Postprint of: Niedźwiecki M., Ciołek M., On Noncausal Identification of Nonstationary Multivariate Autoregressive Processes, IEEE TRANSACTIONS ON SIGNAL PROCESSING, Vol. 67, iss. 3 (2019), pp. 769-782, DOI: 10.1109/TSP.2018.2885480

On noncausal identification of nonstationary multivariate autoregressive processes

Maciej Niedźwiecki, *Senior Member, IEEE,* and Marcin Ciołek, *Member, IEEE,*

***Abstract*—The problem of identification of nonstationary multivariate autoregressive processes using noncausal local estimation schemes is considered and a new approach to joint selection of the model order and the estimation bandwidth is proposed. The new selection rule, based on evaluation of pseudoprediction errors, is compared with the previously proposed one, based on the modified Akaike's final prediction error criterion.**

I. Introduction

ESTIMATION bandwidth and model order are two important design parameters which should be carefully chosen to successfully identify (track) nonstationary autoregressive processes. In this paper we will show that both tasks can be accomplished jointly by means of minimizing the pseudoprediction error based statistic.

A. Problem statement

Consider the problem of identification of a nonstationary discrete-time m-dimensional signal $\{\mathbf{y}(t), t = \ldots, -1, 0, 1, \ldots\}$, $\mathbf{y}(t) = [y_1(t), \ldots, y_m(t)]^{\mathrm{T}}$, governed by the following time-varying vector autoregressive (VAR) model of order n

$$\mathbf{y}(t) = \sum_{i=1}^{n} \mathbf{A}_i(t)\mathbf{y}(t-i) + \mathbf{e}(t), \quad \mathrm{cov}[\mathbf{e}(t)] = \boldsymbol{\rho}(t) \qquad (1)$$

where $\mathbf{A}_1(t), \ldots, \mathbf{A}_n(t)$ denote time-varying $m \times m$ matrices of autoregressive coefficients, and $\{\mathbf{e}(t)\}$ denotes a sequence of zero-mean independent random variables with a time-varying covariance matrix $\boldsymbol{\rho}(t)$.

Due to their simplicity and good predictive capabilities, autoregressive models have found their way to a large number of practical applications in different fields such as biomedicine [1], [2], [3], [4], telecommunications [5], [6], and geophysics [7], [8], [9], among many others. Since parameters of autoregressive models have usually no physical significance, identification of such models is not a goal in itself – it is the means of solving practical problems, in the sense that the corresponding solutions depend explicitly on the estimates of model coefficients.

We will focus on noncausal local estimation techniques, which have recently gained strong support from the theory of locally stationary processes developed by Dahlhaus [10], [11], [12]. Noncausality means that local parameter estimates at a given time instant t are based on both "past" observations $\{\mathbf{y}(s), s \leq t\}$ and "future" observations $\{\mathbf{y}(s), s \geq t\}$ of the analyzed signal $\{\mathbf{y}(s)\}$. Even though not suitable for real-time applications, such as adaptive signal prediction, noncausal estimation schemes can be used in applications that are not time-critical in the sense that the model-based decisions can be postponed, or at least delayed by a certain number of sampling intervals. Parametric spectrum estimation, adaptive predictive coding and signal reconstruction are good examples of such applications. The main advantage of noncausal estimation compared with the causal one is due to significant reduction of the bias component of the mean squared parameter estimation error (MSE). Owing to this property, one can extend the size of the local analysis window which results also in reduction of the variance component of MSE.

This work was partially supported by the National Science Center under the agreement UMO-2015/17/B/ST7/03772. Computer simulations were carried out at the Academic Computer Centre in Gdańsk. Both authors are with the Gdańsk University of Technology, Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control, Narutowicza 11/12, Gdańsk, Poland, maciekn@eti.pg.edu.pl , marcin.ciolek@pg.edu.pl.

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the authors. The material includes Matlab codes and input data allowing one to reproduce simulation results shown in the paper. This material is 1.3 MB in size.

B. Motivation and state of the art

When identifying a nonstationary VAR process, one has to make two important decisions. First, one has to choose the most appropriate structure of the VAR model, i.e., decide how many and which autoregressive coefficients are significant enough to deserve estimation. Second, one should decide upon the estimation bandwidth, i.e., the frequency range in which parameter changes can be tracked "successfully" [13]. Estimation bandwidth is inversely proportional to the effective number of signal samples incorporated in the local estimation procedure, often called its estimation memory. Both choices are important and directly influence identification results.

The model structure should be sufficiently complex to comply with the spectral richness of the analyzed time series, as well as with its hidden inter-channel dependency pattern, but not overly complex to avoid estimation of "insignificant" model parameters.

The choice of the estimation bandwidth is equally important. It is known that the optimal bandwidth, trading off the bias and variance components of the mean squared parameter estimation error, depends on the degree of signal nonstationarity [10].

Finally, we note that when the analyzed signal is nonstationary both decisions should be made locally (since process characteristics and their rate of variation may change over time) and jointly (since they are linked via the principle of parsimony, according to which the number of estimated parameters should stay in a reasonable proportion to the effective number of samples used for their estimation [13]).

In the classical formulation, the problem of selection of the model structure is usually simplified and stated as the problem

of selection of the model order n. In the stationary case the problem of order selection can be solved by identifying models of different orders and then choosing the best fitting variant using one of the available order selection criteria which penalize high-order solutions – such as the Akaike's final prediction error (FPE) and information (AIC) criteria [14], [15], the Schwarz's Bayesian information criterion (BIC) [16], or the Rissanen's minimum description length criterion [17]. Under local stationarity assumptions some of these criteria can be extended to localized estimators – see e.g. [18].

Most of the more recent work on model structure selection is focused on the problem of identification of sparse VAR models, i.e., models in which some, if not most, of the entries of the matrices of autoregressive coefficients are set to zero and not estimated. Such a formulation is usually referred to as a subset autoregression [19]. However, unlike the classical "estimate, then select" approach described earlier, the sparse identification procedures achieve their goal in a direct way by minimizing the extended measure of fit which, in addition to the classical component such as the weighted sum of squared modeling errors, includes a model sparseness enforcing regularizer. When the regularizing term takes the form of the l_1 norm of the vector of estimated coefficients, such solution is known as the LASSO (least absolute shrinkage and selection operator) estimator [20], or – in a more general formulation – group LASSO estimator [21]. Although the majority of research on sparse identification is focused on the time-invariant case, some extensions to the time-varying case are also available – see e.g. [22].

The problem of selection of the estimation bandwidth is specific to identification of nonstationary processes and less explored than the problem of model structure selection. In the univariate case bandwidth scheduling can be based on the intersection of confidence intervals (ICI) rule proposed in [23] and further developed in [24]. However, no extension of this approach to multivariate processes seems to exist. In our recent paper [25] we proposed a solution to the bandwidth selection problem based on the modified FPE criterion. This solution will be further explained below.

C. The model stability issue

While in some applications such as short-term adaptive prediction or model-based Granger causality analysis [26], stability of a local VAR model is not a *sine qua non* applicability condition, in some other ones, such as predictive coding of signals (where the autoregressive model is used to generate signals based on the encoded model parameters), signal simulation (where the model is used to generate artificial data similar to the analyzed one) or signal interpolation [27], it is the obvious requirement. There are also applications, such as parametric spectrum estimation [28], where the local model instability does not seem to have a noticeable influence on the obtained estimation results but it makes the estimation procedure conceptually defective (note that whenever this happens the spectral estimates are evaluated in terms of parameters of an unstable model). Therefore, for practical reasons, one is usually interested in creating models that fulfill the following uniform stability condition

At all time instants t, all zeros of the characteristic polynomial

$$\mathcal{B}(z^{-1},t) = \det\left[\mathbf{I} - \sum_{i=1}^{n} \mathbf{A}_i(t) z^{-i}\right] \tag{2}$$

are uniformly bounded away from (remain strictly inside) the unit circle in the complex plane.

Remark

Condition (2) is not sufficient. To prove stability some additional, smoothness constraints must be imposed on parameter trajectories. In the univariate case strict analysis of conditions that guarantee uniform exponential stability of the time-varying autoregressive model can be found in [29]. When models are obtained via signal identification, (2) is usually a sufficient condition for their practical applicability.

In spite of its practical importance, the model stability issue is surprisingly absent from the statistical literature on identification of VAR processes. It is known that only a few existing estimation schemes (such as the Yule-Walker type algorithms or normalized lattice algorithms [30]) guarantee that the condition (2) is met by the resultant models. In particular, there are no stability guarantees if the VAR model is sought in the unconstrained sparse form (unless a specific sparsity structure is enforced, as in [31]). In the univariate case model stability can be reinstated by projecting unstable poles of the forming filter into the stability region, and by modifying the variance of the driving noise accordingly. However, no such procedure seems to exist in the multivariate case.

D. Contribution and relation to the previous work

The paper presents two asymptotically equivalent local identification methods – noncausal variants of the weighted least squares estimators and stability-preserving weighted Yule-Walker estimators – with joint estimation bandwidth and model order adaptation. As shown in [25], the problem of joint bandwidth-order optimization can be solved by means of applying the modified Akaike's FPE criterion. In the same paper we have demonstrated that the cross-validation approach based on comparison of local leave-one-out signal interpolation errors is not suitable for the purpose of bandwidth optimization. In the current contribution we will suggest replacement of interpolation errors with pseudoprediction errors. We will show that if the cross-validation approach is reformulated in this way, the obtained results are comparable with those provided by the FPE approach. The paper extends results obtained earlier for univariate processes, presented in [32].

E. Glossary of the most frequently used abbreviations and symbols

FPE	final prediction error
NWLS	noncausal weighted least squares
NWYW	noncausal weighted Yule-Walker
PPE	pseudoprediction error

k	bandwidth parameter
K	number of bandwidth variants
$\mathcal{K}$	the set of bandwidth parameters
L_k	effective window width
m	signal dimension
M	half-width of the decision window
n	model order
N	maximum model order
$\mathcal{N}$	the set of model orders
N_k	equivalent window width
t	discrete time
$w_k(i)$	weighting sequence
$v_k(i)$	data taper
$\widehat{(\cdot)}$	NWLS estimate
$\widetilde{(\cdot)}$	NWYW estimate

II. Shorthand notation

Note that the l-th component of the multivariate signal $\{\mathbf{y}(t)\}$ can be written down in the form

$$y_l(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\boldsymbol{\theta}_l(t) + e_l(t), \quad l = 1, \ldots, m \tag{3}$$

where $\boldsymbol{\varphi}(t) = [\mathbf{y}^{\mathrm{T}}(t-1), \ldots, \mathbf{y}^{\mathrm{T}}(t-n)]^{\mathrm{T}}$ denotes the $mn \times 1$ regression vector and $\boldsymbol{\theta}_l(t) = [a_{l1,1}(t), \ldots, a_{lm,1}(t), \ldots, a_{l1,n}(t), \ldots, a_{lm,n}(t)]^{\mathrm{T}}$ denotes the $mn \times 1$ vector of parameters characterizing the l-th signal "channel". Note also that equation (1) can be rewritten in the following equivalent shorthand forms

$$\begin{aligned} \mathbf{y}(t) &= [\mathbf{A}_1(t) \ldots \mathbf{A}_n(t)]\boldsymbol{\varphi}(t) + \mathbf{e}(t) \\ &= \begin{bmatrix} \boldsymbol{\theta}_1^{\mathrm{T}}(t) \\ \vdots \\ \boldsymbol{\theta}_m^{\mathrm{T}}(t) \end{bmatrix} \boldsymbol{\varphi}(t) + \mathbf{e}(t) = \boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{\theta}(t) + \mathbf{e}(t) \end{aligned} \tag{4}$$

where $\boldsymbol{\Phi}(t)$ is the $m^2n \times m$ matrix of the form

$$\boldsymbol{\Phi}(t) = \mathbf{I}_m \otimes \boldsymbol{\varphi}(t) = \begin{bmatrix} \boldsymbol{\varphi}(t) & & \mathbf{O} \\ & \ddots & \\ \mathbf{O} & & \boldsymbol{\varphi}(t) \end{bmatrix}$$

($\otimes$ denotes Kronecker product of the corresponding vectors/matrices) and

$$\boldsymbol{\theta}(t) = [\boldsymbol{\theta}_1^{\mathrm{T}}(t) \ldots \boldsymbol{\theta}_m^{\mathrm{T}}(t)]^{\mathrm{T}} = \mathrm{vec}\{[\mathbf{A}_1(t) \ldots \mathbf{A}_n(t)]^{\mathrm{T}}\}$$

is the $m^2n \times 1$ vector made up of all autoregressive coefficients.

III. Noncausal weighted least squares estimators

A. Estimation scheme

The noncausal weighted least squares (NWLS) estimate of $\boldsymbol{\theta}(t)$ is given in the form

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_k(t) &= \arg\min_{\boldsymbol{\theta}} \sum_{i=-k}^{k} v_k(i) \parallel \mathbf{y}(t+i) - \boldsymbol{\Phi}^{\mathrm{T}}(t+i)\boldsymbol{\theta} \parallel^2 \\ &= \mathbf{G}_k^{-1}(t)\mathbf{r}_k(t) \end{aligned} \tag{5}$$

where

$$\begin{aligned} \mathbf{G}_k(t) &= \sum_{i=-k}^{k} v_k(i)\boldsymbol{\Phi}(t+i)\boldsymbol{\Phi}^{\mathrm{T}}(t+i) \\ \mathbf{r}_k(t) &= \sum_{i=-k}^{k} v_k(i)\boldsymbol{\Phi}(t+i)\mathbf{y}(t+i) \end{aligned} \tag{6}$$

and $\{v_k(i), i = -k, \ldots, k\}$, $v_k(0) = 1$, denotes a nonnegative, symmetric bell-shaped window of width $2k+1$ used for localization purposes – as a result the estimates evaluated at the instant t depend more heavily on the most recent measurements than on the measurements collected in the remote past and future. We will further assume that $v_k(i) = f(i/k)$, where $f(\cdot)$ is the analog window generating function defined on the interval $[-1, 1]$. Owing to the fact that the regression matrix $\mathbf{G}_k(t)$ is block diagonal with identical blocks

$$\mathbf{G}_k(t) = \mathbf{I}_m \otimes \mathbf{R}_k(t) = \begin{bmatrix} \mathbf{R}_k(t) & & \mathbf{O} \\ & \ddots & \\ \mathbf{O} & & \mathbf{R}_k(t) \end{bmatrix}$$

where

$$\mathbf{R}_k(t) = \sum_{i=-k}^{k} v_k(i)\boldsymbol{\varphi}(t+i)\boldsymbol{\varphi}^{\mathrm{T}}(t+i),$$

the NWLS estimate can be expressed and evaluated in the decomposed form as follows

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_k(t) &= [\widehat{\boldsymbol{\theta}}_{1|k}^{\mathrm{T}}(t), \ldots, \widehat{\boldsymbol{\theta}}_{m|k}^{\mathrm{T}}(t)]^{\mathrm{T}} \\ \widehat{\boldsymbol{\theta}}_{l|k}(t) &= \arg\min_{\boldsymbol{\theta}_l} \sum_{i=-k}^{k} v_k(i)[y_l(t+i) - \boldsymbol{\varphi}^{\mathrm{T}}(t+i)\boldsymbol{\theta}_l]^2 \\ &= \mathbf{R}_k^{-1}(t)\mathbf{r}_{l|k}(t), \quad l = 1, \ldots, m \end{aligned} \tag{7}$$

where

$$\mathbf{r}_{l|k}(t) = \sum_{i=-k}^{k} v_k(i)y_l(t+i)\boldsymbol{\varphi}(t+i).$$

This means that the local estimates (7), which can be obtained by considering each channel separately, coincide with the global estimate (5) obtained by considering all channels jointly. This property, owned to the fact that all channels share the same regression vector $\boldsymbol{\varphi}(t)$, was noticed for the first time in [33] (for least squares estimators).

The NWLS estimate of $\boldsymbol{\rho}(t)$ can be obtained from

$$\begin{aligned} \widehat{\boldsymbol{\rho}}_k(t) &= \frac{1}{L_k} \sum_{i=-k}^{k} v_k(i)[\mathbf{y}(t+i) - \boldsymbol{\Phi}^{\mathrm{T}}(t+i)\widehat{\boldsymbol{\theta}}_k(t)] \times \\ &\qquad \times [\mathbf{y}(t+i) - \boldsymbol{\Phi}^{\mathrm{T}}(t+i)\widehat{\boldsymbol{\theta}}_k(t)]^{\mathrm{T}} \\ &= \frac{1}{L_k}\Big\{\mathbf{S}_k(t) - [\widehat{\boldsymbol{\theta}}_{1|k}(t) \ldots \widehat{\boldsymbol{\theta}}_{m|k}(t)]^{\mathrm{T}} \times \\ &\qquad \times [\mathbf{r}_{1|k}(t) \ldots \mathbf{r}_{m|k}(t)]\Big\} \end{aligned} \tag{8}$$

where

$$\mathbf{S}_k(t) = \sum_{i=-k}^{k} v_k(i)\mathbf{y}(t+i)\mathbf{y}^{\mathrm{T}}(t+i)$$

and the quantity

$$L_k = \sum_{i=-k}^{k} v_k(i) \cong k \int_{-1}^{1} f(x)dx$$

denotes the effective window width.

B. Recursive computability

Even though, in principle, any bell-shaped function $f(\cdot)$ can be used for window generation purposes, from the practical viewpoint the most interesting alternatives are those that allow for recursive computation of $\widehat{\boldsymbol{\theta}}_k(t)$ and $\widehat{\boldsymbol{\rho}}_k(t)$. We will show that the Hann (raised cosine) window

$$v_k(i) = \frac{1}{2}\left[1 + \cos\frac{\pi i}{k+1}\right] = \frac{1}{2}\left\{1 + \mathrm{Re}\left[e^{j\frac{\pi i}{k+1}}\right]\right\} \qquad (9)$$

meets this requirement. To see that this is the case let

$$\mathbf{U}_k(t) = \sum_{i=-k}^{k} \boldsymbol{\varphi}(t+i)\boldsymbol{\varphi}^{\mathrm{T}}(t+i)$$

$$\mathbf{W}_k(t) = \sum_{i=-k}^{k} e^{j\frac{\pi i}{k+1}} \boldsymbol{\varphi}(t+i)\boldsymbol{\varphi}^{\mathrm{T}}(t+i)\ .$$

Observe that both quantities defined above are recursively computable

$$\begin{aligned}\mathbf{U}_k(t+1) &= \mathbf{U}_k(t) - \boldsymbol{\varphi}(t-k)\boldsymbol{\varphi}^{\mathrm{T}}(t-k) \\ &\quad + \boldsymbol{\varphi}(t+k+1)\boldsymbol{\varphi}^{\mathrm{T}}(t+k+1) \\ \mathbf{W}_k(t+1) &= e^{-j\frac{\pi}{k+1}}\mathbf{W}_k(t) + \boldsymbol{\varphi}(t-k)\boldsymbol{\varphi}^{\mathrm{T}}(t-k) \\ &\quad + e^{j\frac{\pi k}{k+1}}\boldsymbol{\varphi}(t+k+1)\boldsymbol{\varphi}^{\mathrm{T}}(t+k+1)\end{aligned}$$

and that

$$\mathbf{R}_k(t) = \frac{1}{2}\,\mathbf{U}_k(t) + \frac{1}{2}\,\mathrm{Re}[\mathbf{W}_k(t)]\ .$$

The quantities $\mathbf{r}_{1|k}(t), \ldots, \mathbf{r}_{m|k}(t)$ and $\mathbf{S}_k(t)$ can be computed recursively in an analogous way.

Remark 1

Since the sliding-window subtract-add recursive algorithms for updating the quantities $\mathbf{R}_k(t)$ and $\mathbf{S}_k(t)$ are not exponentially stable but only marginally stable, they diverge at a slow (linear) rate as the number of steps becomes large. For this reason both quantities should be periodically reset by direct (nonrecursive) computation.

C. Estimation bandwidth scheduling

Estimation bandwidth, i.e., the frequency range in which signal parameters can be tracked "successfully" [13] depends on the window size. For small values of k, NWLS estimators quickly adapt to parameter changes (small estimation bias) at the cost of increased variability (large estimation variance). The opposite happens for large values of k. Therefore, to guarantee good tracking results, the bandwidth parameter k should be adjusted so as to trade off the bias and variance components of the mean squared parameter estimation error.

For unknown and/or time-varying rate of signal nonstationarity such bandwidth optimization can be carried out using parallel estimation schemes. In this approach several estimation algorithms, equipped with different bandwidth settings $k \in \mathcal{K} = \{k_1, \ldots, k_K\}$, are simultaneously run and compared. At each time instant only one of the competing algorithms is selected, i.e., the estimated parameter and variance trajectories have the form $\widehat{\boldsymbol{\theta}}_{\widehat{k}(t)}(t)$ and $\widehat{\boldsymbol{\rho}}_{\widehat{k}(t)}(t)$, respectively, where

$$\widehat{k}(t) = \arg\min_{k\in\mathcal{K}} J_k(t) \qquad (10)$$

and $J_k(t)$ denotes the local decision statistic. For comments on the recommended choice of $\mathcal{K}$ see Remark 3 at the end of Section IV.B.

As shown in our recent paper [25], Section 4.2, the problem of bandwidth selection can be solved using the modified Akaike's final prediction error criterion. Taking this approach, one can adopt

$$J_k(t) = \mathrm{FPE}_k(t) = \left[\frac{1+\frac{mn}{N_k}}{1-\frac{mn}{N_k}}\right]^m \det\widehat{\boldsymbol{\rho}}_k(t) \qquad (11)$$

where

$$N_k = \frac{\left[\sum_{i=-k}^{k} v_k(i)\right]^2}{\sum_{i=-k}^{k} v_k^2(i)} \cong k\,\frac{\left[\int_{-1}^{1} f(x)dx\right]^2}{\int_{-1}^{1} f^2(x)dx} \qquad (12)$$

denotes the so-called equivalent window width.

In the same paper, we have demonstrated that the cross-validation approach based on evaluation of local leave-one-out signal interpolation errors (see Section 4.1 in [25]) is not suitable for the purpose of estimation bandwidth selection. In the current contribution we will show that if interpolation errors are replaced with pseudoprediction errors, one obtains a well-behaved bandwidth selection criterion. Following [34], we will introduce the notion of a holey NWLS estimator

$$\widehat{\boldsymbol{\theta}}_k^{\circ}(t) = \arg\min_{\boldsymbol{\theta}} \sum_{\substack{i=-k \\ i\neq 0}}^{k} v_k(i) \parallel \mathbf{y}(t+i) - \boldsymbol{\Phi}^{\mathrm{T}}(t+i)\boldsymbol{\theta} \parallel^2 . \qquad (13)$$

In the original formulation, developed for the purpose of identification of nonstationary finite impulse response (FIR) systems, the holey estimator of $\boldsymbol{\theta}(t)$ completely eliminates from the estimation process the measurement $\mathbf{y}(t)$ collected at the instant t. Note that while this is true in the FIR case, where the regression vector $\boldsymbol{\varphi}(t)$ is made up of past input measurements, in the autoregressive case exclusion of the term $v_k(0) \parallel \mathbf{y}(t) - \boldsymbol{\Phi}^{\mathrm{T}}(t)\boldsymbol{\theta} \parallel^2$ from the sum in (13) does not guarantee independence[1] of the estimate $\widehat{\boldsymbol{\theta}}_k^{\circ}(t)$ of $\mathbf{y}(t)$, simply because $\mathbf{y}(t)$ is a component of the regression vectors $\boldsymbol{\varphi}(t+1), \ldots, \boldsymbol{\varphi}(t+n)$ [and hence a component of the matrices $\boldsymbol{\Phi}(t+1), \ldots, \boldsymbol{\Phi}(t+n)$] included in (13).

Based on $\widehat{\boldsymbol{\theta}}_k^{\circ}(t)$, one can compute a pseudoprediction $\widehat{\mathbf{y}}_k^{\circ}(t)$ of $\mathbf{y}(t)$, and the corresponding pseudoprediction error (PPE)

$$\begin{aligned}\boldsymbol{\varepsilon}_k^{\circ}(t) &= [\varepsilon_{1|k}^{\circ}(t), \ldots, \varepsilon_{m|k}^{\circ}(t)]^{\mathrm{T}} = \mathbf{y}(t) - \boldsymbol{\Phi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}_k^{\circ}(t) \\ &= \mathbf{y}(t) - \sum_{i=1}^{n} \widehat{\mathbf{A}}_i^{\circ}(t)\mathbf{y}(t-i). \qquad (14)\end{aligned}$$

[1] Hereafter the term "independent" is used in its deterministic sense.

The name "pseudoprediction" refers to the fact that the vector of parameter estimates $\widehat{\boldsymbol{\theta}}_k^\circ(t)$ depends on $\mathbf{y}(t)$, which is the predicted quantity.

The proposed decision statistics, based on evaluation of the local pseudoprediction errors, take the form

$$J_k(t) = \mathrm{PPE}_k(t) = \det\left\{\sum_{i=-M}^{M} \boldsymbol{\varepsilon}_k^\circ(t+i)[\boldsymbol{\varepsilon}_k^\circ(t+i)]^{\mathrm{T}}\right\} \tag{15}$$

or

$$\begin{aligned} J_k(t) = \mathrm{PPE}_k^*(t) &= \mathrm{tr}\left\{\sum_{i=-M}^{M} \boldsymbol{\varepsilon}_k^\circ(t+i)[\boldsymbol{\varepsilon}_k^\circ(t+i)]^{\mathrm{T}}\right\} \\ &= \sum_{i=-M}^{M} \| \boldsymbol{\varepsilon}_k^\circ(t+i) \|^2 \end{aligned} \tag{16}$$

where $(2M+1) > m$ is the width of the local decision window $[t-M, t+M]$ centered at t. The value of M is a user-dependent "knob". It should be sufficiently large to avoid erratic behavior of the decision rule, but not overly large to guarantee its adaptivity. In many practical situations $M \in [10, 25]$ is a good choice. Generally, the results of selection are pretty insensitive to the adopted value of M.

We will show that the PPE/PPE* statistics can be evaluated without implementing the holey estimator.

Proposition 1

It holds that

$$\boldsymbol{\varepsilon}_k^\circ(t) = \frac{\boldsymbol{\varepsilon}_k(t)}{1 - b_k(t)} \tag{17}$$

where

$$\begin{aligned} \boldsymbol{\varepsilon}_k(t) &= \mathbf{y}(t) - \boldsymbol{\Phi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}_k(t) \\ b_k(t) &= \boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{R}_k^{-1}(t)\boldsymbol{\varphi}(t). \end{aligned}$$

Proof

It is straightforward to show that

$$\widehat{\boldsymbol{\theta}}_k^\circ(t) = \begin{bmatrix} \widehat{\boldsymbol{\theta}}_{1|k}^\circ(t) \\ \vdots \\ \widehat{\boldsymbol{\theta}}_{m|k}^\circ(t) \end{bmatrix}$$

where

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_{l|k}^\circ(t) &= \arg\min_{\boldsymbol{\theta}_l} \sum_{\substack{i=-k \\ i\neq 0}}^{k} v_k(i)[y_l(t+i) - \boldsymbol{\varphi}^{\mathrm{T}}(t+i)\boldsymbol{\theta}_l]^2 \\ &= [\mathbf{R}_k^\circ(t)]^{-1}\mathbf{r}_{l|k}^\circ(t) \end{aligned}$$

and [since $v_k(0) = 1$]

$$\begin{aligned} \mathbf{R}_k^\circ(t) &= \mathbf{R}_k(t) - \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t) \\ \mathbf{r}_{l|k}^\circ(t) &= \mathbf{r}_{l|k}(t) - y_l(t)\boldsymbol{\varphi}(t). \end{aligned}$$

Exploiting the fact that $\mathbf{r}_{l|k}(t) = \mathbf{R}_k(t)\widehat{\boldsymbol{\theta}}_{l|k}(t)$, and using the matrix inversion lemma [35], one arrives at the expression

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_{l|k}^\circ(t) = \left[\mathbf{R}_k^{-1}(t) + \frac{\mathbf{R}_k^{-1}(t)\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{R}_k^{-1}(t)}{1 - \boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{R}_k^{-1}(t)\boldsymbol{\varphi}(t)}\right] \times \\ \times [\mathbf{R}_k(t)\widehat{\boldsymbol{\theta}}_k(t) - y_l(t)\boldsymbol{\varphi}(t)]. \end{aligned}$$

Substituting this expression into

$$\varepsilon_{l|k}^\circ(t) = y_l(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}_{l|k}^\circ(t)$$

one obtains

$$\begin{aligned} \varepsilon_{l|k}^\circ(t) &= \frac{y_l(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\widehat{\boldsymbol{\theta}}_{l|k}(t)}{1 - \boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{R}_k^{-1}(t)\boldsymbol{\varphi}(t)} \\ & l = 1, \ldots, m \end{aligned}$$

which is identical with (17).

IV. Noncausal weighted Yule-Walker estimators

NWLS estimators do not guarantee uniform stability of the resultant models. The stability problem can be overcome if estimation is carried out using noncausal weighted Yule-Walker (NWYW) estimators.

A. NWLS embedding

We will show that the NWYW estimators can be reinterpreted as local least squares estimates obtained for the tapered data sequence

$$\mathbf{y}_k(t+i|t) = w_k(i)\mathbf{y}(t+i), \quad i \in [-k, k]$$

where $\{w_k(i), i = -k, \ldots, k\}$, $w_k(0) = 1$, is a nonnegative, symmetric, bell-shaped data taper. Similarly as in the case of NWLS estimators, we will assume that $w_k(i) = g(i/k)$, where $g(\cdot)$ is the continuous time taper generation function defined on $[-1, 1]$. Suppose that the data sequence $\mathbf{y}(t-k), \ldots, \mathbf{y}(t+k)$ is extended with n zero samples at its beginning and at its end, and that the data taper $w_k(t-k), \ldots, w_k(t+k)$ is extended likewise. Note that under such extensions it holds that $\mathbf{y}_k(t+i|t) = 0$ for $i \in [-k-n, -k-1]$ and $i \in [k+1, k+n]$. Finally, let $\boldsymbol{\varphi}_k(t+i|t) = [\mathbf{y}_k^{\mathrm{T}}(t+i-1|t), \ldots, \mathbf{y}_k^{\mathrm{T}}(t+i-n|t)]^{\mathrm{T}}$.

Consider the following local least squares estimates of $\boldsymbol{\theta}_1(t), \ldots, \boldsymbol{\theta}_m(t)$

$$\begin{aligned} \widetilde{\boldsymbol{\theta}}_{l|k}(t) &= \arg\min_{\boldsymbol{\theta}_l} \sum_{i=-k}^{k+n} [y_{l|k}(t+i|t) - \boldsymbol{\varphi}_k^{\mathrm{T}}(t+i|t)\boldsymbol{\theta}_l]^2 \\ &= \mathbf{Q}_k^{-1}(t)\mathbf{q}_{l|k}(t), \quad l = 1, \ldots, m \end{aligned} \tag{18}$$

where

$$\begin{aligned} \mathbf{Q}_k(t) &= \sum_{i=-k}^{k+n} \boldsymbol{\varphi}_k(t+i|t)\boldsymbol{\varphi}_k^{\mathrm{T}}(t+i|t) \\ \mathbf{q}_{l|k}(t) &= \sum_{i=-k}^{k+n} y_{l|k}(t+i|t)\boldsymbol{\varphi}_k(t+i|t). \end{aligned}$$

First of all, note that $\mathbf{Q}_k(t)$ is a block Toeplitz matrix of the form

$$\mathbf{Q}_k(t) = \begin{bmatrix} \mathbf{P}_{0|k}(t) & \mathbf{P}_{1|k}(t) & \dots & \mathbf{P}_{n-1|k}(t) \\ \mathbf{P}^{\mathrm{T}}_{1|k}(t) & \mathbf{P}_{0|k}(t) & & \vdots \\ \vdots & & \ddots & \mathbf{P}_{1|k}(t) \\ \mathbf{P}^{\mathrm{T}}_{n-1|k}(t) & \dots & \mathbf{P}^{\mathrm{T}}_{1|k}(t) & \mathbf{P}_{0|k}(t) \end{bmatrix}$$

where

$$\mathbf{P}_{l|k}(t) = \sum_{i=-k+l}^{k} \mathbf{y}_k(t+i|t)\mathbf{y}_k^{\mathrm{T}}(t+i-l|t).$$

Note also that

$$\begin{bmatrix} \mathbf{q}^{\mathrm{T}}_{1|k}(t) \\ \vdots \\ \mathbf{q}^{\mathrm{T}}_{m|k}(t) \end{bmatrix} = \sum_{i=-k}^{k+n} \mathbf{y}_k(t+i|t)\boldsymbol{\varphi}_k^{\mathrm{T}}(t+i|t) = [\mathbf{P}_{1|k}(t) \dots \mathbf{P}_{n|k}(t)]$$

and

$$\begin{bmatrix} \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{1|k}(t) \\ \vdots \\ \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{m|k}(t) \end{bmatrix} = [\widetilde{A}_{1|k}(t) \dots \widetilde{A}_{n|k}(t)]$$

Using (18) and exploiting the fact that $\mathbf{Q}_k(t) = \mathbf{Q}_k^{\mathrm{T}}(t)$, one obtains

$$\begin{bmatrix} \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{1|k}(t) \\ \vdots \\ \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{m|k}(t) \end{bmatrix} \mathbf{Q}_k(t) = \begin{bmatrix} \mathbf{q}^{\mathrm{T}}_{1|k}(t) \\ \vdots \\ \mathbf{q}^{\mathrm{T}}_{m|k}(t) \end{bmatrix}$$

which, after combining all previous results, can be rewritten in the form

$$[\widetilde{A}_{1|k}(t) \dots \widetilde{A}_{n|k}(t)]\mathbf{Q}_k(t) = [\mathbf{P}_{1|k}(t) \dots \mathbf{P}_{n|k}(t)]. \quad (19)$$

Using a similar technique, one can show that

$$\begin{aligned} \widetilde{\boldsymbol{\rho}}_k(t) &= \frac{1}{L_k} \sum_{i=-k}^{k+n} [\mathbf{y}_k(t+i|t) - \boldsymbol{\Phi}_k^{\mathrm{T}}(t+i|t)\widetilde{\boldsymbol{\theta}}_k(t)] \times \\ &\qquad \times [\mathbf{y}_k(t+i|t) - \boldsymbol{\Phi}_k^{\mathrm{T}}(t+i|t)\widetilde{\boldsymbol{\theta}}_k(t)]^{\mathrm{T}} \\ &= \frac{1}{L_k}\Big\{\mathbf{P}_{0|k}(t) - \begin{bmatrix} \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{1|k}(t) \\ \vdots \\ \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{m|k}(t) \end{bmatrix} [\mathbf{q}_{1|k}(t) \dots \mathbf{q}_{m|k}(t)]\Big\} \\ &= \frac{1}{L_k}\Big[\mathbf{P}_{0|k}(t) - \sum_{i=1}^{n} \widetilde{\mathbf{A}}_{i|k}(t)\mathbf{P}^{\mathrm{T}}_{i|k}(t)\Big] \end{aligned} \quad (20)$$

where $\boldsymbol{\Phi}_k(t+i|t) = \mathbf{I}_m \otimes \boldsymbol{\varphi}_k(t+i|t)$.

Since relationships (19) and (20) can be recognized as Yule-Walker equations for a stationary VAR process (provided that the true autocorrelation matrices $\mathrm{E}[\mathbf{y}(t)\mathbf{y}^{\mathrm{T}}(t-l)]$ are replaced with their local (tapered) estimates $\mathbf{P}_{l|k}(t)/L_k$), the quantities $\widetilde{\boldsymbol{\theta}}_k(t)$ and $\widetilde{\boldsymbol{\rho}}_k(t)$ can be interpreted as noncausal weighted Yule-Walker estimators. Furthermore, it can be shown that when the process $\{\mathbf{y}(t)\}$ is persistently exciting in some deterministic or stochastic sense, the matrix $\mathbf{Q}_k(t)$ is positive definite[2] at all times t, which guarantees that the obtained models satisfy the uniform stability condition (2) – see Complement C8.6 in [35].

We note that when $n \ll k$, it holds that $\boldsymbol{\varphi}_k(t+i|t) \cong w_k(i)\boldsymbol{\varphi}(t+i)$, and since $\mathbf{y}_k(t+i|t) = w_k(i)\mathbf{y}(t+i)$, one arrives at $\widetilde{\boldsymbol{\theta}}_k(t) \cong \widehat{\boldsymbol{\theta}}_k(t)$ provided that

$$v_k(i) = w_k^2(i), \quad i \in [-k, k]\,. \quad (21)$$

This means that under the condition (21) the NWYW estimators will yield approximately the same results as the NWLS estimators. Note also that when $v_k(i)$ is the raised cosine window (9), the "equivalent" data taper is

$$w_k(i) = \sqrt{v_k(i)} = \cos\frac{\pi i}{2(k+1)}\,. \quad (22)$$

The cosinusoidal taper of this form allows for recursive computation of the quantities $\mathbf{Q}_k(t)$ and $\mathbf{q}_{l|k}(t)$, $l = 1, \dots, m$ – see Section 3.4 in [25].

Remark 2

Statistical properties of weighted (tapered) Yule-Walker estimators, which belong to a more general class of tapered Whittle estimators, were studied by Dahlhaus [36]. As shown in [36], tapering allows one to reduce both estimation bias and estimation variance of classical Yule-Walker estimators. For short data frames the improvement may be significant [25].

B. Estimation bandwidth scheduling

Paralleling developments made in Section III, we will define the holey estimator $\widetilde{\boldsymbol{\theta}}_k^{\circ}(t)$ in the form

$$\widetilde{\boldsymbol{\theta}}_k^{\circ}(t) = \arg\min_{\boldsymbol{\theta}} \sum_{\substack{i=-k \\ i\neq 0}}^{k+n} \| \mathbf{y}_k(t+i|t) - \boldsymbol{\Phi}_k^{\mathrm{T}}(t+i|t)\boldsymbol{\theta} \|^2 \quad (23)$$

and the corresponding pseudoprediction error in the form

$$\begin{aligned} \boldsymbol{\eta}_k^{\circ}(t) &= [\eta^{\circ}_{1|k}(t), \dots, \eta^{\circ}_{m|k}(t)]^{\mathrm{T}} = \mathbf{y}(t) - \boldsymbol{\Phi}^{\mathrm{T}}(t)\widetilde{\boldsymbol{\theta}}_k^{\circ}(t) \\ &= \mathbf{y}(t) - \sum_{i=1}^{n} \widetilde{\mathbf{A}}^{\circ}_{i|k}(t)\mathbf{y}(t-i) \end{aligned}$$

leading to

$$J_k(t) = \mathrm{PPE}_k(t) = \det\left\{\sum_{i=-M}^{M} \boldsymbol{\eta}_k^{\circ}(t+i)[\boldsymbol{\eta}_k^{\circ}(t+i)]^{\mathrm{T}}\right\} \quad (24)$$

or

$$\begin{aligned} J_k(t) = \mathrm{PPE}_k^*(t) &= \mathrm{tr}\left\{\sum_{i=-M}^{M} \boldsymbol{\eta}_k^{\circ}(t+i)[\boldsymbol{\eta}_k^{\circ}(t+i)]^{\mathrm{T}}\right\} \\ &= \sum_{i=-M}^{M} \| \boldsymbol{\eta}_k^{\circ}(t+i) \|^2\,. \end{aligned} \quad (25)$$

[2]Since $\mathbf{x}^{\mathrm{T}}\mathbf{Q}_k(t)\mathbf{x} = \sum_{i=-k}^{k+n}[\mathbf{x}^{\mathrm{T}}\boldsymbol{\varphi}_k(t+i|t)]^2 \geq 0$ for any mn-dimensional vector $\mathbf{x}$, the matrix $\mathbf{Q}_k(t)$ is "by construction" nonnegative definite.

We will show that, similarly as it was done in Section III, both statistics can be evaluated without implementing the holey estimator.

Proposition 2

It holds that

$$\boldsymbol{\eta}_k^{\circ}(t) = \boldsymbol{\eta}_k(t) + \frac{c_k(t)}{1-d_k(t)}\,\boldsymbol{\gamma}_k(t) \tag{26}$$

where

$$\begin{aligned}
\boldsymbol{\eta}_k(t) &= \mathbf{y}(t) - \boldsymbol{\Phi}^{\mathrm{T}}(t)\widetilde{\boldsymbol{\theta}}_k(t)\\
\boldsymbol{\gamma}_k(t) &= \mathbf{y}(t) - \boldsymbol{\Phi}_k^{\mathrm{T}}(t|t)\widetilde{\boldsymbol{\theta}}_k(t)\\
c_k(t) &= \boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{Q}_k^{-1}(t)\boldsymbol{\varphi}_k(t|t)\\
d_k(t) &= \boldsymbol{\varphi}_k^{\mathrm{T}}(t|t)\mathbf{Q}_k^{-1}(t)\boldsymbol{\varphi}_k(t|t).
\end{aligned}$$

Proof

Note that

$$\widetilde{\boldsymbol{\theta}}_k^{\circ}(t) = \begin{bmatrix} \widetilde{\theta}_{1|k}^{\circ}(t) \\ \vdots \\ \widetilde{\theta}_{m|k}^{\circ}(t) \end{bmatrix}$$

where

$$\begin{aligned}
\widetilde{\boldsymbol{\theta}}_{l|k}^{\circ}(t) &= [\mathbf{Q}_k^{\circ}(t)]^{-1}\mathbf{q}_{l|k}^{\circ}(t)\\
&= [\mathbf{Q}_k(t) - \boldsymbol{\varphi}_k(t|t)\boldsymbol{\varphi}_k^{\mathrm{T}}(t|t)]^{-1}\times\\
&\qquad \times [\mathbf{q}_{l|k}(t) - y_l(t)\boldsymbol{\varphi}_k(t|t)]\\
&= \left[\mathbf{Q}_k^{-1}(t) + \frac{\mathbf{Q}_k^{-1}(t)\boldsymbol{\varphi}_k(t|t)\boldsymbol{\varphi}_k^{\mathrm{T}}(t|t)\mathbf{Q}_k^{-1}(t)}{1-\boldsymbol{\varphi}_k^{\mathrm{T}}(t|t)\mathbf{Q}_k^{-1}(t)\boldsymbol{\varphi}_k(t|t)}\right]\times\\
&\qquad \times [\mathbf{Q}_k(t)\widetilde{\boldsymbol{\theta}}_{l|k}(t) - y_l(t)\boldsymbol{\varphi}_k(t|t)]\\
&= \widetilde{\boldsymbol{\theta}}_{l|k}(t) - \frac{\mathbf{Q}_k^{-1}(t)\boldsymbol{\varphi}_k(t|t)}{1-\boldsymbol{\varphi}_k^{\mathrm{T}}(t|t)\mathbf{Q}_k^{-1}(t)\boldsymbol{\varphi}_k(t|t)}\times\\
&\qquad \times [y_l(t) - \boldsymbol{\varphi}_k^{\mathrm{T}}(t|t)\widetilde{\boldsymbol{\theta}}_{l|k}(t)]
\end{aligned}$$

Based on this result, one obtains

$$\begin{aligned}
\eta_{l|k}^{\circ}(t) &= y_l(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\widetilde{\boldsymbol{\theta}}_{l|k}^{\circ}(t) = y_l(t) - \boldsymbol{\varphi}^{\mathrm{T}}(t)\widetilde{\boldsymbol{\theta}}_{l|k}(t)\\
&\quad + \frac{\boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{Q}_k^{-1}(t)\boldsymbol{\varphi}_k(t|t)}{1-\boldsymbol{\varphi}_k^{\mathrm{T}}(t|t)\mathbf{Q}_k^{-1}(t)\boldsymbol{\varphi}_k(t|t)}\,[y_l(t) - \boldsymbol{\varphi}_k^{\mathrm{T}}(t|t)\widetilde{\boldsymbol{\theta}}_{l|k}(t)]\\
&= \eta_{l|k}(t) + \frac{c_k(t)}{1-d_k(t)}\,\gamma_{l|k}(t), \quad l = 1,\ldots,m
\end{aligned}$$

which is equivalent to (26). ∎

When $n \ll k$, it holds that $\boldsymbol{\varphi}_k(t|t) \cong \boldsymbol{\varphi}(t)$, leading to $\boldsymbol{\eta}_k(t) \cong \boldsymbol{\gamma}_k(t)$, $c_k(t) \cong d_k(t)$ and

$$\boldsymbol{\eta}_k^{\circ}(t) \cong \frac{\boldsymbol{\gamma}_k(t)}{1-d_k(t)}$$

which resembles (17).

It is important to note that the holey NWYW estimator defined above differs from the one proposed in our earlier paper [25] – in the latter case dependence of $\widetilde{\boldsymbol{\theta}}_k^{\circ}(t)$ on $\mathbf{y}(t)$ was completely removed (at the cost of some extra bias errors).

Similarly as in the NWLS case, bandwidth selection can be also performed using the modified Akaike's FPE criterion, namely, by setting

$$J_k(t) = \mathrm{FPE}_k(t) = \left[\frac{1+\frac{mn}{N_k}}{1-\frac{mn}{N_k}}\right]^m \det\widetilde{\boldsymbol{\rho}}_k(t) \tag{27}$$

where

$$N_k = \frac{\left[\sum_{i=-k}^{k} w_k^2(i)\right]^2}{\sum_{i=-k}^{k} w_k^4(i)} \cong k\,\frac{\left[\int_{-1}^{1} g^2(x)dx\right]^2}{\int_{-1}^{1} g^4(x)dx}\,.$$

Remark 3

For NWYW estimators the dependence of the mean squared parameter estimation error on k was studied in [10] using the concept of infill asymptotics. In this approach the discrete time parameter trajectory of an AR process is regarded as a result of sampling a prototype continuous time trajectory. When a fixed-length time interval is sampled over a finer and finer grid of points as the sample size increases, one arrives at a family of increasingly stationary AR processes amenable to asymptotic analysis. Assuming that all processes constituting this family obey the uniform stability condition (2), and that the prototype trajectory is sufficiently smooth (uniformly bounded first, second and third derivatives) the following formula can be derived [10]

$$\mathrm{E}\left[\,\|\,\widetilde{\boldsymbol{\theta}}_k(t) - \boldsymbol{\theta}(t)\,\|^2\right] \cong \frac{b_1(t)}{k} + b_2(t)k^4 \tag{28}$$

where the first term on the right side of (28) corresponds to the variance component of MSE, and the second term is its (squared) bias component. Since the coefficients $b_1(t) > 0$ and $b_2(t) > 0$ are usually unknown, this result is of little practical value. However, it allows one to optimize bandwidth parameters of the parallel estimation scheme. Based on (28), it can be shown that to maximize the overall estimation bandwidth of such a scheme, the parameters $k_1,\ldots,k_K$ should form a geometric progression [25], i.e.,

$$k_{i+1} = \gamma^i k_1, \quad i = 1,\ldots,K-1$$

where the scaling coefficient $\gamma > 1$ depends on the assumed acceptable performance degradation margin and *does not* depend on $b_1(t)$ and $b_2(t)$. As shown in [25], for the acceptable degradation margin of 10% the corresponding value of γ is equal to 1.57.

As to the choice of the number of algorithms working in parallel, it often suffices to take $K = 3$, i.e., to combine the short memory algorithm, the nominal memory algorithm and the long memory one.

Due to asymptotic equivalence of NWLS and NWYW estimators, the recommendations presented above carry on to parallel estimation schemes made up of NWLS algorithms.

V. Joint selection of estimation bandwidth and model order

So far we have considered the situation where the order of autoregression n is known *a priori*. If this is not the case, one

can consider a family of VAR models of different orders and different bandwidth settings

$$\widehat{\boldsymbol{\theta}}_{n|k}(t) = \mathrm{vec}\{[\widehat{\mathbf{A}}_{1,n|k}(t)\dots\widehat{\mathbf{A}}_{n,n|k}(t)]^{\mathrm{T}}\},\ \widehat{\boldsymbol{\rho}}_{n|k}(t)$$
$$n \in \mathcal{N} = \{1,\dots,N\},\quad k \in \mathcal{K}$$

obtained using the NWLS approach. Selection of the most appropriate values of n and k can be based on minimization of the pseudoprediction error statistic

$$\{\widehat{n}(t),\widehat{k}(t)\} = \arg\min_{\substack{n\in\mathcal{N}\\k\in\mathcal{K}}} \mathrm{PPE}_{n|k}(t) \tag{29}$$

where

$$\mathrm{PPE}_{n|k}(t) = \det\left\{\sum_{i=-M}^{M} \boldsymbol{\varepsilon}^{\circ}_{n|k}(t+i)[\boldsymbol{\varepsilon}^{\circ}_{n|k}(t+i)]^{\mathrm{T}}\right\} \tag{30}$$

and $\boldsymbol{\varepsilon}^{\circ}_{n|k}(t)$ denotes pseudoprediction error, which – according to Proposition 1 – can be expressed as an appropriately scaled version of an easier to compute error

$$\boldsymbol{\varepsilon}_{n|k}(t) = \mathbf{y}(t) - \sum_{i=1}^{n} \widehat{\mathbf{A}}_{i,n|k}(t)\mathbf{y}(t-i).$$

Alternatively, one can use the "trace" version of (30).

As shown in [25], the same goal can be achieved by minimizing over n and k the FPE statistic

$$\mathrm{FPE}_{n|k}(t) = \left[\frac{1+\frac{mn}{N_k}}{1-\frac{mn}{N_k}}\right]^{m} \det\widehat{\boldsymbol{\rho}}_{n|k}(t). \tag{31}$$

Finally, one can consider mixed strategies, where FPE is used for model order selection and PPE for bandwidth selection or *vice versa*.

Extensions of the techniques described above to NWYW estimators is straightforward.

VI. Computational aspects

The NWYW and NWLS approaches are computationally attractive. As already shown, in both cases for a given k computations can be carried out in a time-recursive way.

The Whittle-Wiggins-Robinson (WWR) algorithm [35], which can be used to solve Yule-Walker equations (19)-(20) is order-recursive, which means that as a byproduct of computation of $\widetilde{\boldsymbol{\theta}}_{n|k}(t) = \mathrm{vec}\{[\widetilde{\mathbf{A}}_{1,n|k}(t)\dots\widetilde{\mathbf{A}}_{n,n|k}(t)]^{\mathrm{T}}\}$, $\widetilde{\boldsymbol{\rho}}_{n|k}(t)$, one obtains parameters of all lower-order models $\widetilde{\boldsymbol{\theta}}_{i|k}(t)$, $\widetilde{\boldsymbol{\rho}}_{i|k}(t)$, $i<n$. Additionally, as shown in Complement 8.6 of [35], the block lower-triangular Cholesky factors of the matrices $\mathbf{Q}^{-1}_{i|k}(t), i=1,\dots,n$ – the order-dependent variants of $\mathbf{Q}^{-1}_{k}(t)$ – can be expressed in terms of the quantities evaluated by the WWR algorithm. Hence, there is no need to perform matrix inversion when calculating pseudoprediction errors according to (26). The per sample computational load of the parallel estimation scheme incorporating NWYW algorithms is of order $O(KN^2m^3)$. The additional cost of carrying out model selection using the FPE/PPE approach is of order $O(KNm^3)$ and constitutes a small fraction of the overall computational load.

The square-root order-recursive algorithm for computation of the NWLS estimates $\widehat{\boldsymbol{\theta}}_{i|k}(t)$, $\widehat{\boldsymbol{\rho}}_{i|k}(t)$ and $\mathbf{R}^{-1}_{i|k}(t)$, $i=1,\dots,n$, is also a classical one and can be found e.g. in [37].

VII. Simulation results

It is known that every zero-mean stationary VAR process characterized by parameters $\{\boldsymbol{\rho},\mathbf{A}_1,\dots,\mathbf{A}_n\}$ has an equivalent lattice representation $\{\boldsymbol{\sigma},\boldsymbol{\Delta}_1,\dots,\boldsymbol{\Delta}_n\}$, where $\boldsymbol{\sigma} = \mathrm{E}[\mathbf{y}(t)\mathbf{y}^{\mathrm{T}}(t)]$ and $\boldsymbol{\Delta}_i, i=1,\dots,n$ denote the matrices of normalized reflection coefficients (partial autocorrelation coefficients) obeying the following stability condition [$s_{\max}(\cdot)$ denotes the maximum singular value of the respective matrix]

$$s_{\max}(\boldsymbol{\Delta}_i) < 1,\ i=1,\dots,n.$$

Generation of a two-dimensional ($m=2$) nonstationary VAR process was based on three time-invariant "anchor" models specified in the lattice form and obtained by means of identification of three different fragments of a stereo audio recording: the 2-nd order model $M_2 = \{\boldsymbol{\sigma}_2,\boldsymbol{\Delta}_{1,2},\boldsymbol{\Delta}_{2,2}\}$, the 4-th order model $M_4 = \{\boldsymbol{\sigma}_4,\boldsymbol{\Delta}_{1,4},\dots,\boldsymbol{\Delta}_{4,4}\}$ and the 6-th order model $M_6 = \{\boldsymbol{\sigma}_6,\boldsymbol{\Delta}_{1,6},\dots,\boldsymbol{\Delta}_{6,6}\}$ – all parameters are listed below.

M_2:

$$\boldsymbol{\Delta}_{1,2} = \begin{bmatrix} 0.9808 & 0.0375 \\ -0.0333 & 0.9806 \end{bmatrix}, \boldsymbol{\Delta}_{2,2} = \begin{bmatrix} -0.9864 & 0.0043 \\ 0.0056 & -0.9838 \end{bmatrix}$$

$$\boldsymbol{\sigma}_2 = \begin{bmatrix} 0.0025 & -0.0002 \\ -0.0002 & 0.0013 \end{bmatrix}$$

M_4:

$$\boldsymbol{\Delta}_{1,4} = \begin{bmatrix} 0.6869 & -0.0502 \\ 0.6162 & 0.4843 \end{bmatrix}, \boldsymbol{\Delta}_{2,4} = \begin{bmatrix} -0.7596 & 0.4759 \\ -0.2216 & -0.5751 \end{bmatrix}$$
$$\boldsymbol{\Delta}_{3,4} = \begin{bmatrix} 0.7596 & 0.1349 \\ -0.0863 & 0.6461 \end{bmatrix}, \boldsymbol{\Delta}_{4,4} = \begin{bmatrix} -0.4311 & 0.1303 \\ -0.2740 & -0.4002 \end{bmatrix}$$

$$\boldsymbol{\sigma}_4 = \begin{bmatrix} 0.0004366 & 0.0000151 \\ 0.0000151 & 0.0001626 \end{bmatrix}$$

M_6:

$$\boldsymbol{\Delta}_{1,6} = \begin{bmatrix} 0.9895 & -0.0017 \\ 0.0094 & 0.9868 \end{bmatrix}, \boldsymbol{\Delta}_{2,6} = \begin{bmatrix} -0.5170 & -0.6811 \\ -0.5255 & -0.4615 \end{bmatrix}$$
$$\boldsymbol{\Delta}_{3,6} = \begin{bmatrix} 0.3435 & 0.3403 \\ -0.4453 & 0.2691 \end{bmatrix}, \boldsymbol{\Delta}_{4,6} = \begin{bmatrix} -0.3377 & -0.3732 \\ 0.3144 & -0.4107 \end{bmatrix}$$
$$\boldsymbol{\Delta}_{5,6} = \begin{bmatrix} 0.4616 & 0.3769 \\ -0.3874 & 0.3688 \end{bmatrix}, \boldsymbol{\Delta}_{6,6} = \begin{bmatrix} -0.2836 & -0.2263 \\ -0.0509 & -0.3226 \end{bmatrix}$$

$$\boldsymbol{\sigma}_6 = \begin{bmatrix} 0.0041 & 0.0032 \\ 0.0032 & 0.0065 \end{bmatrix}.$$

The simulation scenario is symbolically depicted in Fig. 1. According to this scenario, in the time intervals $[1,t_1]$, $[t_2,t_3)$, $[t_3,t_4]$, $[t_5,t_6)$ and $[t_6,T_{\mathrm{sim}}]$ the simulated process was governed by the time-invariant anchor models M_2, M_4, M_6, M_4 and M_2, respectively, and in the intervals (t_1,t_2) and (t_4,t_5) – by time-varying models obtained by morphing the model M_2 into M_4, and the model M_6 into M_4, respectively. Finally, the generating model was subject to two jump changes, from M_4 to M_6 at the instant t_3, and from M_4 to M_2 at the instant t_6.

Morphing of the model $\{\boldsymbol{\sigma}^A,\boldsymbol{\Delta}^A_1,\dots,\boldsymbol{\Delta}^A_n\}$, valid at the instant t_A, into the model $\{\boldsymbol{\sigma}^B,\boldsymbol{\Delta}^B_1,\dots,\boldsymbol{\Delta}^B_n\}$, valid at the

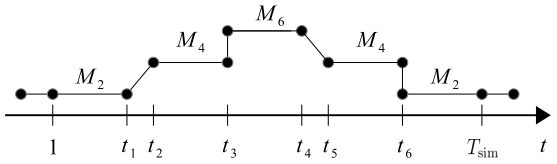

Fig. 1: Simulation scenario.

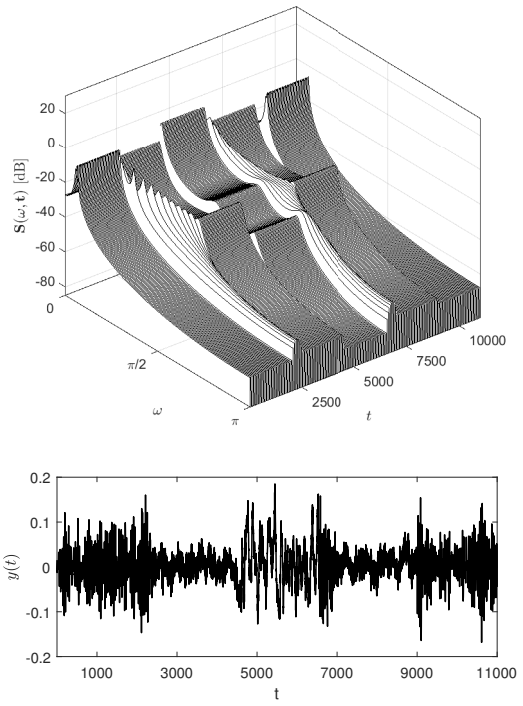

Fig. 2: Single-channel evolutionary spectrum of the simulated nonstationary autoregressive process (upper figure) and a typical process realization (lower figure).

instant t_B, was realized using the following model stability preserving rule

$$\begin{aligned}\boldsymbol{\sigma}(t) &= \mu(t)\boldsymbol{\sigma}^A + [1-\mu(t)]\boldsymbol{\sigma}^B \\ \boldsymbol{\Delta}_i(t) &= \mu(t)\boldsymbol{\Delta}_i^A + [1-\mu(t)]\boldsymbol{\Delta}_i^B \\ i &= 1,\ldots,n \\ \mu(t) &= \frac{t_B - t}{t_B - t_A}\ , \quad t \in (t_A, t_B)\end{aligned}$$

after replacing the nonexistent reflection coefficients with zeros (if applicable).

The evolutionary spectrum (defined below) of the generated process and a typical process realization are shown in Fig. 2.

The parallel estimation scheme was made up of 60 constant-bandwidth ($k_1 = 225$, $k_2 = 337$, $k_3 = 505$) constant-order ($n = 1,\ldots,20$) NWYW algorithms. The cosinusoidal taper (22) was applied ($N_1 = 300$, $N_2 = 450$, $N_3 = 675$). Note that, as recommended in [25], the equivalent memory settings form a geometric progression. The width $2M+1$ of the decision window was set to 35. Generation of $\mathbf{y}(t)$ was started 1000 time instants prior to $t = 1$ and was continued for 1000 time

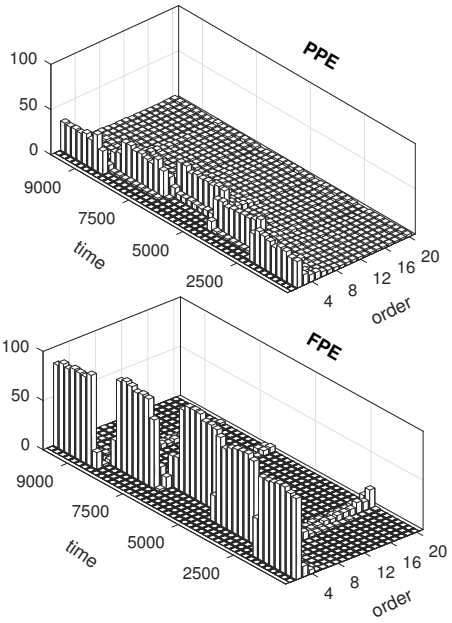

Fig. 3: Histograms of model order estimates obtained for 100 process realizations ($N = 20, T_{\rm sim} = 11000$). Each time bin covers 250 samples.

instants after $t = T_{\rm sim}$, so that all competing algorithms had enough data to start their operation at the instant 1 and finish it at the instant $T_{\rm sim}$.

To check performance of the compared algorithms under different degrees of signal nonstationarity, 3 different values of $T_{\rm sim}$ were considered: $T_{\rm sim} = 5500$ (S_1 – high speed of parameter variation), $T_{\rm sim} = 11000$ (S_2 – medium speed of parameter variation, two times smaller than S_1) and $T_{\rm sim} = 22000$ (S_3 – low speed of parameter variation, four times smaller than S_1).

Two performance measures were used to evaluate estimation results: the squared parameter estimation error

$$d_{\rm PAR}(t) = \| \widetilde{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t) \|^2$$

and the relative entropy rate

$$\begin{aligned}d_{\rm RER}(t) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \Big\{ &\mathrm{tr}\left[\left(\mathbf{S}(\omega,t) - \widehat{\mathbf{S}}(\omega,t)\right)\widehat{\mathbf{S}}^{-1}(\omega,t)\right] \\ &- \log\det\left[\mathbf{S}(\omega,t)\widehat{\mathbf{S}}^{-1}(\omega,t)\right]\Big\}\, d\omega\end{aligned}$$

where $\mathbf{S}(\omega,t)$ denotes the uniquely defined instantaneous spectral density function of the time-varying VAR process [11]

$$\begin{aligned}\mathbf{S}(\omega,t) &= \boldsymbol{\mathcal{A}}^{-1}[e^{-j\omega},\boldsymbol{\theta}(t)]\ \boldsymbol{\rho}(t)\ \boldsymbol{\mathcal{A}}^{-\mathrm{T}}[e^{j\omega},\boldsymbol{\theta}(t)] \\ \boldsymbol{\mathcal{A}}[e^{-j\omega},\boldsymbol{\theta}(t)] &= \mathbf{I} - \sum_{i=1}^{n}\mathbf{A}_i(t)e^{-ij\omega} \\ j &= \sqrt{-1}, \quad \omega \in (-\pi,\pi]\end{aligned} \tag{32}$$

and $\widehat{\mathbf{S}}(\omega,t)$ is the spectral density estimate obtained by means of replacing $\boldsymbol{\theta}(t)$ and $\boldsymbol{\rho}(t)$ in (32) with the corresponding estimates. The relative entropy rate is an extension, to the

multivariate case, of the well-known Itakura-Saito spectral distortion measure [38].

Evaluation of different estimation schemes is based on comparison of the mean scores obtained after combined time and ensemble averaging of $d_{\mathrm{PAR}}(t)$ and $d_{\mathrm{RER}}(t)$ (over $t \in [1, T_{\mathrm{sim}}]$ and 100 independent realizations of $\{\mathbf{y}(t)\}$): Table I shows the mean squared parameter estimation errors, Table II – the (squared) bias component of parameter MSE, and Table III – the corresponding spectral estimation errors. In each table the columns 2-4 show results obtained for the fixed-bandwidth fixed-order algorithms (for different values of k and n). The columns 5-6 show results yielded for a fixed value of n by the bandwidth-adaptive algorithms based on the pseudoprediction error criterion (PPE) and the final prediction error criterion (FPE). Finally, the columns 7-9 show results yielded (for different values of N) by 3 joint order- and bandwidth-adaptive algorithms: the one based exclusively on the pseudoprediction error criterion (PPE), the one based exclusively on the final prediction error criterion (FPE), and the one incorporating the two-stage mixed PPE/FPE strategy (Mix). In the latter case the FPE criterion is used (first) for model order selection for each value of k

$$\widehat{n}_k(t) = \arg\min_{n \in \mathcal{N}} \mathrm{FPE}_{n|k}(t), \quad k \in \mathcal{K}$$

and the PPE criterion is used (subsequently) for bandwidth selection

$$\widehat{k}(t) = \arg\min_{k \in \mathcal{K}} \mathrm{PPE}_{\widehat{n}_k(t)|k}(t).$$

The final decision takes the form $\{\widehat{n}_{\widehat{k}(t)}(t), \widehat{k}(t)\}$.

In our second experiment the simulation scenario depicted in Fig. 1 was preserved, but the assignment of models to the intervals $[1, t_1]$, $[t_2, t_3)$, $[t_3, t_4]$, $[t_5, t_6)$ and $[t_6, T_{\mathrm{sim}}]$ was randomized, namely, the model assigned to each interval was randomly selected (with equal probabilities) from the set $\mathcal{M} = \{M_2, M_4, M_6\}$, the only restriction being that each two consecutive models had to be different (to avoid "fictitious" transitions from M_i to M_i, $i = 2, 4, 6$). Tables IV and V show the scores obtained by averaging the results of 100 random assignment tests described above. Note that these scores are very similar, both from the qualitative and quantitative point of view, to the results presented in Tables I and III, respectively.

The following conclusions can be drawn after analysis of the content of Tables I–V:

1) When the model order is not underestimated ($n \geq 6$) both bandwidth-adaptive schemes yield, in almost all cases, better results than the fixed-bandwidth algorithms they are made up of. For fast and medium-speed parameter variations the PPE criterion yields better results than the FPE criterion.
2) When applied to joint order and bandwidth selection the FPE criterion yields in most cases better results than the PPE criterion. However, since FPE seems to have better order selection properties (see e.g. Fig. 3), and PPE has better bandwidth selection properties (see point 1 above), the best results are obtained when the mixed strategy is applied, i.e., when both criteria are combined. Moreover, when the model order is not underestimated ($n, N \geq 6$) such a mixed strategy yields better results than any of the considered fixed-order fixed-bandwidth algorithms (for all considered values of N).
3) All adaptive schemes achieve an almost optimal (fifty-fifty) balance between the bias and variance components of the parameter estimation MSE (the variance components of MSE can be calculated by subtracting from the values shown in Table I the corresponding values shown in Table II). For fixed-bandwidth algorithms the situation is different. As expected, for smaller values of k MSE is dominated by variance errors, while for larger values of k – by bias errors.

Spectrum estimation results obtained using the mixed strategy are shown in Fig. 4.

The results obtained when the trace version (25) of the PPE criterion is used (not shown in Tables I–III) are slightly inferior to those obtained when the determinant version (24) is applied.

Finally, we note that when the condition of equivalence (21) holds true, the results obtained using the NWLS approach (skipped because of the lack of space) are very similar, both in the quantitative and qualitative sense, to those presented above. The only important difference is that the time-varying models obtained using this approach are not guaranteed to be uniformly stable.

VIII. Conclusion

The problem of joint selection of the model order and estimation bandwidth for the purpose of noncausal identification of nonstationary multivariate autoregressive processes was considered and its new solution, based on evaluation of pseudoprediction errors, was proposed. It was shown that the best estimation results can be obtained when the pseudoprediction based estimation bandwidth scheduling is combined with model order selection based on the modified Akaike's final prediction error statistic.

References

[1] S.G. Fabri, K.P. Camilleri, and T. Cassar, "Parametric modelling of EEG data for the identification of mental tasks," *Biomed. Eng. Trends in Electron., Commun. & Software*, (A. Laskovski (Ed.), pp. 367-386, 2011.

[2] A. Schlögl, *The Electroencephalogram and the Adaptive Autoregressive Model: Theory and Applications*. Aachen, Germany: Shaker Verlag, 2000.

[3] T. Wada, M. Jinnouchi, and Y. Matsumura, "Application of autoregressive modelling for the analysis of clinical and other biological data," *Ann. Inst. Statist. Math.*, vol. 40, pp. 211-227, 1998.

[4] A.A. Zaman, M. Ferdjallah, and A. Khamayseh, A. "Muscle fatigue analysis for healthy adults using TVAR model with instantaneous frequency estimation," *Proc. IEEE 38th Southeast. Symp. Syst. Theory (SSST)*. Cookeville, TN, USA, 244-247, 2006.

[5] K.E. Baddour and N.C. Beaulieu, "Autoregressive models for fading channel simulation," *IEEE Trans. Wireless Comm.*, vol. 4, pp. 1650–1662, 2005.

[6] J.F. Hayes and T.V.J. Ganesh Babu, *Modeling and Analysis of Telecommunication Networks*. Wiley, 2004.

[7] P. Lesage, F. Glangeaud, and J. Mars, "Applications of autoregressive models and time-frequency analysis to the study of volcanic tremor and long-period events," *J. Volc. Geotherm. Res.*, vol. 114, pp. 391-417, 2002.

[8] C. Li and R.L. Nowack, "Application of autoregressive extrapolation to seismic tomography," *Bull. Seism. Soc. Amer.*, pp. 1456-1466, 2004.

[9] D. Brillinger, E.A. Robinson, and F.P. Schoenberg, Eds., *Time Series Analysis and Applications to Geophysical Systems*. Springer, 2012.

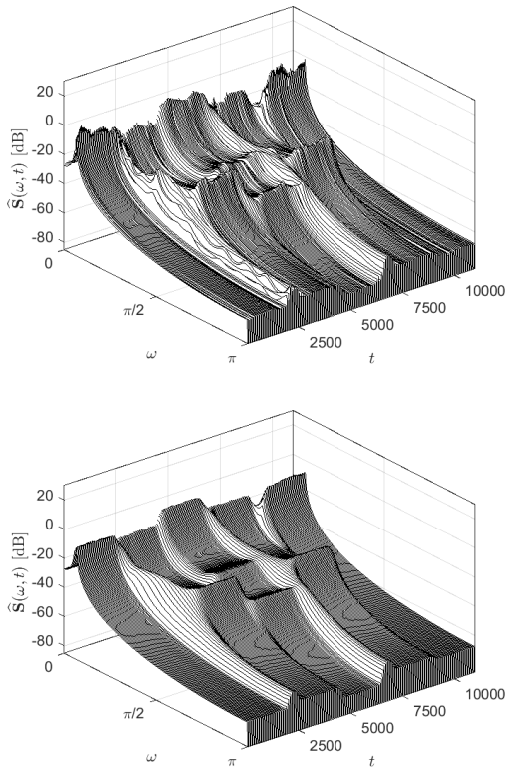

Fig. 4: Estimated evolutionary spectrum of the simulated nonstationary autoregressive process ($T_{\mathrm{sim}} = 11000$) obtained for a single process realization (upper figure) and averaged over 100 realizations (lower figure).

[10] R. Dahlhaus and L. Giraitis, "On the optimal segment length for parameter estimates for locally stationary time series," *J. Time Series Anal.*, vol. 19, pp. 629–655, 1998.

[11] R. Dahlhaus, "Local inference for locally stationary time series based on the empirical spectral measure," *J. Econometrics*, vol. 151, pp. 101–112, 2009.

[12] R. Dahlhaus, "Locally stationary processes," *Handbook Statist.*, vol. 25, pp. 1–37, 2012.

[13] M. Niedźwiecki, *Identification of Time-varying Processes*. New York: Wiley, 2000.

[14] H. Akaike, "Statistical predictor identification," *Ann. Inst. Statist. Math.*, vol. 22, pp. 203–217, 1970.

[15] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. 19, pp. 716–723, 1974.

[16] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, pp. 461–464, 1978.

[17] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[18] M. Niedźwiecki, "On the localized estimators and generalized Akaike's criteria," *IEEE Trans. Automat. Contr.*, vol. 29, pp. 970–983, 1984.

[19] J.H.W. Penm and R.D. Torrell, "Multivariate subset autoregression," *Comm. Statist. B*, vol. 13, pp. 449–461, 1984.

[20] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. R. Stat. Soc. Series B*, vol. 58, pp. 267–288, 1996.

[21] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. R. Stat. Soc. Series B*, vol. 68, pp. 49-67, 2006.

[22] M. Kalli and J.E. Griffin, "Time-varying sparsity in dynamic regression models," *J. Econometrics*, vol. 178, pp. 779–793, 2014.

[23] A. Goldenshluger and A. Nemirovski, "On spatial adaptive estimation of nonparametric regression," *Math. Meth. Stat.*, vol. 6, pp. 135–170, 1997.

[24] V. Katkovnik, "A new method for varying adaptive bandwidth selection," *IEEE Trans. Signal Process.*, vol. 47, pp. 2567–2571, 1999.

TABLE I: Parameter estimation MSE

Fast parameter variations

S_1	Nonadaptive			Bandwidth adaptive		Bandwidth and order adaptive		
n/N	k_1	k_2	k_3	PPE	FPE	PPE	FPE	Mix
1	21.401	21.343	21.264	21.385	21.383	21.385	21.383	21.385
2	16.835	16.997	17.268	16.856	16.883	16.866	16.883	16.856
3	10.930	11.285	11.890	10.875	11.011	10.787	10.997	10.862
4	5.772	5.979	6.353	5.717	5.810	5.721	5.781	5.688
5	2.654	2.985	3.625	2.597	2.714	2.498	2.620	2.507
6	1.546	2.018	2.934	1.471	1.596	1.234	1.418	1.307
7	1.758	2.106	2.895	1.627	1.743	1.308	1.419	1.318
8	1.997	2.230	2.891	1.802	1.904	1.375	1.445	1.365
9	2.268	2.413	3.013	2.010	2.118	1.440	1.485	1.403
10	2.549	2.604	3.150	2.226	2.337	1.505	1.523	1.432
11	2.841	2.818	3.312	2.444	2.559	1.561	1.569	1.463
12	3.141	3.022	3.448	2.665	2.777	1.619	1.593	1.490
13	3.435	3.230	3.600	2.881	2.995	1.666	1.617	1.505
14	3.754	3.455	3.771	3.108	3.229	1.713	1.648	1.522
15	4.034	3.653	3.927	3.304	3.431	1.755	1.688	1.538
16	4.335	3.861	4.085	3.515	3.643	1.796	1.725	1.555
17	4.642	4.072	4.239	3.726	3.854	1.834	1.750	1.567
18	4.970	4.296	4.402	3.947	4.085	1.874	1.782	1.581
19	5.269	4.510	4.572	4.146	4.293	1.908	1.812	1.591
20	5.577	4.731	4.748	4.353	4.508	1.939	1.847	1.605

Medium-speed parameter variations

S_2	Nonadaptive			Bandwidth adaptive		Bandwidth and order adaptive		
n/N	k_1	k_2	k_3	PPE	FPE	PPE	FPE	Mix
1	21.446	21.410	21.368	21.434	21.430	21.434	21.430	21.434
2	16.663	16.724	16.827	16.670	16.678	16.675	16.678	16.670
3	10.546	10.656	10.878	10.502	10.527	10.482	10.512	10.487
4	5.530	5.583	5.742	5.472	5.505	5.500	5.473	5.442
5	2.253	2.296	2.519	2.154	2.195	2.126	2.097	2.066
6	0.949	1.028	1.361	0.804	0.850	0.741	0.675	0.644
7	1.171	1.145	1.396	0.945	0.996	0.805	0.687	0.655
8	1.415	1.292	1.466	1.100	1.155	0.861	0.708	0.673
9	1.688	1.478	1.590	1.276	1.333	0.920	0.724	0.687
10	1.943	1.643	1.696	1.437	1.489	0.973	0.742	0.703
11	2.225	1.834	1.825	1.613	1.663	1.024	0.755	0.711
12	2.505	2.020	1.947	1.780	1.824	1.071	0.765	0.717
13	2.772	2.197	2.066	1.943	1.980	1.117	0.773	0.720
14	3.039	2.376	2.191	2.104	2.140	1.157	0.782	0.725
15	3.316	2.561	2.315	2.265	2.304	1.194	0.789	0.728
16	3.610	2.758	2.451	2.434	2.471	1.232	0.797	0.731
17	3.904	2.948	2.575	2.599	2.635	1.269	0.802	0.732
18	4.199	3.145	2.712	2.761	2.801	1.299	0.808	0.732
19	4.493	3.335	2.842	2.919	2.964	1.330	0.814	0.734
20	4.790	3.529	2.977	3.079	3.131	1.358	0.823	0.737

Slow parameter variations

S_3	Nonadaptive			Bandwidth adaptive		Bandwidth and order adaptive		
n/N	k_1	k_2	k_3	PPE	FPE	PPE	FPE	Mix
1	21.465	21.441	21.416	21.454	21.451	21.454	21.451	21.454
2	16.618	16.645	16.686	16.622	16.620	16.625	16.620	16.622
3	10.458	10.490	10.574	10.420	10.413	10.414	10.398	10.405
4	5.465	5.448	5.486	5.405	5.408	5.447	5.377	5.375
5	2.119	2.061	2.090	2.000	2.009	1.992	1.909	1.909
6	0.716	0.637	0.690	0.534	0.547	0.550	0.369	0.370
7	0.956	0.781	0.761	0.676	0.692	0.608	0.384	0.387
8	1.208	0.936	0.847	0.819	0.834	0.663	0.392	0.399
9	1.468	1.111	0.964	0.970	0.985	0.716	0.401	0.409
10	1.731	1.286	1.080	1.118	1.129	0.766	0.406	0.414
11	1.999	1.464	1.199	1.269	1.277	0.814	0.410	0.419
12	2.261	1.637	1.314	1.408	1.413	0.858	0.411	0.421
13	2.534	1.821	1.438	1.556	1.558	0.902	0.412	0.422
14	2.810	1.998	1.552	1.692	1.693	0.941	0.413	0.423
15	3.084	2.179	1.671	1.830	1.828	0.978	0.414	0.424
16	3.374	2.368	1.795	1.971	1.970	1.013	0.415	0.425
17	3.655	2.553	1.918	2.110	2.109	1.048	0.415	0.425
18	3.939	2.740	2.043	2.247	2.244	1.080	0.416	0.426
19	4.232	2.931	2.168	2.385	2.383	1.112	0.416	0.426
20	4.526	3.123	2.296	2.526	2.528	1.141	0.417	0.427

TABLE II: Bias component of parameter estimation MSE

Fast parameter variations

S_1	Nonadaptive			Bandwidth adaptive		Bandwidth and order adaptive		
n/N	k_1	k_2	k_3	PPE	FPE	PPE	FPE	Mix
1	21.382	21.331	21.256	21.368	21.365	21.368	21.365	21.368
2	16.783	16.948	17.219	16.797	16.832	16.805	16.832	16.797
3	10.766	11.127	11.727	10.703	10.855	10.560	10.854	10.702
4	5.566	5.822	6.224	5.528	5.641	5.419	5.640	5.527
5	2.280	2.701	3.392	2.270	2.408	2.070	2.401	2.262
6	0.945	1.567	2.569	0.959	1.105	0.570	1.098	0.951
7	0.877	1.454	2.371	0.890	1.027	0.570	1.012	0.875
8	0.837	1.371	2.201	0.851	0.969	0.568	0.975	0.858
9	0.833	1.355	2.170	0.851	0.967	0.567	0.963	0.850
10	0.821	1.335	2.141	0.840	0.958	0.567	0.957	0.844
11	0.821	1.330	2.128	0.840	0.959	0.568	0.955	0.841
12	0.814	1.313	2.095	0.837	0.947	0.568	0.944	0.839
13	0.812	1.306	2.079	0.837	0.944	0.569	0.938	0.837
14	0.813	1.303	2.073	0.842	0.947	0.569	0.936	0.836
15	0.813	1.302	2.069	0.846	0.949	0.570	0.935	0.834
16	0.812	1.298	2.062	0.850	0.950	0.570	0.935	0.834
17	0.810	1.291	2.050	0.854	0.948	0.571	0.932	0.833
18	0.814	1.291	2.047	0.862	0.951	0.571	0.931	0.832
19	0.813	1.290	2.043	0.868	0.954	0.572	0.931	0.832
20	0.815	1.290	2.043	0.877	0.958	0.572	0.932	0.832

Medium-speed parameter variations

S_2	Nonadaptive			Bandwidth adaptive		Bandwidth and order adaptive		
n/N	k_1	k_2	k_3	PPE	FPE	PPE	FPE	Mix
1	21.426	21.398	21.360	21.417	21.412	21.417	21.412	21.417
2	16.623	16.691	16.797	16.630	16.642	16.634	16.642	16.630
3	10.424	10.551	10.779	10.390	10.419	10.346	10.419	10.389
4	5.347	5.456	5.649	5.328	5.368	5.296	5.367	5.327
5	1.911	2.060	2.350	1.898	1.942	1.831	1.938	1.895
6	0.404	0.653	1.095	0.408	0.457	0.244	0.451	0.401
7	0.362	0.587	0.999	0.365	0.417	0.242	0.407	0.355
8	0.339	0.550	0.936	0.344	0.394	0.237	0.391	0.342
9	0.338	0.548	0.929	0.345	0.393	0.236	0.388	0.340
10	0.336	0.541	0.914	0.343	0.390	0.237	0.386	0.339
11	0.336	0.539	0.910	0.346	0.391	0.238	0.385	0.338
12	0.338	0.538	0.905	0.349	0.392	0.238	0.384	0.337
13	0.339	0.537	0.900	0.352	0.392	0.239	0.382	0.337
14	0.341	0.537	0.897	0.355	0.394	0.240	0.382	0.337
15	0.342	0.536	0.895	0.359	0.396	0.240	0.381	0.337
16	0.344	0.536	0.892	0.362	0.397	0.241	0.381	0.337
17	0.346	0.536	0.890	0.365	0.398	0.241	0.380	0.337
18	0.350	0.538	0.891	0.371	0.402	0.242	0.380	0.337
19	0.354	0.540	0.891	0.377	0.405	0.242	0.380	0.337
20	0.357	0.540	0.889	0.381	0.408	0.242	0.380	0.337

Slow parameter variations

S_3	Nonadaptive			Bandwidth adaptive		Bandwidth and order adaptive		
n/N	k_1	k_2	k_3	PPE	FPE	PPE	FPE	Mix
1	21.445	21.428	21.408	21.437	21.434	21.437	21.434	21.437
2	16.586	16.621	16.667	16.594	16.594	16.595	16.594	16.594
3	10.361	10.417	10.514	10.340	10.336	10.324	10.335	10.339
4	5.294	5.332	5.405	5.282	5.285	5.287	5.285	5.281
5	1.789	1.838	1.938	1.776	1.781	1.759	1.779	1.774
6	0.181	0.278	0.448	0.183	0.189	0.120	0.186	0.180
7	0.162	0.246	0.397	0.164	0.170	0.118	0.166	0.159
8	0.151	0.227	0.366	0.153	0.159	0.115	0.158	0.152
9	0.154	0.227	0.365	0.155	0.160	0.114	0.156	0.151
10	0.155	0.226	0.361	0.157	0.161	0.114	0.156	0.150
11	0.157	0.226	0.359	0.158	0.161	0.114	0.155	0.150
12	0.161	0.228	0.359	0.161	0.164	0.114	0.155	0.150
13	0.164	0.229	0.358	0.163	0.166	0.114	0.155	0.150
14	0.168	0.232	0.358	0.167	0.169	0.114	0.155	0.150
15	0.171	0.233	0.358	0.169	0.171	0.115	0.155	0.150
16	0.176	0.235	0.359	0.173	0.174	0.115	0.155	0.150
17	0.177	0.236	0.359	0.175	0.175	0.116	0.155	0.150
18	0.180	0.238	0.360	0.178	0.178	0.116	0.155	0.150
19	0.183	0.239	0.360	0.180	0.180	0.116	0.155	0.150
20	0.186	0.241	0.360	0.183	0.182	0.116	0.155	0.150

TABLE III: Averaged relative entropy rate (RER)

Fast parameter variations

S_1	Nonadaptive			Bandwidth adaptive		Bandwidth and order adaptive		
n/N	k_1	k_2	k_3	PPE	FPE	PPE	FPE	Mix
1	2.874	2.834	2.858	2.840	2.882	2.840	2.882	2.840
2	0.879	0.995	1.183	0.887	0.892	0.888	0.892	0.887
3	0.477	0.568	0.719	0.479	0.499	0.490	0.498	0.478
4	0.289	0.360	0.481	0.286	0.315	0.302	0.313	0.284
5	0.214	0.285	0.404	0.208	0.240	0.224	0.236	0.204
6	0.191	0.258	0.375	0.184	0.216	0.199	0.210	0.177
7	0.198	0.261	0.374	0.188	0.222	0.199	0.213	0.178
8	0.205	0.265	0.376	0.194	0.229	0.200	0.216	0.179
9	0.213	0.270	0.378	0.200	0.236	0.201	0.219	0.180
10	0.221	0.275	0.381	0.206	0.242	0.203	0.221	0.182
11	0.230	0.280	0.384	0.212	0.248	0.204	0.223	0.182
12	0.239	0.285	0.387	0.218	0.255	0.205	0.224	0.183
13	0.247	0.290	0.390	0.225	0.260	0.206	0.226	0.184
14	0.256	0.295	0.393	0.231	0.266	0.207	0.228	0.185
15	0.266	0.301	0.396	0.238	0.273	0.208	0.230	0.185
16	0.276	0.307	0.400	0.245	0.279	0.209	0.232	0.186
17	0.286	0.313	0.403	0.252	0.286	0.210	0.234	0.187
18	0.297	0.319	0.407	0.259	0.293	0.211	0.236	0.188
19	0.307	0.325	0.412	0.267	0.300	0.212	0.238	0.188
20	0.318	0.332	0.416	0.274	0.307	0.213	0.240	0.189

Medium-speed parameter variations

S_2	Nonadaptive			Bandwidth adaptive		Bandwidth and order adaptive		
n/N	k_1	k_2	k_3	PPE	FPE	PPE	FPE	Mix
1	2.848	2.775	2.756	2.800	2.841	2.800	2.841	2.800
2	0.750	0.791	0.878	0.749	0.745	0.749	0.745	0.749
3	0.373	0.404	0.473	0.367	0.371	0.374	0.371	0.367
4	0.203	0.224	0.279	0.193	0.201	0.203	0.199	0.191
5	0.129	0.148	0.202	0.117	0.125	0.126	0.121	0.113
6	0.108	0.124	0.176	0.093	0.101	0.101	0.095	0.086
7	0.115	0.127	0.177	0.097	0.105	0.102	0.095	0.086
8	0.123	0.131	0.179	0.101	0.109	0.104	0.096	0.087
9	0.131	0.136	0.182	0.105	0.114	0.105	0.096	0.087
10	0.139	0.141	0.184	0.110	0.118	0.106	0.097	0.087
11	0.148	0.146	0.187	0.115	0.122	0.107	0.097	0.088
12	0.157	0.151	0.190	0.120	0.127	0.109	0.098	0.088
13	0.166	0.156	0.193	0.125	0.131	0.110	0.098	0.088
14	0.175	0.161	0.196	0.130	0.135	0.111	0.099	0.088
15	0.184	0.167	0.200	0.135	0.140	0.112	0.099	0.088
16	0.194	0.172	0.203	0.140	0.144	0.113	0.100	0.089
17	0.205	0.178	0.206	0.145	0.149	0.114	0.100	0.089
18	0.215	0.184	0.209	0.150	0.154	0.115	0.100	0.089
19	0.226	0.190	0.213	0.155	0.159	0.116	0.101	0.089
20	0.236	0.196	0.216	0.160	0.163	0.117	0.101	0.089

Slow parameter variations

S_3	Nonadaptive			Bandwidth adaptive		Bandwidth and order adaptive		
n/N	k_1	k_2	k_3	PPE	FPE	PPE	FPE	Mix
1	2.832	2.749	2.711	2.781	2.822	2.781	2.822	2.781
2	0.696	0.701	0.730	0.688	0.686	0.688	0.686	0.688
3	0.332	0.333	0.355	0.322	0.321	0.325	0.320	0.321
4	0.170	0.166	0.182	0.156	0.157	0.163	0.155	0.155
5	0.096	0.091	0.105	0.080	0.081	0.085	0.077	0.076
6	0.077	0.068	0.080	0.056	0.058	0.062	0.051	0.050
7	0.085	0.072	0.082	0.060	0.061	0.063	0.051	0.050
8	0.093	0.077	0.085	0.065	0.065	0.065	0.051	0.051
9	0.101	0.081	0.087	0.069	0.069	0.066	0.051	0.051
10	0.109	0.087	0.090	0.073	0.073	0.068	0.051	0.051
11	0.118	0.092	0.094	0.077	0.076	0.069	0.052	0.051
12	0.128	0.097	0.097	0.082	0.080	0.070	0.052	0.051
13	0.137	0.102	0.100	0.086	0.084	0.071	0.052	0.051
14	0.146	0.108	0.103	0.090	0.087	0.073	0.052	0.051
15	0.156	0.114	0.106	0.094	0.091	0.074	0.052	0.051
16	0.166	0.119	0.110	0.098	0.095	0.075	0.052	0.051
17	0.177	0.125	0.113	0.102	0.099	0.076	0.052	0.051
18	0.187	0.131	0.117	0.107	0.103	0.077	0.052	0.051
19	0.198	0.137	0.120	0.111	0.107	0.078	0.052	0.051
20	0.209	0.143	0.124	0.115	0.111	0.079	0.052	0.051

[25] M. Niedźwiecki, M. Ciołek, and Y. Kajikawa, "On adaptive covariance and spectrum estimation of locally stationary multivariate processes," *Automatica*, vol. 82, pp. 1–12, 2017.
[26] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*. New York: Springer-Verlag, 2005.
[27] M. Niedźwiecki, M. Ciołek and K. Cisowski, "Elimination of impulsive disturbances from stereo audio recordings using vector autoregressive modeling and variable-order Kalman filtering," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 23, pp. 970–981, 2015.
[28] S. Kay, *Modern Spectrum Estimation: Theory and Application*, New Jersey: Prentice-Hall, 1999.
[29] E. Moulines, P. Priouret, and F. Roueff, "On recursive estimation for time-varying autoregressive processes," *Ann. Statist.*, vol. 33, pp. 2610–2654, 2005.
[30] D.T.L. Lee, M. Morf, and B. Friedlander, "Recursive least squares ladder estimation algorithms," *IEEE Trans. Acoust., Speech, Signal Process.* , vol. 29, pp. 627–641, 1981.
[31] M. Niedźwiecki and M. Ciołek, "Sparse vector autoregressive modeling of audio signals and its application to the elimination of impulsive disturbances", *Proc. 17th IFAC Symposium on System Identification*, Beijin, China, pp. 1202-1207, 2015.
[32] M. Niedźwiecki and M. Ciołek, "New results on estimation bandwidth adaptation", *Proc. 18h IFAC Symposium on System Identification*, Stockholm, Sweden, pp. 933–938, 2018.
[33] A. Zellner, "An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias," *Journ. Am. Statist. Ass.*, vol. 57, pp. 348–368, 1962.
[34] M. Niedźwiecki, "Locally-adaptive Kalman smoothing and its application to identification of nonstationary stochastic systems," *IEEE Trans. Sign. Process.*, vol. 60, pp. 48–59, 2012.
[35] T. Söderström and P. Stoica, *System Identification*. Englewoods Cliffs NJ: Prentice-Hall, 1988.
[36] R. Dahlhaus, "Small sample effects in time series analysis: a new asymptotic theory and a new estimate," *Ann. Statist.*, vol. 16, pp. 808–841, 1988.
[37] M. Karny, "Bayesian estimation of model order," *Problems of Control and Information Theory*, vol. 9, pp. 33–46, 1980.
[38] F. Itakura and S. Saito, "A statistical method for estimation of speech spectral density and formant frequencies," *Electron. Commun. Jap.*, vol. 53-A, pp. 36–43, 1970.

Maciej Niedźwiecki (M'08, SM'13) received the M.Sc. and Ph.D. degrees from the Technical University of Gdańsk, Gdańsk, Poland and the Dr.Hab. (D.Sc.) degree from the Technical University of Warsaw, Warsaw, Poland, in 1977, 1981 and 1991, respectively. He spent three years as a Research Fellow with the Department of Systems Engineering, Australian National University, 1986-1989. In 1990 - 1993 he served as a Vice Chairman of Technical Committee on Theory of the International Federation of Automatic Control (IFAC). He is the author of the book *Identification of Time-varying Processes* (Wiley, 2000). His main areas of research interests include system identification, statistical signal processing and adaptive systems.

Dr. Niedźwiecki is currently a member of the IFAC committees on Modeling, Identification and Signal Processing and on Large Scale Complex Systems, and a member of the Automatic Control and Robotics Committee of the Polish Academy of Sciences (PAN). He works as a Professor and Head of the Department of Automatic Control, Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology.

Marcin Ciołek (M'17) received the M.Sc. and Ph.D. degrees from the Gdańsk University of Technology (GUT), Gdańsk, Poland, in 2010 and 2017, respectively. Since 2017, he has been working as an Adjunct Professor in the Department of Automatic Control, Faculty of Electronics, Telecommunications and Informatics, GUT. His professional interests include speech, music and biomedical signal processing.

TABLE IV: Averaged parameter estimation errors obtained in the randomized model selection experiment

Fast parameter variations

S_1	Nonadaptive			Bandwidth adaptive		Bandwidth and order adaptive		
n/N	k_1	k_2	k_3	PPE	FPE	PPE	FPE	Mix
1	28.727	28.685	28.629	28.714	28.713	28.714	28.713	28.714
2	23.129	23.210	23.347	23.130	23.139	23.140	23.139	23.130
3	16.417	16.718	17.245	16.374	16.431	16.292	16.419	16.363
4	9.136	9.405	9.877	9.106	9.143	9.073	9.116	9.080
5	3.898	4.373	5.235	3.859	3.959	3.600	3.880	3.785
6	1.658	2.252	3.354	1.587	1.718	1.298	1.576	1.457
7	1.921	2.375	3.337	1.789	1.920	1.414	1.619	1.492
8	2.185	2.505	3.341	1.966	2.097	1.491	1.638	1.518
9	2.515	2.719	3.464	2.202	2.330	1.568	1.686	1.560
10	2.839	2.938	3.615	2.431	2.558	1.642	1.725	1.591
11	3.161	3.157	3.769	2.659	2.776	1.711	1.762	1.613
12	3.491	3.378	3.911	2.884	2.987	1.775	1.799	1.641
13	3.814	3.587	4.047	3.100	3.197	1.835	1.830	1.665
14	4.125	3.803	4.209	3.310	3.401	1.887	1.862	1.684
15	4.457	4.029	4.371	3.526	3.613	1.939	1.895	1.703
16	4.794	4.263	4.550	3.747	3.831	1.992	1.941	1.727
17	5.123	4.484	4.704	3.955	4.036	2.033	1.967	1.742
18	5.472	4.729	4.887	4.172	4.262	2.078	2.008	1.762
19	5.818	4.975	5.071	4.392	4.484	2.116	2.037	1.773
20	6.191	5.241	5.267	4.630	4.728	2.158	2.080	1.786

Medium-speed parameter variations

S_2	Nonadaptive			Bandwidth adaptive		Bandwidth and order adaptive		
n/N	k_1	k_2	k_3	PPE	FPE	PPE	FPE	Mix
1	28.876	28.850	28.820	28.866	28.862	28.866	28.862	28.866
2	23.201	23.233	23.288	23.199	23.196	23.206	23.196	23.199
3	16.265	16.349	16.528	16.226	16.211	16.220	16.198	16.215
4	8.926	8.986	9.149	8.886	8.874	8.917	8.846	8.859
5	3.407	3.507	3.812	3.332	3.348	3.250	3.265	3.257
6	0.976	1.104	1.518	0.845	0.880	0.833	0.738	0.714
7	1.249	1.257	1.581	1.022	1.064	0.911	0.775	0.746
8	1.549	1.432	1.658	1.203	1.251	0.982	0.795	0.765
9	1.841	1.620	1.773	1.380	1.427	1.045	0.816	0.783
10	2.161	1.830	1.907	1.571	1.614	1.107	0.829	0.793
11	2.488	2.047	2.052	1.768	1.808	1.167	0.840	0.801
12	2.814	2.264	2.202	1.962	1.994	1.227	0.857	0.811
13	3.135	2.480	2.346	2.147	2.176	1.278	0.866	0.816
14	3.462	2.698	2.492	2.327	2.355	1.326	0.878	0.822
15	3.775	2.902	2.629	2.497	2.524	1.371	0.888	0.826
16	4.108	3.123	2.780	2.675	2.697	1.415	0.897	0.830
17	4.445	3.350	2.933	2.854	2.874	1.455	0.905	0.832
18	4.796	3.585	3.095	3.041	3.059	1.495	0.913	0.835
19	5.162	3.823	3.255	3.228	3.245	1.539	0.921	0.837
20	5.500	4.046	3.407	3.400	3.422	1.573	0.925	0.838

Slow parameter variations

S_3	Nonadaptive			Bandwidth adaptive		Bandwidth and order adaptive		
n/N	k_1	k_2	k_3	PPE	FPE	PPE	FPE	Mix
1	28.901	28.881	28.863	28.892	28.889	28.892	28.889	28.892
2	23.167	23.182	23.203	23.167	23.157	23.170	23.157	23.167
3	16.236	16.263	16.332	16.199	16.172	16.203	16.160	16.187
4	8.863	8.857	8.906	8.812	8.792	8.863	8.765	8.786
5	3.228	3.201	3.276	3.123	3.122	3.098	3.040	3.046
6	0.729	0.676	0.764	0.556	0.567	0.641	0.425	0.424
7	1.015	0.847	0.848	0.723	0.737	0.708	0.453	0.450
8	1.306	1.033	0.956	0.887	0.906	0.770	0.467	0.466
9	1.602	1.229	1.084	1.053	1.071	0.830	0.478	0.478
10	1.915	1.435	1.216	1.223	1.239	0.890	0.483	0.485
11	2.227	1.644	1.356	1.391	1.404	0.944	0.489	0.490
12	2.543	1.855	1.496	1.559	1.566	0.996	0.494	0.495
13	2.857	2.064	1.634	1.718	1.719	1.047	0.498	0.499
14	3.174	2.272	1.772	1.877	1.872	1.093	0.501	0.502
15	3.493	2.484	1.912	2.037	2.024	1.138	0.504	0.504
16	3.830	2.704	2.058	2.201	2.182	1.179	0.506	0.505
17	4.143	2.908	2.193	2.349	2.324	1.218	0.507	0.506
18	4.466	3.120	2.332	2.501	2.471	1.254	0.508	0.507
19	4.801	3.339	2.478	2.658	2.628	1.289	0.509	0.507
20	5.139	3.559	2.622	2.810	2.782	1.321	0.509	0.508

TABLE V: Averaged relative entropy rates obtained in the randomized model selection experiment

Fast parameter variations

S_1	Nonadaptive			Bandwidth adaptive		Bandwidth and order adaptive		
n/N	k_1	k_2	k_3	PPE	FPE	PPE	FPE	Mix
1	2.755	2.712	2.718	2.721	2.758	2.721	2.758	2.721
2	0.906	1.002	1.166	0.912	0.915	0.912	0.915	0.912
3	0.536	0.611	0.739	0.535	0.553	0.549	0.552	0.534
4	0.343	0.403	0.510	0.338	0.363	0.356	0.361	0.337
5	0.217	0.278	0.386	0.211	0.237	0.228	0.233	0.208
6	0.179	0.238	0.343	0.171	0.198	0.189	0.192	0.165
7	0.185	0.239	0.341	0.175	0.203	0.189	0.194	0.165
8	0.192	0.242	0.341	0.179	0.209	0.190	0.196	0.166
9	0.199	0.246	0.342	0.184	0.214	0.191	0.198	0.166
10	0.207	0.251	0.344	0.190	0.220	0.192	0.199	0.167
11	0.216	0.255	0.346	0.195	0.225	0.194	0.200	0.168
12	0.224	0.260	0.348	0.201	0.230	0.195	0.202	0.169
13	0.233	0.265	0.351	0.207	0.236	0.196	0.204	0.169
14	0.242	0.270	0.354	0.213	0.242	0.197	0.205	0.170
15	0.251	0.275	0.357	0.220	0.248	0.198	0.207	0.170
16	0.262	0.281	0.361	0.226	0.254	0.199	0.209	0.171
17	0.271	0.287	0.364	0.233	0.260	0.200	0.211	0.172
18	0.282	0.293	0.368	0.240	0.266	0.201	0.213	0.173
19	0.293	0.299	0.372	0.247	0.273	0.202	0.215	0.173
20	0.304	0.306	0.376	0.254	0.280	0.203	0.217	0.174

Medium-speed parameter variations

S_2	Nonadaptive			Bandwidth adaptive		Bandwidth and order adaptive		
n/N	k_1	k_2	k_3	PPE	FPE	PPE	FPE	Mix
1	2.727	2.662	2.641	2.687	2.720	2.687	2.720	2.687
2	0.796	0.826	0.896	0.793	0.791	0.794	0.791	0.793
3	0.446	0.468	0.522	0.440	0.443	0.447	0.442	0.439
4	0.265	0.278	0.322	0.255	0.261	0.264	0.259	0.253
5	0.139	0.152	0.197	0.127	0.133	0.135	0.130	0.123
6	0.102	0.112	0.155	0.086	0.093	0.096	0.087	0.081
7	0.109	0.115	0.156	0.090	0.097	0.097	0.088	0.081
8	0.116	0.120	0.158	0.094	0.101	0.098	0.088	0.082
9	0.124	0.124	0.160	0.099	0.105	0.100	0.088	0.082
10	0.133	0.129	0.163	0.103	0.109	0.101	0.089	0.082
11	0.141	0.134	0.165	0.108	0.113	0.102	0.089	0.082
12	0.150	0.139	0.168	0.113	0.117	0.104	0.090	0.082
13	0.159	0.144	0.171	0.117	0.122	0.105	0.090	0.083
14	0.169	0.149	0.174	0.122	0.126	0.106	0.091	0.083
15	0.178	0.154	0.177	0.127	0.130	0.107	0.091	0.083
16	0.188	0.160	0.181	0.131	0.134	0.108	0.091	0.083
17	0.198	0.166	0.184	0.136	0.139	0.109	0.092	0.083
18	0.208	0.171	0.187	0.141	0.143	0.110	0.092	0.083
19	0.219	0.177	0.191	0.146	0.148	0.111	0.092	0.083
20	0.230	0.183	0.194	0.151	0.153	0.112	0.093	0.083

Slow parameter variations

S_3	Nonadaptive			Bandwidth adaptive		Bandwidth and order adaptive		
n/N	k_1	k_2	k_3	PPE	FPE	PPE	FPE	Mix
1	2.735	2.660	2.623	2.689	2.725	2.689	2.725	2.689
2	0.747	0.749	0.772	0.739	0.737	0.739	0.737	0.739
3	0.413	0.411	0.427	0.402	0.402	0.406	0.401	0.402
4	0.239	0.231	0.243	0.225	0.226	0.231	0.224	0.223
5	0.113	0.105	0.116	0.096	0.097	0.101	0.093	0.092
6	0.076	0.065	0.074	0.055	0.057	0.062	0.051	0.050
7	0.083	0.069	0.076	0.059	0.060	0.063	0.051	0.050
8	0.091	0.074	0.079	0.063	0.064	0.065	0.051	0.051
9	0.100	0.079	0.081	0.067	0.068	0.066	0.051	0.051
10	0.108	0.084	0.084	0.071	0.071	0.068	0.051	0.051
11	0.117	0.089	0.087	0.075	0.075	0.069	0.052	0.051
12	0.126	0.094	0.090	0.079	0.078	0.070	0.052	0.051
13	0.135	0.100	0.093	0.083	0.082	0.072	0.052	0.051
14	0.145	0.105	0.097	0.087	0.086	0.073	0.052	0.051
15	0.155	0.111	0.100	0.091	0.089	0.074	0.052	0.051
16	0.165	0.116	0.103	0.095	0.093	0.075	0.052	0.051
17	0.175	0.122	0.107	0.099	0.097	0.076	0.052	0.051
18	0.186	0.128	0.110	0.104	0.100	0.077	0.052	0.051
19	0.197	0.134	0.114	0.108	0.104	0.078	0.052	0.051
20	0.208	0.140	0.117	0.112	0.108	0.079	0.052	0.051