Inference for Covariate-Adjusted Sparse to Dense Functional Data with Application to Testing Differences in Energy Prices

Dominik Liebl
Institute for Financial Economics and Statistics, University of Bonn

March 29, 2017

Abstract

We reconsider the nonparametric estimation of mean and covariance functions for covariate-adjusted functional data. Inspired by the recent literature, we differentiate between "sparse" and "dense" functional data – depending on the relative order of m (discretization points per function) and n (number of functions). For each of the two asymptotic scenarios we derive optimal multiple bandwidth expressions and asymptotic normality results. Additionally, we propose practical rule-of-thumb bandwidth approximations and a hybrid ("sparse-to-dense") solution for inference about the mean function. The finite sample properties of the sparse, dense, and sparse-to-dense asymptotic normality results are investigated through simulation studies. Our theoretical results are applied in a real data study where we estimate and test the differences in the electricity prices before and after Germany's abrupt nuclear phaseout after the nuclear disaster in Fukushima Daiichi, Japan, in mid-March 2011.

Keywords: functional data, local linear kernel estimation, multiple bandwidth selection, time series analysis, electricity spot prices, nuclear power phaseout

1 Introduction

On March 15, 2011, Germany showed an abrupt reaction to the nuclear disaster in Fukushima Daiichi, Japan, and shut down 40% of its nuclear power plants – permanently. This substantial loss of cheap (in terms of marginal costs) nuclear power raised concerns about increases in electricity prices and subsequent problems for industry and households; however, empirical studies do not report any significant price differences before and after this partial nuclear phaseout (see, e.g., Nestle, 2012). This is a surprising finding that we want to reconsider in our application in Section 6.

We consider data from the European Power Exchange (EPEX), where it is important to note that all hourly electricity spot prices of a day i are settled simultaneously at 12 am the day before (Grimm et al., 2008). It is common in the literature to differentiate between electricity spot prices of "off-peak hours" (from 1am to 8am and 9pm to 12pm) and "peak hours" (from 9am to 8pm). In this paper we focus on peak-hours, since these show the largest variations in electricity spot prices. Detailed information on the data sources are given in Section 4 of our supplemental paper.

It is quite natural to approach the empirical analysis using a functional data perspective. Pricing in power markets is well explained by the so-called merit order model. This model assumes that the spot prices at electricity exchanges are based on the marginal generation costs of the last power plant that is required to cover the demand. The resulting merit order curve reflects the increasing generation costs of the installed power plants. Often, nuclear and lignite power plants cover the minimal demand for electricity. Higher demand is mostly served by hard coal and gas fired power plants. The merit order model is a so-called fundamental market model (see, e.g., Burger et al., 2008, Ch. 4) and most important for the explanation of electricity spot prices in the literature on energy economics. Well known references using the merit order model are Burger et al. (2004), Sensfuß et al. (2008), Hirth (2013), and Cludius et al. (2014).

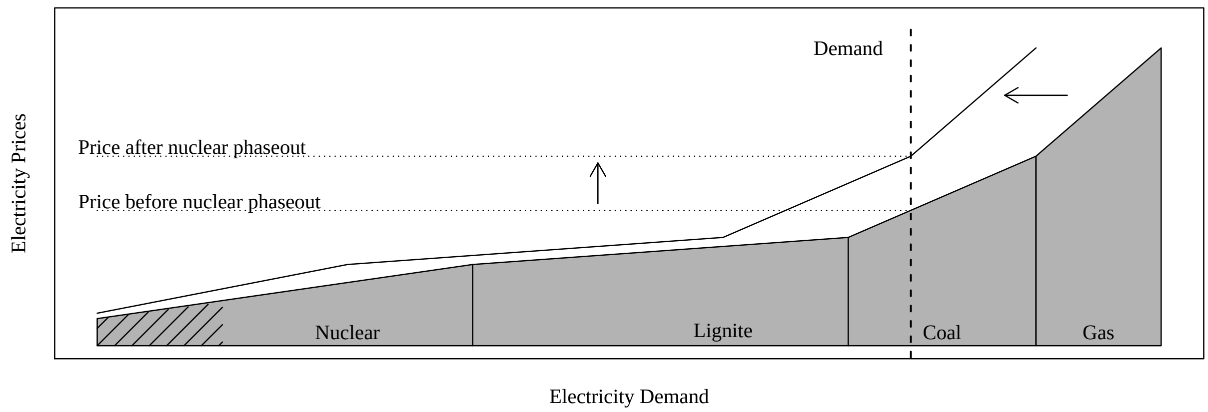

Figure 1: Sketch of the merit order curve and the theoretical price effect of a nuclear power phaseout. The dashed region signifies the proportion of phased out nuclear power plants.

The plot in Figure 1 sketches the merit order curve of the German electricity market and is in line with Cludius et al. (2014). The interplay of the inverse demand curve (dashed line) with the merit order curve determines the electricity prices. It is assumed that the electricity demand is price-inelastic in the short-term perspective of a day-ahead market. This assumption is regularly found in the literature on electricity spot prices (see, e.g., Sensfuß et al., 2008) and confirmed, in all material respects, in empirical studies (see, e.g., Lijesen, 2007).

The (daily-varying) merit order curve and the daily pricing scheme justify our view on the hourly electricity spot prices Y_{ij} of day $i = 1, \ldots, n$ at peak hour $j = 1, \ldots, m$. We interpret them as $m = 12$ discretization points of a covariate-adjusted empirical merit-order curve, or price curve, $X_i(., z)$ plus an additive noise term, i.e.,

$$Y_{ij} = X_i(U_{ij}, Z_i) + \epsilon_{ij}, \tag{1}$$

where $U_{ij} \in \mathbb{R}$ denotes the value of electricity demand at day i and peak hour j, the covariate $Z_i \in \mathbb{R}$ is the daily mean air temperature, and $\epsilon_{ij} \in \mathbb{R}$ is a classic statistical error term. The price function is assumed to be square integrable covariate-adjusted random function, i.e., $X_i(., z) \in L^2[a(z), b(z)]$, where $[a(z), b(z)] \subset \mathbb{R}$ denotes the covariate-adjusted support compact of $X_i(., z)$.

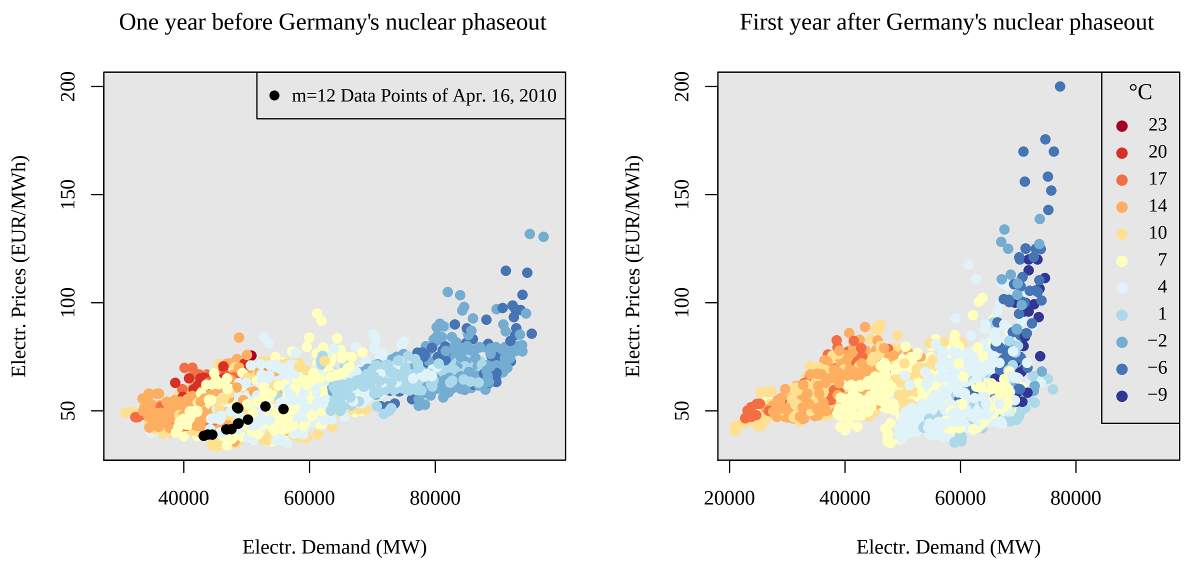

Figure 2: Scatter plots of hourly electricity spot prices Y_{ij} against hourly electricity demand values U_{ij} (gray filled circles), where colors encode the covariate Z_i of daily mean air temperature values. The left plot shows the data from one year before Germany's partial nuclear phaseout, i.e., from March 15, 2010 to March 14, 2011; the right plot shows the data from one year after, i.e., from March 15, 2011 to March 14, 2012.

The temperature covariate Z_i is the most important exogenous factor in the energy market and allows us to control for seasonal differences. The analyzed data are shown in Figure 2, where each circle point shows one data pair (Y_{ij}, U_{ij}) and the color of the circle point shows the temperature covariate Z_i. The covariate Z_i clearly influences the location and the spread of both the hourly electricity spot prices Y_{ij} and the hourly electricity demand U_{ij}. In fact, realizations of U_{ij} are only observed within temperature-specific sub-intervals, i.e., $U_{ij} \in [a(Z_i), b(Z_i)]$, which motivates to our modeling assumption that $X_i(., z) \in L^2[a(z), b(z)]$.

With only $m = 12$ observed data points $(Y_{i1}, U_{i1}), \ldots, (Y_{im}, U_{im})$ per unobserved function $X_i(., Z_i)$ it is not advisable to apply the commonly used "pre-smoothing approach". Therefore, we use the alternative "pooling approach" and estimate the covariate-adjusted mean and covariance functions directly from the pooled data points $\{(Y_{ij}, U_{ij}, Z_i), 1 \leq i \leq n, 1 \leq j \leq m\}$ using the Local Linear Kernel (LLK) estimators of Jiang and Wang (2010).

The main objective in our application in Section 6 is to test for differences in the covariate-adjusted mean price functions before and after Germany's partial nuclear phase-out. Jiang and Wang (2010) have derived – among many other important results – asymptotic normality results for their LLK estimators which might be useful for our testing problem. Though, it turns out that their asymptotic normality results do not provide practically useful approximations. Their asymptotic variance expressions clearly underestimate the actual variances in "large-n, small-m" data scenarios of practical relevance; see also our simulation study in Section 5.

The reason for this is that Jiang and Wang (2010) derive their asymptotic results using a "finite-m" asymptotic, where $n \to \infty$ while m remains bounded, i.e., $m \leq$ const. Though, the finite-m asymptotic neglects an additional functional data specific variance term which is typically not negligible in practice. The additional variance term becomes visible under a double asymptotic where both n and m are allowed to diverge. Zhang et al. (2016) demonstrate this for the case of classic functional data, i.e., without considering covariate adjustments. They show – among many other important results – that the additional functional data specific variance term becomes the leading variance term if m diverges only slightly faster than $n^{1/4}$, i.e., if $m/n^{1/4} \to \infty$.

Inspired by the work of Zhang et al. (2016), we reconsider the LLK estimators of Jiang and Wang (2010) under such a double asymptotic. Here it turns out that the functional data specific variance term becomes the leading variance term if m diverges only slightly faster than $n^{1/5}$, i.e., if $m/n^{1/5} \to \infty$. This is even an order of magnitude earlier than in the situation without covariates considered by Zhang et al. (2016).

A further remarkable result in Zhang et al. (2016) is that their LLK estimators attain a parametric $\sqrt{n}$-convergence rate if $m/n^{1/4} \to \infty$, where even the bias components become negligible if the bandwidth is chosen appropriately. That is, their nonparametric LLK estimators behave then as if they were classic parametric moment estimators applied to a sample of n fully observed functions.

We derive the corresponding results for the LLK estimators of Jiang and Wang (2010) which are, though, slightly different: If $m/n^{1/5} \to \infty$, the LLK estimator for the bivariate mean function $\mu(u, z) = \mathbb{E}(X_i(u, z))$ and the LLK estimator for the trivariate covariance function $\gamma(u_1, u_2, z) = \mathrm{Cov}(X_i(u_1, z), X_i(u_2, z))$ behave like LLK estimators for *uni*variate regression functions with Z_i as the only covariate. That is, the LLK estimators behave then as if the sample of n functions $(X_1(., Z_1), \ldots, X_n(., Z_n))$ were fully observed such that smoothing has to be done only in Z-direction. This is essentially equivalent to the results in Zhang et al. (2016), but with the additional complexity of estimating the mean and covariance functions *conditionally* on the covariate Z_i.

We also derive the explicit multiple bandwidth expressions for the different asymptotic scenarios $m/n^{1/5} \to 0$ and $m/n^{1/5} \to \infty$ (hereinafter referred to as "sparse" and "dense" covariate-adjusted functional data). In the "dense" case, this leads to rather unconventional optimal bandwidth expressions with different convergence rates for the bandwidths in U- and Z-direction. Effectively, this imposes an under-smoothing in U-direction, which then guarantees that the U-related bias and variance components become negligible in comparison to the Z-related bias and variance components. This under-smoothing strategy is already well-known in the literature on the "pre-smoothing approach" (see, e.g., Benko et al., 2009). Our optimal multiple bandwidth expressions allow now to implement this under-smoothing strategy also for the "pooling approach".

Although we present several new theoretical results, their proofs follow from straight forward reconsiderations of (simplified versions of) the asymptotic normality results in

Theorems 3.2 and 3.4 of Jiang and Wang (2010) under a (simplified version of) the double asymptotic used in Zhang et al. (2016). For convenience, the proofs are provided in our supplemental paper. Of central importance in this paper is the practical implementation of the theoretical results. For this we suggest rule-of-thumb approximations to the unknown quantities in the optimal bandwidth, bias and variance expressions and propose practically feasible test statistics that allow for inference about the mean function for "sparse" and "dense" covariate-adjusted functional data. Additionally, we propose a third test statistic for the intermediate case of "sparse-to-dense" covariate-adjusted functional data. We compare the small-sample performances of all three test statistics in our simulation study. Our theoretical results are of direct use in our real data application where we test for differences in the covariate-adjusted mean electricity prices before and after Germany's partial nuclear phaseout.

The literature on covariate-adjusted functional data is scarce. The first research on covariate-adjusted functional data is found in Cardot (2007), who considers the case of fully observed random functions. To date the only other article concerned with covariate-adjusted functional data is that of Jiang and Wang (2010).

The rest of this paper is structured as follows: The next section introduces the considered regression models and LLK estimators. Section 3 presents our assumptions and asymptotic results. Section 4 introduces rule-of-thumb approximations to our theoretical bandwidth expressions as well as practical approximations to the unknown components in the bias and variance terms. Our simulation study is found in Section 5. Section 6 contains the real data study and Section 7 concludes.

2 Nonparametric regression models and estimators

Let X_i^c denote the centered random function $X_i^c(U_{ij}, Z_i) = X_i(U_{ij}, Z_i) - \mathbb{E}(X_i(U_{ij}, Z_i)|\mathbf{U}, \mathbf{Z})$, where $\mathbf{U} = (U_{11}, \dots, U_{nm})^\top$ and $\mathbf{Z} = (Z_1, \dots, Z_n)^\top$. Model (1) can then be written as a nonparametric regression model with the bivariate mean function $\mu(U_{ij}, Z_i) = \mathbb{E}(X_i(U_{ij}, Z_i)|\mathbf{U}, \mathbf{Z})$ as its regression function, i.e.,

$$Y_{ij} = \mu(U_{ij}, Z_i) + X_i^c(U_{ij}, Z_i) + \epsilon_{ij}, \tag{2}$$

where $X_i^c(., z)$, U_{ij}, and Z_i are assumed to be a stationary weakly dependent functional and univariate time series. The error term ϵ_{it} is a classic iid error term with mean zero and finite variance $\mathbb{V}(\epsilon_{it}) = \sigma_\epsilon^2$. Note that Model (2) has a rather unusual composed error term $X_i^c(U_{ij}, Z_i) + \epsilon_{ij}$ consisting of a functional and a scalar component.

Likewise we can define the following nonparametric regression model with the trivariate covariance function $\gamma(U_{ij}, U_{ik}, Z_i) = \mathrm{Cov}(X_i(U_{ij}, Z_i), X_i(U_{ik}, Z_i)|\mathbf{U}, \mathbf{Z})$ as its regression function:

$$C_{ijk} = \gamma(U_{ij}, U_{ik}, Z_i) + \tilde{X}_i^c(U_{ij}, U_{ik}, Z_i) + \varepsilon_{ijk} \tag{3}$$

with $j \neq k \in \{1, \dots, m\}$, where the "raw covariances" C_{ijk}, the centered random function $\tilde{X}_i^c$, and the scalar error term ε_{ijk} are defined as

$$\begin{aligned} C_{ijk} &= (Y_{ij} - \mu(U_{ij}, Z_i))(Y_{ik} - \mu(U_{ik}, Z_i)), \qquad (4)\\ \tilde{X}_i^c(U_{ij}, U_{ik}, Z_i) &= X_i^c(U_{ij}, Z_i)\, X_i^c(U_{ik}, Z_i) - \gamma(U_{ij}, U_{ik}, Z_i), \text{ and}\\ \varepsilon_{ijk} &= X_i^c(U_{ij}, Z_i)\epsilon_{ik} + X_i^c(U_{ik}, Z_i)\epsilon_{ij} + \epsilon_{ij}\epsilon_{ik}. \end{aligned}$$

The scalar error term ε_{ijk} is uncorrelated, but, by contrast to ϵ_{ij}, heteroscedastic with $\mathbb{V}(\varepsilon_{ijk}) = \sigma_\varepsilon^2(u_1, u_2, z)$, where $\sigma_\varepsilon^2(u_1, u_2, z) = \gamma(u_1, u_1, z)\, \sigma_\epsilon^2 + \gamma(u_2, u_2, z)\sigma_\epsilon^2 + \sigma_\epsilon^4$. Note that $\mathbb{E}(\varepsilon_{ijk}) \neq 0$ for all $j = k$, therefore all raw covariance points C_{ijk} with $j = k$ need to be excluded (see also Yao et al., 2005). Correspondingly, the number of raw covariance points for a time point i is given by $M = m^2 - m$, which makes it necessary that $m \geq 2$.

Equivalently to Model (2), Model (3) has a composed error term $\tilde{X}_i^c(U_{ij}, U_{ik}, Z_i) + \varepsilon_{ijk}$ consisting of a functional and a scalar component. The functional error components in Models (2) and (3) lead to the additional functional data specific variance terms in the asymptotic variance expressions of the LLK estimators.

We estimate the mean function $\mu(u, z)$ using the LLK estimator $\hat{\mu}(u, z; h_{\mu,U}, h_{\mu,Z})$ defined as the following locally weighted least squares estimator (see, e.g., Ruppert and Wand, 1994):

$$\hat{\mu}(u, z; h_{\mu,U}, h_{\mu,Z}) = e_1^\top \left([\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz}[\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]\right)^{-1} [\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz}\mathbf{Y}, \qquad (5)$$

where the vector $e_1 = (1, 0, 0)^\top$ selects the estimated intercept parameter and $[\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]$ is a partitioned $nm \times 3$ dimensional data matrix with typical rows $(1, U_{ij} - u, Z_i - z)$. The $nm \times nm$ dimensional diagonal weighting matrix $\mathbf{W}_{\mu,uz}$ holds the bivariate multiplicative kernel weights $K_{\mu,h_{\mu,U},h_{\mu,Z}}(U_{ij} - u, Z_i - z) = h_{\mu,U}^{-1}\,\kappa(h_{\mu,U}^{-1}(U_{ij} - u))\; h_{\mu,Z}^{-1}\,\kappa(h_{\mu,Z}^{-1}(Z_i - z))$, where κ is a univariate symmetric probability density function (pdf) with compact support $\mathrm{supp}(\kappa) = [-1, 1]$, such as, e.g., the univariate Epanechnikov kernel. The usual kernel constants are denoted by $\nu_2(K_\mu) = (\nu_2(\kappa))^2$, with $\nu_2(\kappa) = \int u^2\kappa(u)du$, and $R(K_\mu) = R(\kappa)^2$, with $R(\kappa) = \int \kappa(u)^2 du$. All vectors and matrices are filled in correspondence with the nm dimensional vector $\mathbf{Y} = (Y_{11}, Y_{12}, \ldots, Y_{n,m-1}, Y_{n,m})^\top$.

The LLK estimator for the covariance function $\gamma(u_1, u_2, z)$ is defined correspondingly as

$$\begin{aligned}&\hat{\gamma}(u_1, u_2, z; h_{\gamma,U}, h_{\gamma,Z}) = \\ &= e_1^\top \left([\mathbf{1}, \mathbf{U}_{u_1}, \mathbf{U}_{u_2}, \mathbf{Z}_z]^\top \mathbf{W}_{\gamma,u_1u_2z}[\mathbf{1}, \mathbf{U}_{u_1}, \mathbf{U}_{u_2}, \mathbf{Z}_z]\right)^{-1} [\mathbf{1}, \mathbf{U}_{u_1}, \mathbf{U}_{u_2}, \mathbf{Z}_z]^\top \mathbf{W}_{\gamma,u_1u_2z}\hat{\mathbf{C}}, \qquad (6)\end{aligned}$$

where $e_1 = (1, 0, 0, 0)^\top$ and $[\mathbf{1}, \mathbf{U}_{u_1}, \mathbf{U}_{u_2}, \mathbf{Z}_z]$ is a $nM \times 4$ dimensional data matrix with typical rows $(1, U_{ij} - u_1, U_{ik} - u_2, Z_i - z)$. The $nM \times nM$ dimensional diagonal weighting matrix $\mathbf{W}_{\gamma,u_1u_2z}$ holds the trivariate multiplicative kernel weights $K_{\gamma,h_{\gamma,U},h_{\gamma,Z}}(U_{ij} - u_1, U_{ik} - u_2, Z_i - z) = h_{\gamma,U}^{-1}\,\kappa(h_{\gamma,U}^{-1}(U_{ij} - u_1))\; h_{\gamma,U}^{-1}\,\kappa(h_{\gamma,U}^{-1}(U_{ik} - u_2))\; h_{\gamma,Z}^{-1}\,\kappa(h_{\gamma,Z}^{-1}(Z_i - z))$, where κ is as defined above. The usual kernel constants are $\nu_2(K_\gamma) = (\nu_2(\kappa))^3$ and $R(K_\gamma) = R(\kappa)^3$ All vectors and matrices are filled in correspondence with the nM dimensional vector $\hat{\mathbf{C}} = (\hat{C}_{112}, \hat{C}_{113}, \ldots, \hat{C}_{n,m,m-2}, \hat{C}_{n,m,m-1})^\top$, where the empirical raw covariances $\hat{C}_{ijk}$'s are defined as

$$\hat{C}_{ijk} = (Y_{ij} - \hat{\mu}(U_{ij}, Z_i; h_{\mu,U}, h_{\mu,Z}))(Y_{ik} - \hat{\mu}(U_{ik}, Z_i; h_{\mu,U}, h_{\mu,Z})). \tag{7}$$

It might be considered as restrictive to use only two bandwidths $h_{\gamma,U}$ and $h_{\gamma,Z}$ for the trivariate nonparametric estimator $\hat{\gamma}(u_1, u_2, z; h_{\gamma,U}, h_{\gamma,Z})$. However, the use of a common bandwidth for the u_1- and the u_2-direction is not restrictive at all, since the random variables U_{ij} and U_{ik} are having the same scale.

3 Assumptions and asymptotic results

Before we present our asymptotic results, we list our basic assumptions which are essentially equivalent to those in Ruppert and Wand (1994) with some straight forward adjustments to our functional data and time series context.

A-AS (Asymptotic Scenario): $nm \to \infty$, where $m = m_n \geq 2$ such that $m_n \sim n^\theta$ with $0 \leq \theta < \infty$. Hereby, "$m_n \sim n^\theta$" denotes that the two sequences m_n and n^θ are asymptotically equivalent, i.e., that $0 < \lim_{n\to\infty}(m(n)/n^\theta) < \infty$.

Remark on Assumption A-AS: This assumption is a simplified version of the asymptotic setup of Zhang et al. (2016). The case $\theta = 0$ implies that m is bounded which corresponds to the finite-m asymptotic as considered by Yao et al. (2005) and Jiang and Wang (2010). For $0 < \theta < \infty$ we are able to consider any scenario from sparsely to densely sampled functional data. As in Zhang et al. (2016), our results also apply to a random $m \equiv m_i$, but then conditionally on the values m_i.

A-RD (Random Design) The triple (Y_{ij}, U_{ij}, Z_i) has the same distribution as (Y, U, Z) with pdf f_{YUZ}, where $f_{YUZ}(y, u, z) > 0$ for all $y \in \mathbb{R}$ and $(u, z) \in [a(z), b(z)] \times [z_{\min}, z_{\max}]$ and zero else. The set $\{(u, z) \in \mathbb{R}^2$ s.t. $[a(z), b(z)] \times [z_{\min}, z_{\max}]\}$ is a compact subset of $\mathbb{R}^2$. Equivalently, $(C_{ijk}, U_{ij}, U_{ik}, Z_i)$ has the same distribution as (C, U, U', Z) with pdf f_{CUUZ}, where $f_{CUUZ}(c, u, u', z) > 0$ for all $c \in \mathbb{R}$ and $(u, u', z) \in [a(z), b(z)]^2 \times [z_{\min}, z_{\max}]$ and zero else. The set $\{(u, u', z) \in \mathbb{R}^3$ s.t. $[a(z), b(z)]^2 \times [z_{\min}, z_{\max}]\}$ is a compact subset of $\mathbb{R}^3$.

A-SM (Smoothness) The pdf $f_{YUZ}(y, u, z)$ and its marginals are continuously differentiable. All second-order derivatives of μ and γ are continuous. For estimating μ: the autocovariance functions $\gamma_l((u_1, z_1), (u_2, z_2)) = \mathbb{E}(X_i^c(u_1, z_1)X_{i+l}^c(u_2, z_2)$, $l \geq 0$, are continuously differentiable for all points within their supports. For estimating γ: the autocovariance functions $\tilde{\gamma}_l((u_1, u_2, z_1), (u_1', u_2', z_1')) = \mathbb{E}(\tilde{X}_i^c(u_1, u_2, z_1)\tilde{X}_{i+l}^c(u_1', u_2', z_1')$, $l \geq 0$, are continuously differentiable for all points within their supports.

A-TM (Time Series & Moments) X_i, U_{ij}, and Z_i are strictly stationary, ergodic, and weakly dependent time series having autocovariances that converge uniformly to zero at a geometrical rate. Furthermore, it is assumed that $\mathbb{E}((X_i(u, z))^4) < \infty$ for all (u, z) and that $\mathbb{E}(\epsilon_{ij}^2) < \infty$.

A-BW (Bandwidths) $h_{\mu,U}, h_{\mu,Z} \to 0$ and $(nm)h_{\mu,U}h_{\mu,Z} \to \infty$ as $nm \to \infty$. $h_{\mu,U}, h_{\mu,Z} \to 0$ and $(nM)h_{\mu,U}^2 h_{\mu,Z} \to \infty$ as $nM \to \infty$, where $M = m^2 - m$.

The following two Theorems 3.1 and 3.2 build the basis of our theoretical results.

Theorem 3.1 (Bias and Variance of $\hat{\mu}$) *Let u and z be interior points of* $\mathrm{supp}(f_{UZ})$. *Under our setup the conditional asymptotic bias and variance of the LLK estimator in*

Eq. (5) *are then given by*

(i) $\text{Bias}\left\{\hat{\mu}(u,z;h_{\mu,U},h_{\mu,Z})|\mathbf{U},\mathbf{Z}\right\} = B^{\mu}(u,z)\,(1+o_p(1))$ *with*

$$B^{\mu}(u,z) = \frac{1}{2}\,\nu_2(K_\mu)\,\left(h^2_{\mu,U}\,\mu^{(2,0)}(u,z) + h^2_{\mu,Z}\,\mu^{(0,2)}(u,z)\right),$$

where $\mu^{(k,l)}(u,z) = (\partial^{k+l}/(\partial u^k \partial z^l))\mu(u,z)$.

(ii) $\mathbb{V}\left\{\hat{\mu}(u,z;h_{\mu,U},h_{\mu,Z})|\mathbf{U},\mathbf{Z}\right\} = (S_1^{\mu}(u,z) + S_2^{\mu}(u,z))\,(1+o_p(1))$ *with*

$$S_1^{\mu}(u,z) = (nm)^{-1}\left[h^{-1}_{\mu,U}h^{-1}_{\mu,Z}\,R(K_\mu)\frac{\gamma(u,u,z)+\sigma^2_\epsilon}{f_{UZ}(u,z)}\right] \text{ and}$$

$$S_2^{\mu}(u,z) = n^{-1}\left[\left(\frac{m-1}{m}\right)h^{-1}_{\mu,Z}\,R(\kappa)\frac{\gamma(u,u,z)}{f_Z(z)}\right].$$

Theorem 3.2 (Bias and Variance of $\hat{\gamma}$) *Let u_1, u_2, and z be interior points of* $\text{supp}(f_{UUZ})$. *Under our setup the conditional asymptotic bias and variance of the LLK estimator in Eq.* (6) *are then given by*

(i) $\text{Bias}\left\{\hat{\gamma}(u_1,u_2,z;h_{\gamma,U},h_{\gamma,Z})|\mathbf{U},\mathbf{Z}\right\} = B^{\gamma}(u_1,u_2,z)\,(1+o_p(1))$ *with*

$$B^{\gamma}(u_1,u_2,z) = \frac{1}{2}\nu_2(K_\gamma)\left(h^2_{\gamma,U}\left(\gamma^{(2,0,0)}(u_1,u_2,z) + \gamma^{(0,2,0)}(u_1,u_2,z)\right) + h^2_{\gamma,Z}\gamma^{(0,0,2)}(u_1,u_2,z)\right),$$

where $\gamma^{(k,l,m)}(u_1,u_2,z) = (\partial^{k+l+m}/(\partial u_1^k\,\partial u_2^l\,\partial z^m))\gamma(u_1,u_2,z)$.

(ii) $\mathbb{V}\left\{\hat{\gamma}(u_1,u_2,z;h_{\gamma,U},h_{\gamma,Z})|\mathbf{U},\mathbf{Z}\right\} = (S_1^{\gamma}(u_1,u_2,z) + S_2^{\gamma}(u_1,u_2,z))\,(1+o_p(1))$ *with*

$$S_1^{\gamma}(u_1,u_2,z) = (nM)^{-1}\left[h^{-2}_{\gamma,U}h^{-1}_{\gamma,Z}\,R(K_\gamma)\frac{\tilde{\gamma}((u_1,u_2),(u_1,u_2),z) + \sigma^2_\varepsilon(u_1,u_2,z)}{f_{UUZ}(u_1,u_2,z)}\right] \text{ and}$$

$$S_2^{\gamma}(u_1,u_2,z) = n^{-1}\left[\left(\frac{M-1}{M}\right)h^{-1}_{\gamma,Z}\,R(\kappa)\frac{\tilde{\gamma}((u_1,u_2),(u_1,u_2),z)}{f_Z(z)}\right],$$

$$\begin{aligned} \textit{where}\quad \tilde{\gamma}((u_1,u_2),(u_1,u_2),z) &= \text{Cov}(\tilde{X}_i^c(u_1,u_2,z),\tilde{X}_i^c(u_1,u_2,z)) \quad \textit{and} \\ \sigma^2_\varepsilon(u_1,u_2,z) &= \gamma(u_1,u_1,z)\,\sigma^2_\epsilon + \gamma(u_2,u_2,z)\sigma^2_\epsilon + \sigma^4_\epsilon. \end{aligned}$$

The proofs of Theorems 3.1 and 3.2 can be found in Section 1 of the supplemental paper.

The above bias expressions correspond to the classic bias results (see, e.g., Ruppert and Wand, 1994). Though, the variance expressions are more interesting. The first variance terms $S_1^\mu(u,z)$ and $S_1^\gamma(u_1,u_2,z)$ are equivalent to those of Jiang and Wang (2010) who consider the LLK estimators under the finite-m asymptotic. The second variance terms $S_2^\mu(u,z)$ and $S_2^\gamma(u_1,u_2,z)$ are negligible under a finite-m asymptotic, but generally not negligible when considering a double asymptotic. The corresponding variance terms for the case of functional data without covariate-adjustments can be found in the Thorems 3.1 and 3.2 in Zhang et al. (2016).

Note that the variance effects due to the autocorrelations of our time series context are not first order relevant. The reason for this is that our LLK estimators localize with respect to the U- and Z-variables and not with respect to the time axis. The resulting decorrelation effect is often referred to as the "whitening window" property (see, e.g., Fan and Yao, 2003, Ch. 5.3).

Whether the first variance terms, S_1^μ and S_1^γ, or the second variance terms, S_2^μ and S_2^γ, are the leading variance terms depends on the bandwidth choices and on how fast m diverges with n, i.e., on the value of θ in $m \sim n^\theta$; see Assumption A-AS. In order to determine the decisive θ values we postulate optimal bandwidth choices determined from minimizing the usual asymptotic mean integrated squared error (AMISE) criteria (see, e.g., Fan and Gijbels, 1996), i.e.,

$$\begin{aligned}
\mathrm{AMISE}_{\hat{\mu}} &= \int \Big([\mathrm{Bias}\{\hat{\mu}(u,z)|\mathbf{U},\mathbf{Z}\}]^2 + \mathbb{V}\{\hat{\mu}(u,z)|\mathbf{U},\mathbf{Z}\} \Big) f_{UZ}(u,z)\, d(u,z) \\
\mathrm{AMISE}_{\hat{\gamma}} &= \int \Big([\mathrm{Bias}\{\hat{\gamma}(u_1,u_2,z)|\mathbf{U},\mathbf{Z}\}]^2 + \mathbb{V}\{\hat{\gamma}(u_1,u_2,z)|\mathbf{U},\mathbf{Z}\} \Big) f_{UUZ}(u_1,u_2,z)\, d(u_1,u_2,z).
\end{aligned}$$

In anticipation of some of our results: Under AMISE optimal bandwidth choices, the discriminating θ-threshold is given by $\theta = 1/5$. That is, if $m/n^{1/5} \to 0$ (or $M/n^{1/5} \to 0$), the first variance term S_1^μ (or S_1^γ) is the leading variance term. This asymptotic scenario comprises situations where m (or M) is eventually negligible, i.e., very small in comparison to n. Therefore, we refer to this case as "sparse" covariate-adjusted functional data; see also Section 3.1.

If, however, $m/n^{1/5} \to \infty$ (or $M/n^{1/5} \to \infty$), then the second variance term S_2^μ (or S_2^γ)

is the leading variance term. Note that this asymptotic scenario comprises quite different situations where m (or M) is allowed to be eventually smaller than n, comparable to n, or larger than n. We refer to this asymptotic scenario as "dense" covariate-adjusted functional data; see also Section 3.2.

Depending on the asymptotic properties of their LLK estimators, Zhang et al. (2016) differentiate between "non-dense" (there $m/n^{1/4} \to 0$), "dense" (there $m/n^{1/4} \to C$), and "ultra-dense" (there $m/n^{1/4} \to \infty$) functional data. We do not consider their case of "dense" functional data (here $m/n^{1/5} \to C$), since this leads to a stalemate with variance terms having equal orders of magnitudes which makes it impossible to derive the explicit optimal bandwidth expressions.

3.1 Sparse covariate-adjusted functional data

The explicit AMISE optimal bandwidth expressions for the case of leading first variance terms, S_1^{μ} and S_1^{γ}, can be found in the following two Theorems 3.3 and 3.4. Particularly for the covariance estimator, the derivation of the multiple AMISE optimal bandwidths is a bit tedious, but nevertheless follows the usual steps for minimizing the classical bias-variance trade-off.

Theorem 3.3 (Optimal bandwidths for $\hat{\mu}$) *Let $m/n^{1/5} \to 0$. Under our setup the AMISE optimal bandwidths for estimating the covariate-adjusted mean function are then given by*

$$h_{\mu,U}^{S} = \left(\frac{R(K_\mu)\, Q_{\mu,1}\, \mathcal{I}_{\mu,ZZ}^{3/4}}{nm\, (\nu_2(K_\mu))^2 \left[\mathcal{I}_{\mu,UU}^{1/2}\, \mathcal{I}_{\mu,ZZ}^{1/2} + \mathcal{I}_{\mu,UZ} \right] \mathcal{I}_{\mu,UU}^{3/4}} \right)^{1/6} \tag{8}$$

$$h_{\mu,Z}^{S} = \left(\frac{\mathcal{I}_{\mu,UU}}{\mathcal{I}_{\mu,ZZ}} \right)^{1/4} h_{\mu,U}^{S}, \tag{9}$$

$$\begin{aligned} \textit{where} \qquad Q_{\mu,1} &= \int \left(\gamma(u,u,z)+\sigma_\epsilon^2\right)\, d(u,z), \\ \mathcal{I}_{\mu,UU} &= \int \left(\mu^{(2,0)}(u,z)\right)^2\, f_{UZ}(u,z)\, d(u,z), \\ \mathcal{I}_{\mu,ZZ} &= \int \left(\mu^{(0,2)}(u,z)\right)^2\, f_{UZ}(u,z)\, d(u,z), \quad \textit{and} \\ \mathcal{I}_{\mu,UZ} &= \int \mu^{(2,0)}(u,z)\mu^{(0,2)}(u,z)\, f_{UZ}(u,z)\, d(u,z). \end{aligned}$$

Theorem 3.4 (Optimal bandwidths for $\hat{\gamma}$) *Let $M/n^{1/5} \to 0$. Under our setup the AMISE optimal bandwidths for estimating the covariate-adjusted covariance function are then given by*

$$h^S_{\gamma,U} = \left(\frac{R(K_\gamma)\; Q_{\gamma,1}\; 4\,\sqrt{2}\,\mathcal{I}^{3/2}_{\gamma,ZZ}}{nM\; (\nu_2(K_\gamma))^2\; \left(2\; (\nu_2(K_\gamma))^2\, \mathcal{I}_{\gamma,U_{(1)}Z} + C_{\mathcal{I}}\right)\; \left(C_{\mathcal{I}} - \mathcal{I}_{\gamma,U_{(1)}Z}\right)^{3/2}} \right)^{1/7} \tag{10}$$

$$h^S_{\gamma,Z} = \left(\frac{C_{\mathcal{I}} - \mathcal{I}_{\gamma,U_{(1)}Z}}{2\,\mathcal{I}_{\gamma,ZZ}} \right)^{1/2} h^S_{\gamma,U}, \tag{11}$$

$$\begin{aligned} \textit{where} \quad C_{\mathcal{I}} &= (\mathcal{I}^2_{\gamma,U_{(1)}Z} + 4\,(\mathcal{I}_{\gamma,U_{(1)}U_{(1)}} + \mathcal{I}_{\gamma,U_{(1)}U_{(2)}})\,\mathcal{I}_{\gamma,ZZ})^{1/2}, \\ Q_{\gamma,1} &= \int (\tilde{\gamma}((u_1,u_2),(u_1,u_2),z) + \sigma^2_\varepsilon(u_1,u_2,z))\; d(u_1,u_2,z) \\ \mathcal{I}_{\gamma,U_{(1)}U_{(1)}} &= \int \left(\gamma^{(2,0,0)}(u_1,u_2,z)\right)^2\, f_{UUZ}(u_1,u_2,z)\, d(u_1,u_2,z), \\ \mathcal{I}_{\gamma,U_{(1)}U_{(2)}} &= \int \left(\gamma^{(2,0,0)}(u_1,u_2,z)\gamma^{(0,2,0)}(u_1,u_2,z)\right)\, f_{UUZ}(u_1,u_2,z)\, d(u_1,u_2,z), \\ \mathcal{I}_{\gamma,U_{(1)}Z} &= \int \gamma^{(2,0,0)}(u_1,u_2,z)\gamma^{(0,0,2)}(u_1,u_2,z)\, f_{UUZ}(u_1,u_2,z)\, d(u_1,u_2,z), \quad \textit{and} \\ \mathcal{I}_{\gamma,ZZ} &= \int \left(\gamma^{(0,0,2)}(u_1,u_2,z)\right)^2\, f_{UUZ}(u_1,u_2,z)\, d(u_1,u_2,z). \end{aligned}$$

The proofs of the above two Theorems 3.3 and 3.4 follow that of Herrmann et al. (1995) and can be found in Section 2 of the supplemental paper. The superscript "S" suggest that we are considering the case of "Sparse" covariate-adjusted functional data.

The above AMISE optimal bandwidths are of the well-known magnitudes for bi- and trivariate nonparametric estimators. The following two Corollaries 3.1 and 3.2 contain our asymptotic normality results for the estimators $\hat{\mu}$ and $\hat{\gamma}$ for sparse covariate-adjusted functional data.

Corollary 3.1 (Sparse - asymptotic normality of $\hat{\mu}$) *Let $m/n^{1/5} \to 0$, assume optimal bandwidth choices, i.e., $h_{\mu,U} = h^S_{\mu,U}$ and $h_{\mu,Z} = h^S_{\mu,Z}$, and let u and z be interior points of* $\mathrm{supp}(f_{UZ})$. *Under our setup the LLK estimator $\hat{\mu}(u,z)$ in Eq.* (5) *is then asymptotically normal, i.e.,*

$$T^S_\mu(u,z) = \left(\frac{\hat{\mu}(u,z; h^S_{\mu,U}, h^S_{\mu,Z}) - B^\mu(u,z; h^S_{\mu,U}, h^S_{\mu,Z}) - \mu(u,z)}{\sqrt{S^\mu_1(u,z; h^S_{\mu,U}, h^S_{\mu,Z})}} \right) \overset{a}{\sim} N(0,1),$$

and efficient with convergence rate $\hat{\mu}(u,z; h^S_{\mu,U}, h^S_{\mu,Z})|\mathbf{U},\mathbf{Z} = \mathcal{O}_p\left((nm)^{-1/3}\right)$.

Corollary 3.2 (Sparse - asymptotic normality of $\hat{\gamma}$) *Let $M/n^{1/5} \to 0$, assume optimal bandwidth choices, i.e., $h_{\gamma,U} = h^S_{\gamma,U}$ and $h_{\gamma,Z} = h^S_{\gamma,Z}$, and let u_1, u_2 and z be interior points of* $\mathrm{supp}(f_{UUZ})$. *Under our setup the LLK estimator in Eq.* (6) *is then asymptotically normal, i.e.,*

$$T^S_\gamma(u,z) = \left(\frac{\hat{\gamma}(u_1,u_2,z; h^S_{\gamma,U}, h^S_{\gamma,Z}) - B^\gamma(u_1,u_2,z; h^S_{\gamma,U}, h^S_{\gamma,Z}) - \gamma(u_1,u_2,z)}{\sqrt{S^\gamma_1(u_1,u_2,z; h^S_{\gamma,U}, h^S_{\gamma,Z})}} \right) \overset{a}{\sim} N(0,1)$$

and efficient with convergence rate $\hat{\gamma}(u_1,u_2,z; h^S_{\gamma,U}, h^S_{\gamma,Z})|\mathbf{U},\mathbf{Z} = \mathcal{O}_p\left((nM)^{-2/7}\right)$.

The proofs of the above two Corollaries 3.1 and 3.2 follow directly from Theorems 3.1, 3.2, 3.3, and 3.4 and can be found in Section 2 of the supplemental paper.

3.2 Dense covariate-adjusted functional data

If the second variance summands, S^μ_2 and S^γ_2, are the leading variance terms, it is possible to archive fast univariate convergence rates for the bi- and trivariate estimators $\hat{\mu}(u,z)$ and $\hat{\gamma}(u_1,u_2,z)$. By contrast to the preceding section, however, it is impossible to determine the optimal bandwidths by using only the leading variance terms, namely, S^μ_2 and S^γ_2. The trick is to determine the bandwidth expressions in a hierarchical manner: The optimal Z-bandwidths $h^D_{\mu,Z}$ and $h^D_{\gamma,Z}$ must be derived by optimizing with respect to the leading bias and variance terms, which are both purely Z-related. Given the optimal Z-bandwidths, the optimal U-bandwidths $h^D_{\mu,U}$ and $h^D_{\gamma,U}$ can be determined by optimizing the subsequent lower-order bias and variance terms. This leads to the following two Theorems 3.5 and 3.6.

Theorem 3.5 (Optimal bandwidths for $\hat{\mu}$) *Let $m/n^{1/5} \to \infty$. Under our setup the AMISE optimal bandwidths for estimating the covariate-adjusted mean function are then given by*

$$h^D_{\mu,Z} = \left(\frac{R(\kappa)\, Q_{\mu,2}}{n\ (\nu_2(K_\mu))^2\, \mathcal{I}_{\mu,ZZ}} \right)^{1/5} \quad \text{and} \tag{12}$$

$$h^D_{\mu,U} = \left(\frac{R(K_\mu)\, Q_{\mu,1}}{nm\ (\nu_2(K_\mu))^2\, \mathcal{I}_{\mu,UZ}} \right)^{1/3} \left(h^D_{\mu,Z}\right)^{-1}, \tag{13}$$

$$\begin{aligned} \text{where}\quad Q_{\mu,1} &= \int \left(\gamma(u,u,z) + \sigma^2_\epsilon\right)\, d(u,z), \\ Q_{\mu,2} &= \int \gamma(u,u,z)\, f_U(u)\, d(u,z), \\ \mathcal{I}_{\mu,ZZ} &= \int (\mu^{(0,2)}(u,z))^2\, f_{UZ}(u,z)\, d(u,z), \quad \text{and} \\ \mathcal{I}_{\mu,UZ} &= \int \mu^{(2,0)}(u,z)\mu^{(0,2)}(u,z)\, f_{UZ}(u,z)\, d(u,z). \end{aligned}$$

Theorem 3.6 (Optimal bandwidths for $\hat{\gamma}$) *Let $M/n^{1/5} \to \infty$. Under our setup the AMISE optimal bandwidths for estimating the covariate-adjusted covariance function are then given by*

$$h^D_{\gamma,Z} = \left(\frac{R(\kappa)\, Q_{\gamma,2}}{n\ (\nu_2(K_\gamma))^2\, \mathcal{I}_{\gamma,ZZ}} \right)^{1/5} \quad \text{and} \tag{14}$$

$$h^D_{\gamma,U} = \left(\frac{R(K_\gamma)\, Q_{\gamma,1}}{nM\ (\nu_2(K_\gamma))^2\, \mathcal{I}_{\gamma,U_{(1)}Z}} \right)^{1/4} \left(h^D_{\gamma,Z}\right)^{-3/4}, \tag{15}$$

$$\begin{aligned} \text{where}\quad Q_{\gamma,1} &= \int \left(\tilde{\gamma}((u_1,u_2),(u_1,u_2),z) + \sigma^2_\varepsilon(u_1,u_2,z)\right)\, d(u_1,u_2,z), \\ Q_{\gamma,2} &= \int \tilde{\gamma}((u_1,u_2),(u_1,u_2),z)\, f_{UU}(u_1,u_2)\, d(u_1,u_2,z), \\ \mathcal{I}_{\gamma,ZZ} &= \int (\gamma^{(0,0,2)}(u_1,u_2,z))^2\, f_{UUZ}(u_1,u_2,z)\, d(u_1,u_2,z), \quad \text{and} \\ \mathcal{I}_{\gamma,U_{(1)}Z} &= \int \gamma^{(2,0,0)}(u_1,u_2,z)\gamma^{(0,0,2)}(u_1,u_2,z)\, f_{UUZ}(u_1,u_2,z)\, d(u_1,u_2,z). \end{aligned}$$

The proofs of the above two Theorems 3.5 and 3.6 can be found in Section 3 of the supplemental paper. The superscript "D" suggest that we are considering the case of "Dense" covariate-adjusted functional data.

A surprising result is that the AMISE optimal U-bandwidths in Eqs. (13) and (15) are in a sense **anti**-proportional to each other: The larger the Z-bandwidths, the smaller the U-bandwidths, given fixed sample sizes n and m (and M). This is completely contrary to the classical multiple bandwidth results where the single bandwidths are directly proportional to each other.

In order to explain this finding, observe that a larger Z-bandwidth means that more functions $X_i(., Z_i)$ are used for computing (local) averages. Though, taking averages reduces variance and therefore we can afford some further increase in variance by using a smaller U-bandwidth ("under-smoothing" $\Rightarrow$ smaller-bias & larger-variance) for the sake of an overall better estimation performance through an overall lower bias. This under-smoothing strategy was already found to be optimal in the literature on the "pre-smoothing approach" (see, e.g., Benko et al., 2009). Our optimal multiple bandwidth expressions above allow now to implement this under-smoothing strategy also for the "pooling approach".

The following two Corollaries 3.3 and 3.4 contain our asymptotic normality results for the estimators $\hat{\mu}$ and $\hat{\gamma}$ for dense covariate adjusted functional data.

Corollary 3.3 (Dense - asymptotic normality of $\hat{\mu}$) *Let $m/n^{1/5} \to \infty$, assume optimal bandwidth choices, i.e., $h_{\mu,U} = h^D_{\mu,U}$ and $h_{\mu,Z} = h^D_{\mu,Z}$, and let u and z be interior points of* $\operatorname{supp}(f_{UZ})$. *Under our setup the LLK estimator in Eq.* (5) *is then asymptotically normal, i.e.,*

$$T^D_\mu(u,z) = \left(\frac{\hat{\mu}(u,z;h^D_{\mu,U},h^D_{\mu,Z}) - B^\mu(u,z;h^D_{\mu,U},h^D_{\mu,Z}) - \mu(u,z)}{\sqrt{S^\mu_2(u,z;h^D_{\mu,Z})}} \right) \overset{a}{\sim} N(0,1),$$

and efficient with convergence rate $\hat{\mu}(u,z;h^D_{\mu,U},h^D_{\mu,Z})|\mathbf{U},\mathbf{Z} = \mathcal{O}_p\left(n^{-2/5}\right)$.

Corollary 3.4 (Dense - asymptotic normality of $\hat{\gamma}$) *Let $M/n^{1/5} \to \infty$, assume optimal bandwidth choices, i.e., $h_{\gamma,U} = h^D_{\gamma,U}$ and $h_{\gamma,Z} = h^D_{\gamma,Z}$, and let u_1, u_2 and z be interior points of* $\operatorname{supp}(f_{UUZ})$. *Under our setup the LLK estimator in Eq.* (6) *is then asymptotically normal, i.e.,*

$$T^D_\gamma(u,z) = \left(\frac{\hat{\gamma}(u_1,u_2,z;h^D_{\gamma,U},h^D_{\gamma,Z}) - B^\gamma(u_1,u_2,z;h^D_{\gamma,U},h^D_{\gamma,Z}) - \gamma(u_1,u_2,z)}{\sqrt{S^\gamma_2(u_1,u_2,z;h^D_{\gamma,Z})}} \right) \overset{a}{\sim} N(0,1),$$

and efficient with convergence rate $\hat{\gamma}(u_1,u_2,z;h^D_{\gamma,U},h^D_{\gamma,Z})|\mathbf{U},\mathbf{Z} = \mathcal{O}_p\left(n^{-2/5}\right)$.

The proofs of the above two Corollaries 3.3 and 3.4 follow directly from Theorems 3.1, 3.2, 3.5, and 3.6 and can be found in Section 3 of the supplemental paper.

Note that convergence rates of the mean and covariance estimators are both that of a univariate LLK estimator (i.e., $\mathcal{O}(n^{-2/5})$). This is not surprising, since even if we could observe the random functions $X_i(.,Z_i)$ perfectly we are left with an univariate nonparametric smoothing problem due to the covariate Z.

3.3 Sparse-to-dense covariate-adjusted functional data

The main objective in our application is to test for differences in the covariate-adjusted mean price functions before and after Germany's partial nuclear phaseout. Therefore, we propose the following small-sample correction of the asymptotic normality result in Corollary 3.3:

Corollary 3.5 (Sparse-to-dense - asymptotic normality $\hat{\mu}$) *Under the same setup as in Corollary 3.3, we have that*

$$T^{SD}_\mu(u,z) = \left(\frac{\hat{\mu}(u,z;h^D_{\mu,U},h^D_{\mu,Z}) - B^\mu(u,z;h^D_{\mu,U},h^D_{\mu,Z}) - \mu(u,z)}{\sqrt{S^\mu_1(u,z;h^D_{\mu,U},h^D_{\mu,Z},h^D_{\gamma,U},h^D_{\gamma,Z}) + S^\mu_2(u,z;h^D_{\mu,Z},h^D_{\gamma,U},h^D_{\gamma,Z})}} \right) \overset{a}{\sim} N(0,1).$$

The proof of this result follows directly from observing that $S^\mu_2(u,z;h^D_{\mu,Z},h^D_{\gamma,U},h^D_{\gamma,Z}) = o(S^\mu_1(u,z;h^D_{\mu,U},h^D_{\mu,Z},h^D_{\gamma,U},h^D_{\gamma,Z}))$ which implies that $T^D_\mu(u,z)$ and $T^{SD}_\mu(u,z)$ are asymptotically equivalent, i.e., that $|(T^D(u,z)/T^{SD}(u,z)) - 1| = o_p(1)$.

As demonstrated in our simulation study in Section 5, the additional, asymptotically negligible, variance term S_1^{μ} in $T_{\mu}^{SD}(u,z)$ serves as a very effective small-sample correction.

4 Bandwidth, bias, and variance approximations

4.1 Bandwidth approximations

The bandwidth expressions in Eqs. (8), (9), (10), (11), (12), (13), (14), and (15) are infeasible as they depend on the following unknown quantities: $\mathcal{I}_{\mu,UU}$, $\mathcal{I}_{\mu,UZ}$, $\mathcal{I}_{\mu,ZZ}$, $Q_{\mu,1}$, $Q_{\mu,2}$, $\mathcal{I}_{\gamma,U_{(1)}U_{(1)}}$, $\mathcal{I}_{\gamma,U_{(1)}Z}$, $\mathcal{I}_{\gamma,ZZ}$, $Q_{\gamma,1}$, and $Q_{\gamma,2}$. Following Fan and Gijbels (1996) we suggest approximating the unknown quantities using global polynomial regression models (see below). We propose the following rule-of-thumb bandwidth approximations for the infeasible bandwidths in (8) and (9):

$$\hat{h}^S_{\mu,U} = \left(\frac{R(K_\mu)\, Q_{\mu,1}\, \hat{\mathcal{I}}^{3/4}_{\mu_{\text{poly}},ZZ}}{nm\ (\nu_2(K_\mu))^2\ \left[\hat{\mathcal{I}}^{1/2}_{\mu_{\text{poly}},UU}\, \hat{\mathcal{I}}^{1/2}_{\mu,ZZ} + \hat{\mathcal{I}}_{\mu_{\text{poly}},UZ}\right] \hat{\mathcal{I}}^{3/4}_{\mu_{\text{poly}},UU}} \right)^{1/6} \tag{16}$$

$$\hat{h}^S_{\mu_{\text{poly}},Z} = \left(\frac{\hat{\mathcal{I}}_{\mu_{\text{poly}},UU}}{\hat{\mathcal{I}}_{\mu_{\text{poly}},ZZ}} \right)^{1/4} \hat{h}^S_{\mu,U}. \tag{17}$$

The rule-of-thumb approximations for the infeasible bandwidths in (10) and (11) are

$$\hat{h}^S_{\gamma,U} = \left(\frac{R(K_\gamma)\ \hat{Q}_{\gamma_{\text{poly}},1}\ 4\, \sqrt{2}\, \hat{\mathcal{I}}^{3/2}_{\gamma_{\text{poly}},ZZ}}{nM\ (\nu_2(K_\gamma))^2\ \left(2\ (\nu_2(K_\gamma))^2\ \hat{\mathcal{I}}_{\gamma_{\text{poly}},U_{(1)}Z} + \hat{C}_{\mathcal{I}}\right)\ \left(\hat{C}_{\mathcal{I}} - \hat{\mathcal{I}}_{\gamma_{\text{poly}},U_{(1)}Z}\right)^{3/2}} \right)^{1/7} \tag{18}$$

$$\hat{h}^S_{\gamma,Z} = \left(\frac{\hat{C}_{\mathcal{I}} - \hat{\mathcal{I}}_{\gamma_{\text{poly}},U_{(1)}Z}}{2\, \hat{\mathcal{I}}_{\gamma,ZZ}} \right)^{1/2} \hat{h}^S_{\gamma,U}, \tag{19}$$

where $\hat{C}_{\mathcal{I}} = (\hat{\mathcal{I}}^2_{\gamma_{\text{poly}},U_{(1)}Z} + 4\, (\hat{\mathcal{I}}_{\gamma_{\text{poly}},U_{(1)}U_{(1)}} + \hat{\mathcal{I}}_{\gamma_{\text{poly}},U_{(1)}U_{(2)}})\, \hat{\mathcal{I}}_{\gamma_{\text{poly}},ZZ})^{1/2}$.

The rule-of-thumb approximations for the infeasible bandwidths in (12) and (13) are

$$\hat{h}^D_{\mu,Z} = \left(\frac{R(\kappa)\, \hat{Q}_{\mu_{\text{poly}},2}}{n\ (\nu_2(K_\mu))^2\ \hat{\mathcal{I}}_{\mu_{\text{poly}},ZZ}} \right)^{1/5} \quad \text{and} \tag{20}$$

$$\hat{h}^D_{\mu,U} = \left(\frac{R(K_\mu)\, \hat{Q}_{\mu_{\text{poly}},1}}{nm\ (\nu_2(K_\mu))^2\ \hat{\mathcal{I}}_{\mu_{\text{poly}},UZ}} \right)^{1/3} \left(\hat{h}^D_{\mu,Z}\right)^{-1}. \tag{21}$$

The rule-of-thumb approximations for the infeasible bandwidths in (14) and (15) are

$$\hat{h}^{D}_{\gamma,Z} = \left(\frac{R(\kappa)\,\hat{Q}_{\gamma_{\text{poly}},2}}{n\,\left(\nu_2(K_\gamma)\right)^2\,\hat{\mathcal{I}}_{\gamma_{\text{poly}},ZZ}}\right)^{1/5} \quad \text{and} \tag{22}$$

$$\hat{h}^{D}_{\gamma,U} = \left(\frac{R(K_\gamma)\,\hat{Q}_{\gamma_{\text{poly}},1}}{nM\,\left(\nu_2(K_\gamma)\right)^2\,\hat{\mathcal{I}}_{\gamma_{\text{poly}},U_{(1)}Z}}\right)^{1/4} \left(\hat{h}^{D}_{\gamma_{\text{poly}},Z}\right)^{-3/4}. \tag{23}$$

The above rule-of-thumb bandwidth expressions for the mean function are based on the following approximations:

$$\begin{aligned}
\hat{\mathcal{I}}_{\mu_{\text{poly}},UU} &= \textstyle\int_{\text{supp}(f_{UZ})}(\hat{\mu}^{(2,0)}_{\text{poly}}(u,z))^2\,\hat{f}_{UZ}(u,z)\,d(u,z),\\
\hat{\mathcal{I}}_{\mu_{\text{poly}},UZ} &= \textstyle\int_{\text{supp}(f_{UZ})}\hat{\mu}^{(2,0)}_{\text{poly}}(u,z)\hat{\mu}^{(0,2)}_{\text{poly}}(u,z)\,\hat{f}_{UZ}(u,z)\,d(u,z),\\
\hat{\mathcal{I}}_{\mu_{\text{poly}},ZZ} &= \textstyle\int_{\text{supp}(f_{UZ})}(\hat{\mu}^{(0,2)}_{\text{poly}}(u,z))^2\,\hat{f}_{UZ}(u,z)\,d(u,z),\\
\hat{Q}_{\mu_{\text{poly}},1} &= \textstyle\int_{\text{supp}(f_{UZ})}\hat{\gamma}^{\text{ND}}_{\text{poly}}(u,u,z)\,d(u,z), \quad \text{and}\\
\hat{Q}_{\mu_{\text{poly}},2} &= \textstyle\int_{\text{supp}(f_{UZ})}\hat{\gamma}_{\text{poly}}(u,u,z)\,\hat{f}_{U}(u)\,d(u,z).
\end{aligned}$$

The above rule-of-thumb bandwidth expressions for the covariance function are based on the following approximations:

$$\begin{aligned}
\hat{\mathcal{I}}_{\gamma_{\text{poly}},U_{(1)}U_{(1)}} &= \textstyle\int_{\text{supp}(f_{UZ})}(\hat{\gamma}^{(2,0,0)}_{\text{poly}}(u,u,z))^2\,\hat{f}_{UZ}(u,z)\,d(u,z),\\
\hat{\mathcal{I}}_{\gamma_{\text{poly}},U_{(1)}Z} &= \textstyle\int_{\text{supp}(f_{UUZ})}\hat{\gamma}^{(2,0,0)}_{\text{poly}}(u,u,z)\hat{\gamma}^{(0,0,2)}_{\text{poly}}(u,u,z)\,\hat{f}_{UUZ}(u,u,z)\,d(u,u,z),\\
\hat{\mathcal{I}}_{\gamma_{\text{poly}},ZZ} &= \textstyle\int_{\text{supp}(f_{UUZ})}(\hat{\gamma}^{(0,0,2)}_{\text{poly}}(u,u,z))^2\,\hat{f}_{UUZ}(u,u,z)\,d(u,u,z),\\
\hat{Q}_{\gamma_{\text{poly}},1} &= \textstyle\int_{\text{supp}(f_{UUZ})}\hat{\hat{\gamma}}^{\text{ND}}_{\text{poly}}((u_1,u_2),(u_1,u_2),z)\,d(u_1,u_2,z), \quad \text{and}\\
\hat{Q}_{\gamma_{\text{poly}},2} &= \textstyle\int_{\text{supp}(f_{UUZ})}\hat{\hat{\gamma}}_{\text{poly}}((u_1,u_2),(u_1,u_2),z)\,\hat{f}_{UU}(u_1,u_2)\,d(u_1,u_2,z).
\end{aligned}$$

The estimates $\hat{\mu}_{\text{poly}}$, $\hat{\gamma}^{\text{ND}}_{\text{poly}}$, $\hat{\gamma}_{\text{poly}}$, $\hat{\hat{\gamma}}^{\text{ND}}_{\text{poly}}$, $\hat{\hat{\gamma}}_{\text{poly}}$, $\hat{\mu}^{(2,0)}_{\text{poly}}$, $\hat{\mu}^{(0,2)}_{\text{poly}}$, $\hat{\gamma}^{(2,0,0)}_{\text{poly}}$, and $\hat{\gamma}^{(0,0,2)}_{\text{poly}}$ are the ordinary least squares estimates (and their derivatives) of the following polynomial regression models:

$\boldsymbol{\mu}_{\textbf{poly}}$**:** The model $\mu_{\text{poly}}(u,z)$ is fitted via regressing Y_{ij} on powers (each up to the fourth power) of U_{ij}, Z_i, and $U_{ij}\cdot Z_i$ for all $i\in\{1,\dots,n\}$ and $j\in\{1,\dots,m\}$, i.e., $Y_{ij} = \mu_{\text{poly}}(U_{ij},Z_i)+\text{error}_{ij}$, where $\mu_{\text{poly}}(U_{ij},Z_i) := \beta_0+\sum_{q=1}^{4}\left(\beta^U_q U^q_{ij}+\beta^Z_q Z^q_{ij}+\beta^{UZ}_q(U_{ij}Z_i)^q\right)$.

$\boldsymbol{\gamma}_{\mathbf{poly}}$: The model $\gamma_{\text{poly}}(u_1, u_2, z)$ is fitted via regressing C_{ijk}^{poly} on powers (each up to the fourth power) of U_{ij}, U_{ik}, Z_i, $U_{ij} \cdot Z_i$, and $U_{ik} \cdot Z_i$ for all $i \in \{1, \ldots, n\}$ and all $j, k \in \{1, \ldots, m\}$ with $j \neq k$, i.e., $C_{ijk}^{\text{poly}} = \gamma_{\text{poly}}(U_{ij}, U_{ik}, Z_i) + \text{error}_{ijk}$, where $C_{ijk}^{\text{poly}} := (Y_{ij} - \mu_{\text{poly}}(U_{ij}, Z_i))(Y_{jt} - \mu_{\text{poly}}(U_{ik}, Z_i))$ and
$\gamma_{\text{poly}}(U_{ij}, U_{ik}, Z_i) := \beta_0 + \sum_{q=1}^{4} \left(\beta_q^{U,1} U_{ij}^q + \beta_q^{U,2} U_{ik}^q + \beta_q^Z Z_i^q + \beta_q^{UZ,1} (U_{ij} Z_i)^q + \beta_q^{UZ,2} (U_{ik} Z_i)^q\right)$.

$\tilde{\boldsymbol{\gamma}}_{\mathbf{poly}}$: The model $\tilde{\gamma}_{\text{poly}}((u_1, u_2), (u_3, u_4), z)$ is fitted via regressing $\mathbb{C}_{ijk\ell m}^{\text{poly}}$ on powers (each up to the fourth power) of U_{ij}, U_{ik}, $U_{i\ell}$, U_{im}, and Z_i for all $i \in \{1, \ldots, n\}$ and all $j, k, \ell, m \in \{1, \ldots, m\}$ such that ($j \neq \ell$ AND $k \neq m$), i.e., $\mathbb{C}_{ijk\ell m}^{\text{poly}} = \tilde{\gamma}_{\text{poly}}((U_{ij}, U_{ik}), (U_{i\ell}, U_{im}), Z_i) + \text{error}_{ijk\ell m}$, where $\mathbb{C}_{ijk\ell m}^{\text{poly}} := (C_{ijk}^{\text{poly}} - \gamma_{\text{poly}}(U_{ij}, U_{ik}, Z_i))(C_{i\ell m}^{\text{poly}} - \gamma_{\text{poly}}(U_{i\ell}, U_{im}, Z_i))$ and
$\tilde{\gamma}_{\text{poly}}((U_{ij}, U_{ik}), (U_{i\ell}, U_{im}), Z_i) := \beta_0 + \sum_{q=1}^{4} \left(\beta_q^{U,1} U_{ij}^q + \beta_q^{U,2} U_{ik}^q + \beta_q^{U,3} U_{i\ell}^q + \beta_q^{U,4} U_{im}^q + \beta_q^Z Z_i^q + \right.$ $\left. + \beta_q^{UZ,1} (U_{ij} Z_i)^q + \beta_q^{UZ,2} (U_{ik} Z_i)^q + \beta_q^{UZ,3} (U_{i\ell} Z_i)^q + \beta_q^{UZ,4} (U_{im} Z_i)^q\right)$.

$\boldsymbol{\gamma}_{\mathbf{poly}}^{\mathbf{S}}$: The model $\gamma_{\text{poly}}^{\text{ND}}(u, u, z)$ is fitted via regressing the noise contaminated diagonal values C_{ij}^{poly} on powers (each up to the fourth power) of U_{ij}, and Z_i for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$, i.e., $C_{ij}^{\text{poly}} = \gamma_{\text{poly}}^{\text{ND}}(U_{ij}, U_{ij}, Z_i) + \text{error}_{ij}$, where $C_{ij}^{\text{poly}} := (Y_{ij} - \mu_{\text{poly}}(U_{ij}, Z_i))^2$ and $\gamma_{\text{poly}}^{\text{ND}}(U_{ij}, U_{ij}, Z_i) := \beta_0 + \sum_{q=1}^{4} \left(\beta_q^U U_{ij}^q + \beta_q^Z Z_i^q + \beta_q^{UZ} (U_{ij} Z_i)^q\right)$. The "ND" in $\gamma_{\text{poly}}^{\text{ND}}$ suggest that we are estimating the noise contaminated diagonal values $\gamma(u, u, z) + \sigma_\epsilon$.

$\tilde{\boldsymbol{\gamma}}_{\mathbf{poly}}^{\mathbf{S}}$: The model $\tilde{\gamma}_{\text{poly}}^{\text{ND}}(u_1, u_2, z)$ is fitted via regressing the noise contaminated diagonal values $\mathbb{C}_{ijk}^{\text{poly}}$ on powers (each up to the fourth power) of U_{ij}, U_{ik}, and Z_i for all $i \in \{1, \ldots, n\}$, and $j, k \in \{1, \ldots, m\}$, i.e., $\mathbb{C}_{ijk}^{\text{poly}} = \tilde{\gamma}_{\text{poly}}^{\text{ND}}(U_{ij}, U_{ik}, Z_i) + \text{error}_{ijk}$, where $\mathbb{C}_{ijk}^{\text{poly}} := (C_{ijk}^{\text{poly}} - \gamma_{\text{poly}}(U_{ij}, U_{ik}, Z_i))^2$ and $\tilde{\gamma}_{\text{poly}}^{\text{ND}}(U_{ij}, U_{ik}, Z_i) := \beta_0 + \sum_{q=1}^{4} \left(\beta_q^{U,1} U_{ij}^q + \beta_q^{U,2} U_{ik}^q + \beta_q^Z Z_i^q\right)$. The "ND" in $\tilde{\gamma}_{\text{poly}}^{\text{ND}}$ suggest that we are estimating the noise contaminated diagonal values $\tilde{\gamma}((u_1, u_2), (u_1, u_2), z) + \sigma_\varepsilon(u_1, u_2, z)$.

Estimates of the densities f_{UZ} and f_{UUZ} are computed as kernel density estimates, where we use the default implementations of the R-package ks (Duong et al., 2007).

Remark It is important to specify the models "$\boldsymbol{\mu}_{\text{poly}}$" and "$\boldsymbol{\gamma}_{\text{poly}}$" using interaction terms, since otherwise their partial derivatives $\hat{\mu}^{(2,0)}_{\text{poly}}$, $\hat{\mu}^{(0,2)}_{\text{poly}}$, $\hat{\gamma}^{(2,0,0)}_{\text{poly}}$, and $\hat{\gamma}^{(0,0,2)}_{\text{poly}}$ would degenerate.

4.2 Bias and variance approximations

In our application, we are interested in doing inference about the mean functions before and after Germany's partial nuclear phaseout. For this, we want to use our sparse, dense, and sparse-to-dense asymptotic normality results in Corollaries 3.1, 3.3, and 3.5. This requires us to approximate the unknown quantities $\mu^{(2,0)}$ and $\mu^{(0,2)}$ in the bias term B^{μ} and the unknown quantities $\gamma(u,u,z)+\sigma^2_\epsilon$ and $\gamma(u,u,z)$ in the variance terms S^{μ}_1 and S^{μ}_2.

For approximating the unknown quantities $\mu^{(2,0)}$ and $\mu^{(0,2)}$ in the bias term B_{μ}, we propose to use local polynomial kernel estimators of $\mu^{(2,0)}$ and $\mu^{(0,2)}$, i.e.:

$$\hat{B}^{\mu}(u,z;h_{\mu,U},h_{\mu,Z}) = \tfrac{\nu_2(K_\mu)}{2}\Big(h^2_{\mu,U}\hat{\mu}^{(2,0)}\big(u,z;g_{\mu,U},g_{\mu,Z}\big) + h^2_{\mu,Z}\hat{\mu}^{(0,2)}\big(u,z;g_{\mu,U},g_{\mu,Z}\big)\Big) \tag{24}$$

where $\hat{\mu}^{(2,0)}$ and $\hat{\mu}^{(0,2)}$ are local polynomial (order 3) kernel estimators $\mu^{(2,0)}$ and $\mu^{(0,2)}$:

$$\hat{\mu}^{(2,0)}(u,z;g_{\mu,U},g_{\mu,Z}) = 2!\, e_3^\top \left([\mathbf{1},\mathbf{U}_u^{1:3},\mathbf{Z}_z^{1:3}]^\top \mathbf{W}_{\mu,uz}[\mathbf{1},\mathbf{U}_u^{1:3},\mathbf{Z}_z^{1:3}]\right)^{-1}[\mathbf{1},\mathbf{U}_u^{1:3},\mathbf{Z}_z^{1:3}]^\top \mathbf{W}_{\mu,uz}\mathbf{Y}$$
$$\hat{\mu}^{(0,2)}(u,z;g_{\mu,U},g_{\mu,Z}) = 2!\, e_6^\top \left([\mathbf{1},\mathbf{U}_u^{1:3},\mathbf{Z}_z^{1:3}]^\top \mathbf{W}_{\mu,uz}[\mathbf{1},\mathbf{U}_u^{1:3},\mathbf{Z}_z^{1:3}]\right)^{-1}[\mathbf{1},\mathbf{U}_u^{1:3},\mathbf{Z}_z^{1:3}]^\top \mathbf{W}_{\mu,uz}\mathbf{Y},$$

with $e_3^\top=(0,0,1,0,0,0,0)$, $e_6^\top=(0,0,0,0,0,1,0)$, $\mathbf{U}_u^{1:3}=[\mathbf{U}_u,\mathbf{U}_u^2,\mathbf{U}_u^3]$, $\mathbf{Z}_z^{1:3}=[\mathbf{Z}_z,\mathbf{Z}_z^2,\mathbf{Z}_z^3]$, and $\mathbf{W}_{\mu,uz}=\operatorname{diag}(\dots,g^{-1}_{\mu,U}\,\kappa(g^{-1}_{\mu,U}(U_{ij}-u))\;g^{-1}_{\mu,Z}\,\kappa(g^{-1}_{\mu,Z}(Z_i-z)),\dots)$.

Estimating $\mu^{(2,0)}(u,z)$ and $\mu^{(0,2)}(u,z)$ requires larger smoothing parameters than when estimating $\mu(u,z)$ (see, e.g., Gasser and Müller, 1984). In order to account for this, we simply use $g_{\mu,U}=2h_{\mu,U}$ and $g_{\mu,Z}=2h_{\mu,Z}$. This very practical approach is successfully applied in Härdle and Bowman (1988) and performs very well in our simulation study. A more advanced bandwidth scaling method might be based on Eq. (3.21) in Fan and Gijbels (1996), though, this was not considered here as it would involve further impractical global polynomial regressions.

For approximating the unknown quantities $\gamma(u,u,z)+\sigma_\epsilon^2$ and $\gamma(u,u,z)$ in the variance terms S_1^μ and S_2^μ, we propose to use LLK estimators, i.e.:

$$\hat{S}_1^\mu(u,z;\hat{h}_{\mu,U}^S,\hat{h}_{\mu,Z}^S,\hat{h}_{\gamma,U}^S,\hat{h}_{\gamma,Z}^S) = (nm)^{-1}\left[(\hat{h}_{\mu,U}^S\hat{h}_{\mu,Z}^S)^{-1}R(K_\mu)\frac{\hat{\gamma}^{\mathrm{ND}}(u,u,z;\hat{h}_{\gamma,U}^S,\hat{h}_{\gamma,Z}^S)}{\hat{f}_{UZ}(u,z)}\right] \quad (25)$$

$$\hat{S}_2^\mu(u,z;\hat{h}_{\mu,Z}^D,\hat{h}_{\gamma,U}^D,\hat{h}_{\gamma,Z}^D) = n^{-1}\left[\left(\frac{m-1}{m}\right)(\hat{h}_{\mu,Z}^D)^{-1}\,R(\kappa)\frac{\hat{\gamma}(u,u,z;\hat{h}_{\gamma,U}^D,\hat{h}_{\gamma,Z}^D)}{\hat{f}_Z(z)}\right], \quad (26)$$

where $\hat{\gamma}(u,u,z;\hat{h}_{\gamma,U}^D,\hat{h}_{\gamma,Z}^D)$ is the LLK estimator as defined in (6), and where the estimator of the Noisy Diagonal (ND), $\hat{\gamma}^{\mathrm{ND}}(u,u,z;\hat{h}_{\gamma,U}^S,\hat{h}_{\gamma,Z}^S) \approx \{\gamma(u,u,z)+\sigma_\epsilon^2\}$, is defined as the following LLK estimator:

$$\hat{\gamma}^{\mathrm{ND}}(u,u,z;\hat{h}_{\gamma,U}^S,\hat{h}_{\gamma,Z}^S) = e_1^\top\left([\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top\mathbf{W}_{\gamma^S,uz}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]\right)^{-1}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top\mathbf{W}_{\gamma^S,uz}\hat{\mathbf{C}},$$

where $\hat{\mathbf{C}} = (\hat{C}_{111},\ldots,\hat{C}_{nmm})^\top$. Note that $\hat{\gamma}^{\mathrm{ND}}$ is equivalent to the LLK estimator "$\hat{V}$" in Jiang and Wang (2010).

5 Simulation

We generate data from $Y_{ij} = \mu(U_{ij},Z_i) + X_i^c(U_{ij},Z_i) + \epsilon_{ij}$, where $U_{ij} \sim \mathrm{Unif}(0,1)$, $Z_i \sim \mathrm{Unif}(0,1)$, $\epsilon_{ij} \sim \mathrm{N}(0,1/2)$, $X_i^c(u,z) = \xi_{i1}\psi_1(u,z) + \xi_{i2}\psi_2(u,z)$, $\xi_{i1} \sim \mathrm{N}(0,2)$, $\xi_{i2} \sim \mathrm{N}(0,1)$, $\gamma(u_1,u_2,z) = 2\psi_1(u_1,z)\psi_1(u_2,z) + \psi_2(u_1,z)\psi_2(u_2,z)$, $m \in \{5,10,15\}$, and $n = 100$. We use the following two specifications for the mean and the basis functions ψ_1 and ψ_2:

DGP-1 $\mu(u,z) = 1+2u^2+2z^3+2(uz)^2$, $\psi_1(u,z) = u^2+z^2+(uz)^2$, $\psi_2(u,z) = u^3+z^3+(uz)^3$

DGP-2 $\mu(u,z) = 5\sin(\pi uz/2)$, $\psi_1(u,z) = \sin(\pi uz)$, $\psi_2(u,z) = \sin(2\pi uz)$

The main objective in our application is to test for differences in the covariate-adjusted mean price functions before and after Germany's partial nuclear phaseout. Therefore, we investigate the approximation properties of the finite sample distributions of the following three test statistics for the mean function $\mu(u,z)$:

- "Sparse":

$$\hat{T}^S_\mu(u,z) = \left(\frac{\hat{\mu}(u,z;\hat{h}^S_{\mu,U},\hat{h}^S_{\mu,Z}) - \hat{B}^\mu(u,z;\hat{h}^S_{\mu,U},\hat{h}^S_{\mu,Z}) - \mu(u,z)}{\sqrt{\hat{S}^\mu_1(u,z;\hat{h}^S_{\mu,U},\hat{h}^S_{\mu,Z},\hat{h}^S_{\gamma,U},\hat{h}^S_{\gamma,Z})}} \right) \tag{27}$$

which is based on Corollary 3.1, the sparse rule-of-thumb bandwidths (16), (17), (18), (19), the bias approximation (24), and the variance approximation (25).

- "Dense":

$$\hat{T}^D_\mu(u,z) = \left(\frac{\hat{\mu}(u,z;\hat{h}^D_{\mu,U},\hat{h}^D_{\mu,Z}) - \hat{B}^\mu(u,z;\hat{h}^D_{\mu,U},\hat{h}^D_{\mu,Z}) - \mu(u,z)}{\sqrt{\hat{S}^\mu_2(u,z;\hat{h}^D_{\mu,Z},\hat{h}^D_{\gamma,U},\hat{h}^D_{\gamma,Z})}} \right), \tag{28}$$

which is based on Corollary 3.3, the dense rule-of-thumb bandwidths (20), (21), (22), (23), the bias approximation (24), and the variance approximation (26).

- "Sparse-to-Dense":

$$\hat{T}^{SD}_\mu(u,z) = \left(\frac{\hat{\mu}(u,z;\hat{h}^D_{\mu,U},\hat{h}^D_{\mu,Z}) - \hat{B}^\mu(u,z;\hat{h}^D_{\mu,U},\hat{h}^D_{\mu,Z}) - \mu(u,z)}{\sqrt{\hat{S}^\mu_1(u,z;\hat{h}^D_{\mu,U},\hat{h}^D_{\mu,Z},\hat{h}^D_{\gamma,U},\hat{h}^D_{\gamma,Z}) + \hat{S}^\mu_2(u,z;\hat{h}^D_{\mu,Z},\hat{h}^D_{\gamma,U},\hat{h}^D_{\gamma,Z})}} \right), \tag{29}$$

which is based on Corollary 3.5, the dense rule-of-thumb bandwidths (20), (21), (22), (23), the bias approximation (24), and the variance approximations (26) and (25), but with (25) using the dense rule-of-thumb bandwidths.

For each of in total 500 Monte-Carlo repetitions we compute realizations of $\hat{T}^S_\mu(u,z)$, $\hat{T}^D_\mu(u,z)$, and $\hat{T}^{SD}_\mu(u,z)$ at the point $(u,z) = (0.5, 0.5)$. This point is chosen, since the mean functions have non-zero second order derivative at this point (i.e., $\mu^{(2,0)}_{\text{DGP-1}}(0.5,0.5) = 5$, $\mu^{(0,2)}_{\text{DGP-1}}(0.5,0.5) = 7$, and $\mu^{(2,0)}_{\text{DGP-2}}(0.5,0.5) = \mu^{(0,2)}_{\text{DGP-2}}(0.5,0.5) = -1.180292$). This allows us to asses the small-sample performance of the bias-correction terms $\hat{B}_\mu$. Note that the investigation of the finite sample distributions of $\hat{T}^S_\mu(u,z)$, $\hat{T}^D_\mu(u,z)$, and $\hat{T}^{SD}_\mu(u,z)$ includes also an investigation of the finite sample performance of the rule-of-thumb bandwidths for the covariance function.

	DGP-1			DGP-2		
	$m = 5$	$m = 10$	$m = 15$	$m = 5$	$m = 10$	$m = 15$
Sparse	0.33	0.51	0.49	0.49	0.74	0.73
Dense	0.31	0.14	0.11	0.34	0.21	0.11
Sparse-to-Dense	**0.08**	**0.04**	**0.06**	**0.16**	**0.12**	**0.07**

Table 1: Kullback-Leibler distances between the empirical distributions of the approximate test statistics $\hat{T}^S(u, z)$, $\hat{T}^D(u, z)$, and $\hat{T}^{SD}(u, z)$, at $(u, z) = (0.5, 0.5)$, and the standard normal distribution.

In order to assess the small-sample performance of the empirical distributions of the test statistics $\hat{T}^S_\mu(u, z)$, $\hat{T}^D_\mu(u, z)$, and $\hat{T}^{SD}_\mu(u, z)$, we compute their kernel density estimates $\hat{f}^S_T$, $\hat{f}^D_T$, and $\hat{f}^{SD}_T$ using Gaussian kernels and normal-reference-bandwidths. In Table 1 we present the Kullback-Leibler distances $\mathrm{KL}(\hat{f}|\phi)$ for $\hat{f} \in \{\hat{f}^S_T, \hat{f}^D_T, \hat{f}^{SD}_T\}$ from the standard normal distribution ϕ, where $\mathrm{KL}(\hat{f}|\phi) := \int \hat{f}(s) \cdot \log(\hat{f}(s)/\phi(s))ds$. The results clearly show that the sparse test statistic $\hat{T}^S_\mu(u, z)$ is outperformed by the dense and sparse-to-dense test statistics $\hat{T}^D_\mu(u, z)$ and $\hat{T}^{SD}_\mu(u, z)$. Furthermore, the sparse-to-dense test statistic $\hat{T}^{SD}_\mu(u, z)$ shows the uniformly best performance for both DGPs and for all chosen sample sizes.

In order to get also a visual impression of the empirical small-sample distributions of the approximate test statistics $\hat{T}^S(u, z)$, $\hat{T}^D(u, z)$, and $\hat{T}^{SD}(u, z)$, we show their kernel density estimates in Figure 3. The empirical distributions of the sparse test statistic $\hat{T}^S(u, z)$ show the fattest tails under both DGPs and for all chosen sample sizes. In contrast, the empirical distributions of the sparse-to-dense test statistic $\hat{T}^{SD}(u, z)$ is closest to the target density ϕ and performs surprisingly good in all of the considered sample sizes. The additional, asymptotically negligible, variance term $\hat{S}^\mu_1$ in $\hat{T}^{SD}_\mu(u, z)$ obviously serves as a very effective small-sample correction.

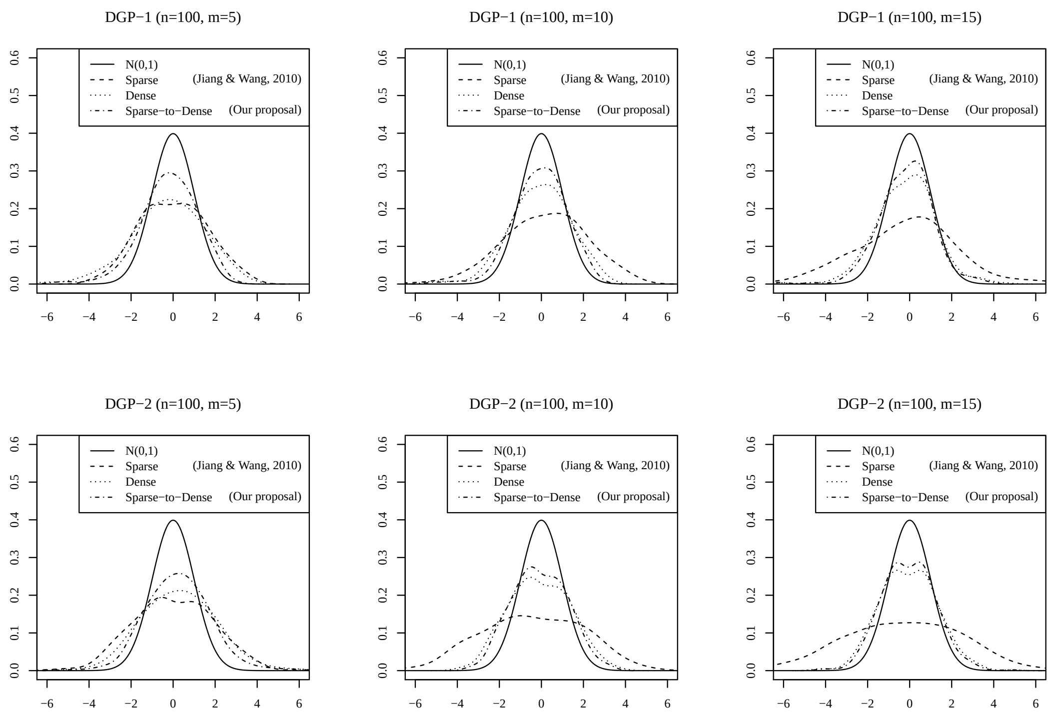

Figure 3: Kernel density estimates (Gaussian kernel and normal-reference-bandwidths) of the empirical distributions of $\hat{T}^{S}(u,z)$, $\hat{T}^{D}(u,z)$, and $\hat{T}^{SD}(u,z)$, with $u=0.5$ and $z=0.5$, versus the density function of the standard normal distribution.

6 Application

In our real data study, we analyze electricity spot prices of the German power market traded at the European Energy Power Exchange (EPEX). The EPEX spot price is of fundamental importance as a benchmark and reference price for many other markets, such as over-the-counter and forward markets (Grimm et al. (2008), Ch. 1). The data for our analysis come from four different sources that are described in detail in Section 4 of the supplemental paper.

6.1 Preliminary descriptions and limitations

The German electricity market, like many others, provides purchase guarantees for renewable energy sources (RES). Therefore, the relevant variable for pricing is electricity demand minus electricity infeeds from RES. Correspondingly, in our application U_{ij} refers to *residual* electricity demand defined as $U_{ij} = \texttt{Elect.Demand}_{it} - \texttt{RES}_{it}$, where $\texttt{RES}_{it} = \texttt{Wind.Infeed}_{it} + \texttt{Solar.Infeed}_{it}$. The effect of further RES such as biomass is still negligible for the German electricity market.

There are some clear limitations in our empirical results. The fundamental price drivers are (residual) electricity demand, temperate, and prices for uranium, lignite, coal, and gas. Our approach, however, only allows to control for the nonlinear effects of electricity (residual) demand and temperature. Therefore, we cannot disentangle shifts in the mean prices into their demand- and supply-side components. Furthermore, we cannot isolate the mean price effects that are due to Germany's nuclear phase out as this event was accompanied by an increase in Germany's resource prices which further increased the electricity prices. That is, we cannot identify the underlying causalities of changes in the conditional mean prices.

Nevertheless, only after controlling for the nonlinear effects of electricity demand we can reasonably compare electricity mean prices. Electricity prices are unit prices measured in "euros per 1megawatt·1hour" (EUR/MWh). But unit prices are only meaningful in relation with the total amount of demanded units. For instance, a unit price of 40EUR/MWh for in total 6000MW is effectively much cheaper than a unit price of 40EUR/MWh for in total 3000MW. Therefore, a reasonable comparison of electricity spot prices necessarily has to condition on electricity demand. Using our statistical model we can compute these conditional means which is not a trivial task due to the involved nonlinearities.

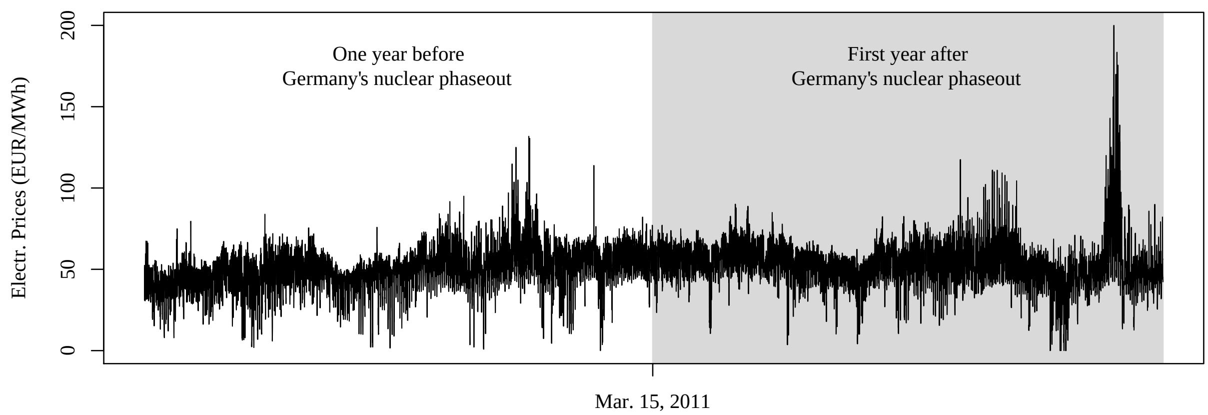

Figure 4: Time series of Germany's hourly electricity spot prices traded on the European Energy Exchange.

6.2 Empirical analysis

On March 15, 2011, just after the nuclear meltdown in Fukushima Daiichi, Japan, Germany decided to switch to a renewable energy economy and initiated this by an immediate and permanent shutdown of about 40% its nuclear power plants. This substantial loss of nuclear power with its low marginal production costs raised concerns about increases in electricity prices and subsequent problems for industry and households. However, empirical studies that build upon univariate time series analysis do not report any clear price effects (see, e.g., Nestle, 2012). A look at the univariate time series of Germany's hourly electricity spot prices, as shown in Figure 4, confirms this finding: Except for the very high prices at the end of the first year after Germany's (partial) nuclear phaseout, it is impossible to identify obvious mean shifts. Though, as discussed above, such a naive approach without conditioning on electricity demand does not allow for a reasonable comparison of the electricity mean prices.

In order to estimate and to test the conditional mean prices one year before and one year after Germany's nuclear phaseout, we apply our local linear estimators in Eqs. (5) and (6) to the two data sets shown in Figure 2. For estimation we use the dense rule-of-

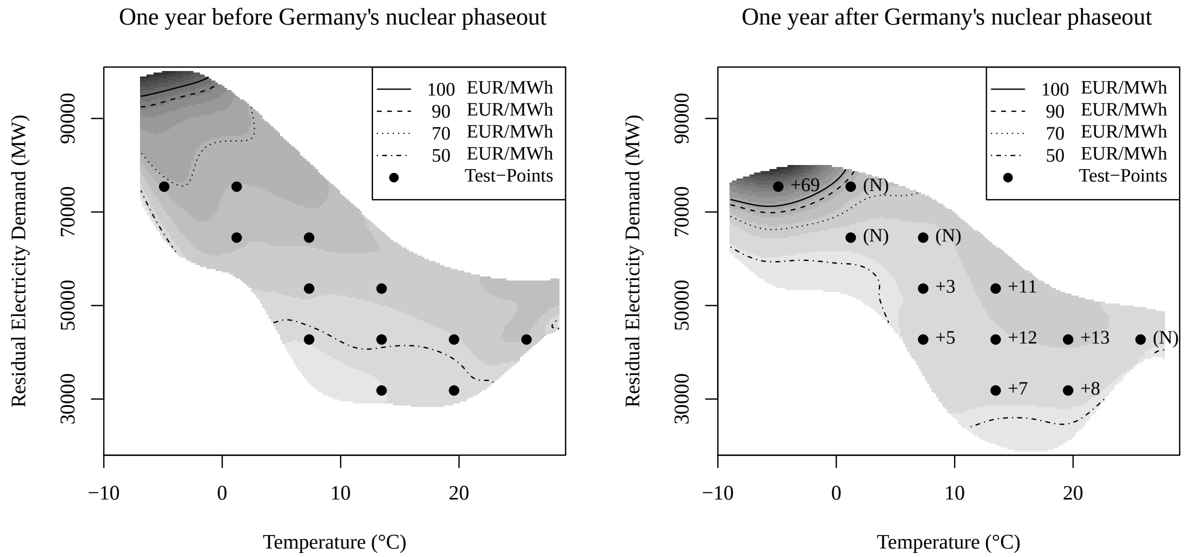

Figure 5: Comparison of the estimated mean functions one year before and after Germany's nuclear phaseout. Insignificant test-results are marked by "(N)". Significant test-results are marked by the numerical differences between the mean functions.

thumb bandwidths in Eq.s (20), (21), (22), and (23), which are very practical here as the usual automatic bandwidth selection procedures such as cross-validation and generalized cross-validation are known to perform badly for correlated data.

Figure 5 shows the mean estimates one year before, say $\hat{\mu}_B(u, z; \hat{h}^D_{\mu_B,U}, \hat{h}^D_{\mu_B,Z})$, and one year after, say $\hat{\mu}_A(u, z; \hat{h}^D_{\mu_A,U}, \hat{h}^D_{\mu_A,Z})$, Germany's (partial) nuclear phaseout. It is striking that the supports of the mean functions are relatively complicated objects. This explains our modeling assumption of z-dependent domains, namely, $X_i(., z) \in L^2[a(z), b(z)]$. The boundary functions $a(z)$ and $b(z)$ of the kidney-shaped supports are estimated using the local linear estimation approach of Martins-Filho and Yao (2007).

In both periods the estimated mean prices range from about 40 to 100 EUR/MWh which is consistent with the ranges of the univariate time series in Figure 4. The *conditional* mean prices for given values of electricity demand and temperature are, however, higher in the year after Germany's (partial) nuclear phaseout. These differences in the conditional means range from about +5 EUR/MWh for moderate factor values, which relate to the relatively flat middle part of the merit order curve, up to +69 EUR/MWh for very high values of (residual) demand at which the merit order curve is very steep.

In order to test the hypothesis of equal means (i.e., H_0: $\mu_B(u,z) = \mu_A(u,z)$) against the alternative of strictly greater mean values in the year after Germany's nuclear phaseout (i.e., H_1: $\mu_B(u,z) < \mu_A(u,z)$), we use our sparse-to-dense test statistic $T^{SD}_{\mu}(u,z)$ applied to regular grid-points that cover the intersection of the supports of both mean functions. This grid $(u_j, z_j)_{j\geq 1}$ of in total $G = 12$ grid-points is shown by the black-filled circle points in Figure 5. The test procedure is conducted using the following Bonferroni-adjusted asymptotic confidence intervals computed separately for each period $P \in \{A, B\}$ and each grid point $j = 1, \dots, G$:

$$\hat{\mu}_P(u_j, z_j; \hat{h}^D_{\mu_P,U}, \hat{h}^D_{\mu_P,Z}) - \hat{B}^{\mu_P}(u_j, z_j; \hat{h}^D_{\mu_P,U}, \hat{h}^D_{\mu_P,Z}) \pm \widehat{se}_{P,j} \cdot z_{\alpha/G} \quad \text{with}$$

$$\widehat{se}_P = \sqrt{\hat{S}^{\mu_P}_1(u_j, z_j; \hat{h}^D_{\mu_P,U}, \hat{h}^D_{\mu_P,Z}, \hat{h}^D_{\gamma_P,U}, \hat{h}^D_{\gamma_P,Z}) + \hat{S}^{\mu_P}_2(u_j, z_j; \hat{h}^D_{\mu_P,Z}, \hat{h}^D_{\gamma_P,U}, \hat{h}^D_{\gamma_P,Z})}$$

where $z_{\alpha/G}$ denotes the $100(1-\alpha/G)$th percentile of the standard normal distribution with Bonferroni-adjusted significance level $\alpha/G = 0.05/12$ in order to account for the multiple testing problem. As we are only concerned with positive shifts of the mean prices, we conduct a one-sided significance test. A shift in the mean at a grid point (u_i, z_j) is significant if the lower one-sided confidence "interval" of $\hat{\mu}_A(u_j, z_j)$ is strictly greater than the upper one-sided confidence "interval" of $\hat{\mu}_B(u_j, z_j)$, i.e., if

$$\left(\hat{\mu}_A(u_j, z_j) - \widehat{se}_{A,j} \cdot z_{\alpha/G}\right) - \left(\hat{\mu}_B(u_j, z_j) + \widehat{se}_{B,j} \cdot z_{\alpha/G}\right) > 0.$$

The chosen grid size of $G = 12$ grid points is found to be appropriate for balancing the variance inflation effect due to the Bonferroni adjustment and the need to capture enough information which allows to compare the mean function by means of point-wise comparisons. The above confidence bands can be seen as rather conservative due the Bonferroni-adjustment which is known to be conservative. The significant mean shifts at the chosen grid points are depicted in the right plot of Figure 5. The numerical values stand for the amount of the price shift measured in EUR/MWh. Insignificant mean shifts are marked by "(N)". Figure 6 in the supplemental paper shows the confidence intervals which serve as basis for our test-decisions.

Figure 5 gives empirical evidence for a further important issue: Germany managed to reduce its residual demand for electricity in the year after the nuclear phaseout, which is reflected by the lower support of the mean function for low temperatures values in the second period. This reduction was mainly due to a politically promoted higher amount of electricity infeeds from RES and obviously helped to avoid the occurrence of some very high electricity prices.

7 Conclusion

The theoretical side of our paper is concerned with the nonparametric estimation of the conditional mean and covariance function of a stationary time series of weakly dependent random functions with covariate-adjustments. Using a double asymptotic we investigate all cases from sparsely sampled to densely sampled functional data which takes into account the vague cases typically found in applications. A specific emphasis lies on the derivation of AMISE optimal bandwidths for the multivariate local linear estimators under this double asymptotic which leads to some non-classical bandwidth expressions. It turns out that these new bandwidth expressions allow the local linear estimators to *inherently* under-smooth the functional data for the sake of a better estimate of the mean or covariance function. The practical side of our paper applies our bias, variance, and bandwidth results in order to test for differences in conditional mean prices before and after Germany's abrupt nuclear phaseout in mid-March 2011.

SUPPLEMENTARY MATERIAL

Supplemental Paper: Contains the proofs or our theoretical results, a description of the data sources, and a plot of the confidence intervals.

R-Codes: Contains all R-codes of our simulation study and our real data application.

Data: Contains a simulated data set which closely resembles the data set used in our real data application.

References

Benko, M., W. Härdle, and A. Kneip (2009). Common functional principal components. *The Annals of Statistics 37*(1), 1–34.

Burger, M., B. Graeber, and G. Schindlmayr (2008). *Managing Energy Risk: An Integrated View on Power and Other Energy Markets* (1. ed.). Wiley.

Burger, M., B. Klar, A. Müller, and G. Schindlmayr (2004). A spot market model for pricing derivatives in electricity markets. *Quantitative Finance 4*(1), 109–122.

Cardot, H. (2007). Conditional functional principal components analysis. *Scandinavian Journal of Statistics 34*(2), 317–335.

Cludius, J., H. Hermann, F. C. Matthes, and V. Graichen (2014). The merit order effect of wind and photovoltaic electricity generation in germany 2008–2016: Estimation and distributional implications. *Energy Economics 44*, 302–313.

Duong, T. et al. (2007). ks: Kernel density estimation and kernel discriminant analysis for multivariate data in r. *Journal of Statistical Software 21*(7), 1–16.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modelling and its Applications* (1. ed.), Volume 66 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC.

Fan, J. and Q. Yao (2003). *Nonlinear Time Series* (1. ed.). Springer Series in Statistics. Springer.

Gasser, T. and H.-G. Müller (1984). Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics*, 171–185.

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

Härdle, W. and A. W. Bowman (1988). Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands. *Journal of the American Statistical Association 83*(401), 102–110.

Herrmann, E., J. Engel, M. Wand, and T. Gasser (1995). A bandwidth selector for bivariate kernel regression. *Journal of the Royal Statistical Society. Series B (Methodological) 57*(1), 171–180.

Hirth, L. (2013). The market value of variable renewables: The effect of solar wind power variability on their relative price. *Energy Economics 38*, 218–236.

Jiang, C.-R. and J.-L. Wang (2010). Covariate adjusted functional principal components analysis for longitudinal data. *The Annals of Statistics 38*(2), 1194–1226.

Lijesen, M. G. (2007). The real-time price elasticity of electricity. *Energy Economics 29*(2), 249–258.

Martins-Filho, C. and F. Yao (2007). Nonparametric frontier estimation via local linear regression. *Journal of Econometrics 141*(1), 283–319.

Nestle, U. (2012). Does the use of nuclear power lead to lower electricity prices? an analysis of the debate in germany with an international perspective. *Energy Policy 41*(0), 152–160.

Ruppert, D. and M. Wand (1994). Multivariate locally weighted least squares regression. *The Annals of Statistics 22*(3), 1346–1370.

Sensfuß, F., M. Ragwitz, and M. Genoese (2008). The merit-order effect: A detailed analysis of the price effect of renewable electricity generation on spot market prices in germany. *Energy policy 36*(8), 3086–3094.

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

Zhang, X., J.-L. Wang, et al. (2016). From sparse to dense functional data and beyond. *The Annals of Statistics 44*(5), 2281–2321.

Supplemental Paper for:

Inference for Covariate-Adjusted Sparse to Dense Functional Data with an Application to Testing Differences in Energy Prices

Dominik Liebl

Institute for Financial Economics and Statistics

University of Bonn

Adenauerallee 24-42

53113 Bonn, Germany

Outline:

Section 1 contains the proofs of the Theorems 3.1 and 3.2. Section 2 contains the proofs of the Theorems 3.3 and 3.4. The data sources are described in detail in Section 4. Finally, Section 4.1 contains the plot of the confidence intervals that have been used for our test-decision in the real data application.

1 Proofs of Theorems 3.1 and 3.2

1.1 Proof of Theorem 3.1

Our proof of Theorem 3.1 generally follows that of Ruppert and Wand (1994)[1], and differs only from the latter reference as we consider additionally, first, a conditioning variable Z_i, second, a functional valued error term, and, third, a time series context.

[1] Ruppert, D. and M. Wand (1994). Multivariate locally weighted least squares regression. The Annals of Statistics 22 (3), 13461370.

Proof of Theorem 3.1, part *(i)*: Let u and z be interior points of $\mathrm{supp}(f_{UZ})$ and define $\mathbf{H}_\mu = \mathrm{diag}(h^2_{\mu,U}, h^2_{\mu,Z})$, $\mathbf{U} = (U_{11}, \ldots, U_{nm})^\top$, and $\mathbf{Z} = (Z_1, \ldots, Z_n)^\top$. Using a Taylor-expansion of μ around (u,z), the conditional bias of the estimator $\hat{\mu}(u,z;\mathbf{H})$ can be written as

$$\mathbb{E}(\hat{\mu}(u,z;\mathbf{H}_\mu) - \mu(u,z)|\mathbf{U},\mathbf{Z}) = \frac{1}{2} e_1^\top \left((nm)^{-1}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]\right)^{-1} \times \\ \times (nm)^{-1}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz}\left(\boldsymbol{\mathcal{Q}}_\mu(u,z) + \mathbf{R}_\mu(u,z)\right), \tag{30}$$

where $\boldsymbol{\mathcal{Q}}_\mu(u,z)$ is a $nm \times 1$ vector with typical elements

$$(U_{ij} - u, Z_i - z)\boldsymbol{\mathcal{H}}_\mu(u,z)(U_{ij} - u, Z_i - z)^\top \in \mathbb{R}$$

with $\boldsymbol{\mathcal{H}}_\mu(u,z)$ being the Hessian matrix of the regression function $\mu(u,z)$. The $nm \times 1$ vector $\mathbf{R}_\mu(u,z)$ holds the remainder terms as in Ruppert and Wand (1994).

Next we derive asymptotic approximations for the 3×3 matrix $\left((nm)^{-1}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]\right)^{-1}$ and the $3{\times}1$ matrix $(nm)^{-1}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz}\boldsymbol{\mathcal{Q}}_\mu(u,z)$ of the right hand side of Eq. (30). Using standard procedures from kernel density estimation it is easy to derive that

$$(nm)^{-1}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z] = \\ \begin{pmatrix} f_{UZ}(u,z) + o_p(1) & \nu_2(K_\mu)\mathbf{D}_{f_{UZ}}(u,z)^\top \mathbf{H}_\mu + o_p(\mathbf{1}^\top \mathbf{H}_\mu) \\ \nu_2(K_\mu)\mathbf{H}_\mu \mathbf{D}_{f_{UZ}}(u,z) + o_p(\mathbf{H}_\mu \mathbf{1}) & \nu_2(K_\mu)\mathbf{H}_\mu f_{UZ}(u,z) + o_p(\mathbf{H}_\mu) \end{pmatrix},$$

where $\mathbf{1} = (1,1)^\top$ and $\mathbf{D}_{f_{UZ}}(u,z)$ is the vector of first order partial derivatives (i.e., the gradient) of the pdf f_{UZ} at (u,z). Inversion of the above block matrix yields

$$\left((nm)^{-1}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]\right)^{-1} = \tag{31}$$

$$\begin{pmatrix} (f_{UZ}(u,z))^{-1} + o_p(1) & -\mathbf{D}_{f_{UZ}}(u,z)^\top (f_{UZ}(u,z))^{-2} + o_p(\mathbf{1}^\top) \\ -\mathbf{D}_{f_{UZ}}(u,z)(f_{UZ}(u,z))^{-2} + o_p(\mathbf{1}) & (\nu_2(K_\mu)\mathbf{H}_\mu f_{UZ}(u,z))^{-1} + o_p(\mathbf{H}_\mu) \end{pmatrix}.$$

The 3×1 matrix $(nm)^{-1}[\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz} \boldsymbol{\mathcal{Q}}_\mu(u,z)$ can be partitioned as following:

$$(nm)^{-1}[\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz} \boldsymbol{\mathcal{Q}}_\mu(u,z) \quad = \quad \begin{pmatrix} \texttt{upper element} \\ \\ \texttt{lower bloc} \end{pmatrix},$$

where the 1×1 dimensional `upper element` can be approximated by

$$\begin{aligned} &(nm)^{-1} \sum_{it} K_{\mu,h}(U_{ij} - u, Z_i - z)(U_{ij} - u, Z_i - z) \boldsymbol{\mathcal{H}}_\mu(u,z)(U_{ij} - u, Z_i - z)^\top \\ &= (\nu_2(\kappa))^2 \, tr\left\{\mathbf{H}_\mu \boldsymbol{\mathcal{H}}_\mu(u,z)\right\} f_{UZ}(u,z) + o_p(tr(\mathbf{H}_\mu)) \end{aligned} \tag{32}$$

and the 2×1 dimensional `lower bloc` is equal to

$$\begin{aligned} &(nm)^{-1} \sum_{it} \left\{K_{\mu,h}(U_{ij} - u, Z_i - z)(U_{ij} - u, Z_i - z) \boldsymbol{\mathcal{H}}_\mu(u,z)(U_{ij} - u, Z_i - z)^\top\right\} \times \\ &\times (U_{ij} - u, Z_i - z)^\top = O_p(\mathbf{H}_\mu^{3/2} \mathbf{1}). \end{aligned} \tag{33}$$

Plugging the approximations of Eqs. (31)-(33) into the first summand of the conditional bias expression in Eq. (30) leads to the following expression

$$\begin{aligned} &\frac{1}{2} e_1^\top ((nm)^{-1}[\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz}[\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z])^{-1} (nm)^{-1}[\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz} \boldsymbol{\mathcal{Q}}_\mu(u,z) = \\ &= \frac{1}{2} (\nu_2(\kappa))^2 \, tr\left\{\mathbf{H}_\mu \boldsymbol{\mathcal{H}}_\mu(u,z)\right\} + o_p(tr(\mathbf{H}_\mu)). \end{aligned}$$

Furthermore, it is easily seen that the second summand of the conditional bias expression in Eq. (30), which holds the remainder term, is given by

$$\frac{1}{2} e_1^\top ((nm)^{-1}[\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz}[\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z])^{-1} (nm)^{-1}[\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz} \mathbf{R}_\mu(u,z) = o_p(tr(\mathbf{H}_\mu)).$$

Summation of the two latter expressions yields the asymptotic approximation of the conditional bias

$$\mathbb{E}(\hat{\mu}(u,z;\mathbf{H}_\mu) - \mu(u,z) | \mathbf{U}, \mathbf{Z}) = \frac{1}{2} (\nu_2(\kappa))^2 \, tr\left\{\mathbf{H}_\mu \boldsymbol{\mathcal{H}}_\mu(u,z)\right\} + o_p(tr(\mathbf{H}_\mu)).$$

This is our bias statement of Theorem 3.1 part *(i)*.

Proof of Theorem 3.1, part *(ii)*: In the following we derive the conditional variance of the local linear estimator $\mathbb{V}(\hat{\mu}(u,z;\mathbf{H}_\mu)|\mathbf{U},\mathbf{Z}) =$

$$\begin{aligned}
=&e_1^\top([\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top\mathbf{W}_{\mu,uz}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z])^{-1}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top\mathbf{W}_{\mu,uz}\operatorname{Cov}(\mathbf{Y}|\mathbf{U},\mathbf{Z})\,\mathbf{W}_{\mu,uz}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]\times\\
&\times([\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top\mathbf{W}_{\mu,uz}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z])^{-1}u_1
\end{aligned}\tag{34}$$

$$\begin{aligned}
=&e_1^\top((nm)^{-1}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top\mathbf{W}_{\mu,uz}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z])^{-1}\times\\
&\times((nm)^{-2}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top\mathbf{W}_{\mu,uz}\operatorname{Cov}(\mathbf{Y}|\mathbf{U},\mathbf{Z})\,\mathbf{W}_{\mu,uz}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z])\times\\
&\times((nm)^{-1}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top\mathbf{W}_{\mu,uz}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z])^{-1}u_1,
\end{aligned}$$

where $\operatorname{Cov}(\mathbf{Y}|\mathbf{U},\mathbf{Z})$ is the $nm\times nm$ matrix with typical elements

$$\operatorname{Cov}(Y_{ij},Y_{\ell k}|U_{ij},U_{\ell k},Z_i,Z_\ell)=\gamma_{|i-\ell|}((U_{ij},Z_i),(U_{\ell k},Z_\ell))+\sigma_\epsilon^2\mathbb{1}(i=\ell\text{ and }j=k)$$

with $\mathbb{1}(.)$ being the indicator function.

We begin with analyzing the 3×3 matrix

$$(nm)^{-2}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top\mathbf{W}_{\mu,uz}\operatorname{Cov}(\mathbf{Y}|\mathbf{U},\mathbf{Z})\,\mathbf{W}_{\mu,uz}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]$$

using the following three Lemmas 1.1-1.3.

Lemma 1.1 *The upper-left scalar (block) of the matrix* $(nm)^{-2}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top\mathbf{W}_{\mu,uz}\operatorname{Cov}(\mathbf{Y}|\mathbf{U},\mathbf{Z})\mathbf{W}_{\mu,uz}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]$ *is given by*

$$\begin{aligned}
&(nm)^{-2}\mathbf{1}^\top\mathbf{W}_{\mu,uz}\operatorname{Cov}(\mathbf{Y}|\mathbf{U},\mathbf{Z})\mathbf{W}_{\mu,uz}\mathbf{1}\\
=\;&(nm)^{-1}f_{UZ}(u,z)|\mathbf{H}_\mu|^{-1/2}R(K_\mu)\left(\gamma(u,u,z)+\sigma_\epsilon^2\right)(1+O_p(tr(\mathbf{H}_\mu^{1/2})))\\
+\;&n^{-1}(f_{UZ}(u,z))^2\left[\left(\frac{m-1}{m}\right)h_{\mu,Z}^{-1}R(\kappa)\frac{\gamma(u,u,z)}{f_Z(z)}+c(u,z)\right](1+O_p(tr(H^{1/2})))\\
=\;&O_p((nm)^{-1}|\mathbf{H}_\mu|^{-1/2})+O_p(n^{-1}h_{\mu,Z}^{-1}).
\end{aligned}$$

where $c(u,z)=2\sum_{l=1}^{n-1}\gamma_l((u,z),(u,z))$ *is the sum of the autocovariance functions. The assumed geometric decay rate of the autocovariances (Assumptions A-TS) implies that there exists a constant* C*,* $0<C<\infty$*, such that* $0\le|c(u,z)|\le C$*.*

Lemma 1.2 *The 1×2 dimensional upper-right block of the matrix* $(nm)^{-2}[\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz} \mathrm{Cov}(\mathbf{Y}|\mathbf{U}, \mathbf{Z}) \mathbf{W}_{\mu,uz} [\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]$ *is given by*

$$\begin{aligned}
& (nm)^{-2} \mathbf{1}^\top \mathbf{W}_{\mu,uz} \mathrm{Cov}(\mathbf{Y}|\mathbf{U}, \mathbf{Z}) \mathbf{W}_{\mu,uz} \begin{pmatrix} (U_{11} - u, Z_1 - z) \\ \vdots \\ (U_{nm} - u, Z_n - z) \end{pmatrix} \\
= \;& (nm)^{-1} f_{UZ}(u,z) |\mathbf{H}_\mu|^{-1/2} (\mathbf{1}^\top \mathbf{H}_\mu^{1/2}) R(K_\mu) \left(\gamma(u,u,z) + \sigma_\epsilon^2\right) (1 + O_p(tr(\mathbf{H}_\mu^{1/2}))) \\
+ \;& n^{-1} (f_{UZ}(u,z))^2 (\mathbf{1}^\top \mathbf{H}_\mu^{1/2}) \left[\left(\frac{m-1}{m} \right) h_{\mu,Z}^{-1} R(\kappa) \frac{\gamma(u,u,z)}{f_Z(z)} + c_r \right] (1 + O_p(tr(\mathbf{H}_\mu^{1/2}))) \\
= \;& O_p((nm)^{-1} |\mathbf{H}_\mu|^{-1/2} (\mathbf{1}^\top \mathbf{H}_\mu^{1/2})) + O_p(n^{-1} (\mathbf{1}^\top \mathbf{H}_\mu^{1/2}) h_{\mu,Z}^{-1}),
\end{aligned}$$

where $c(u,z) = 2 \sum_{l=1}^{n-1} \gamma_l((u,z),(u,z))$ is the sum of the autocovariance functions. The assumed geometric decay rate of the autocovariances (Assumptions A-TS) implies that there exists a constant C, $0 < C < \infty$, such that $0 \le |c(u,z)| \le C$.

The 2×1 dimensional lower-left block of the matrix $(nm)^{-2}[\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz} \mathrm{Cov}(\mathbf{Y}|\mathbf{U}, \mathbf{Z}) \mathbf{W}_{\mu,uz} [\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]$ *is simply the transposed version of this result.*

Lemma 1.3 *The 2×2 lower-right block of the matrix* $(nm)^{-2}[\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz} \mathrm{Cov}(\mathbf{Y}|\mathbf{U}, \mathbf{Z}) \mathbf{W}_{\mu,uz} [\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]$ *is given by*

$$\begin{aligned}
& (nm)^{-2} \left(((U_{11} - u), (Z_1 - z))^\top, \ldots, ((U_{nm} - u), (Z_n - z)^\top) \right) \times \\
& \times \mathbf{W}_{\mu,uz} \mathrm{Cov}(\mathbf{Y}|\mathbf{U}, \mathbf{Z}) \mathbf{W}_{\mu,uz} \begin{pmatrix} (U_{11} - u, Z_1 - z) \\ \vdots \\ (U_{nm} - u, Z_n - z) \end{pmatrix} \\
= \;& (nm)^{-1} f_{UZ}(u,z) |\mathbf{H}_\mu|^{-1/2} \mathbf{H}_\mu R(K_\mu) \left(\gamma(u,u,z) + \sigma_\epsilon^2\right) (1 + O_p(tr(\mathbf{H}_\mu^{1/2}))) \\
+ \;& n^{-1} (f_{UZ}(u,z))^2 \mathbf{H}_\mu \left[\left(\frac{m-1}{m} \right) h_{\mu,Z}^{-1} R(\kappa) \frac{\gamma(u,u,z)}{f_Z(z)} + c_r \right] (1 + O_p(tr(\mathbf{H}_\mu^{1/2}))) \\
= \;& O_p((nm)^{-1} |\mathbf{H}_\mu|^{-1/2} \mathbf{H}_\mu) + O_p(n^{-1} \mathbf{H}_\mu h_{\mu,Z}^{-1})
\end{aligned}$$

where $c(u,z) = 2 \sum_{l=1}^{n-1} \gamma_l((u,z),(u,z))$ is the sum of the autocovariance functions. The *assumed geometric decay rate of the autocovariances (Assumptions A-TS) implies that there exists a constant C, $0 < C < \infty$, such that $0 \leq |c(u,z)| \leq C$.*

Using the approximations for the bloc-elements of the matrix $(nm)^{-2}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz}\mathrm{Cov}(\mathbf{Y}|\mathbf{U},\mathbf{Z})\mathbf{W}_{\mu,uz}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]$, given by the Lemmas 1.1-1.3, and the approximation for the matrix $\left((nm)^{-1}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]\right)^{-1}$, given in (31), we can approximate the conditional variance of the bivariate local linear estimator, given in (34). Some tedious yet straightforward matrix algebra leads to $\mathbb{V}(\hat{\mu}(u,z;\mathbf{H}_\mu)|\mathbf{U},\mathbf{Z}) =$

$$(nm)^{-1}|\mathbf{H}_\mu|^{-1/2}\left\{\frac{R(K_\mu)\left(\gamma(u,u,z)+\sigma_\epsilon^2\right)}{f_{UZ}(u,z)}\right\}(1+o_p(1))$$
$$+\,n^{-1}\left[\left(\frac{m-1}{m}\right)h_{\mu,Z}^{-1}R(\kappa)\frac{\gamma(u,u,z)}{f_Z(z)}+c_r\right](1+o_p(1))\,,$$

which is asymptotically equivalent to our variance statement of Theorem 3.1 part *(ii)*.

Next we proof Lemma 1.1; the proofs of Lemmas 1.2 and 1.3 can be done correspondingly. To show Lemma 1.1 it will be convenient to split the sum such that $(nm)^{-2}\mathbf{1}^\top\mathbf{W}_{\mu,uz}\mathrm{Cov}(\mathbf{Y}|\mathbf{U},\mathbf{Z})\mathbf{W}_{\mu,uz}\mathbf{1} = s_1+s_2+s_3$. Using standard procedures from kernel density estimation leads to

$$\begin{aligned}
s_1 &= (nm)^{-2}\sum_{it}(K_{\mu,h}(U_{ij}-u,Z_i-z))^2\,\mathbb{V}(Y_{ij}|\mathbf{U},\mathbf{Z}) && (35)\\
&= (nm)^{-1}|\mathbf{H}_\mu|^{-1/2}f_{UZ}(u,z)R(K_\mu)\left(\gamma(u,u,z)+\sigma_\epsilon^2\right)+O((nm)^{-1}|\mathbf{H}_\mu|^{-1/2}\,tr(\mathbf{H}_\mu^{1/2})),\\
s_2 &= (nm)^{-2}\sum_{jk}\sum_{\substack{i\ell\\ i\neq\ell}}K_{\mu,h}(U_{ij}-u,Z_i-z)\,\mathrm{Cov}(Y_{ij},Y_{\ell k}|\mathbf{U},\mathbf{Z})\,K_{\mu,h}(U_{\ell k}-x,Z_\ell-z) && (36)\\
&= n^{-1}(f_{UZ}(u,z))^2c(u,z)+O_p(n^{-1}tr(\mathbf{H}_\mu^{1/2}))\\
s_3 &= (nm)^{-2}\sum_{\substack{ij\\ i\neq j}}\sum_{t}h_{\mu,U}^{-1}\kappa(h_{\mu,U}^{-1}(U_{ij}-u))(h_{\mu,Z}^{-1}\kappa(h_{\mu,Z}^{-1}(Z_i-z)))^2\,\mathrm{Cov}(Y_{ij},Y_{jt}|\mathbf{U},\mathbf{Z})\times && (37)\\
&\quad\times h_{\mu,U}^{-1}\kappa(h_{\mu,U}^{-1}(U_{ik}-x))\\
&= n^{-1}(f_{UZ}(u,z))^2\left[\left(\frac{m-1}{m}\right)h_{\mu,Z}^{-1}R(\kappa)\frac{\gamma(u,u,z)}{f_Z(z)}\right]+O_p(n^{-1}tr(\mathbf{H}_\mu^{1/2})),
\end{aligned}$$

where $c(u,z) = 2\sum_{l=1}^{n-1}\gamma_l((u,z),(u,z))$ is the sum of the autocovariance functions. Summing up (35)-(36) leads to the result in Lemma 1.1. Lemmas 1.2 and 1.3 differ from Lemma

1.1 only with respect to the additional factors $\mathbf{1}^\top \mathbf{H}_\mu^{1/2}$ and $\mathbf{H}_\mu$. These come in due to the usual substitution step for the additional data parts $(U_{ij} - u, Z_i - z)$.

1.2 Proof of Theorem 3.2

When using the (unknown) theoretical raw-covariances C_{ijk} as defined in Eq. (4), the proof of Theorem 3.2 follows the same arguments as in the proof of Theorem 3.1. Therefore, this proof is skipped. The only additional issue we need to consider, is that considering the empirical raw-covariances $\hat{C}_{ijk}$, as defined in Eq. (7), will not result in different rates.

Observe that we can expand $\hat{C}_{ijk}$ as following:

$$
\begin{aligned}
\hat{C}_{ijk} = C_{ijk} &+ (Y_{ij} - \mu(U_{ij}, Z_i))(\mu(U_{ik}, Z_i) - \hat{\mu}(U_{ik}, Z_i)) \\
&+ (Y_{jt} - \mu(U_{ik}, Z_i))(\mu(U_{ij}, Z_i) - \hat{\mu}(U_{ij}, Z_i)) \\
&+ (\mu(U_{ij}, Z_i) - \hat{\mu}(U_{ij}, Z_i))(\mu(U_{ik}, Z_i) - \hat{\mu}(U_{ik}, Z_i)).
\end{aligned}
$$

Focusing for simplicity on AMISE optimal bandwidth choices, it follows from Corollaries 3.1 and 3.3 that

$$
\left(\hat{\mu}(U_{ij}, Z_i; h_{\mu,U,\mathrm{opt}},\, h_{\mu,Z,\mathrm{opt}}) - \mu(U_{ij}, Z_i)\right)\big|\mathbf{U}, \mathbf{Z} = \begin{cases} O_p\left((nm)^{-2/3}\right) & \text{if } 0 \leq \theta \leq 1/5 \\ O_p\left(n^{-4/5}\right) & \text{if } \theta > 1/5, \end{cases}
$$

for all i and j, where $h_{\mu,U,\mathrm{opt}}$ and $h_{\mu,Z,\mathrm{opt}}$ denote the θ-specific AMISE optimal bandwidth choices as defined in Eqs. (8), (9), (12), and (13). Therefore, under our setup we have that

$$
\left(\hat{C}_{ijk} - C_{ijk}\right)\big|\mathbf{U}, \mathbf{Z} = \begin{cases} O_p\left((nm)^{-2/3}\right) & \text{if } 0 \leq \theta \leq 1/5 \\ O_p\left(n^{-4/5}\right) & \text{if } \theta > 1/5, \end{cases}
$$

which is of an order of magnitude smaller than the approximation errors of $\hat{\gamma}$ derived under the usage of the theoretical raw covariates C_{ijk}, namely, $\mathcal{O}_p((nM)^{-2/7})$ for $0 \leq \theta \leq 1/5$ and $\mathcal{O}_p(n^{-2/5})$ for $\theta > 1/5$; see also Corollaries 3.2 and 3.4 .

2 Proofs of Theorems 3.3 and 3.4 and Corollaries 3.1 and 3.2

2.1 Proof of Theorem 3.3

The AMISE function (i.e., the AMISE function with leading S_1^μ variance term) for the local linear estimator $\hat{\mu}$ is given by

$$\begin{aligned}\text{AMISE}_{\hat{\mu}}\,(h_{\mu,U}, h_{\mu,Z}) &= (nm)^{-1}\, h_{\mu,U}^{-1}\, h_{\mu,Z}^{-1}\, R(K_\mu)\, Q_{\mu,1}+ \\ &+ \frac{1}{4}\,(\nu_2(K_\mu))^2\,\left[h_{\mu,U}^4\,\mathcal{I}_{\mu,UU} + 2\,h_{\mu,U}^2\,h_{\mu,Z}^2\,\mathcal{I}_{\mu,UZ} + h_{\mu,Z}^4\,\mathcal{I}_{\mu,ZZ}\right],\end{aligned} \tag{38}$$

$$\begin{aligned}\text{where}\qquad Q_{\mu,1} &= \int\left(\gamma(u,u,z)+\sigma_\epsilon^2\right)\,d(u,z),\\ \mathcal{I}_{\mu,UU} &= \int\left(\mu^{(2,0)}(u,z)\right)^2\, f_{UZ}(u,z)\,d(u,z),\\ \mathcal{I}_{\mu,ZZ} &= \int\left(\mu^{(0,2)}(u,z)\right)^2\, f_{UZ}(u,z)\,d(u,z),\quad\text{and}\\ \mathcal{I}_{\mu,UZ} &= \int \mu^{(2,0)}(u,z)\mu^{(0,2)}(u,z)\, f_{UZ}(u,z)\,d(u,z).\end{aligned}$$

This is a known expression for the AMISE function of a two-dimensional local linear estimator with a diagonal bandwidth matrix (see, e.g., Herrmann et al. (1995)[2]) and follows from the formulas in Wand and Jones (1994)[3]. Minimizing the above AMISE function with respect to $h_{\mu,U}$ and $h_{\mu,Z}$ leads to the optimal bandwidth expressions in Theorem 3.3 which correspond to the results in Herrmann et al. (1995)[4].

It follows directly from the Theorem 3.1 that the first variance summand S_1^μ is the leading variance term if the following order relation holds:

$$n^{-1}\,h_{\mu,Z}^{-1} = o\left(n^{-(1+\theta)}\,h_{\mu,U}^{-1}\,h_{\mu,Z}^{-1}\right), \tag{39}$$

[2]Herrmann, E., J. Engel, M. Wand, and T. Gasser (1995). A bandwidth selector for bi-variate kernel regression. Journal of the Royal Statistical Society. Series B (Methodological) 57 (1), 171–180.

[3]Wand, M. and M. Jones (1994). Multivariate plug-in bandwidth selection. Computational Statistics 9 (2), 97–116.

[4]Herrmann, E., J. Engel, M. Wand, and T. Gasser (1995). A bandwidth selector for bi- variate kernel regression. Journal of the Royal Statistical Society. Series B (Methodological) 57 (1), 171180.

where we used that by Assumption A-AS $nm \sim n^{1+\theta}$. Plugging the AMISE optimal bandwidth rates of Theorem 3.3 into the order relation of Eq. (39) leads to the corresponding θ values of $0 \leq \theta < 1/5$ which describe the here considered case of "very sparse to sparse" functional data.

2.2 Proof of Theorem 3.4

The corresponding AMISE function (i.e., the AMISE function with leading S_1^{γ} variance term) for the local linear estimator $\hat{\gamma}$ is given by

$$\begin{aligned} \text{AMISE}_{\hat{\gamma}}\,(h_{\gamma,U}, h_{\gamma,Z}) = (nM)^{-1}\, h_{\gamma,U}^{-2}\, h_{\gamma,Z}^{-1}\, R(K_{\gamma})\, Q_{\gamma,1} + \\ + \frac{1}{4}\,(\nu_2(K_{\gamma}))^2\, \left[2\, h_{\gamma,U}^4\, (\mathcal{I}_{\gamma,U_{(1)}U_{(2)}} + \mathcal{I}_{\gamma,U_{(1)}U_{(2)}}) + 4\, h_{\gamma,U}^2\, h_{\gamma,Z}^2\, \mathcal{I}_{\gamma,U_{(1)}Z} + h_{\gamma,Z}^4\, \mathcal{I}_{\gamma,ZZ}\right], \end{aligned} \tag{40}$$

$$\begin{aligned} \text{where} \quad Q_{\gamma,1} &= \int (\tilde{\gamma}((u_1,u_2),(u_1,u_2),z) + \sigma_{\varepsilon}^2(u_1,u_2,z))\; d(u_1,u_2,z) \\ \mathcal{I}_{\gamma,U_{(1)}U_{(1)}} &= \int \left(\gamma^{(2,0,0)}(u_1,u_2,z)\right)^2\, f_{UUZ}(u_1,u_2,z)\, d(u_1,u_2,z), \\ \mathcal{I}_{\gamma,U_{(1)}U_{(2)}} &= \int \left(\gamma^{(2,0,0)}(u_1,u_2,z)\gamma^{(0,2,0)}(u_1,u_2,z)\right)\, f_{UUZ}(u_1,u_2,z)\, d(u_1,u_2,z), \\ \mathcal{I}_{\gamma,U_{(1)}Z} &= \int \gamma^{(2,0,0)}(u_1,u_2,z)\gamma^{(0,0,2)}(u_1,u_2,z)\, f_{UUZ}(u_1,u_2,z)\, d(u_1,u_2,z), \quad \text{and} \\ \mathcal{I}_{\gamma,ZZ} &= \int \left(\gamma^{(0,0,2)}(u_1,u_2,z)\right)^2\, f_{UUZ}(u_1,u_2,z)\, d(u_1,u_2,z). \end{aligned}$$

Equation (40) again follows from the formulas in Wand and Jones (1994)[5] and additionally by using the following equalities:

$\mathcal{I}_{\gamma,U_{(1)}U_{(1)}} = \mathcal{I}_{\gamma,U_{(2)}U_{(2)}}$, $\mathcal{I}_{\gamma,U_{(1)}U_{(2)}} = \mathcal{I}_{\gamma,U_{(2)}U_{(1)}}$, and $\mathcal{I}_{\gamma,U_{(1)}Z} = \mathcal{I}_{\gamma,U_{(2)}Z}$ due to the symmetry of the covariance function, where the expressions $\mathcal{I}_{\gamma,U_{(2)}U_{(2)}}$, $\mathcal{I}_{\gamma,U_{(2)}U_{(1)}}$, and $\mathcal{I}_{\gamma,U_{(2)}Z}$ are defined equivalently to their above defined counterparts.

Minimizing the above AMISE function with respect to $h_{\gamma,U}$ and $h_{\gamma,Z}$ leads to the optimal bandwidth expressions in Theorem 3.4. This is much more cumbersome than for the case of the mean function μ, but can easily done using, e.g., a computer algebra system.

[5]Wand, M. and M. Jones (1994). Multivariate plug-in bandwidth selection. Computational Statistics 9 (2), 97116.

It follows directly from the Theorem 3.2 that the first variance summand S_1^γ is the leading variance terms if the following order relation holds:

$$n^{-1}\, h_{\gamma,Z}^{-1} = o\left(n^{-(1+2\theta)}\, h_{\gamma,U}^{-2}\, h_{\gamma,Z}^{-1}\right), \tag{41}$$

where we used that by Assumption A-AS $nM \sim n^{1+2\theta}$. Plugging the AMISE optimal bandwidth rates of Theorem 3.4 into the order relation of Eq. (41) leads to the corresponding θ values of $0 \leq \theta < 1/5$ which describe the here considered case of "very sparse to sparse" functional data. Observe that the same θ-threshold value of 1/5 applies to both estimators $\hat{\mu}$ and $\hat{\gamma}$.

2.3 Proofs of Corollaries 3.1 and 3.2

Corollaries 3.1 and 3.2 follow directly from Theorems 3.1 and 3.2 and from applying a central limit theorem for strictly stationary ergodic times series such as Theorem 9.5.5 of Karlin and Taylor (1975)[6]. Efficiency as well as the convergence rate are direct consequences from using the optimal bandwidths $h^S_{\mu,U}$, $h^S_{\mu,Z}$, $h^S_{\gamma,U}$, and $h^S_{\gamma,Z}$.

[6] Karlin, S., & Taylor, H. M. (1975). A first course in stochastic processes (2d ed). New York: Academic.

3 Proofs of Theorems 3.5 and 3.6 and Corollaries 3.3 and 3.4

3.1 Proof of Theorem 3.5

The AMISE function of $\hat{\mu}$ including both variance terms S_1^μ and S_2^μ is given by

$$\mathrm{AMISE}_{\hat{\mu}}\left(h_{\mu,U}, h_{\mu,Z}\right) = \underbrace{\overbrace{(nm)^{-1}\, h_{\mu,U}^{-1}\, h_{\mu,Z}^{-1}\, R(K_\mu)\, Q_{\mu,1}}^{\int S_1^\mu(u,z) f_{UZ}(u,z) d(u,z)}}_{\text{2nd Order}} + \underbrace{\overbrace{n^{-1}\, h_{\mu,Z}^{-1}\, R(\kappa)\, Q_{\mu,2}}^{\int S_2^\mu(u,z) f_{UZ}(u,z) d(u,z)}}_{\text{1st Order}} + \tag{42}$$

$$+ \frac{1}{4}\,(\nu_2(K_\mu))^2 \left[\underbrace{h_{\mu,U}^4\, \mathcal{I}_{\mu,UU}}_{\text{3rd Order}} + \underbrace{2\, h_{\mu,U}^2\, h_{\mu,Z}^2\, \mathcal{I}_{\mu,UZ}}_{\text{2nd Order}} + \underbrace{h_{\mu,Z}^4\, \mathcal{I}_{\mu,ZZ}}_{\text{1st Order}}\right],$$

$$\begin{aligned}
\text{where}\quad \mathcal{I}_{\mu,UU} &= \int (\mu^{(2,0)}(u,z))^2\, f_{UZ}(u,z)\, d(u,z),\\
\mathcal{I}_{\mu,ZZ} &= \int (\mu^{(0,2)}(u,z))^2\, f_{UZ}(u,z)\, d(u,z),\\
\mathcal{I}_{\mu,UZ} &= \int \mu^{(2,0)}(u,z)\mu^{(0,2)}(u,z)\, f_{UZ}(u,z)\, d(u,z),\\
Q_{\mu,1} &= \int \left(\gamma(u,u,z) + \sigma_\epsilon^2\right)\, d(u,z),\quad \text{and}\\
Q_{\mu,2} &= \int \gamma(u,u,z)\, f_U(u)\, d(u,z).
\end{aligned}$$

Note that it is impossible to derive explicit AMISE optimal U- and Z-bandwidth expressions through minimizing Eq. (42) simultaneously for both bandwidths. Since the second variance term S_2^μ is assumed to be the leading variance term, the lowest possible AMISE value can be achieved if there exists a U-bandwidth which, first, allows us to profit from the (partial) annulment of the U-related bias-variance trade-off, but, second, assures that the second variance term S_2^γ remains the leading variance term.

The first requirement is achieved if the U-bandwidth is of a smaller order of magnitude than the Z-bandwidth, i.e., if $h_{\mu,U} = o(h_{\mu,Z})$. This restriction makes those bias components that depend on $h_{\mu,U}$ asymptotically negligible, since it implies that $h_{\mu,U}^2\, h_{\mu,Z}^2 = o(h_{\mu,Z}^4)$ and therefore that $h_{\mu,U}^4 = o(h_{\mu,U}^2\, h_{\mu,Z}^2)$. The latter two strict inequalities lead to the order relations between the three bias terms as indicated in Eq. (42). The second requirement is achieved if the U-bandwidth does not converge to zero too fast, namely if $mh_{\mu,U} \rightarrow \infty$, which implies the order relation between the two variance terms as indicated in Eq. (42).

Let us initially assume that is possible to find an U-bandwidth that fulfills both the above requirements, namely $h_{\mu,U} = o(h_{\mu,Z})$ and $nh_{\mu,U} \rightarrow \infty$. With such an U-bandwidth we can make use of the order relations indicated in Eq. (42). That is, instead of directly minimizing the AMISE function in Eq. (42) both bandwidths, we can minimize the following simpler, but asymptotically equivalent AMISE function, which depends only on the Z-bandwidth:

$$\mathrm{AMISE}_{\hat{\mu}}^{\text{1st Order}}\,(h_{\mu,Z}) = n^{-1}\, h_{\mu,Z}^{-1}\, R(\kappa)\, Q_{\mu,2} + \frac{1}{4}\,(\nu_2(K_\mu))^2\; h_{\mu,Z}^4\, \mathcal{I}_{\mu,ZZ}.$$

The above equation is minimized by the following Z-bandwidth

$$h_{\mu,Z}^D = \left(\frac{R(\kappa)\, Q_{\mu,2}}{n\,\left(\nu_2(K_\mu)\right)^2\, \mathcal{I}_{\mu,ZZ}}\right)^{1/5},$$

which is that of Eq. (12) in Theorem 3.5.

Though, we still need to find U-bandwidth that fulfills the postulated requirements. To do so we suggest plugging-in the above optimal Z-bandwidth into the AMISE function in Eq. (42) and minimize the (then classical) bias-variance trade-off between the asymptotic second order terms, which leads to the following expression for the U-bandwidths:

$$h_{\mu,U}^D = \left(\frac{R(K_\mu)\, Q_{\mu,1}}{nm\,\left(\nu_2(K_\mu)\right)^2\, \mathcal{I}_{\mu,UZ}}\right)^{1/3} \left(h_{\mu,Z}^D\right)^{-1},$$

which is that of Eq. (13) in Theorem 3.5.

In order to check whether this U-bandwidth actually fulfills the two necessary requirements, we apply some rearrangements. Using that by Assumption AS $m \sim n^\theta$, leads to the following more transparent presentation of the bandwidth rates:

$$h_{\mu,Z}^D \sim m^{-1/(5\,\theta)} \quad \text{and} \quad h_{\mu,U}^D \sim m^{-\eta_\mu(\theta)} \quad \text{with} \quad \eta_\mu(\theta) = \frac{1}{3} + \frac{2}{15\,\theta} \tag{43}$$

With Eq. (43) it is now easily verified that the necessary requirements, i.e., that $h_{\mu,U,\text{AMISE}} = o(h_{\mu,Z}^D)$ and $mh_{\mu,U}^D \rightarrow \infty$, are fulfilled iff $\theta > 1/5$.

3.2 Proof of Theorem 3.6

The AMISE expression of $\hat{\gamma}$ including both variance terms S_1^γ and S_2^γ is given by

$$\begin{aligned}\text{AMISE}_{\hat{\gamma}}\left(h_{\gamma,U}, h_{\gamma,Z}\right) = &\overbrace{(nM)^{-1}\, h_{\gamma,U}^{-2}\, h_{\gamma,Z}^{-1}\, R(K_\gamma)\, Q_{\gamma,1}}^{\text{2nd Order}} + \overbrace{n^{-1}\, h_{\gamma,Z}^{-1}\, R(\kappa)\, Q_{\gamma,2}}^{\text{1st Order}} + \\ &+\frac{1}{4}\left(\nu_2(K_\gamma)\right)^2 \left[\underbrace{2\, h_{\gamma,U}^4 \left(\mathcal{I}_{\gamma,U_{(1)}U_{(1)}} + \mathcal{I}_{U_{(1)}U_{(2)}}\right)}_{\text{3rd Order}} + \underbrace{4\, h_{\gamma,U}^2\, h_{\gamma,Z}^2\, \mathcal{I}_{\gamma,U_{(1)}Z}}_{\text{2nd Order}} + \underbrace{h_{\gamma,Z}^4\, \mathcal{I}_{\gamma,ZZ}}_{\text{1st Order}}\right],\end{aligned} \tag{44}$$

$$\begin{aligned}\text{where}\quad \mathcal{I}_{\gamma,U_{(1)}U_{(1)}} &= \int \left(\gamma^{(2,0,0)}(u_1,u_2,z)\right)^2\, f_{UUZ}(u_1,u_2,z)\, d(u_1,u_2,z),\\ \mathcal{I}_{\gamma,U_{(1)}U_{(2)}} &= \int \left(\gamma^{(2,0,0)}(u_1,u_2,z)\gamma^{(0,2,0)}(u_1,u_2,z)\right)\, f_{UUZ}(u_1,u_2,z)\, d(u_1,u_2,z),\\ \mathcal{I}_{\gamma,U_{(1)}Z} &= \int \gamma^{(2,0,0)}(u_1,u_2,z)\gamma^{(0,0,2)}(u_1,u_2,z)\, f_{UUZ}(u_1,u_2,z)\, d(u_1,u_2,z),\\ \mathcal{I}_{\gamma,ZZ} &= \int \left(\gamma^{(0,0,2)}(u_1,u_2,z)\right)^2\, f_{UUZ}(u_1,u_2,z)\, d(u_1,u_2,z),\\ Q_{\gamma,1} &= \int \left(\tilde{\gamma}((u_1,u_2),(u_1,u_2),z) + \sigma_\varepsilon^2(u_1,u_2,z)\right)\, d(u_1,u_2,z),\quad \text{and}\\ Q_{\gamma,2} &= \int \tilde{\gamma}((u_1,u_2),(u_1,u_2),z)\, f_{UU}(u_1,u_2)\, d(u_1,u_2,z)\end{aligned}$$

By the same reasoning as in the preceding section, we fist determine assumptions on the U-bandwidth that maintain the order relation between the two variance terms as indicated in Eq. (44). The first requirement is that $h_{\gamma,U} = o(h_{\gamma,Z})$. This restriction makes those bias components that depend on $h_{\gamma,U}$ asymptotically negligible, since it implies that $h_{\gamma,U}^2\, h_{\gamma,Z}^2 = o(h_{\gamma,Z}^4)$ and therefore that $h_{\gamma,U}^4 = o(h_{\gamma,U}^2\, h_{\gamma,Z}^2)$. The latter leads to the order relations between the three bias terms as indicated in Eq. (44). The second requirement is that the U-bandwidth does not converge to zero too fast, namely that $Mh_{\gamma,U}^2 \to \infty$, which implies the order relation between the first two variance terms as indicated in Eq. (44).

This allows us to minimize the following simpler, but asymptotically equivalent AMISE function, which depends only on the Z-bandwidth:

$$\text{AMISE}_{\hat{\gamma}}^{\text{1st Order}}\left(h_{\gamma,Z}\right) = n^{-1}\, h_{\gamma,Z}^{-1}\, R(\kappa)\, Q_{\gamma,2} + \frac{1}{4}\left(\nu_2(K_\gamma)\right)^2 h_{\gamma,Z}^4\, \mathcal{I}_{\gamma,ZZ}.$$

The above equation is minimized by the following Z-bandwidth

$$h^D_{\gamma,Z} = \left(\frac{R(\kappa)\, Q_{\gamma,2}}{n\ (\nu_2(K_\gamma))^2\ \mathcal{I}_{\gamma,ZZ}} \right)^{1/5},$$

which is that of Eq. (14) in Theorem 3.6.

Parallel to the preceding section, we determine the U-bandwidth by plugging-in the above optimal Z-bandwidth into the AMISE function in Eq. (44) and by minimizing the (then classical) bias-variance trade-off between the asymptotic second order terms, which leads to the following expression for the U-bandwidths:

$$h^D_{\gamma,U} = \left(\frac{R(K_\gamma)\, Q_{\gamma,1}}{nM\ (\nu_2(K_\gamma))^2\ \mathcal{I}_{\gamma,U_{(1)}Z}} \right)^{1/4} \left(h^D_{\gamma,Z}\right)^{-3/4},$$

which is that of Eq. (15) in Theorem 3.6.

In order to check whether this U-bandwidth actually fulfills the two necessary requirements, we apply some rearrangements. Using that by Assumption AS $m \sim n^\theta$ and that by construction $M \sim m^2$, leads to the following more transparent presentation of the bandwidth rates:

$$h^D_{\gamma,Z} \sim M^{-1/(10\,\theta)} \quad \text{and} \quad h^D_{\gamma,U} \sim M^{-\eta_\gamma(\theta)} \quad \text{with} \quad \eta_\gamma(\theta) = \frac{1}{4} + \frac{1}{20\,\theta}. \tag{45}$$

With Eq. (45) it is now easily verified that the necessary requirements, i.e., that $h_{\gamma,U,\text{AMISE}} = o(h_{\gamma,Z,\text{AMISE}})$ and $Mh^2_{\gamma,U,\text{AMISE}} \to \infty$, are fulfilled iff $\theta > 1/5$.

3.3 Proofs of Corollaries 3.3 and 3.4

Corollaries 3.3 and 3.4 follow directly from Theorems 3.1 and 3.2 and from applying a central limit theorem for strictly stationary ergodic times series such as Theorem 9.5.5 of Karlin and Taylor (1975)[7]. Efficiency as well as the convergence rate are direct consequences from using the optimal bandwidths $h^D_{\mu,U}$, $h^D_{\mu,Z}$, $h^D_{\gamma,U}$, and $h^D_{\gamma,Z}$.

[7]Karlin, S., & Taylor, H. M. (1975). A first course in stochastic processes (2d ed). New York: Academic.

4 Data Sources

The data for our analysis come from four different sources. Hourly spot prices of the German electricity market are provided by the European Energy Power Exchange (EPEX) (`www.epexspot.com`), hourly values of Germany's gross electricity demand and electricity exchanges with other countries are provided by the European Network of Transmission System Operators for Electricity (`www.entsoe.eu`), German wind and solar power infeed data are provided by the transparency platform of the European energy exchange (`www.eex-transparency.com`), and German air temperature data are available from the German Weather Service (`www.dwd.de`).

The data dimensions are given by $n = 24$, $N = 552$, $T = 261$, and $T = 262$, where the latter two numbers are the number of working days one year before and one year after Germany's nuclear phaseout. Very few (0.2%) of the data tuples (Y_{ij}, U_{ij}, Z_i) with prices $Y_{ij} > 200$ EUR/MWh are considered as outliers and removed. Such extreme prices cannot be explained by the merit order model, since the marginal costs of electricity production usually do not exceed the value of about 200 EUR/MWh. Prices above this threshold are often referred to as "price spikes" and have to be modeled using different approaches (Ch. 4 in Burger (2008)[8]).

4.1 Plotting the confidence intervals

Figure 6 shows the confidence intervals which serve as basis for our test-decisions.

[8]Burger, M., B. Graeber, and G. Schindlmayr (2008). Managing Energy Risk: An Integrated View on Power and Other Energy Markets (1. ed.). Wiley.

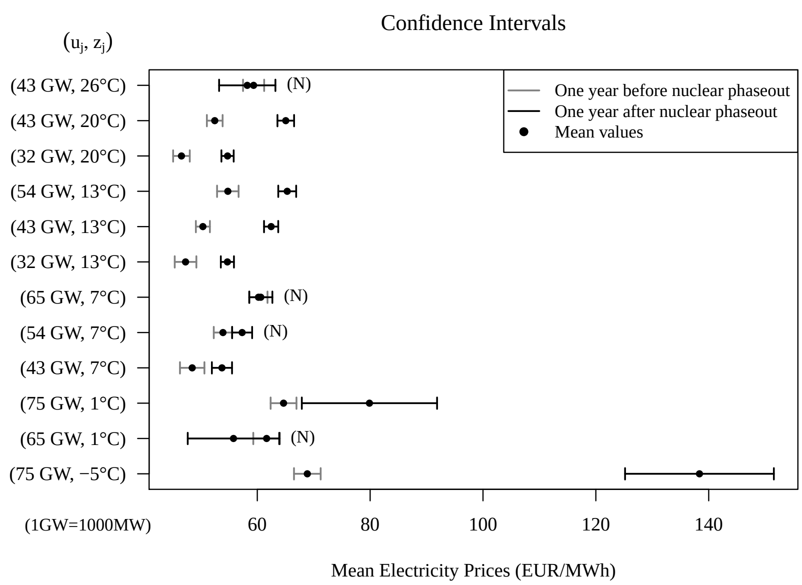

Figure 6: Confidence intervals used for testing point-wise differences in the mean functions $\mu_A(u_j, z_j)$ and $\mu_B(u_j, z_j)$. Insignificant results are marked by “(N)”.