

Partially Observed Functional Data: The Case of Systematically Missing Parts

Dominik Liebl^{*a} and Stefan Rameseder^b

^a*Department of Statistics, University of Bonn*

^b*Department of Econometrics, University of Regensburg*

August 1, 2018

Abstract

New estimators for the mean and the covariance function for partially observed functional data are proposed 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. The requirements of the estimation procedure can be tested using a sequential multiple hypothesis test procedure. An extensive simulation study compares the new estimators with the classical estimators from the literature in different missing data scenarios. The proposed methodology is motivated by the practical problem of estimating the mean price curve in the German Control Reserve Market. In this auction market, price curves are only partially observable and the underlying missing data mechanism depends on systematic trading strategies which clearly violate the missing-completely-at-random assumption. In contrast to the classical estimators, the new estimators lead to useful estimates of the mean and covariance functions. Supplementary materials are provided online¹.

Keywords. Functional data analysis, missing data, fundamental theorem of calculus.

^{*}Corresponding author: Dominik Liebl (dliebl@uni-bonn.de), Department of Statistics, University of Bonn, Adenauerallee 24-42, 53113 Bonn, Germany

¹Supplementary materials: R-package **PartiallyFD** and R-scripts for reproducing the simulation study and the real data application.

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. (2014) and Gromenko et al. (2017) propose a functional regression model for incomplete curves. Liebl (2013) models and forecasts partially observed price functions and Liebl and Kneip (2018) 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). 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 introducing 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, we propose a practical approach for testing the considered violations of the MCAR assumption. Section 4 contains our simulation study and the application is found in Section 5. Proofs and further derivations can be found in Appendix A.

2 Methodology

2.1 General Setup

We consider an i.i.d. sample of differentiable random functions X_1, \dots, 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 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

$\mathfrak{B} = \{\psi_1, \psi_2, \dots\}$ 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. Common 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 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 . For a start, we consider the following Violation (V) of the MCAR assumption:

(V) $X \not\perp O$, but $X^{S^\perp} \perp O$; 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 2.3, we introduce more general versions of Violation (V) and discuss how to apply our estimation strategies accordingly.

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:

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

$$\hat{\sigma}(s, t) = \frac{I_2(s, t)}{\sum_{i=1}^n U_i(s, t)} \sum_{i=1}^n U_i(s, t) [X_i(s) - \hat{\mu}(s)] [X_i(t) - \hat{\mu}(t)], \quad (2)$$

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 $\hat{\mu}(t)$ 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 $\hat{\sigma}(s, t)$ 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

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

with $|\Delta(t)| > 0$, i.e., $\mathbb{E}[\xi_{i1} | O_i(t)] \neq \mathbb{E}[\xi_{i1}]$, iff $p(t) < 1$; an essentially equivalent bias expression applies to $\hat{\sigma}(s, t)$ (see also Proposition A.1 in Appendix 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 $\hat{\mu}_{\text{FTC}}(t)$ and $\hat{\sigma}_{\text{FTC}}(s, t)$ for the mean and covariance functions.

Definition 2.1 (FTC-mean) *The FTC mean estimator is defined as*

$$\hat{\mu}_{\text{FTC}}(t) = \begin{cases} \hat{\mu}(t) & \text{if } t \in [0, d_{\min}] \\ \int_{d_{\min}}^t \hat{\mu}^{(1)}(z)dz + \hat{\mu}(d_{\min}) & \text{if } t \in (d_{\min}, 1], \end{cases} \quad (3)$$

where

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

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

Definition 2.2 (FTC-covariance) *The FTC covariance estimator is defined as*

$$\hat{\sigma}_{\text{FTC}}(s, t) = \begin{cases} \hat{\sigma}(s, t) & \text{if } (s, t) \in [0, d_{\min}]^2 \\ \int_{d_{\min}}^t \hat{\sigma}^{(0,1)}(s, z_2)dz_2 + \hat{\sigma}(s, d_{\min}) & \text{if } (s, t) \in [0, d_{\min}] \times (d_{\min}, 1] \\ \int_{d_{\min}}^s \hat{\sigma}^{(1,0)}(z_1, t)dz_1 + \hat{\sigma}(d_{\min}, t) & \text{if } (s, t) \in (d_{\min}, 1] \times [0, d_{\min}] \\ \int_{d_{\min}}^t \int_{d_{\min}}^s \hat{\sigma}^{(1,1)}(z_1, z_2)dz_1dz_2 + \int_{d_{\min}}^s \hat{\sigma}^{(1,0)}(z_1, d_{\min})dz_1 + \int_{d_{\min}}^t \hat{\sigma}^{(0,1)}(d_{\min}, z_2)dz_2 + \hat{\sigma}(d_{\min}, d_{\min}) & \text{if } (s, t) \in (d_{\min}, 1]^2, \end{cases} \quad (4)$$

where

$$\hat{\sigma}^{(\ell, k)}(s, t) = \frac{I_2(s, t)}{\sum_{i=1}^n U_i(s, t)} \sum_{i=1}^n U_i(s, t) \left[X_i^{(\ell)}(s) - \hat{\mu}^{(\ell)}(s) \right] \left[X_i^{(k)}(t) - \hat{\mu}^{(k)}(t) \right], \quad \ell, k \in \{0, 1\}.$$

A step-by-step derivation of $\hat{\sigma}_{\text{FTC}}(s, t)$ can be found in parts I-III of the proof of Theorem 2.1 (see Appendix A). The following theorem states the consistency of the FTC-estimators under Violation (V) of the MCAR assumption:

Theorem 2.1 *Under our setup and under Violation (V) of the MCAR assumption, the estimators $\hat{\mu}_{\text{FTC}}(t)$, defined in (3), and $\hat{\sigma}_{\text{FTC}}(s, t)$, defined in (4), are pointwise \sqrt{n} -consistent estimators of $\mu(t)$ and $\sigma(s, t)$ for all $s, t \in [0, 1]$.*

2.3 Generalizations

Motivated by our real data application, we considered so far 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 $d_f \in [0, 1]$, i.e., $p(d_f) = 1$. By noting $\int_a^b = -\int_b^a$, the generalized FTC mean estimator is defined as

$$\hat{\mu}_{\text{FTC}}^*(t) = \begin{cases} \int_{d_f}^t \hat{\mu}^{(1)}(z) dz + \hat{\mu}(d_f) & \text{for all } t > d_f \\ \hat{\mu}(d_f) & \text{for } t = d_f \\ -\int_t^{d_f} \hat{\mu}^{(1)}(z) dz + \hat{\mu}(d_f) & \text{for all } t < d_f. \end{cases} \quad (5)$$

Correspondingly, the generalized FTC covariance estimator is defined as

$$\begin{aligned} \hat{\sigma}_{\text{FTC}}^*(s, t) &= \\ &= \int_{d_f}^t \int_{d_f}^s \hat{\sigma}^{(1,1)}(z_1, z_2) dz_1 dz_2 + \int_{d_f}^s \hat{\sigma}^{(1,0)}(z_1, d_f) dz_1 + \int_{d_f}^t \hat{\sigma}^{(0,1)}(d_f, z_2) dz_2 + \hat{\sigma}(d_f, d_f) \end{aligned} \quad (6)$$

for all $s, t > d_f$, where for cases with $s < d_f$ and/or $t < d_f$ one needs to replace “ $\int_{d_f}^s$ ” with “ $-\int_s^{d_f}$ ” and/or “ $\int_{d_f}^t$ ” by “ $-\int_t^{d_f}$ ”, and where $\hat{\sigma}_{\text{FTC}}^*(s, t) = \hat{\sigma}(d_f, d_f)$ for $s = t = d_f$. The estimators in (5) and (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 $S_K = \text{span}\{\psi_1, \dots, \psi_K\}$ with monomial basis functions $\psi_j(t) = t^{j-1}$ for $j = 1, \dots, K$ 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) $X \not\perp O$, but $X^{S_K^\perp} \perp O$; i.e, the dependency between X and O manifests only in the dependency between O and the first K monomial components $\xi_1\psi_1, \dots, \xi_K\psi_K$ 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 $\hat{\mu}^{(K)}$ by repeatedly applying the FTC:

$$\begin{aligned} \tilde{\mu}_{\text{FTC}}^{(K-1)}(t) &= \int_{d_f}^t \hat{\mu}^{(K)}(z) dz + \hat{\mu}^{(K-1)}(d_f), \\ \tilde{\mu}_{\text{FTC}}^{(K-2)}(t) &= \int_{d_f}^t \tilde{\mu}_{\text{FTC}}^{(K-1)}(z) dz + \hat{\mu}^{(K-2)}(d_f), \\ &\vdots \\ \text{and } \tilde{\mu}_{\text{FTC}}(t) &= \int_{d_f}^t \tilde{\mu}_{\text{FTC}}^{(1)}(z) dz + \hat{\mu}(d_f), \end{aligned}$$

where $\tilde{\mu}_{\text{FTC}}(t)$ 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 $\hat{\sigma}^{(K,K)}$:

$$\begin{aligned}
\tilde{\sigma}_{\text{FTC}}^{(K-1,K-1)}(s, t) &= \int_{d_f}^t \int_{d_f}^s \hat{\sigma}^{(K,K)}(z_1, z_2) dz_1 dz_2 + \int_{d_f}^s \hat{\sigma}^{(K,K-1)}(z_1, d_f) dz_1 + \\
&\quad + \int_{d_f}^t \hat{\sigma}^{(K-1,K)}(d_f, z_2) dz_2 + \hat{\sigma}^{(K-1,K-1)}(d_f, d_f), \\
\tilde{\sigma}_{\text{FTC}}^{(K-2,K-2)}(s, t) &= \int_{d_f}^t \int_{d_f}^s \tilde{\sigma}_{\text{FTC}}^{(K-1,K-1)}(z_1, z_2) dz_1 dz_2 + \int_{d_f}^s \int_{d_f}^t \hat{\sigma}^{(K,K-2)}(z_1, d_f) dz_1 dz_2 \\
&\quad + \int_{d_f}^t \int_{d_f}^s \hat{\sigma}^{(K-2,K)}(d_f, z_2) dz_2 dz_2 + 2\hat{\sigma}^{(K-1,K-2)}(d_f, d_f) + \hat{\sigma}^{(K-2,K-2)}(d_f, d_f), \\
&\quad \vdots \\
\text{and } \tilde{\sigma}_{\text{FTC}}(s, t) &= \int_{d_f}^t \int_{d_f}^s \tilde{\sigma}_{\text{FTC}}^{(1,1)}(z_1, z_2) dz_1 dz_2 + \int_{d_f}^s \int_{d_f}^t \hat{\sigma}^{(0,1)}(z_1, d_f) dz_1 dz_2 \\
&\quad + \int_{d_f}^t \int_{d_f}^s \hat{\sigma}^{(0,1)}(d_f, z_2) dz_2 dz_2 + 2\hat{\sigma}^{(1,0)}(d_f, d_f) + \hat{\sigma}^{(0,0)}(d_f, d_f),
\end{aligned}$$

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

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 $\xi_{i1}, \xi_{i2}, \dots$. The coefficients $\xi_{i1}, \xi_{i2}, \dots$, however, refer to the complete trajectories and, therefore, cannot be computed from the partially observed trajectories. Furthermore, existing procedures that allow to predict the coefficients $\xi_{i1}, \xi_{i2}, \dots$ 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 $\xi_{i1}^{[0, d_{\min}]}, \xi_{i2}^{[0, d_{\min}]}, \dots$ 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 $\xi_{i1}^{[0, d_{\min}]}, \xi_{i2}^{[0, d_{\min}]}, \dots$ will be informative about the infeasible coefficients $\xi_{i1}, \xi_{i2}, \dots$. In fact, Liebl and Kneip (2018) argue in a related context that the coefficients $\xi_{i1}^{[0, d_{\min}]}, \xi_{i2}^{[0, d_{\min}]}, \dots$ contain the same information as $\xi_{i1}, \xi_{i2}, \dots$ 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 ψ_1, \dots, ψ_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 $i = 1, \dots, n$, and second, we use $J = \text{median}\{J_1^{\text{BIC}}, \dots, J_n^{\text{BIC}}\}$ 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 coefficients $\xi_{i1}^{[0, d_{\min}]}, \dots, \xi_{iJ}^{[0, d_{\min}]}$. In order to test for possible correlations between $\xi_{i1}^{[0, d_{\min}]}, \dots, \xi_{iJ}^{[0, d_{\min}]}$ and d_i , we use the bootstrap-based multiple testing of Romano and Wolf (2005) which allows 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 1), where we substitute the ξ_{ji} with the approximated versions of the $\xi_{ji}^{[0, d_{\min}]}$. We investigate the set of hypotheses $B = \{H_{0,1}, \dots, H_{0,J}\}$ with $H_{0,j} : \beta_j = 0, j = 1, \dots, J$, using standard squared t -test statistics $\hat{t}_j^2 = (\hat{\beta}_j / \hat{s}_{\hat{\beta}_j})^2$. 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 $H_{j,0}, j = 1, \dots, J$, 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 point towards the MCAR assumption.

(V): Only $H_{1,0} : \beta_1 = 0$ will be rejected, all other $H_{j,0}, j = 2, \dots, J$, cannot be rejected. That is, only ξ_{i1} significantly correlates with d_i . This test decision point 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 1) is initialized by setting $B = B_1$, where $\mathbf{d} = (d_1, \dots, d_n)$, $\Xi = (\xi_1, \dots, \xi_J) \in \mathbb{R}^{n \times J}$, and R denotes the number of bootstrap replications:

Algorithm 1 Romano Wolf Stepdown Procedure for Multiple Hypothesis Testing.

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

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

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

A further simple visual tool for checking Violation (V) is to compare the estimates of the classical estimators, $\hat{\mu}(t)$ and $\hat{\sigma}(s, t)$, with those of the FTC estimators, $\hat{\mu}_{\text{FTC}}(t)$ and $\hat{\sigma}_{\text{FTC}}(s, t)$. Under the MCAR assumption the two estimates of the mean function and the two estimates of the covariance function will essentially coincide, however, under Violation (V) the pairs of estimates will differ.

4 Simulation

In the following simulation study, we assess the finite sample properties of our FTC-estimators $\hat{\mu}_{\text{FTC}}$ and $\hat{\sigma}_{\text{FTC}}$, and investigate performance of the multiple testing procedure described in Algorithm 1. 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 $\psi_{2k+1}(t) = \sqrt{2} \cos(2\pi kt)$, $k = 1, \dots, (J-1)/2$, with $J = 5$ and draw uncorrelated coefficients $\xi_{ij}, j = 1, \dots, J$, from a normal distribution with means $\mathbb{E}[\xi_{ij}] = \mu_j$ and variances $\mathbb{V}(\xi_{ij}) = \lambda_j$, where $\mu_1 = 5, \mu_2 = 2, \mu_3 = 0, \dots, \mu_5 = 0$, and $\lambda_1 = 10, \lambda_2 = 8, \dots, \lambda_5 = 2$. 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 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., $\xi_{i1} \not\perp d_i$, and independent (Ind.), i.e., $\xi_{i1} \perp d_i$, setups. The Dep.-DGPs represent versions of Violation (V) of the MCAR assumption and the Ind.-DGPs fulfill the MCAR assumption.

Table 1: Data Generating Processes.

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

In each of the DGPs, ξ_{i1} is drawn from $N(\mu_1, \lambda_1)$. 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

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

where $d_i^* = \Phi_{\mu_1, \lambda_1}(\xi_{i1})$ and where $q_{\text{emp}, 0.98}$ denotes the empirical 98 percent quantile of the sample $\{d_1^*, \dots, d_n^*\}$. 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, \dots, 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, 2017), 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 $J = 3, 5, \dots, J_{\max}$ with $J_{\max} = 51$.

Table 2 contains our simulation results for our FTC-estimators $\hat{\mu}_{\text{FTC}}$ and $\hat{\sigma}_{\text{FTC}}$. We report the (doubly) integrated squared bias $(\int_0^1 \int_0^1 \text{Bias}^2(\cdot))$ and variance $(\int_0^1 \int_0^1 \text{Var}(\cdot))$ for $\hat{\mu}_{\text{FTC}}$ ($\hat{\sigma}_{\text{FTC}}$) and compare them with the estimation errors of the classical estimators $\hat{\mu}$ ($\hat{\sigma}$). The results clearly demonstrate a systematic bias of the classical estimators $\hat{\mu}$ and $\hat{\sigma}$ under the Dep. DGPs (see the Dep.-Dis. and Dep.-Con. columns), which is non-vanishing as n increases. By contrast, our FTC-estimators show essentially no bias—despite Violation

Table 2: Integrated squared bias and variance.

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

(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 $\hat{\sigma}_{\text{FTC}}$ 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. 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). 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. 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: Selection errors for Null, (V) and Other (in %) – $J_{\max} = 51$.

n	Dep.-Dis.			Dep.-Con.			Ind.-Dis.			Ind.-Con.		
	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

* $J_{\max} = 31$.

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 has an important role in maintaining grid stability, which is one of the basic responsibilities of electricity markets. An effective tool to guarantee grid stability are reserve capacities. Therefore, the German Federal Network Agency (FNA), a public institution, determines each week a sufficiently large amount of reserve capacity d_i that has to be bought, by law, by the Grid Control Cooperation² (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.

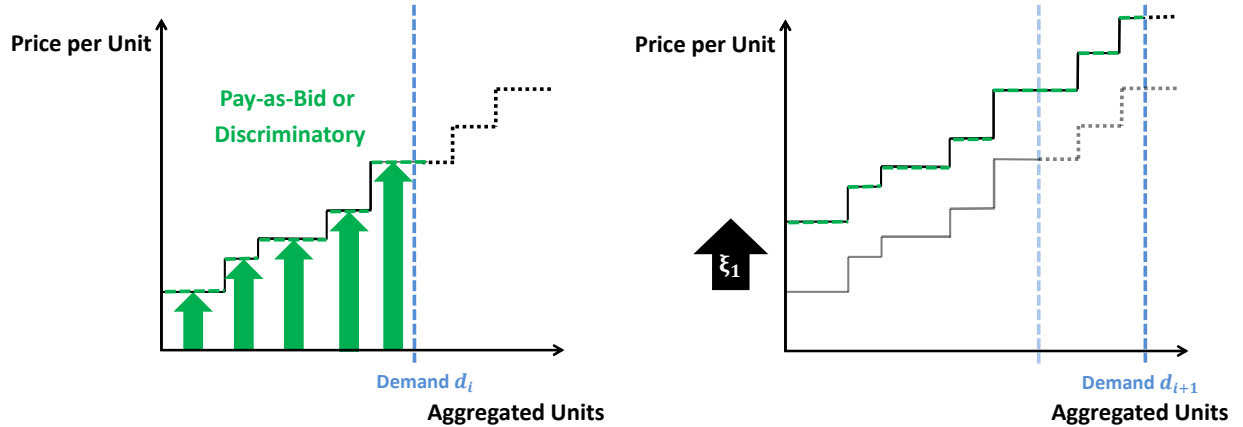


Figure 1: Left: The Pay-as-bid pricing mechanism, where each successful bidder receives the price of his bid. Right: A vertical shift dependency of the price curves due to changes in the announced amount of reserve capacity d_i .

The German Control Reserve Market uses a so-called “pay-as-bid” auction, where every

²The 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.

successful bidder (i.e., the electricity providers) receives his own (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 1 and can be clearly seen in the data (Figure 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.

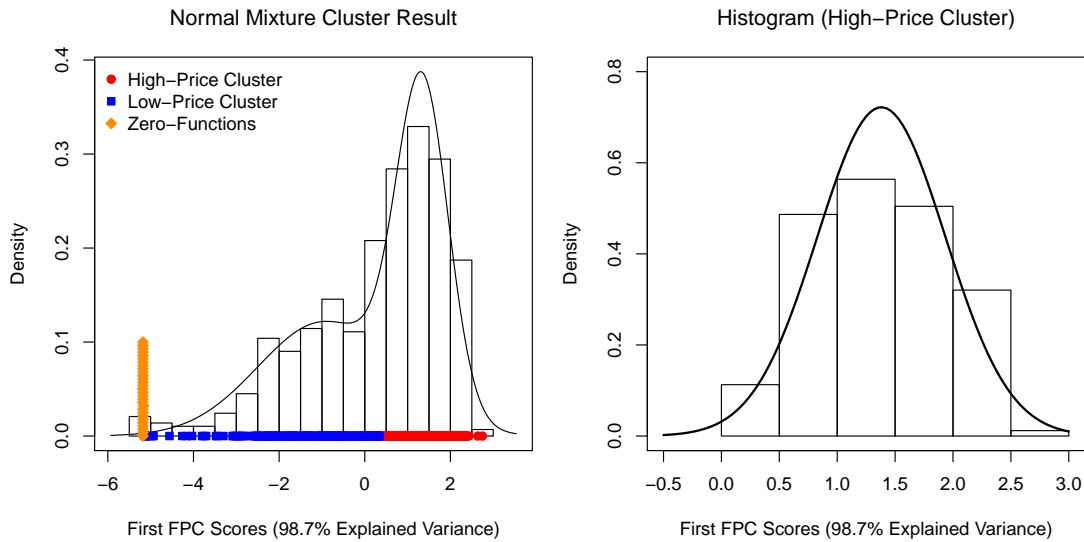


Figure 2: Distribution of the first principal component scores of the log-price curves computed with respect to the fully observed interval $[a, d_{\min}]$ of the original sample.

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 (≥ 20.000 EUR/MW) and are removed. In order to account for some very large 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 accompanying 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 $t_1 = 0, t_2 = 5, \dots, t_{501} = 2500$ 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 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 is needed in our practical test procedure described in Section 3. Standard normality tests reject the null hypothesis of normality due to the hard thresholding 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 2).

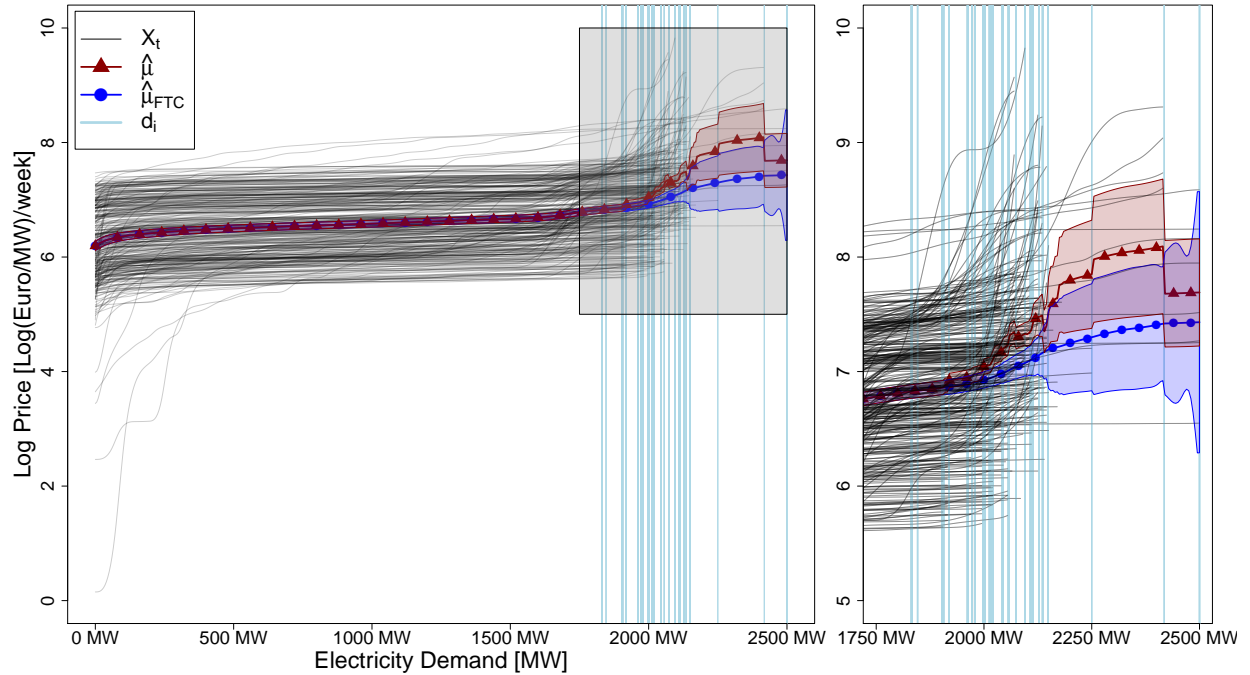


Figure 3: Classical mean (solid with triangles) and FTC mean estimates (solid with circles) computed from $n = 337$ partially observed price curves.

We test the functional data sample for Violation (V) as described in Section 3 where the procedure leads to scenario (V), i.e., the identification of Violation (V). For robustness, we additionally checked $J_{\max} = 51 \pm 10$, which also lead to Violation (V). To estimate

the mean function, we use our FTC-estimator $\hat{\mu}_{\text{FTC}}$ as defined in (3) and compare the estimation result with that from using the classical estimator $\hat{\mu}$. The difference between the classical estimator $\hat{\mu}(t)$, which is inconsistent under Violation (V), and our FTC-estimator $\hat{\mu}_{\text{FTC}}(t)$ is obvious: the systematically missing parts of the curves lead to a positive bias in $\hat{\mu}(t)$ 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 $\hat{\mu}$ results also in a defective estimate of the covariance function when using the classical covariance estimator $\hat{\sigma}$. 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.

Acknowledgements

The authors thank Prof. Rolf Tschernig (University of Regensburg) for his valuable suggestions which helped to improve this research work. Many thanks go also to two anonymous referees for their constructive comments which had a very positive impact on this paper too.

A Appendix

A.1 Methodology

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

Proof of Proposition A.1:

$$\begin{aligned} \mathbb{E}[\hat{\mu}(t)] &= \mathbb{E}[\mathbb{E}[\hat{\mu}(t) \mid O_1(t), \dots, O_n(t)]] \\ &= \mathbb{E}\left[\frac{I_1(t)}{\sum_{i=1}^n O_i(t)} \sum_{i=1}^n O_i(t) \left(\mathbb{E}[\xi_{i1} \mid O_i(t)] \psi_1(t) + \sum_{j \geq 2} \mu_j \psi_j(t)\right)\right] \\ &= \mathbb{E}\left[\frac{I_1(t)}{\sum_{i=1}^n O_i(t)} \sum_{i=1}^n O_i(t) \sum_{j \geq 2} \mu_j \psi_j(t)\right] + \mathbb{E}\left[\frac{I_1(t)}{\sum_{i=1}^n O_i(t)} \sum_{i=1}^n O_i(t) \mathbb{E}[\xi_{i1} \mid O_i(t)] \psi_1(t)\right] \end{aligned}$$

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

$$\begin{aligned} \mathbb{E}\left[\frac{I_1(t)}{\sum_{i=1}^n O_i(t)} \sum_{i=1}^n O_i(t) \sum_{j \geq 2} \mu_j \psi_j(t)\right] &= \mathbb{E}\left[\frac{I_1(t)}{\sum_{i=1}^n O_i(t)} \sum_{i=1}^n O_i(t) \left(\sum_{j \geq 1} \mu_j \psi_j(t) - \mu_1 \psi_1(t)\right)\right] \\ &= \mu(t) - \mu_1 \psi_1(t) \end{aligned}$$

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

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

which results in

$$\begin{aligned} \mathbb{E} \left[\frac{I_1(t)}{\sum_{i=1}^n O_i(t)} \sum_{i=1}^n O_i(t) \mathbb{E} [\xi_{i1} | O_i(t)] \psi_1(t) \right] &= \mathbb{E} \left[\frac{I_1(t)}{\sum_{i=1}^n O_i(t)} \sum_{i=1}^n O_i(t) (\mu_1\psi_1(t) + \Delta(t)) \right] \\ &= \mu_1\psi_1(t) + \Delta(t) \end{aligned}$$

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

Proof of Theorem 2.1: First, we focus on $\hat{\mu}_{\text{FTC}}(t)$. 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 $\hat{\mu}$ is unbiased and consistent.

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

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

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

$$\hat{\mu}^{(1)}(z) = \frac{J(z)}{\sum_{i=1}^n O_i(z)} \sum_{i=1}^n X_i^{(1)}(z) O_i(z) dz \quad (7)$$

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 $\hat{\mu}^{(1)}(z)$ for $z \in [d_{\min}, 1]$. It follows by the Weak Law of Large Numbers that $\hat{\mu}^{(1)}(z) \rightarrow_p \mu^{(1)}(z)$ as $n \rightarrow \infty$. For the second summand in (3)—since $p(d_{\min}) = 1$ —we see that $n^{-1} \sum_{i=1}^n X_i(d_{\min}) \rightarrow_p \mu(d_{\min})$ as $n \rightarrow \infty$. With the Continuous Mapping Theorem (CMT) we finally deduce the unbiasedness and consistency of $\hat{\mu}_{\text{FTC}}(t)$ for all $t \in [0, 1]$. This reasoning is later used to show consistency in the case of $\hat{\sigma}_{\text{FTC}}(s, t)$ as well.

In the case of $\hat{\sigma}_{\text{FTC}}(s, t)$, 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 $\hat{\sigma}(s, t)$ as in (2) is consistent and therefore $\hat{\sigma}_{\text{FTC}}(s, t)$ 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

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

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 after (7), applied to

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

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:

$$\sigma(s, t) = \int_{d_{\min}}^t \frac{\partial}{\partial z_2} \sigma(s, z_2) dz_2 + \sigma(s, d_{\min}) \quad (8)$$

$$\sigma(s, z_2) = \int_{d_{\min}}^s \frac{\partial}{\partial z_1} \sigma(z_1, z_2) dz_1 + \sigma(d_{\min}, z_2) \quad (9)$$

$$\int_{d_{\min}}^t \frac{\partial}{\partial z_2} \sigma(d_{\min}, z_2) dz_2 = \sigma(d_{\min}, t) - \sigma(d_{\min}, d_{\min}) \quad (10)$$

Plugging (9) into (8) and using the linearity of the operations involved yields

$$\begin{aligned} \sigma(s, t) &= \int_{d_{\min}}^t \frac{\partial}{\partial z_2} \left(\int_{d_{\min}}^s \frac{\partial}{\partial z_1} \sigma(z_1, z_2) dz_1 + \sigma(d_{\min}, z_2) \right) dz_2 + \sigma(s, d_{\min}) \\ \Leftrightarrow \sigma(s, t) &= \int_{d_{\min}}^t \int_{d_{\min}}^s \frac{\partial^2}{\partial z_1 \partial z_2} \sigma(z_1, z_2) dz_1 dz_2 + \int_{d_{\min}}^t \frac{\partial}{\partial z_2} \sigma(d_{\min}, z_2) dz_2 + \sigma(s, d_{\min}). \end{aligned} \quad (11)$$

Plugging (10) into (11) yields

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

Replacing $\sigma(d_{\min}, t)$ with a corresponding version of (8) and $\sigma(s, d_{\min})$ with a corresponding version of (9) yields

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

Simplifying and using the notation of Definition 2.2 leads to:

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

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

$$\hat{\sigma}^{(\ell,k)}(s, t) = \frac{I_2(s, t)}{\sum_{i=1}^n U_i(s, t)} \sum_{i=1}^n U_i(s, t) \left[X_i^{(\ell)}(s) - \hat{\mu}^{(\ell)}(s) \right] \left[X_i^{(k)}(t) - \hat{\mu}^{(k)}(t) \right], \quad \ell, k \in \{0, 1\}.$$

References

- 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.
- Engelmann, D. and V. Grimm (2009). Bidding behaviour in multi-unit auctions—an experimental investigation. *The Economic Journal* 119(537), 855–882.
- Ferraty, F. and P. Vieu (2006). *Nonparametric Functional Data Analysis - Theory and Practice*. Springer.
- 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.
- Gellar, J. E., E. Colantuoni, D. M. Needham, and C. M. Crainiceanu (2014). 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.
- Grimm, V., A. Ockenfels, and G. Zoettl (2008). Strommarktdesign: Zur Ausgestaltung der Auktionsregeln an der EEX. *Zeitschrift für Energiewirtschaft* (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.
- 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.
- 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.
- Kraus, D. (2015). Components and completion of partially observed functional data. *Journal of the Royal Statistical Society* 77(4), 777–801.
- 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 A. Kneip (2018). *On the optimal reconstruction of partially observed functional data*. arXiv:1710.10099.
- Little, R. J. and D. B. Rubin (2014). *Statistical analysis with missing data*. John Wiley & Sons.
- Ramsay, J. and B. Silverman (2005). *Functional Data Analysis* (2. ed.). 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 (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- 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.
- 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.
- 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.