Identification of Vector Autoregressive Models with Granger and Stability Constraints

Bogdan Dumitrescu
Department of Automatic Control and Computers
University Politehnica of Bucharest
Bucharest, Romania
bogdan.dumitrescu@acse.pub.ro

Ciprian Doru Giurcăneanu
Department of Statistics
University of Auckland
Auckland, New Zealand
c.giurcaneanu@auckland.ac.nz

Yixia Ding
Department of Statistics
University of Auckland
Auckland, New Zealand
ydin024@aucklanduni.ac.nz

***Abstract*—In this work, we introduce an iterative method for the estimation of vector autoregressive (VAR) models with Granger and stability constraints. When the order of the model (p) and the Granger sparsity pattern (GSP) are not known, the newly proposed method is integrated in a two-stage approach. An information theoretic (IT) criterion is used in the first stage for selecting the value of p. In the second stage, a set of possible candidates for GSP are produced by applying the Wald test, and the best one is chosen with an IT criterion. In experiments with synthetic data, we demonstrate that our method yields more accurate forecasts than the state-of-art algorithm that is based on convex optimization and fits models which are guaranteed to be stable.**

***Index Terms*—Vector autoregressive models, Granger causality, stability, convex optimization, information theoretic criteria**

I. Introduction

Motivation: Granger causality has been widely employed in the analysis of multivariate signals. When the signal is represented by the vector-valued random variable $\boldsymbol{y}$, "a component y_b Granger-causes the component y_a if the use of the former improves the forecasts of the latter" [1]. The condition takes a particularly neat form in the case of the vector autoregressive (VAR) processes, where it reduces to finding the sparsity pattern for the matrix coefficients of the model. In recent years, the popularity of the sparse VAR-models has increased (see, for example, [2]–[4] and the references therein).

The methods that are routinely applied for estimating VAR-models with Granger constraints do not guarantee the stability of the solution. In [5], after briefly mentioning the previous attempts at fitting VAR-models with Granger and stability constraints, the authors formulate the estimation problem as a convex optimization problem that can be easily solved with CVX [6]. The major drawback is that the constraints for stability are too conservative and this can potentially reduce the performance when the model learned from the training data is used to predict the values in the test data.

Contributions and organization of the paper: After presenting in Sec. II the VAR-identification problem (i) with Granger constraints and (ii) with Granger and stability constraints, we introduce in Sec. III a novel estimation method. The new method is iterative and relies on the theoretical result in Theorem 1 that provides a sufficient stability condition, much more flexible than that from [5].

In [5], the Granger constraints are sparsity patterns produced by applying the Wald test [7], with various significance values. This does not lead to an automatic procedure for the identification of the model. Because we aim to eliminate the subjective decisions, we propose a two-stage approach. In the first stage, the order of the model is selected by an information theoretic (IT) criterion, and in the second stage the sparsity pattern is chosen from the set of the candidates yielded by the Wald test. In the second stage, the selection is again done by using an IT criterion. The complete procedure is described in Sec. IV.

The new method is compared with the state-of-the-art in Sec. V, where we conduct experiments with simulated data. The superiority of our proposal stems from the fact that it yields VAR-models that are guaranteed to be stable and, at the same time, produces accurate forecasts. Sec. VI concludes the paper.

Notation: We use bold letters for both vectors and matrices; $\mathbf{I}$ is the identity matrix of appropriate size. The symbol y_a denotes the a-th entry of an arbitrary vector $\boldsymbol{y}$. For a matrix $\boldsymbol{A}$, the symbol $(\boldsymbol{A})_{ab}$ denotes its entry that it is located at the intersection of the a-th row with the b-th column. The operator for transposition is $(\cdot)^T$. The symbols $||\cdot||_2$, $||\cdot||_\infty$ and $||\cdot||_F$ are used for the ℓ_2-norm, ℓ_∞-norm and the Frobenius-norm, respectively. The notation $\log(\cdot)$ is used for the natural logarithm. The symbol $F_{\chi^2_p}(\cdot)$ denotes the cumulative distribution function of a chi-squared random variable (with p degrees of freedom). For any integer $m > 1$, we employ the notation $1:m$ for the set $\{1, \ldots, m\}$.

II. Estimation problem

Consider a stationary and stable VAR process of order $p > 0$, for which the difference equation is [1]:

$$\boldsymbol{y}(n) = \boldsymbol{A}_1\boldsymbol{y}(n-1) + \ldots + \boldsymbol{A}_p\boldsymbol{y}(n-p) + \boldsymbol{\varepsilon}(n),\ n = 1:N, \tag{1}$$

where $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p$ are matrix coefficients of size $K \times K$. The vectors $\{\boldsymbol{\varepsilon}(n)\}_{n=1}^N$ are independently and identically distributed, and they are drawn from a K-variate Gaussian distribution with zero mean vector and non-singular covariance matrix $\boldsymbol{\Sigma}$. The vectors $\{\boldsymbol{y}(n)\}_{n=1-p}^0$ are assumed to be constant.

978-9-0827-9703-9/19/$31.00

A fundamental result says that, for some $a, b \in \mathcal{K}$, where $\mathcal{K} = \{1, \ldots, K\}$, y_b does *not* Granger-cause y_a if and only if $(\boldsymbol{A}_i)_{ab} = 0$ for $i = 1 : p$ [1]. In order to write the estimation problem as an optimization problem with constraints, we collect in $\mathcal{G}$ all the pairs (a, b) with the property that y_b does not Granger-cause y_a. We assume for the moment that $\mathcal{G}$ is known. Additionally, we concatenate the VAR-coefficients in a $K \times pK$ matrix

$$\boldsymbol{A} = \left[\begin{array}{cccc} \boldsymbol{A}_1 & \boldsymbol{A}_2 & \ldots & \boldsymbol{A}_p \end{array} \right], \tag{2}$$

and use the measured vectors for building the matrices

$$\boldsymbol{Y} = \left[\begin{array}{ccc} \boldsymbol{y}(p+1) & \boldsymbol{y}(p+2) & \ldots \quad \boldsymbol{y}(N) \end{array} \right], \tag{3}$$

$$\boldsymbol{H} = \left[\begin{array}{cccc} \boldsymbol{y}(p) & \boldsymbol{y}(p+1) & \ldots & \boldsymbol{y}(N-1) \\ \boldsymbol{y}(p-1) & \boldsymbol{y}(p) & \ldots & \boldsymbol{y}(N-2) \\ \vdots & \vdots & \ldots & \vdots \\ \boldsymbol{y}(1) & \boldsymbol{y}(2) & \ldots & \boldsymbol{y}(N-p) \end{array} \right]. \tag{4}$$

Hence, the VAR-identification problem subject to Granger constraints can be written as follows (see also the discussion in [5]):

$$\begin{array}{ll} \min_{\boldsymbol{A}} & \|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{H}\|_F^2 \\ \text{s.t.} & (\boldsymbol{A}_i)_{ab} = 0, \ i = 1 : p, \ \text{if } (a, b) \in \mathcal{G} \end{array} \tag{5}$$

This is a constrained least squares (LS) problem whose solution can be easily obtained, but the main drawback is that the estimated model is not guaranteed to be stable. For circumventing this difficulty, the authors of [5] proposed to estimate the matrix coefficients of the VAR-model by solving the problem

$$\begin{array}{ll} \min_{\boldsymbol{A}} & \|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{H}\|_F^2 \\ \text{s.t.} & (\boldsymbol{A}_i)_{ab} = 0, \ i = 1 : p, \ \text{if } (a, b) \in \mathcal{G} \\ & \|\boldsymbol{A}\|_\infty \leq 1 \end{array} \tag{6}$$

The second constraint comes from a sufficient stability condition associated with the state matrix of the pK-dimensional VAR(1) process corresponding to (1). This matrix is

$$\left[\begin{array}{ccccc} \boldsymbol{A}_1 & \boldsymbol{A}_2 & \ldots & \boldsymbol{A}_{p-1} & \boldsymbol{A}_p \\ \boldsymbol{I} & 0 & \ldots & 0 & 0 \\ 0 & \boldsymbol{I} & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & \boldsymbol{I} & 0 \end{array} \right]$$

and its first K rows are equal to $\boldsymbol{A}$. The model resulting from (6) is stable and satisfies the Granger constraints, but the accuracy of the estimation is affected by the fact that the stability constraint is too conservative. In the next section, we introduce a novel method for solving the problem that it is of interest for us. The key point is a theoretical result that allows to fit iteratively stable models to the data.

III. Iterative estimation

We denote

$$\boldsymbol{C}(\boldsymbol{z}) = \sum_{i=0}^{p} \boldsymbol{C}_i \boldsymbol{z}^{-i}, \tag{7}$$

a causal matrix polynomial with $K \times K$ coefficients. In our case $\boldsymbol{C}_0 = \boldsymbol{I}$, but the next result holds for any $\boldsymbol{C}_0$. Let us assume that $\boldsymbol{C}(\boldsymbol{z})$ is fixed and stable, in the sense that the VAR process (1) associated with it is stable. Let

$$\boldsymbol{D}(\boldsymbol{z}) = \sum_{i=1}^{p} \boldsymbol{D}_i \boldsymbol{z}^{-i} \tag{8}$$

be another matrix polynomial, now variable. Note that its free coefficient is zero.

Theorem 1: The polynomial $\boldsymbol{C} + \boldsymbol{D}$ is stable if the symmetric polynomial

$$\boldsymbol{R}(\boldsymbol{z}) = 2\boldsymbol{C}(\boldsymbol{z}^{-1})^T \boldsymbol{C}(\boldsymbol{z}) + \boldsymbol{C}(\boldsymbol{z}^{-1})^T \boldsymbol{D}(\boldsymbol{z}) + \boldsymbol{D}(\boldsymbol{z}^{-1})^T \boldsymbol{C}(\boldsymbol{z}) \tag{9}$$

is positive definite on the unit circle.

This is a sufficient stability condition, that describes a stability domain around $\boldsymbol{C}(\boldsymbol{z})$, which can be considered its center (although the domain typically has not a spherical, ellipsoidal or other symmetric shape). The domain is convex. The proof is a direct extension to matrix coefficient of the proof in [8] and is not given here for space reasons.

The condition from Theorem 1 cannot be implemented as such, but in the relaxed condition

$$\boldsymbol{R}(\boldsymbol{z}) - \lambda \boldsymbol{I} \text{ is sum-of-squares}, \tag{10}$$

where λ is a small positive constant that ensures some robustness to the stability condition and has also the role to improve the numerical conditioning. The domain described by (10) is also convex. The properties of trigonometric polynomials that are sum-of-squares are discussed in [8]; especially important here is their parameterization with linear matrix inequalities (LMI), which allows easy integration in convex optimization problems. So, given a stable polynomial $\boldsymbol{C}(\boldsymbol{z})$ with $\boldsymbol{C}_0 = \boldsymbol{I}$ and denoting

$$\boldsymbol{A}(\boldsymbol{z}) = \boldsymbol{I} - \sum_{i=1}^{p} \boldsymbol{A}_i \boldsymbol{z}^{-i}, \tag{11}$$

adding stability constraints to (5) leads to the problem

$$\begin{array}{ll} \min_{\boldsymbol{A},\boldsymbol{D}} & \|\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{H}\|_F^2 \\ \text{s.t.} & (\boldsymbol{A}_i)_{ab} = 0, \ i = 1 : p, \ \text{if } (a, b) \in \mathcal{G} \\ & \boldsymbol{A}(\boldsymbol{z}) = \boldsymbol{C}(\boldsymbol{z}) + \boldsymbol{D}(\boldsymbol{z}) \\ & \text{relation } (10) \text{ holds}, \\ & \text{with } \boldsymbol{R}(\boldsymbol{z}) \text{ defined by } (9) \end{array} \tag{12}$$

Using the parametrization of a sum-of-squares polynomial as an LMI, this problem can be transformed into a semidefinite programming problem, which can be solved efficiently and conveniently with CVX and Pos3Poly [9]. Note that the coefficients of $\boldsymbol{A}(\boldsymbol{z})$ and $\boldsymbol{D}(\boldsymbol{z})$ appear linearly in (12).

Problem (12) can be solved several times. The first time we take either $\boldsymbol{C}(\boldsymbol{z}) = \boldsymbol{I}$, which is the trivial initialization, or take

Algorithm 1 Two-Stage Approach

Stage1 [Select $\hat{p}$]:
  for all $p \in \{1, \ldots, p_{\max}\}$ **do**
    Fit VAR(p) to $\boldsymbol{Y}$ and compute $\mathrm{ITC}(\boldsymbol{Y}; p)$.
  end for
  $\hat{p} \leftarrow \arg \min_{1 \le p \le p_{\max}} \mathrm{ITC}(\boldsymbol{Y}; p)$.
Stage2 [Select $\widehat{\mathcal{G}}$]:
  for all $\alpha \in \mathcal{P}$ **do**
    $\mathrm{Th} \leftarrow F^{-1}_{\chi^2_{\hat{p}}}(1-\alpha)$
    for all $(a, b) \in \mathcal{K}^2$ **do**
      Compute Wald statistic W_{ab}.
      Get the Granger sparsity pattern:
      $(\mathcal{G}_\alpha)_{ab} \leftarrow$ 1 if $W_{ab} > \mathrm{Th}$
      $(\mathcal{G}_\alpha)_{ab} \leftarrow$ 0 otherwise
    end for
    Fit to the data a VAR-model with order $\hat{p}$ and sparsity pattern $\mathcal{G}_\alpha$. Enforce stability.
    Compute $\mathrm{ITC}(\boldsymbol{Y}; \hat{p}, \mathcal{G}_\alpha)$.
  end for
  $\widehat{\mathcal{G}} \leftarrow \arg \min_{\mathcal{G}_\alpha} \mathrm{ITC}(\boldsymbol{Y}; \hat{p}, \mathcal{G}_\alpha)$.

$\boldsymbol{C}(\boldsymbol{z})$ equal to the solution of the Granger problem (6), which has simpler (and more conservative) stability conditions. In the next iterations, we take $\boldsymbol{C}(\boldsymbol{z})$ equal to the solution $\boldsymbol{A}(\boldsymbol{z})$ of the previous problem (12). Since the stability domain is built around the previous solution, the value of the objective is guaranteed to decrease from an iteration to the next.

IV. Two-stage approach

All the methods presented above assume that the order p and the sparsity pattern $\mathcal{G}$ are known. Because this is not true in practical applications, we propose a two-stage approach. In the first stage, an estimate $\hat{p}$ of the order is obtained by using an IT criterion. In the second stage, the estimated model of order $\hat{p}$ is used for generating a set of possible candidates for $\mathcal{G}$. To this end, we apply the Wald test for the significance values $\mathcal{P} = \{0.001, 0.002, 0.005, 0.010, 0.020, 0.025, 0.050, 0.100, 0.200, 0.975, 0.995\}$. In this way, we get at most eleven different candidates from which we select the sparsity pattern $\widehat{\mathcal{G}}$ by using again an IT criterion, which is not necessarily the same as the one employed in the first stage. A complete description of this procedure can be found in Algorithm 1.

The use of the two-stage approach is advantageous because it reduces the computational complexity as the convex optimization problems like those in (5), (6), (12) are solved for only a limited number of candidates. The same idea was already applied in [10] in a different context, where in the second stage have been fitted models with a certain sparsity pattern for the inverse spectral density matrix (ISDM). Enforcing sparsity for the model coefficients is different from enforcing sparsity for ISDM. This is why we resort to the Wald test that was already applied in [5], where the interested reader can find all the details related to its calculation. However, the discussion in [5] is limited to the influence of the significance value on Type I and Type II errors. Here we propose a method for selecting $\widehat{\mathcal{G}}$. The use of IT criteria is important because the distributional properties of the Wald statistics on which the test is based are derived under asymptotic assumptions, and hence should be used with caution when the sample size is small or moderate [7]. A possible alternative is the cross-validation, but this leads to a higher computational burden.

In the first stage, we apply the following IT criteria for selecting the order of the model: (i) Schwarz Bayesian Criterion (SBC), which is the name used in time series literature for the Bayesian Information Criterion (BIC) [11]; (ii) Final Prediction Error (FPE) that was introduced by Akaike in [12]; (iii) Corrected Akaike Information Criterion (AICc) [13], a variant of the Akaike Information Criterion (AIC) [14] designed for small sample sizes; (iv) Renormalized Maximum Likelihood (RNML) that was derived and analyzed in [10], [15] by relying on the theory from [16].

In the second stage, we apply the same criteria and, as in [10], [17], [18], we alter them by replacing the number of parameters with the effective number of parameters (N_{ef}). In contrast to the aforementioned references, we do not count the non-zero entries of the ISDM, but the non-zero entries of the matrix coefficients and, because of that, $N_{\mathrm{ef}} = (K^2 - N_0)\hat{p}$, where N_0 stands for the number of zeros in the sparsity pattern $\widehat{\mathcal{G}}$. An important difference with respect to Stage 1 is that, in Stage 2, the number of potential candidates can be much larger. In the first stage, the number of the candidates is given by the maximum possible value of the order ($p_{\max}$), while in the second stage is 2^{K^2}. In order to take into consideration this aspect, we implement two *extended* criteria, namely the extended BIC (EBIC) and the extended RNML (ERNML). Note that EBIC is obtained by adding the term $4N_{\mathrm{ef}} \log K$ to BIC. Similarly, ERNML is computed by adding $2N_{\mathrm{ef}} \log K$ to RNML. More details about the two criteria can be found in [18], [19].

V. Numerical examples

We conduct two experiments, called Exp_1 and Exp_2. In Exp_1, we simulate data for a VAR-model with $K = 10$ and $p = 3$ (large number of components and small order). In Exp_2, $K = 3$ and $p = 10$ (small number of components and large order). In both experiments, the sparsity pattern $\mathcal{G}$ is selected such that the entries on the main diagonal are non-zero and at most 60% of the K^2 entries are equal to zero. All non-zero entries are drawn independently from a Gaussian distribution with zero-mean and standard deviation equal to 0.25.

The number of trials is N_{tr} = 200 as we generate 200 datasets for Exp_1 and Exp_2, respectively. Four different values of the sample size are considered for each dataset: $N \in \{200, 300, 400, 500\}$. For each trial, a new set of matrix coefficients are simulated as explained above. In all cases, the covariance matrix for the driven noise is $\boldsymbol{\Sigma} = 0.5\boldsymbol{I}$, and the noise samples are independent with respect to the randomly generated entries of the coefficients. All generated models are guaranteed to be stable. In the Matlab implementation, we use

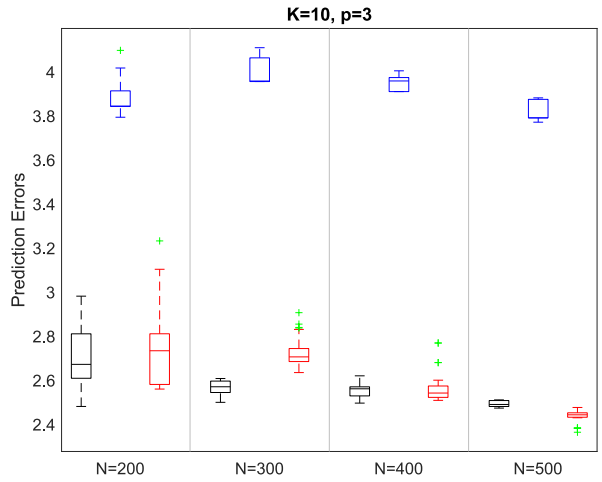

Fig. 1: Results of prediction in Exp_1. For each sample size (N), a different color is used for each boxplot, depending on the convex optimization problem solved in the training phase: the first boxplot (in black) is for (5), the second one (in blue) is for (6) and the last one (in red) is for (12).

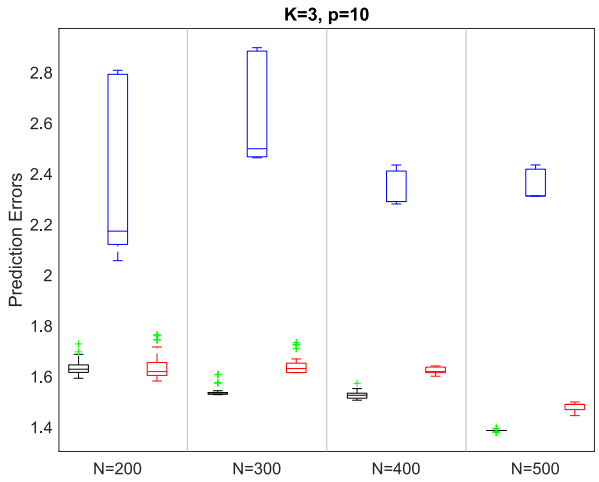

Fig. 2: Results of prediction in Exp_2: All graphical conventions are the same as in Fig. 1. Note that the range of the values on the vertical axis of the plot is different from the range in Fig. 1.

for simulating the data and for fitting the model in the first stage some functions from the package written for [20]. We take $p_{\max} = 15$ in Algorithm 1. The iterative estimation based on problem (12) is initialized with the solution of (6). We take $\lambda = 0.1$ in (10).

In each experiment, we use the model estimated from the first N measurements in order to forecast the values at the time moments $N+1, \ldots, N+h$, where $h = 100$. Note that the sample size for the training set varies, whereas the number of samples within the test set is constant. Let $\boldsymbol{e}^{\text{tr},n}_{\text{C}_1,\text{C}_2,\text{ME}}$ be the vector of the prediction errors at time moment n in trial tr; C_1 and C_2 denote the IT criteria applied in the first and in the second stage; ME stands for the estimation method used in training, which corresponds to one of the convex optimization problems in (5), (6), (12). For the sake of completeness, we also allow C_1 to be an oracle knowing the true order of the VAR-model. Similarly, C_2 can be an oracle that is knowledgeable about the sparsity pattern $\mathcal{G}$.

For each triple $(\text{C}_1, \text{C}_2, \text{ME})$ and for each trial tr, we compute $\text{MSE}^{\text{tr}}_{\text{C}_1,\text{C}_2,\text{ME}} = \dfrac{1}{K}\dfrac{1}{h}\displaystyle\sum_{n=N+1}^{N+h} \left\| \boldsymbol{e}^{\text{tr},n}_{\text{C}_1,\text{C}_2,\text{ME}} \right\|_2^2$. Then we calculate $median\left(\text{MSE}^{1}_{\text{C}_1,\text{C}_2,\text{ME}}, \ldots, \text{MSE}^{N_{\text{tr}}}_{\text{C}_1,\text{C}_2,\text{ME}}\right)$ and report the results in [21, Tables 25-32]. The same results are presented by using boxplots in Fig. 1 for Exp_1 and in Fig. 2 for Exp_2. Note that for a given sample size N we show in the boxplot corresponding to a method ME the results for all the pairs (C_1, C_2) of IT criteria. The most important conclusion that can be drawn from the two figures is that the forecast performance of the iterative estimation method that we propose is similar to that of the classical identification method based on LS with Granger constraints. The method from [5] produces the worse prediction results.

Both our method and the one from [5] have the advantage of yielding only stable models. According to the results reported in [21, Tables 17-24], the empirical probability that "LS with Granger constraints" produces an unstable model can be as large as 0.325 in Exp_2 when $N = 200$. For the same sample size, in Exp_1, the empirical probability is at most 0.185.

Another important aspect concerns the capability of recovering the true sparsity pattern. Let $\text{D}^{\text{tr}}_{\text{C}_1,\text{C}_2,\text{ME}}$ denote the number of the positions where $\widehat{\mathcal{G}}$ estimated in trial tr is different from $\mathcal{G}$. The significance of C_1, C_2 and ME is the same as before. We use this quantity for measuring the errors that occur in the estimation of the sparsity pattern. This means that we implicitly assume that Type I errors and Type II errors are equally important. For other applications, one may employ a dissimilarity measure that gives different weights to the two types of errors. Furthermore, we calculate $\dfrac{1}{K^2}\dfrac{1}{N_{\text{tr}}}\displaystyle\sum_{\text{tr}=1}^{N_{\text{tr}}} \text{D}^{\text{tr}}_{\text{C}_1,\text{C}_2,\text{ME}}$ and show the results in [21, Tables 9-16]. The same values are also displayed in Figs. 3-4 by using boxplots. It is evident from those figures that the accuracy in estimating the sparsity pattern is almost the same disregarding the convex optimization problem that it is solved. As expected, the accuracy improves when the sample size N raises. Interestingly enough, the outliers corresponding to modest estimation results in Exp_1 for $N < 400$ are because of the use of AICc in the selection of the sparsity pattern (see again [21, Tables 9-10]). This indicates that it is not recommended to apply AICc in the second stage of the algorithm when N is small.

In spite of the difficulties related to the use of AICc in the second stage, AICc performs well when it is employed to select the order of the model. According to [21, Tables 1-8], FPE is the best and AICc is the second best in correctly

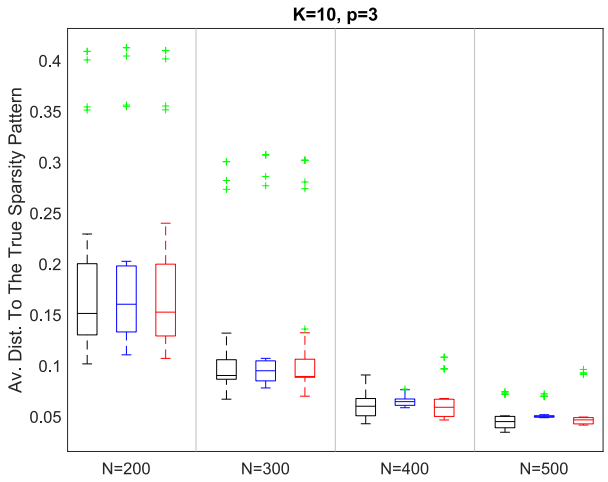

Fig. 3: Exp_1 - Average distances from the estimated sparsity pattern to the true one. For each sample size (N), a different color is used for each boxplot. The significance of the colors is the same as in Fig. 1.

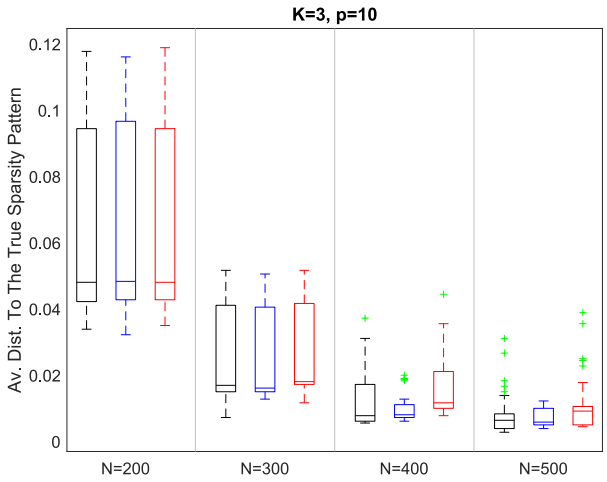

Fig. 4: Exp_2 - Average distances from the estimated sparsity pattern to the true one. All graphical conventions are the same as in Fig. 3.

estimating the order p. Both SBC and RNML tend to underestimate the order. This is surprising because RNML was found to be superior to other IT criteria in estimating the order of VAR-models having a certain sparsity pattern for the ISDM [10], [15].

VI. Final remarks

The novel estimation method introduced in this work is superior to "LS with Granger constraints" because it guarantees the stability of the fitted model. We have provided empirical evidence that it is also superior to the method from [5] because it yields better predictions. Another outcome of the experimental results is that SBC should be used with caution for selecting the order of sparse VAR-models when the sample size is small or moderate. The same applies to the use of AICc for choosing the Granger sparsity pattern. The experimental results can be reproduced by using the Matlab code available at https://www.stat.auckland.ac.nz/%7Ecgiu216/PUBLICATIONS.htm.

References

[1] H. Lutkepöhl, *New Introduction to Multiple Time Series Analysis*, John Wiley & Sons, 2005.

[2] R.A. Davis, P. Zang, and T. Zheng, "Sparse vector autoregressive modeling," *Journal of Computational and Graphical Statistics*, pp. 1–53, 2015.

[3] J. Songsiri, "Sparse autoregressive model estimation for learning Granger causality in time series," in *2013 IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, May 2013, pp. 3198–3202.

[4] E. Siggiridou and D. Kugiumtzis, "Granger causality in multivariate time series using a time-ordered restricted vector autoregressive model," *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1759–1773, 2016.

[5] N. Raksasri and J. Songsiri, "Guaranteed stability of autoregressive models with Granger causality learned from Wald tests," *Engineering Journal*, vol. 21, no. 6, pp. 23–36, 2017.

[6] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming," http://cvxr.com/cvx, Sept. 2010.

[7] A.C. Cameron and P.K. Trivedi, *Microeconometrics: methods and applications*, Cambridge University Press, 2005.

[8] B. Dumitrescu, *Positive trigonometric polynomials and signal processing applications*, Springer, 2007.

[9] B.C. Şicleru and B. Dumitrescu, "POS3POLY – a MATLAB Preprocessor for Optimization with Positive Polynomials," *Optimization and Engineering*, vol. 14, no. 2, pp. 251–273, 2013.

[10] S. Maanan, B. Dumitrescu, and C.D. Giurcăneanu, "Conditional independence graphs for multivariate autoregressive models by convex optimization: Efficient algorithms," *Signal Processing*, vol. 133, pp. 122–134, 2017.

[11] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, pp. 461–464, 1978.

[12] H. Akaike, "Autoregressive model fitting for control," *Annals of the Institute of Statistical Mathematics*, vol. 23, pp. 163–180, 1971.

[13] C.M. Hurvich and C.L. Tsai, "A corrected Akaike information criterion for vector autoregressive model selection," *Journal of Time Series Analysis*, vol. 14, pp. 271–279, 1993.

[14] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. AC-19, pp. 716–723, Dec. 1974.

[15] S. Maanan, B. Dumitrescu, and C. D. Giurcăneanu, "Renormalized maximum likelihood for multivariate autoregressive models," in *Proceedings of 2016 European Signal Processing Conference (EUSIPCO 2016)*, Budapest, Hungary, 2016, pp. 193–197.

[16] J. Rissanen, *Information and complexity in statistical modeling*, Springer-Verlag, 2007.

[17] J. Songsiri, J. Dahl, and L. Vandenberghe, "Graphical models of autoregressive processes," in *Convex Optimization in Signal Processing and Communications*, D.P. Palomar and Y.C. Eldar, Eds., pp. 89–116. Cambridge Univ. Press, 2010.

[18] S. Maanan, B. Dumitrescu, and C.D. Giurcăneanu, "Maximum entropy expectation-maximization algorithm for fitting latent-variable graphical models to multivariate time series," *Entropy*, vol. 20, no. 1, 2018.

[19] R. Foygel and M. Drton, "Extended Bayesian information criteria for Gaussian graphical models," in *Advances in Neural Information Processing Systems*, J.D. Lafferty, C.K.I. Williams, J. Shawe-Taylor, R.S. Zemel, and A. Culotta, Eds., 2010, vol. 23, pp. 604–612.

[20] A. Neumaier and T. Schneider, "Estimation of parameters and eigenmodes of multivariate autoregressive models," *ACM Trans. Math. Softw.*, vol. 27, pp. 27–57, 2001.

[21] B. Dumitrescu, C.D. Giurcăneanu, and Y. Ding, "Supplementary material to: Identification of vector autoregressive models with Granger and stability constraints," https://www.stat.auckland.ac.nz/%7Ecgiu216/PUBLICATIONS.htm, 2019.