The Applicability of Functional Data in Multi-Unit Auctions

Dissertation zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaft

eingereicht an der

Fakultät für Wirtschaftswissenschaften der Universität Regensburg

vorgelegt von Stefan Rameseder

Berichterstatter:

Prof. Dr. Rolf Tschernig, Universität Regensburg
Prof. Dr. Dominik Liebl, Universität Bonn

Tag der Disputation: 9. November 2023

This dissertation is dedicated to my yet unborn daughter.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and overview

1.1 Introduction

The author recalls an early morning with the managing directors of entelios AG, now EnerNoc AG, a demand-response supplier in the German Power Reserve and Balancing Energy Market on Wednesday, December 18th, 2013. As a mathematician and young econometrician, the author was asked to provide a price for the offering of 20 megawatts (MW). The price would be the company's bid within the secondary power reserve for the first Christmas week starting on December 23rd. The mechanism was a frequential, discriminatory or pay-as-bid multi-unit auction. While the maximum accepted price from week before was approximately 800 €/MW, the last year's price for the Christmas week was approximately 8,200 €/MW. If a (risk-taking) bid of 8,000 €/MW is locked in and the maximum accepted price will be 7,999 €/MW, the offer will not be accepted, meaning that the company will not receive anything, in particular not 160,000 €. If a (risk-averse) bid of 100 €/MW is provided and the maximum accepted price will be 5,000 €/MW, the company will receive 2,000 €, while the bidder with the highest accepted bid will secure 100,000 € for 20 MW.

1.1.1 Functional Data Analysis

The challenge to empirically model frequential multi-unit auctions was the starting point of this thesis and led the author to a branch of statistics called "functional data analysis" (FDA), which dates back to Ramsay (1982). FDA's growing importance over the last two decades is not only due to increased computational power, but mainly due to the exponential increase in the availability of high-resolution data recordings, e.g., temperature or growth curves in nature or electricity multi-unit auctions in energy markets. Observations of such functions are

called *functional data*, and FDA refers to the statistical analysis of such data. While in the temperature example, data is generated by continuous processes, which, in principle, can be sampled at arbitrary discretization points, FDA also allows us to exploit properties of differentiable functions generated by, in principle, discrete demand or supply step functions as in the electricity example above. In this case, an unbiased estimator can only be defined via the derivative of the supply function.

This thesis aims to contribute to the FDA scientific field by developing novel methods to approach (multi-unit) auctions from a functional data perspective and prove the added value in practice. Moreover, the implementations are made publicly available as packages for the open-source programming language GNU R to simplify the application of the methods onto other multi-unit auctions, e.g., in other balancing energy markets or on certain treasury bills and bonds markets.

The remainder of this chapter contains a brief introduction to functional data and mutli-unit auctions as well as an overview of the three papers that make up the dissertation.

Functional data analysis is the statistical counterpart of functional analysis and analyzes so called "functional data", i.e., discretized observations of "functional random variables" (FRV). These are random variables taking values in a Hilbert space $L^2(\mathcal{I})$, the study subject of functional analysis, where $\mathcal{I}$ is typically a subset of $\mathbb{R}^n$. In most applications, $\mathcal{I}$ is a closed interval on the real line, e.g., $[0,1]$, the random functions are one-dimensional curves, i.e., $x \in L^2([0,1],\mathbb{R})$, and the FRV is a map $X : (\Omega, \mathcal{F}, \mathcal{P}) \to L^2([0,1],\mathbb{R})$ with $\omega \mapsto x(t)$. As a generalization, there are also applications to random surfaces or, more recently, random functions defined on manifolds.

Next to classical statistical tasks such as mean and variance estimation, FDA is also interested in how variables influence each other. Within FDA, the idea of regression analysis is extended to functional variables and differentiates between three different types of functional regression models: the first type (*scalar-on-function regression*) is a regression model with a functional predictor and a scalar outcome variable. The second type is a model where a functional outcome variable is explained by scalar predictors (*function-on-scalar regression*), and in the third type, a function is regressed on another function (*function-on-function regression*), e.g., in functional autoregressive models.

From a theoretical perspective, this dissertation addresses methodological extensions and questions related to the applicability and interpretability of the first type of model. It further investigates the estimation of mean and variance in the case of partially observed data, i.e., by systematically violating the i.i.d. assumption, and uses this novel mean and variance estimator for a new prediction

algorithm that avoids functional autoregressive models.

For a brief introduction to the *scalar-on-function* regression model, we assume that we observe an i.i.d. sample of tuples (y_i, X_i), $i = 1, \ldots, n$, where X_i are realizations of a functional random variable taking values in the Hilbert space $L^2([0,1])$ and the y_i are generated from the model

$$y_i = \alpha + \int_{[0,1]} \beta(t) X_i(t)\, \mathrm{d}t + \varepsilon_i, \tag{1.1}$$

where ε_i is an scalar-valued i.i.d. error term that has mean zero, finite variance, and is independently distributed from the X_i. The deterministic parameter function $\beta(t)$ is often the quantity of interest because it determines the dependency between X_i and y_i.

Several estimation approaches are available in the literature. Classically, estimation builds on functional principal components (FPCA). In this setting, the spectral theorem is used to approximate both the coefficient function β and the functional observations by the leading eigenfunctions of the random curves' empirical covariance operator as a basis for a finite dimensional function space. On the one hand, this finite dimensional subspace of L^2 is optimal for approximating the functional predictors for a given number of eigenfunctions but, on the other hand, may not be the best fit for approximating β. Alternatively, the estimation of β can also be achieved by using other basis systems, such as splines. Next to estimating β, splines are also used to preprocess the auction data into functional data, see Figure 1.1.

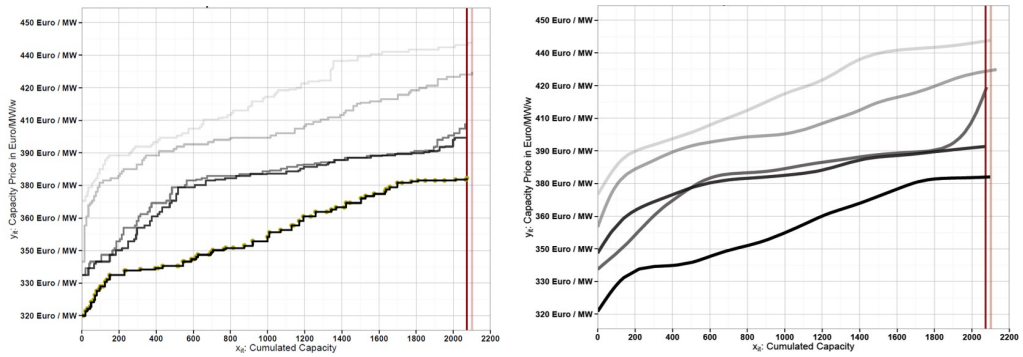

Figure 1.1: LEFT: Five weekly, discrete auction step functions from 19.09.2011 until 17.10.2011 (light gray to black). RIGHT: Representation as functions, i.e., differentiable auction curves.

1.1.2 Multi-Unit Auctions

Auctions are seemingly pervasive, they are used to sell a variety of agricultural commodities and natural resources as well as fine art, real estate, and used cars. In 2022, Google processed approximately 100,000 searches every single second, the results of the requests appearing in the iconic results page—the advertised ones on the top positions, the organic results below. The order of those links is determined at a position auction, a type of auction used for Google AdWords. Besides AdWords, the allocation of reserve capacity in Germany accounts for hundreds of millions of euros every year and is one of the most important indicators for renewable energy investments. This dissertation deals with these auctions by applying functional data methods onto frequential, discriminatory or pay-as-bid, sealed-bid multi-object/unit auctions.

In short, *multi-object* auctions involve the sale of several different objects, e.g., AdWords positions, while *multi-unit* auctions involve the sale of several units of the same, non-differentiable object, e.g., MW or MW/h. If there is one price to pay by every bidder, the mechanism is called uniform pricing, while a pricing mechanism is called *discriminatory* or *pay-as-bid* if every bidder pays their own or different prices. *Sealed-bid* describes the closed format of the auction, i.e., the bids are not known to other bidders. Auctions are referred to as *frequential* if they are held on a regular basis, e.g., every day or every week. For a general introduction to auctions, the author recommends Krishna (2009), while for a structural econometric introduction, Paarsch et al. (2006) provide a standard.

1.2 Overview and contribution

1.2.1 Functional linear regression with points of impact

The first paper was published as Liebl et al. (2020) and is coauthored with Dominik Liebl and Christoph Rust. It extends the literature on an augmented scalar-on-function regression model, namely the functional linear regression model with points of impact. The paper builds on the work of Kneip et al. (2016) and introduces several adjustments to the estimation procedure, improving thereby the finite sample performance. The model suggested in Kneip et al. (2016) is defined by

$$Y_i = \int \beta(t) X_i(t)\,\mathrm{d}t + \sum_{s=1}^{S} \beta_s X_i(\tau_s) + \epsilon_i,\ \ i = 1, \ldots, n. \tag{1.2}$$

Here, the functional predictors not only influence the scalar-valued response variable via the integral, but the functional predictors X_i are also directly related to the outcome variable Y_i at specific influential points $\tau_s \in [0,1], s = 1, \ldots, S$. These so-called *points of impact* are not known a priori and have to be estimated together with the other model parameters $\beta, \beta_1, \ldots, \beta_S$. To identify points of impact, the random functions X_i must not be too smooth, and must possess what Kneip et al. (2016) call *specific local variation*. The principal idea behind identifying points of impact is to search for peaks over discretization points t of the correlation between Y_i and $Z_i(t, \delta)$, where $Z_i(t, \delta)$ is the central second-order difference quotient of $X_i(t)$ with differencing parameter $\delta > 0$. Given a set of potential points of impact, the model can be estimated using standard FDA approaches. Kneip et al. (2016) use a finite-dimensional representation of the curves (functional principal components) and select the model complexity (number of principal components, number of points of impact, differencing parameter δ) by minimizing the Bayes Information Criterion (BIC) *simultaneously* over these three parameters. They also suggest theoretically motivated choices for the number of points of impact and the differencing parameter δ, which, however, are not really suited for practical applications.

In our paper, we suggest several adjustments to the original procedure. First, we use a different basis system and expand β using B-splines. As we argue, the originally proposed FPCA approach is optimal for approximating the functional covariates but may be less well-suited to approximating β, especially if it is a relatively smooth function. Additionally, as we show in a simulation exercise, selecting the number of (FPCA) basis functions and number of points of impact simultaneously may result in severe misspecifications of the model. Second, we

decouple the estimation of the points of impact and the estimation (and choice of complexity) of β by using an alternating procedure that reduces the risk of choosing a misspecified model.

In an extensive simulation study, we show that our procedure outperforms the original approach in an MSE sense. This is particularly true for smaller sample sizes. Besides that, the paper also provides theoretical arguments for these adjustments and treats a case study with real data using data from Google AdWords. This case study also shows the usefulness of the model framework for decomposing the dependency between the functional predictor and the scalar outcome variable into time-global effects (via β) and time-local effects (via points of impact), which facilitates the interpretability of the estimation results at the point-of-impact locations. To make this methodology accessible to practitioners, an R package implementing both the original estimation procedure and the proposed one is available from one of the authors' Github account.

1.2.2 Partially Observed Functional Data: The Case of Systematically Missing Parts

The second paper was published as Liebl and Rameseder (2019) and is coauthored with Dominik Liebl. It provides a generalization of the classical mean and covariance function estimators $\widehat{\mu}(t)$ and $\widehat{\sigma}(s,t)$ for partially observed functional data using a detour via the fundamental theorem of calculus. The new estimators allow for a consistent estimation of the mean and covariance function under specific violations of the missing-completely-at-random assumption (MCAR).

In our setup, the random functions $X_i(t)$ are only observable over random subdomains $D_i = [0, d_i] \subseteq [0,1]$. These subdomains lead to a t-specific observed data indicator $O_i(t)$ defined as $O_i(t) = \mathbb{1}_{t \in D_i}$; i.e., $O_i(t) = 1$ if $X_i(t)$ is observable and $O_i(t) = 0$ if $X_i(t)$ is missing. The MCAR assumption requires independence between the entire random processes O_i and X_i. If they are independent, the mean function can be consistently estimated via $\widehat{\mu}(t) = (I_1(t))(\sum_{i=1}^n O_i(t))^{-1} \sum_{i=1}^n X_i(t)O_i(t)$ with $I_1(t) = \mathbb{1}_{\sum_{i=1}^n O_i(t)>0}$.

However, motivated by the practical problem of estimating the mean price curve in the German Control Reserve (GCR) Market, the defined setup violates the MCAR assumption by assuming that the dependency of X and O manifests only in the dependency between O and the vertical shift component of X. This violation can be explained by systematic trading strategies in the market. In this case, $\widehat{\mu}(t)$ is biased and inconsistent, while we prove that the proposed mean estimator $\widehat{\mu}_{\mathrm{FTC}}(t)$

$$\widehat{\mu}_{\mathrm{FTC}}(t) = \begin{cases} \widehat{\mu}(t) & \text{if} \quad t \in [0, d_{\min}] \\ \int_{d_{\min}}^{t} \widehat{\mu}^{(1)}(z)dz + \widehat{\mu}(d_{\min}) & \text{if} \quad t \in (d_{\min}, 1], \end{cases} \tag{1.3}$$

where

$$\widehat{\mu}^{(k)}(t) = \frac{I_1(t)}{\sum_{i=1}^{n} O_i(t)} \sum_{i=1}^{n} X_i^{(k)}(t) O_i(t),$$

and where $X_i^{(k)}(t)$ denotes the k-th derivative of $X_i(t)$, is consistent.

To test for this type of dependency, a sequential multiple hypothesis test procedure from Romano and Wolf (2005) is introduced. The idea is to test the basis representation $X(t) = \sum_{j \geq 1} \xi_j \psi_j(t)$ for only ξ_{i1} significantly correlating with d_i while all other $\xi_{ij}, j \geq 2$ do not.

An extensive simulation study compares the new estimators with the classical estimators from the literature in different missing data scenarios and ensures the applicability of both the testing procedure and the proposed estimator's performance. In contrast to the classical estimators, the new estimators also lead to useful estimates of the mean and covariance functions in the German Control Reserve Market that are used in the next publication to predict future supply curves. Supplementary materials such as data and codes are provided at the author's Github account.

1.2.3 Forecasting of discriminatory auction curves with underlying missing data mechanism in the German Capacity Reserve and Balance Power Market

The third paper is single authored and addresses the moment mentioned at the beginning of the introduction, i.e., providing bids for the GCR. The proposed, novel method outperforms classical but standard approaches in this market. While electricity spot markets, e.g., the European Energy Exchange (EEX), are extensively investigated by both industry and researchers, balancing markets, e.g., the GCR, play a minor role. Next to financial volumes, the most important reason is the different pricing mechanism, where the latter typically uses exotic auction formats and therefore needs adjusted statistical tools.

The GCR's market design is a weekly, pay-as-bid multi-unit auction with only partially observable supply functions. According to Grimm et al. (2008) and Engelmann and Grimm (2009), it is strategically optimal to predict and bid almost the (unknown) maximum price the GCC needs to cover for the inelastic demand d_t for week t, i.e., the amount of capacity to secure the stability of the

electricity grid. The demand is determined and announced by the Bundesnetzagentur (BNetzA) on a weekly basis and differs significantly during volatile energy consumption periods like Christmas, see for example Figure 1.1 with two different demands in red or Figure 3.5.3 with different demands in turquoise. The partial observability is due to the fact that bids exceeding the maximum accepted prices are not published.

The paper's principal idea of providing a bid is a four-step process:

1. Estimate consistently the mean of all curves up to t for the whole domain $[0, d_{\max}]$ using $\widehat{\mu}_{\mathrm{FTC}}(t)$ to cover for the (tested) violation of the MCAR assumption
2. Forecast a univariate gravity point for $t+1$ that also includes also exogenous variables
3. Create a curve prediction out of the gravity point and the mean estimate
4. Use the evaluation of the functional prediction at a MW position reflecting demand d_{t+1} and risk-aversion as bid

The paper applies this procedure to four different weekly auctions in Secondary Reserve but is also applicable in the German Primary (PR) and Tertiary Reserve (TR) and can be scaled to general discriminatory multi-unit auctions with underlying missing data mechanisms. As part of the modeling process, the paper provides guidance and codes for every step: the transfer of discrete bids into functional data using P-splines, the testing for the violation, the model estimation, and evaluation using specific, highly asymmetric metrics. ling process, the paper provides guidance and codes for every step: the transfer of discrete bids into functional data using P-splines, the testing for the violation, the model estimation, and evaluation using specific, highly asymmetric metrics.

To the author's knowledge, this is the first evaluation of prediction methods for Secondary Reserve (SR) in the GCR which accounted for more than €1.3 billion between 2011 and 2016.

Chapter 2

Improving the estimation of functional linear regression with points of impact

2.1 Introduction

In many practical applications, one is interested in the relationship between a real-valued outcome variable Y_i and a function-valued predictor $\{X_i(t);\ a \leq t \leq b\}$. In our motivating Google AdWords case study, for instance, we aim to explain the numbers of clicks Y_i using impression trajectories $X_i(t)$, where t denotes a certain day within the considered time interval $[a, b]$ of one year and $i = 1, \ldots, n$ indexes the cross section of keywords associated with the considered Google AdWords ad campaign.[1] The economic success of any ad campaign depends on product specific (time-global) seasonalities as well as on (time-local) events. The slowly varying seasonal component could be estimated using the function-valued slope parameter of the classical functional linear regression model (see, e.g., Hall and Horowitz, 2007). The presence of time-local effects, however, harms such a simple estimation approach (see the right plot in Figure 2.2 for notable examples). Therefore, we use the recent functional linear regression models with so-called Points of Impact (PoI) that allow us to identify and to control for time-local effects.

Point of impact models were originally introduced by McKeague and Sen (2010), who argue that these models are better to interpret than the classical

[1] Online ad campaigns use text corpora populations of relevant search keywords (for instance, `outdoor jacket`, `mountain boots`, etc., in the case of an outdoor equipment campaign) to identify potential customers by their Google searches (see Section 2.4 for more details).

functional linear regression models. Indeed, several convincing real data applications are presented in the related work of Lindquist and McKeague (2009). The method of Kneip et al. (2016) generalizes the original point of impact model by adding a classical functional linear regression component. While the original point of impact model captures only time-local effects, the augmented point of impact model of Kneip et al. (2016) also allows for time-global effects. In our paper we present a new and relevant case study where time-local as well as time-global effects are important for modeling the outcome.

As demonstrated in our simulation study, however, the finite sample performance of the estimation procedure proposed by Kneip et al. (2016) is very sensitive to the performance of the involved model selection. Therefore, we propose an adjusted sequential estimation algorithm that leads to significantly improved and more robust estimation results by using a refined model selection procedure.

The functional linear regression model with PoIs of Kneip et al. (2016) is related to several other works in the literature. Identifiability and estimation of points of impact was originally studied by McKeague and Sen (2010). The authors focus on a one-point of impact model without a functional linear model component; however, the possibility of a partial model misspecification by an additional functional linear model component is also discussed theoretically. Ferraty et al. (2010) and Poß et al. (2020) allow for multiple PoIs within a nonparametric model, but both do not consider a functional linear model component. Matsui and Konishi (2011) consider the extraction of local information within functional linear regressions using a LASSO-type approach, but do not estimate global components. Torrecilla et al. (2016) focus on a classification context, and Fraiman et al. (2016) consider feature selection for functional data at a more general level. Our estimation algorithm uses the penalized smoothing splines estimator for functional linear regression models proposed by Crambes et al. (2009). The related literature is extensive and the following examples are by no means exhaustive. Cardot et al. (2007) consider functional linear regression with errors-in-variables, Crambes et al. (2009) address optimality issues, Goldsmith et al. (2010) focus on penalized smoothing splines within a mixed model framework, and Maronna and Yohai (2013) propose a robust version of the penalized smoothing splines estimator. Scalar-on-function regression models are successfully applied to solve important practical problems. Chiou (2012) proposes a functional regression model for predicting traffic flows. Goldsmith et al. (2012) introduce a penalized functional regression model to explore the relationship between cerebral white matter tracts in multiple-sclerosis patients. Koeppe et al. (2014) consider regularized functional linear regression for brain image data. Gellar et al. (2014a) and Gromenko et al. (2017) propose functional

regression models for incomplete curves. An overview article on methods for scalar-on-function regression is found in Reiss et al. (2017). Readers with a general interest in Functional Data Analysis (FDA) are referred to the textbooks of Ramsay and Silverman (2005), Ferraty and Vieu (2006), Horváth and Kokoszka (2012), and Hsing and Eubank (2015). To the best of our knowledge, we are the first to use methods from FDA to analyze data from an online ad campaign; however, there are several contributions in FDA on related applications. Reddy and Dass (2006) use a classical functional linear regression model to analyze online art auctions, Liu and Müller (2008) analyze eBay auction prices using methods for sparse functional data, Wang et al. (2008) forecast eBay auction prices, Wang et al. (2008) develop a model for the price dynamics at eBay using differential equation models, and Zhang et al. (2010) consider real-time forecasting of eBay auctions using functional K-nearest neighbors.

The rest of the paper and our contributions are structured as follows. The next section (Section 2.2) contains our methodology. In Section 2.2.1, we begin with a short presentation of the original procedure of Kneip et al. (2016). In Section 2.2.2, we introduce our three main proposals (1. Sequential model selection and estimation, 2. Smoothing splines estimator, and 3. Standardizations) which we use to stabilize and improve the estimation procedure of Kneip et al. (2016). The implementation of our estimation algorithm is presented in Section 2.2.3. Section 2.3 contains our simulation results, our case study on analyzing Google AdWords data is found in Section 2.4 and Section 2.5 contains the conclusion. Appendix 2.A presents further simulation results. The online supplement supporting this article contains the R-package `FunRegPoI` and the R-codes for reproducing our simulation study and the real data application.

2.2 Methodology

We formally consider the following functional linear regression model with PoIs introduced by Kneip et al. (2016):

$$Y_i = \int_a^b \beta(t) X_i(t)\,\mathrm{d}t + \sum_{s=1}^{S} \beta_s X_i(\tau_s) + \epsilon_i, \quad i = 1, \ldots, n. \tag{2.1}$$

Here, $(Y_1, X_1), \ldots, (Y_n, X_n)$ denote an i.i.d. sample of scalar response variables $Y_i \in \mathbb{R}$ and random predictor functions $X_i \in L^2([a,b])$, where $\mathbb{E}[Y_i] = 0$ and $\mathbb{E}[X_i(t)] = 0$ for all $t \in [a,b]$. Without loss of generality, we set $[a,b] = [0,1]$. The i.i.d. error term ϵ_i has mean zero, variance $\mathbb{E}[\epsilon_i^2] = \sigma_\epsilon^2 < \infty$, and is independent of X_i. The assumption that Y_i and X_i have mean zero is only for notational simplicity; for the estimation, however, we will explicitly denote the centering of the data.

The function-valued slope parameter $\beta \in L^2([0,1])$ in Model (2.1) describes the time-global influences of X_i on Y_i. The scalar-valued slope parameters $\beta_s \in \mathbb{R}$ take into account the time-local influences where the corresponding (unknown) time-points τ_s denote the locations of the PoIs. The estimation algorithm described below addresses the estimation of all unknown model parameters, namely, the global slope coefficient β, the local influences of the PoIs $\beta_1, \ldots, \beta_S$, and the set of PoI locations $\mathcal{T} = \{\tau_1, \ldots, \tau_S\}$.

In the following, we introduce our basic notation. The functions $X_i(t)$ are observed at p equidistant grid points $t_1, \ldots, t_p$ with $t_j = (j-1)/(p-1)$. For non-equidistant designs, this can always be achieved by pre-smoothing the data. In $\mathbf{Y} = (Y_1, \ldots, Y_n)' \in \mathbb{R}^n$, we collect all observations of the response variable Y_i, and in $\mathbf{X} = (X_i(t_j))_{ij} \in \mathbb{R}^{n \times p}$, we collect all discretizations $X_i(t_j), i = 1, \ldots, n, j = 1, \ldots, p$. Furthermore, let $\mathbf{Y}^c$ and $\mathbf{X}^c$ define the centered versions of $\mathbf{Y}$ and $\mathbf{X}$, i.e., $\mathbf{Y}^c = (Y_1^c, \ldots, Y_n^c)'$, $\mathbf{X}^c = (X_i^c(t_j))_{ij}$, where $Y_i^c = Y_i - \overline{\mathbf{Y}}$, $X_i^c(t_j) = X_i(t_j) - \overline{\mathbf{X}}_j$, $\overline{\mathbf{Y}} = n^{-1} \sum_{i=1}^n Y_i$, $\overline{\mathbf{X}}_j = n^{-1} \sum_{i=1}^n X_i(t_j)$.

2.2.1 The original procedure of Kneip et al. (2016)

In this section, we briefly describe the estimation and model selection procedures proposed in Kneip et al. (2016). Later, in Section 2.2.2, we describe our adjustments to improve the original procedure and explain why these adjustments result in superior estimation performances.

To estimate the potential PoIs $\widetilde{\tau}_s, s = 1, \ldots, \widetilde{S}$, Kneip et al. (2016) propose a local maxima search (over t_j) based on the sample version $|n^{-1} \sum_{i=1}^n Z_{X_i}(t_j; \delta) Y_i|$ of the cross-moment $|\,\mathbb{E}[Z_{X_i}(t;\delta) Y_i]|$, where $Z_{X_i}(t;\delta) = X_i(t) - (X_i(t-\delta) + X_i(t+$

Algorithm 1 Search Potential Points of Impact Algorithm

procedure SEARCHPOTPOI($\delta \in \mathcal{D} = (0, \delta_{\max}]$, $\mathbf{X} = \mathbf{X}^c$, $\mathbf{Y} = \mathbf{Y}^c$)
  Given δ, define the index $k_\delta \in \mathbb{N}$ such that $1 \leq k_\delta < (p-1)/2 \iff \delta \approx k_\delta/(p-1)$.
  Restrict the set of possible grid indices, i.e., define $\mathcal{J}_{0,\delta} = \{k_\delta + 1, \ldots, p - k_\delta\}$.
  For each index $j \in \mathcal{J}_{0,\delta}$, calculate $Z_{X_i}(t_j; \delta) = X_i(t_j) - \frac{1}{2}(X_i(t_j - \delta) + X_i(t_j + \delta))$.
  while $\mathcal{J}_{s,\delta} \neq \emptyset$, iterate over $s = 1, 2, 3, \ldots$, and **do**
    Determine the index $j_s \in \mathcal{J}_{s-1,\delta}$ of the empirical maximum of $Z_X(t; \delta)Y$, i.e.,
$$j_s = \underset{j \in \mathcal{J}_{s-1,\delta}}{\operatorname{argmax}} \left| \frac{1}{n} \sum_{i=1}^{n} Z_{X_i}(t_j; \delta) Y_i \right|.$$
    Define the s-th potential impact point $\widetilde{\tau}_s = t_{j_s}$ as grid point at index j_s.
    Eliminate all points in an environment of size $\sqrt{\delta}$ around $\widetilde{\tau}_s$, i.e., define
$$\mathcal{J}_{s,\delta} = \{j \in \mathcal{J}_{s-1,\delta} \mid |t_j - \widetilde{\tau}_s| \geq \sqrt{\delta}/2\}.$$
  end while
  return $\widetilde{\mathcal{T}} = \{\widetilde{\tau}_1, \ldots, \widetilde{\tau}_{\widetilde{S}}\}$
end procedure

$\delta))/2$ is the central second-order difference quotient of $X_i(t)$ with $\delta > 0$. The statistic $Z_{X_i}(t; \delta)$ acts as a filter on $X_i(t)$ that uncovers the local-specific variance component of the process $X_i(t)$; see the left plot in Figure 2.1.

The existence of a local-specific variance component in X_i is crucial for the estimation procedure of Kneip et al. (2016) and allows us to show the identifiability of the points of impact and of the model parameters model parameters (see Kneip et al., 2016, Theorem 1). Processes that have a local-specific variance component are typically rough stochastic processes (for instance, Brownian motions, Ornstein-Uhlenbeck processes, etc.), i.e., processes with covariance functions that are sufficiently non-smooth at the diagonal (see Kneip et al., 2016, Theorem 3). Kneip et al. (2016) use a parameter $0 < \kappa < 2$ to quantify the smoothness of the covariance function at the diagonal and propose an estimator $\widehat{\kappa}$ to decide in practice, whether the covariance function is sufficiently non-smooth at the diagonal. The reader is referred to Section 2.4 for an application of this procedure.

The estimation procedure proposed by Kneip et al. (2016) to detect potential PoIs is formally described in Algorithm 1. In each iteration, $s = 1, 2, \ldots$, one PoI is selected by the global maximum of the trajectory of $|n^{-1} \sum_{i=1}^{n} Z_{X_i}(t_j; \delta) Y_i|$ over $j \in \mathcal{J}_{s-1,\delta}$, where $\mathcal{J}_{s-1,\delta} \subset \{1, \ldots, p\}$ denotes an index set defined in Algorithm 1

(see lines 3 and 8). Once a PoI is selected, the algorithm eliminates the grid points within a $\sqrt{\delta}/2$-neighborhood around the selected PoI (see line 8 of Algorithm 1). The algorithm terminates when $\mathcal{J}_{s,\delta}$ is the empty set. The elimination step in line 8 is necessary for providing a consistent estimation procedure.

The selection of the first PoI is shown in the middle plot of Figure 2.1. The first elimination step is shown in the right plot of Figure 2.1, where the second PoI, $\widetilde{\tau}_2$, is determined by the global maximum of the remaining parts of the trajectory of $|n^{-1}\sum_{i=1}^{n} Z_{X_i}(t_j;\delta)Y_i|$ over $j \in \mathcal{J}_{1,\delta}$.

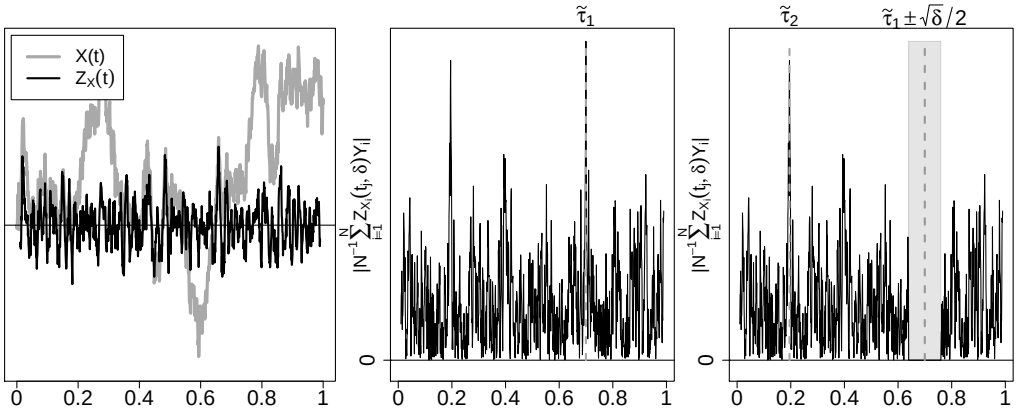

Figure 2.1: LEFT: Trajectories of $X_i(t_j)$ and $Z_{X_i}(t_j;\delta)$, with $\delta = 0.01$. MIDDLE: Trajectory of $|n^{-1}\sum_{i=1}^{n} Z_{X_i}(t_j;\delta)Y_i|$, with first choice $\widetilde{\tau}_1$. RIGHT: Visualization of the second iteration of the `searchPotPoI`-Algorithm.

To estimate the model coefficients for given PoIs $\widehat{\tau}_s$, Kneip et al. (2016) propose an FPCA-based estimation procedure using the approximate model $Y_i \approx \int_0^1 \beta_K(t)X_{i,K}(t)\,\mathrm{d}t + \sum_{s=1}^{\widehat{S}} \beta_s X_i(\widehat{\tau}_s) + \epsilon_i$, where $\beta_K(t) \approx \beta(t)$ and $X_{i,K}(t) \approx X_i(t)$ are K-dimensional approximations based on the first K eigenfunctions of the empirical covariance operator of X_i (see Kneip et al., 2016, Eq. (6.1)). Besides the smoothing parameter, K, one needs to choose a good value of the tuning parameter δ and a subset $\widehat{\mathcal{T}} \subseteq \widetilde{\mathcal{T}}$ of the set of potential PoIs $\widetilde{\mathcal{T}}$ from Algorithm 1. For selecting $\widehat{\mathcal{T}} \subseteq \widetilde{\mathcal{T}}$, Kneip et al. (2016) propose an asymptotic cut-off approach and data-driven Bayesian Information Criterion (BIC)-based approach. In this paper, we focus on the data-driven BIC-based approach as this approach performs clearly better than the asymptotic approach (see Table 1 in Kneip et al., 2016). Moreover, the asymptotic cut-off parameter is hardly applicable in practice as it depends on a generally unknown constant $A > \sqrt{2}$ (see Table 4 in Kneip et al., 2016).

Kneip et al. (2016) propose an infeasible version and a more general feasible version of their data-driven BIC-based procedure to select K, δ, and $\widehat{\mathcal{T}} \subseteq \widetilde{\mathcal{T}}$. The

infeasible strategy is used in their simulation study where the authors perform a BIC-based selection of K and $\widehat{\mathcal{T}}$, and set $\delta = 1/\sqrt{n}$. This naive parametrization of δ is appropriate in their simulation study, but can be arbitrarily bad in practice. The more general strategy is used in the application section of Kneip et al. (2016) where the authors optimize the $\mathrm{BIC}(K, \widehat{\mathcal{T}}, \delta)$ *simultaneously* over K, subsets $\widehat{\mathcal{T}} \subset \widetilde{\mathcal{T}}$, and a fine grid of $\delta \in (0, \delta_{\max}]$. In this paper, we only focus on the latter general model selection strategy since this is practically the most relevant strategy proposed in Kneip et al. (2016) which is not based on unknown constants or naive choices of tuning parameters.

2.2.2 Improving the procedure of Kneip et al. (2016)

In this section, we explain our three main proposals for improving the estimation procedure of Kneip et al. (2016): 1. Sequential model selection and estimation, 2. Smoothing splines estimator, and 3. Standardizations. Later, in Section 2.2.3, we describe the implementation of our estimation algorithm, which builds upon these proposals.

1. Sequential model selection and estimation. Estimating the model parameters in Model (2.1) bears the substantial risk of an omitted-variable-bias since not incorporating the (unknown) true PoI locations τ_s can result in a heavily biased estimator $\widehat{\beta}(t)$ (see the right plot in Figure 2.2 for a noteworthy example). This is a critical issue in practice, and our simulation results show that the original estimation procedure of Kneip et al. (2016) may suffer severely from such biases.

The underlying problem is that the selection of the number $\widehat{S}$ of PoIs and their locations $\widetilde{\tau}_1, \ldots, \widetilde{\tau}_{\widehat{S}}$ and the selection of the smoothing parameter, K, for estimating $\beta(\cdot)$ are two ambiguous selection problems. It is easy to trade model complexities between the empirical PoI model component and the empirical functional model component without affecting the model fit. This results in a quite delicate model selection problem, which generally leads to unstable estimates when trying to solve both selection problems simultaneously as suggested in Kneip et al. (2016).

Let us explain the reason for this instability by considering the following two extreme situations—both approximating the regression Model (2.1):

- Let $K \gg 0$ and $\widehat{S} = 0$. For very large K, the estimator $\widehat{\beta}_K(t)$ is flexible enough, such that

$$\int_0^1 \widehat{\beta}_K(t) X_{i,K}(t)\,\mathrm{d}t \approx \int_0^1 \beta(t) X_i(t)\,\mathrm{d}t + \sum_{s=1}^{S} \beta_s X_i(\tau_s).$$

In this case, $\widehat{\beta}_K(t)$ approximates $\beta(t)$, except at the points of impact locations $t = \tau_s$, where $\widehat{\beta}_K(t)$ approximates $\beta_s X_i(\tau_s)$, i.e., where $\int_{\tau_s-h}^{\tau_s+h} \widehat{\beta}_K(t) X_{i,K}(t)\,\mathrm{d}t \approx \beta_s X_i(\tau_s)$ with, e.g., $h = 0.01$ (see the right plot in Figure 2.2 for examples of such estimates $\widehat{\beta}_K(t)$).

- Let $K = 0$ and $\widehat{S} \gg 0$. A large set of PoI candidates $X_i(\widehat{\tau}_1), \ldots, X_i(\widehat{\tau}_{\widehat{S}})$ leads to a very flexible linear model, such that

$$\sum_{s=1}^{\widehat{S}} \widehat{\beta}_s X_i(\widehat{\tau}_s) \approx \int_0^1 \beta(t) X_i(t)\,\mathrm{d}t + \sum_{s=1}^{S} \beta_s X_i(\tau_s).$$

  In this case, $\sum_{s=1}^{\widehat{S}} \widehat{\beta}_s X_i(\widehat{\tau}_s)$ acts like a Riemann sum for approximating the integral $\int_0^1 \beta(t) X_i(t)\,\mathrm{d}t$, except for the $\widehat{\beta}_s$-values at $\widehat{\tau}_s \approx \tau_s$, where $\widehat{\beta}_s X_i(\widehat{\tau}_s) \approx \beta_s X_i(\tau_s)$.

These two extreme situations demonstrate that there is a certain ambiguity between the model selection parameters K and $\widehat{S} = |\widehat{\mathcal{T}}(\delta)|$ that allows for shifting the model-complexities between the integral-part and the PoI-part of the empirical model. This ambiguity generally leads to unstable model selections when optimizing $\mathrm{BIC}(K, \widehat{\mathcal{T}}, \delta)$ *simultaneously* over K, subsets $\widehat{\mathcal{T}} \subset \widetilde{\mathcal{T}}$, and δ—as proposed in Kneip et al. (2016). As a consequence, one gets unstable estimates of $\beta(\cdot)$ caused by omitted-variable biases in $\widehat{\beta}(\cdot)$, as shown in the right plot in Figure 2.2.

To stabilize the model selection procedure we propose a sequential selection and estimation procedure (see Section 2.2.3). In the first ("Pre-select") step, our procedure pre-selects all potential points of impact $\widetilde{\mathcal{T}} = \{\widetilde{\tau}_1, \ldots, \widetilde{\tau}_{\widetilde{S}}\}$ while ignoring the estimation of the functional parameter $\beta(\cdot)$. In the second ("Estimate") step, our procedure estimates the model parameters, $\beta(\cdot)$ and β_s, given the pre-selected points of impact.

The theoretical justification for this sequential approach is given by the following result that holds under the assumptions of Kneip et al. (2016) and implies that the points of impact can be estimated consistently without knowledge (or pre-estimation) of the slope function $\beta(\cdot)$ (see Lemmas 3 and 4 in the supplementary paper of Kneip et al., 2016):

$$\begin{aligned} |\mathbb{E}(Z_{\delta,i}(t) Y_i)| &= \beta_r c(\tau_s) \delta^\kappa + o(\delta^\kappa) \quad \text{if } t_j = \tau_s \text{ for some } s = 1, \ldots, S \\ |\mathbb{E}(Z_{\delta,i}(t) Y_i)| &= O(\delta^2) \qquad\qquad\qquad\quad \text{if } t \notin \{\tau_1, \ldots, \tau_S\} \end{aligned} \tag{2.2}$$

as $\delta \to 0$, where $0 < \kappa < 2$ and $0 < c(\tau_r) < \infty$ are constants specific to the considered process X_i.

That is, the trajectory of $|\mathbb{E}(Z_{\delta,i}(t_j))Y_i)|$, $j = 1, \ldots, p$, will have peaks at grid points $t_j \approx \tau_r$, even without knowledge (or pre-estimation) of the slope function $\beta(\cdot)$. Consequently, Step 1 ("Pre-select") of our algorithm (Section 2.2.3) leads to a consistent point of impact selection if, for instance, $\delta^\kappa \sim n^{-1}$, since $|\mathbb{E}(Z_{\delta,i}(t_j))Y_i)|$ can be consistently estimated using its empirical counterpart $|n^{-1}\sum_{i=1}^n Z_{\delta,i}(t_j))Y_i|$ for all $j = 1, \ldots, p$ as $n \to \infty$.

Using the consistently pre-selected points of impact in Step 2 ("Estimate") leads to a more stable estimation of the model parameters, $\beta(\cdot)$ and β_s, as it avoids a simultaneous selection of the PoIs and the smoothing parameters. The further Steps 3-5 (see the overview in Section 2.2.3) of our selection and estimation algorithm contain repetitions of the selection and estimation steps (Step 1 and Step 2). These repetitions are asymptotically irrelevant, but further improve the estimation results in practice (see Section 2.3).

2. Smoothing splines estimator. Deviating from Kneip et al. (2016), we propose using a penalized smoothing splines estimator. The FPCA-based estimator, proposed by Kneip et al. (2016), is only optimal under the restrictive assumption of a structural link between the functional regression parameter, $\beta(\cdot)$, and the functional regressor, X (see Assumptions (3.1-3.3) in Hall and Horowitz, 2007). However, this link does not necessarily hold in applications and also cannot be tested in practice.[2] Therefore, we propose using the penalized smoothing splines estimator of Crambes et al. (2009). While this estimator achieves minimax optimal rates under similar structural link assumptions (see Crambes et al. , 2009), it is also known to perform well if these structural assumptions do not hold since the spline basis system has some very general approximation properties (see, for instance, De Boor, 1978, and Crambes et al., 2009).

3. Standardizations. The standardization of the curves in Step 1 ("Pre-select") and Step 3 ("Sub-select") of our algorithm scales the trajectories of the process Z_δ by the inverse of the pointwise standard deviation of the process X. From an asymptotic perspective, this is irrelevant, since this scaling only leads to different constants $c(\tau_r)$ in (2.2). However, standardization of the data is a typical preprocessing step in model selection problems leading to more homogenous signals which further improves the selection results in practice (see Section 2.3).

2.2.3 The PES-ES estimation algorithm

Our estimation algorithm is built up from the following three Pre-select, Estimate, and Sub-select (PES) steps:

[2] Remember that the FPCA-basis, based on the eigendecomposition of the covariance operator of X, is the optimal empirical basis to approximate X, but generally not the optimal basis to approximate $\beta(\cdot)$.

1. **Pre-select:** Pre-select potential PoIs $\widetilde{\mathcal{T}} = \{\widetilde{\tau}_1, \ldots, \widetilde{\tau}_{\widetilde{S}}\}$. (See Section 2.2.3)
2. **Estimate:** Estimate the function- and scalar-valued slope parameters $\beta, \beta_1, \ldots, \beta_{\widetilde{S}}$ given the set of potential PoIs $\widetilde{\mathcal{T}}$. (See Section 2.2.3)
3. **Sub-select:** Sub-select PoIs from the set of potential PoIs $\widetilde{\mathcal{T}}$. (See Section 2.2.3)

Typically, the estimation step (Step 2) leads to inefficient estimators $\widehat{\beta}(\cdot)$, but avoids omitted-variable biases. Inefficient, because $\widetilde{\mathcal{T}}$ tends to contain many redundant PoI locations ($\widetilde{S} > S$), which reduces the number of degrees of freedom. We reduce the risk of omitted-variable biases because the large set of potential PoIs $\widetilde{\mathcal{T}}$ has a high likelihood of containing the true PoI locations (as explained in more detail in Section 2.2.2). Our final PES-ES algorithm, described in Section 2.2.3, uses a repetition of the latter two Estimate-Sub-select (ES) steps, which can result in a further improvement of the estimation results (see Section 2.3).

Pre-Select PoIs

To select potential PoIs, we use Algorithm 1 with the difference that instead of using centered observations of the functions $\mathbf{X}^c$, we use the pointwise standardized curves $\mathbf{X}^{st}$ as input of the algorithm, where $X_i^{st}(t_j) = X_i^c(t_j)/\mathrm{sd}(\mathbf{X}_j)$ and $\mathrm{sd}(\mathbf{X}_j) = (n^{-1}\sum_{i=1}^n (X_i(t_j) - \overline{\mathbf{X}}_j)^2)^{1/2}$. As described in Section 2.2.2, this is irrelevant from an asymptotic point of view, but typically stabilizes and improves the PoI selection in practice.

Estimate Slope Parameters

To estimate the slope parameters—given the pre-selected PoIs $\widetilde{\mathcal{T}}$—we adapt the penalized smoothing splines estimator proposed by Crambes et al. (2009) in order to incorporate PoIs. Let us initially recap the situation of Model (2.1) without PoIs ($S = 0$, $\mathcal{T} = \emptyset$), as considered by Crambes et al. (2009). Their estimator of $\beta(\cdot)$, evaluated at the grid points $t_1, \ldots, t_p$, is given by

$$\left(\widehat{\beta}^{\rho}(t_1), \ldots, \widehat{\beta}^{\rho}(t_p)\right) = \frac{1}{n}\left(\frac{1}{np}\mathbf{X}^{c\prime}\mathbf{X}^c + \rho\mathbf{A}\right)^{-1}\mathbf{X}^{c\prime}\mathbf{Y}^c, \tag{2.3}$$

where the penalty matrix $\mathbf{A} = \mathbf{P} + p\mathbf{A}^\star$ is composed of a non-classical projection matrix $\mathbf{P}$ and a classical regularization matrix $\mathbf{A}^\star$. The non-classical $p \times p$ projection matrix $\mathbf{P} = \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'$, with $\mathbf{W} = (t_j^l)_{j,l} \in \mathbb{R}^{p\times m}$ is introduced by Crambes et al. (2009) in order to guarantee uniqueness of their estimator, where t_j^l denotes the lth power of the grid point t_j with $j = 1, \ldots, p$ and $l = 0, \ldots, m-1$.

Following the usual convention, we set $m = 2$, which results in the classical choice of *cubic* splines. The classical $p \times p$ regularization matrix $\mathbf{A}^\star$ is defined as

$$\mathbf{A}^\star = \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\left(\int_0^1 \mathbf{b}^{(2)}(t)\mathbf{b}^{(2)}(t)'\,\mathrm{d}t\right)(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}',$$

where $\mathbf{b}(t) = (b_1(t), \ldots, b_p(t))'$ are natural cubic spline basis functions, $\mathbf{b}^{(2)}(t)$ denotes their second derivatives, and $\mathbf{B}$ is a $p \times p$ matrix with elements $b_i(t_j), i, j = 1, \ldots, p$. For the implementation of the natural cubic spline basis functions, we use the `ns`-function contained in the R-package `splines`.

In order to incorporate the pre-selected PoIs, we need to extend the matrices $\mathbf{X}^c$ and $\mathbf{A}$. The extended data matrix is given by $\mathbf{X}^c_{\widetilde{\mathcal{T}}} = (\mathbf{X}^c, p\mathbf{X}^c(\widetilde{\tau}_1), \ldots, p\mathbf{X}^c(\widetilde{\tau}_{\widetilde{S}})) \in \mathbb{R}^{n\times(p+\widetilde{S})}$, where $\mathbf{X}^c(\widetilde{\tau}_s) = (X_1^c(\widetilde{\tau}_s), \ldots, X_n^c(\widetilde{\tau}_s))' \in \mathbb{R}^n$. The extended penalty matrix is given by

$$\mathbf{A}_{\widetilde{\mathcal{T}}} = \begin{pmatrix} \mathbf{A} & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{(p+\widetilde{S})\times(p+\widetilde{S})},$$

where all entries with respect to the PoIs are zero (see Goldsmith et al., 2010, for an equivalent extension of the penalty matrix). The augmented estimator of $\beta(t_1), \ldots, \beta(t_p)$ and $\beta_1, \ldots, \beta_S$,

$$\widehat{\beta}^\rho_{\widetilde{\mathcal{T}}} = \left(\widehat{\beta}^\rho_{\widetilde{\mathcal{T}}}(t_1), \ldots, \widehat{\beta}^\rho_{\widetilde{\mathcal{T}}}(t_p), \widehat{\beta}^\rho_{\widetilde{\mathcal{T}},1}, \ldots, \widehat{\beta}^\rho_{\widetilde{\mathcal{T}},\widetilde{S}}\right) = \frac{1}{n}\left(\frac{1}{np}\mathbf{X}^{c\prime}_{\widetilde{\mathcal{T}}}\mathbf{X}^c_{\widetilde{\mathcal{T}}} + \rho\mathbf{A}_{\widetilde{\mathcal{T}}}\right)^{-1}\mathbf{X}^{c\prime}_{\widetilde{\mathcal{T}}}\mathbf{Y}^c, \tag{2.4}$$

depends on the included set of PoIs $\widetilde{\mathcal{T}}$ and on the smoothing parameter ρ. In order to determine an optimal smoothing parameter, we use the following Generalized Cross-Validation (GCV) criterion, as proposed by Crambes et al. (2009):

$$\mathrm{GCV}(\rho) = \frac{\frac{1}{n}\mathrm{RSS}(\widehat{\beta}^\rho_{\widetilde{\mathcal{T}}})}{\left(1 - \frac{1}{n}\mathrm{Tr}(\mathbf{H}^c_{\rho,\widetilde{\mathcal{T}}})\right)^2}. \tag{2.5}$$

Here, the Residual Sum of Squares (RSS) is defined as $\mathrm{RSS}(\widehat{\beta}^\rho_{\widetilde{\mathcal{T}}}) = ||\mathbf{Y}^c - \mathbf{H}^c_{\rho,\widetilde{\mathcal{T}}}\mathbf{Y}^c||^2$, where $||.||$ denotes the Euclidean norm, and the smoother matrix $\mathbf{H}^c_{\rho,\widetilde{\mathcal{T}}}$ is defined as $\mathbf{H}^c_{\rho,\widetilde{\mathcal{T}}} = (np)^{-1}\mathbf{X}^c_{\widetilde{\mathcal{T}}}((np)^{-1}\mathbf{X}^{c\prime}_{\widetilde{\mathcal{T}}}\mathbf{X}^c_{\widetilde{\mathcal{T}}} + \rho\mathbf{A}_{\widetilde{\mathcal{T}}})^{-1}\mathbf{X}^{c\prime}_{\widetilde{\mathcal{T}}}$. Our final estimator for the slope parameters is given by the GCV-optimized version of (2.4),

$$\widehat{\beta}_{\widetilde{\mathcal{T}}} = \left(\widehat{\beta}_{\widetilde{\mathcal{T}}}(t), \widehat{\beta}_{\widetilde{\mathcal{T}},1}, \ldots, \widehat{\beta}_{\widetilde{\mathcal{T}},\widetilde{S}}\right) = \left(\widehat{\beta}^{\rho_{\mathrm{GCV}}}_{\widetilde{\mathcal{T}}}(t), \widehat{\beta}^{\rho_{\mathrm{GCV}}}_{\widetilde{\mathcal{T}},1}, \ldots, \widehat{\beta}^{\rho_{\mathrm{GCV}}}_{\widetilde{\mathcal{T}},\widetilde{S}}\right), \quad t \in \{t_1, \ldots, t_p\}, \tag{2.6}$$

where $\rho_{\text{GCV}} = \text{argmin}_{\rho \in (0, \rho_{\max}]} \text{GCV}(\rho)$.

Sub-Select PoIs

This part of our estimation algorithm is aimed at selecting the true PoIs from the pre-selected set of potential PoIs $\widetilde{\mathcal{T}} = \widetilde{\mathcal{T}}(\delta)$ given the estimate $\widehat{\beta}_{\widetilde{\mathcal{T}}}$ in (2.6). This sub-selection is performed by minimizing the following BIC over subsets $\mathcal{R} \subseteq \widetilde{\mathcal{T}}(\delta)$:

$$\widehat{\mathcal{T}} = \underset{\mathcal{R} \subseteq \widetilde{\mathcal{T}}(\delta)}{\text{argmin}} \, \text{BIC}(\mathcal{R}), \quad \text{where}$$

$$\text{BIC}(\mathcal{R}) = n \log\left(\frac{\text{RSS}(\mathcal{R})}{n}\right) + \log(n) \cdot S_{\mathcal{R}}, \quad \text{with} \quad S_{\mathcal{R}} = |\mathcal{R}|. \tag{2.7}$$

Here, $\text{RSS}(\mathcal{R})$ is made up of the residuals from regressing the $\widehat{\beta}_{\widetilde{\mathcal{T}}}$-neutralized $Y_{i,\widehat{\beta}_{\widetilde{\mathcal{T}}}} = Y_i^c - \int_0^1 \widehat{\beta}_{\widetilde{\mathcal{T}}}(t) X_i^c(t) \, \mathrm{d}t$ onto $X_i^{st}(\widetilde{\tau}_s), \ldots, X_i^{st}(\widetilde{\tau}_{S_{\mathcal{R}}})$, with $\{\widetilde{\tau}_1, \ldots, \widetilde{\tau}_{S_{\mathcal{R}}}\} = \mathcal{R}$, where $\widehat{\beta}_{\widetilde{\mathcal{T}}}(t)$ is the estimate of $\beta(\cdot)$ defined as the first element in the vector of estimates (2.6).

For optimizing $\text{BIC}(\mathcal{R})$ over $\mathcal{R} \subseteq \widetilde{\mathcal{T}}(\delta)$, we use a directed search strategy taking into account the information content in $\widetilde{\mathcal{T}} = \{\widetilde{\tau}_1, \ldots, \widetilde{\tau}_{\widetilde{S}}\}$. By construction, the order of the PoI locations $\widetilde{\tau}_1, \ldots, \widetilde{\tau}_{\widetilde{S}}$ reflects a decreasing signal-to-noise ratio and, therefore, a decreasing quality of the estimates. This suggests minimizing $\text{BIC}(\mathcal{R})$ using a directed search strategy where $\text{BIC}(\mathcal{R})$ is evaluated consecutively at the sets $\mathcal{R} = \{\widetilde{\tau}_1\}$, $\mathcal{R} = \{\widetilde{\tau}_1, \widetilde{\tau}_2\}, \ldots, \mathcal{R} = \{\widetilde{\tau}_1, \ldots, \widetilde{\tau}_{\widetilde{S}}\}$.

The full PES-ES estimator

Our estimation algorithm, PES-ES, consists of the above described Pre-select-Estimate-Sub-select (PES) steps and uses a repetition of the latter two Estimate-Sub-select (ES) steps:

1. **P**re-Select	$\widetilde{\mathcal{T}} = \widetilde{\mathcal{T}}(\delta)$	(Section 2.2.3)
2. **E**stimate	$\widehat{\beta}_{\widetilde{\mathcal{T}}}$	(Section 2.2.3)
3. **S**ub-Select	$\widehat{\mathcal{T}} \subseteq \widetilde{\mathcal{T}}$	(Section 2.2.3)
4. re**E**stimate	$\widehat{\beta}_{\widehat{\mathcal{T}}}$	(Section 2.2.3, with $\widetilde{\mathcal{T}}$ replaced by $\widehat{\mathcal{T}}$)
5. re**S**ub-Select	$\widehat{\mathcal{T}}_{\text{re}} \subseteq \widehat{\mathcal{T}}$	(Section 2.2.3, with $\widetilde{\mathcal{T}}$ replaced by $\widehat{\mathcal{T}}$)

Note that the entire PES-ES algorithm depends on the initially pre-selected set of potential PoIs $\widetilde{\mathcal{T}}(\delta)$ and, therefore, on the choice of δ. In the following, we write $\widehat{\mathcal{T}}_{\text{re}}(\delta)$ in order to emphasize this entire dependency on δ. We follow Kneip et al. (2016) and determine an optimal δ by minimizing the BIC. For each δ-value on a fine grid in $(0, \delta_{\max}]$, we run the entire PES-ES algorithm and select

the optimal δ by,

$$\delta_{\text{BIC}} = \underset{\delta \in (0, \delta_{\max}]}{\text{argmin}} \ \text{BIC}(\delta), \quad \text{with}$$

$$\text{BIC}(\delta) = n \log \left(\frac{\text{RSS}(\widehat{\mathcal{T}}_{\text{re}}(\delta))}{n} \right) + \log(n) \cdot \text{edf}(\widehat{\mathcal{T}}_{\text{re}}(\delta)), \tag{2.8}$$

where $\text{RSS}(\widehat{\mathcal{T}}_{\text{re}}(\delta)) = ||\mathbf{Y}^c - \mathbf{H}^c_{\rho_{\text{GCV}}, \widehat{\mathcal{T}}_{\text{re}}(\delta)} \mathbf{Y}^c||^2$ with smoother matrix $\mathbf{H}^c_{\rho_{\text{GCV}}, \widehat{\mathcal{T}}_{\text{re}}(\delta)}$ defined as $\mathbf{H}^c_{\rho_{\text{GCV}}, \widehat{\mathcal{T}}_{\text{re}}(\delta)} = (np)^{-1} \mathbf{X}^c_{\widehat{\mathcal{T}}_{\text{re}}(\delta)} ((np)^{-1} \mathbf{X}^{c\prime}_{\widehat{\mathcal{T}}_{\text{re}}(\delta)} \mathbf{X}^c_{\widehat{\mathcal{T}}_{\text{re}}(\delta)} + \rho_{\text{GCV}} \mathbf{A})^{-1} \mathbf{X}^{c\prime}_{\widehat{\mathcal{T}}_{\text{re}}(\delta)}$ and effective degrees of freedom $\text{edf}(\widehat{\mathcal{T}}_{\text{re}}(\delta)) = \text{Tr}(\mathbf{H}^{c\prime}_{\rho_{\text{GCV}}, \widehat{\mathcal{T}}_{\text{re}}(\delta)} \mathbf{H}^c_{\rho_{\text{GCV}}, \widehat{\mathcal{T}}_{\text{re}}(\delta)})$; see Hastie and Tibshirani (1990), Ch. 3.5 for an overview of possible definitions of edf.

2.3 Simulations

In the following simulation study, we assess the finite sample properties of our PES-ES algorithm. The original estimation procedure proposed by Kneip, Poss, and Sarda (2016), abbreviated as KPS, serves as our main benchmark, and its implementation is described in Section 2.2.1. The smoothing splines estimator (2.3) by Crambes, Kneip, and Sarda (2009), abbreviated hereafter as CKS, serves as a challenging benchmark for our NoPoI data generating process (i.e., a functional linear regression model *without* points of impact). Section 2.3.1 introduces the considered data generating processes and presents our simulation results.

We aim to provide an in-depth assessment of our PES-ES estimation algorithm. Therefore, in order to assess the improvements that are due to the final ES (Estimation and Subselection) step, we compare the PES-ES results with those of the reduced PES estimation algorithm without the final ES step. We also show that an additional second repetition of the ES step (PES-2ES) does not lead to a significant improvement of our estimation algorithm.

Kneip et al. (2016) arbitrarily set $K_{\max} = 6$, which is, however, too small for our simulation study where $K_{\max} = 6$ often becomes a binding upper optimization threshold. The choice of $K_{\max}$ is crucial since it constrains the magnitude of possible omitted-variable biases in $\widehat{\beta}_K(t)$. The same issue applies to $\rho_{\min}$ when optimizing the GCV in (2.5) over $\rho \in [\rho_{\min}, \rho_{\max}]$ with $\rho_{\min} \approx 0$. Therefore, we choose very conservative optimization intervals $[K_{\min}, K_{\max}] = [1, 150]$ and $[\rho_{\min}, \rho_{\max}] = [10^{-6}, 200]$.

2.3.1 Data Generating Processes and Simulation Results

We consider five different Data Generating Processes (DGPs), as described in Table 2.1. The DGPs Easy and Complicated represent a simple and a more complex version of Model (2.1). The Complicated DGP is challenging due to the closeness of the PoI locations τ_1 and τ_2, which may trigger omitted-variable biases in $\widehat{\beta}(t)$ when omitting either τ_1 or τ_2. The two further DGPs NoPoI ($\mathcal{T} = \emptyset$) and OnlyPoI ($\beta(t) \equiv 0$) are used to check the robustness of our PES-ES algorithm against model-misspecifications.

Table 2.1: Data Generating Processes.

DGP	$\beta(t)$	S	$\mathcal{T} = \{\tau_1, \ldots, \tau_S\}$	$\{\beta_1, \ldots, \beta_S\}$
Easy	$\beta(t) = -(t-1)^2 + 2$	2	$\{0.3, 0.6\}$	$\{-3, 3\}$
Complicated	$\beta(t) = -5(t-0.5)^3 - t + 1$	3	$\{0.3, 0.4, 0.6\}$	$\{-3, 3, 3\}$
OnlyPoI	$\beta(t) \equiv 0$	2	$\{0.3, 0.6\}$	$\{-3, 3\}$
NoPoI	$\beta(t) = -(t-1)^2 + 2$	0	$\emptyset$	$\emptyset$

For each DGP and two sample sizes ($n = 250$ and $n = 500$), we generate 1000 replications of n functions $X_i(t)$ observed at $p = 300$ equidistant points $t_1, \ldots, t_p$ in $[0, 1]$. In Appendix 2.A we additionally present simulation results for $p = 500$. The functions $X_i(t)$ are standard Brownian Motions, and the dependent variables Y_i are generated according to Model (2.1) with $\epsilon_i \sim N(0, 0.125^2)$. Our simulation is implemented using the statistical language R (R Core Team, 2017a), and the R-codes for reproducing the simulation results are part of the online supplement supporting this article.

The upper panel of Table 2.2 reports the integrated squared bias and the integrated variance for the estimator $\widehat{\beta}(t)$ of $\beta(t)$. The integrated squared bias is computed as $\int_0^1 (\bar{\beta}(t) - \beta(t))^2 \, \mathrm{d}t$, where $\bar{\beta}(t) = 1000^{-1} \sum_{r=1}^{1000} \widehat{\beta}_r(t)$ is the mean of the estimates over all replications. The integrated variance is computed as $1000^{-1} \int_0^1 \sum_{r=1}^{1000} (\widehat{\beta}_r(t) - \bar{\beta}(t))^2 \, \mathrm{d}t$. The lower panel of Table 2.2 reports the average squared bias $S^{-1} \sum_{s=1}^{S} (\bar{\beta}_s - \beta_s)^2$, with $\bar{\beta}_s = 1000^{-1} \sum_{r=1}^{1000} \widehat{\beta}_{r,s}$, and the average variance $S^{-1} \sum_{s=1}^{S} 1000^{-1} \sum_{r=1}^{1000} (\widehat{\beta}_{r,s} - \bar{\beta}_s)^2$ for the PoI coefficient estimators $\widehat{\beta}_s$, conditionally on the event that τ_s was correctly found[3], where a single τ_s is considered to be found correctly if $|\widehat{\tau}_s - \tau_s| < 0.01$. The latter requirement corresponds to an estimation precision of only ± 3 grid points, which is substantially more challenging than the matching requirement originally used in Kneip et al. (2016). The shades of gray in Table 2.2 show the ranking of the mean squared error (MSE); the lowest/highest MSE (=Bias^2 + Var.) has the darkest/lightest gray-scale.

[3]Note that it is impossible to compute estimation errors for non-found τ_s.

Table 2.2: Squared bias and variance of the estimators. Shades of gray show the ranking of the Mean Squared Error (MSE): lowest/highest MSE has the darkest/lightest gray-scale.

		Easy		Complicated		NoPoI		OnlyPoI	
$\int \widehat{\beta}(t)$		Bias^2	Var.	Bias^2	Var.	Bias^2	Var.	Bias^2	Var.
$n = 250$	PES	0.02	0.22	0.21	1.98	0.00	0.02	0.00	0.08
	PES-ES	0.02	0.24	0.16	1.68	0.00	0.01	0.00	0.06
	PES-2ES	0.02	0.25	0.16	1.66	0.00	0.01	0.00	0.06
	KPS	2.81	51.17	155.17	303.03	0.01	0.02	0.05	6.65
	CKS	-	-	-	-	0.00	0.00	-	-
$n = 500$	PES	0.01	0.06	0.05	0.55	0.00	0.01	0.00	0.01
	PES-ES	0.00	0.05	0.04	0.38	0.00	0.01	0.00	0.01
	PES-2ES	0.00	0.05	0.04	0.38	0.00	0.01	0.00	0.01
	KPS	0.35	16.69	91.32	245.88	0.01	0.01	0.00	0.5
	CKS	-	-	-	-	0.00	0.00	-	-
$\frac{1}{S}\sum \widehat{\beta}_s$									
$n = 250$	PES	0.01	0.1	0.01	0.09	-	-	0.00	0.02
	PES-ES	0.00	0.08	0.01	0.06	-	-	0.00	0.02
	PES-2ES	0.00	0.08	0.01	0.06	-	-	0.00	0.02
	KPS	0.03	0.54	1.59	4	-	-	0.00	0.06
$n = 500$	PES	0.00	0.02	0.00	0.01	-	-	0.00	0.02
	PES-ES	0.00	0.01	0.00	0.01	-	-	0.00	0.00
	PES-2ES	0.00	0.01	0.00	0.01	-	-	0.00	0.00
	KPS	0.01	0.2	0.78	2.92	-	-	0.00	0.01

The simulation results for the slope parameters $\beta(t)$ and $\beta_1, \ldots, \beta_S$ in the upper and lower panel of Table 2.2 show that the smoothing-spline-based estimation algorithms PES and PES-ES clearly outperform the FPCA-based KPS estimator. The final ES-step in the PES-ES algorithm aims to remove further falsely selected point of impact candidates. This is advantageous in all DGPs, except for the Easy DGP with $n = 250$ and the NoPoI DGP, where PES-ES and PES achieve essentially equivalent results. Note that the final ES-step is particularly advantageous for the Complicated DGP and the smaller sample size $n = 250$, where KPS shows a very poor performance. Only in this particular case,

Table 2.3: Percentage of replications with correct detection of all points of impact $\tau_1, \ldots, \tau_S$.

		300 grid points			500 grid points		
		Easy	Compl.	OnlyPoI	Easy	Compl.	OnlyPoI
$n = 250$	PES	97.5	77.4	99	97	83.8	99.1
	PES-ES	97.6	79.3	99.2	97.3	85.3	99.2
	PES-2ES	97.6	79.3	99.2	97.3	85.8	99.2
	KPS	89.7	19.3	98.7	89.5	24	98.5
$n = 500$	PES	99.3	94.6	99.9	99.3	94	99.9
	PES-ES	99.4	95.7	99.9	99.2	95.3	99.9
	PES-2ES	99.4	95.8	99.9	99.2	95.3	99.9
	KPS	96.9	37.2	100	97	41.9	99.5

one additional second repetition of the ES-step (PES-2ES) further reduces the variance. This improvement, however, is not substantial and does not justify the additionally involved computational burden of PES-2ES. PES-ES also performs very well in the NoPoI and the OnlyPoI DGPs, where PES-ES is actually a misspecified estimation procedure. In the case of NoPoI, PES-ES performs almost as well as the corresponding (minimax-optimal) benchmark-estimator CKS, and in the case of OnlyPoI, PES-ES is the best performing method.

Table 2.3 reports for each estimator and sample size the percentage of replications where all PoI locations $\tau_1, \ldots, \tau_S$ are found correctly. The left part of the table contains the results for functions observed on $p = 300$ grid points and the right part for $p = 500$ grid points. PES-ES and PES-2ES outperform all competitors, except in the case of OnlyPoI with $n = 500$, where all estimation procedures show essentially the same performance. Again, the difference between PES(-(2)ES) and KPS is particularly large for the smaller sample size $n = 250$ and the Complicated DGP. Increasing the resolution of the grid from $p = 300$ to $p = 500$ does not change the results. Similarly, the increased resolution also does not affect the precision of the estimate for the slope parameter $\beta(\cdot)$ and $\beta_1, \ldots, \beta_s$, see Table 2.A.2 in Appendix 2.A.

To show the performance boost of using standardized data for locating the potential PoIs (as described at the end of Section 2.2.3), we report the simulation results without standardizing the data (see Tables 2.A.1 and 2.A.3 in Appendix 2.A). The results show that the standardization of the data is beneficial for the Complicated DGP. Table 2.A.4 in Appendix 2.A shows the simulation results for the Complicated DGP, but with different noise-to-signal ratios, that is, with

different values for the error variance in model (2.1). PES(-ES) still outperforms KPS significantly; however, it turns out that the difference becomes less pronounced as the noise-to-signal ratio increases.

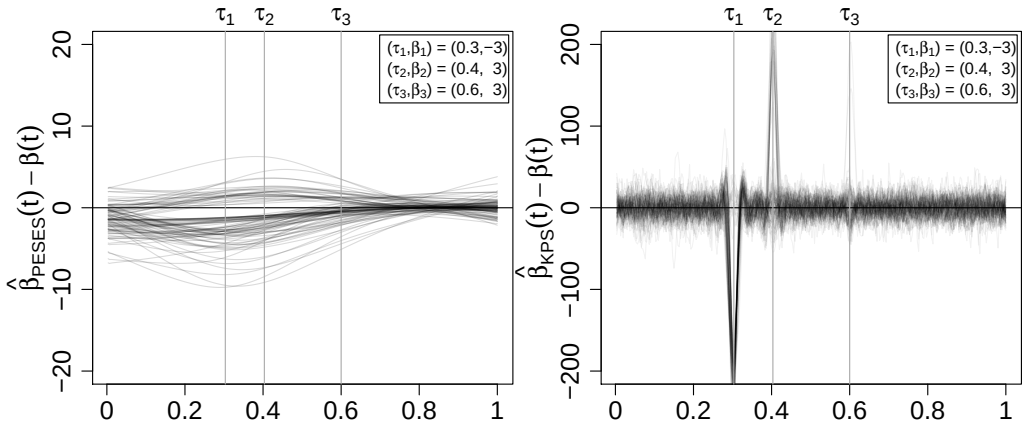

Figure 2.2: Pointwise deviations $\widehat{\beta}(t) - \beta(t)$ of the 10% largest L^2 distances $\int_0^1 (\widehat{\beta}(t) - \beta(t))^2 \, \mathrm{d}t$ for the Complicated DGP. Note that the scales of the two y-axes differ by a factor of 10.

2.4 Application

To illustrate the practical importance of the functional linear regression model with points of impact, we present an application to data from Google AdWords, which is the most popular online advertising platform and of fundamental importance for Alphabet's (Google's parent company) economic success (in 2014, 90 percent of Alphabet's sales came from AdWords). Online advertising, in turn, is the most important branch of today's advertising industry, with an expected U.S. revenue of 60 billion USD in 2016 (Doty et al., 2021). The case study described below is motivated by the needs of *Crealytics*, the company that generously provided the data. Today this company uses the described method—with some further confidential enhancements—to support their daily business.

The main pricing mechanism at Google AdWords is the so-called Pay-Per-Click (PPC) mechanism. Here, advertisers (e.g., an online outdoor shop in our application) can bid for a sponsored "impression" to be displayed along with Google's search results when a user conducts a search query related to a specific keyword (e.g., `outdoor jacket`)[4]. The basic building block of an online ad

[4]Sponsored impressions link to the advertised homepage—they are similar to, but distinguishable from ordinary Google search results.

campaign is a text corpus of (hundreds, thousands, or ten-thousands of, etc.) keywords related to the advertised products.

The limited number of sponsored impressions is allocated by an auction. Advertisers whose impression appears on the display are chosen according to their ad-rank, which is basically their original bid, i.e., the maximum "costs-per-click" an advertiser is willing to pay times the quality score, a discrete metric (from 1, the lowest, to 10, the best) determining the relevance of an advertiser's impression. Google AdWords auctions are time continuous and an advertiser only pays if a user clicks on the displayed impression. (See Geddes, 2014, for an in-depth introduction to Google AdWords.)

The bidding process is usually based on bidding software that evaluate specific key-figures. One of the most important key-figures is the so-called Click-Though Rate (CTR), which is defined as the daily number of clicks per impression. The CTR estimates the current probability of receiving a click on a sponsored impression and therefore plays an important role in assisting the bidding process on a short-term basis (Geddes, 2014).

The economic success of ad campaigns, however, also depends on long-sighted bidding strategies taking into account product specific (time-global) seasonalities as well as (time-local) events, such as the importance of Valentine's Day for an online flower shop. Unfortunately, existing key-figures such as the CTR only provide a daily perspective and are not suitable for assisting in the implementation of long-sighted bidding strategies. Therefore, the functional linear regression model with points of impact is a suitable methodology to identify the (global and local) functional relationship between the *yearly* clicks and the *yearly* trajectories of daily impressions—leading to a long-sighted version of the CTR.

As a yearly measure of clicks, we use the logarithmized yearly sums of clicks, i.e., $Y_i = \log(C_i)$ with $C_i := \sum_{t=1}^{365} \texttt{clicks}_{it}$, where i indexes the ith keyword of the considered ad campaign. As a yearly measure of impressions, we use the yearly trajectories of daily logarithmized numbers of impressions, i.e., $X_i(t) = \log(\mathcal{I}_i(t))$ with $\mathcal{I}_i(t) := \texttt{impressions}_{it}$, where $t = 1, \ldots, 365$ indexes the days of the considered year. Our application uses data from a real Google AdWords campaign run by an online store selling outdoor equipment in the year from April 1st, 2012 to March 31st, 2013. The left plot in Figure 2.3 shows all trajectories $X_i(t)$ of the considered ad campaign. The middle plot shows three exemplary (logarithmized) impression trajectories $X_i(t)$. The right panel shows the (logarithmized) yearly sum of clicks Y_i, received on the impressions of the ith keyword.

The data are provided by *Crealytics* (www.crealytics.com), an online advertising service provider with offices in Berlin (Germany), London (UK), and New

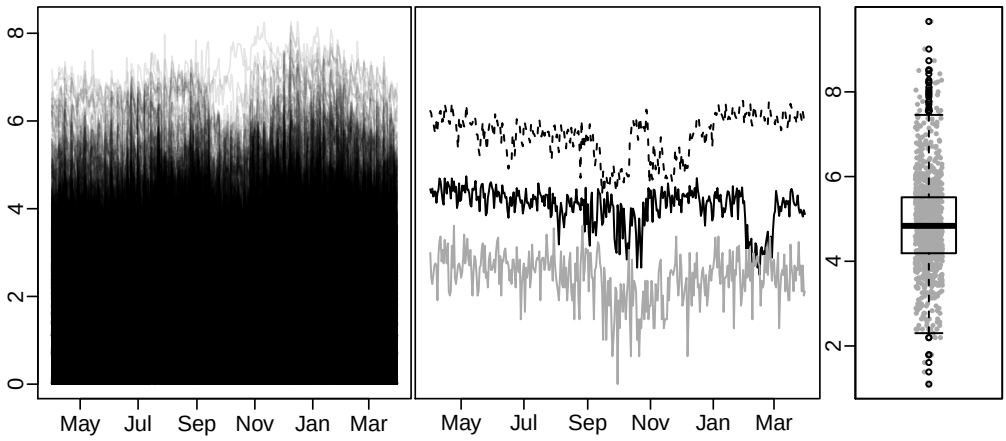

Figure 2.3: LEFT: Yearly trajectories of daily logarithmized numbers of impressions. MIDDLE: Three exemplary trajectories $X_i(t)$. RIGHT: Logarithmized yearly clicks Y_i.

York City (USA). The considered ad campaign is that of an online store selling outdoor equipment. (For reasons of confidentiality, we cannot publish the company's name). A lot of keywords received no impression during the considered time span of 365 days from April 1st, 2012, to March 31st, 2013. Therefore, we consider only the well established and relevant keywords that have been used on at least 320 days within the considered time span—leading to $n = 903$ trajectories observed at $p = 365$ grid points. The very few missing values in the logarithmized impression trajectories are imputed by zeros since a missing value means that the corresponding keyword did not receive an impression.

The considered functional linear regression model with PoIs in (2.1) is identifiable if the covariance function of the function-valued explanatory variable X_i is sufficiently non-smooth at the diagonal (see Section 2.2.1 and Theorem 3 in Kneip et al., 2016). Kneip et al. (2016) propose the following consistent estimator $\widehat{\kappa}$ for their κ controlling the smoothness at the diagonal of the covariance function:

$$\widehat{\kappa} = \log_2 \left(\frac{(1/(p-2k_\delta)) \sum_{j \in \mathcal{J}_{0,\delta}} \sum_{i=1}^{n} Z_{\delta,X_i}(t_j)^2}{(1/(p-2k_\delta)) \sum_{j \in \mathcal{J}_{0,\delta}} \sum_{i=1}^{n} Z_{\delta/2,X_i}(t_j)^2} \right).$$

An estimate of $\widehat{\kappa} < 2$ indicates identifiability, which is clearly fulfilled in our case where $\widehat{\kappa} = 0.03$.

The estimation results from applying our PES-ES estimation algorithm and the originally proposed KPS procedure are summarized in Figure 2.4 and Table 2.4. In the case of the PES-ES estimate, the function-valued slope parameter $\widehat{\beta}(t)$ shows a peak in the late summer and a pronounced negative trend towards the

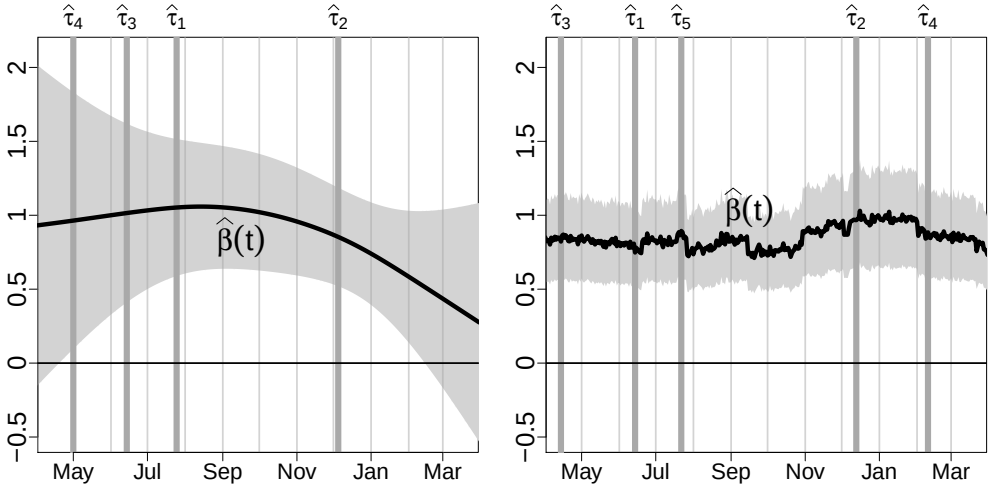

Figure 2.4: Result of the PES-ES (left panel) and KPS (right panel) estimate for $\beta(\cdot)$. The variabilities of the estimators are visualized using the gray shaded bands (see Remark 1).

Table 2.4: Estimate of PoI parameters β_r

PES-ES			KPS		
Location	Coef.	St.Err.	Location	Coef.	St.Err.
($\widehat{\tau}_4$) May 01	-0.17***	0.04	($\widehat{\tau}_3$) April 14	-0.10**	0.03
($\widehat{\tau}_3$) June 14	0.22***	0.03	($\widehat{\tau}_1$) June 14	0.22***	0.03
($\widehat{\tau}_1$) July 25	-0.15***	0.03	($\widehat{\tau}_5$) July 22	-0.17***	0.03
($\widehat{\tau}_2$) December 05	0.01	0.03	($\widehat{\tau}_2$) December 13	0.06*	0.03
			($\widehat{\tau}_4$) February 10	-0.11***	0.03

end of the considered period. The shape of $\widehat{\beta}(t)$ is in line with our expectations since the demand for outdoor equipment is generally greater during the summer months than during the winter months. The negative trend towards the end of the considered period is due to the strongly increased competition for outdoor equipment ads in Google AdWords during the considered period. Additionally, the estimation procedure identifies four PoIs (in order of the magnitude of $|\widehat{\beta}_s|$): June 14th ($\widehat{\tau}_3$; $\widehat{\beta}_3 = 0.22$), May 1st ($\widehat{\tau}_4$; $\widehat{\beta}_4 = -0.17$), July 25th ($\widehat{\tau}_1$; $\widehat{\beta}_1 = -0.15$), and December 5th ($\widehat{\tau}_2$; $\widehat{\beta}_2 = 0.01$), where the effect of the PoI at $\widehat{\tau}_2$ seems to be of lower importance.

Remark. *Drawing inference about the function-valued slope coefficient and the PoI parameters is a difficult issue in regression models with functional predictors.*

This is due to the fact that estimation in such models involves an ill-posed inversion problem and the estimator of the function-valued slope parameter is not asymptotically normal in the strong topology (Cardot et al., 2007). In addition, it is difficult to construct confidence regions for random elements in infinite dimensional Hilbert spaces with proper coverage probability (Choi and Reimherr, 2018). All we can do is to visualize the variability of the estimator that is due to the error term ϵ_i. For this purpose, we approximate the sampling variance of the composite parameter vector $\beta^{\rho}_{\widetilde{\mathcal{T}}}$ using Eq. (15.16) in Ramsay and Silverman (2005), Ch. 15, and show Bonferroni-adjusted Gaussian (invalid) confidence intervals in Figure 2.4.

The PoI $\widehat{\tau}_3$ on June 14th, with coefficient $\widehat{\beta}_3 = 0.22$, summarizes two positive effects. On the one hand, the store started a contest on May 23rd, 2012, giving away outdoor gear. This contest ended on June 13th, i.e., one day before the PoI which resulted in an increased click-through ratio of contest participants looking for the winners. On the other hand, the closest competitor started the spring sale, which led to a spillover bringing many interested buyers onto the homepage to compare prices.

The two other significant PoIs are explained by effects specific to the German calendar (about 80 percent of the customers live in Germany). The PoI $\widehat{\tau}_4$ on May 1st, with coefficient $\widehat{\beta}_4 = -0.17$, marks Labor Day (commemorating the Haymarket Riot in Chicago in 1886), a national holiday in Germany which is typically an opportunity for family outings. Similar in interpretation, the PoI $\widehat{\tau}_1$ on July 25th, with coefficient $\widehat{\beta}_1 = -0.15$, marks the beginning of the official summer holidays in Baden-Württemberg and Lower Saxony—two large German states. Both PoIs show a negative sign, which is due to a higher volume in search queries related to outdoor activities; however, the users do not click on the sponsored impressions since they do not intend to buy something—they are only searching the Internet for (free) information on hiking trails etc., which results in a lower CTR.

In contrast to the PES-ES estimate of $\beta(\cdot)$, the KPS estimate of $\beta(\cdot)$ is difficult to interpret and does not fit to our expectations (see right panel of Fig. 2.4): the trajectory is very unstable, does not show the expected peak in late summer, and does not show the plausible negative trend towards the end of the considered period. Regarding the PoI selections, the KPS approach identifies essentially the same PoIs as the PES-ES approach, but favors one additional PoI at February 10 (see Table 2.4), which has a significant negative impact ($\widehat{\beta}_4 = -0.11$) on the outcome variable. This additional PoI may reflect a compensation for the missing negative trend in the KPS estimate of $\beta(\cdot)$; see our discussion in Section 2.2.2.

The log-transformations in $Y_i = \log(C_i)$ and $X_i(t) = \log(\mathcal{I}_i(t))$ allow us to

interpret the estimated slope coefficients as elasticities. Taking derivatives with respect to $\mathcal{I}_i(t)$ at a single time point t leads to the following *time-local* elasticity:

$$\frac{\%\Delta C_i}{\%\Delta \mathcal{I}_i(t)} \approx \begin{cases} \widehat{\beta}_s & \text{if} \quad t = \widehat{\tau}_s \\ 0 & \text{else.} \end{cases}$$

That is, time-local changes in $\mathcal{I}_i(t)$ generally have no (i.e., practically negligible) effects on the yearly clicks C_i, except at PoIs, i.e., if $t = \widehat{\tau}_1, \ldots, \widehat{\tau}_{\widehat{S}}$. For instance, a 1% increase in the impressions at the time point of the after-contest PoI ($t = \widehat{\tau}_3$) causes (on average) a 0.22% ($\widehat{\beta}_3 = 0.22$) increase in the yearly clicks.

The function-valued slope parameter $\widehat{\beta}(t)$ does not contribute to the time-local elasticities; however, it determines the elasticities with respect to time-global changes in the impressions, for instance, over the course of a month. The following Riemann sum allows for a simple, approximative approach to interpret such *time-global* elasticities:

$$\widehat{\log(C_i)} \approx \frac{1}{365} \sum_{t=1}^{365} \widehat{\beta}(t) \log\big(\mathcal{I}_i(t)\big) + \sum_{s=1}^{\widehat{S}} \widehat{\beta}_s \log\big(\mathcal{I}_i(\widehat{\tau}_s)\big).$$

For instance, the total elasticity of C_i with respect to $\mathcal{I}_i(t)$ for *all* $t \in$ August is given by

$$\sum_{t \in \text{August}} \frac{\%\Delta C_i}{\%\Delta \mathcal{I}_i(t)} \approx \frac{1}{365} \sum_{t \in \text{August}} \widehat{\beta}(t) + \sum_{s=1}^{\widehat{S}} \widehat{\beta}_s \mathbf{1}_{(\widehat{\tau}_s \in \text{August})},$$

where $\mathbf{1}_{(\texttt{TRUE})} = 1$ and $\mathbf{1}_{(\texttt{FALSE})} = 0$. That is, a 1% increase in the impressions $\mathcal{I}_i(t)$, simultaneously for all $t \in$ August, causes a 0.1% increase in the yearly clicks since $365^{-1} \sum_{t \in \text{August}} \widehat{\beta}(t) + \sum_{s=1}^{\widehat{S}} \widehat{\beta}_s \mathbf{1}_{(\widehat{\tau}_s \in \text{August})} \approx 0.1$. Hence, the time-global August-elasticity is half the size of the elasticity of the after-contest PoI. This is absolutely plausible since the super-imposed influence of the contest and the spillover definitely outperforms a high-season month such as August in terms of clicks-per-impressions.

2.5 Conclusion

In this work we propose an improved algorithm for estimating the unknown model components of the functional linear regression model with points of Kneip et al. (2016). Our estimation algorithm decouples the estimation of the points of impact from the estimation of the function-valued slope parameter. The first

step of the estimation algorithm allows for a consistent estimation of the points of impact without knowledge (or pre-estimation) of the slope function. Given the consistent estimates of the points of impact, the second step of the estimation algorithm consists of an essentially classical estimation of the function-valued slope parameter. For this latter step, we propose a generalization of the penalized smoothing splines estimator of Crambes et al. (2009), which allows us to incorporate the estimates of the points of impact. A further minor finite sample improvement is achieved by repeating the estimation of the points of impact, given the estimate of the function-valued slope parameter from the second step and by a final repetition of the estimation of the slope parameter, given the updated estimates of the points of impact.

The new estimation algorithm significantly improves the original estimation procedure by Kneip et al. (2016). Using an extensive simulation study, we assess the robustness of our estimation algorithm for different data generating processes, different signal-to-noise ratios, different sample sizes and different sampling resolutions for discretizing the function-valued predictors.

The paper was originally motivated by an interesting case study on a Google AdWords ad campaign. Our proposed functional linear regression model with points of impact allows for data-based insights into the (time-global) seasonal factors and the (time-local) events influencing the yearly number of clicks on impressions of the considered Google AdWords online ad campaign.

2.A Appendix

Additional simulation setups

Table 2.A.1: Squared bias and variance of the estimators. Lowest/highest MSE has the darkest/lightest gray-scale. **Scenario:** No standardization of the functions in preselection step and $p = 300$ grid points.

		Easy		Complicated		NoPoI		OnlyPoI	
$\int \widehat{\beta}(t)$		Bias^2	Var.	Bias^2	Var.	Bias^2	Var.	Bias^2	Var.
$n = 250$	PES	0.05	0.89	2.02	12.36	0.00	0.02	0.00	0.08
	PES-ES	0.05	0.74	1.81	11.64	0.00	0.01	0.00	0.07
	PES-2ES	0.04	0.72	1.85	11.69	0.00	0.01	0.00	0.06
	KPS	3.98	60.37	139.62	301.13	0.01	0.02	0.12	10.97
$n = 500$	PES	0.01	0.23	0.87	6.04	0.00	0.01	0.00	0.01
	PES-ES	0.01	0.2	0.89	5.55	0.00	0.01	0.00	0.02
	PES-2ES	0.01	0.19	0.9	5.46	0.00	0.01	0.00	0.02
	KPS	0.69	23.39	82.42	241.99	0.01	0.01	0.01	1.77
$\frac{1}{S}\sum \widehat{\beta}_s$									
$n = 250$	PES	0.02	0.55	0.06	0.42	-	-	0.00	0.02
	PES-ES	0.02	0.46	0.04	0.33	-	-	0.00	0.02
	PES-2ES	0.02	0.45	0.04	0.32	-	-	0.00	0.02
	KPS	0.04	0.65	1.01	3.31	-	-	0.00	0.14
$n = 500$	PES	0.00	0.1	0.03	0.14	-	-	0.00	0.00
	PES-ES	0.00	0.09	0.02	0.11	-	-	0.00	0.00
	PES-2ES	0.00	0.09	0.02	0.1	-	-	0.00	0.00
	KPS	0.01	0.22	0.49	2.25	-	-	0.00	0.02

Table 2.A.2: Squared bias and variance of the estimators. Lowest/highest MSE has the darkest/lightest gray-scale. **Scenario:** With standardization of the functions in preselection step and $p = 500$ grid points.

		Easy		Complicated		NoPoI		OnlyPoI	
$\int \widehat{\beta}(t)$		Bias^2	Var.	Bias^2	Var.	Bias^2	Var.	Bias^2	Var.
$n = 250$	PES	0.04	0.37	0.16	1.43	0.01	0.04	0.00	0.06
	PES-ES	0.03	0.3	0.09	0.94	0.00	0.02	0.00	0.04
	PES-2ES	0.04	0.3	0.08	0.94	0.00	0.02	0.00	0.03
	KPS	2.62	46.83	135.19	288.08	0.01	0.02	0.09	8.14
	CKS	-	-	-	-	0.00	0.01	-	-
$n = 500$	PES	0.01	0.09	0.06	0.41	0.00	0.02	0.00	0.01
	PES-ES	0.01	0.08	0.05	0.38	0.00	0.01	0.00	0.02
	PES-2ES	0.01	0.08	0.04	0.38	0.00	0.01	0.00	0.02
	KPS	0.43	17.82	91.76	238.1	0.01	0.01	0.01	2.71
	CKS	-	-	-	-	0.00	0.00	-	-
$\frac{1}{S}\sum \widehat{\beta}_s$									
$n = 250$	PES	0.01	0.08	0.01	0.02	-	-	0.00	0.04
	PES-ES	0.01	0.11	0.01	0.02	-	-	0.00	0.05
	PES-2ES	0.01	0.11	0.01	0.02	-	-	0.00	0.05
	KPS	0.03	0.54	1.04	3.26	-	-	0.00	0.14
$n = 500$	PES	0.00	0.01	0.00	0.01	-	-	0.00	0.00
	PES-ES	0.00	0.03	0.00	0.01	-	-	0.00	0.00
	PES-2ES	0.00	0.03	0.00	0.01	-	-	0.00	0.00
	KPS	0.01	0.16	0.62	2.39	-	-	0.00	0.05

Table 2.A.3: Squared bias and variance of the estimators. Lowest/highest MSE has the darkest/lightest gray-scale. **Scenario:** No standardization of the functions in preselection step and $p = 500$ grid points.

		Easy		Complicated		NoPoI		OnlyPoI	
$\int \widehat{\beta}(t)$		Bias^2	Var.	Bias^2	Var.	Bias^2	Var.	Bias^2	Var.
$n = 250$	PES	0.09	1.18	1.32	11.64	0.00	0.02	0.00	0.08
	PES-ES	0.06	0.94	1.36	10.69	0.00	0.01	0.00	0.06
	PES-2ES	0.06	0.9	1.32	10.42	0.00	0.01	0.00	0.06
	KPS	2.05	41.9	145.2	291.28	0.01	0.02	0.04	6.06
$n = 500$	PES	0.02	0.26	0.48	4.67	0.00	0.01	0.00	0.01
	PES-ES	0.01	0.18	0.46	4.1	0.00	0.01	0.00	0.01
	PES-2ES	0.01	0.18	0.48	4.04	0.00	0.01	0.00	0.01
	KPS	0.47	19.09	81.27	229.75	0.01	0.01	0.01	2.77
$\frac{1}{S}\sum \widehat{\beta}_s$									
$n = 250$	PES	0.01	0.23	0.04	0.23	-	-	0.00	0.02
	PES-ES	0.01	0.28	0.04	0.22	-	-	0.00	0.04
	PES-2ES	0.01	0.3	0.04	0.22	-	-	0.00	0.04
	KPS	0.03	0.5	1.17	3.2	-	-	0.00	0.06
$n = 500$	PES	0.00	0.13	0.02	0.08	-	-	0.00	0.00
	PES-ES	0.00	0.12	0.02	0.07	-	-	0.00	0.00
	PES-2ES	0.00	0.11	0.02	0.09	-	-	0.00	0.00
	KPS	0.01	0.2	0.66	2.49	-	-	0.00	0.04

Table 2.A.4: Mean squared bias and variance. Lowest/highest MSE has the darkest/lightest gray-scale. DGP Complicated with different standard deviations σ_ϵ.

		$\sigma_\epsilon = 0.5$		$\sigma_\epsilon = 1$		$\sigma_\epsilon = 2$		$\sigma_\epsilon = 5$	
$\int \widehat{\beta}(t)$		Bias^2	Var.	Bias^2	Var.	Bias^2	Var.	Bias^2	Var.
$n = 250$	PES	0.22	1.78	0.28	3.76	0.58	13.76	0.67	31.37
	PES-ES	0.2	1.67	0.23	2.37	0.3	9.21	0.23	38.64
	KPS	19.51	95.41	1.52	35.09	0.56	27.38	0.21	46.26
$n = 500$	PES	0.12	0.67	0.15	1.36	0.29	5.59	0.5	21.84
	PES-ES	0.09	0.31	0.13	0.55	0.2	2.9	0.27	21.9
	KPS	11.47	80.98	0.6	20.17	0.2	10.18	0.14	25.74
$\frac{1}{\lvert S\rvert}\sum \widehat{\beta}_s$									
$n = 250$	PES	0.01	0.05	0.02	0.18	0.03	1.04	2.51	9.02
	PES-ES	0.01	0.05	0.01	0.18	0.01	0.69	1.9	11.37
	KPS	0.31	2.16	0.17	2.15	0.11	3.08	1.79	14.06
$n = 500$	PES	0.00	0.02	0.01	0.06	0.01	0.27	0.97	6.06
	PES-ES	0.00	0.02	0.01	0.05	0.01	0.2	0.42	5.86
	KPS	0.19	1.61	0.04	0.9	0.03	1	0.62	8.33

Table 2.A.5: Percentage of replications with correct detection of all points of impact $\tau_1, \ldots, \tau_S$.

		300 grid points			500 grid points		
		Easy	Compl.	OnlyPoI	Easy	Compl.	OnlyPoI
$n = 250$	PES	86.1	28.4	98.3	85	28	98.6
	PES-ES	87.4	30.4	98.3	86.5	30.4	98.8
	PES-2ES	87.5	30.4	98.3	86.6	30.8	98.8
	KPS	87.9	25.7	98.4	90.9	24.3	98.7
$n = 500$	PES	97.6	54.5	100	97.4	58.9	99.9
	PES-ES	97.8	56.2	100	97.8	61.1	99.9
	PES-2ES	97.8	56.2	100	97.8	61.2	99.9
	KPS	95.6	44.5	99.8	96.9	45.4	99.4

Chapter 3

Partially Observed Functional Data: The Case of Systematically Missing Parts

3.1 Introduction

The classical literature on Functional Data Analysis (FDA) focuses on the analysis of functions where each function $X_i(t)$ is observable for all $t \in [a, b]$ (see, for instance, the textbooks Ramsay and Silverman, 2005, Ferraty and Vieu, 2006, Horváth and Kokoszka, 2012, and Hsing and Eubank, 2015). However, this regular situation does not apply to many functional data sets of practical relevance where the functions X_i are only partially observable, i.e., where $X_i(t)$ is only observable for $t \in D_i$ with $D_i \subset [a, b]$ describing the i-specific observable subset of the total domain.

The latter situation is often referred to as fragmented, truncated, incomplete, or partially observed functional data and its practical relevance has triggered a series of research works dealing with different aspects of this problem. Delaigle and Hall (2013) propose a "shift-and-connect" procedure to reconstruct and classify fragmentary functional data and Delaigle and Hall (2016) incorporate a Markov chain model. Zhou et al. (2014) model truncated functional data using warping functions. Goldberg et al. (2014) and Kraus (2015) use functional linear regression models to predict the missing parts. Gellar et al. (2014b) and Gromenko et al. (2017) propose a functional regression model for incomplete curves. Liebl (2013) models and forecasts partially observed price functions and Kneip and Liebl (2019) consider optimal reconstructions of partially observed functional data. All these works, however, make use of the so-called

Missing-Completely-At-Random (MCAR) assumption, i.e., the assumption that the missing data mechanism is independent from all other stochastic components of relevance (see, for instance, Little and Rubin, 2014, Ch. 1). All of the above cited works lead to inconsistent results if the MCAR assumption is violated. To the best of our knowledge, we are the first to consider specific violations of the MCAR assumption in the context of partially observed functional data.

The classical missing data literature can be divided into the following four—not mutually exclusive—classes (see Little and Rubin, 2014, Ch. 1): first, procedures that discard all incompletely recorded units on the premise that the MCAR assumption holds; second, procedures that use a re-weighting of the data in order to adjust for the missing data; third, procedures that impute missing values using simple estimators (e.g., mean imputation); and fourth, procedures that are based on model assumptions. In this work, we add a fifth class that exclusively applies to functional data due to its general availability of derivatives. By using a detour via the fundamental theorem of calculus, we propose an estimation procedure that allows the consistent estimation of the mean and covariance function under specific violations of the MCAR assumption. In order to test for the considered violations of the MCAR assumption, we propose the application of the sequential multiple hypothesis testing procedure of Romano and Wolf (2005).

Note that we focus on the regular case of functional data where the single functions are only partially observed, but where the observed parts are fully observed. That is, we do not consider the case of sparse functional data where one observes only a few real, possibly noise contaminated, discretization points per function (see James et al., 2000, James and Sugar, 2003, Yao et al., 2005a, and Yao et al., 2005b, among others). Our estimation procedure requires the availability of derivatives and, therefore, cannot directly deal with the case of sparse functional data.

Our work is motivated by a real data set from energy economics with partially observed price curves $X_i(t)$. For each $X_i(t)$ we only observe the initial part for all $t \in [a, d_i]$; however, the final part with $t \in (d_i, b]$ is missing. The specific market structure incentivizes a systematic bidding strategy that results in a missing data mechanism where the random variable d_i correlates with the overall level of the price curves X_i. Larger values of d_i are associated with price curves having an overall high price level and vice versa (see Figure 3.5.3). The general exposition of our methodology is geared towards the data situation in this real data application, which provides a simple and instructive walk-along setup. However, we generalize this setup also for broader violations of the MCAR assumption and broader missing data designs.

The rest of the paper is structured as follows. The next section introduces the

setup under consideration, our statistical methodology, and generalizations. In Section 3.3, we propose a practical approach for testing the considered violations of the MCAR assumption. Section 3.4 contains our simulation study and the application is found in Section 3.5. Proofs and further derivations can be found in Appendix 3.A.

3.2 Methodology

3.2.1 General Setup

We consider an i.i.d. sample of differentiable random functions $X_1, \ldots, X_n$ each with the same distribution as X with values in the separable Hilbert space $L^2([a,b])$, where we set $[a,b] = [0,1] \subset \mathbb{R}$ without loss of generality. We assume that $\mathbb{E}[||X||_2^4] < \infty$, where $||X||_2^2 = \int_0^1 X(t)^2 dt$. The mean and covariance functions are denoted by $\mu(t) = \mathbb{E}[X(t)]$ and $\sigma(s,t) = \mathbb{E}[(X(s) - \mu(s))(X(t) - \mu(t))]$.

Motivated by our real data application we consider the following missing data mechanism: the random functions X_i are only observable over random subdomains $D_i = [0, d_i] \subseteq [0,1]$. Here, d_i are i.i.d. copies of a real random variable d with density f_d, where $f_d(t) > 0$ for $t \in [d_{\min}, 1]$ and zero else, with deterministic $0 < d_{\min} < 1$. The random subdomains $D_i = [0, d_i]$ lead to a t-specific observed data indicator $O_i(t)$ defined as $O_i(t) = \mathbb{1}_{t \in D_i}$; i.e., $O_i(t) = 1$ if $X_i(t)$ is observable and $O_i(t) = 0$ if $X_i(t)$ is missing. Let $p(t) = \mathbb{P}(O_i(t) = 1)$ denote the probability of observing functions covering t. Observe that under our setup $p(t) = 1$ for all $t \in [0, d_{\min}]$ and $0 < p(t) < 1$ for all $t \in (d_{\min}, 1]$. That is, the random functions do not contain missing parts over the lower interval $[0, d_{\min}]$, but may have missing endings over the upper interval $(d_{\min}, 1]$.

The MCAR assumption requires independence between the entire random processes O_i and X_i. Under our setup, however, a violation of the MCAR assumption can only affect the upper interval $(d_{\min}, 1]$, i.e., the part of the domain where the functions may have missing endings. That is, any violation of the MCAR assumption is without effect in the lower interval $[0, d_{\min}]$, where the functions do not contain missing observations; see Figure 3.5.3 for a real data example.

Our estimators use a detour via the fundamental theorem of calculus which allows us to address some practically relevant violations of the MCAR assumption. To formalize these violations, further notation needs to be introduced. Decompose $X(t) = \sum_{j \geq 1} \xi_j \psi_j(t)$, where $\mathfrak{B} = \{\psi_1, \psi_2, \ldots\}$ forms a deterministic orthogonal basis system of $L^2([0,1])$ with $\psi_1 \equiv 1$, and where $\xi_j = \int_0^1 X(t)\psi_j(t)dt$, with $j \geq 1$, denote the—not necessarily centered—random basis coefficients. Com-

mon basis systems which fulfill these assumptions are, for instance, a Fourier basis system or Legendre polynomials after applying an appropriate scaling of the domain. Denoting $S = \text{span}\{1\}$ and $S^\perp = \text{span}\{\psi_2, \psi_3, \dots\}$ allows us to separate the functions $X = X^S + X^{S^\perp}$, where X^S and $X^{S^\perp}$ are the orthogonal projections of X on S and $S^\perp$. With these notations, we can formalize the Violation (V) of the MCAR assumption:

(V) $X \not\perp\!\!\!\perp O$, but $X^{S^\perp} \perp\!\!\!\perp O$; i.e, the dependency between X and O manifests only in the dependency between O and the vertical shift component $\xi_1\psi_1 = \xi_1$ of X.

Note that Violation (V) causes distortions of the mean and covariance functions for every $t \in (d_{\min}, 1]$ since $\mathbb{E}[\xi_1|O(t)] \neq \mathbb{E}[\xi_1]$ for all $t \in (d_{\min}, 1]$. For $t \in [0, d_{\min}]$, however, Violation (V) is ineffective since $O(t) = 1$ almost surely, such that $\mathbb{E}[\xi_j|O(t)] = \mathbb{E}[\xi_j]$ for all $j \geq 1$ and all $t \in [0, d_{\min}]$.

This relatively simple Violation (V) of the MCAR assumption is motivated by our real data application and allows for a comprehensible and pithy explanation of our estimation strategies. Below, in Section 3.2.3, we introduce more general versions of Violation (V) and discuss how to apply our estimation strategies accordingly.

3.2.2 FTC-Estimators

In the case of partially observed functional data, it is impossible to use the classical mean estimator, $\bar{X}(t) = n^{-1}\sum_{i=1}^n X_i(t)$, and covariance estimator, $n^{-1}\sum_{i=1}^n (X_i(t) - \bar{X}(t))(X_i(s) - \bar{X}(s))$ since the value of $X_i(t)$ may not be observed. Therefore, Delaigle and Hall (2013) and Kraus (2015) propose the use of the following estimators:

$$\widehat{\mu}(t) = \frac{I_1(t)}{\sum_{i=1}^n O_i(t)} \sum_{i=1}^n X_i(t)O_i(t), \tag{3.1}$$

$$\widehat{\sigma}(s,t) = \frac{I_2(s,t)}{\sum_{i=1}^n U_i(s,t)} \sum_{i=1}^n U_i(s,t)\left[X_i(s) - \widehat{\mu}(s)\right]\left[X_i(t) - \widehat{\mu}(t)\right], \tag{3.2}$$

where the existence functions $I_1(t) = \mathbb{1}_{\sum_{i=1}^n O_i(t)>0}$ and $I_2(s,t) = \mathbb{1}_{\sum_{i=1}^n U_i(s,t)>0}$ for $U_i(s,t) = O_i(s)O_i(t)$ are necessary to prevent divisions by zero through defining $0/0 = \text{NA}$.

Violation (V) of the MCAR assumption implies that the estimator $\widehat{\mu}(t)$ is biased and inconsistent for all $t \in (d_{\min}, 1]$, i.e., for all t with $p(t) < 1$. The same applies to the estimator $\widehat{\sigma}(s,t)$ if at least one of the arguments s or t is

an element of the critical set $(d_{\min}, 1]$. Simple rearrangements using the law of iterated expectations show that under our setup

$$\mathbb{E}[\widehat{\mu}(t)] = \mu(t) + \Delta(t), \text{ where } \Delta(t) = \mathbb{E}\left[\frac{I_1(t)}{\sum_{i=1}^n O_i(t)} \sum_{i=1}^n O_i(t)\mathbb{E}\left[\xi_{i1} \mid O_i(t)\right]\right] - \mu_1,$$

with $|\Delta(t)| > 0$, i.e., $\mathbb{E}\left[\xi_{i1} \mid O_i(t)\right] \neq \mathbb{E}\left[\xi_{i1}\right]$, iff $p(t) < 1$; an essentially equivalent bias expression applies to $\widehat{\sigma}(s,t)$ (see also Proposition 3.A.1 in Appendix 3.A).

Therefore, we propose the following detour via the Fundamental Theorem of Calculus (FTC) which leads to consistent estimators of the mean and covariance functions under Violation (V) of the MCAR assumption. The FTC states that one can decompose a differentiable function $f(t)$ as $f(t) = \int_0^t f^{(1)}(z)dz + f(0)$, where $f^{(1)}$ denotes the first derivative of f. This motivates our new estimators $\widehat{\mu}_{\text{FTC}}(t)$ and $\widehat{\sigma}_{\text{FTC}}(s,t)$ for the mean and covariance functions.

Definition 3.2.1 (FTC-mean)**.** *The FTC mean estimator is defined as*

$$\widehat{\mu}_{\text{FTC}}(t) = \begin{cases} \widehat{\mu}(t) & \text{if} \quad t \in [0, d_{\min}] \\ \int_{d_{\min}}^t \widehat{\mu}^{(1)}(z)dz + \widehat{\mu}(d_{\min}) & \text{if} \quad t \in (d_{\min}, 1], \end{cases} \tag{3.3}$$

where

$$\widehat{\mu}^{(k)}(t) = \frac{I_1(t)}{\sum_{i=1}^n O_i(t)} \sum_{i=1}^n X_i^{(k)}(t) O_i(t),$$

and where $X_i^{(k)}(t)$ denotes the k-th derivative of $X_i(t)$.

Definition 3.2.2 (FTC-covariance)**.** *The FTC covariance estimator is defined as*

$$\widehat{\sigma}_{\mathrm{FTC}}(s,t) = \begin{cases} \widehat{\sigma}(s,t) & \text{if } (s,t) \in [0,d_{\min}]^2 \\ \int_{d_{\min}}^{t} \widehat{\sigma}^{(0,1)}(s,z_2)dz_2 + \widehat{\sigma}(s,d_{\min}) & \text{if } (s,t) \in [0,d_{\min}] \times (d_{\min},1] \\ \int_{d_{\min}}^{s} \widehat{\sigma}^{(1,0)}(z_1,t)dz_1 + \widehat{\sigma}(d_{\min},t) & \text{if } (s,t) \in (d_{\min},1] \times [0,d_{\min}] \\ \int_{d_{\min}}^{t}\int_{d_{\min}}^{s} \widehat{\sigma}^{(1,1)}(z_1,z_2)dz_1dz_2 + \\ \int_{d_{\min}}^{s} \widehat{\sigma}^{(1,0)}(z_1,d_{\min})dz_1 + \\ \int_{d_{\min}}^{t} \widehat{\sigma}^{(0,1)}(d_{\min},z_2)dz_2 + \\ \widehat{\sigma}(d_{\min},d_{\min}) & \text{if } (s,t) \in (d_{\min},1]^2, \end{cases} \tag{3.4}$$

where

$$\widehat{\sigma}^{(\ell,k)}(s,t) = \frac{I_2(s,t)}{\sum_{i=1}^{n} U_i(s,t)} \sum_{i=1}^{n} U_i(s,t) \left[X_i^{(\ell)}(s) - \widehat{\mu}^{(\ell)}(s)\right] \left[X_i^{(k)}(t) - \widehat{\mu}^{(k)}(t)\right]$$

with $\ell, k \in \{0,1\}$.

A step-by-step derivation of $\widehat{\sigma}_{\mathrm{FTC}}(s,t)$ can be found in parts I-III of the proof of Theorem 1 (see Appendix 3.A). The following theorem states the consistency of the FTC-estimators under Violation (V) of the MCAR assumption:

Theorem 1. *Under our setup and under Violation (V) of the MCAR assumption, the estimators* $\widehat{\mu}_{\mathrm{FTC}}(t)$*, defined in (3.3), and* $\widehat{\sigma}_{\mathrm{FTC}}(s,t)$*, defined in (3.4), are pointwise* $\sqrt{n}$*-consistent estimators of* $\mu(t)$ *and* $\sigma(s,t)$ *for all* $s,t \in [0,1]$.

3.2.3 Generalizations

Motivated by our real data application, we so far considered a relatively simple missing data design that can be characterized by the following two restrictions: first, $O_i(t)$ only depends on vertical shifts in $X_i(t)$, and second, the domain $[0,1]$ is divided into two subdomains $[0,d_{\min}]$ and $(d_{\min},1]$, where the first is covered by all functions, i.e., $p(t) = 1$ for all $t \in [0,d_{\min}]$, and where the second contains the missing parts, i.e., $0 < p(t) < 1$ for all $t \in (d_{\min},1]$. In the following, we generalize both of these restrictions.

The latter domain restriction can be weakened considerably, since the FTC estimators need only one point where the full sample of curves is observed. Denote this point by $d_{\mathrm{f}} \in [0,1]$, i.e., $p(d_{\mathrm{f}}) = 1$. By noting $\int_a^b = -\int_b^a$, the generalized

FTC mean estimator is defined as

$$\widehat{\mu}^{\star}_{\mathrm{FTC}}(t) = \begin{cases} \int_{d_\mathrm{f}}^{t} \widehat{\mu}^{(1)}(z)dz + \widehat{\mu}(d_\mathrm{f}) & \text{for all } t > d_\mathrm{f} \\ \widehat{\mu}(d_\mathrm{f}) & \text{for } t = d_\mathrm{f} \\ -\int_{t}^{d_\mathrm{f}} \widehat{\mu}^{(1)}(z)dz + \widehat{\mu}(d_\mathrm{f}) & \text{for all } t < d_\mathrm{f}. \end{cases} \tag{3.5}$$

Correspondingly, the generalized FTC covariance estimator is defined as

$$\begin{aligned} \widehat{\sigma}^{\star}_{\mathrm{FTC}}(s,t) &= \\ &= \int_{d_\mathrm{f}}^{t}\int_{d_\mathrm{f}}^{s} \widehat{\sigma}^{(1,1)}(z_1,z_2)dz_1dz_2 \\ &+ \int_{d_\mathrm{f}}^{s} \widehat{\sigma}^{(1,0)}(z_1,d_\mathrm{f})dz_1 + \int_{d_\mathrm{f}}^{t} \widehat{\sigma}^{(0,1)}(d_\mathrm{f},z_2)dz_2 + \widehat{\sigma}(d_\mathrm{f},d_\mathrm{f}) \end{aligned} \tag{3.6}$$

for all $s,t > d_\mathrm{f}$, where for cases with $s < d_\mathrm{f}$ and/or $t < d_\mathrm{f}$ one needs to replace "$\int_{d_\mathrm{f}}^{s}$" with "$-\int_{s}^{d_\mathrm{f}}$" and/or "$\int_{d_\mathrm{f}}^{t}$" by "$-\int_{t}^{d_\mathrm{f}}$", and where $\widehat{\sigma}^{\star}_{\mathrm{FTC}}(s,t) = \widehat{\sigma}(d_\mathrm{f},d_\mathrm{f})$ for $s = t = d_\mathrm{f}$. The estimators in (3.5) and (3.6) can be trivially adjusted for scenarios when the observation mechanism contains larger fragments of fully observed samples.

In order to generalize Violation (V), define $S_K = \mathrm{span}\{\psi_1,\ldots,\psi_K\}$ with monomial basis functions $\psi_j(t) = t^{j-1}$ for $j = 1,\ldots,K$ and let $S_K^{\perp}$ denote the orthogonal complement of S_K in $L^2([0,1])$. Additionally, we need to assume that X is K-times differentiable. The following violation of the MCAR assumption generalizes Violation (V):

(V^K) $X \not\perp\!\!\!\perp O$, but $X^{S_K^{\perp}} \perp\!\!\!\perp O$; i.e, the dependency between X and O manifests only in the dependency between O and the first K monomial components $\xi_1\psi_1,\ldots,\xi_K\psi_K$ of X.

Our estimation procedure can be applied to this more general Violation (V^K) of the MCAR assumption using the following recursion, which back-transforms

the consistent estimator $\widehat{\mu}^{(K)}$ by repeatedly applying the FTC:

$$\begin{aligned}
\widetilde{\mu}_{\mathrm{FTC}}^{(K-1)}(t) &= \int_{d_\mathrm{f}}^{t} \widehat{\mu}^{(K)}(z)dz + \widehat{\mu}^{(K-1)}(d_\mathrm{f}),\\
\widetilde{\mu}_{\mathrm{FTC}}^{(K-2)}(t) &= \int_{d_\mathrm{f}}^{t} \widetilde{\mu}_{\mathrm{FTC}}^{(K-1)}(z)dz + \widehat{\mu}^{(K-2)}(d_\mathrm{f}),\\
&\vdots\\
\text{and}\quad \widetilde{\mu}_{\mathrm{FTC}}(t) &= \int_{d_\mathrm{f}}^{t} \widetilde{\mu}_{\mathrm{FTC}}^{(1)}(z)dz + \widehat{\mu}(d_\mathrm{f}),
\end{aligned}$$

where $\widetilde{\mu}_{\mathrm{FTC}}(t)$ denotes the (back-transformed) estimator of $\mu(t)$.

An equivalent, yet more tedious recursion can be applied in order to back-transform the consistent estimator $\widehat{\sigma}^{(K,K)}$:

$$\begin{aligned}
\widetilde{\sigma}_{\mathrm{FTC}}^{(K-1,K-1)}(s,t) &= \int_{d_\mathrm{f}}^{t}\int_{d_\mathrm{f}}^{s} \widehat{\sigma}^{(K,K)}(z_1,z_2)dz_1dz_2 + \int_{d_\mathrm{f}}^{s} \widehat{\sigma}^{(K,K-1)}(z_1,d_\mathrm{f})dz_1 +\\
&\quad + \int_{d_\mathrm{f}}^{t} \widehat{\sigma}^{(K-1,K)}(d_\mathrm{f},z_2)dz_2 + \widehat{\sigma}^{(K-1,K-1)}(d_\mathrm{f},d_\mathrm{f}),\\
\widetilde{\sigma}_{\mathrm{FTC}}^{(K-2,K-2)}(s,t) &= \int_{d_\mathrm{f}}^{t}\int_{d_\mathrm{f}}^{s} \widetilde{\sigma}_{\mathrm{FTC}}^{(K-1,K-1)}(z_1,z_2)dz_1dz_2\\
&\quad + \int_{d_\mathrm{f}}^{s}\int_{d_\mathrm{f}}^{t} \widehat{\sigma}^{(K,K-2)}(z_1,d_\mathrm{f})dz_1dz_1\\
&\quad + \int_{d_\mathrm{f}}^{t}\int_{d_\mathrm{f}}^{t} \widehat{\sigma}^{(K-2,K)}(d_\mathrm{f},z_2)dz_2dz_2\\
&\quad + 2\widehat{\sigma}^{(K-1,K-2)}(d_\mathrm{f},d_\mathrm{f}) + \widehat{\sigma}^{(K-2,K-2)}(d_\mathrm{f},d_\mathrm{f}),\\
&\vdots\\
\text{and}\quad \widetilde{\sigma}_{\mathrm{FTC}}(s,t) &= \int_{d_\mathrm{f}}^{t}\int_{d_\mathrm{f}}^{s} \widetilde{\sigma}_{\mathrm{FTC}}^{(1,1)}(z_1,z_2)dz_1dz_2 + \int_{d_\mathrm{f}}^{s}\int_{d_\mathrm{f}}^{t} \widehat{\sigma}^{(0,1)}(z_1,d_\mathrm{f})dz_1dz_1\\
&\quad + \int_{d_\mathrm{f}}^{t}\int_{d_\mathrm{f}}^{t} \widehat{\sigma}^{(0,1)}(d_\mathrm{f},z_2)dz_2dz_2 + 2\widehat{\sigma}^{(1,0)}(d_\mathrm{f},d_\mathrm{f}) + \widehat{\sigma}^{(0,0)}(d_\mathrm{f},d_\mathrm{f}),
\end{aligned}$$

where $\widetilde{\sigma}_{\mathrm{FTC}}(s,t)$ denotes the (back-transformed) estimator of $\sigma_{\mathrm{FTC}}(s,t)$.

3.3 A Practical Approach to Test Violation (V)

In practice, there are situations where Violation (V) is essentially known from the context such as, for instance, in our real data application. Generally, however, practitioners would like to test for Violation (V) in order to gain confidence for

applying our FTC-estimators. In the following, we focus on testing Violation (V); however, the testing procedure can be easily generalized to the more general Violation (V^K) as well.

Unfortunately, a proper test for Violation (V) cannot—to the best of our knowledge—be achieved using existing procedures. A proper test procedure needs to investigate possible dependencies between O_i and the coefficients ξ_{i1}, ξ_{i2},.... The coefficients $\xi_{i1}, \xi_{i2}, \ldots$, however, refer to the complete trajectories and, therefore, cannot be computed from the partially observed trajectories. Furthermore, existing procedures that allow us to predict the coefficients $\xi_{i1}, \xi_{i2}, \ldots$ from the partially observed trajectories require the possibly violated MCAR assumption (see, for instance, Kraus, 2015, or Yao et al., 2005a).

Therefore, we can only propose a practical test procedure which will be useful for well-structured functional data, but generally cannot be applied in the case of complex structured functional data. The idea is to use the feasible coefficients $\xi_{i1}^{[0,d_{\min}]}, \xi_{i2}^{[0,d_{\min}]}, \ldots$ with respect to the sub-part of the domain $[0, d_{\min}]$ over which the full sample of functions is observed. For simply structured functional data, the coefficients $\xi_{i1}^{[0,d_{\min}]}, \xi_{i2}^{[0,d_{\min}]}, \ldots$ will be informative about the infeasible coefficients $\xi_{i1}, \xi_{i2}, \ldots$. In fact, Kneip and Liebl (2019) argue in a related context that the coefficients $\xi_{i1}^{[0,d_{\min}]}, \xi_{i2}^{[0,d_{\min}]}, \ldots$ contain the same information as $\xi_{i1}, \xi_{i2}, \ldots$ if the functional data are such that equality over $[0, d_{\min}]$ implies equality over the total domain $[0, 1]$; i.e., if $X_i(t) = X_j(t)$ for all $t \in [0, d_{\min}]$ and $i \neq j$ implies that $X_i(t) = X_j(t)$ for all $t \in [0, 1]$. The latter might hold for simply structured functional data and is fulfilled, for instance, for finite dimensional random functions $X_i(t) = \sum_{j=1}^{K} \xi_{ij}\psi_k(t)$, as long as the basis functions $\psi_1, \ldots, \psi_K$ are linearly independent over $[0, d_{\min}]$.

To identify cases falling under Violation (V), we propose the use of the following simple, practical procedure. Assuming Gaussian random functions X_i, it suffices to consider correlations in order to check the independence assumption in Violation (V). Under Violation (V), the upper threshold d_i of the partially observed domain $[0, d_i]$ is correlated with ξ_{i1}, but mutually uncorrelated with ξ_{ij} for all $j \geq 2$. In order to detect such a correlation structure, we propose the projection of the commonly observed parts of the random functions $X_i(t)$ for all $t \in [0, d_{\min}]$ onto a $J < \infty$ dimensional Fourier basis system, where the number of basis functions J is selected using the Bayesian Information Criterion (BIC): first, we select BIC-optimal numbers of basis functions J_i^{BIC} for all $i = 1, \ldots, n$, and second, we use $J = \text{median}\{J_1^{\text{BIC}}, \ldots, J_n^{\text{BIC}}\}$ as an overall trade-off between over- and underspecification.

Projecting the random functions $X_i(t)$ for all $t \in [0, d_{\min}]$ onto a $J < \infty$ dimensional Fourier basis system allows us to approximate the feasible coeffi-

cients $\xi_{i1}^{[0,d_{\min}]}, \ldots, \xi_{iJ}^{[0,d_{\min}]}$. In order to test for possible correlations between $\xi_{i1}^{[0,d_{\min}]}, \ldots, \xi_{iJ}^{[0,d_{\min}]}$ and d_i, we use the bootstrap-based multiple testing of Romano and Wolf (2005), which allows us to control for the multiple comparisons problem involved.

The procedure of Romano and Wolf (2005) controls the Family-Wise Error Rate (FWER) under an arbitrary set of dependence structures among the test statistics. Furthermore, their procedure allows for a so-called strong control of the FWER, i.e., the false rejection rate α is guaranteed for any combination of true and non-true null hypotheses in opposite to weak control, where a false rejection rate is only guaranteed if all null hypotheses are true (see Romano and Wolf, 2005, for more details).

We apply their sequential testing procedure to the multiple linear regression model $d_i = \beta_0 + \sum_{j=1}^{J} \beta_j \xi_{ji} + u_i$ (see Algorithm 2), where we substitute the ξ_{ji} with the approximated versions of the $\xi_{ij}^{[0,d_{\min}]}$. We investigate the set of hypotheses $B = \{\mathrm{H}_{0,1}, \ldots, \mathrm{H}_{0,J}\}$ with $\mathrm{H}_{0,j} : \beta_j = 0$, $j = 1, \ldots, J$, using standard squared t-test statistics $\hat{t}_j^2 = (\widehat{\beta}_j / \hat{s}_{\widehat{\beta}_j})^2$. Although there are a total of 2^J possibly different test outcomes, we are only interested in the following three different types of outcomes:

Null: The overall null: all $H_{j,0}, j = 1, \ldots, J$, cannot be rejected, i.e., all parameters β_j are insignificant and their basis coefficients ξ_{ij} do not correlate with d_i. This test decision points towards the MCAR assumption.

(V): Only $H_{1,0} : \beta_1 = 0$ will be rejected; all other $H_{j,0}, j = 2, \ldots, J$, cannot be rejected. That is, only ξ_{i1} significantly correlates with d_i. This test decision points towards Violation (V) of the MCAR assumption.

Other Any of the remaining $(2^J - 2)$ different test decisions. These test decisions point towards more complicated violations of the MCAR assumption.

Given n observations, J hypotheses, and a significance level α, the procedure proposed by Romano and Wolf (2005) is basically to bootstrap the statistics and calculate the bootstrapped p-value. For our special case, we can use a simplified version of the original algorithm (see Romano and Wolf, 2005, Algorithm 1 and Equation (26)). Our algorithm (see Algorithm 2) is initialized by setting $B = B_1$, where $\mathbf{d} = (d_1, \ldots, d_n)$, $\boldsymbol{\Xi} = (\xi_1, \ldots, \xi_J) \in \mathbb{R}^{n \times J}$, and R denotes the number of bootstrap replications:

A further simple visual tool for checking Violation (V) is to compare the estimates of the classical estimators, $\widehat{\mu}(t)$ and $\widehat{\sigma}(s,t)$ with those of the FTC estimators, $\widehat{\mu}_{\mathrm{FTC}}(t)$ and $\widehat{\sigma}_{\mathrm{FTC}}(s,t)$. Under the MCAR assumption the two estimates of the mean function and the two estimates of the covariance function

Algorithm 2 Romano Wolf Stepdown Procedure for Multiple Hypothesis Testing.

procedure RomanoWolf($\mathbf{d}, \boldsymbol{\Xi}, \alpha, R$)
  Calculate the test statistics $\hat{t}_j^2, j = 1, \ldots, J$, for the sample $(\mathbf{d}, \boldsymbol{\Xi})$.
  Perform a model-based bootstrap to obtain statistics $\hat{t}_{j,r}^2, j = 1, \ldots, J, r = 1, \ldots, R$.
  for $j = 1, \ldots, J$ hypotheses **do**
    Define $\hat{t}_{b_j}^2$ the maximum of all (remaining) test statistics in B_j with $b_j = \#B_j$.
    Define the p-value for b_j hypotheses as

$$\widehat{p}_{R,b_j} = \frac{1}{R}\left[1 + \sum_{r=1}^{R-1} \mathbb{1}\{\hat{t}_{j,r}^2 \geq \hat{t}_j^2\}\right]$$

    if $\widehat{p}_{R,b_j} > \alpha$ **then**
      stop and accept all remaining hypotheses in B_j,
    else
      reject the b_j-th hypothesis and define a new set of nulls $B_{j+1} = B_j \setminus H_{0,b_j}$.
    end if
  end for
  return $B = B_j$ the final set of hypotheses and $p = \widehat{p}_{R,b_j}$ the p-value.
end procedure

will essentially coincide; however, under Violation (V) the pairs of estimates will differ.

3.4 Simulation

In the following simulation study, we assess the finite sample properties of our FTC-estimators $\widehat{\mu}_{\text{FTC}}$ and $\widehat{\sigma}_{\text{FTC}}$ and investigate performance of the multiple testing procedure described in Algorithm 2. For a feasible data generation $X_i(t) = \sum_{j \geq 1} \xi_{ij}\psi_j(t)$, we use a finite dimensional Fourier basis $\psi_1(t) = 1$, $\psi_{2k}(t) = \sqrt{2}\sin(2\pi kt)$, and $\psi_{2k+1}(t) = \sqrt{2}\cos(2\pi kt)$, $k = 1, \ldots, (J-1)/2$, with $J = 5$ and draw uncorrelated coefficients $\xi_{ij}, j = 1, \ldots, J$, from a normal distribution with means $\mathbb{E}[\xi_{ij}] = \mu_j$ and variances $\mathbb{V}(\xi_{ij}) = \lambda_j$, where $\mu_1 = 5, \mu_2 = 2$, $\mu_3 = 0, \ldots, \mu_5 = 0$, and $\lambda_1 = 10, \lambda_2 = 8, \ldots, \lambda_5 = 2$. This results in the mean function $\mu(t) = 5 + 2 \cdot \sqrt{2} \cdot \sin(2\pi t)$ and the covariance function $\sigma(s,t) = \sum_{j=1}^{J} \lambda_j \psi_j(s)\psi_j(t)$.

We consider four different Data Generating Processes (DGPs), as described

in Table 3.4.1. The DGPs represent discrete (Dis.), i.e., $d_i \in \{0.5, 1\}$, and continuous (Con.), i.e., $d_i \in [0.5, 1]$, distributions for d_i, as well as dependent (Dep.), i.e., $\xi_{i1} \not\perp\!\!\!\perp d_i$, and independent (Ind.), i.e., $\xi_{i1} \perp\!\!\!\perp d_i$, setups. The Dep.-DGPs represent versions of Violation (V) of the MCAR assumption and the Ind.-DGPs fulfill the MCAR assumption.

Table 3.4.1: Data Generating Processes.

	Dep.-Dis.	Dep.-Con.	Ind.-Dis.	Ind.-Con.
ξ_{i1}	$\mathrm{N}(\mu_1, \lambda_1)$	$\mathrm{N}(\mu_1, \lambda_1)$	$\mathrm{N}(\mu_1, \lambda_1)$	$\mathrm{N}(\mu_1, \lambda_1)$
d_i	$\mathrm{sgn}_{\{0.5,1\}}(\xi_{i1}^c)$	$\mathrm{Unif}^{\mathrm{sc}}(\xi_{i1}; \mu_1, \lambda_1)$	$\mathrm{Ber}_{\{0.5,1\}}(0.5)$	$\mathrm{Unif}([0.5, 1])$

In each of the DGPs, ξ_{i1} is drawn from $\mathrm{N}(\mu_1, \lambda_1)$. For the Dep.-Dis. DGP, we define $d_i = 0.5$ if $\xi_{i1}^c = \xi_{i1} - \mu_1 < 0$ and $d_i = 1$ if $\xi_{i1}^c > 0$. In the Dep.-Con. DGP, we define d_i as a transformed version of ξ_{i1} using the following two steps: first, we use the uniform integral transformation to transform ξ_{i1} to a uniformly (on $[0, 1]$) distributed random variable $\Phi_{\mu_1,\lambda_1}(\xi_{i1})$, where Φ_{μ_1,λ_1} denotes the distribution function of the Normal distribution with parameters μ_1 and λ_1. Second, we project all values of $\Phi_{\mu_1,\lambda_1}(\xi_{i1})$ smaller 0.5 onto 0.5 and force the 2 percent largest values onto 1, which results in a mixture (discrete and continuous) distribution with point masses at 0.5 and 1. Formally, this means

$$d_i = \begin{cases} 0.5 & \text{if } d_i^* \leq 0.5 \\ d_i^* & \text{if } d_i^* \in (0.5, q_{\mathrm{emp},0.98}) \\ 1 & \text{if } d_i^* \geq q_{\mathrm{emp},0.98}, \end{cases}$$

where $d_i^* = \Phi_{\mu_1,\lambda_1}(\xi_{i1})$ and where $q_{\mathrm{emp},0.98}$ denotes the empirical 98 percent quantile of the sample $\{d_1^*, \ldots, d_n^*\}$. The structure of the Dep.-Dis. DGP resembles our application and represents a quite challenging example in contrast to DGPs with bigger masses at 1. For the Ind.-Dis. DGP we draw d_i from an adjusted Bernoulli (Ber) distribution on $\{0.5, 1\}$ with parameter $p = 0.5$, and for the Ind.-Con. DGP we draw d_i from a standard uniform distribution.

For each DGP, we generate 500 replications of $n = 50, 150, 250$, and 500 functions $X_i(t)$ evaluated at equidistant grid points $t_1, \ldots, t_p$ in $[0, 1]$ with $p = 501$, where all values $X_i(t_j)$ with $t_j > d_i$ are considered missing and replaced by NAs. Our simulation is implemented using the statistical language R (R Core Team, 2017b), where we make use of the package `fda` (Ramsay et al., 2014) to create the Fourier system. In order to maintain the sin-cosine pairs of the Fourier system, we consider only odd numbers of basis functions $J = 3, 5, \ldots, J_{\max}$ with

$J_{\max} = 51$.

Table 3.4.2: Integrated squared bias and variance.

		Dep.-Dis.		Dep.-Con.		Ind.-Dis.		Ind.-Con.	
n		Bias	Var.	Bias	Var.	Bias	Var.	Bias	Var.
50	$\widehat{\mu}_{\text{FTC}}$	0.0	1.0	0.0	1.0	0.0	0.9	0.0	1.0
	$\widehat{\mu}$	3.2	0.8	6.0	1.3	0.0	0.9	0.1	1.9
150	$\widehat{\mu}_{\text{FTC}}$	0.0	0.3	0.0	0.4	0.0	0.3	0.0	0.3
	$\widehat{\mu}$	3.2	0.3	6.8	0.5	0.0	0.3	0.0	0.6
250	$\widehat{\mu}_{\text{FTC}}$	0.0	0.2	0.0	0.2	0.0	0.2	0.0	0.2
	$\widehat{\mu}$	3.1	0.2	6.7	0.3	0.0	0.2	0.0	0.3
500	$\widehat{\mu}_{\text{FTC}}$	0.0	0.1	0.0	0.1	0.0	0.1	0.0	0.1
	$\widehat{\mu}$	3.2	0.1	6.7	0.1	0.0	0.1	0.0	0.2
50	$\widehat{\sigma}_{\text{FTC}}$	0.1	40.7	0.5	47.1	0.4	43.8	0.7	44.1
	$\widehat{\sigma}$	31.3	27.1	43.2	38.4	0.3	36.7	7.6	57.6
150	$\widehat{\sigma}_{\text{FTC}}$	0.1	13.3	0.1	15.7	0.1	15.1	0.1	15.0
	$\widehat{\sigma}$	30.4	8.7	38.0	16.6	0.0	12.2	0.7	28.3
250	$\widehat{\sigma}_{\text{FTC}}$	0.1	7.9	0.1	9.2	0.1	9.4	0.1	8.8
	$\widehat{\sigma}$	30.7	5.4	38.7	10.2	0.0	7.5	0.4	17.7
500	$\widehat{\sigma}_{\text{FTC}}$	0.1	4.3	0.0	4.8	0.1	4.7	0.1	4.6
	$\widehat{\sigma}$	0.1	4.3	0.0	4.8	0.1	4.7	0.1	8.7

Table 3.4.2 contains our simulation results for our FTC-estimators $\widehat{\mu}_{\text{FTC}}$ and $\widehat{\sigma}_{\text{FTC}}$. We report the (doubly) integrated squared bias $\left(\int_0^1\right)\int_0^1 \text{Bias}^2(\cdot)$ and variance $\left(\int_0^1\right)\int_0^1 \text{Var}(\cdot)$ for $\widehat{\mu}_{\text{FTC}}$ ($\widehat{\sigma}_{\text{FTC}}$) and compare them with the estimation errors of the classical estimators $\widehat{\mu}$ ($\widehat{\sigma}$). The results clearly demonstrate a systematic bias of the classical estimators $\widehat{\mu}$ and $\widehat{\sigma}$ under the Dep. DGPs (see the Dep.-Dis. and Dep.-Con. columns), which is non-vanishing as n increases. In contrast, our FTC-estimators show essentially no bias—despite Violation (V) of the MCAR assumption. If the MCAR assumption holds, both estimators lead to equivalent results up to some minor numerical issues; the slightly negative performance of the FTC-estimator $\widehat{\sigma}_{\text{FTC}}$ in the Ind.-Dis. DGP is due to computing numerical integrals and the disadvantage of the classical estimators in the Ind.-Con. DGP is due to the influence of ξ_1 with the largest variance λ_1 on the outer domains in $[d_{\min}, 1]$, where only a small percentage of the sample is observed.

To demonstrate the performance of our practical approach for identifying Violation (V), we present simulation results for the adapted multiple testing procedure proposed by Romano and Wolf (2005); see Table 3.4.3. We use a significance level of $\alpha = 0.05$ and report the percentages of replications falling

under the scenarios Null, (V), and Other (see Section 3.3). The simulation results provide a convincing picture of our practical identification procedure. The procedure complies with the strong control of the FWER in all DGPs with $n \geq 150$, where the true scenario (V) for the Dep. DGPs and the true scenario Null for the Ind. DGPs are only rejected in approximately 5% of the replications. For $n = 50$ the procedure shows a very poor performance, since selecting a basis system with up to $J_{\max} = 51$ basis elements in such a small sample typically leads to misspecified basis choices resulting in unstable test results. However, the test results stabilize when using a reduced number of $J_{\max} = 31$ basis elements for $n = 50$; see second row in Table 3.4.3. From our experience with this method, one should apply this procedure with $J_{\max} = 51$ for sample sizes $n \geq 150$, but a reduced number of basis elements $31 \leq J_{\max} < 51$ for $50 \leq n < 150$.

Table 3.4.3: Selection errors for Null, (V) and Other (in %) – $J_{\max} = 51$.

	Dep.-Dis.			Dep.-Con.			Ind.-Dis.			Ind.-Con.		
n	Null	(V)	Other	Null	(V)	Other	Null	(V)	Other	Null	(V)	Other
50	85.0	15.0	0.0	49.6	50.4	0.0	100.0	0.0	0.0	100.0	0.0	0.0
50*	0.0	97.2	2.8	2.0	95.0	3.0	93.8	1.2	5.0	98.6	0.2	1.2
150	0.0	99.0	1.0	0.0	97.8	2.2	98.2	0.2	1.6	97.6	0.2	2.2
250	0.0	96.6	3.4	0.0	99.0	1.0	97.4	0.4	2.2	97.6	0.6	1.8
500	0.0	95.6	4.4	0.0	94.6	5.4	96.0	0.6	3.4	95.2	0.2	4.6

* $J_{\max} = 31$.

3.5 German Control Reserve Market Study

In this section, we use the above described estimation procedure to estimate the mean price curve of control power prices from the German Control Reserve Market. The German Control Reserve Market plays an important role in maintaining grid stability, which is one of the basic responsibilities of electricity markets. Reserve capacities are an effective tool for guaranteeing grid stability. Each week, the German Federal Network Agency (FNA), a public institution, determines a sufficiently large amount of reserve capacity d_i that has to be bought, by law, by the Grid Control Cooperation[1] (GCC). The purpose of the German Control Reserve Market is to generate market prices at which the GCC has to buy the dictated reserve capacities from the electricity providers.

The German Control Reserve Market uses a so-called "pay-as-bid" auction, where every successful bidder (i.e., the electricity providers) receives his own

[1]The GCC is a merger of the four German Transport System Operators (TSOs) www.50hertz.com, www.amprion.net, www.transnetbw.de, and www.tennettso.de.

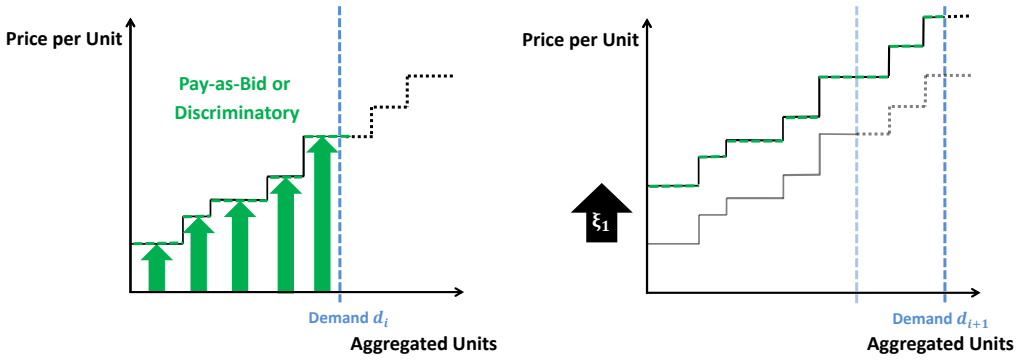

Figure 3.5.1: LEFT: The Pay-as-bid pricing mechanism, where each successful bidder receives the price of his bid. RIGHT: A vertical shift dependency of the price curves due to changes in the announced amount of reserve capacity d_i.

(sealed offered) price at which the bidder is willing to provide a certain amount of electricity. In this market design it is optimal to bid the (unknown) maximum price the GCC is eventually forced to pay (see, e.g., Grimm et al., 2008, or Engelmann and Grimm, 2009). Therefore, all bidders try to predict this unknown maximum price and the best predictor is the publicly announced amount of reserve capacity d_i. If the FNA dictates a large value of d_i, bidders tend to increase their prices, which leads to vertical shifts in the price curve and vice versa. That is, price curves X_i that are observed over larger sub-domains $[0, d_i]$ tend to have overall higher price levels and vice versa. This relationship is shown schematically in the right plot of Figure 3.5.1 and can be clearly seen in the data (Figure 3.5.3). This relatively simple dependency between the sub-domains $[0, d_i]$ and the level-shifts in the price curves X_i motivated the above described Violation (V) of the MCAR assumption.

Data, Preprocessing, and Exploratory Analysis: The data are freely available from `www.regelleistung.net` and we provide the data set as well as implementations of our estimators in the accompanying R-package `PartiallyFD`. The weekly data covers the time horizon from June 27th, 2011, to April, 17th, 2017. Two special auctions (January 2nd, 2012 and December 31st, 2012) resulted in extreme prices ($\geq$ 20,000 EUR/MW) and have been removed. In order to account for some very high prices in the data, we analyze logarithmized price curves, where all prices lower than 1 EUR/MW are mapped to 1 EUR/MW. To obtain a functional data sample, we pre-smooth each logarithmized price curve using cubic monotone P-splines in order to keep the monotonicity and the differentiability of the bid curves. The necessary function `monotSpline` is included in the ac-

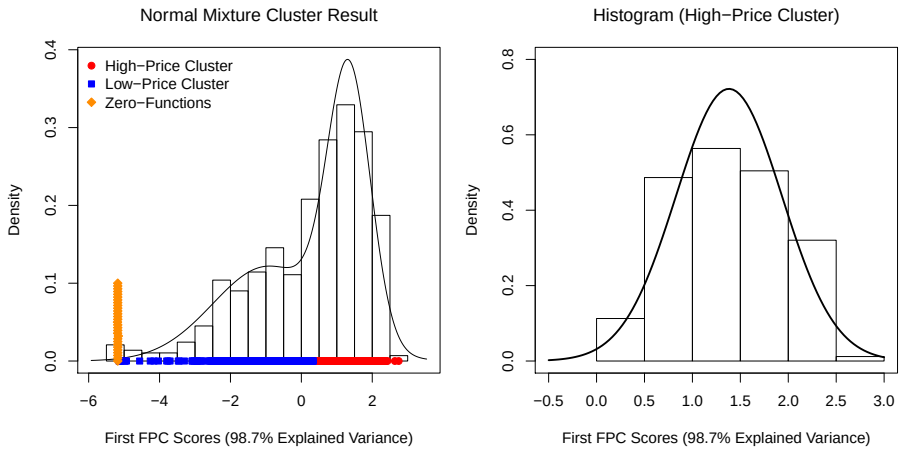

Figure 3.5.2: Distribution of the first principal component scores of the log-price curves computed with respect to the fully observed interval $[a, d_{\min}]$ of the original sample.

companying R-package `PartiallyFD`. All functions are observed over the initial interval $[a, d_{\min}] = [0\text{ MW}, 1832\text{ MW}]$, but are only partially observed over the final interval $(d_{\min}, b] = (1832\text{ MW}, 2500\text{ MW}]$. As in our simulation study, we evaluate the curves at $p = 501$ equidistant points $t_1 = 0, t_2 = 5, \ldots, t_{501} = 2500$ over the complete domain $[a, b] = [0\text{ MW}, 2500\text{ MW}]$; evaluation points within the missing sub-domains $[d_i, b]$ are filled by NAs. The log-price curves are relatively simply structured random functions and their first Functional Principal Component (FPC), computed with respect to the fully observed lower interval, accounts for 98.7% of the total variance. (See, for instance, Ramsay and Silverman, 2005, Ch. 8, for an introduction to FPC Analysis). This allows for an exploratory data analysis with respect to the first FPC scores which reveals that the original sample of log-price curves is bi-modal and has an additional point-mass at zero-price functions. Therefore, we remove the zero-price functions from the sample and perform a normal mixture cluster analysis on the first FPC scores using the R-package `mclust` of Fraley and Raftery (2002). The result of this exploratory data analysis is shown in Figure 3.5.2. For our further analysis we focus on the practically relevant high-price cluster with $n = 337$ price curves. Normality is not necessary for our estimators to be consistent under Violation (V), but it is needed in our practical test procedure described in Section 3.3. Standard normality tests reject the null hypothesis of normality due to the hard threshold used to allocate the data into the single clusters; however, the empirical distribution of the high-price cluster is not substantially far from a normal

distribution (see Figure 3.5.2).

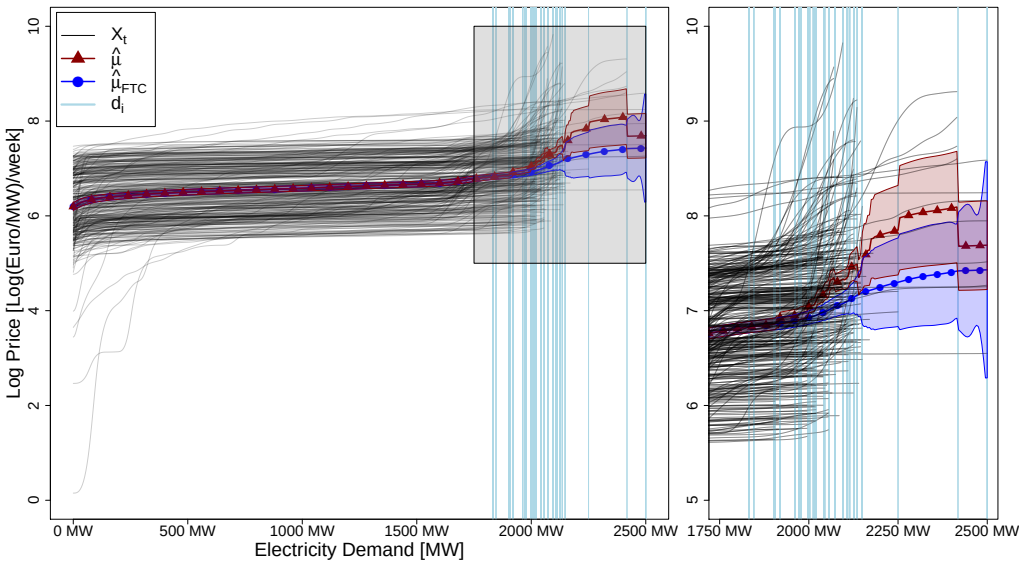

Figure 3.5.3: Classical mean (solid with triangles) and FTC mean estimates (solid with circles) computed from $n = 337$ partially observed price curves.

We test the functional data sample for Violation (V) as described in Section 3.3 where the procedure leads to scenario (V), i.e., the identification of Violation (V). For robustness, we additionally check $J_{\max} = 51 \pm 10$, which also leads to Violation (V). To estimate the mean function, we use our FTC-estimator $\widehat{\mu}_{\mathrm{FTC}}$ as defined in (3.3) and compare the estimation result with that from using the classical estimator $\widehat{\mu}$. The difference between the classical estimator $\widehat{\mu}(t)$, which is inconsistent under Violation (V), and our FTC-estimator $\widehat{\mu}_{\mathrm{FTC}}(t)$ is obvious: the systematically missing parts of the curves lead to a positive bias in $\widehat{\mu}(t)$ resulting in an implausible, non-monotonous, non-smooth mean curve. In contrast, our FTC-estimator leads to a smooth and monotonous mean curve, which is perfectly plausible given the monotonicity of the price curves. Nonetheless, the confidence intervals are overlapping, but this is not surprising given the sparseness of the data at the critical upper ending of the domain. The bias in the classical mean estimator $\widehat{\mu}$ also results in a defective estimate of the covariance function when using the classical covariance estimator $\widehat{\sigma}$. A plot of the covariance surface is omitted for reasons of space, but can be produced using the R-scripts of the online supplementary materials.

3.A Appendix

Proposition 3.A.1. *Under (B) and under Violation (V), the estimator $\widehat{\mu}(t)$ is biased, i.e., $\mathbb{E}[\widehat{\mu}(t)] = \mu(t) + \Delta(t)$, where $\Delta(t) = \mathbb{E}\left[\mathbb{E}\left[\xi_{i1} \mid O_i(t)\right]\right] - \mu_1\psi_1(t)$.*

Proof of Proposition 3.A.1: For simplicity, denote the indicator percentage $\mathrm{IP}(t) = \frac{I_1(t)}{\sum_{i=1}^n O_i(t)}$, then

$$\begin{aligned}
\mathbb{E}\left[\widehat{\mu}(t)\right] =& \mathbb{E}\left[\mathbb{E}\left[\widehat{\mu}(t) \mid O_1(t), \ldots, O_n(t)\right]\right] \\
=& \mathbb{E}\left[\mathrm{IP}(t) \sum_{i=1}^n O_i(t) \left(\mathbb{E}\left[\xi_{i1} \mid O_i(t)\right] \psi_1(t) + \sum_{j\geq 2} \mu_j \psi_j(t)\right)\right] \\
=& \mathbb{E}\left[\mathrm{IP}(t) \sum_{i=1}^n O_i(t) \sum_{j\geq 2} \mu_j \psi_j(t)\right] \\
+& \mathbb{E}\left[\mathrm{IP}(t) \sum_{i=1}^n O_i(t) \mathbb{E}\left[\xi_{i1} \mid O_i(t)\right] \psi_1(t)\right]
\end{aligned}$$

The first summand deviates from $\mu(t)$ only by $\mu_1\psi_1(t)$:

$$\begin{aligned}
\mathbb{E}\left[\mathrm{IP}(t) \sum_{i=1}^n O_i(t) \sum_{j\geq 2} \mu_j \psi_j(t)\right] =& \mathbb{E}\left[\mathrm{IP}(t) \sum_{i=1}^n O_i(t) \left(\sum_{j\geq 1} \mu_j \psi_j(t) - \mu_1\psi_1(t)\right)\right] \\
=& \mu(t) - \mu_1\psi_1(t)
\end{aligned}$$

The second part deviates from $\mu_1\psi_1(t)$ by $\Delta(t)$ defined via

$$\Delta(t) = \mathbb{E}\left[\frac{I_1(t)}{\sum_{i=1}^n O_i(t)} \sum_{i=1}^n O_i(t) \mathbb{E}\left[\xi_{i1} \mid O_i(t)\right]\right] - \mu_1\psi_1(t)$$

which results in

$$\begin{aligned}
\mathbb{E}\left[\mathrm{IP}(t) \sum_{i=1}^n O_i(t) \mathbb{E}\left[\xi_{i1} \mid O_i(t)\right] \psi_1(t)\right] =& \mathbb{E}\left[\mathrm{IP}(t) \sum_{i=1}^n O_i(t) \left(\mu_1\psi_1(t) + \Delta(t)\right)\right] \\
=& \mu_1\psi_1(t) + \Delta(t)
\end{aligned}$$

The bias of $\widehat{\sigma}(s,t)$ can be derived in a similar manner.

Proof of Theorem 1: First, we focus on $\widehat{\mu}_{\mathrm{FTC}}(t)$. When $t \in [0, d_{\min}]$, we observe $p(t) = 1$, i.e., 100% of the sample. Therefore, the MCAR assumption is not violated and $\widehat{\mu}$ is unbiased and consistent.

For $t \in (d_{\min}, 1]$, the estimator is defined as

$$\widehat{\mu}_{\text{FTC}}(t) = \int_{d_{\min}}^{t} \widehat{\mu}^{(1)}(z)dz + \widehat{\mu}(d_{\min})$$

by decomposing $\widehat{\mu}(t)$ via the FTC. A direct application of the product rule delivers the representation of

$$\widehat{\mu}^{(1)}(z) = \frac{J(z)}{\sum_{i=1}^{n} O_i(z)} \sum_{i=1}^{n} X_i^{(1)}(z) O_i(z) dz \tag{3.7}$$

as the mean estimator of the first derivatives $X_i^{(1)}(t)$.

Under our setup with $\psi_1 \equiv 1$ and under Violation (V), the derivatives $X_i^{(1)}(z) = \sum_{j \geq 2} \xi_{ij} \psi_j^{(1)}(z)$ are independent from $O_i(z)$ such that there is no violation of the MCAR when estimating $\mu^{(1)}(z)$ by $\widehat{\mu}^{(1)}(z)$ for $z \in [d_{\min}, 1]$. It follows by the Weak Law of Large Numbers that $\widehat{\mu}^{(1)}(z) \to_p \mu^{(1)}(z)$ as $n \to \infty$. For the second summand in (3.3)—since $p(d_{\min}) = 1$—we see that $n^{-1} \sum_{i=1}^{n} X_i(d_{\min}) \to_p \mu(d_{\min})$ as $n \to \infty$. With the Continuous Mapping Theorem (CMT) we finally deduce the unbiasedness and consistency of $\widehat{\mu}_{\text{FTC}}(t)$ for all $t \in [0, 1]$. This reasoning is later used to show consistency in the case of $\widehat{\sigma}_{\text{FTC}}(s, t)$ is as well.

In the case of $\widehat{\sigma}_{\text{FTC}}(s, t)$, we go stepwise through all combinations for $s, t \in [0, d_{\min}] \cup (d_{\min}, 1]$.

I: For $s, t \in [0, d_{\min}]$.
Here, $p(t) = 1$, i.e., there is no missing data and the standard estimator $\widehat{\sigma}(s, t)$ as in (3.2) is consistent and therefore $\widehat{\sigma}_{\text{FTC}}(s, t)$ as well.

II: Without loss of generality, we assume $s \in [0, d_{\min}]$ and $t \in (d_{\min}, 1]$.
Then, the estimator is defined as

$$\widehat{\sigma}_{\text{FTC}}(s, t) = \int_{d_{\min}}^{t} \widehat{\sigma}^{(0,1)}(s, z_2) dz_2 + \widehat{\sigma}(s, d_{\min}),$$

where $\widehat{\sigma}^{(k,\ell)}(s, t) = (\partial^{k+\ell} / (\partial s^k \partial t^\ell)) \widehat{\sigma}(s, t)$ with $k, \ell \in \mathbb{N}$. Hence, using the same arguments as for (3.7), applied to

$$\widehat{\sigma}^{(0,1)}(s, z_2) = \frac{I_2(s, z_2)}{\sum_{i=1}^{n} U_i(s, z_2)} \sum_{i=1}^{n} U_i(s, z_2) \Big[X_i(s) - \widehat{\mu}(s) \Big] \Big[X_i^{(1)}(z_2) - \widehat{\mu}^{(1)}(z_2) \Big]$$

and to $\widehat{\sigma}(s, d_{\min})$, we can deduce the consistency of $\widehat{\sigma}^{(0,1)}(s, t)$ and $\widehat{\sigma}(s, d_{\min})$ for $s \in [0, d_{\min}]$ and $z_2 \in (d_{\min}, 1]$.
Parallel reasoning holds for the case where $s \in (d_{\min}, 1]$ and $t \in [0, d_{\min}]$.

III: For $s, t \in (d_{\min}, 1]$.
In order to derive the estimator for the case where $s, t \in (d_{\min}, 1]$, observe that the following three equations hold due to the FTC:

$$\sigma(s,t) = \int_{d_{\min}}^{t} \frac{\partial}{\partial z_2}\sigma(s,z_2)dz_2 + \sigma(s,d_{\min}) \tag{3.8}$$

$$\sigma(s,z_2) = \int_{d_{\min}}^{s} \frac{\partial}{\partial z_1}\sigma(z_1,z_2)dz_1 + \sigma(d_{\min},z_2) \tag{3.9}$$

$$\int_{d_{\min}}^{t} \frac{\partial}{\partial z_2}\sigma(d_{\min},z_2)dz_2 = \sigma(d_{\min},t) - \sigma(d_{\min},d_{\min}) \tag{3.10}$$

Plugging (3.9) into (3.8) and using the linearity of the operations involved yields

$$\begin{aligned}\sigma(s,t) &= \int_{d_{\min}}^{t} \frac{\partial}{\partial z_2}\left(\int_{d_{\min}}^{s} \frac{\partial}{\partial z_1}\sigma(z_1,z_2)dz_1 + \sigma(d_{\min},z_2)\right)dz_2 + \sigma(s,d_{\min}) \\ \Leftrightarrow \sigma(s,t) &= \int_{d_{\min}}^{t}\int_{d_{\min}}^{s} \frac{\partial^2}{\partial z_1 \partial z_2}\sigma(z_1,z_2)dz_1dz_2 + \int_{d_{\min}}^{t} \frac{\partial}{\partial z_2}\sigma(d_{\min},z_2)dz_2 \\ &\quad + \sigma(s,d_{\min}).\end{aligned} \tag{3.11}$$

Plugging (3.10) into (3.11) yields

$$\begin{aligned}\sigma(s,t) = &\int_{d_{\min}}^{t}\int_{d_{\min}}^{s} \frac{\partial^2}{\partial z_1 \partial z_2}\sigma(z_1,z_2)dz_1dz_2 + \sigma(d_{\min},t) \\ &+ \sigma(s,d_{\min}) - \sigma(d_{\min},d_{\min}).\end{aligned} \tag{3.12}$$

Replacing $\sigma(d_{\min},t)$ with a corresponding version of (3.8) and $\sigma(s,d_{\min})$ with a corresponding version of (3.9) yields

$$\begin{aligned}\sigma(s,t) = &\int_{d_{\min}}^{t}\int_{d_{\min}}^{s} \frac{\partial^2}{\partial z_1 \partial z_2}\sigma(z_1,z_2)dz_1dz_2 \\ &+ \int_{d_{\min}}^{t} \frac{\partial}{\partial z_2}\sigma(d_{\min},z_2)dz_2 + \sigma(d_{\min},d_{\min}) \\ &+ \int_{d_{\min}}^{s} \frac{\partial}{\partial z_1}\sigma(z_1,d_{\min})dz_1 + \sigma(d_{\min},d_{\min}) \\ &- \sigma(d_{\min},d_{\min}).\end{aligned}$$

Simplifying and using the notation of Definition 3.2.2 leads to:

$$\sigma(s,t) = \int_{d_{\min}}^{t}\int_{d_{\min}}^{s} \sigma^{(1,1)}(z_1,z_2)dz_1dz_2 + \int_{d_{\min}}^{t} \sigma^{(0,1)}(d_{\min},z_2)dz_2 + \int_{d_{\min}}^{s} \sigma^{(1,0)}(z_1,d_{\min})dz_1 + \sigma(d_{\min},d_{\min}),$$

where, following the equivalent arguments as used above, $\sigma^{(1,1)}(z_1,z_2)$, $\sigma^{(0,1)}(d_{\min},z_2)$, $\sigma^{(1,0)}(z_1,d_{\min})$, and $\sigma(d_{\min},d_{\min})$ can be estimated consistently using

$$\widehat{\sigma}^{(\ell,k)}(s,t) = \frac{I_2(s,t)}{\sum_{i=1}^{n} U_i(s,t)} \sum_{i=1}^{n} U_i(s,t) \left[X_i^{(\ell)}(s) - \widehat{\mu}^{(\ell)}(s)\right] \left[X_i^{(k)}(t) - \widehat{\mu}^{(k)}(t)\right]$$

with $\ell, k \in \{0,1\}$.

Chapter 4

Forecasting of discriminatory auction curves with underlying missing data mechanism in the German Capacity Reserve and Balance Power Market

4.1 Introduction

Capacity reserve and balancing power markets are essential for both electricity and policy makers. Costs for capacity provision and instantaneously retrievable energy play an important role in evaluating the costs for frequency control, the integration of renewable energies, and future investment opportunities. Since, e.g., electricity storage is currently prohibitively expensive, grid operators need these markets to efficiently acquire capacity to permanently balance generation and demand. While uniform pricing mechanisms result in one uniform price, the comprehension of costs in discriminatory auction mechanisms is significantly more complicated. Therefore, the understanding of the structure and dependencies of whole curves rather than single statistics of these curves is crucial for the future cost development and, in particular, predictions.

In this paper, we first give a brief introduction to auction econometrics, particularly the different strategies behind uniform versus discriminatory or "pay-as-bid" auctions. The first format typically involves standard mechanisms, see, e.g., Liebl (2013), while the latter involves the prediction of "maximal" prices closely related to the X-model approach, see, e.g., Kulakov (2020). To motivate

the problem, we briefly explain the German Auction Design for the GCR and give a detailed description of the data. Later, we give a short introduction to Functional Data Analysis (FDA) which analyzes statistics of function-valued random variables where we focus on partially observed functional data as described in Liebl and Rameseder (2019). FDA plays an important role in auction forecasting, see, e.g., Wang et al. (2008), and in forecasting electricity spot prices, see Liebl (2013). On the other hand, Ziel and Steinert (2016) use multivariate regime-switching methods to predict independent demand and supply curves at the EEX to provide the intersection—the well-known "X-model"—as price forecast. There are two functional requirements in the following case which do not allow for their approach: the first is to allow for a continuum of regimes and the second is to use a detour via derivatives. For a general summary of different forecasting approaches, we refer to Weron (2014) and Weron (2007) for an overview. Ziel and Steinert (2016) and Ziel and Weron (2018) provide a multivariate but not applicable relative of our forecasting. Our method will be explained in detail. We conclude with an evaluation of four different auctions with different methods and specifically auction-adjusted metrics.

4.1.1 Auctions

In the following we are dealing with a sequential, sealed-bid, discriminatory price, multi-unit auction. We recommend Krishna (2009) for an exact definition and introduction to the extensive amount of theoretical research in auction theory. There is—probably due to the absence of widely available data—a sparse amount of empirical literature: Paarsch et al. (2006) provide an introduction and Hendricks and Paarsch (1995) give a research summary; the authors of Shmueli and Jank (2005) started a series of papers applying functional data to ebay auctions; however, their findings are not applicable since their ebay data cannot be considered as multi-unit auction. There is an important step from auction literature typically concerned with individual interactions to anonymous online auctions for which we refer to Lucking-Reiley (1999). Nevertheless, this approach is negligible for our purposes due to the sealed-bid mechanism.

We recommend Kwasnica and Sherstyuk (2013) for a theoretical survey of multi-unit auctions; from an empirical perspective, we recommend Hortaçsu (2011) for applied research in the case of multi-unit auctions from a structural point of view. Since this approach tries to find structural parameters by estimating individual bid functions, we are not able to follow such an approach due to the anonymous bids.

4.1.2 The German Capacity Reserve and Balancing Energy Market (GCR)

Typically, electrical grids have to fulfill the property that consumption and production of electricity equal each other in every moment. While in the classical view consumption defined how much electricity has to be produced, i.e., electrical producers had to adjust to cover the demand, smart technologies like intelligent grids and storage opportunities allow for more flexibility on both sides. Nevertheless, shortages or exuberances of production and/or demand can never be completely prohibited, which is the reason why the policy maker typically generates a market to cover for these cases.

In Germany, this market is called the "Capacity Reserve and Balancing Energy Market" (GCR) and is operated by the Grid Control Cooperation[1] (GCC). It is designed for two purposes: ex ante, it insures that the grid provides enough power (capacity reserve) to guarantee 99.9% of all shortcomings and congestions and, ex post, it allocates enough positive and negative balancing energy to stabilize the grid after a shortcoming and exuberance again.

It is technically structured into three different reserve qualities; primary (PR), secondary (SR), and tertiary reserve (TR)—for a detailed description, we recommend Just (2015). In our application, we focus on SR, which is the biggest market. SR consists of four auctions, two for the provision of power and energy—we abbreviate them as POS (NEG) in the case of a shortcoming (exuberance)—and two for peak and off-peak, which we abbreviate with HT and NT. This results in four weekly hold auctions POS HT, POS NT, NEG HT, and NEG NT. We focus on NEG HT, i.e., auctions for exuberances during peak. In addition, we also provide results for the others in the GIT repository. For a very detailed economic description, see Ocker et al. (2015).

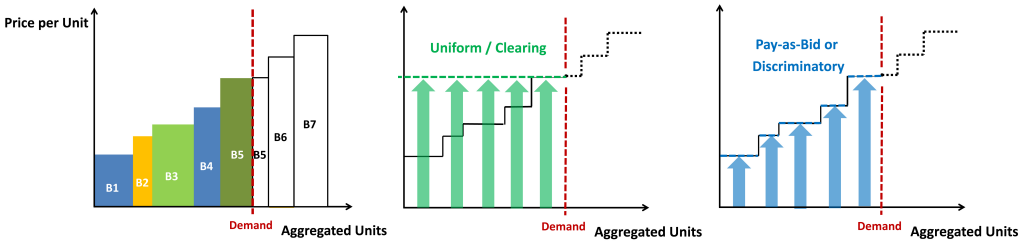

Figure 4.1.1: LEFT: Aggregation of the (censored) supply curve with differently colored market participants. MIDDLE: The uniform pricing mechanism. RIGHT: A variant of discriminatory or "pay-as-bid" pricing, all with inelastic demand.

[1]The GCC is a merger of the four German Transport System Operators (TSOs) www.50hertz.com, www.amprion.net, www.transnetbw.de, and www.tennettso.de.

The German Control Reserve auction uses so-called pay-as-bid pricing. Every bidder—typically an energy producer—provides his own sealed-offered price at which the bidder is willing to provide a certain amount of electricity. These bids are sorted according to the price (see Figure 4.1.1, left). Depending on the demand, bids are successful if their accumulated supply position is below the demand and they receive the fee they offered their electricity for (see Figure 4.1.1, right). In this market design it is strategically optimal to bid almost the (unknown) maximum price the GCC is eventually forced to pay (see, e.g., Grimm et al. (2008) or Engelmann and Grimm (2009)). This is in contrast to uniform pricing in which it is optimal to bid the own costs plus an epsilon (see Figure 4.1.1, middle).

4.2 Data

4.2.1 Data generating process

In the main paper, we focus on applying our method onto the NEG HT auctions. The raw data can be downloaded at GCC's homepage. We provide collected weekly raw bids—from 2011-06-27 until 2017-04-17—for capacity reserve prices here, in this case: "SRL_NEG_HT_LP". In addition, the GIT repository provides all codes and data for all steps and all SR auctions, which leads to similarly positive results as in the SR NEG HT auction.

The available raw data has the following structure: each week on Wednesday, every prequalified and anonymous auction participant can provide multiple bids $b_i^t, i = 1, \ldots, n(t)$. Here, the i-th bid b_i^t of total number of bids $n(t)$ for the auction at date $t = 1, \ldots, T$ is a tuple (τ_i^t, x_i^t) consisting of a number of supplied capacity $\tau_t^i \in \mathbb{N} > 5$ (in MW) and a capacity price $x_t^i \in \mathbb{R}_0^+$ (in €/MW/w). The demand of capacity reserve d_t at auction t is chosen exogenously by the Transport System Operators (TSOs) and is known to all bidders in advance (see Figure 4.A.2).

A bid b_i^t is called "winning"—the bids left of the demand in Figure 4.1.1, right—i.e., the bidder receives the fee to provide capacity reserve if the cumulative capacity of all lower and equal bids is below the demand d_t. Formally, the i-th bid b_t^i is a winning bid, i.e., $Z_t(b_t^i) = 1$ defined via

$$Z_t(b_i^t) = I(x_i^t \leq x_{\max}^t), \quad \text{where} \quad x_{\max}^t = \max_{1 \leq i \leq n(t)} \{x_i^t \mid \sum_{k=1}^{i-1} \tau_k^t < d_t\} \tag{4.1}$$

The "winning" function $Z_t(\cdot)$ will be used to evaluate our forecasting algorithm (see Section 4.3.7).

We want to provide a detailed view of the data with multiple graphs. Since the GCC does not provide any information about non-winning bids, i.e., $\{x_i^t \mid Z_t(b_t^i) = 0\}$, we can only present the winning bids (see, e.g., the black dots in Figure 4.A.1).

For an overview of the temporal development of the auction curves, consider Figure 4.2.1, a censored plot with prices of up to 3500 /MW/w—for the other auctions, see Figure 4.A.3. This graph type shows a two-dimensional projection of bids onto bins—the sum of MW aggregated in a bin are color-coded: the darker the color, the more MW at such a price bin and the flatter the curve in this price range. Figure 4.A.1, left, shows exemplary the winning bids and the demanded capacity d_t (2240 MW) for one week starting on Monday 2012-02-02. The demand over time can be seen in Figure 4.A.2—the obvious dependency during Christmas weeks is due to higher uncertainty of electricity production and consumption during holidays.

All bidders try to predict this unknown maximum price, i.e., the highest price amongst the winning bids, and the best predictor is the publicly announced amount of reserve capacity d_t. If the Federal Network Agency (FNA) dictates a large value of d_t, bidders tend to increase their prices, which leads to vertical shifts in the price curve and vice versa (see Figure 4.3.1). That is, prices x_i^t that are observed over larger sub-domains $[0, d_t]$ tend to have overall higher price levels and vice versa. This relationship is shown schematically in the right plot of Figure 4.3.1 and can be clearly seen in the data (see Figure 4.3.2 or this .gif). This relatively simple dependency between the sub-domains $[0, d_t]$ and the level-shifts in the prices x_i^t motivated Liebl and Rameseder (2019) to develop new mean and covariance estimators, which we exploit for maximum price forecasting.

4.2.2 Data preprocessing

We smooth the discrete data points using a monotone P-Spline interpolation allowing us to define the bid curves $X_t(\tau) = \sum_{j=1}^m \hat{\beta}_{jt}\delta_j(\tau)$ with cubic B-splines δ_j for each t as

$$\min_{\beta} \sum_{i=1}^{n(t)} \left[x_i^t - X_t(\tau_i^t)\right]^2 + \lambda \int_0^{d_t} \left[X_t^{(2)}(z)\right]^2 dz \quad \text{w.r.t. } \mathbf{C}\beta \geq \mathbf{0} \tag{4.2}$$

While the linear restrictions $\mathbf{C}$ onto β ensure monotonicity of the resulting functions, the penalty additionally smooths the step curves which guarantees differentiability that is needed by the method of Liebl and Rameseder (2019). The penalty parameter λ is chosen according to Generalized Cross Validation (Meyer (2012), Eq. (3), R-Code can be found at the author's GitHub repository). We

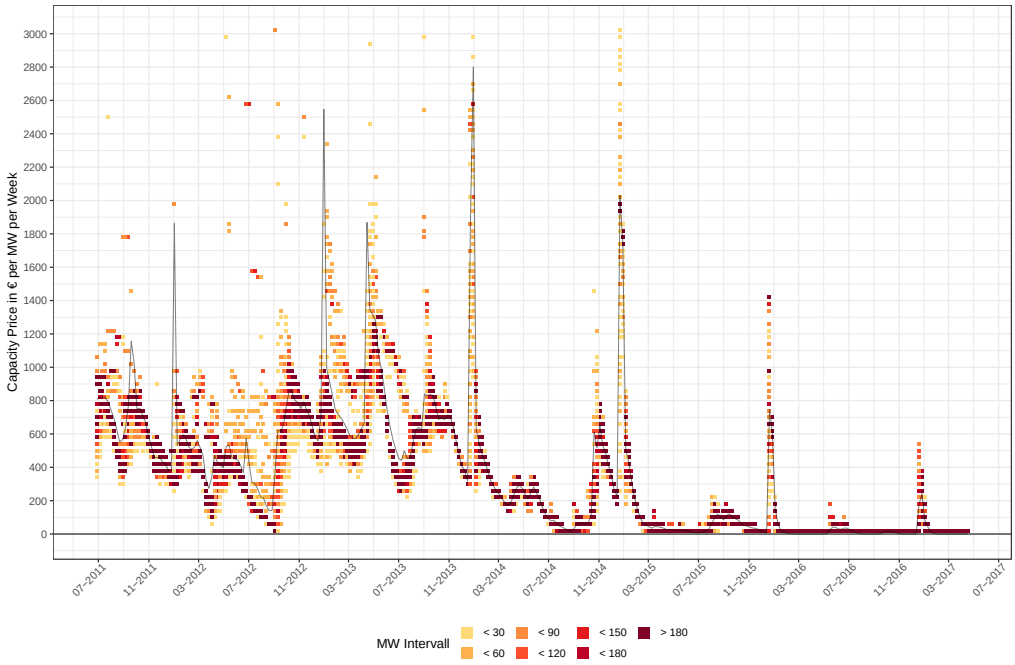

Figure 4.2.1: $T = 304$ discrete and supply curves in NEG HT up to prices of 3,500 Euros per MW per week. The coloring describes the amount of MW—all capacity within a price range of 20 €/MW is accumulated. Darker colors indicate flat supply curves while brighter colors represent steep functions. The black line shows the MW-weighted average of the prices $x_t^{\text{wavg}} = (\sum_{i=1}^{n(t)} x_i^t * \tau_i^t) / \sum_{i=1}^{n(t)} \tau_i^t$ and alludes to atypical behavior when extreme prices above 3,500 Euros were realized.

would like to refer to Ziel (2016) for an alternative LASSO-approach.

4.3 Methodology

4.3.1 Summary

In this paper, we are interested in forecasting the "optimal" price, i.e., the maximum accepted price in the next week (see Figure 4.1.1, right). Mathematically, it is the one-step-ahead prediction $x_{t+1|t}^{\max} = \mathbb{E}[x_{t+1}^{\max} | X_t(\cdot), \dots, X_1(\cdot)]$ of maximum winning prices given all past bid curves to provide bids slightly lower than this estimate—if $x_{t+1|t}^{\max} > x_{t+1}^{\max}$, the bid will not be accepted. Our proposed FDA algorithm at t to predict the price curve at $t+1$ can be described in the following four steps:

1. Step: Calculate the consistent mean estimator of all curves up to t (see 4.3.3).
2. Step: Forecast a univariate gravity point for $t+1$ (see 4.3.4).
3. Step: Build the predicted curve by adding the mean function to the gravity point (see 4.3.5).
4. Step: Predict a price by inserting an accumulated MW position into the curve in $t+1$ (see 4.3.6).

4.3.2 Functional Data Analysis

In the most general form of the FDA framework, a random variable X has realizations $X_1, \ldots, X_T$ in a function space, e.g., a separable Hilbert space $L^2([a,b])$ over a compact domain $[a,b] \subset \mathbb{R}$. Assuming $\mathbb{E}[||X||_2^4] < \infty$, the first two (functional) moments $\mu(\tau) = \mathbb{E}[X(\tau)]$ and $\sigma(\rho,\tau) = \mathbb{E}[(X(\rho) - \mu(\rho))(X(\tau) - \mu(\tau))]$ are well-defined.

4.3.3 Mean function estimation

The missing data mechanism does not allow to use the classical functional mean estimator $\hat{\mu}(t) = n^{-1} \sum_{t=1}^{n} X_t(\tau)$ (see a sketch of the problem in Figure 4.3.1). This is due to the strong dependency of the bidding behavior on the preannounced and deterministic demand d_t and the asymmetric information, i.e., non-winning bids are not published by the GCC. This violation is formalized in Violation (V) of Liebl and Rameseder (2019). We therefore apply the mean function estimation technique proposed by Liebl and Rameseder (2019), who also provide a practical approach to test their underlying assumptions (see both estimators in Figure 4.3.2 where the classical is clearly biased due to the missing data mechanism). The Fundamental Theorem of Calculus (FTC) estimator originally introduced by Liebl and Rameseder (2019), which uses the detour via the well-known theorem, is defined as

$$\widehat{\mu}^{\mathrm{FTC}}(\tau) = \begin{cases} \widehat{\mu}(\tau) & \text{if} \quad \tau \in [0, d_{\min}], \\ \int_{d_{\min}}^{\tau} \widehat{\mu}^{(1)}(z)dz + \widehat{\mu}(d_{\min}) & \text{if} \quad \tau \in (d_{\min}, d_{\max}]. \end{cases} \tag{4.3}$$

In all of our applications, $d_{\min} = 1832$ MW, i.e., the minimum demanded capacity by the FNA over all periods and auctions (see Figure 4.A.2). We can now briefly repeat the adjusted theorem by Liebl and Rameseder (2019):

Theorem 2. *Under our setup, i.e. Violation (V), the estimator $\widehat{\mu}^{\mathrm{FTC}}(t)$ is a pointwise $\sqrt{n}$-consistent estimator of $\mu(t)$ for all $t \in [0, d_{max}]$ where $d_{max} = 2500$ MW for all auctions (see Figure 4.A.2).*

Note that even if the classical estimator is unbiased, the FTC estimator is also unbiased.

We will later use this technique of functional mean estimation for our out-of-sample evaluation of our forecasting performance. This means that for the forecast in $t+1$, we will estimate the FTC $\widehat{\mu}_t^{\mathrm{FTC}}(\tau)$ of the sample of curves $X_1, \ldots, X_t$.

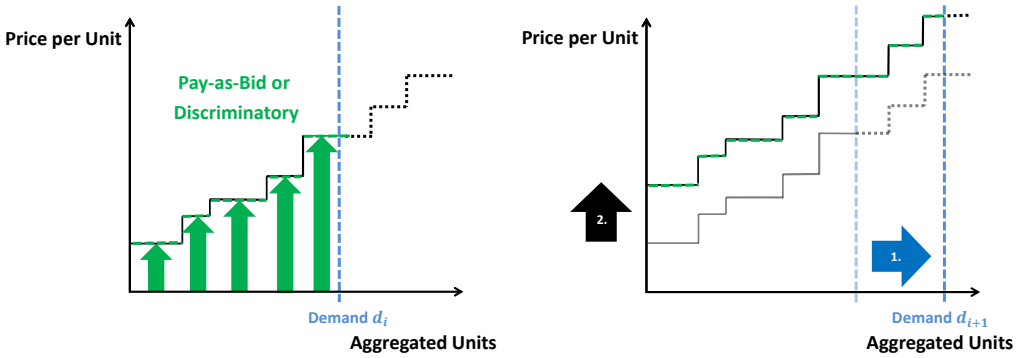

Figure 4.3.1: LEFT: The pay-as-bid pricing mechanism, where each successful bidder receives the price of his (solid) bid, are published by the FNA. RIGHT: The pre-announced increased amount of reserve capacity d_t (1.) leads to a vertical shift dependency of the price curves (2.) where the additional (solid) winning bids are published, whereas the dotted predecessing were not.

4.3.4 Univariate forecasting of gravity points

The definition of the FTC estimator in Equation 4.3 involves the antiderivative which is unique up to a constant. To gain a representative reconstruction of the bidding function, these constants, i.e., points where the mean function is hung up, need to be chosen. We call these constants gravity points in the following.

We evaluate a large set of models where a model is a combination of a gravity point with a set of exogenous variables. The set of gravity points G consists of

G_1 the minimum capacity price $x_t^{\min} = X_t(0)$,

G_2 the maximum capacity price $x_t^{\max} = X_t(d_t)$,

G_3 the median capacity price x_t^{med},

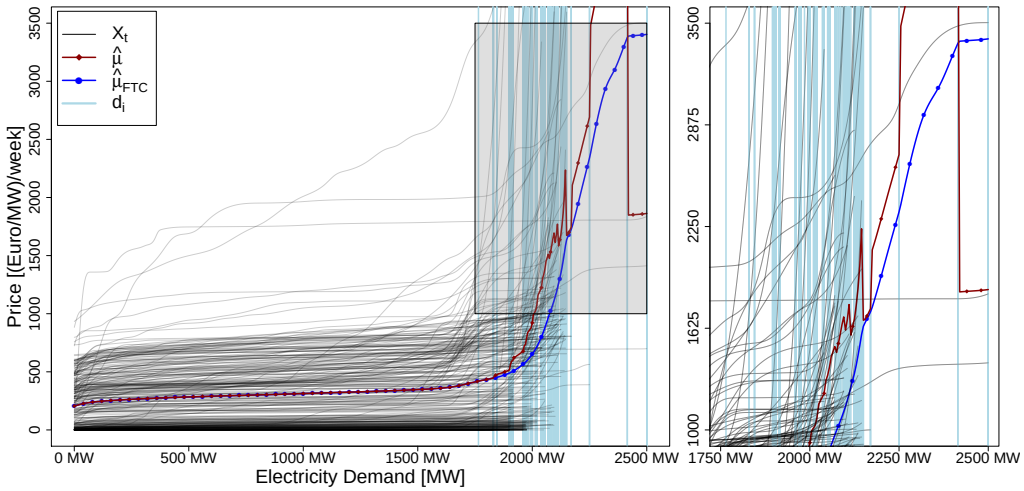

Figure 4.3.2: Classical mean (solid with triangles) and FTC mean estimates (solid with circles) computed from $n = 304$ partially observed price curves in NEG HT.

G_4 the capacity-weighted average price

$$w_t = (\sum_{i=1}^{n(t)} x_i^t * \tau_i^t) / \sum_{i=1}^{n(t)} \tau_i^t = (\int_0^{d_t} X_t(\tau) d\tau)/d_t, \text{ and}$$

G_5 exemplary MW positions $\tau_p \in [0, d_{\min}]$, i.e., $x_t^{\mathrm{p}} = X_t(\tau_p)$. Here, we use $\tau_p = 1250$ MW.

As a set of exogenous variables we use different combinations E of

E_1 the demanded capacity d_t,

E_2 a binary variable for the two Christmas weeks (see Figure 4.A.2)

E_3 the spread $x_t^{\max} - x_t^{\min}$.

Therefore, a model m directly relates to a combination of G and E, i.e., $m \in G \times E$.

For the prediction with model $m = (G_i, E_j), i = 1, \ldots, 5, j = 1, \ldots, 3$, we use the auto.arima function of R (see Hyndman and Khandakar (2008)), i.e.,

$$x_t^m \sim \text{ARIMAx(p,d,q)}. \tag{4.4}$$

The dependent variable is the gravity point G_i given a regressor combination E_j. We draw one-step mean predictions, i.e., $x_{t+1|t}^m = \mathbb{E}[x_{t+1}^m | X_t(\cdot), \ldots, X_1(\cdot)]$.

4.3.5 Building up whole bidding functions

To provide a forecast value of our supply curve depending on the model m, we combine the estimate of the gravity point $x^m_{t+1|t}$ with the estimate of the unconditional mean by centering the curve around this point, i.e.,

$$X_{t+1|t}(\tau) = x^m_{t+1|t} + \widehat{\mu}^{\mathrm{FTC}}_t(\tau) - s(\widehat{\mu}^{\mathrm{FTC}}_t(\tau)) \tag{4.5}$$

where the last term accounts for the centering by subtracting the statistic of the mean function estimate, e.g., in case of G_1 with the minimum statistic: $s(\widehat{\mu}^{\mathrm{FTC}}_t(\tau)) = \widehat{\mu}^{\mathrm{FTC}}_t(0)$

4.3.6 Choice of the accumulated MW position

Providing a bid means choosing a $\tau^\star$ and evaluating $X_{t+1|t}(\tau^\star)$, which represents the predicted price at the accumulated MW position $\tau^\star$. The bid is a winning bid if $X_{t+1|t}(\tau^\star) < x^{\max}_{t+1}$. Since $X_{t+1|t}$ is monotonically increasing, i.e., $\tau_1 > \tau_2$ implies $X_{t+1|t}(\tau_1) > X_{t+1|t}(\tau_2)$, selecting a larger $\tau^\star \in [0, d_{t+1}]$ means a higher price at the risk of exceeding the maximum accepted price $x^{\max}_{t+1}$ and therefore dropping out of the auction.

To formalize this, we define $p_{t+1} \in [0, 1]$. For each week, we will choose $\tau^\star(p)$ as the percentage of the predetermined demand in $t + 1$, i.e., $\tau^\star_{t+1}(p) = p \cdot d_{t+1}$ (see Figure 4.A.2, left). We talked to three different SR prequalified bidders and all are risk averse in the sense that they typically aim at the lower half of the bidding curve. We therefore use $p = 40\%$, i.e., $\tau^\star_{t+1} = 0.4 \cdot d_{t+1}$ as the choice of the accumulated MW position for our model.

4.3.7 Metrics

Standard forecast performance measures, e.g., MSEP or MAE, assume symmetric loss functions of the forecast consumer. When forecasting highly inhomogeneous value domains, e.g., in retail sales, MAPE and its relatives are well-established metrics. However, our bidder faces the following problem:

- If the forecast is above the maximum winning bid, i.e., $X_{t+1|t}(\tau^\star) > x^{\max}_{t+1} \Leftrightarrow Z_{t+1}(X_{t+1|t}(\tau^\star)) = 0$, the bidder does not need to provide his capacity because cheaper bidders provide the full demand d_t—the bidder leaves the auction empty-handed.

- If the forecast is below the maximum winning bid, i.e., $X_{t+1|t}(\tau^\star) < x^{\max}_{t+1} \Leftrightarrow Z_{t+1}(X_{t+1|t}(\tau^\star)) = 1$, the bidder receives the fee for which he provides his capacity.

Most of the bidders therefore evaluate their forecasting strategies in two KPIs:

- **Acceptance Rate (AR)** of model m over horizon H:

$$AR(m;H) = \sum_{h=1}^{H} Z_{t+h}(x^m_{t+h|t+h-1})/H$$

- **Performance Rate (PR):**

$$PR(m;H) = \sum_{h=1}^{H} Z_{t+h}(x^m_{t+h|t+h-1})x^m_{t+h|t+h-1} / \sum_{h=1}^{H} w_{t+h}$$

While $AR(m;H) \in [0,1]$ counts the number of winning bids, e.g., $AR(m;H) = 1$ constitutes an always-winning model, the performance $PR(m;H) \geq 0$ relates the model performance to the weighted average. $PR(m;H) > 1$ signifies that the model performs on average better than the realized weighted averages. Note that due to the information asymmetry, the weighted average is censored by only taking winning bids into account. This means that all non-winning bids with return zero would highly decrease the denominator in $PR(m;H)$ and therefore increase the performance, i.e., a $PR(m;H) = 1$ means having a strategy better than the average.

4.3.8 Benchmark models

We consider forecasting algorithms typically used in the literature for benchmarking.

- Naive Model (NM)
  We use the naive forecast $x^m_{t+1|t} = x^m_t$ where the forecast equals the gravity point G_j from the last observed auction in t.

- Automated ARIMA (AA)
  There is a huge amount of software providing automated identification of ARIMA models from which Hyndman and Khandakar (2008) provides the most prominent open-source algorithm. While we forecast the gravity points to build up whole predictions of the bidding functions, we also use the univariate predictions of gravity points as benchmarks (see Section 4.3.4).

- Prophet (PR)
  As a serious competitor we use Facebooks Prophet (see Taylor and Letham

(2018)), a state-of-the-art automated forecasting algorithm. Prophet belongs to the class of Generalized Additive Models (GAMs) (see Hastie and Tibshirani (1987)), which decompose the time series into deterministic trend, periodicity, and irregularities. In the case of Prophet, time is the only regressor. Prophet is fit by L-BFGS (see Byrd et al. (1995)) and has some advantages over ARIMA; in particular, it is more flexible in terms of multiple periods which is used to model the irregularly periodic behavior in the bidding curves (see Figure 4.A.3). As different benchmarking models, we apply Prophet onto the univariate time series of the gravity points G.

4.4 Application and results

Table 4.4.1 provides the forecast performance measured in the metrics of Section 4.3.7 for the competitor models presented in Section 4.3.8 and our proposed FDA model. We have chosen to use "2015-05-25" to "2017-04-17" as our out-of-sample horizon, i.e., $H = 100$. Although the table shows only the forecast results for NEG HT, similarly positive results for the other auctions can be easily reproduced.

Table 4.4.1: Acceptance Rate (AR) in % and Performance Rate (PR) in % in NEG HT.

		FDA		NM		PR		AA	
G_i	E_j	AC	PR	AC	PR	AC	PR	AC	PR
1	1,2	99	111	55	23	67	59	69	72
2	1,2	92	109	52	21	72	62	72	79
3	1,2	100	114	63	58	81	78	76	69
4	1,2	97	102	60	57	83	77	79	77
5	1,2	95	99	60	60	80	73	75	73
1	1,2,3	92	89			69	42	57	55
2	1,2,3	95	91			70	46	62	58
3	1,2,3	98	90			78	55	65	60
4	1,2,3	92	92			79	59	58	59
5	1,2,3	91	85			72	56	62	65

While the naive model NM is not able to capture the dynamics in time or the dependency of the demand at all, the automated univariate models perform better and are comparable. However, none of them is close to the high Accep-

tance Rate and Performance Rate of the FDA model perfectly suited for such data. Note that $\sum_{t=1}^{T} w_t \approx 110,000\,€/\text{MW}$ in NEG HT, i.e., the difference of FDA in $(G_3, E_{1,2})$ to its closed competitor AA in $(G_2, E_{1,2})$ is approximately $38,500\,€/\text{MW}$.

4.5 Conclusion

The present research aims to establish a flexible forecasting approach for discriminatory multi-unit auctions with an underlying missing data mechanism and to show its superior performance compared to classical models. The forecasting performance justifies the model, which also can be used for other types of analyses. Although the market seems exotic, the assumptions behind the modeling approach are not very restrictive and can be easily generalized, e.g., to all kinds of bid or ask curves.

For further research, we recommend two points: (1) other strategies behind the choice of the accumulated MW position $\tau^\star$ can be investigated. This directly relates to (2): we do not incorporate variable or fixed costs of providing capacity due to the fact that the German Market consists of two auctions, one for capacity reserve and the other for balancing energy. To implement these costs in the forecasts decisions, we need to model both markets at once. This is, however, common practice within the prequalified bidders we have talked to.

In conclusion, current research evaluates potential designs of capacity reserve and balancing energy markets, in particular auctions. This includes assumptions about bidding behavior and pricing mechanisms. While the bidding behavior in uniform price auctions is empirically well-understood, the present research provides an approach for the empirical modeling in the case of discriminatory pricing. The goal behind capacity reserve and balancing energy markets is to obtain a very high level of grid stability at justifiable costs.

4.A Appendix

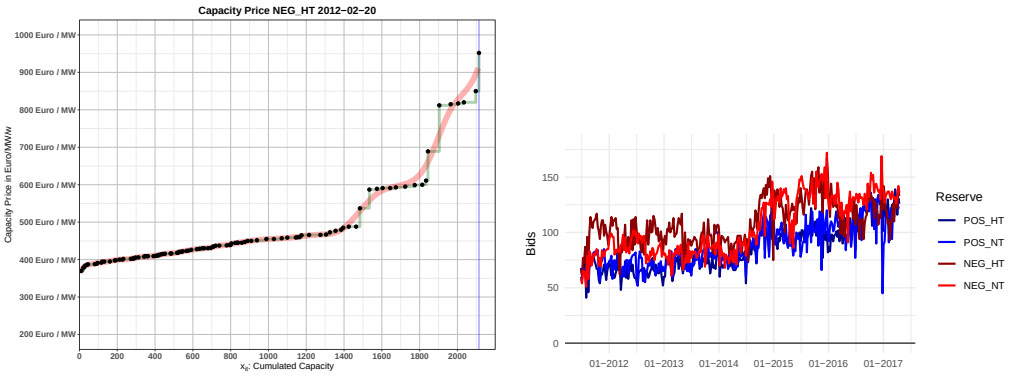

Figure 4.A.1: Left: Winning raw bids (dots), resulting step curve (green), and P-spline (red) of the "NEG_HT" auction on 2012-02-20. Right: Amount of winning bids per week over time.

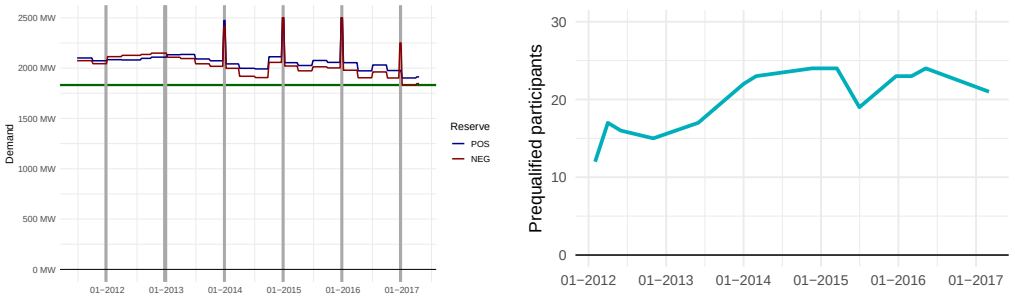

Figure 4.A.2: Left: Time series of $d_{\max}$ for SR POS and NEG. Right: Amount of prequalified bidders over time.

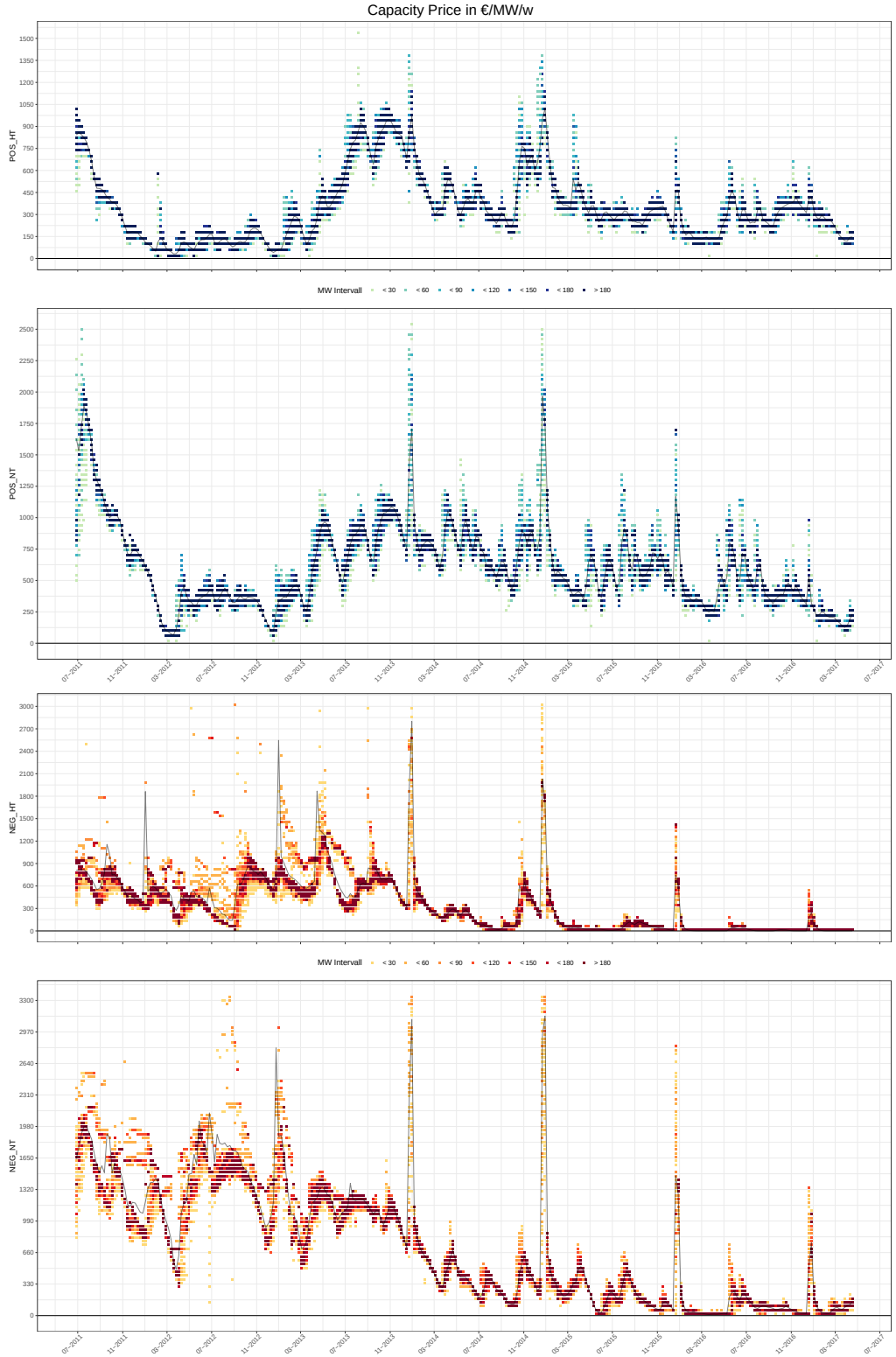

Figure 4.A.3: $T = 304$ discrete supply curves for all four auctions “POS_HT", “POS_NT", “NEG_HT", and “NEG_NT"—all capped at maximum prices of 1,500, 2,500, 3,000, and 3,300 Euros per MW per week. Weighted average prices are drawn as black lines.

Conclusion and outlook

The present thesis aims to contribute to the literature on functional data analysis (FDA), in particular the prediction challenge mentioned in the beginning of the Introduction 1.1. It demonstrates that FDA provides a beneficial toolkit for empirical questions in relation to auctions by applying and extending FDA methods. All the chapters of this thesis deal with real auction data collected and preprocessed by the author. While analyzing the data, fundamental problems related to the applicability and interpretability are generalized and investigated. In this context and always motivated by real data, the thesis makes empirical but also methodological contributions.

On the methodological side, the thesis contributes to the existing literature by improving the estimation procedure for the functional regression model with points of impact, an augmented scalar-on-function regression model that can be interpreted more easily. Another methodological contribution is a generalized mean and covariance estimator for non-randomly missing data and a test procedure to test for applicability. The empirical contribution addresses a beneficial forecast approach for partially observable supply functions in pay-as-bid multi-unit auctions, which allows for strategically optimal bids.

In the first paper (see Chapter 2.1) an improved estimation algorithm for the functional regression model with points of impact is developed. This model can be used to decompose the effect of a functional predictor on a scalar outcome variable into the classical global component and a time-specific local component. Motivated by a problem involving real data, where the dependency between yearly trajectories of daily impressions and aggregated clicks is modeled, the paper proposes an improved estimation algorithm taking into account that in finite samples, there is an ambiguity between global and local effects. More specifically, the proposed procedure decouples the estimation of the model parameters and the selection of points of impact. Additionally, instead of the classical FPCA estimator, our procedure uses a smoothing spline estimator for the functional coefficient. An extensive simulation study shows a substantial improvement in terms of the precision of the estimation.

The second paper (see Chapter 3.1) is motivated by the mean and covariance estimation in an auction market with partially observable supply functions. Classical FDA mean and covariance estimators lead to *obviously* inconsistent results due to non-randomly missing data caused by strategic bidding behavior. Our generalizations utilize the nature of FDA by using its properties such as continuity and differentiability in the form of the fundamental theorem of calculus: by using classical estimators onto the derivatives and reintegrating over the partially observable domains, certain dependencies can be bypassed, which leads to consistent estimators under certain assumptions. A procedure to test for the assumption is successfully developed and extensively evaluated in a simulation study. The paper provides useful estimates for mean and covariance functions that serve substantially in the prediction method proposed in the last paper.

The third paper (see Chapter 4.1) addresses the predictions of the strategically optimal maximum price in pay-as-bid multi-unit auctions with partially observable supply curves. The forecasting procedure models and evaluates the predictions of whole auction curves since functional alternatives are not applicable due to the missing data mechanism. The predictions are evaluated in four auctions in the German Capacity Reserve and Balance Power Market (GCR) and outperform classical methods due to the nature of the setup: non-randomly missing data and quantile predictions. The method was used from end of 2017 to 2019. Unfortunately, at the time of this writing, the paper is still under review.

Still obsessed with the challenge of the Introduction 1.1, the author is keen to formulate a more general tool box to predict arbitrary multi-unit auctions. This includes the following extensions for future research:

Empirically, there is a significant amount of multi-unit auctions with data available to the author, e.g., the primary and tertiary reserve in the GCR. There are also other markets, e.g., the fragmented US energy market provides an enormous set of different grid reliability mechanisms, e.g., New York (NYISO), New England (ISO-NE), Midwestern ISO, Ontario (IESO), Pennsylvania, New Jersey, and Maryland (PJM), Texas (ERCOT), ..., with open available data.

Methodologically, there are a lot of opportunities to generalize the prediction approach even further. From a prediction perspective, the choice of the gravity point, the choice of the position $\tau^{\star}_{t+1}$ (see Section 4.3.6), and the approach if $d_{t+1} > d = \max\{d_1, \ldots, d_t\}$ should be further investigated. In addition, the author envisions incorporating functional dependencies in the form of the functional regression with PoIs and other exogenous regressors into the prediction approach, e.g., the influence of past bidding shapes and EEX price signals onto the shape of the supply functions. Since some of the US market mechanisms also involve different non-randomly missing data mechanisms, the author also asks the question

of whether for a specific set of missing data mechanisms, there could be basis representations $X(t) = \sum_{j\geq 1} \xi_j \psi_j(t)$ to follow similar approaches as proposed in the second paper. Econometrically, the connection of FDA with structural parameters can be established and discussed, e.g., the analysis of market power, which is a significant topic in restructured electricity markets.

Acknowledgments

Firstly, I want to express my gratitude to Prof. Rolf Tschernig. Despite missing an econometric background, he believed in me and provided me the opportunity to work on a doctoral thesis in this field. During my time at the Chair of Econometrics, he created a diverse environment to stimulate, combine, and support new ideas across universities, scientific fields, and cultures; his high standards have made me a better human being.

I am also utterly grateful to Prof. Dominik Liebl. His holistic knowledge and interest, not only in Functional Data Analysis, provided a wonderful atmosphere of support, sparring, and collaboration including drive and focus to streamline the fast prototyping of new ideas; for me, this is the up-to-date definition of the perfect scientific environment. I also like to thank Christoph Rust: his hands-on mentality in no matter what type of technology or algorithm was always one of the strongest motivators in an economy department.

Many colleagues and fellow Ph.D. students at the University of Regensburg have substantially supported my work, not only by providing invaluable and insightful comments but also by being a source of motivation. I deeply appreciate to have worked with Binh Nguyen, Ben Booker, Roland Weigand, Kathrin Kagerer, Joachim Schnurbus, and Stephan Huber.

Finally, I am profoundly grateful to my family, in particular my grandfater Josef, who, unfortunately, is not able to see the final dissertation. Without their understanding, patience, and unwavering support all this would not have been possible.

Bibliography

Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing 16*(5), 1190–1208.

Cardot, H., C. Crambes, A. Kneip, and P. Sarda (2007). Smoothing splines estimators in functional linear regression with errors-in-variables. *Computational Statistics & Data Analysis 51*(10), 4832–4848.

Cardot, H., A. Mas, and P. Sarda (2007). CLT in functional linear regression models. *Probability Theory and Related Fields 138*(3-4), 325–361.

Chiou, J.-M. (2012). Dynamical functional prediction and classification, with application to traffic flow prediction. *The Annals of Applied Statistics 6*(4), 1588–1614.

Choi, H. and M. Reimherr (2018). A geometric approach to confidence regions and bands for functional parameters. *Journal of the Royal Statistical Society: Series B 80*(1), 239–260.

Crambes, C., A. Kneip, and P. Sarda (2009). Smoothing splines estimators for functional linear regression. *The Annals of Statistics 37*(1), 35–72.

De Boor, C. (1978). *A practical guide to splines*, Volume 27. Springer.

Delaigle, A. and P. Hall (2013). Classification using censored functional data. *Journal of the American Statistical Association 108*(504), 1269–1283.

Delaigle, A. and P. Hall (2016). Approximating fragmented functional data by segments of Markov chains. *Biometrika 103*(4), 779–799.

Doty, D., K. Sruoginis, and D. Silverman (2021). IAB/PwC Internet advertising revenue report. https://www.iab.com/wp-content/uploads/2022/04/IAB_Internet_Advertising_Revenue_Report_Full_Year_2021.pdf.

Engelmann, D. and V. Grimm (2009). Bidding behaviour in multi-unit auctions–an experimental investigation. *The Economic Journal 119*(537), 855–882.

Ferraty, F., P. Hall, and P. Vieu (2010). Most-predictive design points for functional data predictors. *Biometrika 97*(4), 807–824.

Ferraty, F. and P. Vieu (2006). *Nonparametric Functional Data Analysis - Theory and Practice*. New York: Spinger.

Fraiman, R., Y. Gimenez, and M. Svarc (2016). Feature selection for functional data. *Journal of Multivariate Analysis 146*, 191–208.

Fraley, C. and A. E. Raftery (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association 97*(458), 611–631.

Geddes, B. (2014). *Advanced Google AdWords*. John Wiley & Sons.

Gellar, J. E., E. Colantuoni, D. M. Needham, and C. M. Crainiceanu (2014a). Variable-domain functional regression for modeling ICU data. *Journal of the American Statistical Association 109*(508), 1425–1439.

Gellar, J. E., E. Colantuoni, D. M. Needham, and C. M. Crainiceanu (2014b). Variable-domain functional regression for modeling ICU data. *Journal of the American Statistical Association 109*(508), 1425–1439.

Goldberg, Y., Y. Ritov, and A. Mandelbaum (2014). Predicting the continuation of a function with applications to call center data. *Journal of Statistical Planning and Inference 147*, 53–65.

Goldsmith, J., J. Bobb, C. M. Crainiceanu, B. Caffo, and D. Reich (2010). Penalized functional regression. *Journal of Computational and Graphical Statistics 20*(4), 830–851.

Goldsmith, J., C. M. Crainiceanu, B. Caffo, and D. Reich (2012). Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 61*(3), 453–469.

Grimm, V., A. Ockenfels, and G. Zoettl (2008). Strommarktdesign: Zur Ausgestaltung der Auktionsregeln an der EEX. *Zeitschrift für Energiewirtschaft 32*(3), 147–161.

Gromenko, O., P. Kokoszka, and J. Sojka (2017). Evaluation of the cooling trend in the ionosphere using functional regression with incomplete curves. *The Annals of Applied Statistics 11*(2), 898–918.

Hall, P. and J. L. Horowitz (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics 35*(1), 70–91.

Hastie, T. and R. Tibshirani (1987). Generalized additive models: some applications. *Journal of the American Statistical Association 82*(398), 371–386.

Hastie, T. J. and R. J. Tibshirani (1990). *Generalized additive models*, Volume 43. CRC press.

Hendricks, K. and H. J. Paarsch (1995). A survey of recent empirical work concerning auctions. *Canadian Journal of Economics 28*(2), 403–426.

Hortaçsu, A. (2011). Recent progress in the empirical analysis of multi-unit auctions. *International Journal of Industrial Organization 29*(3), 345–349.

Horváth, L. and P. Kokoszka (2012). *Inference for Functional Data with Applications*. Springer.

Hsing, T. and R. Eubank (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. John Wiley & Sons.

Hyndman, R. J. and Y. Khandakar (2008). Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software 27*(3), 122.

James, G. M., T. J. Hastie, and C. A. Sugar (2000). Principal component models for sparse functional data. *Biometrika 87*(3), 587–602.

James, G. M. and C. A. Sugar (2003). Clustering for sparsely sampled functional data. *Journal of the American Statistical Association 98*(462), 397–408.

Just, S. (2015). The german market for system reserve capacity and balancing. *EWL Working Paper 2015*(6), 1–38.

Kneip, A. and D. Liebl (2019). On the optimal reconstruction of partially observed functional data.

Kneip, A., D. Poss, and P. Sarda (2016). Functional linear regression with points of impact. *The Annals of Statistics 44*(1), 1–30.

Koeppe, R., J. Zhu, B. Nan, and X. Wang (2014). Regularized 3d functional regression for brain image data via haar wavelets. *The Annals of Applied Statistics 8*(2), 1045–1064.

Kraus, D. (2015). Components and completion of partially observed functional data. *Journal of the Royal Statistical Society 77*(4), 777–801.

Krishna, V. (2009). *Auction theory*. Academic press.

Kulakov, S. (2020). X-model: Further development and possible modifications. *Forecasting 2*(1), 20–35.

Kwasnica, A. M. and K. Sherstyuk (2013). Multiunit auctions. *Journal of economic surveys 27*(3), 461–490.

Liebl, D. (2013). Modeling and forecasting electricity spot prices: A functional data perspective. *The Annals of Applied Statistics 7*(3), 1562–1592.

Liebl, D. and S. Rameseder (2019). Partially observed functional data: The case of systematically missing parts. *Computational Statistics & Data Analysis 131*, 104–115.

Liebl, D., S. Rameseder, and C. Rust (2020). Improving estimation in functional linear regression with points of impact: Insights into google adwords. *Journal of Computational and Graphical Statistics 29*(4), 814–826.

Lindquist, M. A. and I. W. McKeague (2009). Logistic regression with brownian-like predictors. *Journal of the American Statistical Association 104*(488), 1575–1585.

Little, R. J. and D. B. Rubin (2014). *Statistical analysis with missing data*. John Wiley & Sons.

Liu, B. and H.-G. Müller (2008). *Functional data analysis for sparse auction data*, pp. 269–290. Statistical Methods in eCommerce Research (eds W. Jank and G. Shmueli) John Wiley & Sons.

Lucking-Reiley, D. (1999). Using field experiments to test equivalence between auction formats: Magic on the internet. *American Economic Review 89*(5), 1063–1080.

Maronna, R. A. and V. J. Yohai (2013). Robust functional linear regression based on splines. *Computational Statistics & Data Analysis 65*, 46–55.

Matsui, H. and S. Konishi (2011). Variable selection for functional regression models via the L1 regularization. *Computational Statistics & Data Analysis 55*(12), 3304–3310.

McKeague, I. W. and B. Sen (2010). Fractals with point impact in functional linear regression. *The Annals of Statistics 38*(4), 2559–2586.

Meyer, M. C. (2012). Constrained penalized splines. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique 40*(1), 190–206.

Ocker, F., K.-M. Ehrhart, and M. Ott (2015). An economic analysis of the german secondary balancing power market.

Paarsch, H. J., H. Hong, et al. (2006). *An introduction to the structural econometrics of auction data*. The MIT Press.

Poß, D., D. Liebl, A. Kneip, H. Eisenbarth, T. D. Wager, and L. Feldman Barrett (2020). Super-consistent estimation of points of impact in nonparametric regression with functional predictors.

Ramsay, J. (1982, December). When the data are functions. *Psychometrika 47*(4), 379–396.

Ramsay, J. and B. Silverman (2005). *Functional Data Analysis* (2. ed.). New York: Springer.

Ramsay, J. O., H. Wickham, S. Graves, and G. Hooker (2014). fda: Functional data analysis. R package version 2.4.4.

R Core Team (2017a). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

R Core Team (2017b). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reddy, S. K. and M. Dass (2006). Modeling on-line art auction dynamics using functional data analysis. *Statistical Science 21*(2), 179–193.

Reiss, P. T., J. Goldsmith, H. L. Shang, and R. T. Ogden (2017). Methods for scalar-on-function regression. *International Statistical Review 85*(2), 228–249.

Romano, J. P. and M. Wolf (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association 100*(469), 94–108.

Shmueli, G. and W. Jank (2005). Visualizing online auctions. *Journal of Computational and Graphical Statistics 14*(2), 299–319.

Taylor, S. J. and B. Letham (2018). Forecasting at scale. *The American Statistician 72*(1), 37–45.

Torrecilla, J. L., J. R. Berrendero, and A. Cuevas (2016). Variable selection in functional data classification: a maxima-hunting proposal. *Statistica Sinica 26*(2), 619–638.

Wang, S., W. Jank, and G. Shmueli (2008). Explaining and forecasting online auction prices and their dynamics using functional data analysis. *Journal of Business & Economic Statistics 26*(2), 144–160.

Wang, S., W. Jank, G. Shmueli, and P. Smith (2008). Modeling price dynamics in eBay auctions using differential equations. *Journal of the American Statistical Association 103*(483), 1100–1118.

Weron, R. (2007). *Modeling and forecasting electricity loads and prices: A statistical approach*, Volume 403. John Wiley & Sons.

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.

Yao, F., H.-G. Müller, and J.-L. Wang (2005a). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association 100*(470), 577–590.

Yao, F., H.-G. Müller, and J.-L. Wang (2005b). Functional linear regression analysis for longitudinal data. *The Annals of Statistics 33*(6), 2873–2903.

Zhang, S., W. Jank, and G. Shmueli (2010). Real-time forecasting of online auctions via functional k-nearest neighbors. *International Journal of Forecasting 26*(4), 666–683.

Zhou, R. R., N. Serban, N. Gebraeel, and H.-G. Müller (2014). A functional time warping approach to modeling and monitoring truncated degradation signals. *Technometrics 56*(1), 67–77.

Ziel, F. (2016). Forecasting electricity spot prices using lasso: On capturing the autoregressive intraday structure. *IEEE Transactions on Power Systems 31*(6), 4977–4987.

Ziel, F. and R. Steinert (2016). Electricity price forecasting using sale and purchase curves: The x-model. *Energy Economics 59*, 435–454.

Ziel, F. and R. Weron (2018). Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Economics 70*, 396–420.