International Journal of Forecasting 32 (2016) 1340–1351

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Constrained functional time series: Applications to the Italian gas market

Antonio Canale [a,*], Simone Vantini [b]

[a] Department of Economics and Statistics, University of Turin and Collegio Carlo Alberto, Italy
[b] MOX, Department of Mathematics, Politecnico di Milano, Italy

ARTICLE INFO

Keywords:
Autoregressive model
Demand and offer model
Energy forecasting
Functional data analysis
Functional ridge regression

ABSTRACT

Motivated by market dynamic modelling in the Italian Natural Gas Balancing Platform, we propose a model for analyzing time series of functions, subject to equality and inequality constraints at the two edges of the domain, respectively, such as daily demand and offer curves. Specifically, we provide the constrained functions with suitable pre-Hilbert structures, and introduce a useful isometric bijective map that associates 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, and estimate the model via the 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 could be generalized to any situation in which one has to deal with functions that are subject to the above constraints which evolve over time.

1. Introduction

Energy markets in general, and natural gas markets in particular, are emerging fields that pose a great variety of forecasting problems, including load forecasting (Hong, 2014), price forecasting (Weron, 2014), daily price curve profile forecasting (Chen & Li, 2015), consumption forecasting (Brabec, Konár, Pelikán, & Malı, 2008), and many others. Motivated by price prediction in the Italian natural gas balancing market, this paper proposes a model for forecasting the day-to-day evolution of supply and demand curves. The proposed model is innovative from both the methodological and applied perspectives.

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, which clearly does not help gas traders with either the forecasting of future prices or decision making and bidding. At the same time, while the usual forecasting methods, such as classical time series analysis, produce useful predictions of scalar quantities of interest (e.g., prices), they do not provide the insights into the market that are given by the supply and demand model. Furthermore, in markets with a moderate number of traders, the effect of a single offer or demand cannot be incorporated directly into either the inferential procedure or what-if simulations. For all of these reasons, the prediction of the entire supply and demand curves, and hence of their intersection, can be of strong interest.

We deal with this problem using a functional data analysis (FDA) approach. FDA is an extremely useful set of tools for dealing with data that can be modeled as functions, such as our demand and supply curves; for a quick introduction, refer to Ferraty and Vieu (2006), Ramsay and

* Corresponding author.
E-mail address: antonio.canale@unito.it (A. Canale).

http://dx.doi.org/10.1016/j.ijforecast.2016.05.002
0169-2070/

Silverman (2002, 2005), or Sørensen, Goldsmith, and Sangalli (2013). However, our approach differs from the most common FDA framework in two ways. 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 has 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). To the best of our knowledge, the present work is the first to tackle the temporal dependence jointly with constraints pertaining to monotonicity, boundedness, and values of the function at the boundary of the domain. We refer to this joint framework henceforth as *constrained functional time series*.

Before going into detail about the mathematical modeling and the estimation method that we propose herein, we provide a brief overview of the state of the art pertaining to both monotonic estimation and functional time series estimation.

The problem of having monotonic estimates of unknown functions that are observed only at a few sparse points in the domain, possibly with some measurement error, has been being tackled in the literature for many decades, even before the recent outbreak of FDA. Isotonic regression was the first approach presented in the literature, and has been the most common approach to this issue for years (see for example Mammen, 1991; Mammen & Thomas-Agnan, 1999; Mukerjee, 1988; Passow & Roulier, 1977; Ramsay, 1988; Winsberg & Ramsay, 1980, 1981). The basic idea is to introduce a flexible functional basis (e.g., splines) for representing the function and estimating 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 either an I-spline basis with a positive constraint on the coefficients or a B-spline basis with equally spaced knots and a monotonicity constraint on the coefficients. A similar approach has been proposed in the framework of kernel regression (Hall & Huang, 2001; Henderson, List, Millimet, Parmeter, & Price, 2008), where the kernels are modified locally in order to achieve monotonicity. Another approach is the projection method (Bloch & Silverman, 1997; Friedman & Tibshirani, 1984; Mammen, Marron, Turlach, & Wand, 2001), where the unknown monotonic function is estimated in an unconstrained fashion and then projected onto the convex subspace of the monotonic functions.

The approach that we are going to use here instead of these is in line with the so-called transform/back-transform method. This method is in common use in the FDA literature, and was initially proposed by Ramsay and Silverman (2002, 2005). Basically, the idea is to transform the functions so as to perform an unconstrained estimation, and then back-transform the estimated function to the convex subspace of the monotonic functions. Some very recent work (Boogaart, Egozcue, & Pawlowsky-Glahn, 2014; Egozcue, Díaz-Barrero, & Pawlowsky-Glahn, 2006; Menafoglio, Guadagnini, & Secchi, 2014) focusing on modeling the cumulative distribution functions of absolutely continuous random variables (inspired by the pioneering work on compositional data by Aitchison, 1982) formalized this approach by imposing a suitable Hilbert structure on the set of probability density functions and an isometric bijective map on L^2 for transforming and back-transforming functional data and conveniently mapping the entire statistical analysis in a linear subspace of L^2 (i.e., the zero-mean L^2 functions). The present paper, on the other hand, imposes 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 that satisfy an equality constraint on one edge of the domain and an inequality constraint on the other, with an associated isometric bijective map to L^2 that allows us to model the temporal dependence in an unconstrained framework. In the remainder of this paper, we will refer to this as the $\mathcal{M}^2$ space. To the best of our knowledge, this is the first time that a geometry in a functional space has been introduced and formalized in order to obtain a sound theoretical framework for modeling the temporal dependencies among constrained functional data.

The literature dealing with the temporal dependencies among functional data is more recent, dating to the end of last century. The pioneering contribution to the topic is that of Bosq (1991), who derived a functional Yule–Walker estimator for time-dependent functional data. Functional autoregressive models (FAR) are the most commonly used approach for modeling temporal dependencies among functional data, due to both their ease of interpretation and their good performances in applications (Elezović, 2009). In FAR models, autoregressive parameters are replaced by Hilbert–Schmidt operators, and thus, model estimation is defined by the estimation of the autoregressive operators. Various different methods have been presented in the literature; for recent surveys, see Hormann and Kokoszka (2012) and Horváth and Kokoszka (2012). Autoregressive operators are linked directly with lagged autocovariance operators (e.g., Kargin & Onatski, 2008), and thus one possible approach would be to estimate the lagged autocovariance operators from the functional time series, and then to estimate the autoregressive operators accordingly. However, sample autocovariance operators are typically replaced by reduced rank approximations because of the infinite dimensionality of functional data, and in order to obtain more stable estimates. A spread approach relies on functional principal component decomposition (e.g., Aue, Norinho, & Hörmann, 2015; Hyndman & Shang, 2009; Hyndman & Ullah, 2007; Shang, 2013) and the use of reduced numbers of principal components. Other alternative reduced rank approximations that have been presented in the literature are based on wavelet expansions of the original data (Antoniadis & Sapatinas, 2003) and on predictive factors (Kargin & Onatski, 2008).

Another approach to the estimation of the autoregressive operators is the direct minimization of the mean squared error of prediction. However, the minimization problem has to be approached with some care in order to avoid over-fitting, due to the infinite dimensionality of functional data. For instance, Fan and Zhang (2000) and

Elezović (2009) used a two-step approach: (i) 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), which is actually a continuous family of point-wise scalar autoregressive models; and (ii) smooth the autoregressive functions thus obtained, to take into account the effects of neighboring points of the domain in determining the value at a given point of the function in the next period.

The present paper, on the other hand, targets the minimization of the mean squared error of prediction directly, and from a perspective that is more consistent with current research in FDA, by introducing a penalty term involving the squared Hilbert–Schmidt norm of the autoregressive operators into the objective function. Under this approach, a full rank operator is obtained by shrinking the set of degenerative solutions (that one would obtain without penalty) toward a temporal independence scenario (that one would obtain by setting the penalty constant to infinity). Specifically, 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 (as are in fact used in the application) comes naturally, and thus is not detailed here.

The rest of the paper is structured as follows. Section 2 begins by introducing the space $\mathcal{M}^2(a,b)$ and an isometric bijective map to a subspace of $L^2(a,b)$, and describes the $\mathcal{M}^2$-FAR model in detail, with a particular emphasis on model estimation and the choice of the autoregressive order. Section 3 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 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 also imply that $\mathcal{M}^2(a,b) \subset L^2(a,b)$ and that all functions belonging to $\mathcal{M}^2(a,b)$ increase monotonically and are bounded. Note that if condition $g(b) < 1$ is replaced with $g(b) = 1$, we obtain exactly the same conditions that are required to define the pre-Hilbert space that leads to the definition of the Bayes space geometry introduced and developed by Boogaart et al. (2014), Egozcue et al. (2006), and Menafoglio et al. (2014). With respect to the curves studied in those works (that are valued one at the right edge of the domain), the curves that we are dealing with are subject to a right-censoring effect, which means that they are valued less than one 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 $g \in \mathcal{M}^2(a,b)$,

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

is its image, which belongs to $L^2(a,b)$. By applying the exponential function and integrating between a and $s \in [a,b]$ on both sides of Eq. (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 looks at g as a cumulative distribution function of a scalar absolutely-continuous random variable and maps it to the natural logarithm of the corresponding hazard function. Thus, we will call it the log-hazard transformation and its inverse transformation the anti-log-hazard transformation, and will refer to them as 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 introduced for mathematical tractability, which allows it to be used in different application contexts as well.

In the rest of this 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 that defines 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

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

- 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

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

Note that the neutral element of addition $\oplus$ is $1 - e^{-(s-a)}$ (i.e., the cumulative distribution function of the exponential distribution with a 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 $\mathrm{logH}\left(\mathcal{M}^2(a,b)\right)$ embedded in $L^2(a,b)$. Specifically, 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 $\langle \cdot,\cdot \rangle_{\mathcal{M}^2} : \mathcal{M}^2(a,b) \times \mathcal{M}^2(a,b) \to \mathbb{R}$, given by

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

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 in Eq. (5) are defined as:

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

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

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 Eqs. (3) and (4), 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 Eq. (5). For example, monotonic non-decreasing step-wise functions belong to the closure of $\mathcal{M}^2(a,b)$, 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, requiring the closing of the space and – relying on the separability of $L^2(a,b)$ – the defining of the operations of addition and scalar multiplication consistently with Eqs. (3) and (4) on the closure. Nevertheless, this extension (which would make logH an isometric bijective map over the entire space $L^2(a,b)$) is outside the scope of this work with 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 Lemma 1 shows, the predictions provided by the estimated model are linear combinations, in the sense defined in Eqs. (3) and (4), of functions in $\mathcal{M}^2(a,b)$ that are guaranteed to be in $\mathcal{M}^2(a,b)$. Fig. 1 summarizes all relationships between $\mathcal{M}^2(a,b)$ and $L^2(a,b)$.

2.2. Functional autoregressive model

Here, we describe the model that we use for dealing with temporal dependence. The model can be formulated coherently on either the original data in the space $\mathcal{M}^2(a,b)$ or the log-hazard transformed data in $L^2(a,b)$. To help the intuition, we report the latter formulation here, as it is the one that is used in practice for computation. We will denote this model $\mathcal{M}^2$-FAR. Let $\{f_t\}_{t=1}^T$ be a collection of random functions in $L^2(a,b)$ (here, the log-hazard transformed functions $f_t = \mathrm{logH}(g_t)$) generated sequentially through discrete time t. We assume that $f_t(s)$ depends on the values that are assumed earlier by the random functions potentially appearing in the sequence at each domain

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

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

location $s \in [a,b]$. Let us model this temporal dependence conditionally through a (non-concurrent) autoregressive functional time series. Specifically, a functional autoregressive model of order p (FAR(p)) is defined as

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

or equivalently

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

where the Hilbert–Schmidt operator Ψ_j plays the role of the jth 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 order p can be selected by following various different approaches. For example, Kokoszka and Reimherr (2013) propose a multistage testing procedure that exploits the FAR representation in terms of functional principal components, and derives the approximated distributions of suitable test statistics. Another approach extends classical time series identification tools to the functional framework. In scalar time series analysis, it is 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 of evaluating the autocorrelation and partial autocorrelation functions for different values of the lag, and deciding which (if any) autoregressive or moving average components should be used (Box, Jenkins, & Reinsel, 2013). We perform a similar investigation in the FAR framework by introducing measures of the functional autocorrelation and functional partial autocorrelation that play the same roles as their scalar counterparts. Define the functional autocorrelation function of lag k of a functional time series $\{f_t\}_{t=1}^T$ by

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

which expresses the correlation between $f_t(s)$ and $f_{t+k}(u)$. Even though one could integrate Eq. (10) and obtain a scalar measure of the autocorrelation, which can be plotted as a function of k in a standard correlogram, we instead focused on the visual comparison of the functional autocorrelation function in Eq. (10) along k, in order to obtain a better understanding of the dependence across functions observed at different times. In analogy with the definition of scalar partial autocorrelation, define 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 of the functional autoregressive model of order p in Eq. (8) at time $t+k$, and $\{r_t^{p*}\}$ are the functional residuals of the functional autoregressive model of order k in Eq. (8) at time t, fitted to the reversed series.

2.3. Model estimation and prediction

As was mentioned in the introduction, we estimate the lagged autoregressive operators Ψ_j and the non-centrality function α by direct minimization. Specifically, we obtain the estimates as the solution of the following penalized minimization problem:

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

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 the predictions to the data. The second term, on the other hand, is the sum of the squared Hilbert–Schmidt norms of the lagged autoregressive operators Ψ_j. The lower this term, the lower the autocorrelation associated with the estimated model. Finally, λ is a 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., models in which the effect of the last p functions on the present function is weaker). In fact, 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, thus trivially leading to the prediction of future curves using 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 to an interpolating model, and in particular (consistently with Theorem 1 below) to the one with the minimal Hilbert–Schmidt norms of the lagged autoregressive operators Ψ_j. As is common practice in FDA, a suitable value of the penalty constant λ can be selected either by minimizing the prediction residual sum of squares on a test sample or 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. On the other hand, if more robust estimates are desired, larger values of λ might be favoured. Either way, the following theorem proves the existence and uniqueness of the estimators of α and Ψ_j for $j = 1, \ldots, p$, for any choice of the value of λ, and also provides 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 $s, u \in [a,b]$, there is always a unique solution of the minimization problem in Eq. (12), which is equal to:*

$$\hat{\psi}_j(s,u) = \sum_{t=p+1}^{T} \left(\mathbb{P}_\lambda^{-1}(\mathbf{f}_t - \bar{\mathbf{f}}) \right)\big((j-1)(b-a)+u\big) \times \left(f_t(s) - \bar{f}_{[0]}(s) \right);$$

$$\hat{\alpha}(s) = \bar{f}_{[0]}(s) - \sum_{j=1}^{p} \int_a^b \hat{\psi}_j(s,u)\bar{f}_{[j]}(u)\mathrm{d}u;$$

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 $\sum_{t=p+1}^{T} \big\{ \big(\mathbf{f}_t(\tilde{s}) - \bar{\mathbf{f}}(\tilde{s})\big)\big(\mathbf{f}_t(\tilde{u}) - \bar{\mathbf{f}}(\tilde{u})\big) + \lambda \big\}$ and with $\tilde{s}$ and $\tilde{u} \in \big(a, b+p(b-a)\big)$.

Finally, before moving on to the Italian natural gas market application that motivated this research, we want to point out the coherence of the entire working pipeline, 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 that the joint use of the geometry introduced in Section 2.1 and the model estimation procedure described in this section guarantees that the plug-in predictions will satisfy all of the constraints that characterize functions that belong 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 estimation method (i.e., the minimization of Eq. (12)) proposed here is introduced for the

estimation of functional autoregressive models. Nevertheless, it is trivial to extend it to the estimation of functional-to-functional non-concurrent regression models simply by 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 $x_{1i}, \ldots, x_{pi}$:

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

with $\{y_i\}_{i=1,\ldots,n}$ being the functional responses and $\{x_{ji}\}_{\{j=1,\ldots,p \;\; i=1,\ldots,n\}}$ being the functional regressors. This makes our estimation method a functional generalization of ridge regression (Hastie, Tibshirani, & Friedman, 2009).

3. Application to the Italian natural gas balancing platform

3.1. Context

The natural gas market has been studied and discussed extensively over the last decade, from the economic, political, and environmental viewpoints. In Europe, for example, several legislative and infrastructural measures have been undertaken for the regulation of the market. Two such measures are the legal splitting of pipeline managers and gas shippers, and the legislation requiring obligatory third party access to transmission, distribution, storage and liquefied natural gas capacity data (European Union, 2003).

While favouring a liberal market, such measures have created several new logistic challenges. In Italy, and various other markets, the national pipeline is no longer controlled solely by 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 various different exporting countries, such as Algeria or Russia. The gas is then consumed by civil, industrial, and thermo-electric stations across the country. The role of the pipeline manager, Snam s.p.a., is to balance injections and consumption via storage or other measures. In fact, the responsibility for possible imbalances is assigned to each shipper, which has to predict its injection and withdrawal amounts and communicate these forecasts to Snam daily, with a penalty for any imbalances, whether positive or negative.

In this setting, the Italian Natural Gas Balancing Platform (PB-GAS) was introduced in December 2011, with the ultimate aim of achieving a self-balancing system. The PB-GAS provides a platform on which Snam, shippers, and traders can perform a physical balancing of the pipeline network and procure gas on a day-ahead basis. It is a system in which gas operators sell and buy natural gas virtually, regardless 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. Each day, Snam submits a demand bid or supply offer for a volume of gas that corresponds to the overall imbalance of the system, with a price equal to 23 or 0 Euro/GJ respectively, while the operators submit demand bids and supply offers for the storage resources that they have available.

In this situation, demand bids and supply offers are sorted by price, from the highest to the lowest and from the lowest to the highest respectively, so that the demand and offer curves are obtained as the cumulative sum of the quantities in GJ. Henceforth, we refer to offer curves rather than supply curves, since the former are built from the actual prices and quantities that the traders offer in the auction, while the latter represent individual traders' marginal cost curves. The selection of bids/offers accepted on the PB-GAS is based on an 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 be always on the left of the intersection, and thus to exchange the volume of the imbalance deterministically.

As part of balancing the network, shippers can also take advantage of the market from a speculative perspective, buying natural gas at a lower price or selling excess gas at a higher price, with respect to their benchmark supplying indexes. It is clear that forecasting tools are critically important for individual shippers' decision-making. For example, simple price prediction is the first procedure to implement. However, the pointwise or interval price forecast alone is of limited utility when the shape of tomorrow's curves, and thus the resulting equilibrium price, can be modified strongly by the effect of a single bid. It would be much more useful to have predictions of the entire demand and offer curves. Such tools would allow traders to see the effects of their bids on the shape of tomorrow's curve, and on the price itself, directly.

3.2. Data description

The data used in our analysis are from the first thirteen months of the PB-GAS, namely from December 1st, 2011, to December 31st, 2012. The data are available from the website of Gestore Mercati Energetici (2013). The original data are reported in XML format, with a single entry representing an awarded bid, with its own code, date, trader name, type (sell or buy), awarded price, and awarded quantity. We build the offer (and demand) curve for each day by ordering the selling bids increasingly (decreasingly) by price and obtaining the values of the quantities by cumulating each single awarded quantity.

Before applying the model described in Section 2, the raw data are 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 for the range of investigation. Specifically, the smoothed versions of the offer (and demand) curves are obtained by means of a local polynomial regression, as implemented in the *R* function `locpoly` of `KernSmooth` (Wand and Ripley, 2015). Additional details are reported in the Appendix. The legislative upper bound of 23 Euro/GJ allows us to scale the curves by 23^{-1} without loss of information in order to constrain them vertically between zero and one. The scaled offer and reverse-demand curves naturally satisfy

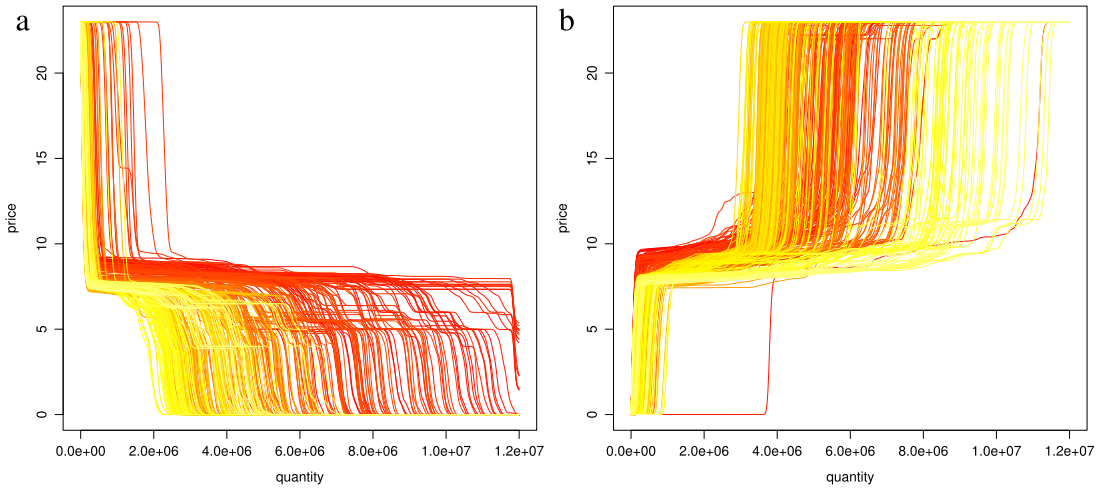


Fig. 2. Smoothed functional time series of (a) demand and (b) offer curves. Color denotes time, with the oldest curves being darkest and the most recent ones brightest. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the requirement 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 make the local polynomial regression output equal to zero for the offer and to one for the demand when the quantity is zero. The smoothed functional time series obtained are plotted in Fig. 2.

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 as a function of quantity. We apply the methodology presented in Section 2 to models $\{1 - g_t^D\}_{t=1}^T$ and $\{g_t^S\}_{t=1}^T$. In particular, we estimate the $\mathcal{M}^2$-FAR(p) models described in the previous section separately for the two time series, for different choices of the autoregressive order p. Moreover, since it is reasonable to assume that the past curves of both demand and supply will influence the future demand and supply curves jointly, we also estimate bivariate $\mathcal{M}^2$-FAR(p) models. As in classical time series analysis, the multivariate extension is naturally:

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

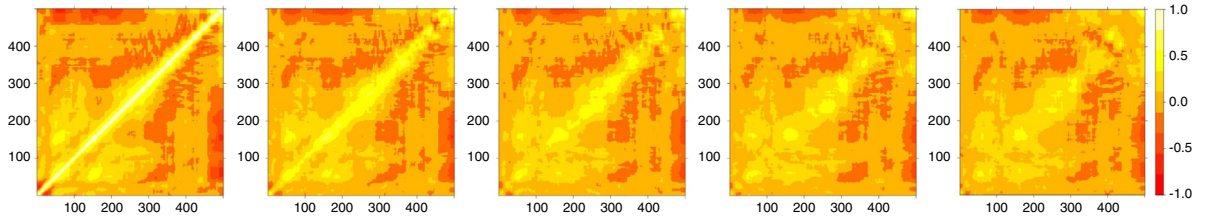
$$f_t^S = \alpha^S + \sum_{j=1}^{p} \Psi_j^{SS} f_{t-j}^S + \sum_{j=1}^{p} \Psi_j^{SD} f_{t-j}^D + \epsilon_t^S,$$

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. For fixed values of p, the latter bivariate FAR (BFAR) model has the computational complexity of a univariate FAR model of order $2p$. The penalization parameter has been fixed at $\lambda = 10^{-7}$ for all p and for both FAR and BFAR models. The particular choice of λ has only a minor impact on the results of the analysis. Indeed, the performances obtained are qualitatively comparable for values of λ ranging from 10^{-5} to 10^{-9}.

As for the choice of the $\mathcal{M}^2$-FAR order p, the method of Kokoszka and Reimherr (2013) cannot be applied in a straightforward way in our setting, which also considers 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 Eqs. (10)–(11) and reported in Figs. 3 and 4. The plots show the sample autocorrelation function and the 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.

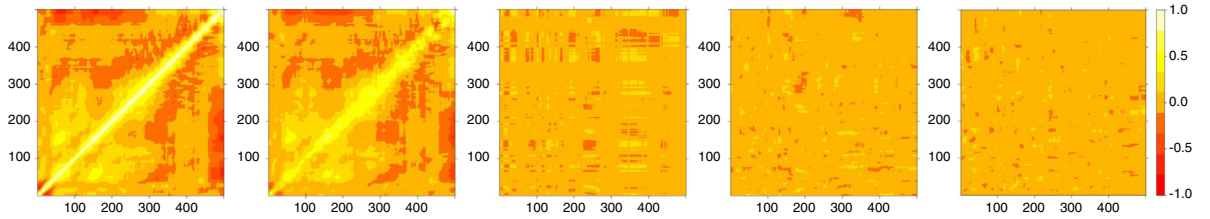
Fig. 3 shows that the autocorrelation is persistent for increasing lags as well, which is a typical feature of scalar autoregressive models. Some details of the autocorrelation function are amenable to an application interpretation. First, a higher autocorrelation is registered in the first part of the curves domain, which is the region in which the demand and offer curves typically intersect. It is clear that yesterday's price influences the bids (and thus the shape of the curves) of today, so it is natural to expect a high autocorrelation in this part of the domain. Second, the autocorrelation also remains high for increasing lags, mostly around the diagonal of the plots in Fig. 3. This means that the values of the curves at point s are influenced mainly by the curves observed previously in the neighborhood of s. The structure described by Fig. 4 suggests that the main dependence of curve t on the past comes from the curve observed at $t - 1$. In fact, the dependence of the curve at time t on the curve at time $t-2$ is basically zero. This suggests that a lag-1 model may be appropriate for fitting our data.

As a second approach, we consider data-driven measures. For example, we can assess the model's goodness-of-fit for increasing values of the autoregressive order p. This also allows us to compare the univariate and bivariate models. The comparison is carried out using different functional measures of the discrepancies between the original and predicted curves. Given our modeling approach,



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

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



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

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

the most natural measure 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, this coincides 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 the goodness-of-fit, namely 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 1–2. The autoregressive order p has similar impacts on the estimation of both the demand and offer curves (i.e., for a fixed value of p, the errors for the demand and offer curves are of roughly the same order of magnitude) for both the $\mathcal{M}^2$-FAR and $\mathcal{M}^2$-BFAR specifications. Specifically, the bivariate model performs dramatically better than either of the univariate ones for a fixed value of p, due to the obvious dependence between the supply and demand curves. In particular, the performances of $\mathcal{M}^2$-BFAR(1) and $\mathcal{M}^2$-FAR(2) are comparable. As a reference, we also fit a functional version of the simple exponential smoothing model (FES), but it is outperformed by all of the $\mathcal{M}^2$-FAR and $\mathcal{M}^2$-BFAR specifications. Note that the models in lines 3–7 of the two tables have the same accuracies to three decimal places for the indexes calculated on the original scale of the functions. Details of the implementation and tuning of the FES model are provided in the Appendix.

As additional measure of the goodness-of-fit, we also consider an application-driven approach. As was discussed in Section 3.1, the prediction of the whole curve is more informative than simple price forecasting. However, price forecasts are a byproduct of our procedure (they can be obtained easily as the intersection of the two predicted curves), and it is desirable that such predictions be reliable. We calculate the predicted daily prices and quantities, and compare them with the real prices and quantities achieved through the daily bids, in terms of root mean squared errors (RMSE), as reported in Table 3.

We now focus on price prediction, since, as has been said, it is likely to be the chief feature of interest to a trader. As benchmarks, we fitted a scalar exponential smoother (SES) to time series of prices using the `ses` function of the *R* package `forecast` (Hyndman, 2016; Hyndman & Khandakar, 2008) with the default specification, along with the FES model. The RMSE of the price from the univariate $\mathcal{M}^2$-FAR(1) model (i.e., 2.03 Euro) is much larger than those of the prices from the FES and SES models (i.e., 0.31 and 0.20 Euro, respectively). On the other hand, the $\mathcal{M}^2$-FAR(2) model outperforms the two benchmark methods, providing an RSME of price of 0.18 Euro. However, if focussing on quantity, the univariate approach based on $\mathcal{M}^2$-FAR(p) model turns out to be unable to achieve an RMSE that is comparable to or better than those from the two benchmark methods. In this sense, the bivariate approach is key to outperforming the benchmark methods for both price and quantity prediction. Indeed, the $\mathcal{M}^2$-BFAR(1) model achieves RMSEs on both prices and quantities (0.18 Euro and $2.16 \cdot 10^5$ GJ, respectively) that are lower than those from the FES and SES models. Given the models' performances for predicting both the offer

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

	$\mathcal{M}^2$-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 and original offer curves.

	$\mathcal{M}^2$-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}

Table 3
RMSEs for the price (in Euro) and quantity (in GJ), obtained as crossing point between the 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

and demand curves and the price and quantity, and also considering the suggestion of Fig. 4, we therefore choose the $\mathcal{M}^2$-BFAR(1) model as the final model for our analysis.

3.4. Trading example

To conclude the analysis, we provide an example of the tremendous additional insights that traders can obtain from the whole-curve forecasts. Consider the prediction for day T, reported in panel (a) of Fig. 5. Suppose that a given trader is aware that he is going to buy a large quantity Q of natural gas tomorrow, for both legislative and logistic reasons (for full details, please refer to the PB-GAS normative, on Gestore Mercati Energetici's website). The trader can lower the price by submitting an extra supply offer for a small quantity, that eventually he is going to buy above Q. For example, assume that he submits an offer of 240,000 GJ at 7.20 Euro. The modified curve is represented by a dotted line in panel (a) of Fig. 5, while panel (b) shows a zoomed view of the 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 help us understand which is the most convenient action, panel (c) of Fig. 5 reports the price obtained as a function of the price and quantity of the extra bid. If a trader wants to move the intersection point, the lower the offered price, the higher the offered quantity needs to be. Evidently, prices above the estimated one affect the shape of the curve after the intersection, and therefore have 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 have proposed a model for functional time series that preserves particular features of the curves, such as monotonicity and the equality and inequality constraints at the edges of the domain. A bijective map that associates each possible constrained function with an unconstrained one is introduced, and an isometry between the space of the constrained functions and L^2, a suitable geometry, is developed in order to achieve this. Specifically, we provide the constrained functions with a suitable pre-Hilbert structure. The transformed curves are then modeled by means of a 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 the functional data and functional predictions, with a penalty term based on the Hilbert–Schmidt squared norm of autoregressive lagged operators. We have proved that a unique solution always exists, 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 generalized easily to the modeling of multivariate functional data or the inclusion of scalar covariates, or other functional predictors that are available at the time of prediction. The inclusion of a trend and/or seasonality term in the model would be tricky in terms of estimation but might lead to further improvements in the predictive

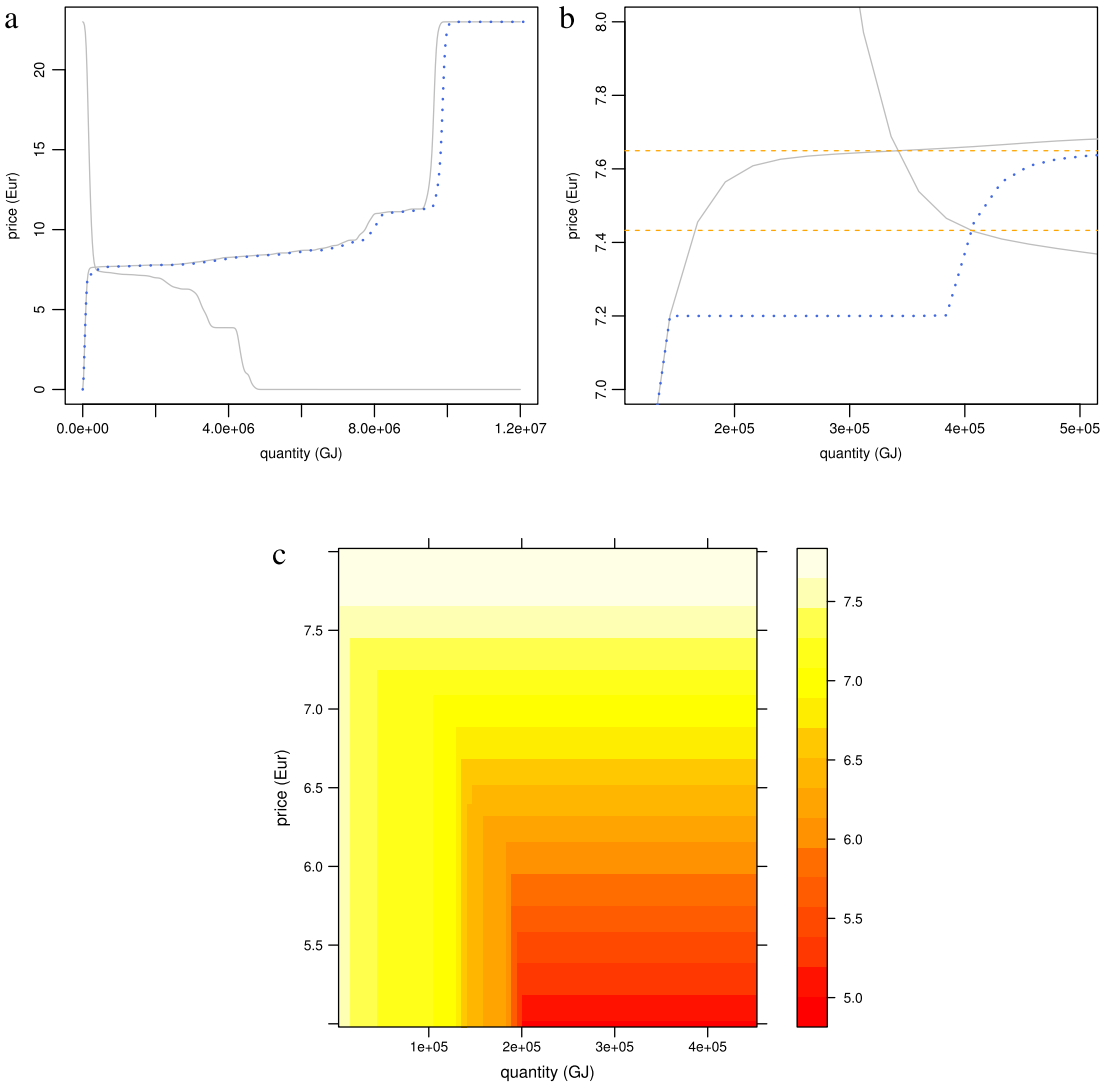


Fig. 5. What-if simulations: (a) curve prediction (continuous lines) and a offer perturbation (dotted line), (b) zooming in on the neighborhood of the intersections, where the horizontal dashed lines represent the prices obtained as default and those after the bid perturbation; (c) price heatmap (brighter colors for a higher resultant price) obtained as a function of the price and quantity of an extra bid. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

performance, and would certainly be worthy of future investigation.

The method has been applied successfully to data on the Italian Natural Gas Balancing Platform, revealing that tomorrow's curves are influenced strongly by those of today. The prediction of tomorrow's curves is of critical interest for gas traders, as it allows for what-if simulations that can assist decision making if one wants to behave speculatively in this market.

Acknowledgments

We thank the referees for their constructive comments, which led to substantial improvements in the manuscript. This work was motivated and partly funded by Moxoff.

Appendix

A.1. Proofs

Proof of Theorem 1. Let us begin by showing 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 Eq. (12) with respect to α, thus trivially obtaining

$$\hat{\alpha} = \frac{1}{T-p} \sum_{t=p+1}^{T} \left(f_t - \sum_{j=1}^{p} \Psi_j f_{t-j} \right) = \bar{f}_{[0]} - \sum_{j=1}^{p} \Psi_j \bar{f}_{[j]}.$$

Hence, the minimization of Eq. (12) can be carried out on the simplified objective function depending only on Ψ_j for $j = 1, \ldots, p$ (obtained using Eq. (12), by replacing α

with $\hat{\alpha}$):

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

The proof of the existence and uniqueness of the minimizers comes by noticing that Ψ_j for $j = 1, \ldots, p$ being Hilbert–Schmidt operators, the second term in Eq. (A.1) 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 in Eq. (A.1) is a positive definite quadratic form with respect to $\{\Psi_j\}_{j=1,\ldots,p}$, and thus it admits a unique minimum. It is obtained by linear combination with the positive coefficients (i.e., one 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, according to the Fubini–Tonelli Theorem, $\|\Psi_j\|_{HS}^2 = \int_a^b\left(\int_a^b\psi_j^2(s,u)\mathrm{d}u\right)\mathrm{d}s$, and thus the minimization of Eq. (A.1) can be carried out separately for each value of $s \in [a,b]$; i.e., minimizing:

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

with respect to $\{\psi_j(s,\cdot)\}_{j=1,\ldots,p}$ for all $s \in [a,b]$. Focusing on the case $p = 1$, the minimization problem in Eq. (A.2) can be seen as a continuous version of a ridge-regression-like minimization. Thus, we have

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

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 obtained directly by chaining the functions $f_{t-1}, \ldots, f_{t-p}$, for $t = p+1, \ldots, T$, 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 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 now show 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 computation, we have

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

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

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

meaning that $\hat{f}_{T+1}(s)$ is equal to

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

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

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

which belongs to $\mathcal{M}^2(a,b)$ being $\mathcal{M}^2(a,b)$ a space vector with respect to the addition in Eq. (3) and the scalar multiplication in Eq. (4). □

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` (Wand and Ripley, 2015), with degree 0 and Gaussian kernel. When dealing with smoothing, the choice of the bandwidth parameter is important. Here, we fixed it to 6000 for all curves. This choice has been made subjectively by eye, but turns out to be an acceptable compromise between smoothness and coherence for the raw data. Despite being subjective, this choice-by-eye is used widely, and is satisfactory in many situations (see Wand & Jones, 1994, Section 3.1). The derivatives of the smoothed functions are obtained via numerical derivation.

A.3. Functional exponential smoother

Exponential smoothing is a useful method for producing one-step-ahead predictions for 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 using least squares

estimation, by means of the `optim` function of *R* and the `ses` function of package `forecast` (Hyndman, 2016). 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. Statistical Methodology*, *44*(2), 139–177.

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

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

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

Boogaart, K. G., Egozcue, J. J., & Pawlowsky-Glahn, V. (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.), *NATO science series C, Nonparametric functional estimation and related topics* (pp. 509–529). Springer.

Box, G., Jenkins, G., & Reinsel, G. (2013). *Wiley series in probability and statistics, Time series analysis: forecasting and control*. Wiley.

Brabec, M., Konár, O., Pelikán, E., & Malý, M. (2008). A nonlinear mixed effects model for the prediction of natural gas consumption by individual customers. *International Journal of Forecasting*, *24*(4), 659–678.

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

Egozcue, J., Díaz-Barrero, J., & Pawlowsky-Glahn, V. (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., & 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*(2), 303–322.

Ferraty, F., & Vieu, P. (2006). *Springer series in statistics, Nonparametric functional data analysis: theory and practice*. Springer.

Friedman, J. H., & Tibshirani, R. (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 13.02.13).

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

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

Henderson, D.J., List, J.A., Millimet, D.L., Parmeter, C.F., & Price, M.K. (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., & Kokoszka, P. (2012). Functional time series. In C. Rao, & T. Subba Rao (Eds.), *Handbook of statistics. Vol. 30* (pp. 155–186). Oxford, UK: Elsevier.

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

Hyndman, R. J. (2016). forecast: Forecasting functions for time series and linear models (R package v7.1). http://cran.r-project.org/package=forecast.

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

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

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

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

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

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

Menafoglio, A., Guadagnini, A., & Secchi, P. (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., & Roulier, J. A. (1977). Monotone and convex spline interpolation. *SIAM Journal on Numerical Analysis*, *14*(5), 904–909.

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

Ramsay, J. O., & Silverman, B. W. (2002). *Springer series in statistics, Applied functional data analysis*. New York: Springer-Verlag, Methods and case studies.

Ramsay, J., & Silverman, B. (2005). *Springer series in statistics, Functional data analysis*. Springer.

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

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

Wand, M., & Ripley, B. (2015). KernSmooth: Functions for Kernel Smoothing Supporting Wand & Jones (1995). R package v2.23-15. http://cran.r-project.org/package=KernSmooth.

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

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