MOX–Report No. 42/2014

Constrained Functional Time Series: an Application to Demand and Supply Curves in the Italian Natural Gas Balancing Platform

Canale, A.; Vantini, S.

MOX, Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox@mate.polimi.it http://mox.polimi.it

Constrained Functional Time Series: an Application to Demand and Supply Curves in the Italian Natural Gas Balancing Platform

Antonio Canale* & Simone Vantini†

October 23, 2014

Abstract

In Italy we have assisted to the recent introduction of the natural gas balancing platform, a system in which gas operators virtually sell and buy natural gas in order to balance the common pipelines network. Basically, the operators daily submit demand bids and supply offers which are eventually sorted according to price. Demand and supply curves are hence obtained by cumulating the corresponding quantities. Motivated by market dynamic modeling in the Italian Natural Gas Balancing Platform, we propose a model to analyze time series of bounded and monotonic functions. In detail, we provide the constrained functions with a suitable pre-Hilbert structure and introduce a useful isometric bijective map associating each possible bounded and monotonic function to an unconstrained. We then introduce a functional-to-functional autoregressive model that we use to predict the entire demand/supply function. We estimate the model by minimizing the squared L^2 distance between functional data and functional predictions with a penalty term based on the Hilbert-Schmidt squared norm of autoregressive lagged operators. We have proved that the solution always exist, unique and that it is linear on the data with respect to the introduced geometry thus

*Department of Economics and Statistics, University of Turin and Collegio Carlo Alberto, Turin, Italy (antonio.canale@unito.it)

†MOX, Department of Mathematics, Politecnico di Milano, Milan, Italy (simone.vantini@polimi.it)

guaranteeing that the plug-in predictions of future entire demand/supply functions satisfy all required constraints. We also provide an explicit expression for estimates and predictions. The approach is of general interest and can be generalized in any situation in which one has to deal with constrained monotonic functions (strictly positive or bounded) which evolve through time (e.g., dose response functions right-censored survival curves or cumulative distribution functions).

Keywords: Functional Data Analysis; Auto Regressive Model; Functional Ridge Regression;

1. INTRODUCTION

Supply and demand curves model is well known microeconomic model of price determination. Basically, it assumes that in a competitive market, the unit price for a particular good is determined at the intersection of two monotone curves having the traded quantity as abscissa and the current price as ordinate. The economic equilibrium, which determines the price, is obtained in that point where the quantity demanded by consumers (at current price) will equal the quantity supplied by producers (at current price). Demand and supply curves can be statistically estimated from price, quantity, and other exogenous variables using least squares or simultaneous equations (see Epple and McCallum 2006, for a recent application). However usual approaches are intrinsically descriptive rather than predictive and, even more, they rely on strict parametric structure for the regression functions.

Motivated by price prediction in Italian natural gas balancing market, we propose a model to forecast supply and demand curves evolving day by day. The approach is of general interest and can be generalized in any situations in which one has to deal with constrained functions which evolve through time. The proposed method is innovative both from the methodology perspective and from the application point of view.

On the one side, in fact, usual application of supply and demand curve model is descriptive and static. This clearly does not help the trader of a particular market to forecast possible price or help him in the decision making on bidding in the future. On the other side, usual forecasting methods, such as classical time series analysis, while producing useful forecast of quantities of interest (e.g., price), they do not provide the

insights of the market given by the supply and demand model. Even more, in markets with a moderate number of traders, the effect of a single non-standard offer or demand cannot be directly incorporated in the inference procedure or in what-if simulations. For all these reasons, the prediction of the supply and demand curves, and hence of their intersection can be of dramatic interest.

Functional data analysis (FDA) provides an extremely useful set of tools to deal with data that can be modeled as functions (e.g., demand and supply curves). Refer to Ramsay and Silverman (2005), Ramsay and Silverman (2002), or Ferraty and Vieu (2006) for a quick introduction to FDA. Differently from the most common framework in FDA, we hereby focus on functions that are constrained (i.e., monotonic, lower and upper bounded, and with an equality constraint on one edge of the domain and an inequality constraint on the other edge) and temporally dependent. From a practical point of view, we indeed aim at predicting future functions given past functions and controlling for shape constraints characterizing both past and future functions.

The current and past literature focussed separately on: (a) the problem of obtaining a constrained estimation of the underlying function given some point-wise evaluations of it and (b) on the problem of modeling functional data with temporal dependence (i.e., functional time series). At our knowledge, the present work is the first one in which constraints pertaining to monotonicity, boundedness, and values of the function at the boundary of the domain and the temporal dependence are jointly tackled. We hencefort refer to this joint framework as *Constrained Functional Time Series*.

Before going into the details of the mathematical modeling and of the estimation method that we propose inhere, we want to give a brief overview on the state of the art pertaining to both monotonic estimation and functional time series estimation.

The problem of having monotonic estimates of unknown functions observed just at some sparse points of the domain with possibly some measurement errors has been tackled in the literature for many decades even before the recent outbreak of functional data analysis. Isotonic regression is the first approach presented in the literature to this purpose and it has been for years the most common approach to this purpose (e.g., Passow and Roulier 1977; Winsberg and Ramsay 1980, 1981; Ramsay 1988; Mukerjee 1988; Mammen 1991; Kelly and Rice 1994; Mammen and Thomas-Agnan 1999). The basic

idea of this approach is to introduce a flexible functional basis (e.g., splines) to represent the function and to estimate the coefficients of the basis expansion by minimizing the residual some of squares under the constraint of monotonicity of the estimated functions. Typical choices rely on the use of an I-spline basis with a positive constraint on the coefficients or on the use of a B-spline basis with equally spaced knots with a monotonicity constraint on the coefficients. A similar approach has been proposed more recently in the framework of kernel regression (Hall and Huang 2001; Henderson et al. 2008). The basic idea of these latter works is to locally modify the local kernels such that the residual sum of squares is minimal and the estimated function is monotonic. Another approach is the projection method (Friedman and Tibshirani 1984; Bloch and Silverman 1997; Mammen et al. 2001). The basic idea of this method is to estimate the unknown function in an unconstrained fashion, then project the estimated function onto the convex subspace of the monotonic functions (i.e., looking for the monotonic function being the closest to the unconstrained estimated function according to a given metric). The approach we are going to use is in line with the transform/back-transform method. This method is very spread in the literature pertaining to functional data analysis and has been firstly proposed by Ramsay and Silverman (2005, 2002). The idea is to find a bijective map from the space of the generic functions to the convex subspace of the monotonic functions. The idea is indeed to transform data such to perform an unconstrained estimation on the transformed data and then back-transform the estimated function to the convex subspace of the monotonic functions. Some very recent works (Egozcue et al. 2006; Menafoglio et al. 2013; Boogaart et al. 2014) focusing on modeling cumulative distribution functions of absolutely continuous random variable (inspired by the pioneering work by Aitchison (1982) on compositional data) formalized this approach by introducing a suitable Hilbert structure on the set of the cumulative distribution functions (i.e., Bayes linear space) and an isometric bijective map to L^2 for transforming and back-transforming functional data and conveniently map the entire statistical analysis in an unconstrained framework (i.e., L^2). In the present paper, instead, we are going to introduce a suitable pre-Hilbert structure on the set of monotonic, lower and upper bounded functions satisfying an equality constraint on one edge of the domain and an inequality constraint on the other one and

an associated isometric bijective map to L^2 allowing us to model temporal dependence in an unconstrained framework. In the rest of the manuscript we will refer to it as the $\mathcal{M}^2$ space. To our knowledge, this is the first time that a geometry in a functional space is introduced and formalized such to obtained a sound theoretical framework to model temporal dependence between constrained functional data.

The literature dealing with the temporal dependence between functional data is instead more recent and dates at the beginning of this century. Functional autoregressive models (FAR) are the most used approach, both for their ease of interpretation and performance in applications (Elezović 2009). FAR models, i.e. autoregressive models in which scalar random variable are replaced with random functions, do not have to be confounded with functional-coefficient autoregressive models (e.g. Chen and Tsay 1993; Fan and Yao 2002) that are instead scalar models in which the dependence between the past values and the current value of the time series is non-linear dependence and estimated in a generalized additive model perspective.

In functional autoregressive models, autoregressive parameters are replaced by Hilbert-Schmidt operators and thus model estimation is declined in the estimation of the autoregressive operators. To this purpose different methods have been presented in the literature, though at the current state of the art none seems to overcome the others (Hormann and Kokoszka 2012).

Autoregressive operators are directly linked with lagged autocovariance operators (e.g., Kargin and Onatski 2008) and thus a first possible approach is to estimate the lagged autocovariance operators from the functional time series and then use the these estimates to estimate the autoregressive operators. Because of the infinite dimensionality of functional data and to obtain more stable estimates, sample autocovariance operators are typically replaced by reduce rank approximations. A very common approach relies on functional principal component decomposition (e.g., Shang 2013) and the use of a reduce number of principal components. This approach, though easy and still widely spread, has been recently shown to fail the comparison in terms of prediction with an alternative reduce rank representation based on predictive factors (Kargin and Onatski 2008). Another reduce rank approximation presented in the literature is based on a wavelet expansion of the original data (Antoniadis and Sapatinas 2003).

A second approach to the estimation of the autoregressive operators is the direct minimization of the mean squared error of prediction. In order to avoid over-fitting due by the infinite dimensionality of functional data, the minimization problem has to be dealt with some care. For instance, Fan and Zhang (2000) and Elezović (2009) used a two-step approach: (i) they estimate a concurrent functional autoregressive model, i.e. a model in which the cross-effects between different parts of the domains are set to zero, this is indeed a continuous family of point-wise scalar autoregressive models, and (ii) they smooth the obtained autoregressive functions to take into account the effects of neighborhood points of the domain in determining the value at a given point of tomorrow function. In the present paper, instead, we will directly target at the minimization of the mean squared error of prediction in a perspective that is more consistent with current research in functional data analysis by introducing in the objective function a penalty term involving the squared Hilbert-Schmidt norm of the autoregressive operators. With this approach, a full rank operator is obtained by shrinking the set the degenerative solutions (that one would obtain without penalty) toward a temporal independence scenario (that one would obtain by setting the penalty constant to infinity). In detail we prove the existence and uniqueness of the estimators and provide their explicit expressions.

The rest of the paper is structured as follows. In Section 2 we first introduce the space $\mathcal{M}^2(a,b)$ and an isometric bijective map to $L^2(a,b)$, and describe in detail the $\mathcal{M}^2$-FAR model we are using, with particular emphasis on model estimation. Section 3 describes instead our motivating context and discuss the application of our methodology to the Italian natural gas balancing market data. Section 4 summarizes the results and discusses possible generalizations.

2. MODEL AND METHODS

2.1 The space $\mathcal{M}^2(a,b)$: geometry and mapping functions

Let $\mathcal{M}^2(a,b)$ be the family of differentiable functions $g : [a,b] \to [0,1]$ such that: (i) $g(a) = 0$, $g(b) < 1$, and (ii) $0 < m_g \leq g'(s) \leq M_g < +\infty$ for all $s \in [a,b]$. Previous conditions imply also that $\mathcal{M}^2(a,b) \subset L^2(a,b)$ and that all functions belonging to $\mathcal{M}^2(a,b)$ are monotonic increasing and bounded. Note that if condition $g(b) < 1$ is

replaced with $g(b) = 1$ we obtain exactly the same conditions required to define the Bayes space geometry introduced and developed in Egozcue et al. (2006); Menafoglio et al. (2013); Boogaart et al. (2014). With respect to the curves studied in these works (that are valued 1 at the right edge of the domain) the curves we are dealing with are subject to a right censoring effect which makes them valued less than 1 at the right edge of the domain of observation.

We first introduce a bijective map from $\mathcal{M}^2(a,b)$ to $L^2(a,b)$ such that for any $g \in \mathcal{M}^2(a,b)$, then

$$f(s) = \log\left(\frac{g'(s)}{1-g(s)}\right) \tag{1}$$

is its image which belongs to $L^2(a,b)$. By applying the exponential function and integrating between a and $s \in [a,b]$ in both sides of equation (1) we obtain the inverse transformation:

$$g(s) = 1 - \exp\left(-\int_a^s \exp\left(f(u)\right) \mathrm{d}u\right). \tag{2}$$

Note that the direct transformation look at g as a cumulative distribution function of a scalar absolutely-contonuous random variable and maps it into the natural logarithm of the corresponding hazard function. We will thus call it log-hazard transformation and its inverse transformation anti-log-hazard transformation and indicate them with logH and logH^{-1}, respectively. In particular we have that constant function in $L^2(a,b)$ are linked to exponential functions in $\mathcal{M}^2(a,b)$: $\forall c \in \mathbb{R}\ f(s) = c \leftrightarrow g(s) = 1 - e^{-e^c(s-a)}$, with the special case of the null function in $L^2(a,b)$ which is linked to the exponential function with unitary decay rate in $\mathcal{M}^2(a,b)$ (i.e., $f(s) = 0 \leftrightarrow g(s) = 1 - e^{-(s-a)}$).

In the rest of the section, we will build a entire geometry on $\mathcal{M}^2(a,b)$ which makes the log-hazard transformation isometric with respect to the geometry induced by the usual inner product in $L^2(a,b)$. We start making $\mathcal{M}^2(a,b)$ a vector space defining the operations of addition and scalar multiplication.

Definition 1. *Let $g_1, g_2 \in \mathcal{M}^2(a,b)$, $\alpha \in \mathbb{R}$. We define*

- *the addition of g_1 and g_2 as the operation $\oplus : \mathcal{M}^2(a,b) \times \mathcal{M}^2(a,b) \to \mathcal{M}^2(a,b)$ given by*

$$g_1 \oplus g_2 = 1 - \exp\left(-\int_a^s \frac{g_1'(u)}{1-g_1(u)} \cdot \frac{g_2'(u)}{1-g_2(u)} du\right), \tag{3}$$

- *the scalar multiplication of g_1 by α as the operation $\odot : \mathbb{R} \times \mathcal{M}^2(a,b) \to \mathcal{M}^2(a,b)$ given by*

$$\alpha \odot g_1 = 1 - \exp\left(- \int_a^s \left(\frac{g_1'(u)}{1 - g_1(u)} \right)^{\alpha} du \right). \tag{4}$$

Note that the neutral element of addition $\oplus$ is $1 - e^{-(s-a)}$ (i.e., the cumulative distribution function of the exponential distribution with unitary decay rate) and the neutral element of scalar multiplication $\odot$ is 1.

We are now introducing a suitable geometry in $\mathcal{M}^2(a,b)$ to make the log-hazard transformation an isometry between $\mathcal{M}^2(a,b)$ and the image of the log-hazard transformation $\text{logH}\left(\mathcal{M}^2(a,b)\right)$ embedded in $L^2(a,b)$. In detail, we are defining an inner product in the functional vector space $\mathcal{M}^2(a,b)$ and the corresponding norm and distance as follows:

Definition 2. *Let $g_1, g_2 \in \mathcal{M}^2(a,b)$. We define the inner product of g_1 and g_2 as $< \cdot, \cdot >_{\mathcal{M}^2}$: $\mathcal{M}^2(a,b) \times \mathcal{M}^2(a,b) \to \mathbb{R}$ given by*

$$< g_1, g_2 >_{\mathcal{M}^2} = \int_a^b \log\left(\frac{g_1'(s)}{1 - g_1(s)} \right) \log\left(\frac{g_2'(s)}{1 - g_2(s)} \right) ds. \tag{5}$$

Definition 3. *Let $g_1, g_2 \in \mathcal{M}^2(a,b)$. The metric $d_{\mathcal{M}^2}(\cdot,\cdot) : \mathcal{M}^2(a,b) \times \mathcal{M}^2(a,b) \to \mathbb{R}_0^+$ and the norm $||\cdot||_{\mathcal{M}^2} : \mathcal{M}^2(a,b) \to \mathbb{R}_0^+$ induced by the inner product (5) are defined as:*

$$d_{\mathcal{M}^2}(g_1, g_2) = \left[\int_a^b \left\{ \log\left(\frac{g_1'(s)}{1 - g_1(s)} \right) - \log\left(\frac{g_2'(s)}{1 - g_2(s)} \right) \right\}^2 ds \right]^{1/2}, \tag{6}$$

$$||g_1||_{\mathcal{M}^2} = \left[\int_a^b \left\{ \log\left(\frac{g_1'(s)}{1 - g_1(s)} \right) \right\}^2 ds \right]^{1/2}. \tag{7}$$

Note that even though the functional vector space $\mathcal{M}^2(a,b)$ is closed with respect to linear combinations of elements as defined in (3) and (4), it is not complete with respect to the metric $d_{\mathcal{M}^2}$ induced by the inner product $< g_1, g_2 >_{\mathcal{M}^2}$ defined in (5). For example, monotonic non-decreasing step-wise functions belong to the closure of $\mathcal{M}^2(a,b)$ and not to $\mathcal{M}^2(a,b)$ itself. This makes $\mathcal{M}^2(a,b)$ just a pre-Hilbert space. It is quite straightforward to make it Hilbert: this would require to close the space and define on the closure the operations of addition and scalar multiplication consistently with (3) and (4). This could be done relying on the separability of $L^2(a,b)$. This

$$
\begin{array}{rcl}
\mathcal{M}^2(a,b) & \xrightarrow{\text{logH}} & \text{logH}\left(\mathcal{M}^2(a,b)\right) \subset L^2(a,b) \\
g_1 & \longmapsto & f_1 = \text{logH}(g_1) \\
 & & \\
g_1 \oplus g_2 & \longleftrightarrow & f_1 + f_2 \\
\alpha \odot g_1 & \longleftrightarrow & \alpha \cdot f_1 \\
 & & \\
< g_1, g_2 >_{\mathcal{M}^2} & = & < f_1, f_2 >_{L^2} \\
d_{\mathcal{M}^2}(g_1, g_2) & = & d_{L^2}(f_1, f_2) \\
||g_1||_{\mathcal{M}^2} & = & ||f_1||_{L^2}
\end{array}
$$

Figure 1: Relations between $\mathcal{M}^2(a,b)$ and image in $L^2(a,b)$ of the map logH.

extension (which would make logH an isometric bijective map with the entire space $L^2(a,b)$ and not only with its image $\text{logH}\left(\mathcal{M}^2(a,b)\right) \subset L^2(a,b)$)) is out of the scope of this work being the pre-Hilbert nature of $\mathcal{M}^2(a,b)$ the minimal condition to make the estimation and prediction process described in the next section self-consistent. Indeed, as detailed in Corollary 1 the predictions provided by the estimated model are linear combinations in the sense defined in (3) and (4) of functions in $\mathcal{M}^2(a,b)$ which are guaranteed to be in $\mathcal{M}^2(a,b)$. Figure 1 summarizes all relations between $\mathcal{M}^2(a,b)$ and $L^2(a,b)$.

2.2 Functional Auto Regressive Model

We here describe the model we use for dealing with temporal dependence. The model can be equivalently formulated on the original data in the space $\mathcal{M}^2(a,b)$ or on the log-hazard transformed data in $L^2(a,b)$. To help the intuition, we here report the latter formulation which is indeed the one used in the practice for computations. We will denote this model as $\mathcal{M}^2$-FAR. Let $\{f_t\}_{t=1}^T$ be a collection of random functions in $L^2(a,b)$ (here the log-hazard transformed functions $f_t = \text{logH}(g_t)$) generated sequentially through discrete time t. We assume that $f_t(s)$ depends on the values assumed by the random functions earlier appearing in the sequence potentially at each domain location $s \in [a,b]$. Let us model this temporal dependence conditionally through an (non-concurrent) autoregressive functional time series. Precisely a (non-concurrent)

functional auto regressive model of order p (i.e., FAR(p)) is defined as

$$f_t = \alpha + \sum_{j=1}^{p} \Psi_j f_{t-j} + \epsilon_t, \tag{8}$$

or equivalently

$$f_t(s) = \alpha(s) + \sum_{j=1}^{p} \int_a^b \psi_j(s,u) f_{t-j}(u) \mathrm{d}u + \epsilon_t(s) \ \forall s \in [a,b], \tag{9}$$

where the Hilbert-Schmidt operator Ψ_j plays the role of the j-th lagged autoregressive parameter. The bi-variate function $\psi_j(u,s) \in L^2(a,b) \times (a,b)$ is its kernel which determines the impact of $f_{t-j}(u)$ on $f_t(s)$. Finally, ϵ_t are the innovation terms which are *i.i.d.* zero-mean finite-variance random functions.

2.3 Model estimation and prediction

As mentioned in the introductory section we estimate the lagged autoregressive operators Ψ_j and the non-centrality function α by direct minimization. In detail, we get the estimates as the solution of the following penalized minimization problem:

$$\min_{\alpha \in L^2 \ \cap \ \{\Psi_j\}_{j=1,\dots,p} \ \subseteq \ \mathrm{HS}} \left(\sum_{t=p+1}^{T} \left|\left| f_t - \left(\alpha + \sum_{j=1}^{p} \Psi_j f_{t-j} \right) \right|\right|_{L^2}^2 + \lambda \sum_{j=1}^{p} ||\Psi_j||_{\mathrm{HS}}^2 \right) . \tag{10}$$

The first term of the objective function is the sum of the squared residuals (between the observed values and their predictions) according to the L^2 metric. The lower this term, the better the fit of predictions to data. The second term is instead the sum of the squared Hilbert-Schmidt norms of the lagged autoregressive operators Ψ_j. The lower this term, the lower is the autocorrelation associated to the estimated model. Finally λ is the positive penalty constant which defines the relative weights of the two terms in the objective function.

It is of interest to discuss the two limit cases $\lambda \to +\infty$ and $\lambda \to 0^+$. As the penalty constant comes bigger the estimated model is pulled towards models with less memory (i.e., model in which the effect of the last p functions on the present function is weaker). Indeed when $\lambda \to +\infty$ the estimated model is the trivial model with $\hat{\Psi}_j = \mathbf{0}$ (i.e., the null operator) for $j = 1, \dots, p$ and $\hat{\alpha} = \frac{1}{T-p} \sum_{t=p+1}^{T} f_t$. In $\mathcal{M}^2(a,b)$ this model

describes a sequence of *i.i.d.* random functions and thus trivially leads to predict future curves with the Fréchet $\mathcal{M}^2$-mean of the observed curves:

$$\hat{g}_{T+1}(s) = 1 - \exp\left(- \int_a^s \prod_{t=p+1}^{T} \left(\frac{g_t'(u)}{1 - g_t(u)} \right)^{\frac{1}{T-p}} \mathrm{d}u \right) \tag{11}$$

At the other extreme, when $\lambda \to 0^+$, the estimated model converges toward an interpolating model, in particular (consistently with Theorem 1) the one with the minimal Hilbert-Schmidt norms of the lagged autocorrelation operators Ψ_j. As common practice in functional data analysis, the choice of a suitable value of the penalty constant λ can be addressed either by minimization of the prediction residual sum of squares on a test sample, by crossvalidation, or even heuristically by looking - in this case - at the smoothness/roughness of the kernels of the estimated lagged autocorrelation operators Ψ_j and/or of the predictions (i.e. the so called Goldilocks' method). If one desires little bias in the estimates, small values of λ might be favored. If instead more robust estimates are desired, larger values of λ might be favored. Anyhow for any choice of the value of λ the following theorem proves the existence and uniqueness of the estimators of α and Ψ_j for $j = 1, \ldots, p$ and provides also their explicit expressions.

Theorem 1 (Existence, uniqueness, and explicit expression of the estimators)**.** *For any $\lambda > 0$, the solution of minimization problem (10) always exists unique for $j = 1, \ldots, p$ and $s, u \in [a, b]$ and is equal to:*

$$\begin{aligned} \hat{\psi}_j(s,u) &= \sum_{t=p+1}^{T} \Big(\mathbb{P}_\lambda^{-1}(\mathbf{f}_t - \bar{\mathbf{f}}) \Big)\big((j-1)(b-a) + u\big) \Big(f_t(s) - \bar{f}_{[0]}(s) \Big); \\ \hat{\alpha}(s) &= \bar{f}_{[0]}(s) - \sum_{j=1}^{p} \int_a^b \hat{\psi}_j(s,u) \bar{f}_{[j]}(u)\, du; \end{aligned}$$

where $\bar{f}_{[j]} = \frac{1}{T-p} \sum_{t=p+1}^{T} f_{t-j}$; $\mathbf{f}_t \in L^2\big(a, b + p(b-a)\big)$ is the function obtained by chaining $f_{t-1}, \ldots, f_{t-p}$, i.e. $\mathbf{f}_t\big((j-1)(b-a) + s\big) = f_{t-j}(s)$; $\bar{\mathbf{f}} = \frac{1}{T-p} \sum_{t=p+1}^{T} \mathbf{f}_t$; and $\mathbb{P}_\lambda : L^2\big(a, b + p(b-a)\big) \to L^2\big(a, b + p(b-a)\big)$ is the HS operator with kernel $\sum_{t=p+1}^{T} \big\{ \big(\mathbf{f}_t(\tilde{s}) - \bar{\mathbf{f}}(\tilde{s})\big)\big(\mathbf{f}_t(\tilde{u}) - \bar{\mathbf{f}}(\tilde{u})\big) + \lambda \big\}$ with $\tilde{s}$ and $\tilde{u} \in \big(a, b + p(b-a)\big)$.

Proof. See Appendix. □

Finally, before moving to the Italian Natural Gas Market application that has urged

this research, we want to point out the coherence of the entire working pipeline starting from the time series $\{g_1, \ldots, g_T\} \in \mathcal{M}^2(a,b)$ to the prediction of the future function g_{T+1}, which is the final aim of the work. The following Corollary states indeed that the joint use of the geometry introduced in Section 2.1 (i.e., the pre-Hilbert space $\mathcal{M}^2(a,b)$) and the model estimation procedure described in this section (i.e., minimization of (10)) guarantees the plug-in predictions to satisfy all constraints characterizing functions belonging to $\mathcal{M}^2(a,b)$.

Corollary 1 (Linearity of predictions)**.** *The plug-in prediction*

$$\hat{g}_{T+1} = \mathrm{logH}^{-1}\left(\hat{\alpha} + \sum_{j=1}^{p} \hat{\Psi}_j \ \mathrm{logH}(g_{T+1-j})\right)$$

is a linear combination in $\mathcal{M}^2(a,b)$ of $\{g_1, \ldots, g_T\}$ and thus belongs to the space $\mathcal{M}^2(a,b)$.

Proof. See Appendix. □

Remark 1. *Note the proposed estimation method (i.e., the minimization of (10)) is here introduced for the estimation of functional autoregressive models. Nevertheless, it is trivial to extend it to the estimation functional-to-functional non-concurrent regression models by simply replacing the function f_t with a generic functional response y_i and the p lagged functions $f_{t-1}, \ldots, f_{t-p}$ with p generic functional regressors $x_{1i}, \ldots, x_{pi}$:*

$$\min_{\alpha \in L^2 \ \cap \ \{\Psi_j\}_{j=1,\ldots,p} \ \subseteq \ HS} \left(\sum_{i=1}^{n} \left|\left| y_i - \left(\alpha + \sum_{j=1}^{p} \Psi_j x_{ji}\right)\right|\right|_{L^2}^2 + \lambda \sum_{j=1}^{p} ||\Psi_j||_{HS}^2 \right) ,$$

with $\{y_i\}_{i=1,\ldots,n}$ being the functional responses and $\{x_j i\}_{\{j=1,\ldots,p \quad i=1\ldots,n\}}$ being the functional regressors. This sets our estimation method as a functional generalization of ridge regression (Hastie et al. 2009).

3. APPLICATION TO ITALIAN NATURAL GAS MARKET

3.1 Context

In the last decade the natural gas market has been extensively studied and discussed from an economic, political and environmental viewpoint. In Europe, for example,

several legislative and infrastructural measures have been undertaken to regulate this market. Among them the legal splitting of pipeline managers and gas shippers and the legislation for obligatory third party access to transmission, distribution, storage and liquefied natural gas capacity (European Union 2003).

Such measures, while favoring a liberal market, have created logistic new challenges. In Italy, likewise other markets, the control of the national pipeline has been split from the national main natural gas shipper causing uncertainty in the physical balancing of the network. Under this scenario, several shippers inject natural gas into the network from different importing countries such as Algeria or Russia. The gas is then consumed by civil, industrial and thermo-electric stations spread through the country. The role of the pipeline manager, in Italy *Snam Rete Gas* (SNAM), is to compensate injections and consumptions via storage or other measures. In fact, the risk of possible unbalance is assigned to each shipper which has to daily predict and communicate to SNAM its injection and consumption forecasts on which a penalty is payed for any positive or negative unbalance. It turns out that to minimize such a risk, natural gas shippers should deal with statistical forecast of consumptions.

With the final aim of having a self balancing system, the Italian natural gas balancing platform (PB-GAS), was introduced in December 2011. The PB-GAS is a system where gas operators and traders virtually sell and buy natural gas in order to balance the common pipeline. The PB-GAS is managed by the energy regulatory *Gestore Mercati Energetici* (GME), with SNAM acting as central counterpart for all daily offers. Every day SNAM submits a demand bid or supply offer for a volume of gas corresponding to the overall imbalance of the system while the operators submit demand bids and supply offers for the storage resources they have available.

In this situation, demand bids and supply offers are sorted from the highest to the lowest and from the lowest to the highest respectively, so that demand and supply curves are obtained as the cumulative sum of the quantities in gigajoules (GJ). The selection of bids/offers accepted on the PB-GAS is based on the auction mechanism so that every offer to the left of the intersection of the two curves is accepted and exchanged at the resulting price.

While balancing the network, each shipper can also take advantage of the market

in a speculative perspective, buying natural gas at lower price or selling exceeding gas at higher price, with respect to their benchmark supplying indexes. It is clear that forecast tools are dramatically important for the decision-making of each shipper. For example, the mere price prediction is a fist procedure to implement. However, the pointwise or interval price forecast alone, is of limited utility when the effects of non-standard bids can strongly modify the tomorrow demand or supply curves shape, and thus the resulting equilibrium price. It would be much more useful to have a prediction of the entire demand and supply curves. With such tools at hand, traders can directly see the effect of their bids on the shape of tomorrow curve and on price itself.

3.2 Data description

The data available refer to the first thirteen monts of the PB-GAS, namely from December 1st, 2011 to December 31st, 2012. The data are available at the website of GME (2013). The original data are reported in a XML table format, where each row represents a single awarded bid with its own code, date, trader name, type (sell or buy), awarded price, and awarded quantity. For each day, we build the supply (and demand) curve ordering the selling bids increasingly (decreasingly) by price and obtaining the value of the quantities by cumulating each single awarded quantity.

Before applying the model described in Section 2 row data has been converted to functional data in $\mathcal{M}^2(a, b)$ with $a = 0$ and $b = 1.2 \times 10^7$, which can be considered as a conservative upper bound of the range of investigation. In detail, the smoothed versions of the supply (and demand) curves are obtained by means of local polynomial regression as implemented in the R function `locpoly` of `KernSmooth` package. To preserve the constraint in zero for each curve we added an artificial point on the negative axis in order to have the local polynomial regression output equal to zero for the supply and to 23 for the demand. The final smoothed curves are then scaled by 23^{-1} to make them constrained between zero and one. As an example of the output of the preprocessing, in Figure 2 the supply (and demand) curve for a particular day are reported in the raw and smoothed versions. The obtained smoothed functional time series are plotted in Figure 3. Their evaluation on a grid of 500 points are provided in the Supplementary Materials.

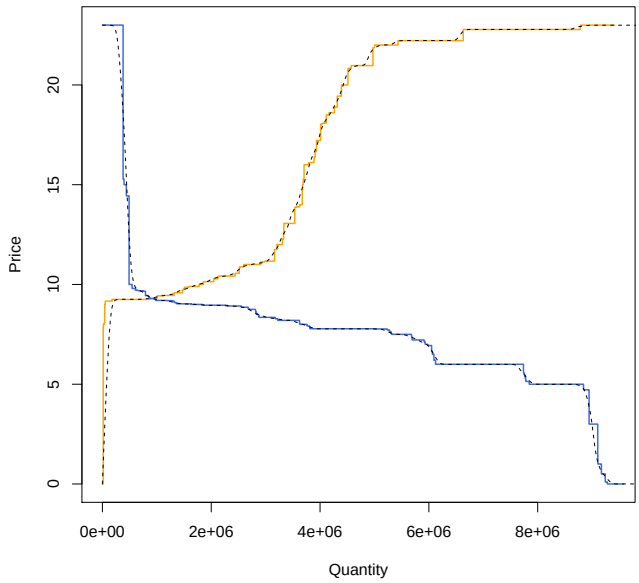
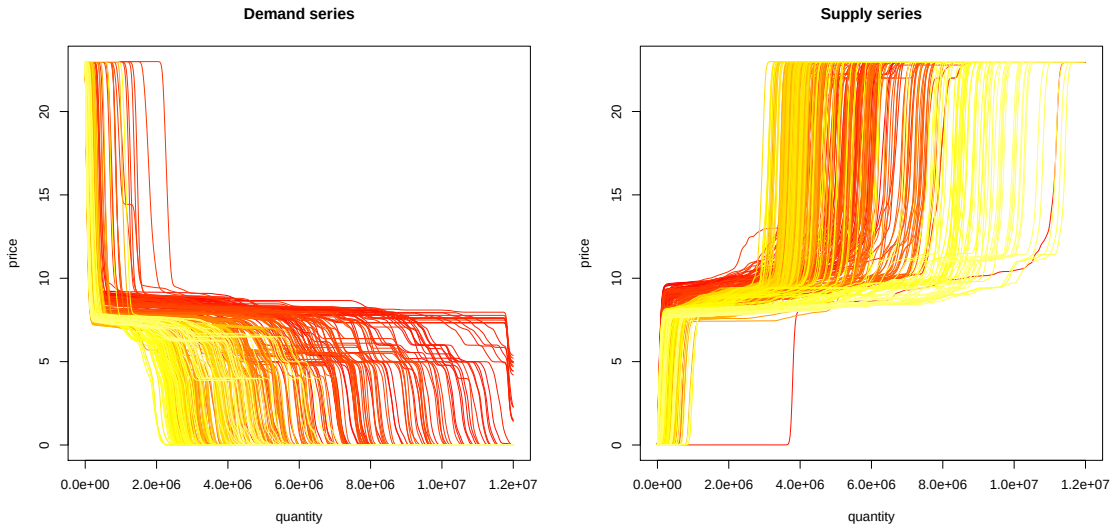


Figure 2: Real demand and supply curves February, 8 2012 (solid lines) and smoothed versions (dashed lines)



(a) (b)

Figure 3: Smoothed functional time series of demand (a) and supply (b) curves. Color denotes time, with the oldest curves in dark and the most recent ones in bright.

3.3 Results

Let $g_t^D(s)$ and $g_t^S(s)$ be the demand and supply curve of day t, respectively, expressing the price in function of quantity. We apply the methodology presented in Section 2 to model $\{g_t^D\}_{t=1}^T$ and $\{1 - g_t^S\}_{t=1}^T$. In particular, we estimated the $\mathcal{M}^2$-FAR(p) models described in the previous section, for different choices of the auto-regressive order p, separately for the two time series. As in classical time series analysis, the model can be easily extended to model both the functional time series jointly, if needed. The penalization parameter is fixed here to $\lambda = 10^{-8}$ for all p.

To asses model goodness of fit, for increasing value of the auto-regressive order p, we compare the $\mathcal{M}^2$ root mean squared error between the estimated curves and the original ones:

$$\mathcal{M}^2\text{-}RMSE = \sqrt{\frac{1}{T-p}\sum_{p+1}^{T} d_{\mathcal{M}^2}(g_t, \hat{g}_t)^2}.$$

Note that thanks, to the isometric nature of the log-hazard transformation, the latter ones coincide with the L^2 root mean squared error between the estimated transformed curves and the original transformed ones:

$$L^2\text{-}RMSE = \sqrt{\frac{1}{T-p}\sum_{p+1}^{T} ||f_t - \hat{f}_t||_{L^2}^2}.$$

The results are reported in Table 1. As a reference, we fit a $\mathcal{M}^2$-FAR(0) model (i.e., a model ignoring the temporal autocorrelation and thus predicting future curves with the Frechét $\mathcal{M}^2$-mean of the curves as defined in (11)). First, we note that the auto-regressive order p has similar impacts in the estimation of both the demand and supply curves (i.e., for fixed p, the errors for the demand and supply curves are roughly of the same order of magnitude). As expected, increasing p the estimates improve, with the larger improvements observed moving from $p = 1$ to $p = 2$ and moderate ones moving from $p = 0$ to $p = 1$ and from $p = 2$ to $p = 3$ suggesting $p = 2$ as a good candidate value.

Further insights for the choice of the $\mathcal{M}^2$-FAR order p could come from extending classical time series identification tools to the functional framework. In scalar time series analysis, it is a common practice to look at the auto-correlation and partial

Table 1: $\mathcal{M}^2$-$RMSE$ between the estimated curves and the original ones.

	Demand	Supply
$\mathcal{M}^2$-FAR(0)	7.98×10^3	5.96×10^3
$\mathcal{M}^2$-FAR(1)	6.12×10^2	4.74×10^2
$\mathcal{M}^2$-FAR(2)	1.05×10^{-3}	6.25×10^{-4}
$\mathcal{M}^2$-FAR(3)	1.08×10^{-4}	6.06×10^{-5}

auto-correlation of the time series prior the analysis. Indeed, following the classical Box-Jenkins approach, the first step of the modeling procedure consists in evaluating the autocorrelation and partial autocorrelation functions for different values of the lag and deciding which (if any) autoregressive or moving average component should be used (Box et al. 2013). To perform a similar investigation in the $\mathcal{M}^2$-FAR framework, we here introduce a measure of functional auto-correlation and of functional partial auto-correlation playing the roles of their scalar correspondents. Let us define the functional autocorrelation function of lag k of a stationary functional time series $\{f_t\}_{t=1}^T$ the function

$$R_k(s,u) = \frac{E[(f_t(s) - E[f_t(s)])(f_{t+k}(u) - E[f_{t+k}(u)])]}{\sqrt{E[(f_t(s) - E[f_t(s)])^2]E[(f_{t+k}(u) - E[f_{t+k}(u)])^2]}}. \tag{12}$$

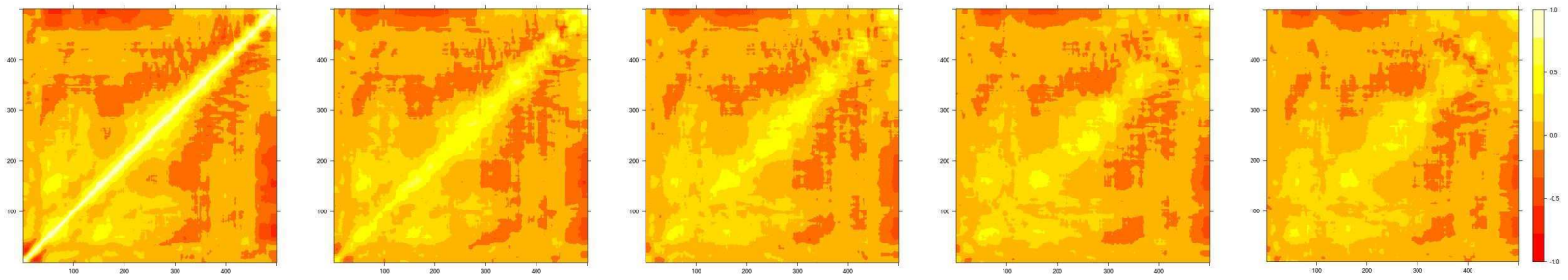
which expresses the correlation between $f_t(s)$ and $f_{t+k}(u)$. Even though one could integrate (12) and obtain a scalar measure of autocorrelation, which can be plotted as a function of k as a standard correlogram, to better understand the dependence across functions observed at different times, we focused on the visual comparison of the functional autocorrelation function (12) along p. In analogy with the definition of scalar partial autocorrelation, define also the functional partial autocorrelation function of order $k > 2$ as

$$\Gamma_k(s,u) = \frac{E[r_t^{k*}(s)\, r_{t+k}^k(u)]}{\sqrt{E[(r_t^{k*}(s))^2]E[(r_{t+k}^k(u))^2]}}, \tag{13}$$

with $\Gamma_0(s,u) = R_0(s,u)$, $\Gamma_1(s,u) = R_1(s,u)$ and where $\{r_{t+k}^p\}$ are the functional residuals at time $t+k$ of the functional auto-regressive model 8 of order p and $\{r_t^{p*}\}$ are the function residuals at time t of the functional auto-regressive model 8 of order k, fitted to the reversed series.

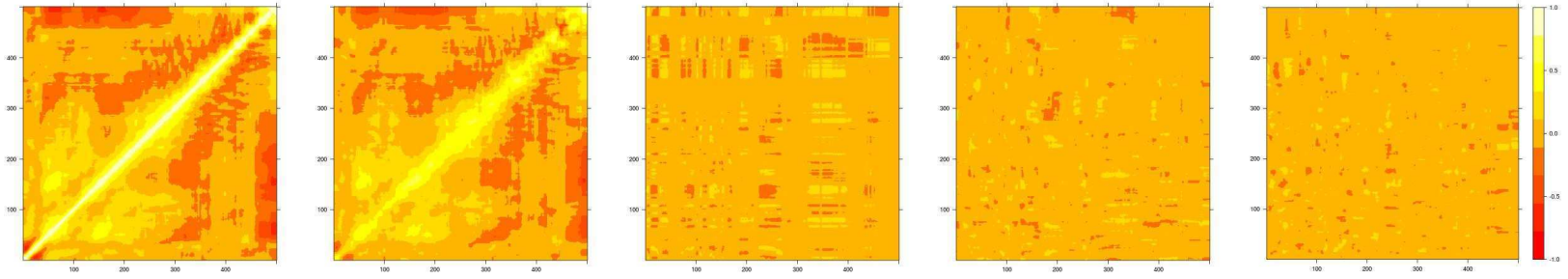
Figures 4 and 5 show the sample autocorrelation function and partial autocorrelation function, respectively, for $k = 0, 1, \ldots 4$ for the demand series only. Qualitatively

similar results were obtained for the supply series. Figure 4 shows that the auto-correlation is persistent also for increasing lags, a typical feature registered in scalar auto-regressive models. Some details of the autocorrelation function are amenable to an application interpretation. First, higher autocorrelation is registered in the first part of the curves domain. This is indeed the region where, typically, the demand and supply curves intersect. It is clear that yesterday's price influences the bids (and thus the curves shape) of today, so it is natural to expect high autocorrelation in this part of the domain. Second, the auto-correlation remains high also for increasing lags mostly around the diagonal of the plots in Figure 4. This means that the value of the curves at point s is mainly influenced by the previous observed curves in a neighborhood of s. The structure described by Figure 5, suggests that the main dependence of curve t from the past, comes from the curve observed at $t-1$. In fact the dependence of the curve at time t from the curve at time $t-2$, is basically zero. This suggests that an $\mathcal{M}^2$-FAR(1) may be sufficiently appropriate to fit our data according to this measure.



(a) $k = 0$ (b) $k = 1$ (c) $k = 2$ (d) $k = 3$ (e) $k = 4$

Figure 4: Sample functional autocorrelation function $k = 0, 1, \ldots 4$ for the demand series.



(a) $k = 0$ (b) $k = 1$ (c) $k = 2$ (d) $k = 3$ (e) $k = 4$

Figure 5: Sample functional partial autocorrelation function $k = 0, 1, \ldots 4$ for the demand series.

Table 2: RMSE for price (in Euro) and quantity (in GJ) obtained as crossing point between estimated curves.

	quantity (GJ)	price (Euro)
$\mathcal{M}^2$-FAR(1)	336430	2.03
$\mathcal{M}^2$-FAR(2)	336053	0.18
$\mathcal{M}^2$-FAR(3)	336053	0.18

As additional measure of goodness of fit, we consider an application driven approach. As motivated in Section 3.1, the whole curve prediction is a more informative tool than the mere price prediction. However, price forecast is a byproduct of our procedure (it can be easily obtained as the intersection of the two predicted curves), and it is desirable that such a prediction is reliable. We calculated the predicted daily prices and quantities, and compare those values with the real prices and quantities resulted in the daily bids, in terms of root mean squared errors, reported in Table 2.

We are now focusing on price prediction, since, as said, it is the most important feature to forecast for a trader. As benchmark we fitted a classical ARIMA model to the scalar time series of prices. After differentiating the series to make it stationary, an inspection of the correlograms suggest an ARMA(1,1) order. The root mean squared error under this specification is of 0.20 Euro. With a simple $\mathcal{M}^2$-FAR(1) the prediction error for the price is greater than 2 Euro, which is much higher than the benchmark prediction. On the contrary, with a $\mathcal{M}^2$-FAR(2), the benchmark prediction error is outperformed obtaining an error of 0.18 Euro. Figure 6, report the original price time series along with the predictions obtained with the standard ARIMA models, and with the $\mathcal{M}^2$-FAR(2) models. Given the performances in predicting both the supply and demand curves and the price and quantity (despite the suggestion of Figures 5), we choose the $\mathcal{M}^2$-FAR(2) model as the final model for the analysis.

To conclude the analysis we give an example of the tremendous additional insights that the whole curve forecast can give to traders. Consider the prediction for day T reported in panel (a) of Figure 7. Suppose that a given trader, is aware that he is going to buy a large quantity Q of natural gas tomorrow urged by legislative and logistic reasons (for all the details, please refer to the PB-GAS normative, in GME (2013) website). To lower the price, the trader can submit an extra non-standard supply offer for a small quantity, that eventually he is going to buy above Q. For

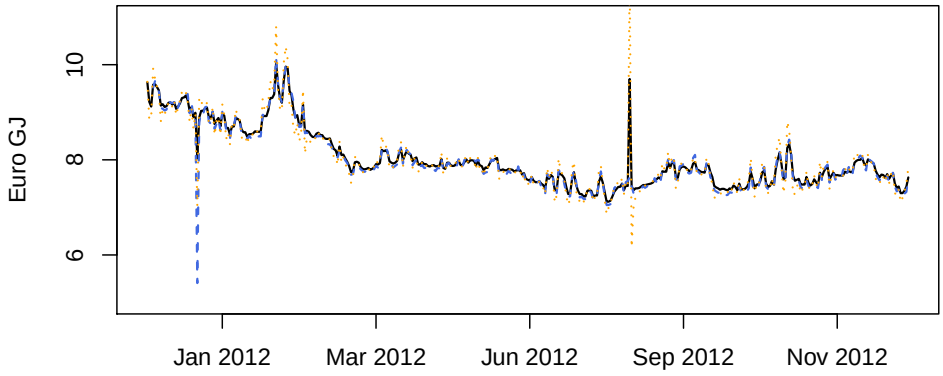
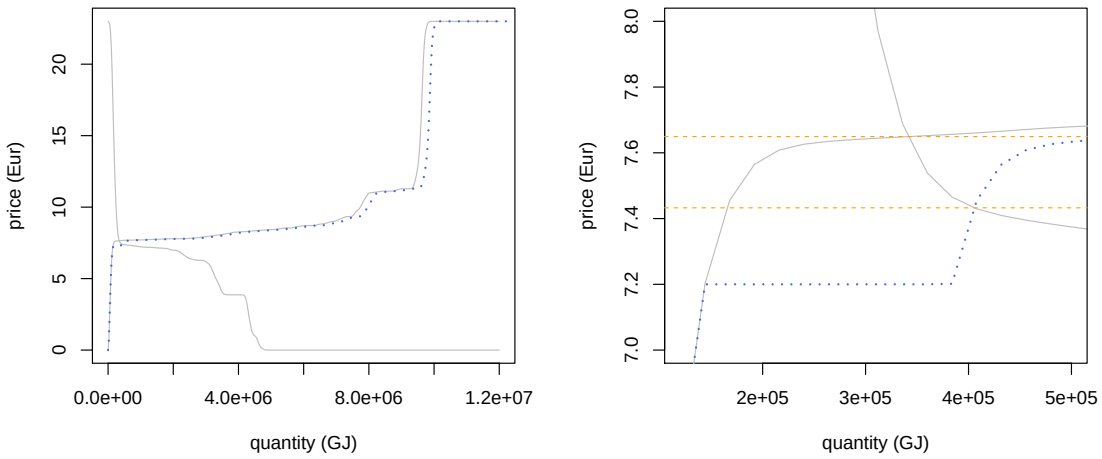


Figure 6: Scalar time series of prices (continuous line) and of the estimated prices with the classical ARIMA approach (dotted line) and as intersection of the curves predicted with a FAR(2) model (dashed lines).

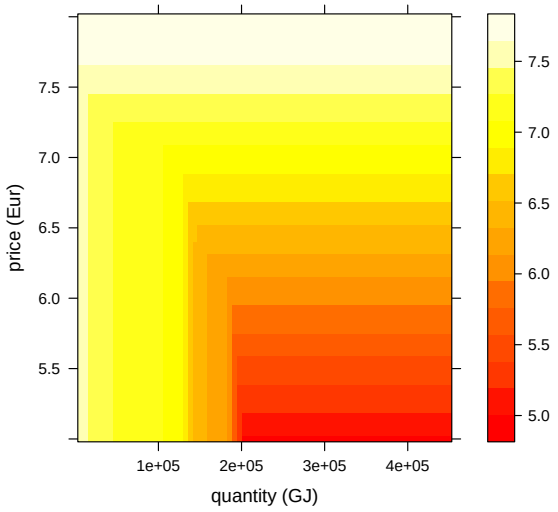
example assume to submit an offer of 240,000 GJ at 7.20 Euro. The modified curve is represented by a dotted line in panel (a) of Figure 7, with panel (b) showing a zoom in a neighborhood of the intersection. In this case the price is lowered from 7.65 to 7.43 Euro leading the trader to save $Q \times 0.22$ Euro. To better understand which is the most convenient action, panel (c) of Figure 7 reports the obtained price in function of price and quantity of the extra non-standard bid. To move the intersection point, the lower the offered price, the higher need to be the offered quantity. Evidently, prices above the estimated one, affect the shape of the curve after the intersection, with no consequences from a practical viewpoint.

4. DISCUSSION

Motivated by the analysis of functional time series of demand and supply curves in the Italian natural gas market, we proposed a model for functional time series, which preserves particular curves features such as monotonicity and bounds on the codomain. A bijective map associating each possible bounded and monotonic function to an unconstrained one is introduced. To make the latter an isometry between the space of monotone increasing and bounded function and L^2, a suitable geometry is introduced. In detail, we provide the constrained functions with a suitable pre-Hilbert structure.



(a) (b)



(c)

Figure 7: What-if simulations: curves prediction (continuous lines) and a supply perturbation (dotted line) (a) and zoom on a neighborhood of the intersections (b) with horizontal dashed lines representing the prices obtained as default and after the non-standard bid perturbation; price heatmap (brighter colors for higher resulted price) obtained as a function of a price and quantity of an extra non-standard bid (c).

The transformed curves are then modeled by means of functional auto regressive model. The autoregressive lagged operators and the non-centrality function of the model are obtained by minimizing the squared L^2 distance between functional data and functional predictions with a penalty term based on the Hilbert-Schmidt squared norm of autoregressive lagged operators. We have proved that the solution always exist, unique and that it is linear on the data with respect to the introduced geometry thus guaranteeing that the plug-in predictions of future functional data satisfy all required constraints. We also provide an explicit expression for estimates and predictions. The model can be easily generalized to include scalar covariates, or other functional predictors available at prediction time.

The methods has been successfully applied to data on the Italian natural gas balancing platform, revealing that tomorrow curves are strongly influenced by those of today and partially by those of yesterday. The prediction of tomorrow's curve is of dramatic interest for gas traders as it allows for what-if simulations that can help the decision making if ones wants act in this market with a speculative behavior.

ACKNOWLEDGEMENT

This work has been motivated and partially funded by MOXOFF srl.

APPENDIX: PROOFS

Proof of Theorem 1. Let us first show that the minimization withe respect to α is trivial. Indeed, for fixed values of Ψ_j for $j = 1, \dots, p$ the minimization of the objective function is obtained by minimizing the first term in equation 10 with respect to α, thus trivially obtaining $\hat{\alpha} = \frac{1}{T-p} \sum_{t=p+1}^{T} \left(f_t - \sum_{j=1}^{p} \Psi_j f_{t-j} \right) = \bar{f}_{[0]} - \sum_{j=1}^{p} \Psi_j f_{[j]}$. Hence the minimization of (10) can be carried out on the simplified objective function depending only on Ψ_j for $j = 1, \dots, p$ (obtained by (10) by replacing α with $\hat{\alpha}$):

$$\sum_{t=p+1}^{T} \left|\left| (f_t - \bar{f}_{[0]}) - \sum_{j=1}^{p} \Psi_j (f_{t-j} - \bar{f}_{[j]}) \right|\right|_{L^2}^2 + \lambda \sum_{j=1}^{p} ||\Psi_j||_{HS}^2 \, . \tag{14}$$

The proof of the existence and uniqueness of the minimizers comes by noticing that being Ψ_j for $j = 1, \dots, p$ Hilbert-Schmidt operators, the second term in (14) can be computed as $\lambda \sum_{j=1}^{p} \sum_{k \in \mathbb{N}} ||\Psi_j \phi_k||_{L^2}^2$ with $\{\phi_k\}_{k \in \mathbb{N}}$ being an arbitrary orthonormal

basis of $L^2(a,b)$. This latter identity points out that the simplified objective function (14) is a positive definite quadratic form in with respect to $\{\Psi_j\}_{j=1,\dots,p}$ and thus it admits a unique minimum. It is indeed obtained by linear combination with positive coefficients (i.e., 1 and λ) of a semi-positive definite quadratic form (i.e., the first term) and a positive definite quadratic form (i.e., the second term).

The explicit expressions of the estimators can be obtained by noticing that, thanks to Fubini-Tonelli Theorem, the $||\Psi_j||^2_{HS} = \int_a^b \left(\int_a^b \psi_j^2(s,u)\mathrm{d}u \right) \mathrm{d}s$, and thus the minimization of (14) can be carried out separately for each value of $s \in [a,b]$, i.e. minimizing:

$$\sum_{t=p+1}^{T} \left\{ \big(f_t(s) - \bar{f}_{[0]}(s)\big) - \sum_{j=1}^{p} \int_a^b \psi_j(s,u)\big(f_{t-j}(u) - \bar{f}_{[j]}(u)\big)\mathrm{d}u \right\}^2 + \lambda \sum_{j=1}^{p} \int_a^b \psi_j^2(s,u)\mathrm{d}u, \tag{15}$$

with respect to $\{\psi_j(s,\cdot)\}_{j=1,\dots,p}$ for all $s \in [a,b]$. Focussing on the case $p=1$, minimization problem (15) can be seen as a continuous version of a ridge-regression-like minimization problem. We thus have

$$\hat{\psi}_1(s,\cdot) = \sum_{t=p+1}^{T} \Big(\mathbb{P}_\lambda^{-1}(f_{t-1} - \bar{f}_{[1]}) \Big)(\cdot) \Big(f_t(s) - \bar{f}_{[0]}(s) \Big),$$

with $\mathbb{P}_\lambda$ being the HS operator with kernel

$$\sum_{t=p+1}^{T} \left\{ \big(f_{t-1}(s) - \bar{f}_{[1]}(s)\big)\big(f_{t-1}(u) - \bar{f}_{[1]}(u)\big) + \lambda \right\}.$$

The explicit solution for $p \geq 2$ is directly obtained by chaining, for $t = p+1, \dots, T$, functions $f_{t-1}, \dots, f_{t-p}$ in a unique function $\mathbf{f}_t$ defined on the auxiliary domain $\big(a, b + p(b-a)\big)$ and replicating the proof as in $p=1$. □

Proof of Corollary 1. The plug-in prediction of f_{T+1} is defined as

$$\hat{f}_{T+1} = \hat{\alpha} + \sum_{j=1}^{p} \hat{\Psi}_j f_{T+1-j}.$$

We are now showing that $\hat{f}_{T+1}$ is a linear combination in $L^2(a,b)$ of $\{f_1, \dots, f_T\}$. Let

$u_j^* = \big((j-1)(b-a)+u\big)$, then by simple computations we have

$$\hat{\alpha}(s) = f_{[0]}(s) - \sum_{t=p+1}^{T}\left[\left\{\sum_{j=1}^{p}\int_a^b \left(\mathbb{P}_\lambda^{-1}(\mathbf{f}_t - \bar{\mathbf{f}})\right)(u_j^*)\bar{f}_{[j]}(u)\mathrm{d}u\right\}\left(f_t(s) - \bar{f}_{[0]}(s)\right)\right],$$

and that:

$$\sum_{j=1}^{p}\big(\hat{\Psi}_j f_{T+1-j}\big)(s) = \sum_{t=p+1}^{T}\left[\left\{\sum_{j=1}^{p}\int_a^b \left(\mathbb{P}_\lambda^{-1}(\mathbf{f}_t - \bar{\mathbf{f}})\right)(u_j^*) f_{T+1-j}(u)\mathrm{d}u\right\}\left(f_t(s) - \bar{f}_{[0]}(s)\right)\right],$$

leading to $\hat{f}_{T+1}(s)$ equal to

$$f_{[0]}(s) + \sum_{t=p+1}^{T}\left[\left\{\sum_{j=1}^{p}\int_a^b \left(\mathbb{P}_\lambda^{-1}(\mathbf{f}_t - \bar{\mathbf{f}})\right)(u_j^*)\Big(f_{T+1-j}(u) - \bar{f}_{[j]}(u)\Big)\mathrm{d}u\right\}\left(f_t(s) - \bar{f}_{[0]}(s)\right)\right].$$

Thanks to the isometry between $\mathcal{M}^2(a,b)$ and $L^2(a,b)$, $\hat{g}_{T+1} = \mathrm{logH}^{-1}(\hat{f}_{T+1})$ is a linear combination in $\mathcal{M}^2(a,b)$ of $\{g_1 = \mathrm{logH}^{-1}(f_1), \ldots, g_T = \mathrm{logH}^{-1}(f_T)\}$ which belongs to $\mathcal{M}^2(a,b)$ being $\mathcal{M}^2(a,b)$ a space vector with respect to addition (3) and scalar multiplication (4). □

REFERENCES

Aitchison, J. (1982), "The statistical analysis of compositional data," *Journal of the Royal Statistical Society. Series B (Methodological)*, 139–177.

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

Bloch, D. and Silverman, B. (1997), "Monotone discriminant functions and their applications in rheumatology," *Journal of American Statistical Association*, 92, 144–153.

Boogaart, K. G., Egozcue, J. J., and Pawlowsky-Glahn, V. (2014), "Bayes Hilbert Spaces," *Australian & New Zealand Journal of Statistics*, 56, 171–194.

Box, G., Jenkins, G., and Reinsel, G. (2013), *Time Series Analysis: Forecasting and Control*, Wiley Series in Probability and Statistics, Wiley.

Chen, R. and Tsay, R. S. (1993), "Functional-coefficient autoregressive models," *Journal of the American Statistical Association*, 88, 298–308.

Egozcue, J., Díaz-Barrero, J., and Pawlowsky-Glahn, V. (2006), "Hilbert space of probability density functions based on Aitchison geometry," *Acta Mathematica Sinica*, 22, 1175–1182.

Elezović, S. (2009), "Functional modelling of volatility in the Swedish limit order book," *Computational Statistics & Data Analysis*, 53, 2107–2118.

Elezović, S. (2009), "Functional modelling of volatility in the Swedish limit order book," *Computational Statistics & Data Analysis*, 53, 2107–2118.

Epple, D. and McCallum, B. T. (2006), "Simultaneous Equation Econometric: The missing example," *Economic Inquiry*, 44, 374–384.

European Union (2003), "Directive 2003/54/EC," *Official Journal of the European Union*, 176, 37–55.

Fan, J. and Yao, Q. (2002), *Nonlinear time series*, vol. 2, Springer.

Fan, J. and Zhang, J.-T. (2000), "Two-step estimation of functional linear models with applications to longitudinal data," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62, 303–322.

Ferraty, F. and Vieu, P. (2006), *Nonparametric Functional Data Analysis: Theory and Practice*, Springer Series in Statistics, Springer.

Friedman, J. H. and Tibshirani, R. (1984), "The monotone smoothing of scatterplots." *Technometrics*, 26, 243–250.

GME (2013), `http://www.mercatoelettrico.org/En/Mercati/Gas/PGas.aspx`, retrieved on 13/02/2013.

Hall, P. and Huang, L. S. (2001), "Nonparametric kernel regression subject to monotonicity constraints," *Annals of Statistics*, 624–647.

Hastie, T., Tibshirani, R., and Friedman, J. (2009), *The elements of statistical learning*, vol. 2, Springer.

Henderson, D. J., List, J. A., Millimet, D. L., Parmeter, C. F., and Price, M. K. (2008), "Imposing Monotonicity Nonparametrically in First-Price Auctions," MPRA Paper 8769, University Library of Munich, Germany.

Hormann, S. and Kokoszka, P. (2012), *Time Series*, Elsevier, Oxford, UK, vol. 30 of *Handbook of Statistics*, chap. Functional time series, pp. 155–186.

Kargin, V. and Onatski, A. (2008), "Curve forecasting by functional autoregression," *Journal of Multivariate Analysis*, 99, 2508–2526.

Kelly, D. and Rice, M. (1994), "Preferences for verb interpretation in children with specific language impairment," *Journal of Speech and Hearing Research*, 37, 182–192.

Mammen, E. (1991), "Estimating a smooth monotone regression function," *The Annals of Statistics*, 724–740.

Mammen, E., Marron, J., Turlach, B., Wand, M., et al. (2001), "A general projection framework for constrained smoothing," *Statistical Science*, 16, 232–248.

Mammen, E. and Thomas-Agnan, C. (1999), "Smoothing splines and shape restrictions," *Scandinavian Journal of Statistics*, 26, 239–252.

Menafoglio, A., Guadagnini, A., and Secchi, P. (2013), "A kriging approach based on Aitchison geometry for the characterization of particle-size curves in heterogeneous aquifers," *Stochastic Environmental Research and Risk Assessment*, 1–17.

Mukerjee, H. (1988), "Monotone nonparametric regression," *The Annals of Statistics*, 741–750.

Passow, E. and Roulier, J. A. (1977), "Monotone and convex spline interpolation," *SIAM J. Numer. Anal.*, 14, 904–909.

Ramsay, J. and Silverman, B. (2005), *Functional Data Analysis*, Springer Series in Statistics, Springer.

Ramsay, J. O. (1988), "Monotone regression splines in action," *Statistical science*, 425–441.

Ramsay, J. O. and Silverman, B. W. (2002), *Applied functional data analysis*, Springer Series in Statistics, Springer-Verlag, New York, methods and case studies.

Shang, H. L. (2013), "Functional time series approach for forecasting very short-term electricity demand," *Journal of Applied Statistics*, 40, 152–168.

Winsberg, S. and Ramsay, J. O. (1980), "Monotonic transformations to additivity using splines," *Biometrika*, 67, 669–674.

— (1981), "Analysis of pairwise preference data using integrated B-splines," *Psychometrika*, 46, 171–186.

MOX Technical Reports, last issues

Dipartimento di Matematica "F. Brioschi",
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

42/2014 Canale, A.; Vantini, S.
Constrained Functional Time Series: an Application to Demand and Supply Curves in the Italian Natural Gas Balancing Platform

41/2014 Esfandiar, B.; Porta, G.; Perotto, S.; Guadagnini, A.
Impact of space-time mesh adaptation on solute transport modeling in porous media

40/2014 Antonietti, P.F.; Mazzieri, I.; Quarteroni, A.
Improving seismic risk protection through mathematical modeling

39/2014 Ghiglietti, A.; Paganoni, A.M.
Statistical inference for functional data based on a generalization of Mahalanobis distance

38/2014 Shen, H.; Truong, Y.; Zanini, P.
Independent Component Analysis for Spatial Stochastic Processes on a Lattice

37/2014 Giuliani, N.; Mola, A.; Heltai, L.; Formaggia,L.
FEM SUPG stabilisation of mixed isoparametric BEMs: application to linearised free surface flows

36/2014 Abb,A.;Bonaventura,L.; Nini, M.;Restelli,M.;
Anisotropic dynamic models for Large Eddy Simulation of compressible flows with a high order DG method

35/2014 Tricerri, P.; Dede', L.; Deparis, S.; Quarteroni, A.; Robertson, A.M.; Sequeira, A.
Fluid-structure interaction simulations of cerebral arteries modeled by isotropic and anisotropic constitutive laws

34/2014 Antonietti, P.F.; Pacciarini, P.; Quarteroni, A.
A discontinuous Galerkin Reduced Basis Element method for elliptic problems

33/2014 Canuto, C.; Simoncini, V.; Verani, M.
Contraction and optimality properties of an adaptive Legendre-Galerkin method: the multi-dimensional case