Constrained Functional Time Series: Applications to the Italian Gas Market

Antonio Canale
Department of Economics and Statistics,
University of Turin and Collegio Carlo Alberto, Italy
and
Simone Vantini
MOX, Department of Mathematics,
Politecnico di Milano, Italy

Abstract

Motivated by market dynamic modelling in the Italian Natural Gas Balancing Platform, we propose a model to analyze time series of functions subject to an equality and inequality constraint at the two edges of the domain, respectively, such as daily demand and offer curves. 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 one. We introduce a functional-to-functional autoregressive model that is used to forecast future demand/offer functions. We estimate the model via minimization of a penalized mean squared error of prediction with a penalty term based on the Hilbert-Schmidt squared norm of autoregressive lagged operators. The approach is of general interest and is suited for generalization in any situation in which one has to deal with functions subject to the above constraints which evolve through time.

Preprint submitted to International journal of Forecasting *May 17, 2016*

Keywords: Autoregressive Model; Demand and Offer Model; Energy Forecasting; Functional Data Analysis; Functional Ridge Regression.

1. Introduction

Energy markets in general, and natural gas markets in particular, are emerging fields which pose a great variety of forecasting problems. Among them load forecasting (**?**), price forecasting (**?**), daily price curve profile forecasting (**?**), consumption forecasting (**?**), and many others. In this paper, 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 proposed model is innovative both from the methodology perspective and from the applied point of view.

The supply and demand curves model is indeed a well known microeconomic model of price determination but its application is typically descriptive and static rather than strategic and predictive. This clearly does not help gas traders to forecast future prices or help them in the decision making and bidding. At the same time, usual forecasting methods, such as classical time series analysis, while producing useful predictions of scalar quantities of interest (e.g., price), 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 offer or demand cannot be directly incorporated in the inferential procedure or in what-if simulations. For all these reasons, the prediction of the entire supply and demand curves, and hence of their intersection, can be of dramatic interest.

We deal with this problem with a functional data analysis (FDA) ap-

proach. FDA is an extremely useful set of tools to deal with data that can be modeled as functions, as our demand and supply curves. Refer to **??**, **?**, or **?** for a quick introduction to FDA. However we innovate over the most common framework in FDA, in two aspects. First, we focus on functions that are constrained (i.e., monotonic and with an equality constraint on one edge of the domain and an inequality constraint on the other edge) and second, we embed such constraints for curves that are temporally dependent. The statistical literature focused separately on: (a) the problem of obtaining a constrained estimation of the underlying function given some point-wise evaluations of it and (b) the problem of modeling functional data with temporal dependence (i.e., functional time series). At our knowledge, the present work is the first 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 henceforth 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 give a brief overview on the state of the art pertaining both to 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 FDA. Isotonic regression is the first approach presented in the literature and it has been for years the most common approach to this purpose (see for example **???????**). 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 sum 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 in the framework of kernel regression (**??**) where the kernels are locally modified to achieve monotonicity. Another approach is the projection method (**???**) where the unknown monotonic function is estimated in an unconstrained fashion and then projected onto the convex subspace of the monotonic functions.

The approach we are going to use here, instead, is in line with the so called transform/back-transform method. This method is a common practice in FDA literature and has been he firstly proposed by **??**. The idea is basically to transform the functions such to perform an unconstrained estimation and then back-transform the estimated function to the convex subspace of the monotonic functions. Some very recent works (**???**) focusing on modeling cumulative distribution functions of absolutely continuous random variable (inspired by the pioneering work by **?** on compositional data) formalized this approach by introducing a suitable Hilbert structure on the set of the probability density functions and an isometric bijective map to L^2 for transforming and back-transforming functional data and conveniently map the entire statistical analysis in a linear subspace of L^2 (i.e., the zero-mean L^2 functions). In the present paper, instead, we are going to introduce a suitable pre-Hilbert structure, i.e. a non-complete vector space provided

with an inner product, 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 obtain 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. A pioneering contribution to the topic is due to **?** which derived a functional Yule-Walker estimator for time dependent functional data. Functional autoregressive models (FAR) are the most used approach to model temporal dependence between functional data, both for their ease of interpretation and performance in applications (**?**). In FAR models, autoregressive parameters are replaced by Hilbert-Schmidt operators and thus model estimation is declined in the estimation of the autoregressive operators. Different methods have been presented in the literature. Recent surveys are reported in **?** and **?**. Autoregressive operators are directly linked with lagged autocovariance operators (e.g., **?**) and thus a first possible approach is to estimate the lagged autocovariance operators from the functional time series and then the autoregressive operators accordingly. Because of the infinite dimensionality of functional data and to obtain more stable estimates, sample autocovariance operators are typically replaced by reduced rank approximations. A spread approach relies on functional principal component decomposition (e.g., **????**)

and the use of a reduced number of principal components. Other alternative reduced rank approximations presented in the literature are based on wavelet expansion of the original data (**?**) and on predictive factors (**?**).

Another 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, **?** and **?** 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 neighbors 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 FDA 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. To help the intuition, the theory is presented in the framework of univariate (i.e., real-valued) functional time series. Its extension to multivariate functional time series (which is in fact used in the application) comes naturally and it

is thus not detailed in the paper.

The rest of the paper is structured as follows. In Section **??** we first introduce the space $\mathcal{M}^2(a,b)$ and an isometric bijective map to a subspace of $L^2(a,b)$, and describe in detail the $\mathcal{M}^2$-FAR model, with particular emphasis on model estimation and choice of the autoregressive order. Section **??** describes our motivating context, discusses the application of our methodology to the Italian Natural Gas Balancing Platform data, and shows the potential of this new approach in terms of new trading opportunities. Section **??** 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 pre-Hilbert space that leads to the definition of the Bayes space geometry introduced and developed in **?**, **?**, and **?**. With respect to the curves studied in those 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 suitable bijective map from $\mathcal{M}^2(a,b)$ to a subspace

of $L^2(a,b)$. This map is 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 (**??**) 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 looks at g as a cumulative distribution function of a scalar absolutely-continuous random variable and maps it into the natural logarithm of the corresponding hazard function. We will thus call it log-hazard transformation, its inverse transformation anti-log-hazard transformation and indicate them with logH and logH^{-1}, respectively. In particular we have that constant functions 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., $f(s) = 0 \leftrightarrow g(s) = 1 - e^{-(s-a)}$). This map is not related to our specific motivating problem but is rather introduced for mathematical tractability. This allows its use also in different application contexts.

In the rest of the section, we will build an 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*

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

- *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*

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

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 $\mathrm{logH}(\mathcal{M}^2(a,b))$ 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.

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*

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

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 (**??**) *are defined as:*

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

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

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 (**??**) and (**??**), it is not complete with respect to the metric $d_{\mathcal{M}^2}$ induced by the inner product $\langle g_1, g_2 \rangle_{\mathcal{M}^2}$ defined in (**??**). 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. Even though it is quite straightforward to make it Hilbert (this would require to close the space and - relying on the separability of $L^2(a, b)$ - define on the closure the operations of addition and scalar multiplication consistently with (**??**) and (**??**)), this extension (which would make logH an isometric bijective map with the entire space $L^2(a, b)$) 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 Lemma **??**, the predictions provided by the estimated model are linear combinations in the sense defined in (**??**) and (**??**) of functions in $\mathcal{M}^2(a, b)$ which are guaranteed to be in $\mathcal{M}^2(a, b)$. Figure **??** summarizes all relations between $\mathcal{M}^2(a, b)$ and $L^2(a, b)$.

2.2. Functional Autoregressive Model

We here describe the model we use for dealing with temporal dependence. The model can be coherently formulated either on the original data in the

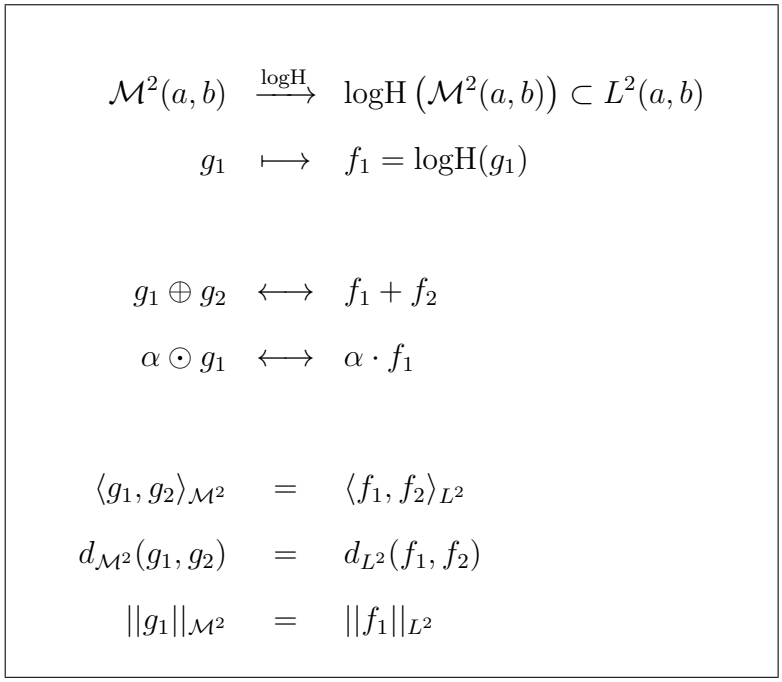


Figure 1: Relations between $\mathcal{M}^2(a,b)$ and image in $L^2(a,b)$ of the map logH.

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 practice for computation. 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 functional autoregressive model of order p (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 bivariate function $\psi_j(s,u) \in L^2\{(a,b)\times(a,b)\}$ 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 and α is a non-centrality function.

The choice of the order p can be done following different approaches. For example, **?** propose a multistage testing procedure. The latter exploits the FAR representation in terms of functional principal components and derive the approximated distribution of a suitable test statistics. Another approach extends classical time series identification tools to the functional framework.

In scalar time series analysis, it is a common practice to look at the autocorrelation and partial autocorrelation 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 (**?**). To perform a similar investigation in the FAR framework, we here introduce a measure of functional autocorrelation and of functional partial autocorrelation playing the roles of their scalar counterparts. Define the functional autocorrelation function of lag k of a functional time series $\{f_t\}_{t=1}^T$ by

$$R_k(s,u) = \frac{E[(f_t(s) - E[f_t(s)])(f_{t+k}(u) - E[f_{t+k}(u)])]}{\sqrt{E[(f_t(s) - E[f_t(s)])^2]E[(f_{t+k}(u) - E[f_{t+k}(u)])^2]}}, \tag{10}$$

which expresses the correlation between $f_t(s)$ and $f_{t+k}(u)$. Even though one could integrate (**??**) 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 (**??**) along k. In analogy with the definition of scalar partial autocorrelation, define also the functional partial autocorrelation function of order $k > 2$ as

$$\Gamma_k(s,u) = \frac{E[r_t^{k*}(s)\, r_{t+k}^k(u)]}{\sqrt{E[(r_t^{k*}(s))^2]E[(r_{t+k}^k(u))^2]}}, \tag{11}$$

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 autoregressive model (**??**) of order p and $\{r_t^{p*}\}$ are the function residuals at time t of the functional autoregressive model (**??**) of order k, fitted to the reversed series.

2.3. Model estimation and prediction

As mentioned in the Introduction 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:

$$\min_{\alpha \in L^2 \,\cap\, \{\Psi_j\}_{j=1,\ldots,p} \,\subseteq\, \mathrm{HS}} \left(\sum_{t=p+1}^{T} \left|\left| f_t - \left(\alpha + \sum_{j=1}^{p} \Psi_j f_{t-j} \right) \right|\right|_{L^2}^2 + \lambda \sum_{j=1}^{p} ||\Psi_j||_{\mathrm{HS}}^2 \right). \tag{12}$$

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 gets bigger the estimated model is shrunk 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 $j = 1, \ldots, p$ 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:

$$\hat{g}_{T+1}(s) = 1 - \exp\left(- \int_a^s \prod_{t=p+1}^{T} \left(\frac{g_t'(s)}{1 - g_t(s)} \right)^{\frac{1}{T-p}} \mathrm{d}s \right). \tag{13}$$

At the other extreme, when $\lambda \to 0^+$, the estimated model converges toward an interpolating model, in particular (consistently with Theorem **??** below) the one with the minimal Hilbert-Schmidt norms of the lagged autoregressive operators Ψ_j. As common practice in FDA, 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 or even heuristically by looking - in this case - at the smoothness/roughness of the kernels of the estimated lagged autoregressive operators Ψ_j and/or of the predicted functions (i.e. the so called Goldilocks' method). If one desires little bias in the estimates, small values of λ might be favoured. If instead more robust estimates are desired, larger values of λ might be favoured. 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. Its proof is reported in the Appendix.

Theorem 1 (Existence, uniqueness, and explicit expression of the estimators)**.** *For any $\lambda > 0$, $j = 1, \ldots, p$, and $s, u \in [a, b]$, the solution of minimization problem (**??**) always exists unique 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* $\tilde{u} \in \big(a, b + p(b-a)\big)$.

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 Lemma states indeed that the joint use of the geometry introduced in Section **??** and the model estimation procedure described in this section guarantees the plug-in predictions to satisfy all constraints characterizing functions belonging to $\mathcal{M}^2(a,b)$. Its proof is reported in the Appendix.

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

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

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

Remark 1. *Note the proposed estimation method (i.e., the minimization of (**??**)) 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* $x_{1i}, \ldots, x_{pi}$*:*

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

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 (**?**).*

3. Application to Italian Natural Gas Balancing Platform

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 (**?**).

Such measures, while favouring 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 exporting countries such as Algeria or Russia. The gas is then consumed by civil, industrial, and thermo-electric stations spread across the country. The role of the pipeline manager, Snam s.p.a., is to compensate injections and consumptions via storage or other measures. In fact, the risk of possible imbalance is assigned to each shipper which has to daily predict and communicate to Snam its injection and withdrawal forecasts on which a penalty is payed for any positive or negative imbalance.

With the final aim of having a self balancing system, the Italian Natural Gas Balancing Platform (PB-GAS), has been introduced in December 2011. The PB-GAS provides a venue for Snam, shippers, and traders to perform physical balancing of the pipeline network and procure gas on a day-ahead

basis. It is a system in which gas operators virtually sell and buy natural gas, independently of its physical location. The PB-GAS is managed by the energy regulatory *Gestore Mercati Energetici* (GME), with Snam acting as central counterpart for all daily offers. The regulation fixes the upper and lower bounds of the price at 0 and 23 Euro/GJ, respectively. Every day Snam submits a demand bid or supply offer for a volume of gas corresponding to the overall imbalance of the system and with a price equal to, respectively, 23 or 0 Euro/GJ 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, according to price, from the highest to the lowest and from the lowest to the highest respectively, so that demand and offer curves are obtained as the cumulative sum of the quantities in GJ. We henceforth refer to offer curves, rather than supply curves since the former are build from the actual prices and quantities that the traders offer in the auction while the latter represent the marginal cost curve of each trader. The selection of bids/offers accepted on the PB-GAS is indeed 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. Bidding a demand (offer) at the maximum (minimum) permitted price, allows Snam to always be on the left of the intersection and thus to deterministically exchange the volume of the imbalance.

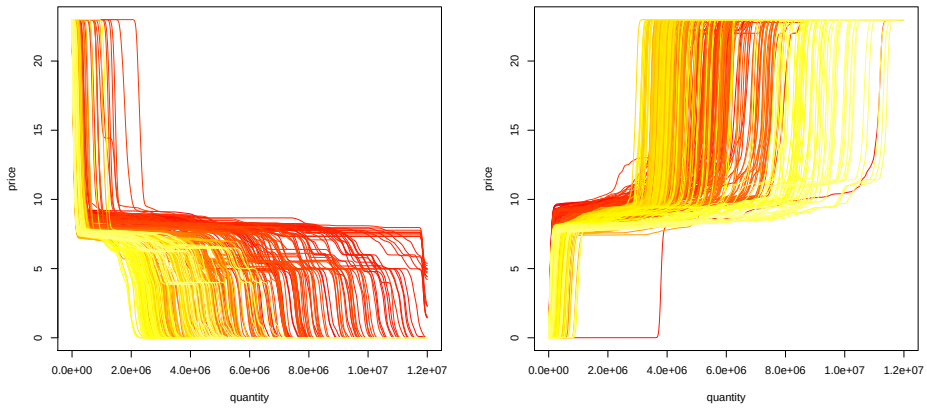
While balancing the network, each shipper can also take advantage of the market in a speculative perspective, buying natural gas at lower price or selling exceeding gas at higher price, with respect to their benchmark supplying indexes. It is clear that forecast tools are dramatically important for the

decision-making of each shipper. For example, the mere price prediction is a first procedure to implement. However, the pointwise or interval price forecast alone, is of limited utility when the effects of a single bids can strongly modify tomorrow's curves shape, and thus the resulting equilibrium price. It would be much more useful to have a prediction of the entire demand and offer curves. With such tools at hand, traders can directly see the effect of their bids on the shape of tomorrow's curve and on price itself.

3.2. Data description

The data used in our analysis refer to the first thirteen months of the PB-GAS, namely from December 1st, 2011 to December 31st, 2012. The data are available at the website of **?**. The original data are reported in XML format, where single entry represents an awarded bid with its own code, date, trader name, type (sell or buy), awarded price, and awarded quantity. For each day, we build the offer (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 **??**, the raw data has been converted to functional data in $\mathcal{M}^2(a, b)$ with $a = 0$ and $b = 1.2 \times 10^7$ GJ, which can be considered as a conservative upper bound of the range of investigation. In detail, the smoothed versions of the offer (and demand) curves are obtained by means of local polynomial regression as implemented in the R function `locpoly` of `KernSmooth`. Additional details are reported in the Appendix. The legislative upper bound at 23 Euro/GJ allows us to scale the curves by 23^{-1} without loss of information so to make them vertically constrained between zero and one. The scaled offer and reverse-



(a) (b)

Figure 2: Smoothed functional time series of demand (a) and offer (b) curves. Color denotes time, with the oldest curves in dark and the most recent ones in bright.

demand curves naturally satisfy the requirements of being exactly zero in a. To preserve this constraint in the smoothed curves, before performing the local polynomial smoothing, we added an artificial point on the negative x-axis for each curve in order to have the local polynomial regression output equal to zero for the offer and to one for the demand when the quantity is zero. The obtained smoothed functional time series are plotted in Figure **??**.

3.3. Results

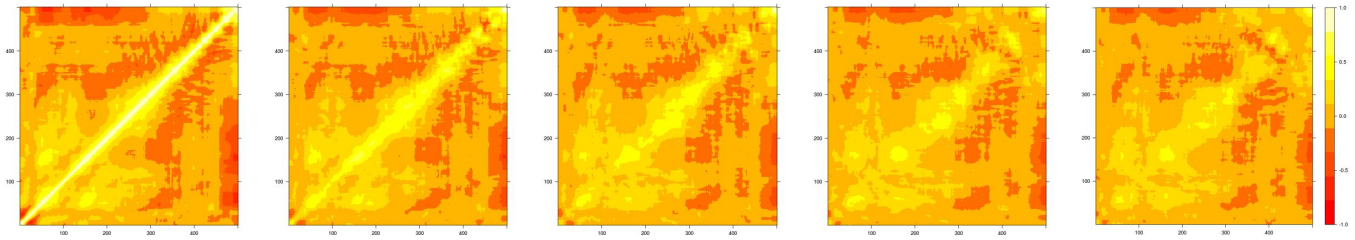
Let $g_t^D(s)$ and $g_t^S(s)$ be the scaled demand and offer curves of day t, respectively, expressing the price in function of quantity. We apply the methodology presented in Section **??** to model $\{1 - g_t^D\}_{t=1}^T$ and $\{g_t^S\}_{t=1}^T$. In particular, we estimated the $\mathcal{M}^2$-FAR(p) models described in the previous section, for different choices of the autoregressive order p, separately for the two time series. Moreover, since it is reasonable to assume that both the past demand and

supply curves influence future demand and supply curves jointly, we estimate also bivariate $\mathcal{M}^2$-FAR(p) models. As in classical time series analysis, the multivariate extension naturally reads:

$$f_t^D = \alpha^D + \sum_{j=1}^{p} \Psi_j^{DD} f_{t-j}^D + \sum_{j=1}^{p} \Psi_j^{DS} f_{t-j}^S + \epsilon_t^D,$$
$$f_t^S = \alpha^S + \sum_{j=1}^{p} \Psi_j^{SS} f_{t-j}^S + \sum_{j=1}^{p} \Psi_j^{SD} f_{t-j}^D + \epsilon_t^S,$$

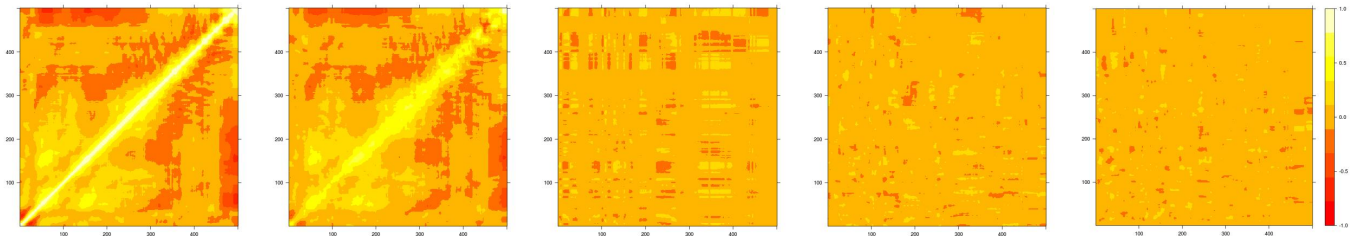
where $\{f_t^D\}_{t=1}^T$ and $\{f_t^S\}_{t=1}^T$ are the the logH transformations of $\{1 - g_t^D\}_{t=1}^T$ and $\{g_t^S\}_{t=1}^T$, respectively; Ψ_j^{DD}, Ψ_j^{SS}, Ψ_j^{DS}, and Ψ_j^{SD} are the lagged operators. The latter bivariate FAR (BFAR) model has, for fixed p, the computational complexity of a univariate FAR model of order $2p$. The penalization parameter has been fixed $\lambda = 10^{-7}$ for all p and both for FAR and BFAR models. The particular choice of λ has only a minor impact on the results of the analysis. Indeed the performance obtained with values of λ ranging from 10^{-5} to 10^{-9} give qualitatively comparable results.

As for the choice of the $\mathcal{M}^2$-FAR order p the method of **?** cannot be applied in a straightforward way in our settings, which also consider bivariate FAR models. Thus, we follow two different approaches. The first one is to look at the functional autocorrelation and functional partial autocorrelation plots obtained from equations (**??**)–(**??**) and reported in Figures **??** and **??**. The plots show the sample autocorrelation function and partial autocorrelation function, respectively, for $k = 0, 1, \ldots, 4$ and for the demand series only. Qualitatively similar results were obtained for the offer series.



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

Figure 3: 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 4: Sample functional partial autocorrelation function $k = 0, 1, \ldots, 4$ for the demand series.

Figure ?? shows that the autocorrelation is persistent also for increasing lags, a typical feature registered in scalar autoregressive 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 offer 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 autocorrelation remains high also for increasing lags mostly around the diagonal of the plots in Figure ??. 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 ??, 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 a lag-1 model may be sufficiently appropriate to fit our data.

As a second approach we consider data-driven measures. For example we can asses the model goodness of fit, for increasing values of the autoregressive order p. This allow us also to compare the univariate and bivariate models. The comparison is carried out using different functional measures of discrepancy between the predicted curves and the original ones. The most natural, given our modeling approach, is the $\mathcal{M}^2$ root mean squared error defined as

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

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

predicted transformed curves and the original transformed ones. We also computed other standard measures of goodness of fit, i.e.: the L^2 root mean squared error, the L^1 mean absolute error and the L^∞ mean absolute error on the original scale, defined as

$$L^2\text{-}RMSE = \sqrt{\frac{1}{T-p}\sum_{p+1}^{T} ||g_t - \hat{g}_t||^2_{L^2}},$$

$$L^1\text{-}MAE = \frac{1}{T-p}\sum_{p+1}^{T} ||g_t - \hat{g}_t||_{L^1},$$

$$L^\infty\text{-}MAE = \frac{1}{T-p}\sum_{p+1}^{T} ||g_t - \hat{g}_t||_{L^\infty}.$$

The results are reported in Tables **??**–**??**. In both the $\mathcal{M}^2$-FAR and $\mathcal{M}^2$-BFAR specifications, the autoregressive order p has similar impact on the estimation of both the demand and offer curves (i.e., for fixed p, the errors for the demand and offer curves are roughly of the same order of magnitude). In detail, due to the obvious dependence between supply and demand curves, for fixed p the bivariate model has a dramatically better performance than the two univariate ones. In particular, the $\mathcal{M}^2$-BFAR(1) has a performances comparable to that of the $\mathcal{M}^2$-FAR(2). As a reference, we also fit a functional version of the simple exponential smoothing (FES) which is anyhow outperformed by all the $\mathcal{M}^2$-FAR and $\mathcal{M}^2$-BFAR specifications. Note that the models in lines 3–7 of both tables have the same accuracy to three digits for the indexes calculated on the original scale of the functions. Details about the implementation and tuning of the FES model are provided in the Appendix.

As additional measure of goodness of fit, we consider also an application

Table 1: Functional errors between the predicted demand curves and the original ones.

	$\mathcal{M}^2$-$RMSE$	L^2-$RMSE$	L^1-MAE	L^∞-MAE
FES	7.00×10^3	4.86×10^3	5.78×10^6	5.09×10^0
$\mathcal{M}^2$-FAR(1)	6.13×10^2	4.08×10^3	4.37×10^6	2.60×10^0
$\mathcal{M}^2$-FAR(2)	1.06×10^{-3}	4.85×10^2	4.00×10^5	1.51×10^0
$\mathcal{M}^2$-FAR(3)	1.09×10^{-4}	4.85×10^2	4.00×10^5	1.51×10^0
$\mathcal{M}^2$-BFAR(1)	3.57×10^{-4}	4.85×10^2	4.00×10^5	1.51×10^0
$\mathcal{M}^2$-BFAR(2)	1.15×10^{-5}	4.85×10^2	4.00×10^5	1.51×10^0
$\mathcal{M}^2$-BFAR(3)	2.36×10^{-6}	4.85×10^2	4.00×10^5	1.51×10^0

Table 2: Functional errors between the predicted offer curves and the original ones.

	$\mathcal{M}^2$-$RMSE$	L^2-$RMSE$	L^1-MAE	L^∞-MAE
FES	6.25×10^3	7.63×10^3	8.90×10^6	8.29×10^0
$\mathcal{M}^2$-FAR(1)	4.74×10^2	3.85×10^3	5.04×10^6	1.91×10^0
$\mathcal{M}^2$-FAR(2)	6.26×10^{-4}	3.53×10^2	2.83×10^5	8.78×10^{-1}
$\mathcal{M}^2$-FAR(3)	6.07×10^{-5}	3.53×10^2	2.83×10^5	8.78×10^{-1}
$\mathcal{M}^2$-BFAR(1)	3.66×10^{-4}	3.53×10^2	2.83×10^5	8.78×10^{-1}
$\mathcal{M}^2$-BFAR(2)	1.32×10^{-5}	3.53×10^2	2.83×10^5	8.78×10^{-1}
$\mathcal{M}^2$-BFAR(3)	2.63×10^{-6}	3.53×10^2	2.83×10^5	8.78×10^{-1}

driven approach. As motivated in Section **??**, 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 (RMSE), reported in Table **??**.

We are now focusing on price prediction, since, as said, it is likely the most important feature to forecast for a trader. As benchmark, along with the FES model we fitted also a scalar exponential smoother (SES) to time series of prices using the `ses` function of the R package `forecast` (**?**) with default specification. The RMSE on price provided by the univariate $\mathcal{M}^2$-FAR(1) models (i.e., 2.03 Euro) is much larger than the ones provided by the FES and SES models (i.e., 0.31 and 0.20 Euro, respectively). The $\mathcal{M}^2$-FAR(2) models outperform instead the two benchmark methods providing an RSME on price of 0.18 Euro. On the contrary, if focussing on quantity, the univariate approach based on $\mathcal{M}^2$-FAR(p) models turns out not to be able to achieve an RMSE on quantity comparable or better than the ones provided by the two benchmark methods. In this sense, the bivariate approach is key to outperform benchmark methods in both price and quantity prediction. Indeed, the $\mathcal{M}^2$-BFAR(1) model provides a RMSE on both price and quantities (i.e., 0.18 Euro and $2.16 \cdot 10^5$ GJ, respectively) lower than the ones provided by the FES and SES models. Given the performances in predicting both the offer and demand curves and the price and quantity and also considering the suggestion of Figures **??**, we choose the $\mathcal{M}^2$-BFAR(1)

Table 3: RMSE for price (in Euro) and quantity (in GJ) obtained as crossing point between estimated curves.

	quantity (GJ)	price (Euro)
FES	347,473	0.31
SES	239,356	0.20
$\mathcal{M}^2$-FAR(1)	336,430	2.03
$\mathcal{M}^2$-FAR(2)	336,053	0.18
$\mathcal{M}^2$-FAR(3)	336,053	0.18
$\mathcal{M}^2$-BFAR(1)	216,228	0.18
$\mathcal{M}^2$-BFAR(2)	216,228	0.18
$\mathcal{M}^2$-BFAR(3)	216,228	0.18

model as final model for the analysis.

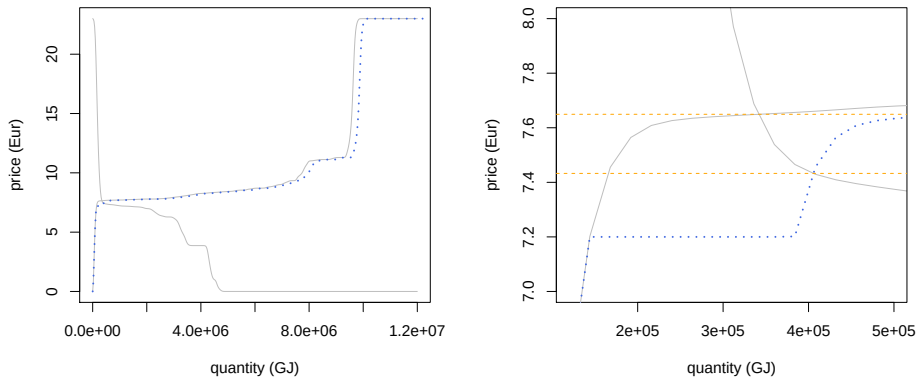
3.4. Trading Example

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 **??**. 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 **?** website). To lower the price, the trader can submit an extra supply offer for a small quantity, that eventually he is going to buy above Q. For 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 **??**, with panel (b) showing a zoom in a neighbourhood of the intersection. In this case the price is lowered from 7.65 to 7.43 Euro

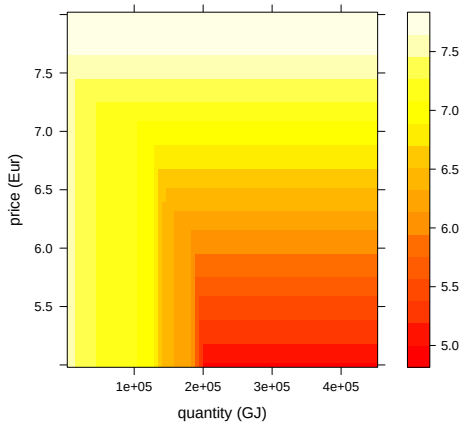
leading the trader to save $Q \times 0.22$ Euro. To better understand which is the most convenient action, panel (c) of Figure **??** reports the obtained price in function of price and quantity of the extra bid. To move the intersection point, the lower the offered price, the higher need to be the offered 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 offer curves in the Italian natural gas market, we proposed a model for functional time series, which preserves particular curves features such as monotonicity and equality and inequality constraints at the edges of the domain. A bijective map associating each possible constrained function to an unconstrained one is introduced. To make the latter an isometry between the space of the constrained functions and L^2, a suitable geometry is introduced. In detail, we provide the constrained functions with a suitable pre-Hilbert structure. The transformed curves are then modeled by means of functional autoregressive 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 exists, 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 explicit expressions for estimates and predictions. The model can be



(a) (b)



(c)

Figure 5: What-if simulations: curves prediction (continuous lines) and a offer 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 bid perturbation; price heatmap (brighter colors for higher resulted price) obtained as a function of a price and quantity of an extra bid (c).

easily generalized to model multivariate functional data or to include scalar covariates, or other functional predictors available at prediction time. More tricky in terms of estimation but for sure worthy of future investigation would be the introduction in the model of a trend and/or seasonality term which might lead to a further improvement in the prediction performance.

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. The prediction of tomorrow's curves is of dramatic interest for gas traders as it allows for what-if simulations that can help the decision making if one wants to act in this market with a speculative behavior.

Acknowledgement

We thank the Referees for for their constructive comments, which led to a substantial improvement of the manuscript. This work has been motivated and partially funded by Moxoff.

Appendix A.

Appendix A.1. Proofs

Proof of Theorem **??**. Let us first show that the minimization with respect to α is trivial. Indeed, for fixed values of Ψ_j for $j = 1, \ldots, p$ the minimization of the objective function is obtained by minimizing the first term in equation (**??**) 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 (**??**) can be carried out on the simplified objective function depending only on Ψ_j for $j = 1, \dots, p$ (obtained by (**??**) by replacing α with $\hat{\alpha}$):

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

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 (**??**) 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 (**??**) 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 $||\Psi_j||_{HS}^2 = \int_a^b \left(\int_a^b \psi_j^2(s, u) \mathrm{d}u \right) \mathrm{d}s$, and thus the minimization of (**??**) can be carried out separately for each value of $s \in [a, b]$, i.e. minimizing:

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

with respect to $\{\psi_j(s, \cdot)\}_{j=1,\dots,p}$ for all $s \in [a, b]$. Focusing on the case $p = 1$, the minimization problem (**??**) can be seen as a continuous version of a

ridge-regression-like minimization. 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 $t = p+1, \ldots, T$, functions $f_{t-1}, \ldots, f_{t-p}$ 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 Lemma **??**. 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 $\{f_1, \ldots, f_T\}$. Let $u_j^* = \big((j-1)(b-a) + u\big)$, then by simple computations we have

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

and that $\sum_{j=1}^{p}\big(\hat{\Psi}_j f_{T+1-j}\big)(s)$ is equal to

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

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

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

Thanks to the isometry between the space $\mathcal{M}^2(a,b)$ and $\log\mathrm{H}\left(\mathcal{M}^2(a,b)\right)$, $\hat{g}_{T+1} = \log\mathrm{H}^{-1}(\hat{f}_{T+1})$ is a linear combination in $\mathcal{M}^2(a,b)$ of

$$\{g_t = \log\mathrm{H}^{-1}(f_t), t = 1, \ldots, T\}$$

which belongs to $\mathcal{M}^2(a,b)$ being $\mathcal{M}^2(a,b)$ a space vector with respect to addition (**??**) and scalar multiplication (**??**). □

Appendix A.2. Details on the smoothing of the raw data

The smoothed versions of the offer (and demand) curves are obtained by means of local polynomial regression as implemented in the R function `locpoly` of `KernSmooth`, with degree 0 and Gaussian kernel. An important parameter, when one has to deal with smoothing is the bandwidth parameter. We fixed the bandwidth to 6,000 for all curves. This choice has been made subjectively by eye and turns out to be an acceptable compromise between smoothness and coherence to the raw data. Despite being subjective, the choice-by-eye is widely used and satisfactory in many situations (see **?**, Section 3.1). The derivative of the smoothed functions are obtained via numerical derivation.

Appendix A.3. Functional exponential smoother

Exponential smoothing is a useful method to produce one-step-ahead predictions in classical time series. A functional version of it can be written as

$$\hat{g}_{t+1}(s) = \alpha \sum_{j=0}^{\infty} (1-\alpha)^j g_{t-j}(s),$$

where $\alpha \in (0,1)$ and thus the bounds and the monotonicity constraint are preserved in the one-step-ahead prediction. In our application, we estimated α by least squares estimation by means of the `optim` function of R and the

`ses` function of package `forecast`. Two different values have been obtained for the two functional time series, namely 0.44 and 0.55 for the demand and offer series, respectively.

References

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society. Series B (Methodological) 44*(2), 139–177.

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

Aue, A., D. D. Norinho, and S. Hörmann (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association 110*(509), 378–392.

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

Boogaart, K. G., J. J. Egozcue, and V. Pawlowsky-Glahn (2014). Bayes hilbert spaces. *Australian & New Zealand Journal of Statistics 56*(2), 171–194.

Bosq, D. (1991). Modelization, nonparametric estimation and prediction for continuous time processes. In G. Roussas (Ed.), *Nonparametric functional estimation and related topics, NATO Science Series C*, pp. 509–529. Springer.

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

Brabec, M., O. Konr, E. Pelikn, and M. Mal (2008). A nonlinear mixed effects model for the prediction of natural gas consumption by individual customers. *International Journal of Forecasting 24*(4), 659 – 678. Energy Forecasting.

Chen, Y. and B. Li (2015). An adaptive functional autoregressive forecast model to predict electricity price curves. *Journal of Business & Economic Statistics* (doi:10.1080/07350015.2015.1092976), 1–56.

Egozcue, J., J. Díaz-Barrero, and V. Pawlowsky-Glahn (2006). Hilbert space of probability density functions based on aitchison geometry. *Acta Mathematica Sinica 22*(4), 1175–1182.

Elezović, S. (2009). Functional modelling of volatility in the swedish limit order book. *Computational Statistics & Data Analysis 53*(6), 2107–2118.

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

Fan, J. and J.-T. Zhang (2000). Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62*(2), 303–322.

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

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

Gestore Mercati Energetici (2013). Italian natural gas trading platform operative details. Available at:`www.mercatoelettrico.org/En/Mercati/Gas/PGas.aspx`, retrieved on 13/02/2013.

Hall, P. and L. S. Huang (2001). Nonparametric kernel regression subject to monotonicity constraints. *Annals of Statistics 29*(3), 624–647.

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

Henderson, D. J., J. A. List, D. L. Millimet, C. F. Parmeter, and M. K. Price (2008). Imposing monotonicity nonparametrically in first-price auctions. MPRA Paper 8769, University Library of Munich, Germany.

Hong, T. (2014). Energy forecasting: Past, present, and future. *Foresight: The International Journal of Applied Forecasting* (32), 43–48.

Hormann, S. and P. Kokoszka (2012). Functional time series. In C. Rao and T. Subba Rao (Eds.), *Handbook of Statistics*, Volume 30, pp. 155–186. Elsevier, Oxford, UK.

Horváth, L. and P. Kokoszka (2012). *Inference for functional data with applications*, Volume 200. Springer Science & Business Media.

Hyndman, R. J. and Y. Khandakar (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software 26*(3), 1–22.

Hyndman, R. J. and H. L. Shang (2009). Forecasting functional time series. *Journal of the Korean Statistical Society 38*(3), 199–211.

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

Kargin, V. and A. Onatski (2008). Curve forecasting by functional autoregression. *Journal of Multivariate Analysis 99*(10), 2508–2526.

Kokoszka, P. and M. Reimherr (2013). Determining the order of the functional autoregressive model. *Journal of Time Series Analysis 34*(1), 116–129.

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

Mammen, E., J. Marron, B. Turlach, M. Wand, et al. (2001). A general projection framework for constrained smoothing. *Statistical Science 16*(3), 232–248.

Mammen, E. and C. Thomas-Agnan (1999). Smoothing splines and shape restrictions. *Scandinavian Journal of Statistics 26*(2), 239–252.

Menafoglio, A., A. Guadagnini, and P. Secchi (2014). A kriging approach based on aitchison geometry for the characterization of particle-size curves in heterogeneous aquifers. *Stochastic Environmental Research and Risk Assessment 28*(7), 1835–1851.

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

Passow, E. and J. A. Roulier (1977). Monotone and convex spline interpolation. *SIAM J. Numer. Anal. 14*(5), 904–909.

Ramsay, J. and B. Silverman (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 B. W. Silverman (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*(1), 152–168.

Sørensen, H., J. Goldsmith, and L. M. Sangalli (2013). An introduction with medical applications to functional data analysis. *Statistics in medicine 32*(30), 5222–5240.

Wand, M. P. and M. C. Jones (1994). *Kernel smoothing*, Volume 60. Crc Press.

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting 30*(4), 1030 – 1081.

Winsberg, S. and J. O. Ramsay (1980). Monotonic transformations to additivity using splines. *Biometrika 67*(3), 669–674.

Winsberg, S. and J. O. Ramsay (1981). Analysis of pairwise preference data using integrated B-splines. *Psychometrika 46*(2), 171–186.