Vol. XX, No. X, pp. xxx-xxx

GENERALIZED EXPONENTIAL SMOOTHING IN PREDICTION OF HIERARCHICAL TIME SERIES

Daniel Kosiorowski[1], **Dominik Mielczarek**[2] **Jerzy P. Rydlewski**[3]
Małgorzata Snarska[4]

ABSTRACT

Shang and Hyndman (2016) proposed grouped functional time series forecasting approach as a combination of individual forecasts using generalized least squares regression. We modify their methodology using generalized exponential smoothing technique for the most disaggregated series in order to obtain more robust predictor. We show some properties of our proposals using simulations and real data related to electricity demand prediction.

Key words: functional time series, hierarchical time series, forecast reconciliation, depth for functional data.

1. Introduction

A problem of optimal reconciliation of forecasts of activities of economic agents partitioned into certain groups and/or levels of hierarchy was considered in the economic as well as econometric literature many times and is still present in a public economic debate (see Kohn (1982), Weale (1988), Hyndman et al. (2011)). National import/export or national added value balances are important examples here. Discrepancies between forecasts at global and local (regional) level are usually thought to be caused by divergent methodologies or different precision of measurements at different hierarchy levels used. A reconciliation of the level and global forecasts is very important issue for policy makers both from fiscal and monetary side. Current fiscal policy is usually modified on tax and duties inflows forecasts basis to create long term social welfare. These estimates must of course include demographic forecasts in regions of a country as well. As for monetary policy decisions, where

[1] Department of Statistics, Cracow University of Economics, Kraków, Poland. E-mail: daniel.kosiorowski@uek.krakow.pl

[2] AGH University of Science and Technology, Faculty of Applied Mathematics, al. A. Mickiewicza 30, 30-059 Krakow, Poland. E-mail: dmielcza@wms.mat.agh.edu.pl

[3] AGH University of Science and Technology, Faculty of Applied Mathematics, al. A. Mickiewicza 30, 30-059 Krakow, Poland. E-mail: ry@agh.edu.pl

[4] Department of Financial Markets, Cracow University of Economics, Kraków, Poland. E-mail: malgorzata.snarska@uek.krakow.pl

the important task is the inflation target and sustainable growth of economy, central banks have to combine and aggregate in a reasonable way not inflation forecasts for particular countries and/or regions, but also include various reporting from banks and financial institutions. The issue is also very important from a particular company point of view in a context of product lines management, consumers portfolio optimization and consumers segmentation. Let us look at the problem of demand forecast reconciliation from an agent based economic perspective and think about economy as resultant of choices and decisions made by individual actors. Suppose now, that one reconsiders classical problem of demand curve forecasting. In the classical framework the problem reduces to an appropriate aggregation of various economic utility measures of agents taking part in a certain market. By aggregate measure economists usually understand a summary measure describing the whole market or economy. When constructing such measures one is usually faced by the difficulty of reconciling less aggregated measures (i.e., behavior of individual economic agent) into a more coarse-grained one being still relative to less aggregated counterparts. Classically one assumes that law of one price, law of demand and supply, no arbitrage or general equilibrium conditions hold. That is why economists usually include abstract concepts like composite good or representative agent, common utility functions. Individual behavior of economic quantity is then assumed to change proportionately to the predefined composite index. The aggregate demand curve is usually then estimated as a simple sum of individual demand curves originating from homogeneous utility curves. Standard approach to economic analysis imply the so called representative agent, i.e. an agent that acts rationally maximizing his individual utility curve. Classical macroeconomic theoretical analysis assumes equality of individual utilities. Since all agents are the same, economy is now viewed as a simple sum of decision of individuals and individual forecasts of future economic states average themselves to aggregate forecasts. It is easy to notice, that real economy is much more complicated. Agents certainly optimize their utilities, but the utilities are rather heterogeneous and their decisions are strongly affected by decisions of other individuals with whom their interact. Choices of agents due to their emotions, mistakes etc. are random to certain degree. Classical economic theory ignores these micro fluctuations. On the other hand side an empirical evidence shows that behaviors of individual consumers, firms and households are very unique and heterogeneous. Introduction of fluctuations and interactions between economic agents into a dynamic economic system results in large number of combination of individual states even when economic agents are faced with binary choice decisions (i.e. buy or not to buy a certain amount of particular good). Introduction of interactions may result in a changing partition of agents into clusters and or levels of hierarchy. A discrepancies between aggregates measuring the same may appear due to the relocation of agents into new clusters and/or levels for example. Many economic phenomena can be described by means of functions (i.a. utility curves, demand curves, development paths). In recent years a very interesting in this coontext statistical methodology for analyzing functional data has been developed (see Bosq (2000), Ramsay et al.(2009), Horvath and Kokoszka (2012), Krzyśko et al. (2013), Shang and Hyndman (2016)). Following the cited authors we consider random curve $X = \{x(t), t \in [0,T]\}$, where T is fixed, as a random element of the separable Hilbert space $L^2([0,T])$ with the inner product $< x,y >= \int x(t)y(t)dt$. The space is equipped with the Borel $\sigma-$algebra. Furthermore, in Bosq's (2000) monography it is proved that probability distributions do exist for functional processes with values in Hilbert space. Functional time series (FTS) is a series of functions indexed by time (e.g. see Fig. 1). Imagine now, that one observes economic data at various granularity e.g. demand for electricity of a single household, single city, state or region, utility curves for different groups of consumers. A hierarchical functional time series is a series of functions grouped at specified levels (household, town, region, whole country), i.e. see Fig. 1. At each level a forecast can be made. A natural problem arises: how to use information obtained at different levels to obtain a reconciliated prediction at all levels?

To understand the idea behind hierarchical functional forecasting better, revert back to our motivational problem of forecasting changes in demand function. Traditionally the demand for every good or service is usually explained by its own price, income and preferences of the buyers depending on economic conditions, the existence of close substitutes, and the quality of the good and service itself. Although there is a vast economic literature on consumer choice theory and aggregate demand, it still remains inconclusive on how to effectively model total demand, when one assumes that each individual has its own value and utility functions, especially when these functions vary in shape and time, individuals can interact with each other and this also influences their individual demand functions. That is why we are going to treat these functions as random allowing these functions to be monitored in time. It is advantageous to look at random functions observed at regular time intervals as functional time series (FTS) as well. Alternatively a functional time series may be constructed by separating an almost continuous time interval into natural consecutive partitions such as hours, days, weeks, months, years.
The problem of hierarchical time series prediction is solved with various ways. Bottom-up method relies on forecasting each of the disaggregated series at the lowest level of the hierarchy, and then using simple aggregation to obtain forecasts at higher level of the hierarchy (see Kahn (1998)). Top-down method involves fore-

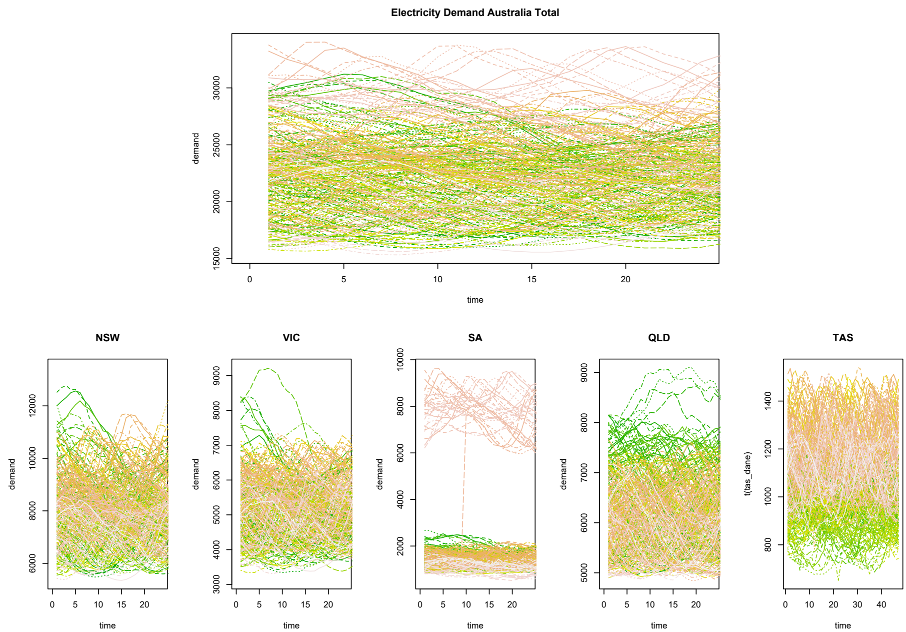

Figure 1: Electricity demand in regions of Australia in 2016 – hierarchic functional time series example.

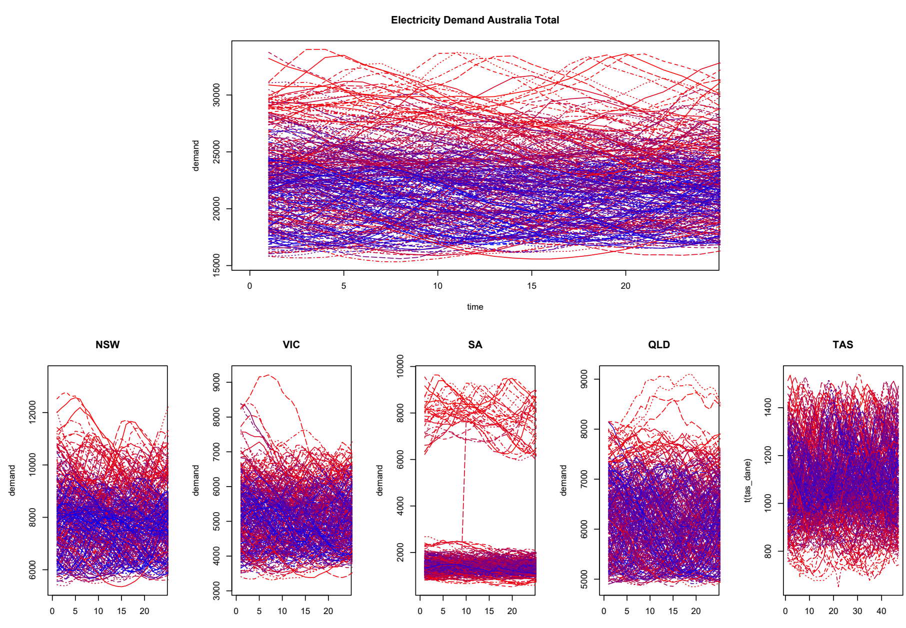

Figure 2: Corrected band depths of curves. Depth value varies from maximal (blue color the deepest curve) to minimal (red color the most peripheral curve).

casting of aggregated series at the top level of the hierarchy, and then using disaggregation to obtain forecasts at lower level of the hierarchy based on historical proportions. Shang and Hyndman (2016), extending the method of Hyndman et al. (2011), considered grouped functional time series forecasting as an optimal combination of individual forecasts using generalized least squares regression with level forecasts treated as regressors. In the context of HTS prediction a general problem arises: which method of forecasting at particular level should be chosen (see Bosq (2000), Besse et al. (2000), Hyndman and Ullah (2007), Hyndman and Shang (2009), Aue et. al. (2015)? Shang and Hyndman (2016) proposed grouped functional time series forecasting approach as a combination of individual forecasts obtained by means of their smart predicting method in which functional time series is reduced to family of one dimensional time series of principal component scores representing of original functional series. As a result of conducted simulation studies we decided to modify their methodology. Instead of using principal component scores forecast methods we decided to propose a certain functional generalization of exponential smoothing technique (see Hyndman et al.(2008) for a theoretical background of the exponential smoothing), i.e. we used moving local medians and moving local averagess for the most disaggregated series in order to obtain more robust predictor than Shang and Hyndman' (2016) predictor. Main aim of the paper is to modify the Shang and Hyndman (2016) predictor so that it could cope with functional outliers and/or it would be elastic enough to adapt to changes in data generating mechanism. The remainder of the paper is as follows: in the second section elements of depth concept are sketched and in the third section our proposals are introduced. Fourth section presents results of simulation as well as empirical studies of properties of our proposals. The paper ends with conclusions, references and short appendix containing R script illustrating forecasts calculation performed using our proposals.

2. Depths for functional data

For obtaining robust hierarchical FTS predictor we focused our attention on the functional data depth concept (Nagy et al. (2016) and Nieto-Reyes and Battey (2016)). Let $\mathbf{X} = \{x_1, ..., x_n\}$ be a sample of continuous curves defined on the compact interval T. Let λ denote the Lebesgue measure and let $a(i_1, i_2) = \{t \in T : x_{i_2} - x_{i_1} \geq 0\}$, where x_{i_1} and x_{i_2} are band delimiting objects. Let $L_{i_1,i_2} = \frac{\lambda(a(i_1,i_2))}{\lambda(T)}$. We have chosen, in our opinion the best depth for considered functional data, namely the corrected generalized band depth (cGBD, see López-Pintado and Jörnsten (2007)).

The cGBD of a curve x with respect to the sample $\mathbf{X}$ is defined as

$$cGBD(x|\mathbf{X}) = \frac{2}{n(n-1)} \sum_{1 \le i_1 < i_2 \le n} \frac{\lambda(A^c(x; x_{i_1}, x_{i_2}))}{\lambda(T)}$$

where

$$A^c(x; x_{i_1}, x_{i_2}) = \{t \in a(i_1, i_2) : x_{i_1}(t) \le x(t) \le x_{i_2}(t)\}, \text{ if } L_{i_1,i_2} \ge \frac{1}{2}$$

or

$$A^c(x; x_{i_1}, x_{i_2}) = \{t \in a(i_2, i_1) : x_{i_2}(t) \le x(t) \le x_{i_1}(t)\}, \text{ if } L_{i_2,i_1} > \frac{1}{2}.$$

Band depth is modified in order to consider only the proportion of the domain where the delimiting curves define a contiguous region which has non–zero width. In order to perform the construction we calculated the depth regions of order α for considered cGBD, i.e. $R_\alpha(P) = \{x : cGBD(x,P) \ge \alpha\}$. α-central regions $R_\alpha(P) = \{x \in L^2([0,T]) : D(x,P) \ge \alpha\}$ for any depth function $D(x,P)$ may be defined, where P denotes probability distribution. Various descriptive characteristics, like scatter, skewness, kurtosis, may be expressed in terms of α-regions. These regions are nested and inner regions contain less probability mass. Following Paindaveine and Van Bever (2013), when defining local depth it will be more appropriate to index the family $\{R_\alpha(P)\}$ by means of probability contents. Consequently, for any $\beta \in (0,1]$ we define the smallest depth region with P-probability equal or larger than β as

$$R^\beta(P) = \bigcap_{\alpha \in A(\beta)} R_\alpha(P),$$

where $A(\beta) = \{\alpha \ge 0 : P(R_\alpha(P)) \ge \beta\}$. The depth regions $R_\alpha(P)$ and $R^\beta(P)$ provide only the deepest point neighborhood. However, we can replace P by its symmetrized version $P_x = \frac{1}{2}P^{\mathbf{X}} + \frac{1}{2}P^{2x-\mathbf{X}}$. For any depth function $D(\cdot,P)$ the corresponding *sample local depth function at the locality level* $\beta \in (0,1]$ is $LD^\beta(x,P^{(n)}) = D(x,P_x{}^{\beta(n)})$, where $P_x^{\beta(n)}$ denotes the empirical measure with those data points that belong to $R_x^\beta(P^{(n)})$. $R_x^\beta(P^{(n)})$ is the smallest sample depth region that contains at least a proportion β of the $2n$ random functions $x_1,...,x_n,2x-x_1,...,2x-x_n$. Depth is always well defined – it is an affine invariant from original depth. For $\beta = 1$ we obtain global depth, while for $\beta \simeq 0$ we obtain extreme localization. As in the population case, our sample local depth will require considering, for any $x \in L^2$, the symmetrized distribution P_x^n which is empirical distribution associated with $x_1,...,x_n,2x-x_1,...,2x-x_n$. Sample properties of the local versions of depths result from general findings presented in (see Zuo and Serfling, 2000). Implementations of local versions of several depths including projection depth, Student, simplicial, L^p depth, regression depth and modified band depth that can be found in free R package *DepthProc* (see Kosiorowski and Zawadzki, 2014). In order to choose the locality parameter β we recommend using expert knowledge related to the number of components or regimes in the considered data. Sample properties of the local versions of depths result from general findings presented in Paindaveine and Van Bever (2013). For other concepts of local depths see e.g. Sguera et al. (2016). Implementations of local versions of several depths including projection depth, Student, simplicial, L^p depth, regression depth and modified band depth can be found in free R package *DepthProc* – see Kosiorowski and Zawadzki (2014). Fig. 2 presents local sample cGBD for data presented on fig. 1 with $\beta = 0.45$ and representing electricity demand curves in 2016 in regions of Australia. Depth values are represented by colors varying from blue (the deepest curves) to red (the most peripheral curves)

3. Our proposals

We consider a sample of N functions $\mathbf{X}^N$=$\{x_i(t), t \in [0,T]\}$. Let $FD^{\beta}(y|\mathbf{X}^N)$ denote sample functional depth of $y(t)$ with locality parameter β, e.g. the functional depth is equal to corrected generalized band depth: $FD = cGBD$. Sample β–local median is defined as

$$MED_{FD^{\beta}}(\mathbf{X}^N) = \underset{i=1,\ldots,N}{\arg\max} FD^{\beta}(x_i|\mathbf{X}^N).$$

Assume now, that we observe a data stream $x_i(t)$, $i = 1,2,\ldots$. We put our proposals forward.

Proposal 1: Moving median concept is employed and generalized exponential smoothing takes the following form, where as a predictor for $(n+1)$th moment we take

$$\hat{x}_{n+1}(t) = MED_{FD^{\beta}}(W_{n,k}),$$

and $W_{n,k}$ denotes a moving window of length k ending in a moment n, i.e., $W_{n,k} = \{x_{n-k+1}(t), \ldots, x_n(t)\}$. For sample of N functions $\mathbf{X}^N$=$\{x_i(t), t \in [0,T]\}$, let $FD^{\beta}(y|\mathbf{X}^N)$ denote sample functional depth of $y(t)$ with locality parameter β. Sample α–trimmed mean with locality parameter β is defined as

$$ave(\alpha,\beta)(\mathbf{X}^N) = ave(x_i : FD^{\beta}(x_i|\mathbf{X}^N) < \alpha),$$

where *ave* denotes sample average.

Proposal 2: In this setup a generalized exponential smoothing technique is used as well. As a predictor for $(n+1)$th moment we take

$$\hat{x}_{n+1}(t) = z \cdot ave(\alpha_1, \beta_1)(W_{n_1,k_1}) + (1-z) \cdot ave(\alpha_1, \beta_1)(W_{n_2,k_2}),$$

where $W_{n,k}$ denotes a moving window of length k ending in a moment n , i.e., $W_{n,k} = \{x_{n-k+1}(t), ..., x_n(t)\}$, $z \in [0,1]$ is a forgetting parameter and $n_2 < n_1$. Thus we consider a closer past represented by W_{n_1} and a further past represented by W_{n_2}. Notice, that lengths of the moving windows k, k_1, k_2 used in Proposals 1 – 2 relate to the analogous forgetting parameters α in the classical exponential smoothing. Additionally, we have in our disposal a resolution parameter β at which we predict a phenomenon.

Proposal 3: Hierarchical FTS predictor

Step 1: At any aggregation level we apply our Proposition 1 or Proposition 2 in order to compute forecasts independently on every level. Hence, generalized exponential smoothing takes the following form

$$\hat{x}_{n+1}^{level}(t) = MED_{FD^{\beta}}(W_{n,k}^{level})$$

or

$$\hat{x}_{n+1}^{level}(t) = z \cdot ave(\alpha_1, \beta_1)(W_{n_1,k_1}^{level}) + (1-z) \cdot ave(\alpha_1, \beta_1)(W_{n_2,k_2}^{level}).$$

Such approach allows us to accommodate expert knowledge and adjust forgetting and resolution parameters to researcher's requirements. For comparison purpose, in Shang and Hyndman (2016) paper authors make predictions using functional regression based on time constant functional principal components scores modeling by means of one–dimensional time series method proposed in Hyndman and Shang (2009).

Step 2 – In this step we consider a whole hierarchy structure of the phenomenon. We made a level forecasts with Proposals 1 or 2. Smart reconciliation of forecasts is conducted then and our reconciled predictor takes a form:

$$\hat{\mathbf{X}}_{n+1}(t) = \mathbf{F}(\hat{x}_{n+1}^{level_{i_1}}, ..., \hat{x}_{n+1}^{level_{i_k}}),$$

where hierarchical structure is described by fixed hierarchy levels $level_{i_1}, ..., level_{i_k}$ and $\mathbf{F}$ is a Generalized Least Squares Estimator (see Shang and Hyndman, 2016). We can write our model in the form

$$R_t = S_t b_t,$$

where R_t is a vector of all series at all levels, b_t is a vector of the most disaggregated data and S_t is a time constant matrix that shows a relation between them. We propose to do the base forecast:

$$\hat{R}_{n+1} = S_{n+1}\beta_{n+1} + \varepsilon_{n+1},$$

where $\hat{R}_{n+1}$ is a matrix of base forecast for all series at all levels, $\beta_{n+1} = E[b_{n+1}|R_1,...,R_n]$ is the unknown mean of the forecast distribution of the most disaggregated series and ε_{n+1} is responsible for errors. We propose to use a generalized least square method as in Shang and Hyndman (2016)

$$\hat{\beta}_{n+1} = \left(S_{n+1}^T W^{-1} S_{n+1}\right)^{-1} S_{n+1}^T W^{-1} \hat{R}_{n+1}$$

modified so that we use a robust estimator of the dispersion matrix W, i.e. instead of diagonal matrix, which contains forecast variances of each series, we use a robust measure of forecast dispersion taking into account dependency structure between the level forecasts. Notice, that a dynamic updating of the dispersion matrix should be considered in further studies.
If we consider a hierarchy as in above Fig. 2, our dispersion matrix takes a form:

$$W = diag\{v^{total}, v^{level_{i1}}, v^{level_i 2}, v^{level_i r}, ..., v^{level_i q}\}$$

where

$$v^{level} = V\left\{\int_0^T \left(x_{nk}^{level} - \hat{x}_n^{level}\right)^2 ds, k \in K_{level}, n = 1, ..., 365\right\}$$

where K_{level} is a number of obs. at considered level in time n, $n = 1, ..., n_h$. If and where V is a robust measure of dispersion. We propose to use $c \cdot MAD$ instead of standard deviation or take into account dependency structure between level series using well known minimum covariance determinant (MCD) or recently proposed PCS robust matrix estimators of multivariate scatter (see Vakili and Schmitt (2014)).

4. Properties of our proposals

Thanks to kindness of prof. Han Lin Shang, who made his R script available for us we calculated optimal combination of forecast predictor add we compared Shang and Hyndman (2016) proposal with our Proposal 3.

We generated samples from SV, GARCH, Wiener, Brown bridge, FAR processes and various mixtures of them. In the simulations we considered several locality parameters differing within the levels of hierarchy, moving window lengths. We considered samples with and without functional outliers. The outliers were defined with respect to the functional boxplot induced by the cGBD i. e. we replaced

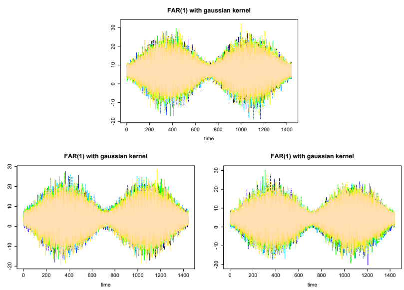

Figure 3: Simulated HTFS consisted of FAR(1) processes

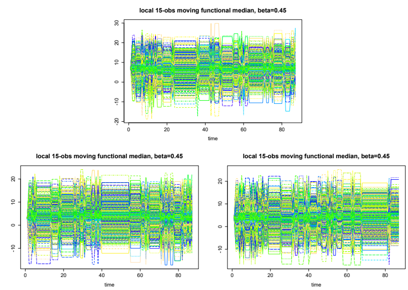

Figure 4: HFTS prediction using our proposal

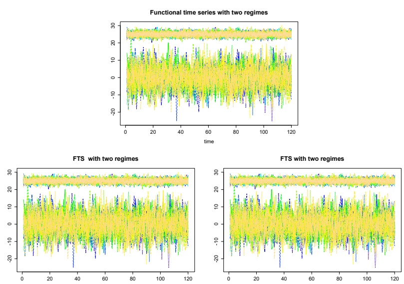

Figure 5: Simulated HTFS consisted of two regime FTS processes

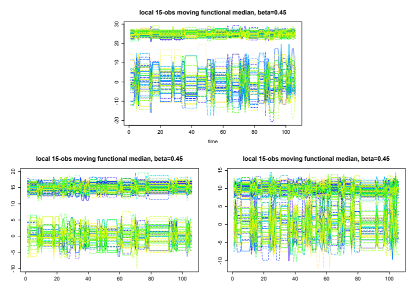

Figure 6: HFTS prediction using our proposal

1%,5%,10% of curves in the samples by means of arbitrary curves being outside a band determined by the functional boxplot whiskers and compared medians and medians of absolute deviations from the medians (MAD) of integrated forecasts errors in these two situations. Fig. 3 presents simulated hierarchical functional time series consisted of three functional autoregression models of order 1 FAR(1) with gaussian kernels and sine–cosine errors design (see Didericksen et al. 2012). Fig. 4 presents corresponding level forecasts obtained by means of local moving median calculated from 15-obs. windows and locality parameters equal to 0.45. Fig. 5 presents simulated hierarchical time series consisted of three processes being mixtures of two stochastic volatility processes (SV). Fig. 6 presents corresponding level forecasts obtained by means of local moving median calculated from 15-obs. window and locality parameters equal to 0.45. We indicated an order of appearance of observations using colors palette starting from yellow and ending on blue. In the appendix we placed a simple script depending on *DepthProc* R package illustrating a general idea of the performed simulations.

In order to check the statistical properties of our proposals we considered empirical data set related to electricity demand in the period from 1 to 31 January 2016 in Australia. The data come from five regions of Australia, denoted with the following symbols: nsw, q, sa, tas, vic. All the considered data was taken from Australian Energy Market Operator https://www.aemo.com.au/. Fig. 7 presents 297 predicted electricity demand curves obtained by means of our Proposal 3 using moving local median (taken from our Proposal 1) to calculate level forecasts, window lengths $k = 10$ observations and locality parameter equal to $\beta = 0.2$. Fig. 8 presents boxplots of 297 integrated squared differences between observed and predicted demand curves in each region and in the whole Australia. Above the boxplots one can find medians of absolute deviations (MAD) for the corresponding 297 forecast error measures. A performance of our proposal was compared with Shang and Hyndman (2016) proposals with respect to sum of integrated squared forecast residuals and with respect to the MAD of integrated squared forecast residuals. We considered also a prediction by means of estimated FAR(1) process, but we did not obtain better forecasts in comparison to Shang and Hyndman (2016) proposal. Our proposal seems to be more robust to functional outliers than their proposal, however. It is not surprising, as the authors made their forecasts basing on nonrobust generalized least squares method. Admittedly, Shang and Hyndman (2016) claimed that their proposal performed better in comparison to bottom-up approach basing on moving medians, but notice, that they considered Fraiman and Muniz global depths only. Moreover thanks to locality parameter adjusting our proposal is more appropriate for detecting the change of regimes in HFTS setup.

4.1. Uncertainty evaluation

Series of functional principal component scores are considered as surrogates of original functional time series (see Aue et al. 2015, Hyndman and Shang (2009)). Several authors postulate using dynamic functional principal components approach in order to take into account time changing dependency structure of described phenomenon (Aue et al. 2015). Notice, that such modification may drastically increase computational complexity of the HFTS procedure. In a context of uncertainty evaluation of our proposals we suggest considering Vinod and de Lacalle (2009) maximum entropy bootstrap for time series approach. Bootstrap methods for FTS were studied among other by Hörmann and Kokoszka (2012), Shang (2016). Similarly as in Shang and Hyndman (2016) we propose to use maximum entropy bootstrap methodology to obtain confidence regions and to conduct statistical inference. The *meboot* and *DepthProc* R packages give the appropriate computational support for that aims.

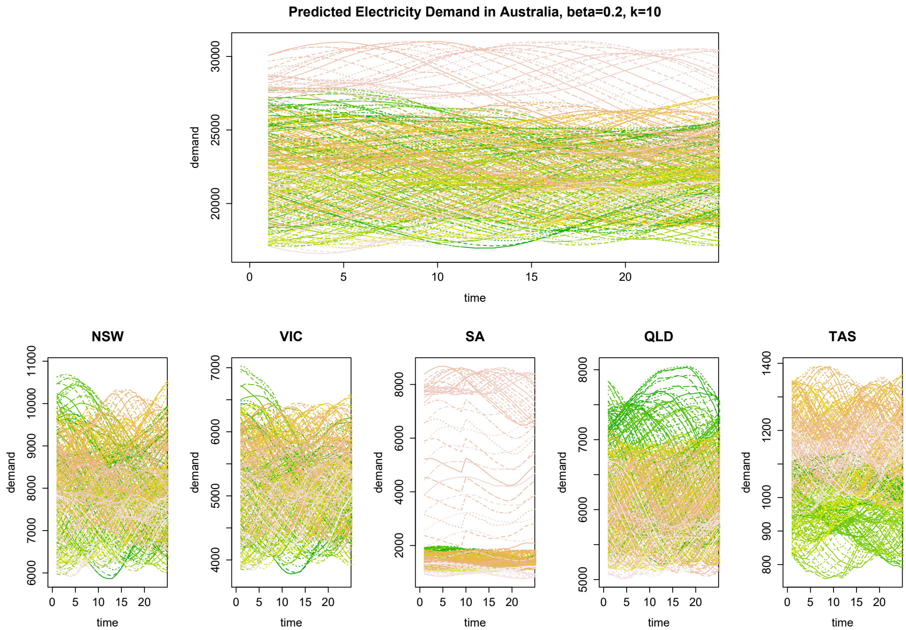

Figure 7: Predicted electricity demand in Australia in 2016

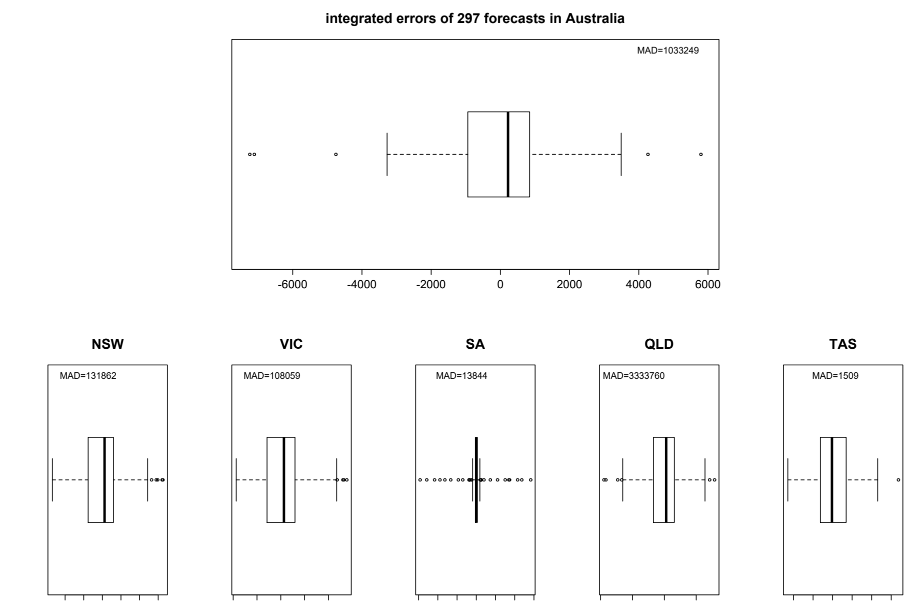

Figure 8: Preditcion errors of electricity demand in Australia in 2016

5. Conclusions

Hierarchical functional time series methodology opens new areas of statistical as well as economical research. E–economy provides a great deal of HFTS data. Our proposal of HFTS predictor basing on local moving functional median performs surprisingly well in comparison to a milestone proposal of Shang and Hyndman (2016). Lengths of the moving windows used in our proposals 1 – 3 relate to the forgetting parameters α's in the classical exponential smoothing. Moreover, we have in our disposal a resolution parameter β at which we predict a phenomenon. When using the locality parameter in our proposal, we can take into account different sensitivity to details , i.e. e.g. number of different regimes of the considered phenomena.

Acknowledgements

JPR research has been partially supported by the AGH local grant no. 15.11.420.038
DM research has been partially supported by the AGH local grant no. 11.11.420.004
MS research was partially supported by NSC Grant no. .OPUS.2015.17.B.HS4.02708,
MS and DK research has been partially supported by Cracow University of Economics local grants

REFERENCES

AUE, A., DUBABART NORINHO, D., HÖRMANN, S. (2015), *On the prediction of stationary functional time series. Journal of the American Statistical Association*, 110(509), 378–392.

BOSQ, D. (2000), *Linear Processes in Function Spaces*. Springer.

DIDERICKSEN,D., KOKOSZKA, P., ZHANG, X. (2012), *Empirical properties of forecasts with the functional autoregressive model. Computational Statistics*, 27(2): 285–298.

FEBRERO-BANDE, M. O., DE LA FUENTE M., (2012), *Statistical Computing in Functional Data Analysis: The R Package fda.usc*, *Journal of Statistical Software*, **51**(4), 1–28.

HORVATH, L., KOKOSZKA, P. (2012), *Inference for Functional Data with Applications*, Springer.

HÖRMANN S., KOKOSZKA P. (2012), *Functional Time Series*, in *Handbook of Statistics: Time Series Analysis – Methods and Applications*, 30, 157–186.

HYNDMAN, R. J., AHMED R. A., ATHANASOPOULOS, G., SHANG, H. L. (2011), *Optimal combination forecasts for hierarchical time series*, *Computational Statistics & Data Analysis*, Volume 55(9), 2579 –2589.

HYNDMAN, R.J., KOEHLER, A.B., ORD, J.K., SNYDER, R.D.(2008) *Forecasting with Exponential Smoothing – The State Space Approach*, Springer.

HYNDMAN, R. J., SHANG, H., L. (2009) *Forecasting functional time series*, *Journal of the Korean Statistical Society* 38(3), 199 –221.

HYNDMAN, R.J., ULLAH, M. (2007) *Robust forecasting of mortality and fertility rates: A functional data approach*, *Computational Statistics & Data Analysis* 51(10), 4942 – 4956.

HYNDMAN, R.J., KOEHLER, A.B., SNYDER, R.D., GROSE, S. (2002) A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3), 439–454.

KAHN, K.B. (1998) *Revisiting top-down versus bottom-up forecasting*, *The Journal of Business Forecasting Methods & Systems* 17(2), 14–19.

KOHN, R. (1982) *When is an aggregate of a time series efficiently forecast by its past*, *Journal of Econometrics* 18(3), 337 – 349.

KOSIOROWSKI, D., ZAWADZKI, Z., (2014) *DepthProc: An R Package for Robust Exploration of Multidimensional Economic Phenomena arXiv: 1408.4542*.

KOSIOROWSKI, D. (2014), *Functional Regression in Short Term Prediction of Economic Time Series*, *Statistics in Transition*, vol. 15, no. 4, 611 –626.

KOSIOROWSKI, D. (2016) *Dilemmas of Robust Analysis of Economic Data Streams*, *Journal of Mathematical Sciences* (Springer), vol. 218, No. 2,167–181, 2016.

KRZYŚKO, M., K. DERĘGOWSKI, K., GÓRECKI T. (2013), *Jądrowa i funkcjonalna analiza głównych składowych*, MSA 2013 plenary lecture.

LÓPEZ-PINTADO, S., JÖRNSTEN, S., R. (2007) *Functional Analysis via Extensions of the Band Depth*, IMS Lecture Notes–Monograph Series Complex Datasets and Inverse Problems: Tomography, Networks and Beyond, Vol. 54, 103–120, Institute of Mathematical Statistics.

NIETO-REYES, A., BATTEY, H. (2016), *A Topologically Valid Definition of Depth for Functional Data*, *Statistical Science* 31(1), 61–79.

PAINDAVEINE, D., G. VAN BEVER, G. (2013), *From Depth to Local Depth: A Focus on Centrality*, *Journal of the American. Statistical Asssociation*, Vol. 108, No. 503, Theory and Methods, 1105–1119.

RAMSAY, J. O., G. HOOKER, G., GRAVES, S. (2009), Functional Data Analysis with R and Matlab, Springer.

NAGY, S., HLUBINKA, D., GIJBELS, I. (2016) *Integrated depth for functional data: Statistical properties and consistency*, *ESIAM Probability and Statistics*.

SHANG, H., L., HYNDMAN, R., J., (2016) *Grouped functional time series forecasting: an application to age-specific mortality rates*, *Journal of Computational and Graphical Statistics*.

SGUERA, C., GALEANO, P., LILLO, R., E. (2016), *Global and local functional depths* arXiv 1607.05042v1.

SHANG, H., L. (2016) *Bootstrap methods for stationary functional time series*, *Statistics and Computing*.

WEALE, M. (1988) *The reconciliation of values, volumes and prices in the national accounts*, *Journal of the Royal Statistical Society A* 151(1) 211–221.

VAKILI K., SCHMITT E. (2014). *Finding multivariate outliers with FastPCS*. *Computational Statistics & Data Analysis*, 69, 54–66.

VINOD, H., D., J. L. DE LACALLE, J., L. (2009) *Maximum entropy bootstrap for time series: the meboot R package*, *Journal of Statistical Software*, 29(5).

Appendix

```
#Simple R script, example showing how to calculate base forecasts for three hierarchy levels
#using moving functional median implemented within the DepthProc R package.
require(DepthProc)
require(RColorBrewer)
require(zoo)
wrapMBD = function(x)depthMedian(x, depth1="MBD",method="Local",beta=0.45)
#Simple stochastic volatility process simulator
#Simple stochastic volatility process simulator #
SV <- function(n, gamma, fi, sigma, delta) {
epsilon <- rnorm(n)
eta <- rnorm(2*n, 0, delta)
h <- c()
h[1] <- rnorm(1)
for (t in 2:(2*n)) {
h[t] <- exp(gamma+fi*(h[t-1]-gamma)+sigma*eta[t]) }
Z <- sqrt(tail(h,n)) * epsilon
return(Z)
}
example <- SV(100, 0, 0.2, 0.5, 0.1)
plot(ts(example))
#functional time series simulator
m.data1<-function(n,a,b) {
M<-matrix(nrow=n,ncol=120)
for (i in 1:n) M[i,]<- a*SV(120,0,0.3,0.5,0.1)+b
M }
m.data.out1<-function(eps,m,n,a,b,c,d){
```

```
H<-rbind(m.data1(m,a,b),m.data1(n,c,d))
ind=sample((m+n),eps)
H1=H[ind,]
H1 }
m <- matrix(c(1, 0, 1, 3, 2, 3, 2, 0), nrow = 2, ncol = 4) m[2,]=c(2,2,3,3) m[1,]=c(0,1,1,0)
#below three functional time series
M2A= m.data.out1(150,3000,7000,5,0,1,25)
M2B= m.data.out1(150,3000,7000,2,0,1,15)
M2C= m.data.out1(150,3000,7000,3,0,1,10)
matplot(t(M2A),type="l",col=topo.colors(151), xlab="time",main="Functional time series
with two regimes")
matplot(t(M2B),type="l",col=topo.colors(151), xlab="time",main="FTS with two regimes")
matplot(t(M2C),type="l",col=topo.colors(151), xlab="time"
,main="FTS with two regimes")
#below moving local medians applied to the above series, window lengths = 15 obs.,
#locality parameters betas = 0.45
result4A = rollapply(t(M2A),width = 15, wrapMBD, by.column = FALSE)
result4B = rollapply(t(M2B),width = 15,wrapMBD, by.column = FALSE)
result4C = rollapply(t(M2C),width = 15, wrapMBD, by.column = FALSE)
matplot(result4A,type="l",col=topo.colors(87), xlab="time",main="local 15-obs moving func-
tional median, beta=0.45")
matplot(result4B,type="l",col=topo.colors(87), xlab="time",main="local 15-obs moving func-
tional median, beta=0.45")
matplot(result4C,type="l",col=topo.colors(87), xlab="time",main="local 15-obs moving func-
tional median, beta=0.45")
```