, Varese.

Pötscher, B. M. (1991). Effects of model selection on inference. *Econometric Theory* **7**, 7–67.

Winker, Peter (1995). Identification of multivariate AR–models by threshold accepting. *Computational Statistics and Data Analysis* **20**(9), 295–307.

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

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