Submitted to the Annals of Applied Statistics
arXiv: arXiv:0000.0000

FINITE SAMPLE CORRECTION FOR TWO-SAMPLE INFERENCE WITH SPARSE COVARIATE ADJUSTED FUNCTIONAL DATA

By Dominik Liebl

University of Bonn

This work is motivated by the problem of testing for differences in the mean electricity prices before and after Germany's abrupt nuclear phaseout after the nuclear disaster in Fukushima Daiichi, Japan, in mid-March 2011. Given the nature of the data, we approach this problem using a Local Linear Kernel (LLK) estimator for the nonparametric mean function of sparse covariate adjusted functional data. We demonstrate that the two-sample test statistic based on existing asymptotic results suffers from severe size-distortions. Motivated by this finding, we propose a theory-based finite sample correction for the considered LLK estimator. Our simulation study shows that this finite sample correction is very effective in eliminating the size-distortions. The practical use of the procedure is demonstrated in our real data application, where we address some open questions on the differences in the mean electricity prices before and after Germany's (partial) nuclear 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 the nuclear phaseout (see, e.g., Nestle, 2012). This surprising finding is reconsidered in our real data application in Section 6.

Pricing at electricity exchanges is explained well by the so-called merit-order model. This model assumes that the spot prices are based on the so-called "merit-order curve"—a monotonically increasing curve reflecting the increasingly ordered generation costs of the installed power plants. The merit-order model is a fundamental market model (see, for instance, Burger, Graeber and Schindlmayr, 2008, Ch. 4) and most important for the explanation of electricity spot prices in the literature on energy economics (see, e.g., Burger et al. (2004), Sensfuß, Ragwitz and Genoese (2008), Hirth (2013),

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

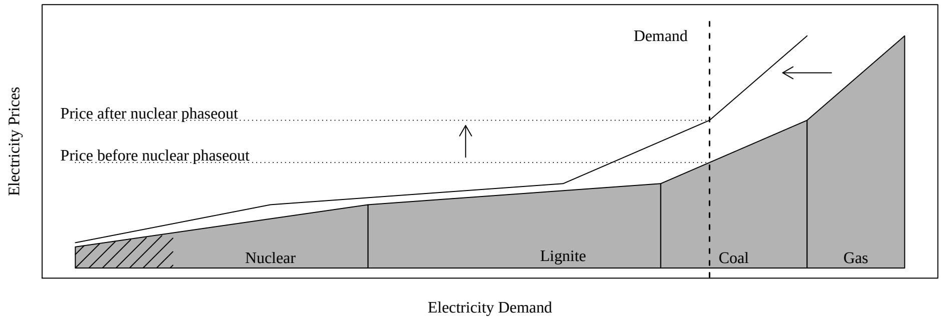

FIG 1. *Sketch of the merit-order curve and the theoretical price effect of the nuclear power phaseout. The dashed region signifies the proportion of phased out nuclear power plants.*

Liebl (2013), Cludius et al. (2014), and Bublitz, Keles and Fichtner (2017)).

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. Electricity demand is assumed to be price-inelastic in the short-term perspective of a spot market. The latter assumption is regularly found in the literature (see, e.g., Sensfuß, Ragwitz and Genoese, 2008) and confirmed in empirical studies (see, e.g., Lijesen, 2007).

We consider electricity spot prices from the European Power Exchange (EPEX), where the hourly electricity spot prices of day i are settled simultaneously at 12 am the day before (see, for instance, Benth, Kholodnyi and Laurence, 2013, Ch. 6). Following the literature, we differentiate between "peak hours" (from 9am to 8pm) and "off-peak hours" (all other hours) and focus on the $m = 12$ peak-hours, since these show the largest variations in electricity prices and electricity demand.

The daily simultaneous pricing scheme at the EPEX results in a daily varying merit-order curve (or simply "price curve") X_i. Therefore, we model the electricity spot price $Y_{ij} \in \mathbb{R}$ of day i and peak hour j as a discretization point of the underlying (unobserved) daily merit-order curve X_i evaluated at the corresponding value of electricity demand $U_{ij} \in \mathbb{R}$,

$$(1.1) \qquad Y_{ij} = X_i(U_{ij}, Z_i) + \epsilon_{ij}, \quad j = 1, \dots, m, \quad i = 1, \dots, n.$$

The shape and the location of the price curve $X_i(., Z_i) \in L^2[a(Z_i), b(Z_i)]$, with $[a(Z_i), b(Z_i)] \subset \mathbb{R}$, are allowed to be adjusted by the (curve-specific) covariate $Z_i \in \mathbb{R}$, where Z_i denotes the daily mean air temperature. The daily mean air temperature is the most important exogenous factor in energy spot markets and of crucial importance to model the domain-shifts in the empirical price curves X_i (see Figure 2). The statistical error term ϵ_{ij} is assumed to be independently and identically distributed (iid) with mean zero and finite variance and assumed to be independent from X_i, U_{ij}, and Z_i.

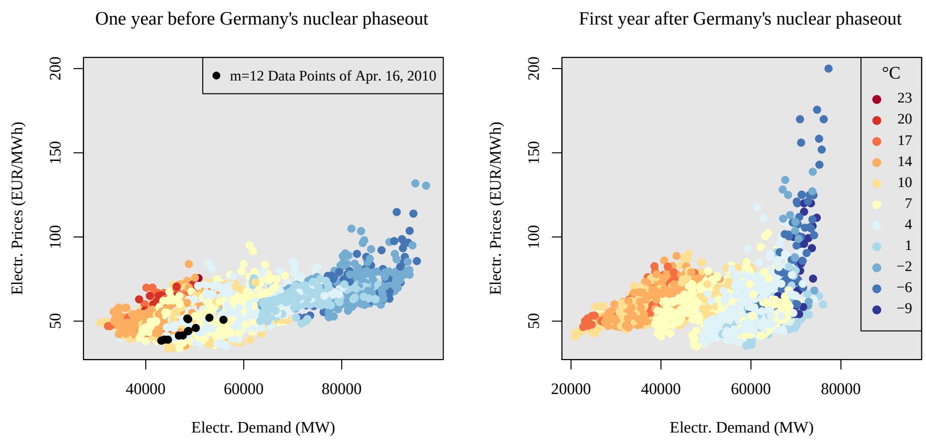

FIG 2. *Scatter plots of the price-demand data pairs* (Y_{ij}, U_{ij})*, where the colors encode the additional covariate of daily mean air temperature* Z_i*. The left plot shows the data from one year before Germany's partial nuclear phaseout, i.e., from March 15, 2010 to March 14, 2011; the right plot shows the data from one year after, i.e., from March 15, 2011 to March 14, 2012.*

The analyzed data are shown in the scatter plots in Figure 2. The covariate Z_i clearly influences the location of the spot prices Y_{ij}, since realizations of U_{ij} are only observed within temperature-specific subintervals, i.e., $U_{ij} \in [a(Z_i), b(Z_i)]$. This observation motivates our modeling assumption that $X_i(., Z_i) \in L^2[a(Z_i), b(Z_i)]$.

Germany's (partial) nuclear phaseout means a shift of the mean merit-order curve resulting in higher electricity spot prices—particularly at hours with large values of electricity demand (see Figure 1). This effect is obvious for very cold days (see Figure 2), though not obvious on other days. In our application we estimate and test the differences in the mean prices for different values of electricity demand and temperature.

The objective of this article is a two-sample test for testing the pointwise null hypothesis of equal means against the alternative of larger mean values after Germany's nuclear phaseout, i.e.,

$$H_0: \ \mu_A(u,z) = \mu_B(u,z) \quad \text{vs} \quad H_1: \ \mu_A(u,z) > \mu_B(u,z),$$

where

$$\mu_A(u,z) = \mathbb{E}\left(X_i^A(u,z)\right) \qquad \text{and} \qquad \mu_B(u,z) = \mathbb{E}\left(X_i^B(u,z)\right)$$

are the mean functions of the random price functions After, $X_i^A(.,z)$, and Before, $X_i^B(.,z)$, Germany's nuclear phaseout.

We estimate the mean functions μ_A and μ_B separately for each period $P \in \{A,B\}$ from the observed data points $\{(Y_{ij}^P, U_{ij}^P, Z_i^P); 1 \leq i \leq n_P, 1 \leq j \leq m\}$ using the Local Linear Kernel (LLK) estimator suggested by Jiang and Wang (2010). This mean estimator of Jiang and Wang (2010) is currently the only nonparametric alternative for sparsely observed covariate adjusted functional data. The pre-smoothing approach suggested by Cardot (2007) does not lead to reliable estimation results as we have only $m = 12$ ("peak hours") data points per function.

Right now the only way to derive a two-sample test for detecting differences in the mean functions from sparse covariate adjusted functional data is to use the asymptotic normality results of Jiang and Wang (2010). We demonstrate, however, that the asymptotic variance expression of Jiang and Wang (2010) severely underestimates the actual variance of the LLK estimator in sparse (i.e., "small-m large-n") data scenarios of practical relevance. This leads to strong size-distortions of the two-sample test statistic resulting in drastic over-rejections of the null hypothesis.

We are able to explain this finding based on our theoretical results. Jiang and Wang (2010) derive their asymptotic results using a "finite-m" asymptotic, where m is bounded while $n \to \infty$. This finite-m asymptotic, however, neglects an additional functional-data-specific variance term which is asymptotically negligible, but typically not negligible in practical sparse or moderately sparse data scenarios (e.g., $n = 200$ and $m = 5$ or $m = 15$). The additional variance term becomes visible under a double asymptotic where $n \to \infty$ and $m \to \infty$, but with m diverging significantly more slowly than n to respect our context of sparse or moderately sparse functional data.

Based on our theoretical results, we contribute a finite sample correction which accounts for the additional functional-data-specific variance term. We examine the finite sample properties of our new test statistic and compare them with those of the test statistic based on the asymptotic results of Jiang and Wang (2010) by means of a simulation study. In our application we use our test statistic in order to test for point-wise differences in the mean electricity prices before and after Germany's (partial) nuclear phaseout.

The literature on covariate adjusted functional data is fairly scarce. Cardot (2007) considers functional principal component analysis for covariate adjusted random functions, though he focuses on the case of dense functional data and does not derive inferential results for his mean estimate. Jiang and Wang (2010) show many theoretical results of fundamental importance for sparse covariate adjusted functional data and we use their point-wise asymptotic normality result for the mean function as a benchmark. Li, Staicu and Bondell (2015) consider a copula-based model and Zhang and Wei (2015) propose an iterative algorithm for computing functional principal components, though neither contributes inferential results for the covariate adjusted mean function.

For the case *without* covariate adjustments there are several papers considering inference for the mean function. In the following we give a non-exhaustive overview. Zhang and Chen (2007) and Hall and Van Keilegom (2007) consider inference for dense functional data. Benko, Härdle and Kneip (2009) develop bootstrap procedures for testing the L^2 distance between mean functions (or eigenfunctions) based on densely or fully observed random functions. Cao, Yang and Todem (2012) derive simultaneous confidence bands for the mean function in the case of dense functional data. Gromenko and Kokoszka (2012) also consider an L^2 based test statistic and Horváth, Kokoszka and Reeder (2013) focus on the case of dependent functional data under a similar framework. Fogarty and Small (2014) generalize the related topic on equivalence testing from scalar to functional data. Vsevolozhskaya, Greenwood and Holodov (2014) consider pairwise comparisons of functional means using fully observed functions. Zhang and Wang (2016) derive asymptotic normality results for the LLK estimator of the mean function for sparse to dense functional data. We emphasize, however, that the existing results for functional data *without* covariate adjustments cannot easily be generalized to account for additional covariate adjustments. Related to our work is also that of Serban (2011) and Gromenko, Kokoszka and Sojka (2017), who consider covariate adjusted, namely, spatio-temporal functional data, though, do not focus on a sparse functional data context. Readers with a general interest in functional data analysis are referred to the textbooks of Ramsay and Silverman (2005), Ferraty and Vieu (2006), Horváth and Kokoszka (2012), and Hsing and Eubank (2015). Our finite sample correction for two-sample inference using the LLK estimator is motivated by our specific real data problem, but we expect it to be also of a more general interest in applied statistics.

The rest of the paper is organized as follows. The next section introduces the nonparametric regression model under consideration and the LLK estimator. Section 3 contains our asymptotic normality result and introduces the new two-sample test statistic with finite sample correction. Section 4 introduces our estimators used to approximate the unknown bias, variance and bandwidth components. Our simulation study is found in Section 5. Section 6 contains the real data study. The proofs of our theoretical results are based on standard arguments in nonparametric statistics and can be found in the online supplement supporting this article (Liebl, 2017).

2. Nonparametric regression model and LLK estimator. In the following, we use a common notation for both samples (A and B) unless a differentiation is required by the context. Without loss of generality, we consider a standardized domain where $(U_{ij}, Z_i) \in [0,1]^2$ such that $X_i(., Z_i) \in L^2([0,1])$. The latter can always be achieved by computing $U_{ij}^{new} = (U_{ij}^{orig} - a(Z_i^{orig}))/(b(Z_i^{orig}) - a(Z_i^{orig}))$ and $Z_i^{new} = (Z_i^{orig} - \min_{1\le i\le n}(Z_i^{orig}))/(\max_{1\le i\le n}(Z_i^{orig}) - \min_{1\le i\le n}(Z_i^{orig}))$. The functional interval borders $a(.)$ and $b(.)$ are unobserved, but can be estimated from the data points (U_{ij}, Z_i) using the LLK estimators of Martins-Filho and Yao (2007).

Let X_i^c denote the centered function $X_i^c(u,z) = X_i(u,z) - \mathbb{E}(X_i(u,z))$. Model (1.1) can then be written as a nonparametric regression model with the conditional mean function $\mu(U_{ij}, Z_i) = \mathbb{E}(X_i(U_{ij}, Z_i)|\mathbf{U}, \mathbf{Z})$, given $\mathbf{U} = (U_{11}, \dots, U_{nm})^\top$ and $\mathbf{Z} = (Z_1, \dots, Z_n)^\top$, i.e.,

$$Y_{ij} = \mu(U_{ij}, Z_i) + X_i^c(U_{ij}, Z_i) + \epsilon_{ij}, \tag{2.1}$$

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 classical iid error term with mean zero, finite variance $\mathrm{V}(\epsilon_{it}) = \sigma_\epsilon^2$, and assumed to be independent from X_s^c, $U_{s\ell}$, and Z_s for all $s = 1, \dots, n$ and $\ell = 1, \dots, m$.

Note that Model (2.1) has a rather unusual composed error term consisting of a functional $X_i^c(U_{ij}, Z_i)$ and a scalar component ϵ_{ij}. The functional error component introduces very strong local correlations, since

$$\mathrm{Corr}\left(Y_{ij}, Y_{ik}|U_{ij} = u_1, U_{ik} = u_2, Z_i = z\right) = \mathrm{Corr}\left(X_i^c(u_1, z), X_i^c(u_2, z)\right) \approx 1,$$

for $u_1 \approx u_2$, which leads to the above mentioned functional-data-specific variance term that makes the finite sample correction necessary.

We estimate the mean function $\mu(u,z)$ using the same LLK estimator as considered in Jiang and Wang (2010). In the following we define the estimator based on a matrix notion:

$$\begin{aligned} \hat{\mu}(u, z; h_{\mu,U}, h_{\mu,Z}) &= \\ &= e_1^\top \left([\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz} [\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]\right)^{-1} [\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz} \mathbf{Y}, \end{aligned} \tag{2.2}$$

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 $nm \times 3$ 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 usual second-order kernel such as, e.g., the Epanechnikov or the Gaussian kernel. The kernel constants are denoted by $\nu_2(K_\mu) = (\nu_2(\kappa))^2$, with $\nu_2(\kappa) = \int_{[0,1]} u^2\kappa(u)du$, and $R(K_\mu) = R(\kappa)^2$, with $R(\kappa) = \int_{[0,1]} \kappa(u)^2 du$. All vectors and matrices are filled in correspondence with the nm dimensional vector $\mathbf{Y} = (Y_{11}, Y_{12}, \ldots, Y_{n,m-1}, Y_{nm})^\top$.

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

- A1 (Asymptotic Scenario) $nm \to \infty$, where $m = m_n \geq 2$ such that $m_n \asymp n^\theta$ with $0 \leq \theta < \infty$. Hereby, "$m_n \asymp n^\theta$" denotes that the two sequences m_n and n^θ are asymptotically equivalent, i.e., that $\lim_{n\to\infty}(m_n/n^\theta) = C$ with constant $0 < C < \infty$.
- A2 (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 $(y, u, z) \in \mathbb{R} \times [0,1]^2$ and zero else. The error term ϵ_{it} is iid and independent from X_s^c, $U_{s\ell}$, and Z_s for all $s = 1, \ldots, n$ and $\ell = 1, \ldots, m$.
- A3 (Smoothness & Kernel) The pdf $f_{YUZ}(y, u, z)$ and its marginals are continuously differentiable. All second-order derivatives of the function μ are continuous. The (auto-)covariance 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. The multiplicative kernel functions K_μ and K_γ are products of second-order kernel functions κ.
- A4 (Moments & Dependency) X_i, U_{ij}, and Z_i are strictly stationary, ergodic, and weakly dependent time-series with auto-covariances that converge uniformly to zero at a geometrical rate. It is assumed that $\mathbb{E}(X_i(u, z)^4) < \infty$, $\mathbb{E}(\epsilon_{ij}) = 0$, $\mathbb{E}(\epsilon_{ij}^2) = \sigma_\epsilon^2 < \infty$ for all (u, z), i, and j.
- A5 (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 $nM \to \infty$.

Remark. Assumption A1 is a simplified version of the asymptotic setup of Zhang and Wang (2016). The case $\theta = 0$ implies that m is bounded which corresponds to a simplified version of the finite-m asymptotic considered by Jiang and Wang (2010). For $0 < \theta < \infty$ we can consider all further scenarios from sparse to dense functional data. In line with our real data application, we consider a deterministic m as also done, for instance, by Hall, Müller and Wang (2006). However, our results are generalizable to a random m using some minor modifications.

The following theorem contains our asymptotic normality result (with and without finite sample correction) for the LLK estimator $\hat{\mu}$ in Eq. (2.2):

THEOREM 3.1 (Asymptotic normality). *Let $m/n^{1/5} \to 0$, let (u,z) be an interior point of $[0,1]^2$, and assume optimal bandwidth rates $h_{\mu,U} \asymp h_{\mu,Z} \asymp (nm)^{-1/6}$. Under Assumptions A1-A5 the LLK estimator $\hat{\mu}$ in Eq. (2.2) is then asymptotically normal, i.e.,*
(i) without finite sample correction:

$$\left(\frac{\hat{\mu}(u,z;h_{\mu,U},h_{\mu,Z}) - B_\mu(u,z;h_{\mu,U},h_{\mu,Z}) - \mu(u,z)}{\sqrt{V_\mu^{\mathrm{I}}(u,z;h_{\mu,U},h_{\mu,Z})}}\right) \overset{a}{\sim} N(0,1)$$

(ii) with finite sample correction:

$$\left(\frac{\hat{\mu}(u,z;h_{\mu,U},h_{\mu,Z}) - B_\mu(u,z;h_{\mu,U},h_{\mu,Z}) - \mu(u,z)}{\sqrt{V_\mu^{\mathrm{I}}(u,z;h_{\mu,U},h_{\mu,Z}) + V_\mu^{\mathrm{II}}(u,z;h_{\mu,Z})}}\right) \overset{a}{\sim} N(0,1)$$

where $\mu \in \{\mu_A, \mu_B\}$,

$$\begin{aligned}
B_\mu(u,z;h_{\mu,U},h_{\mu,Z}) &= \frac{1}{2}\,\nu_2(K_\mu)\,\left(h_{\mu,U}^2\,\mu^{(2,0)}(u,z) + h_{\mu,Z}^2\,\mu^{(0,2)}(u,z)\right),\\
V_\mu^{\mathrm{I}}(u,z;h_{\mu,U},h_{\mu,Z}) &= \frac{1}{nm}\left[\frac{R(K_\mu)}{h_{\mu,U}h_{\mu,Z}}\frac{\gamma(u,u,z)+\sigma_\epsilon^2}{f_{UZ}(u,z)}\right],\\
V_\mu^{\mathrm{II}}(u,z;h_{\mu,Z}) &= \frac{1}{n}\left[\left(\frac{m-1}{m}\right)\frac{R(\kappa)}{h_{\mu,Z}}\frac{\gamma(u,u,z)}{f_Z(z)}\right], \quad \textit{and}\\
\mu^{(k,l)}(u,z) &= (\partial^{k+l}/(\partial u^k \partial z^l))\mu(u,z).
\end{aligned}$$

Theorem 3.1 implies the standard optimal convergence rate $(nm^{-1/3})$ for bivariate LLK estimators. The finite sample correction in result *(ii)* is accomplished by the second variance term $V_\mu^{\mathrm{II}}(u,z;h_{\mu,Z})$ which could be dropped from a pure asymptotic view, since it is asymptotically negligible in comparison to the first variance term, i.e., $V_\mu^{\mathrm{II}}(u,z;h_{\mu,Z}) = o(V_\mu^{\mathrm{I}}(u,z;h_{\mu,U},h_{\mu,Z}))$ under the assumptions of Theorem 3.1. However, this second variance term is typically not negligible in practice and serves as a very effective finite sample correction.

Note that the variance effects due to the autocorrelations from our time-series context are not first-order relevant. The reason for this is that we localize with respect to the variables U and Z and not with respect to time i. The resulting de-correlation effect is referred to as the "whitening window" property (see, for instance, Fan and Yao, 2003, Ch. 5.3).

Result *(i)* of Theorem 3.1 (i.e., without finite sample correction) is essentially equivalent to Theorem 3.2 in Jiang and Wang (2010), who, however, consider the case of a bounded m. In contrast, our asymptotic normality results *(i)* and *(ii)* in Theorem 3.1 hold for all $m \to \infty$, $n \to \infty$ with $m/n^{1/5} \to 0$. This explains our empirical finding in our simulations in Section 5 that the asymptotic normality result with finite sample correction (result *(ii)*) provides very good finite sample approximations for practical "small-m large-n" (e.g., $m = 5$ and $n = 200$) as well as "moderate-m large-n" cases (e.g., $m = 15$ and $n = 200$).

The following corollary follows directly from Theorem 3.1 and contains the asymptotic normality results for our two-sample test statistics (with and without finite sample correction):

COROLLARY 3.1 (Two-sample test statistics). *Under the same conditions as in Theorem 3.1 and under the null hypothesis H_0: $\mu_A(u,z) = \mu_B(u,z)$, the following two two-sample test statistics are asymptotically normal.*

(i) Without finite sample correction:

$$Z_{u,z} = \left(\frac{\hat{\mu}_A(u,z) - B_{\mu_A}(u,z) - \hat{\mu}_B(u,z) + B_{\mu_B}(u,z)}{\sqrt{V^{\mathrm{I}}_{\mu_A}(u,z) + V^{\mathrm{I}}_{\mu_B}(u,z)}} \right) \overset{a}{\sim} N(0,1)$$

(ii) With finite sample correction:

$$Z^c_{u,z} = \left(\frac{\hat{\mu}_A(u,z) - B_{\mu_A}(u,z) - \hat{\mu}_B(u,z) + B_{\mu_B}(u,z)}{\sqrt{V^{\mathrm{I}}_{\mu_A}(u,z) + V^{\mathrm{II}}_{\mu_A}(u,z) + V^{\mathrm{I}}_{\mu_B}(u,z) + V^{\mathrm{II}}_{\mu_B}(u,z)}} \right) \overset{a}{\sim} N(0,1)$$

The dependencies of B_μ, V^{I}_μ, and V^{II}_μ on the bandwidth parameters $h_{\mu,U}$ and $h_{\mu,Z}$ are suppressed for readability reasons.

The test statistics $Z_{u,z}$ and $Z^c_{u,z}$ are infeasible as they depend on the unknown bias, variance, and bandwidth expressions, B_μ, V^{I}_μ, V^{II}_μ, $h_{\mu,U}$ and $h_{\mu,Z}$. In Section 4 we propose estimators $\hat{V}^{\mathrm{I}}_\mu$, $\hat{V}^{\mathrm{II}}_\mu$, and $\hat{B}_\mu$ for the unknown bias and variance terms. For estimating optimal bandwidths we use (bivariate) Generalized Cross Validation (GCV).

Because errors from estimating the bias, variance, and bandwidth components could contaminate our simulation results, we also consider a theoretical benchmark scenario based on the theoretical (usually unknown) bias, variance, and bandwidth expressions. For this, we provide the following theoretically optimal bandwidths, where optimality is with respect to a minimal Asymptotic Mean Integrated Squared Error (AMISE) of $\hat{\mu}$:

THEOREM 3.2 (AMISE optimal bandwidths for $\hat{\mu}$). *Let $m/n^{1/5} \to 0$ and let (u, z) be an interior point of $[0,1]^2$. Under Assumptions A1-A5 the AMISE optimal bandwidths for the LLK estimator $\hat{\mu}$ in Eq. (2.2) are then given by*

$$h_{\mu,U}^{opt} = \left(\frac{R(K_\mu)\, Q_\mu\, (\mathcal{I}_{ZZ}^{\mu})^{3/4}}{nm\ (\nu_2(K_\mu))^2 \left[\mathcal{I}_{UU}^{\mu}{}^{1/2}\, \mathcal{I}_{ZZ}^{\mu}{}^{1/2} + \mathcal{I}_{UZ}^{\mu} \right] \mathcal{I}_{UU}^{\mu}{}^{3/4}} \right)^{1/6} \tag{3.1}$$

$$h_{\mu,Z}^{opt} = \left(\frac{\mathcal{I}_{UU}^{\mu}}{\mathcal{I}_{ZZ}^{\mu}} \right)^{1/4} h_{\mu,U}^{opt}, \quad \textit{where} \tag{3.2}$$

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

The optimal bandwidths rates ($(nm)^{-1/6}$) are well known for bivariate LLK estimators and the bandwidth expressions in Theorem 3.2 are essentially equivalent to those in Herrmann et al. (1995).

The proofs of Theorems 3.1 and 3.2 and Corollary 3.1 require the derivation of the standard bias and variance expression for LLK estimators (see Ruppert and Wand, 1994). To avoid the presentation of lengthy, but classical, mathematical results, the proofs are given as a technical report and can be found in the online supplement supporting this article.

4. Practical approximations.

We approximate the unknown bias term $B_\mu(u, z; h_{\mu,U}, h_{\mu,Z})$ by

$$\begin{aligned}
&\hat{B}_\mu(u,z; h_{\mu,U}, h_{\mu,Z}) = \\
&= \tfrac{\nu_2(K_\mu)}{2} \Big(h_{\mu,U}^2 \hat{\mu}^{(2,0)}\big(u,z; g_{\mu,U}, g_{\mu,Z}\big) + h_{\mu,Z}^2 \hat{\mu}^{(0,2)}\big(u,z; g_{\mu,U}, g_{\mu,Z}\big) \Big),
\end{aligned} \tag{4.1}$$

where the estimates of the second-order partial derivatives $\hat{\mu}^{(2,0)}$ and $\hat{\mu}^{(0,2)}$ are local polynomial (order 3) kernel estimators. That is,

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

with $e_3^\top = (0,0,1,0,0,0,0)$, $\mathbf{U}_u^{1:3} = [\mathbf{U}_u,\mathbf{U}_u^2,\mathbf{U}_u^3]$, $\mathbf{Z}_z^{1:3} = [\mathbf{Z}_z,\mathbf{Z}_z^2,\mathbf{Z}_z^3]$, and diagonal matrix $\mathbf{W}_{\mu,uz}$ with weights $g_{\mu,U}^{-1}\,\kappa(g_{\mu,U}^{-1}(U_{ij}-u))\; g_{\mu,Z}^{-1}\,\kappa(g_{\mu,Z}^{-1}(Z_i-z))$ on its diagonal. The estimator $\hat{\mu}^{(0,2)}$ is defined correspondingly, but with $e_3^\top$ replaced by $e_6^\top = (0,0,0,0,0,1,0)$. For estimating the bandwidths $g_{\mu,U}$ and $g_{\mu,Z}$ we use bivariate GCV based on second-order differences. We follow the procedure of Charnigo and Srinivasan (2015), but use a GCV-penalty instead of the (asymptotically equivalent) C_p-penalty proposed there.

We estimate the unknown first variance term $V_\mu^{\mathrm{I}}(u,z;h_{\mu,U},h_{\mu,Z})$ by

(4.2)
$$\hat{V}_\mu^{\mathrm{I}}(u,z;h_{\mu,U},h_{\mu,Z},h_{\gamma,U},h_{\gamma,Z}) = \frac{1}{nm}\left[\frac{R(K_\mu)}{h_{\mu,U}h_{\mu,Z}}\,\frac{\hat{\gamma}^{\mathrm{ND}}(u,u,z;h_{\gamma,U},h_{\gamma,Z})}{\hat{f}_{UZ}(u,z)}\right],$$

where the LLK estimator $\hat{\gamma}^{\mathrm{ND}}(u,u,z;h_{\gamma,U},h_{\gamma,Z})$ of $\gamma(u,u,z)+\sigma_\epsilon^2$ is defined as following:

$$\hat{\gamma}^{\mathrm{ND}}(u,u,z;h_{\gamma,U},h_{\gamma,Z}) = \\ e_1^\top \left([\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top \mathbf{W}_{\gamma,uz}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]\right)^{-1}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top \mathbf{W}_{\gamma,uz}\hat{\mathbf{C}}^{\mathrm{ND}}, \tag{4.3}$$

with $\hat{\mathbf{C}}^{\mathrm{ND}} = (\hat{C}_{111},\dots,\hat{C}_{ijj}\dots,\hat{C}_{nmm})^\top$ consisting only of the Noisy Diagonal (ND) raw-covariances $\hat{C}_{ijj}^{\mathrm{ND}} = (Y_{ij}-\hat{\mu}(U_{ij},Z_i;h_{\mu,U},h_{\mu,Z}))^2$ for which $\mathbb{E}(\hat{C}_{ijj}|\mathbf{U},\mathbf{Z}) = \gamma(U_{ij},U_{ik},Z_i)+\sigma_\epsilon^2$, neglecting the bias in $\hat{\mu}$. Note that $\hat{\gamma}^{\mathrm{ND}}$ is equivalent to the LLK estimator "$\hat{V}$" in Jiang and Wang (2010).

The unknown second variance term $V_\mu^{\mathrm{II}}(u,z;h_{\mu,Z})$ is estimated by

(4.4)
$$\hat{V}_\mu^{\mathrm{II}}(u,z;h_{\mu,Z},h_{\gamma,U},h_{\gamma,Z}) = \frac{1}{n}\left[\left(\frac{m-1}{m}\right)\frac{R(\kappa)}{h_{\mu,Z}}\,\frac{\hat{\gamma}(u,u,z;h_{\gamma,U},h_{\gamma,Z})}{\hat{f}_Z(z)}\right],$$

where the LLK estimator $\hat{\gamma}(u_1,u_2,z;h_{\gamma,U},h_{\gamma,Z})$ of $\gamma(u_1,u_2,z)$ defined as following:

$$\hat{\gamma}(u_1,u_2,z;h_{\gamma,U},h_{\gamma,Z}) = \\ e_1^\top \left([\mathbf{1},\mathbf{U}_{u_1},\mathbf{U}_{u_2},\mathbf{Z}_z]^\top \mathbf{W}_{\gamma,u_1u_2z}[\mathbf{1},\mathbf{U}_{u_1},\mathbf{U}_{u_2},\mathbf{Z}_z]\right)^{-1}\times \\ [\mathbf{1},\mathbf{U}_{u_1},\mathbf{U}_{u_2},\mathbf{Z}_z]^\top \mathbf{W}_{\gamma,u_1u_2z}\hat{\mathbf{C}}. \tag{4.5}$$

Here, $e_1 = (1,0,0,0)^\top$ 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 $(1, U_{ij} - u_1, U_{ik} - u_2, Z_i - z)$ and $M = m^2 - m$. (The latter explains the requirement of Assumption A1 that $m \geq 2$.) The $nM \times nM$ dimensional diagonal weighting matrix $\mathbf{W}_{\gamma,u_1u_2z}$ holds the trivariate multiplicative kernel weights $K_{\gamma,h_{\gamma,U},h_{\gamma,Z}}(U_{ij} - u_1, U_{ik} - u_2, Z_i - z) = h_{\gamma,U}^{-1}\,\kappa(h_{\gamma,U}^{-1}(U_{ij} - u_1))\; h_{\gamma,U}^{-1}\,\kappa(h_{\gamma,U}^{-1}(U_{ik} - u_2))\; h_{\gamma,Z}^{-1}\,\kappa(h_{\gamma,Z}^{-1}(Z_i - z))$. All vectors and matrices are filled in correspondence with the nM dimensional vector $\hat{\mathbf{C}} = (\hat{C}_{112}, \ldots, \hat{C}_{ijk}, \ldots, \hat{C}_{nm,m-1})^\top$ consisting only of the off-diagonal raw-covariances

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

with $j \neq k \in \{1, \ldots, m\}$ for which $\mathbb{E}(\hat{C}_{ijk}|\mathbf{U}, \mathbf{Z}) = \gamma(U_{ij}, U_{ik}, Z_i)$, neglecting the bias in $\hat{\mu}$. We use bivariate GCV in order to estimate the bandwidth parameters $h_{\mu,U}$, $h_{\mu,Z}$, $h_{\gamma,U}$ and $h_{\gamma,Z}$.

5. Simulation.

We generate data from $Y_{ij} = \mu(U_{ij}, Z_i) + X_i^c(U_{ij}, Z_i) + \epsilon_{ij}$, where $U_{ij} \sim \text{Unif}(0,1)$, $Z_i \sim \text{Unif}(0,1)$, $\epsilon_{ij} \sim \text{N}(0,1)$, $X_i^c(u,z) = \xi_{i1}\psi_1(u,z) + \xi_{i2}\psi_2(u,z)$, $\xi_{i1} \sim \text{N}(0,3)$, and $\xi_{i2} \sim \text{N}(0,2)$, where we consider the following two Data Generating Processes (DGPs):

DGP-1	$\mu_A(u,z) = 5\sin(\pi\, u\, z/2)$	$\psi_1(u,z) = \sin(1\pi uz)$
	$\mu_B(u,z) = \mu_A(u,z) - \Delta$	$\psi_2(u,z) = \sin(2\pi uz)$
DGP-2	$\mu_A(u,z) = 5\sin(\pi uz)$	$\psi_1(u,z) = \sin(2\pi uz)$
	$\mu_B(u,z) = \mu_A(u,z) - \Delta$	$\psi_2(u,z) = \sin(3\pi uz)$

with

$$\Delta = \begin{cases} 0 & \text{if} \quad H_0\colon \mu_A(u,z) = \mu_B(u,z) \\ 1 & \text{if} \quad H_1\colon \mu_A(u,z) > \mu_B(u,z); \end{cases}$$

DGP-2 is the more complex process as its mean and basis functions have a higher curvature.

As sample sizes we consider $n \in \{100, 200\}$ and $m \in \{5, 15\}$. For each sample size, each DGP, and each $\Delta \in \{0, 1\}$ we generate 5000 Monte-Carlo samples. The two-sample test statistics with plugged-in empirical bias and variance expressions from Equations (4.1)-(4.5) and GCV-optimal bandwidths will be denoted as:

$$\widehat{Z} \text{ and } \widehat{Z}^c.$$

Because errors from the plug-in estimates for the bias, variance, and bandwidth components could contaminate our simulation results, we also consider a theoretical benchmark scenario based on the theoretical (usually unknown) bias, variance, and bandwidth expressions in Theorems 3.1 and 3.2. These theoretical benchmark test statistics will be denoted as:

$$Z \text{ and } Z^c.$$

The test statistics are evaluated at $(u, z) = (0.5, 0.5)$. At this point the second order derivatives of the mean functions are non-zero, namely,

$$\begin{array}{ll} \text{DGP-1} & \mu^{(2,0)}(0.5, 0.5) = \mu^{(0,2)}(0.5, 0.5) \approx -1.2 \\ \text{DGP-2} & \mu^{(2,0)}(0.5, 0.5) = \mu^{(0,2)}(0.5, 0.5) \approx -8.7 \end{array}$$

for $\mu \in \{\mu_A, \mu_B\}$, which allows us to asses the performance of the empirical the bias correction in $\widehat{Z}$ and $\widehat{Z}^c$. We consider only one evaluation point, since the whole procedure is computationally extremely demanding. The simulation study is implemented using the statistical computer language R (R Core Team, 2017) and the R-script can be found in the online supplement supporting this article (Liebl, 2017).

TABLE 1

Empirical size ($\Delta = 0$) and size-adjusted power ($\Delta = 1$) in percent.

			$n = 100$				$n = 200$			
			$m = 5$		$m = 15$		$m = 5$		$m = 15$	
			5%	1%	5%	1%	5%	1%	5%	1%
$\Delta = 0$	DGP-1	$\widehat{Z}$	15.3	7.0	21.2	13.5	16.6	8.9	35.1	27.4
		$\widehat{Z}^c$	5.0	0.6	3.9	0.5	3.2	0.4	4.1	0.4
		Z	14.0	6.0	24.3	15.9	13.8	6.3	23.9	15.5
		Z^c	3.7	0.7	3.7	0.8	3.9	0.6	4.3	0.9
	DGP-2	$\widehat{Z}$	23.7	12.9	20.9	15.3	14.8	7.9	21.6	13.0
		$\widehat{Z}^c$	5.2	0.8	4.7	1.0	4.0	0.7	4.2	0.6
		Z	12.3	4.0	20.0	12.0	11.0	4.1	19.8	11.7
		Z^c	4.3	0.8	4.3	0.7	3.9	0.8	4.4	0.8
$\Delta = 1$	DGP-1	$\widehat{Z}$	75.4	60.0	84.7	54.7	99.2	93.9	94.1	84.8
		$\widehat{Z}^c$	74.2	57.6	86.0	57.8	98.9	91.4	94.2	85.4
		Z	89.9	69.1	87.4	62.8	97.8	91.0	96.6	87.8
		Z^c	89.9	69.1	87.4	62.8	97.8	91.0	96.6	87.8
	DGP-2	$\widehat{Z}$	44.8	25.2	50.5	20.6	80.4	56.0	81.4	65.0
		$\widehat{Z}^c$	43.8	23.9	50.3	20.4	74.3	44.6	82.5	66.7
		Z	59.1	31.4	59.2	32.7	77.2	49.5	76.7	52.4
		Z^c	59.1	31.4	59.2	32.7	77.2	49.5	76.7	52.4

The upper panel ($\Delta = 0$) in Table 1 shows the empirical sizes of the two-sample test statistics Z, Z^c, $\widehat{Z}$, and $\widehat{Z}^c$ for nominal sizes $\alpha \in \{5\%, 1\%\}$. The empirical test statistics $\widehat{Z}$ and $\widehat{Z}^c$ perform comparable to the infeasible theoretical benchmark test statistics Z and Z^c. The test statistics with finite sample correction Z^c and $\widehat{Z}^c$ show a conservative testing behavior, i.e., their empirical sizes are generally slightly smaller than the nominal sizes. However, the test statistics without finite sample correction Z and $\widehat{Z}$ drastically over-reject the null hypothesis which makes them non applicable in practice. The reason for this can be seen in Table 2, where we show the empirical variances of the test statistics Z, Z^c, $\widehat{Z}$, and $\widehat{Z}^c$ under the null hypothesis. The test statistics with finite sample correction Z^c and $\widehat{Z}^c$ have—as desired—empirical variances close to, but not greater than one. However, the test statistics without finite sample correction Z and $\widehat{Z}$ have strongly inflated variances.

TABLE 2
Empirical variances of the test statistics under the null hypothesis.

		$n = 100$		$n = 200$	
		$m = 5$	$m = 15$	$m = 5$	$m = 15$
DGP-1	$\widehat{Z}$	2.2	4.8	3.1	7.9
	$\widehat{Z}^c$	0.9	0.9	0.8	0.7
	Z	2.4	5.6	2.3	5.2
	Z^c	0.8	0.9	0.9	0.9
DGP-2	$\widehat{Z}$	2.9	8.0	2.9	4.3
	$\widehat{Z}^c$	0.8	1.2	0.9	0.9
	Z	2.0	4.1	1.9	3.8
	Z^c	0.9	0.9	0.9	1.0

The lower panel ($\Delta = 1$) in Table 1 shows the size-adjusted[1] power for the different test statistics. The size-adjusted power values for $\widehat{Z}$ and $\widehat{Z}^c$ (likewise for Z and Z^c) are all essentially equal for given sample sizes and given DGPs, since the size-adjustments correct for the differences in the variances which equalizes the originally different test-decisions. The size-adjusted power values improve significantly as n increases, but essentially not as m increases. One the one hand, this is likely due to the relatively small increase in m. On the other hand, it is known that as m becomes "large" the LLK estimator becomes independent from m and attains (tendentiously) a $\sqrt{n}$ convergence rate (see, e.g., Li and Hsing (2010) or Zhang and Wang (2016)); this property might also temper the effect of increasing m.

The computational costs for the involved bandwidths selections are very expensive and escalate as the total sample size increases—even though we

[1]Size-adjusted power is computed using the finite sample critical values from our simulations under the null hypothesis. Such size-adjustments are often used to provide a fair comparison of test statistics with and without size-distortions.

use GCV instead of cross-validation. If speed is of concern, one should consider to use simpler procedures for approximating the bias and bandwidth expressions; for instance, rule-of-thumb procedures based on our theoretical bias and bandwidth expressions.

6. Application. 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% of 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.

A look at the univariate time-series of Germany's hourly electricity spot prices in Figure 3, however, does not indicate any obvious global mean shift—only a very strong, but local, mean shift at the end of the year after Germany's (partial) nuclear phaseout which was due to a strong, but temporary, capacity constraint. Indeed, Nestle (2012), who analyzes this univariate time-series, concludes that a change of the share of nuclear power (as induced by Germany's nuclear phaseout) does not have a relevant influence on the electricity market price.

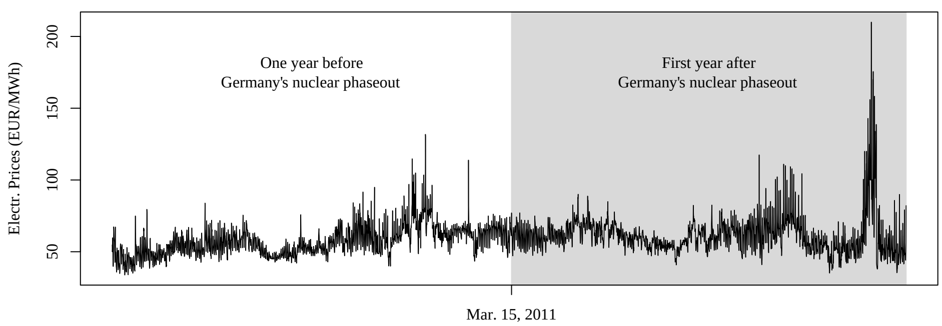

FIG 3. *Time-series of Germany's hourly electricity spot prices.*

The analysis in Nestle (2012), however, contains a severe shortcoming as it does not consider the important control variables of electricity demand and temperature. In contrast, our nonparametric statistical model in Eq. (1.1) allows us to control for the nonlinear effects of electricity demand. By considering the exogenous temperature factor we can control for seasonal differences in the electricity prices and account for the temperature dependent domains $[a(Z_i), b(Z_i)]$.

Data. The data for our analysis come from four different sources that are described in detail in the online supplement supporting this article (Liebl, 2017). The German electricity market, like many others, provides purchase guarantees for renewable energy sources (RES). Therefore, the relevant variable for pricing is electricity demand (or "load") minus electricity infeeds from RES. Additionally, we control for the Net Imports (NI) of electricity from neighboring countries. Correspondingly, in our application U_{ij} refers to *residual* electricity demand defined as $U_{ij} = \texttt{Elect.Demand}_{ij} - \texttt{RES}_{ij} + \texttt{NI}_{ij}$, where $\texttt{RES}_{ij} = \texttt{Wind.Infeed}_{ij} + \texttt{Solar.Infeed}_{ij}$ and $\texttt{NI}_{ij} = \texttt{Elect.Imports}_{ij} - \texttt{Elect.Exports}_{ij}$. The effect of further RES such as biomass is still negligible for the German electricity market. 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 set to $Y_{ij} = 200$ EUR/MWh. Such extreme prices are often referred to as "price spikes" and need to be modeled using different approaches (see, for instance, Ch. 4 in Burger et al., 2004). We analyze the daily peak-hour prices ($m = 12$) of the working days from one year before ($n_B = 242$) and one year after ($n_A = 239$) Germany's partial nuclear phase-out on March 15, 2011. The scatter plots of the two data sets are shown in Figure 2. Legal issues do not allow us to publish the data set[2], though a simulated data set that closely resembles the original data can be found in the online supplement supporting this article (Liebl, 2017).

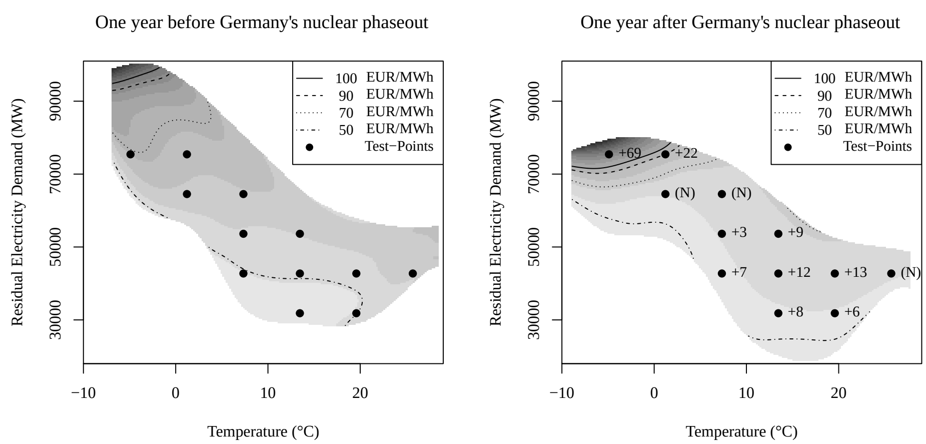

FIG 4. *Contour-plots of the estimated mean functions one year before and after Germany's nuclear phaseout. Insignificant point-wise test results are marked by "(N)"; significant results are marked by the numerical values of the point-wise differences* $\hat{\mu}_A(u, z) - \hat{\mu}_B(u, z)$.

[2]For the review-process I submit the *original* real data set.

Figure 4 shows the mean estimates one year before, $\hat{\mu}_B$, and one year after, $\hat{\mu}_A$, Germany's nuclear phaseout. The kidney-shaped supports of the mean functions $\hat{\mu}_B$ and $\hat{\mu}_A$ are due to the temperature dependent domains $[\hat{a}_B(z), \hat{b}_B(z)]$ and $[\hat{a}_A(z), \hat{b}_A(z)]$, where the empirical boundary functions $\hat{a}(.)$ and $\hat{b}(.)$ are computed using the LLK boundary estimator of Martins-Filho and Yao (2007).

In order to test the null hypothesis H_0: $\mu_A(u,z) = \mu_B(u,z)$ against the alternative H_1: $\mu_A(u,z) > \mu_B(u,z)$, we use our finite sample corrected two-sample test statistic $\widehat{Z}^c_{u,z}$ as described above. The test statistic is evaluated at regular grid-points (u_j, z_j) that cover the intersection of the supports of both mean functions. The total $G = 12$ grid-points are shown by the black-filled circle points in Figure 4. In order to account for the multiple testing we use a Bonferroni-adjusted significance level $\alpha_B = \alpha/G$ where $\alpha = 0.05$. Significant differences at the chosen grid points are depicted by the numerical values in the right plot of Figure 4; insignificant mean shifts are marked by "(N)".

The conditional mean prices for given values of electricity demand and temperature are higher in the year after Germany's nuclear phaseout than the year before. The relatively low and moderate (significant) differences from +3 to +13 EUR/MWh are observed for moderate values of electricity demand. This is in line with our expectations, since the merit-order curve is expected to be relatively flat for moderate values of electricity demand which tempers the price effects due to the nuclear phase out (see Figure 1). The large differences from +22 to +69 EUR/MWh are observed for high values of electricity demand, where the merit-order curve is expected to be very steep and, therefore, the price effects are large (see Figure 1).

Figure 4 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 ($\hat{b}_A(z) < \hat{b}_B(z)$) of the mean function $\hat{\mu}_A$—especially visible for low temperature values $-10 < z < 5$. This reduction in residual demand 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.

Acknowledgements. I would like to thank my colleague Alois Kneip for many stimulating and fruitful discussions which helped to improve this research work. Many thanks go to the student assistants of the Institute of Financial Economics and Statistics of the University of Bonn for coding assistance.

SUPPLEMENTARY MATERIAL

Supplement A: Supplement Paper
(doi: COMPLETED BY THE TYPESETTER; pdf file). This supplement paper contains the proofs of our theoretical results and a detailed description of the data sources.

Supplement B: R-Codes and Data
(doi: COMPLETED BY THE TYPESETTER; zip file). R codes of the simulation study and of the real data application. Simulated data which closely resembles the original data set.

References.

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

Benth, F. E., Kholodnyi, V. A. and Laurence, P. (2013). *Quantitative Energy Finance: Modeling, Pricing, and Hedging in Energy and Commodity Markets*, 1. ed. Springer Science & Business Media.

Bublitz, A., Keles, D. and Fichtner, W. (2017). An analysis of the decline of electricity spot prices in Europe: Who is to blame? *Energy Policy* **107** 323–336.

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

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

Cao, G., Yang, L. and Todem, D. (2012). Simultaneous inference for the mean function based on dense functional data. *Journal of Nonparametric Statistics* **24** 359–377.

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

Charnigo, R. and Srinivasan, C. (2015). A multivariate generalized Cp and surface estimation. *Biostatistics* **16** 311–325.

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

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

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*, 1. ed. *Springer Series in Statistics*. Springer.

Fogarty, C. B. and Small, D. S. (2014). Equivalence testing for functional data with an application to comparing pulmonary function devices. *The Annals of Applied Statistics* **8** 2002–2026.

Gromenko, O. and Kokoszka, P. (2012). Testing the equality of mean functions of ionospheric critical frequency curves. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **61** 715–731.

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

Hall, P., Müller, H. G. and Wang, J. L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics* **34** 1493–1517.

Hall, P. and Van Keilegom, I. (2007). Two-sample tests in functional data analysis starting from discrete data. *Statistica Sinica* **17** 1511–1531.

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

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*, 1. ed. *Springer Series in Statistics* **200**. Springer.

Horváth, L., Kokoszka, P. and Reeder, R. (2013). Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75** 103–122.

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

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

Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics* **38** 3321–3351.

Li, M., Staicu, A.-M. and Bondell, H. D. (2015). Incorporating covariates in skewed functional data models. *Biostatistics* **16** 413–426.

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

Liebl, D. (2017). Supplement to "Finite-Sample Correction for Two-Sample Inference with Sparse Covariate-Adjusted Functional Data".

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

Martins-Filho, C. and Yao, F. (2007). Nonparametric frontier estimation via local linear regression. *Journal of Econometrics* **141** 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** 152–160.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, 2. ed. *Springer Series in Statistics*. Springer.

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

Sensfuss, F., Ragwitz, M. and Genoese, M. (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** 3086–3094.

Serban, N. (2011). A space-time varying coefficient model: The equity of service accessibility. *The Annals of Applied Statistics* **5** 2024–2051.

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

Vsevolozhskaya, O., Greenwood, M. and Holodov, D. (2014). Pairwise comparison of treatment levels in functional analysis of variance with application to erythrocyte hemolysis. *The Annals of Applied Statistics* **8** 905–925.

Zhang, J.-T. and Chen, J. (2007). Statistical inferences for functional data. *The Annals of Statistics* **35** 1052–1079.

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

Zhang, W. and Wei, Y. (2015). Regression based principal component analysis for sparse functional data with applications to screening growth paths. *The Annals of Applied Statistics* **9** 597–620.

Dominik Liebl
Statistische Abteilung
University of Bonn
Adenauerallee 24-26
53113 Bonn, Germany
E-mail: dliebl@uni-bonn.de

Supplement Paper for:

Finite Sample Correction for Two-Sample Inference with Sparse covariate adjusted Functional Data

By Dominik Liebl

APPENDIX A: PROOFS

The following Lemma A.1 builds the basis of our theoretical results.

Lemma A.1 (Bias and Variance of $\hat{\mu}$). *Let (u,z) be an interior point of $[0,1]^2$. Under Assumptions A1-A5 the conditional asymptotic bias and variance of the LLK estimator $\hat{\mu}$ in Eq. (2.2) are then given by*

(i) $\text{Bias}\,\{\hat{\mu}(u,z;h_{\mu,U},h_{\mu,Z})|\mathbf{U},\mathbf{Z}\} = B_\mu(u,z) + o_p(h^2_{\mu,U}+h^2_{\mu,Z})$ *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) $\text{V}\,\{\hat{\mu}(u,z;h_{\mu,U},h_{\mu,Z})|\mathbf{U},\mathbf{Z}\} = \left(V^{\text{I}}_\mu(u,z)+V^{\text{II}}_\mu(u,z)\right)(1+o_p(1))$ *with*
$V^{\text{I}}_\mu(u,z) = (nm)^{-1}\left[h^{-1}_{\mu,U}h^{-1}_{\mu,Z}\,R(K_\mu)\frac{\gamma(u,u,z)+\sigma^2_\epsilon}{f_{UZ}(u,z)}\right]$ *and*
$V^{\text{II}}_\mu(u,z) = n^{-1}\left[\left(\frac{m-1}{m}\right)h^{-1}_{\mu,Z}\,R(\kappa)\frac{\gamma(u,u,z)}{f_Z(z)}\right]$.

Proof of Lemma A.1. Our proof of Lemma A.1 generally follows that of Ruppert and Wand (1994), and differs only from the latter reference as we consider additionally a conditioning variable Z_i, a function-valued error term, and a time-series context.

Proof of Lemma A.1, part *(i)*: For simplicity, consider a second-order kernel function κ with compact support such as the Epanechnikov kernel; this is, of course, without loss of generality. Let (u,z) be a interior point of $[0,1]^2$ and define $\mathbf{H}_\mu = \text{diag}(h^2_{\mu,U},h^2_{\mu,Z})$, $\mathbf{U} = (U_{11},\dots,U_{nm})^\top$, and $\mathbf{Z} = (Z_1,\dots,Z_n)^\top$. 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

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

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 3×1 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. (A.1). Using standard arguments from nonparametric statistics 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

$$\text{(A.2)}\quad \left((nm)^{-1}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top\mathbf{W}_{\mu,uz}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]\right)^{-1}=$$

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

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

$$\text{(A.3)}\quad \begin{aligned}&(nm)^{-1}\sum_{it}K_{\mu,h}(U_{ij}-u,Z_{\mathrm{I}}-z)(U_{ij}-u,Z_{\mathrm{I}}-z)\boldsymbol{\mathcal{H}}_\mu(u,z)(U_{ij}-u,Z_{\mathrm{I}}-z)^\top\\ &=(\nu_2(\kappa))^2\,tr\left\{\mathbf{H}_\mu\boldsymbol{\mathcal{H}}_\mu(u,z)\right\}f_{UZ}(u,z)+o_p(tr(\mathbf{H}_\mu))\end{aligned}$$

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

(A.4)
$$(nm)^{-1} \sum_{it} \left\{ K_{\mu,h}(U_{ij} - u, Z_{\mathrm{I}} - z)(U_{ij} - u, Z_{\mathrm{I}} - z)\boldsymbol{\mathcal{H}}_{\mu}(u, z)(U_{ij} - u, Z_i - z)^{\top} \right\} \times$$
$$\times (U_{ij} - u, Z_{\mathrm{I}} - z)^{\top} = O_p(\mathbf{H}_{\mu}^{3/2}\mathbf{1}).$$

Plugging the approximations of Eqs. (A.2)-(A.4) into the first summand of the conditional bias expression in Eq. (A.1) leads to the following expression

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

Furthermore, it is easily seen that the second summand of the conditional bias expression in Eq. (A.1), which holds the remainder term, is given by

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

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

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

Proof of Lemma A.1, part *(ii)*: In the following we derive the conditional variance of the local linear estimator $\mathrm{V}(\hat{\mu}(u, z; \mathbf{H}_{\mu}) | \mathbf{U}, \mathbf{Z}) =$

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

(A.5)
$$\begin{aligned}
=&e_1^{\top} ((nm)^{-1} [\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]^{\top} \mathbf{W}_{\mu,uz} [\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z])^{-1} \times \\
&\times ((nm)^{-2} [\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]^{\top} \mathbf{W}_{\mu,uz} \,\mathrm{Cov}(\mathbf{Y} | \mathbf{U}, \mathbf{Z})\, \mathbf{W}_{\mu,uz} [\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]) \times \\
&\times ((nm)^{-1} [\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]^{\top} \mathbf{W}_{\mu,uz} [\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z])^{-1} e_1,
\end{aligned}$$

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

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

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 A.2-A.4.

Lemma A.2. *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 $c(u,z) = 2\sum_{l=1}^{n-1} \gamma_l((u,z),(u,z))$. *Under Assumption A4 there exists a constant* C, $0 < C < \infty$, *such that* $0 \le |c(u,z)| \le C$.

Lemma A.3. *The* 1×2 *dimensional upper-right block of the matrix* $(nm)^{-2}[\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz}\operatorname{Cov}(\mathbf{Y}|\mathbf{U}, \mathbf{Z})\mathbf{W}_{\mu,uz}[\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]$ *is given by*

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

where $c(u,z) = 2\sum_{l=1}^{n-1} \gamma_l((u,z),(u,z))$. *Under Assumption A4 there exists a constant* C, $0 < C < \infty$, *such that* $0 \le |c(u,z)| \le C$.

Remark. *The* 2×1 *dimensional lower-left block of the matrix* $(nm)^{-2}[\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,uz}\operatorname{Cov}(\mathbf{Y}|\mathbf{U}, \mathbf{Z})\mathbf{W}_{\mu,uz}[\mathbf{1}, \mathbf{U}_u, \mathbf{Z}_z]$ *is simply the transposed version of the result in Lemma A.3.*

LEMMA A.4. *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*

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

where $c(u,z) = 2\sum_{l=1}^{n-1}\gamma_l((u,z),(u,z))$. Under Assumption A4 there exists a constant C, $0 < C < \infty$, such that $0 \le |c(u,z)| \le C$.

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 A.2-A.4, 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 (A.2), we can approximate the conditional variance of the bivariate local linear estimator, given in (A.5). Some straightforward matrix algebra leads to $\mathrm{V}(\hat{\mu}(u,z;\mathbf{H}_\mu)|\mathbf{U},\mathbf{Z}) =$

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

which is asymptotically equivalent to our variance statement of Lemma A.1 part *(ii)*.

Proof of Lemma A.2. (The proofs of Lemmas A.3 and A.4 can be done correspondingly.) To show Lemma A.2 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

(A.6)

$$
\begin{aligned}
s_1 &= (nm)^{-2}\sum_{it}(K_{\mu,h}(U_{ij}-u, Z_{\mathrm{I}}-z))^2\,\mathrm{V}(Y_{ij}|\mathbf{U},\mathbf{Z}) \\
&= (nm)^{-1}|\mathbf{H}_\mu|^{-1/2}f_{UZ}(u,z)R(K_\mu)\left(\gamma(u,u,z)+\sigma_\epsilon^2\right) + O((nm)^{-1}|\mathbf{H}_\mu|^{-1/2}\,tr(\mathbf{H}_\mu^{1/2}))
\end{aligned}
$$

(A.7)

$$
\begin{aligned}
s_2 &= (nm)^{-2}\sum_{jk}\sum_{\substack{i\ell\\ i\neq\ell}}K_{\mu,h}(U_{ij}-u, Z_{\mathrm{I}}-z)\,\mathrm{Cov}(Y_{ij},Y_{\ell k}|\mathbf{U},\mathbf{Z})\,K_{\mu,h}(U_{\ell k}-x, Z_\ell - z) \\
&= n^{-1}(f_{UZ}(u,z))^2c(u,z) + O_p(n^{-1}tr(\mathbf{H}_\mu^{1/2}))
\end{aligned}
$$

(A.8)

$$s_3 = (nm)^{-2}\sum_{\substack{ij\\ i\neq j}}\sum_{t} h_{\mu,U}^{-1}\kappa(h_{\mu,U}^{-1}(U_{ij}-u))(h_{\mu,Z}^{-1}\kappa(h_{\mu,Z}^{-1}(Z_{\mathrm{I}}-z)))^2\,\mathrm{Cov}(Y_{ij},Y_{jt}|\mathbf{U},\mathbf{Z})\times$$

$$\times\, h_{\mu,U}^{-1}\kappa(h_{\mu,U}^{-1}(U_{ik}-x))$$

$$= n^{-1}(f_{UZ}(u,z))^2\left[\left(\frac{m-1}{m}\right)h_{\mu,Z}^{-1}R(\kappa)\frac{\gamma(u,u,z)}{f_Z(z)}\right] + O_p(n^{-1}tr(\mathbf{H}_\mu^{1/2})),$$

where $c(u,z) = 2\sum_{l=1}^{n-1}\gamma_l((u,z),(u,z))$. Summing up (A.6)-(A.7) leads to the result in Lemma A.2. Lemmas A.3 and A.4 differ from Lemma A.2 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 $(U_{ij}-u, Z_i-z)$.

Proofs of Theorem 3.1 and Corollary 3.1. Theorem 3.1 and Corollary 3.1 follow directly from Lemma A.1 and from applying a central limit theorem for strictly stationary ergodic times series such as Theorem 9.5.5 in Karlin and Taylor (1975).

Proof of Theorem 3.2. Under the assumptions of the Theorem 3.2, the leading variance term is given by V_μ^{I} and the variance term V_μ^{I} is negligible. This leads to the following AMISE function for the local linear estimator $\hat{\mu}$:

(A.9)
$$\mathrm{AMISE}_{\hat{\mu}}\,(h_{\mu,U},h_{\mu,Z}) = (nm)^{-1}\,h_{\mu,U}^{-1}\,h_{\mu,Z}^{-1}\,R(K_\mu)\,Q_\mu +$$
$$+\,\frac{1}{4}\,(\nu_2(K_\mu))^2\,\left[h_{\mu,U}^4\,\mathcal{I}_{UU}^\mu + 2\,h_{\mu,U}^2\,h_{\mu,Z}^2\,\mathcal{I}_{UZ}^\mu + h_{\mu,Z}^4\,\mathcal{I}_{ZZ}^\mu\right],$$

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

This corresponds to the well 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) and follows from the formulas in Wand and Jones (1994). Minimizing the above AMISE function with respect to $h_{\mu,U}$ and $h_{\mu,Z}$ leads to the optimal bandwidth expressions in Theorem 3.2 which are essentially equivalent to those in Herrmann et al. (1995).

Remark. The requirement that $m/n^{1/5}\to\infty$ in Theorem 3.1, Corollary 3.1, and Theorem 3.2 guarantees that the first variance term V_μ^{I} asymptotically dominates the second variance term V_μ^{II}, under the assumption of optimal bandwidth rates; i.e., that the following order relation holds:

(A.10)
$$n^{-1}\,h_{\mu,Z}^{-1} = o\left(n^{-(1+\theta)}\,h_{\mu,U}^{-1}\,h_{\mu,Z}^{-1}\right),$$

where we used that by Assumption A1 $nm \asymp n^{1+\theta}$. Plugging the AMISE optimal bandwidth rates $(nm)^{-1/6}$ (Theorem 3.2) into the order relation of Eq. (A.10) leads to the corresponding θ values of $0 \leq \theta < 1/5$ which imply that $m \to \infty$ and $n \to \infty$ such that $m/n^{1/5} \to 0$.

APPENDIX B: DATA 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).

REFERENCES

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

Karlin, S. and Taylor, H. M. (1975). *A First Course in Stochastic Processes*, 2. ed. Academic Press.

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

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

Dominik Liebl
Statistische Abteilung
University of Bonn
Adenauerallee 24-26
53113 Bonn, Germany
E-mail: dliebl@uni-bonn.de