An Empirical Study for the Estimation of Autoregressive Hilbertian Processes by Wavelet Packet Method

A. Laukaitis

Vilnius Management Academy
J.Basanavičiaus str. 29a, LT-03109 Vilnius, Lithuania
algirdas@vva.lt

Abstract.   In this paper wavelet packet bases are used for an estimation of the autoregressive Hilbertian processes operator. We assume that integral operator kernel can have some singular structures and estimate them by projecting functional processes on suitable bases. Linear methods for continuous-time prediction using Hilbert-valued autoregressive processes are compared with the suggested method on simulated data and on real-life data sets. Statistics of residual partial sums processes and *Ex poste* prediction are used to check the model.

Keywords: autoregressive Hilbertian process, functional data analysis, wavelet packet bases, ill-posed inverse problem, residual partial sums processes.

1   Introduction

In many real life applications, the data and the model can be structured in several ways. In some time series it is possible to interpret the observed data as the realisation of the functional process. Recently there has been much interest in the possibility of using autoregressive Hilbertian model to predict the weather [1], harmonic levels in electrical networks [2] or cash flow in automatic teller machine networks [3]. In all the mentioned papers, the projection on the finite number of principal components of the empirical covariance operator has been used to predict future functional observations. A comprehensive theoretical study of this method has been presented by Bosq [4].

Fourier transform and spline bases have been used in the above-mentioned papers. The authors used a rule of thumb to determine the number of bases. Further multivariate principal components analysis and cross-validation criteria have been used to reduce the dimensionality of the space.

Wavelet bases can be used instead of Fourier transform or spline bases for a less regular functions. Antoniadis and Sapatinas [5] in their paper proposed to exploit some wavelet bases regularity properties by suggesting three linear methods using wavelet
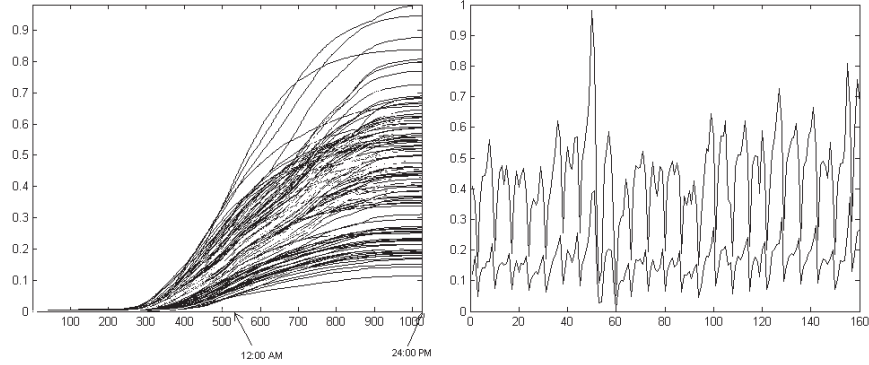
Fig. 1. Bivariate time series.

regularisation techniques. Compared with spline bases, these methods showed a similar prediction errors on climatological time series.

In this paper we are concerned with the autoregressive Hilbertian process that has some singular structures in its integral kernel. For the Gaussian functional noise and autoregressive operator with some singularities, it can be assumed that the covariance operator principal components bases are not the best one for the estimation of the model parameters.

In order to deal with the autoregressive operator singularities we suggest to use wavelet packet bases (WPB) so that the most suitable bases from the dictionary of admissible bases can be found.

Recently, the best basis methods have appeared in a variety of papers in the area of image compression and denoising (see [6] for review). The arguments presented by these papers in favor of the wavelet packet bases can be applied to functional data with some singular structures as well.

The rest of the article is organised as follows. The functional autoregressive process and autoregressive operator kernel with singular structures are introduced in Section 2. The residual partial sums processes and a certain class of functionals of the residuals partial sums are introduced. These functionals serve as tests of stability and as a criterion for parameters estimation. The section concludes with the simulation results, which show that the proposed method improves the prediction results for a certain class of functional processes.

Section 3 presents the standard and non-standard forms of operators decomposition and discusses their extensions with the wavelet packet dictionaries. Beylkin *et al.* [7, 8] have demonstrated that some operator can be presented in the non-standard form as the sparse matrix and there exist algorithms that apply those operators in order of $O(N)$ operations. The estimation of the operator of a first-order functional autoregressive process based on the wavelet packet transformation is presented in Section 4. The uniform moving residual sums statistics are used to test the stability and suitability of the model.

The real life application of the use of the wavelet packet transformation is presented in Section 5. The final section concludes with suggestions for future research.

2 Autoregressive Hilbertian and residual partial sums processes

Let H denote the Hilbert space $L_2(0, 1)$ with the norm $\|x\|^2 = \int_0^1 x^2(t)dt$ and with the inner product $\langle x, y \rangle = \int_0^1 x(t)y(t)dt$. A sequence $(\xi_i = \xi_i(t), t \in [0, 1]; 0 < i \leq N)$ of random variables with values in H is said to follow a Hilbert space valued autoregressive process of the first order (ARH(1)) associated with (ε, ρ) if it is stationary and such that

$$\xi_i = \rho(\xi_{i-1}) + \varepsilon_i, \tag{1}$$

where $(\varepsilon_i(t), t \in [0, 1]; 0 < i \leq N)$ is an H white noise, and operator $\rho \colon H \to H$ is linear and compact [9].

Row residuals $(\widehat{\varepsilon}_i)$ of the model (1) are defined in the usual way by

$$\widehat{\varepsilon}_i = \xi_i - \widehat{\rho}(\xi_{i-1}),$$

where $\widehat{\rho}$ is an estimate of the operator ρ based on $\xi_1, \ldots, \xi_N$.

In order to test the model (1) for stability and suitability we consider the process

$$\widehat{\xi}_N(s) = N^{-1/2}\big(\widehat{S}(\lfloor Ns \rfloor) + (Ns - \lfloor Ns \rfloor)\widehat{\varepsilon}_{\lfloor Ns \rfloor + 1}\big), \quad 0 \leq s \leq 1,$$

where $\widehat{S}(0) = 0$ and

$$\widehat{S}(i) = \sum_{j \leq i} \widehat{\varepsilon}_j, \quad \text{for } 0 < i \leq N.$$

Residual partial sums process weak convergence to a Brownian motion with values in the space H and in stronger topological framework e.g. Hölder space has been investigated in the paper by Laukaitis and Rackauskas [10]. To test ARH(1) model stability, they suggested to use the uniform moving residual sums statistics

$$\widehat{\mathrm{MS}}(N, \alpha) = \max_{0 \leq m < N} (m^{-\alpha}) \max_{1 \leq k \leq N-m} \big\|\widehat{S}(k+m) - \widehat{S}(k) - \frac{m}{N}\widehat{S}(N)\big\|,$$

where $0 < \alpha < 1$. The dependency of the statistic $\widehat{\mathrm{MS}}(N, \alpha)$ from the choice of Hölder parameter α has been presented at [10]. The authors demonstrated that statistics $\widehat{\mathrm{MS}}(N, \alpha)$ is more sensitive to the changes of (1) when Hölder exponent approaches $1/2$.

In this paper the statistic $\widehat{\mathrm{MS}}(N, \alpha)$ was used to test if the data are consistent with the model (1). Our method is based on the choice of the best wavelet basis and this statistic is used to test how well the model fits the data for each wavelet basis.

Let $\beta \colon [0, 1]^2 \to R$ satisfies $\int_0^1 \int_0^1 \beta^2(s, t)dsdt < \infty$. Consider the operator $\rho \colon H \to H$ defined by the kernel β

$$\rho(x) = \int_0^1 \beta(s, t)x(t)dt, \quad \text{for } s \in [0, 1]. \tag{2}$$

We now consider three special examples of the simulation that explains the motivation behind this work.

1. The two randomly generated Dirac basis (peaks) on the surface of the kernel β (one peak on the diagonal). In that case, the autoregressive operator will have the form

$$\rho(x) = A\langle x, v_1\rangle v_1 + B\langle x, v_3\rangle v_2, \tag{3}$$

   where $\{v_1, v_2, v_3\}$ is an orthonormal system in H and $0 < A < 1$ and $0 < B < 1$.

2. As the second example the kernels with singularities of the form $\log|t - s|$ are considered. Applications of such operators have been analyzed by Bradley [11]. For the simulation studies the following operator has been used

$$\beta(s,t) = \begin{cases} -\log|t-s| - 2\log\big(-|t-s|+1\big), & |t-s| \neq 0 \text{ or } |t-s| \neq -1, \\ 100, & |t-s| = 0 \text{ or } |t-s| = -1. \end{cases} \tag{4}$$

3. The Gaussian kernel as an example of a smooth kernel function

$$\beta(s,t) = K \exp\big\{(s^2 + t^2)/2\big\}, \quad s, t \in [0,1]. \tag{5}$$

In addition to the wavelet packet basis approach, two methods presented by Bosq [4] and Antoniadis *et al.* [5] are used in the simulation.

The estimation of the operator ρ is ill-posed. In order to handle it Bosq [4] suggested the projection method on the span of principal components of the covariance operator. Let $\widehat{\Gamma}_0 = \frac{1}{N} \sum_{i=1}^{N} \xi_i \otimes \xi_i$ and $\widehat{\Gamma}_1 = \frac{1}{N} \sum_{i=1}^{N} \xi_i \otimes \xi_{i+1}$ be the empirical covariance and cross-covariance operators and let, H_{i_N} be the span of k_N eigenvectors of $\widehat{\Gamma}_0$ associated with the largest eigenvalues, and let π_{k_N} be the orthogonal projector on this subspace. The $u \otimes v$ defines the tensor product for two fixed elements $u, v \in H$ and is the bounded linear operator form H to H, defined by

$$x \in H \mapsto (u \otimes v)(x) = \langle u, x\rangle v.$$

Let us define the regularized covariance and cross-covariance estimates as follows: $\widetilde{\Gamma}_0 = \pi_{k_N} \widehat{\Gamma}_0 \pi'_{k_N}$ and $\widetilde{\Gamma}_1 = \pi_{k_N} \widehat{\Gamma}_1 \pi'_{k_N}$. The regularized estimator will be

$$\widetilde{\rho} = \pi_{k_N} \widetilde{\Gamma}_1 \widetilde{\Gamma}_0^{-1} \pi'_{k_N}. \tag{6}$$

Under certain assumptions on the covariance operator, this estimator is consistent(for details see [4]). For less regular space Antoniadis *et al.* [5] proposed the method based on the wavelet-vaguelette decomposition and which is defined as follows

$$\widehat{\rho\xi_i} = \sum_{k=1}^{2^{j_0}} \langle \xi_i, \rho^* \widehat{\phi_{j_0 k}}\rangle \phi_{j_0 k} \tag{7}$$

and

$$\rho^* \widehat{\phi_{j_0 k}} = (\widehat{\Gamma}_0^* \widehat{\Gamma}_0 + \lambda I)^{-1} \widehat{\Gamma}_0^* \widehat{\Gamma}_1^* \phi_{j_0 k},$$

where $\widehat{\Gamma}_0^*$ and $\widehat{\Gamma}_1^*$ are the adjoint operator of $\widehat{\Gamma}_0$ and $\widehat{\Gamma}_1$.

Autoregressive operator estimation based on the Wavelet-vaguelette decomposition shows similar prediction accuracy as the covariance operator principal component projection method [5]. A recent study from Laukaitis and Rackauskas [3] confirms those findings.

Table 1 presents our simulation results for the model (1). We assume that the autoregressive operator has one of the three forms presented above. Wiener process is used as the noise component ε_i in our simulations. The methods of the principal components (6), the wavelet-vaguelette decomposition (7) and the wavelet packet bases (see Section 4) are compared by increasing the number of samples from 10^2 to 10^5 and calculating the mean quadratic errors of the forecasted variables. All simulation has been carried in the MATLAB environment.

Table 1. Mean quadratic errors for one-step-ahead forecast of the ξ in model (1)

Model (3)	10^2	10^3	10^4	10^5
PCA	0.09559	0.00398	0.001350	0.00104
Vaguelette	0.09130	0.00348	0.001449	0.00101
WPB	0.00140	0.00106	0.00096	0.00091
Model (4)				
PCA	0.1302	0.06796	0.03352	0.03122
Vaguelette	0.1334	0.05279	0.03218	0.03174
WPB	0.1103	0.05072	0.03083	0.02989
Model (5)				
PCA	0.1331	0.0069	0.0056	0.0025
Vaguelette	0.1366	0.0081	0.0059	0.0026
WPB	0.1497	0.0067	0.00591	0.0027

As we can see from Table 1, while consistent, method of the principal components perform poorly when we have small number of samples and if the best predictors of the future evolution have little to do with the largest principal components.

The examples and arguments provided in the working paper of Kargin and Onatski [12] confirms that the method of functional principal components is not always the best way to estimate autoregressive operator. We see that the method of wavelet packet basis gives more accurate estimation when we deal with the low number of observations.

3   The standard and non-standard form of operator decomposition

Operators with singularities in the kernel has been considered in the paper by Beylkin *et al.* [7, 8]. They investigated operators representation in the so-called non-standard form and demonstrated that for some operators with singularities sparse representation with algorithm of order $O(N)$ can be achieved.

By using wavelet bases the function $\beta(s, t)$ from (2) can be represented in standard and nonstandard forms. The standard form is the representation of an operator in the tensor product basis. Let $\{V_j\}_{j \in Z}$ be a multiresolution approximation [6] of the function $\beta(s, t)$. If there is the coarsest scale n, then we have

$$V_j = V_n \oplus_{j'=j+1}^{j'=n} W_{j'},$$

where subspace W_{j+1} is supplementary of subspace V_{j+1} and is defined as: $V_j = V_{j+1} \oplus W_{j+1}$. In the case of finitely many scales the autoregressive operator is a representation of $\rho_j = P_j \rho P_j$, where P_j is the projection on subspace V_j. Using standard form representation we can decompose autoregressive operator by the set of operators acting between subspaces of different scales: $W_{j'} \to W_j, W_j \to W_{j'}, V_n \to W_{j'}, W_{j'} \to V_n, V_n \to V_n$.

Alternatively, wavelet bases in $L_2(R^2)$, may be constructed using the scaling function in addition to the wavelets. In that case, the triplet of functions $\psi_{j,k}(t) \, \psi_{j,l}(s)$, $\psi_{j,k}(t)\phi_{j,l}(s)$, $\phi_{j,k}(t)\psi_{j,l}(s)$, where $j, k, l \in Z$, forms a basis of $L_2(R^2)$. Representing operators in these bases leads to the non-standard form [8] as a chain of triplets $\rho = \{A_j B_j \Gamma_j\}_{j \in Z}$ acting on the subspaces V_j and W_j: $A_j \colon W_j \to W_j$, $B_j \colon V_j \to W_j$, $\Gamma_j \colon W_j \to V_j$. The operators $\{A_j B_j \Gamma_j\}_{j \in Z}$ are defined as $A_j = Q_j \rho Q_j$, $B_j = Q_j \rho P_j$ and $\Gamma_j = P_j \rho Q_j$. They admit a recursive definition via the relation

$$\rho_j = \begin{pmatrix} A_{j+1} & B_{j+1} \\ \Gamma_{j+1} & \rho_{j+1} \end{pmatrix},$$

where the operators ρ_j, $\rho_j \colon V_j \to V_j$ are defined by $\rho_j = P_j \rho P_j$. If there is a coarsest scale n, then

$$\rho = \{\{A_j B_j \Gamma_j\}_{j \in Z, j \le n}, \rho_n\},$$

where $\rho_n = P_n \rho P_n$. If the number of the scales is finite, then the operators are organised as blocks of a matrix.

In this paper we extend the notion of the standard and non-standard integral operator representation by considering operator decomposition in the wavelet packet basis. Wavelet packet bases has been introduced by Coifman *et al.* [13] as a generalization of the wavelet transform basis. This generalization of the traditional wavelet filter bank structure permits the representation of a function by one of many bases, each of which is constructed by a unique ensemble of scalings and translations of the same wavelet/scaling filter pair.

The motivation for using wavelet packet bases follows from the possibility to have a whole ensemble of localized representations beyond the traditional wavelet transform. Wavelets can isolate functions behavior in both length (time) and frequency.

Following the same approach as in Section 3 both standard and non-standard forms of autoregressive operator can be introduced for wavelet packet bases (more on the functions decomposition by wavelet packet bases can be found in [13] or [6]).

4   Estimation of the autoregressive operator by wavelet packets

Let Γ_0 be the covariance operator of the random curves $\{\xi_i\}$ defined by $\Gamma_0 = E\langle \xi_i \otimes \xi_i \rangle$ so that $\langle \Gamma_0 x, z \rangle = E\langle \xi_i, x \rangle \langle \xi_i, z \rangle$ for $x, z \in H$. Let Γ_1 be the cross-covariance operator for the curves $\{\xi_i\}$ and $\{\xi_{i-1}\}$ defined by $\Gamma_1 = E\langle \xi_{i-1} \otimes \xi_i \rangle$ so that $\langle \Gamma_1 x, z \rangle = E\langle \xi_{i-1}, x \rangle \langle \xi_i, z \rangle$ for $x, z \in H$.

It is easy to see that by multiplying (1) by ξ_{i-1} and taking expectation we have that the following relationship holds:

$$\Gamma_1 = \rho \Gamma_0.$$

An intuitive way for ρ estimation is to substitute the covariance and cross-covariance operators with their estimates

$$\widehat{\Gamma}_0(x) = \frac{1}{n}\sum_{i=1}^{n}\langle \xi_i, x \rangle \xi_i, \quad \widehat{\Gamma}_1(x) = \frac{1}{n-1}\sum_{i=1}^{n-1}\langle \xi_i, x \rangle \xi_{i+1}.$$

In the expression above we have that Γ_0 is singular and we face with an ill-posed problem. As a consequence, obtaining a consistent estimate of ρ requires regularization of the solution. One way to achieve that would be to use projection method on the span of principal components as mentioned above (6). Mallat *et al.* [14] suggested to estimate locally stationary process eigenfunctions of the covariance operator by searching for their close match in the dictionary of local cosines bases. The best orthogonal basis provides a means of quickly computing compact, adaptive function approximations. In the case of the functional autoregressive processes we have the possibility to chose the basis close to the principal components basis if the principal components basis is the best one or to find better if the principal components basis is not optimal basis for the prediction.

Let $\mathcal{D} = (\phi_\gamma)_{\gamma \in N}$ be a dictionary of waveforms. Wavelet packets and local cosine bases can be as examples of such dictionaries.

The algorithm begins by assuming that the operator kernel β can be sparsely approximated by one of the wavelet packet admissible tree bases. From the fast dynamic programming algorithm of Coifman and Wickerhauser [15] follows that the best bases can be found by a bottom-up progression. At the bottom of deeps J there is only one basis for consideration W_J^p because bottom nodes are not subdecomposed. In the space span by basis W_J^p we take the pairs $(\phi_{\gamma_1}, \phi_{\gamma_2})_k$ of bases that maximizes correlation of functional process $\{\xi_i\}$ from (1)

$$\nu(\phi_{\gamma_1}, \phi_{\gamma_2}) = \frac{\langle \phi_{\gamma_1}, \Gamma_1 \phi_{\gamma_2} \rangle^2}{\langle \phi_{\gamma_1}, \Gamma_0 \phi_{\gamma_1} \rangle \langle \phi_{\gamma_2}, \Gamma_0 \phi_{\gamma_2} \rangle}.$$

Starting from $k = 1$ the operator ρ is estimated in the bases $(\phi_{\gamma_1}, \phi_{\gamma_2})_1$. The statistic $\widehat{\mathrm{MS}}(N, \alpha)$ presented above is used to check the model. If we reject the null hypothesis that the residual partial sums process behaves like white noise when we continue increasing the number of pairs $(\phi_{\gamma_1}, \phi_{\gamma_2})_k$. The next step after the model (1) evaluation in the space W_J^p is to follow as in [15] by investigating the spaces $\{W_j^p\}_{0 \leq p < 2^j}$. When we finish the search in the admissible tree bases we get the most sparse solution.

Also, as it has been shown in our simulation example above, there are special cases when we can have basis better then principal components one. The example below illustrates that we can find in the dictionary the basis that is close to principal components basis.

5 Estimation of the payment systems transactions intensity

The operational cost control as well as the financial liquidity control is the motivation for the following analysis. For the banks, it is important to forecast the next day cash flow at any given moment of time. That is why the functional data analysis is suitable for such applications.

We tackle the problem of the payments prediction intensity in Lithuania cards payment market at ATM (automated teller machine is an electronic device that allows a bank's customers to make cash withdrawals) networks between 03.11.2003 and 10.03.2004. Fig. 2 left side displays the functional data. Each function represents the number of transactions. Fig. 2 right side displays bivariate time series that we received after fixing the values from each functional process at 12:00 AM and 24:00 PM.
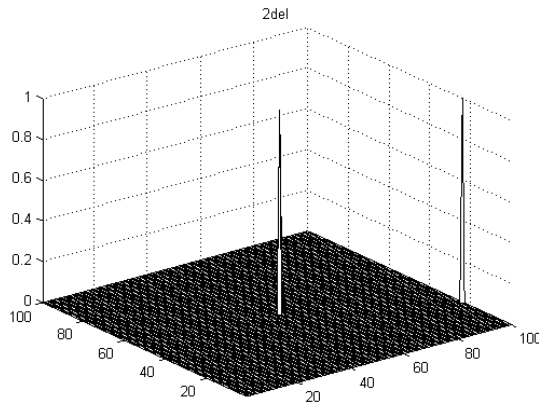

Fig. 2. Bivariate time series.

We use double stochastic Poisson process $(N(t), t \in [0, 1])$ to model the intensity of

the transactions.The corresponding intensities $\Lambda(t), t \in [0, 1]$ are given by

$$\Lambda(t) = \int\limits_0^t \lambda(s)ds = EN(t).$$

By using the observations of the functions $\Lambda(t)$, we build the observations for their derivatives $\lambda(t)$. Namely, we consider

$$\lambda(t_j) = \Lambda(t_j) - \Lambda(t_{j-1}), \quad i = 1, \ldots, n, \;\; j = 1, \ldots, 1024.$$

In the process of the exploratory data analysis we used statistic $\widehat{\mathrm{MS}}$ to judge about the model (1) suitability. This analysis helped us to find that we must censor empirical process by eliminating two weeks of Christmas holiday period (for details see [3]). In Table 2 we show values of the statistic $\widehat{\mathrm{MS}}$ calculated for transactions processes before and after we censored data. The distribution of the statistic $\widehat{\mathrm{MS}}$ can be found in the paper by Laukaitis and Rackauskas [10].

From the results of Table 2 we can say that after adjusting and estimating autoregressive operator the residuals of transactions intensity process follows white noise behavior with 94 % confidential level.

Table 2. Statistic $\widehat{\mathrm{MS}}$ for transactions double stochastic Poisson process data set. Case A – No censoring and no differentiation. Case B – Christmas period censored

a α	0.1	0.2	0.3	0.4	0.45	0.49	0.495
$(A)MS$	12.31	13.31	15.32	17.24	19.23	22.29	22.56
$(B)MS$	0.77	0.85	0.93	1.02	1.07	1.11	1.12

Finally, let us discuss the prediction capability of the model. Consider the prediction problem of the function $\lambda_{n+1}(t), t \in [0, 1]$. For prediction we used the model (1). Different types of the forecasting errors can be considered. The most widely accepted measures of the performance of the estimator of ξ_{n+1} is it's mean integrated square error given by

$$MISE = E \int\limits_0^1 \left(\xi_{n+1}(t) - \widehat{\xi}_{n+1}(t) \right)^2 dt.$$

We evaluated the MISE error by choosing ten days not used in the parameters estimation(see Table 3).

It is clear that the predicted values depends on the number of principal components used (for principal components estimation method) or the number of selected bases (for wavelet packet bases estimation method) in the model. As we can see the most optimal subspace dimension for principal components method is 4 and 7 for wavelet packet bases method.

Table 3. Prediction errors

Error\M	1	2	3	4	5	6	7	8
PCA	5.50	5.02	4.99	4.95	5.05	4.97	5.04	5.07
WPB	5.86	5.53	5.24	5.18	5.05	4.97	4.96	5.06

6 Concluding remarks

We suggested the wavelet packet bases method for functional autoregressive process estimation. As we have seen the method shows better prediction compared with the other known methods when autoregressive kernel exhibits some singular structures and when the number of observations is limited. More mathematical investigation is needed to investigate asymptotic properties of this method. We hope to provide them in the future research work.

References

1. P. Besse, H. Cardot, D. Stephenson, Autoregressive forecasting of some functional climatic variations, *Scandinavian Journal of Statistics*, **27**(4), pp. 673–688, 2000.

2. A. Cavallini, G. C. Montavarri, M. Loggini, O. Lessi, M. Cacciani, Nonparametric prediction of harmonic levels in electrical networks, in: *Proceed. IEEE ICHPS VI, Bologna*, pp. 165–171, 1994.

3. A. Laukaitis, A. Rackauskas, Functional Data Analysis of Payment Systems, *Nonlinear Analysis: Modelling and Control*, **7**(2), pp. 53–68, 2002.

4. D. Bosq, Modelization non-parametric estimation and prediction for continuous time processes, in: *Nonparametric Functional Estimation and Related Topics*, G. Roussas (Ed.), NATO ASI series, Kluwer Academic Publishers, pp. 509–529, 1991.

5. A. Antoniadis, T. Sapatinas T., Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes, *Journal of Multivariate Analysis*, **87**(1), pp. 133–158, 2003.

6. S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1999.

7. G. Fann, G. Beylkin, R. Harrison, K. Jordan, Singular operators in multiwavelet bases , APPM preprint 517, *IBM Journal of Research and Development*, **48**(2), pp. 161–171, 2004.

8. G. Beylkin, R. Coifman, V. Rokhlin, Fast Wavelet Transforms and Numerical Algorithm, *Commun. Pure & Appl. Math.*, **44**, pp. 141–183, 1991.

9. D. Bosq, *Linear Processes in Function Spaces*, in: *Lecture Notes in Statistics*, Vol. 149, Springer-Verlag, Berlin, 2000.

10. A. Laukaitis, A. Rackauskas, Testing changes in Hilbert space autoregressive models (Russian,English), *Lith. Math. J.*, **42**(4), pp. 343–354; translation from *Liet. Mat. Rink.*, **42**(4), pp. 434–447, 2002.

11. K. Bradley, A. Alpert, Class of Bases in L^2 for the Sparse Representation of Integral Operators, *SIAM Journal on Mathematical Analysis*, **24**(1), pp. 246–262, 1993.

12. V. Kargin, A. Onatski, *Dynamics of Interest Rate Curve by Functional Auto-Regression*, Economics Working Paper Archive at WUSTL, 2004.

13. R. R. Coifman, Y. Meyer, M. V. Wickerhauser, Wavelet Analysis and Signal Processing, in: *Wavelets and Their Applications*, M. B. Ruskai et al. (Eds.), Jones and Bartlett, Boston, MA, pp. 153–178, 1992.

14. S. Mallat, G. Papanicolaou, Z. Zhang, Adaptive Covariance Estimation of Locally Stationary Processes, *Annals of Statistics*, **26**(1), pp. 1–47, 1998.

15. R. R. Coifman, M. V. Wickerhauser, Entropy-based algorithms for best basis selection, *IEEE Trans. Inform. Theory*, **38**(2), pp. 713–718, 1992.

16. P. Billingsley, *Convergence of Probability measures*, John Wiley & Sons, New York, 1999.

17. P. Besse, H. Cardot, F. Ferraty, Simultaneous nonparametric regressions of unbalanced longitudinal data, *Computational Statistics and Data Analysis*, **24**, pp. 255–270, 1997.

18. S. Guillas, Rates of convergence of autocorrelation estimates for autoregressive Hilbertian processes, *Stat. Probab. Letters*, **55**(3), pp. 281–291, 2001.