AUSTRIAN JOURNAL OF STATISTICS
Volume 35 (2006), Number 2&3, 371–378

Sparse Parameter Estimation in Overcomplete Time Series Models

Vítězslav Veselý and Jaromír Tonner
Masaryk University, Czech Republic

Abstract: We suggest a new approach to parameter estimation in time series models with large number of parameters. We use a modified version of the Basis Pursuit Algorithm (BPA) by Chen et al. [SIAM Review 43 (2001), No. 1] to verify its applicability to times series modelling. For simplicity we restrict to ARIMA models of univariate stationary time series. After having accomplished and analyzed a lot of numerical simulations we can draw the following conclusions: (1) We were able to reliably identify nearly zero parameters in the model allowing us to reduce the originally badly conditioned overparametrized model. Among others we need not take care about model orders the fixing of which is a common preliminary step used by standard techniques. For short time series paths (100 or less samples) the sparse parameter estimates provide more precise predictions compared with those based on standard maximum likelihood estimators from MATLAB's System Identification Toolbox (IDENT). For longer paths (500 or more) both techniques yield nearly equal prediction paths. (2) As the model usually depends on the estimated parameters, we tried to improve their accuracy by iterating BPA several times.

Keywords: Overcomplete Model, Algorithm.

1 Introduction

Chen, Donoho, and Saunders (1998) deal the problem of sparse representation of vectors (signals) by using special overcomplete (redundant) systems of vectors (atoms) spanning this space. Typically such systems (also called frames or dictionaries) are obtained either by refining existing basis or merging several such bases (refined or not) of various kind (so called packets).

In contrast with vectors which belong to a finite-dimensional space, Veselý (2002) formulates the problem of sparse representation within a more general framework of (even infinite dimensional) separable Hilbert space. Such functional approach allows us to get more precise representation of objects from such space which, unlike vectors, are by their nature not discrete.

In this paper we attack the problem of sparse representation from overcomplete time series models using expansions in the Hilbert space $L^2 := L^2(\Omega, \mathcal{A}, \mathcal{P})$ of random variables defined on the probability space $(\Omega, \mathcal{A}, \mathcal{P})$ and having finite variance. With complex scalars we have inner product defined by $\langle X, Y \rangle := \mathrm{E}X\overline{Y}$. A numerical study demonstrates benefits and limits of this approach when applied to overcomplete AR(I)MA models of univariate (covariance) stationary time series.

2 Recovering Sparse Parameter Estimates from Overcomplete Time Series Models

In this section we shall keep to the notation of Brockwell and Davis (1991), corrected second printing.

2.1 Overcomplete ARMA Model for Stationary Time Series

Let $H_t = \overline{\mathrm{sp}}(\{X_{t+1-j}\}_{j=1}^{\infty})$ be a separable closed space in L^2 spanned by the history of $\{X_t\}$ up to the time t and $P_t : L^2 \to H_t$ the orthogonal projection operator. By orthogonalization of $\{X_t\}$ we get also $H_t = \overline{\mathrm{sp}}(\{Z_{t+1-j}\}_{j=1}^{\infty})$ where $\{Z_t\}$ is uncorrelated, $Z_t = X_t - P_{t-1}X_t$, $t \in \mathbb{Z}$. We shall confine ourselves to $\{X_t\}$ zero-mean stationary with autocovariance function γ, $\{X_t\} \sim \mathit{ARMA}(p,q)$, in which case mean and variance are constant and $\{Z_t\}$ is a white noise, $\{Z_t\} \sim \mathit{WN}(0,\sigma^2)$, $\sigma > 0$. Thus both $\{X_{t+1-j}\}$ and $\{Z_{t+1-j}\}$ are dictionaries in H_t. Merging both dictionaries, we get a new overcomplete dictionary $\{U_{t+1-j}\}_{j=1}^{\infty} = \{X_{t+1-j}\}_{j=1}^{\infty} \cup \{Z_{t+1-j}\}_{j=1}^{\infty}$ in H_t. Fixing P, Q such that $0 \le p \le P \le \infty$, $0 \le q \le Q \le \infty$ we get an overcomplete but still finite atomic decomposition of $\widehat{X}_{t+}$:

$$\widehat{X}_{t+1} = P_t X_{t+1} = \sum_{j=1}^{P} \mathit{\Phi}_j X_{t+1-j} + \sum_{k=1}^{Q} \mathit{\Theta}_k Z_{t+1-k} =: T_t^{P,Q}\xi =: \sum_{i=1}^{P+Q} U_{t+1-i}\xi_i \,, \qquad (1)$$

with atoms $U_{t+1-j} := X_{t+1-j}$ for $j = 1, \dots, P$ and $U_{t+1-P-k} := Z_{t+1-k}$ for $k = 1, \dots, Q$ where $\xi := \{\Phi, \Theta\}$ stands for the corresponding concatenation of coefficient sequences $\Phi := \{\mathit{\Phi}_j\}_{j=1}^{P}$ and $\Theta := \{\mathit{\Theta}_k\}_{k=1}^{Q}$. Clearly $T_t := T_t^{P,Q} : \ell^2(J) \to H_t$, $J := \{1, \dots, P+Q\}$, is a bounded linear operator with closed range space $\mathcal{R}(T_t) = H_t$ of finite dimension. After changing the notation accordingly this model comprises all three commonly used representations, namely

- invertible representation $\widehat{X}_{t+1} = \sum_{j=1}^{\infty}(-\pi_j)X_{t+1-j} =: T_t^{\infty,0}(-\pi)$, when putting $\pi_0 = 1$ and $\pi_j = 0$ for $j < 0$:
  $Z_{t+1} = X_{t+1} - \widehat{X}_{t+1} = \sum_{j=0}^{\infty} \pi_j X_{t+1-j} = \sum_{j=-\infty}^{\infty} \pi_j X_{t+1-j}$;
- causal representation $\widehat{X}_{t+1} = \sum_{k=1}^{\infty} \psi_k Z_{t+1-k} =: T_t^{0,\infty}\psi$, when putting $\psi_0 = 1$ and $\psi_k = 0$ for $k < 0$:
  $X_{t+1} = \widehat{X}_{t+1} + Z_{t+1} = \sum_{k=0}^{\infty} \psi_k Z_{t+1-k} = \sum_{k=-\infty}^{\infty} \psi_k Z_{t+1-k}$;
- overcomplete $\mathit{ARMA}(P,Q)$ representation $\widehat{X}_{t+1} = T_t^{P,Q}\xi$ with finite but sufficiently overestimated orders P, Q: $p \le P < \infty$, $q \le Q < \infty$; the choice $P = Q = 10$ being satisfactory in most cases.

Hereafter we shall deal with the third case in more detail, the sparse solution of which is expected to exclude redundant parameters which are nearly noughts allowing us to approach the original $\mathit{ARMA}(p,q)$ model and its parameter estimates.

As $\mathcal{R}(T_t) = H_t$ is closed, the restriction of adjoint operator T_t^* onto H_t is a topological linear isomorphism T_t^* of H_t onto a closed subspace $H_t' \subseteq \ell^2(J)$, $\dim H_t = \dim H_t'$. Thus instead of (1) we can solve the underdetermined system of $M := P+Q$ linear equations (analogy to normal equations known from linear regression) obtained by applying

T_t^* to both sides of (1):

$$b_t = R_t \xi \qquad \text{or equivalently} \quad b_i(t) = \sum_{j=1}^{M} R_{ij}(t)\xi_j \quad \text{for } i \in J\,. \tag{2}$$

In view of $T_t^*(X) = \{\langle X, U_{t+i-1}\rangle\}_{i=1}^M$, $X \in L^2$, above there is $b_t = [b_i(t)]_{i=1}^M$ where $b_i(t) := \langle X_{t+1}, U_{t+1-i}\rangle = \mathrm{E}\, X_{t+1}\overline{U}_{t+1-i}$ is standing for 2-nd order joint moment of X_{t+1} and i-th atom U_{t+1-i}, and $R_t = [R_{ij}(t)]_{i,j=1}^M$ with $R_{ij}(t) := \langle U_{t+1-j}, U_{t+1-i}\rangle = \mathrm{E}\, U_{t+1-j}\overline{U}_{t+1-i}$ standing for 2-nd order joint moment of j-th and i-th atom which is a covariance of them due to zero mean.

Lemma 1 *Let $\{X_t\}$ be a stationary time series and $i, j \in \mathbb{Z}$ arbitrary. The following holds:*

(1) If $\{X_t\}$ is causal then $\langle X_{t+1-j}, Z_{t+1-i}\rangle = \sigma^2\psi_{i-j}$.

(2) If $\{X_t\}$ is invertible then $\langle X_{t+1-j}, Z_{t+1-i}\rangle = \sum_{k=0}^{\infty}\gamma(i-j+k)\overline{\pi}_k$.

Proof.
(1) Causal representation substituted for X_{t+1-j} yields
$\langle X_{t+1-j}, Z_{t+1-i}\rangle = \langle \sum_{k=0}^{\infty}\psi_k Z_{t+1-j-k}, Z_{t+1-i}\rangle = \sum_{k=0}^{\infty}\psi_k\langle Z_{t+1-j-k}, Z_{t+1-i}\rangle = \sum_{k=0}^{\infty}\psi_k\sigma^2\delta_{i,j+k} = \sigma^2\psi_{i-j}$.
(2) Invertible representation substituted for Z_{t+1-i} yields
$\langle X_{t+1-j}, Z_{t+1-i}\rangle = \langle X_{t+1-j}, \sum_{k=0}^{\infty}\pi_k X_{t+1-i-k}\rangle = \sum_{k=0}^{\infty}\overline{\pi}_k\langle X_{t+1-j}, X_{t+1-i-k}\rangle = \sum_{k=0}^{\infty}\overline{\pi}_k\gamma(t+1-j-(t+1-i-k)) = \sum_{k=0}^{\infty}\overline{\pi}_k\gamma(i-j+k)$. □

The next theorem reveals the entries $b_i := b_i(t)$, $i = 1, \ldots, P+Q$, of the left- hand-side vector $\boldsymbol{b} := b_t$ in (2) and the structure of the matrix $\boldsymbol{R} := (T_t)^* T_t = [R_{ij}]$ which show to be independent of t due to stationarity. That is why we have omitted the subscript t from the notation.

Theorem 1 *If $\{X_t\} \sim ARMA(p,q)$ is zero- mean and causal with autocovariance function $\gamma = \{\gamma(h)\}_{h=0}^{\infty}$, $\gamma(h) := cov(X_{t+h}, X_t) = EX_{t+h}\overline{X}_t = \langle X_{t+h}, X_t\rangle$, then the equation* (2) *attains with $0 \le p \le P < \infty$ and $0 \le q \le Q < \infty$ the form*

$$\boldsymbol{b} = \boldsymbol{R}\boldsymbol{\xi} \quad \text{with} \quad \boldsymbol{b} = \begin{bmatrix} \boldsymbol{\gamma}_P \\ \sigma^2\boldsymbol{\psi}_Q \end{bmatrix},\ \boldsymbol{R} = \begin{bmatrix} \boldsymbol{\Gamma}_P & \sigma^2\boldsymbol{\Psi}^* \\ \sigma^2\boldsymbol{\Psi} & \sigma^2\boldsymbol{I}_Q \end{bmatrix} \quad \text{and} \quad \boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\Phi}_P \\ \boldsymbol{\Theta}_Q \end{bmatrix}, \tag{3}$$

where $\boldsymbol{\gamma}_P := [\gamma(1), \ldots, \gamma(P)]^T$, $\boldsymbol{\psi}_Q := [\psi(1), \ldots, \psi(Q)]^T$, $\sigma^2 = \gamma(0)/\sum_{k=0}^{\infty}|\psi_k|^2 = \gamma(0)/\|\psi\|^2$, $\boldsymbol{\Phi}_P = [\Phi_1, \ldots, \Phi_P]^T$, and $\boldsymbol{\Theta}_Q = [\Theta_1, \ldots, \Theta_Q]^T$.
$\boldsymbol{I}_Q$ is identity matrix of order Q, $\boldsymbol{\Gamma}_P$ and $\boldsymbol{\Psi}$ are Toeplitz matrices:

$$\boldsymbol{\Gamma}_P := [\gamma(i-j)]_{i,j=1}^P = \begin{bmatrix} \gamma(0) & \overline{\gamma(1)} & \cdots & \overline{\gamma(P-1)} \\ \gamma(1) & \gamma(0) & \cdots & \overline{\gamma(P-2)} \\ \vdots & \vdots & \cdots & \vdots \\ \gamma(P-1) & \gamma(P-2) & \cdots & \gamma(0) \end{bmatrix} \quad \text{and} \tag{4}$$

$$\boldsymbol{\Psi} := [\psi(i-j)]_{i,j=1}^{Q,P} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \psi(1) & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ \psi(Q-1) & \psi(Q-2) & \cdots & \cdot \end{bmatrix} \text{ of size } Q \times P. \tag{5}$$

Proof. As $U_{t+1-j} = X_{t+1-j}$ for $j = 1, \ldots, P$ and $U_{t+1-P-k} = Z_{t+1-k}$, then in view of (2) we get for $\boldsymbol{R} =: [R_{ij}] =: \left[\begin{smallmatrix} \boldsymbol{\Gamma}_P & \boldsymbol{W}^* \\ \boldsymbol{W} & \boldsymbol{V}_Q \end{smallmatrix}\right]$:
$\boldsymbol{\Gamma}_P =: [R_{ij}]_{i,j=1}^P$ where $R_{ij} = \langle U_{t+1-j}, U_{t+1-i}\rangle = \langle X_{t+1-j}, X_{t+1-i}\rangle = \gamma(t+1-j-(t+1-i)) = \gamma(i-j) = \overline{\gamma(j-i)}$;
$\boldsymbol{V}_Q = [R_{P+i,P+j}]_{i,j=1}^Q$ where $R_{P+i,P+j} = \langle U_{t+1-P-j}, U_{t+1-P-i}\rangle = \langle Z_{t+1-j}, Z_{t+1-i}\rangle = \sigma^2\delta_{ij}$ which implies $\boldsymbol{V}_Q = \sigma^2\boldsymbol{I}_Q$;
$\boldsymbol{W} = [R_{P+i,j}]_{i,j=1}^{Q,P}$ where $R_{P+i,j} = \langle U_{t+1-j}, U_{t+1-P-i}\rangle = \langle X_{t+1-j}, Z_{t+1-i}\rangle = \sigma^2\psi_{i-j}$ in view of lemma 1(1), which implies $\boldsymbol{W} = \sigma^2\boldsymbol{\Psi}$.
Clearly the above formulas remain valid for $R_{i0} = \langle X_{t+1}, U_{t+1-i}\rangle$ as well when extending the scope of column index by $j = 0$. Then we have also
$b_i = \langle X_{t+1}, U_{t+1-i}\rangle = \langle X_{t+1}, X_{t+1-i}\rangle = R_{i0} = \gamma(i)$ for $i = 1, \ldots, P$;
$b_{P+i} = \langle X_{t+1}, U_{t+1-P-i}\rangle = \langle X_{t+1}, Z_{t+1-i}\rangle = R_{P+i,0} = \sigma^2\psi_i$ for $i = 1, \ldots, Q$.
The relation for σ^2 is well- known (Brockwell and Davis, 1991, eq.(3.2.4)) and easily derived from
$\gamma(0) = \langle X_t, X_t\rangle = \langle\sum_{j=0}^\infty \psi_j Z_{t-j}, \sum_{j=0}^\infty \psi_k Z_{t-k}\rangle = \sum_{j,k=0}^\infty \psi_j\overline{\psi}_k\langle Z_{t-j}, Z_{t-k}\rangle = \sum_{j,k=0}^\infty \psi_j\overline{\psi}_k\sigma^2\delta_{jk} = \sigma^2\sum_{j=0}^\infty|\psi_j|^2$. □

Corollary 1 *If the time series $\{X_t\}$ from theorem 1 is both causal and invertible, then the entries of $\boldsymbol{\psi}_Q$ and $\boldsymbol{\Psi}$ may be evaluated from the invertible representation too:*

$$\psi_i = \frac{1}{\sigma^2}\sum_{k=0}^\infty \gamma(i+k)\overline{\pi}_k \quad \textit{for} \quad i = 0, 1, \ldots \quad \textit{where} \tag{6}$$

$$\sigma^2 = \sum_{k=0}^\infty \gamma(k)\overline{\pi}_k = \langle\gamma, \pi\rangle \quad \textit{taking the scalar product in } \ell^2. \tag{7}$$

Proof. Equating both relations for $\langle X_{t+1-j}, Z_{t+1-i}\rangle$ in lemma 1 with $j = 0$, we get immediately $\sigma^2\psi_i = \sum_{k=0}^\infty \gamma(i+k)\overline{\pi}_k$ and the relation for σ^2 as its special case with $i = 0$ due to $\psi_0 = 1$. □

2.2 Algorithm for Sparse Parameter Estimation

A lot of algorithms have been suggested by various authors (see Chen et al., 1998) for searching sparse representations from the overcomplete ones. In this paper we use a computationally intensive universal multi- stage iterative procedure coded in MATLAB which shows to be robust against propagation of numerical errors when solving inverse problems being extremely badly conditioned. The procedure is based on BPA (Basis Pursuit Algorithm) originally suggested by Chen et al. (1998) for finite- dimensional vectors and later on extended to functional setting by Veselý (2002). The main steps of the algorithm applied to the solution of (2) are [see also Zelinka et al., 2004]:

(A0) Choosing a raw initial estimate $\boldsymbol{\xi}^{(0)}$.

(A1) We improve $\boldsymbol{\xi}^{(0)}$ iteratively by stopping at $\boldsymbol{\xi}^{(1)}$ which satisfies optimality criterion $\|\boldsymbol{b}-\widehat{\boldsymbol{b}}\| \to \min$ within numerical precision $\|\widehat{\boldsymbol{b}}-\boldsymbol{R}\boldsymbol{\xi}^{(1)}\| < \varepsilon/2$, $\widehat{\boldsymbol{b}} := P_{H_t'}\boldsymbol{b}$. Solution $\boldsymbol{\xi}^{(1)}$ is ε- suboptimal but not sparse in general.

(A2) Starting with $\boldsymbol{\xi}^{(1)}$ we are looking for
$\boldsymbol{\xi}^{(2)} = \operatorname{argmin}_{\boldsymbol{\xi}\in\ell^2(J)}\|\boldsymbol{\xi}\|_{\boldsymbol{w},1}$ subject to $\|\boldsymbol{R}\boldsymbol{\xi}^{(1)} - \boldsymbol{R}\boldsymbol{\xi}^{(2)}\| < \varepsilon/2$, which tends to be nearly sparse and is ε- suboptimal due to triangle inequality $\|\widehat{\boldsymbol{b}} - \boldsymbol{R}\boldsymbol{\xi}^{(2)}\| \leq \|\widehat{\boldsymbol{b}} - \boldsymbol{R}\boldsymbol{\xi}^{(1)}\| + \|\boldsymbol{R}\boldsymbol{\xi}^{(1)} - \boldsymbol{R}\boldsymbol{\xi}^{(2)}\| < \varepsilon/2 + \varepsilon/2 = \varepsilon$; $\|\boldsymbol{\xi}\|_{\boldsymbol{w},1} := \sum_{j=1}^{P+Q} w_j|\xi_j|$ is ℓ^1- norm of coefficients weighted in order to balance nonuniform norms of atoms: $w_j = \sqrt{\gamma(0)}$ for $j = 1,\ldots,P$ and $w_j = \sigma$ for $j = P+1,\ldots,P+Q$.

(A3) We construct a sparse and ε- suboptimal solution $\boldsymbol{\xi}^* := \{\xi_j^{(2)}\}_{j\in F^*}$ by choosing zero threshold $\delta > 0$ as large as possible such that $\|\widehat{\boldsymbol{b}} - \boldsymbol{R}\boldsymbol{\xi}^*\| < \varepsilon$ still holds with $F^* = \{\, j \in J \,|\, |\xi_j^{(2)}| \geq \delta\}$.

(A4) Optionally we can repeat step (A1) with J replaced by significantly reduced F^* and new initial estimate $\boldsymbol{\xi}^{(0)} = \boldsymbol{\xi}^*$ from the previous step (A3). We expect to obtain a possibly improved sparse representation $\boldsymbol{\xi}^*$.

Hereafter we refer to this four-step algorithm as to BPA4.

The overall estimation procedure is as follows:

(1) Replace exact autocovariance function γ by its sample estimate $\widehat{\gamma}$.

(2) Compute estimates $\widehat{\psi}$ and $\widehat{\sigma}^2$ from $\widehat{\gamma}$ via iterating *Innovations algorithm* (IA) sufficiently many times (Brockwell and Davis, 1991, §8.3) until $\widehat{\sigma}^2$ and $\widehat{\psi}_i$, $i = 1,\ldots,Q$ are stabilized.

(3) Compute sparse solution $\boldsymbol{\xi}^* = \left[\begin{smallmatrix}\boldsymbol{\Phi}_P^*\\ \boldsymbol{\Theta}_Q^*\end{smallmatrix}\right]$ of eq. (3) via BPA4 with γ, ψ and σ^2 replaced by their estimates. As the initial estimate in step (A0) we can use for example the pseudoinverse solution $\boldsymbol{\xi}^{(0)} = \boldsymbol{R}^+\boldsymbol{b}$.

(4) Optional step. We know from the causal representation (Brockwell and Davis, 1991, eq. (3.3.5)) and from (6) that both ψ and σ^2 are functions of unknown parameters $\boldsymbol{\Phi}_P$ and $\boldsymbol{\Theta}_Q$. Therefore we can reestimate $\widehat{\psi}$ on the basis of sparse solution obtained in step (3) utilizing Lemma 1(1) where we substitute sample estimates. Let $\{x_t\}_{t=1}^n$ denote sample path of $\{X_t\}$ and $\{z_t\}_{t=1}^n$ errors of one- step predictions from the model estimated in step (3). Putting $t = n-1$ and $j = 0$ in Lemma 1(1) we arrive at a possibly improved estimate of ψ:
$\widehat{\psi}_i = \langle X_n, Z_{n-i}\rangle/\sigma^2 \approx \left(\sum_{m=0}^{n-i-1} x_{n-m}\overline{z}_{n-i-m}\right)/\left(\sum_{m=0}^{n-i-1}|z|_{n-i-m}^2\right)$.
This procedure may be iterated several times. If convergence is exhibited then the stabilized solution $\boldsymbol{\Phi}_P^*$ and $\boldsymbol{\Theta}_Q^*$ from the last iteration will be used as the final parameter estimate, otherwise we keep the initial solution $\boldsymbol{\xi}^*$ from step (3).

3 Design of the Numerical Simulation Study

- simulated lengths: 500 and 100 samples ($\boldsymbol{x}$), out of which the leading 300 and 80 ($\boldsymbol{x}_m$), respectively, have been used for parameter estimation; the remaining 200 and 20 for verification ($\boldsymbol{x}_v$);
- simulations were done for several $AR(I)MA(p,q)$ models with varying orders and parameter vectors $\boldsymbol{\Phi}_P$ and $\boldsymbol{\Theta}_Q$ (see Tables below);
- 100 simulations were carried out for every pair (length,model);

- for each simulation the command `predict` from MATLAB's *System Identification toolbox* (IDENT) designed by Ljung (2002) was used to compute one-step predictions on $\boldsymbol{x}_v$ based on exact parameters and on four different estimation techniques:
  (1) single sparse using BPA4: steps (1)–(3) of the algorithm from section 2.2,
  (2) iterated sparse using BPA4:steps (1)–(4) of the algorithm from section 2.2,
  (3) iterated sparse using Moore- Penrose pseudoinverse: the same as (2) except that (A0) was used instead of BPA4, and
  (4) maximum likelihood (ML) estimate using `armax` function from IDENT;
- for every triple (length,model,simulation) the quality of the prediction was evaluated using function `compare` from IDENT:
  (1) standard deviation of one- step prediction errors,
  (2) the percentage of the measured output $\boldsymbol{x}_v$ that was explained by the model.
  Their mean with sample std in parantheses were summarized in Table 1.

Table 1: Summary of simulation results for the ARMA model

Type of estimate for ARMA		$\Phi = 0.50 - 0.80$ $\Theta = 0.60$ $\sigma = 1.50$ $n = 500$	$0.50 - 0.80$ 0.60 1.50 100	$0.90 - 0.80$ 0.60 1.50 500	$0.90 - 0.80$ 0.60 1.50 100 100
Single sparse	$\widehat{\sigma}$	1.629 (0.342)	1.029 (0.257)	1.612 (0.191)	1.049 (0.295)
LSQ est.	%	47.129 (2.704)	66.536 (9.929)	59.565 (6.367)	72.567 (7.689)
Iter. sparse	$\widehat{\sigma}$	1.533 (0.139)	1.112 (0.283)	1.555 (0.091)	1.230 (0.319)
LSQ est.	%	49.805 (3.427)	63.917 (10.628)	61.080 (4.010)	67.285 (11.636)
Iter. sparse	$\widehat{\sigma}$	1.470 (0.109)	1.494 (0.086)	1.498 (0.070)	1.279 (0.788)
inv. matrix	%	51.694 (4.227)	59.518 (18.139)	63.106 (3.399)	62.256 (17.059)
IDENT	$\widehat{\sigma}$	1.484 (0.102)	1.441 (0.243)	1.492 (0.069)	1.451 (0.268)
max. lik.	%	51.290 (4.450)	53.392 (11.175)	62.634 (3.537)	59.591 (13.781)
Exact	$\widehat{\sigma}$	1.483 (0.081)	1.407 (0.232)	1.486 (0.061)	1.429 (0.258)
	%	52.079 (4.503)	54.437 (11.374)	62.881 (3.813)	60.226 (13.581)
Type of estimate		$\Phi = 0.30$ $\Theta = 0.70$ 0.40 $\sigma = 1.50$ $n = 500$	0.30 0.70 0.40 1.50 100	$1.20 - 0.80$ 0.60 1.50 500	$1.20 - 0.80$ 0.60 1.50 100 100
Single sparse	$\widehat{\sigma}$	1.477 (0.080)	1.070 (0.258)	1.659 (0.390)	1.071 (0.313)
LSQ est.	%	37.158 (4.151)	46.570 (13.108)	65.648 (8.187)	76.213 (7.299)
Iter. sparse	$\widehat{\sigma}$	1.471 (0.078)	1.083 (0.266)	1.643 (0.0972)	1.249 (0.395)
LSQ est.	%	37.438 (3.715)	45.777 (14.412)	65.363 (3.6007)	71.397 (11.138)
Iter. sparse	$\widehat{\sigma}$	1.474 (0.076)	1.082 (0.290)	∞ (∞)	∞ (∞)
inv. matrix	%	37.315 (3.756)	45.707 (14.695)	0 (∞)	0 (∞)
IDENT	$\widehat{\sigma}$	1.492 (0.080)	1.433 (0.312)	1.496 (0.075)	1.426 (0.247)
max. lik.	%	36.560 (3.705)	29.422 (13.862)	68.495 (3.339)	67.779 (8.165)
Exact	$\widehat{\sigma}$	1.481 (0.084)	1.417 (0.265)	1.473 (0.073)	1.396 (0.237)
	%	36.593 (3.819)	31.312 (12.176)	69.523 (3.360)	68.882 (8.092)

4 Conclusions

From the table and other experiments we can draw the following conclusions:

- For larger sample sizes (roughly $n > 500$) the ML-estimate from IDENT and the sparse estimate produce practically equal predictions even though the parametrizations of both estimates are typically quite different.
- With decreasing sample size the sparse estimate tends to be superior to the ML-estimate from IDENT as to the precision of predictions.
- The parametrization obtained from a sparse estimate in an overcomplete model cannot be used as an estimate of the parameters in the ideal ARMA model because it is related exclusively to the particular sample path. It sometimes produces even better predictions than the ideal model.
- On the other hand the prediction of any sample path from the model works with parametrization obtained from any other path with any $P \geq p$ and $Q \geq q$ for the same sample size (short or long). This seems to confirm that our procedure constructs a new type of estimator for the time series itself not just an estimate of one particular path.
- The number of significant parameters in the sparse parametrization rarely exceeds the number of parameters from the ideal ARMA model, it happens very often that it is smaller.
- Iterated sparse based on BPA4 preserves the quality of predictions and mostly reduces their uncertainty compared with single sparse.
- Iterated sparse based on MP-pseudoinverse behaves similarly but sometimes fails to converge, which is probably a consequence of higher sensitivity to round- off errors coming from the pseudoinverse of a matrix having accidentally a wrong condition number.

Table 2: Summary of selected simulation results for the ARIMA model

Type of estimate for ARIMA		$\Phi = 0.50 - 0.20$ $\Theta = 0.60$ $\sigma = 1.50$ $[n, D] = 100\ 1$	$0.50 - 0.20$ 0.60 1.50 100 2	$1.20 - 0.80$ 0.60 1.50 100 1	$1.20 - 0.80$ 0.60 1.50 100 2
Single sparse	$\hat{\sigma}$	2.097 (1.236)	8.545 (7.892)	1.649 (0.734)	7.133 (6.002)
LSQ. est.	%	48.030 (13.401)	50.972 (13.095)	76.150 (6.756)	75.557 (8.310)
IDENT	$\hat{\sigma}$	2.876 (1.578)	17.844 (12.569)	2.300 (1.038)	18.250 (12.140)
max. lik.	%	29.748 (14.045)	30.489 (16.822)	66.263 (9.445)	65.612 (9.568)
Exact	$\hat{\sigma}$	2.788 (1.334)	17.196 (11.659)	2.243 (0.981)	18.444 (12.776)
	%	31.465 (11.970)	31.902 (15.620)	66.207 (9.557)	67.031 (9.156)

Analogical numerical study for ARIMA models leads to the same conclusions (see Table 2). This is in accordance with our expectation because predictions in ARIMA model are derived from ARMA- predictions of differenced time series.

Standard statistical techniques numerically stabilize the parameter estimation in an *ARMA*(p, q) model assuming low orders p, q. Our procedure is relaxing this assumption assuming low number of parameters within a possibly higher range of orders. This is less restrictive which may explain better precision of predictions in short time series where there is not enough information inherent in the data to confirm the low order assumption. That is why in situations where one cannot derive the low order assumption from an a priori knowledge, the usage of our technique should be preferred.

A continued research is planned for VARMA where the BPA based technique is promising in revealing sparse structure of parameter matrices which are commonly estimated as being full which, of course, may deteriorate the stability and reliability of estimates and predictions.

Acknowledgements

Vítězslav Veselý was supported by MŠMT ČR under research contract MSM0021622418 and Jaromír Tonner was supported by the Faculty of Economics and Administration of the Masaryk University (Brno, CZ) within PhD research project No. 56 1403.

References

Brockwell, P. J., and Davis, R. A. (1991). *Time series: Theory and Methods* (2-nd ed.). New York: Springer-Verlag.

Chen, S. S., Donoho, D. L., and Saunders, M. A. (1998). Atomic decomposition by basis pursuit. *SIAM Journal of Scientific Computing*, *20*, 33-61. (reprinted in SIAM Review, **43** (2001), no. 1, pp. 129-159)

Ljung, L. (2002, July). *System identification toolbox for use with MATLAB – User's guide.* 24 Prime Park Way, Natick, Mass. 01760.

Veselý, V. (2002). Hilbert-space techniques for spectral representation in terms of overcomplete bases. In I. Horová (Ed.), *Proceedings of the Summer School DATASTAT'2001, Čihák near Žamberk* (Vol. 11, p. 259-273). Masaryk University, Brno (Czech Rep.)[1].

Zelinka, J., Veselý, V., and Horová, I. (2004). Comparative study of two kernel smoothing techniques. In I. Horová (Ed.), *Proceedings of the Summer School DATASTAT'2003, Svratka* (Vol. 15, p. 419-436). Masaryk University, Brno (Czech Rep.).

Authors' address:

Vítězslav Veselý and Jaromír Tonner
Department of Applied Mathematics and Computer Science
Faculty of Economics and Administration, Masaryk University
Lipová 41a
CZ-602 00 Brno
Czech Republic
E-mail: vesely@econ.muni.cz and jtonner@tiscali.cz
http://www.math.muni.cz/˜vesely