The Annals of Applied Statistics
2019, Vol. 13, No. 2, 1128–1146
https://doi.org/10.1214/18-AOAS1230

NONPARAMETRIC TESTING FOR DIFFERENCES IN ELECTRICITY PRICES: THE CASE OF THE FUKUSHIMA NUCLEAR ACCIDENT

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. Taking into account the nature of the data and the auction design of the electricity market, we approach this problem using a Local Linear Kernel (LLK) estimator for the nonparametric mean function of sparse covariate-adjusted functional data. We build upon recent theoretical work on the LLK estimator and propose a two-sample test statistics using a finite sample correction to avoid size distortions. Our nonparametric test results on the price differences point to a Simpson's paradox explaining an unexpected result recently reported in the literature.

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. So far, however, empirical studies are scarce and based on restrictive model assumptions. In this work we add a nonparametric functional data perspective and compare our test results with the existing benchmark results. Our results point to a Simpson's paradox explaining the unexpected result recently reported by Grossi, Heim and Waterson (2017).

Pricing at electricity exchanges is explained well by the merit-order model. This model assumes that spot prices are based on the 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), Chapter 4) and is most important for the explanation of electricity spot prices in the literature on energy economics (see Bublitz, Keles and Fichtner (2017), Burger et al. (2004), Cludius et al. (2014), Grossi, Heim and Waterson (2017), Hirth (2013), Liebl (2013), Sensfuß, Ragwitz and Genoese (2008)).

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 demand curve

Received August 2017; revised November 2018.
Key words and phrases. Electricity spot prices, functional data analysis, local linear kernel estimation, nuclear power phaseout, sparse functional data, time series analysis.

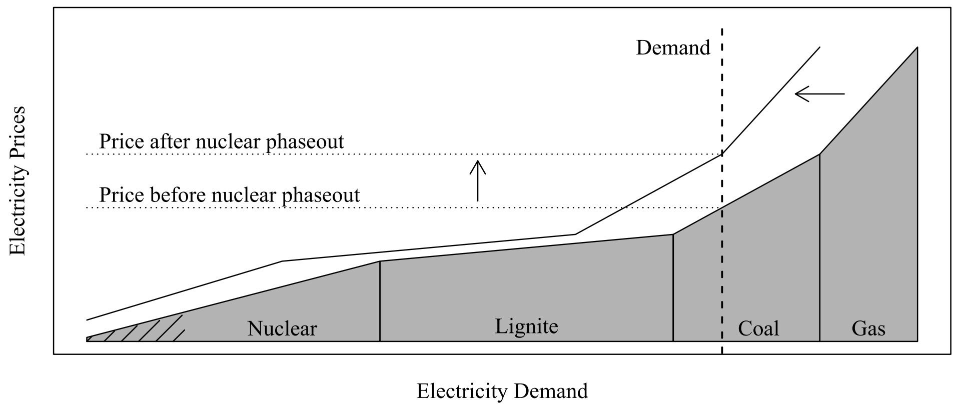

FIG. 1. *Sketch of the merit-order curve and the theoretical price effect of the nuclear power phase-out. The dashed region signifies the proportion of phased out nuclear power plants.*

(dashed line) with the merit-order curve determines the electricity prices, where 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 (2014), Chapter 6). Following the literature, we differentiate between "peak-hours" (from 9 am to 8 pm) 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. However, we do not directly observe the price functions X_i, but only their noisy discretization points (Y_{ij}, U_{ij}), with $j = 1, \dots, m = 12$ (see black points in Figure 2). This data situation with only a few, that is, $m = 12$, irregularly spaced evaluation points, $U_{i1}, \dots, U_{im}$, per function is referred to as sparse functional data (see, e.g., Yao, Müller and Wang (2005)). The smoothness of the underlying price curve X_i induces a high correlation between electricity prices Y_{ij} and Y_{ik} with similar values of electricity demand $U_{ij} \approx U_{ik}$. Ignoring these correlations when doing inference can result in serious size distortions and invalid test decisions (see Liebl (2019a)).

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 $U_{ij} \in \mathbb{R}$,

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

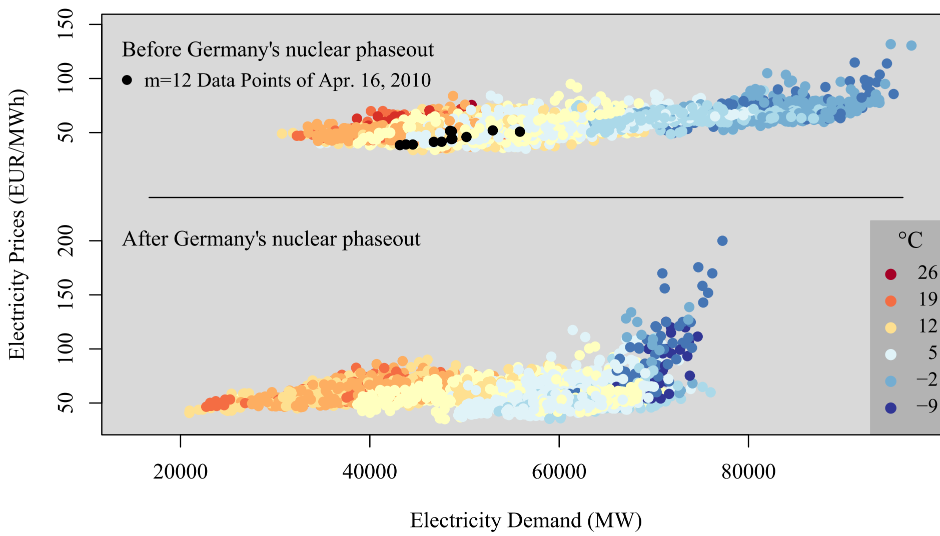

FIG. 2. *Scatter plots of the price-demand data pairs* (Y_{ij}, U_{ij}) *and the the additional covariate of daily mean air temperature* Z_i *(measured in* $^\circ C$*). The upper panel shows the data from one year before Germany's partial nuclear phaseout, that is, from March* 15, 2010 *to March* 14, 2011; *the lower panel shows the data from one year after, that is, from March* 15, 2011 *to March* 14, 2012.

where $Z_i \in \mathbb{R}$ denotes the daily mean air temperature—a covariate of fundamental importance for the shape and the location of the random function $X_i(\cdot, Z_i)$. The statistical error term ϵ_{ij} is assumed to be independently and identically distributed (i.i.d.) with mean zero and finite variance and assumed to be independent from X_i, U_{ij}, and Z_i.

The shape of $X_i(\cdot, Z_i)$ and its location, that is, $X_i(\cdot, Z_i) \in L^2[a(Z_i), b(Z_i)]$, with $[a(Z_i), b(Z_i)] \subset \mathbb{R}$, are both allowed to be functions of the covariate temperature Z_i. The scatter plot of the data triplets (Y_{ij}, U_{ij}, Z_i) is shown in Figure 2. Note that electricity demand U_{ij} is observed within temperature-specific subintervals, that is, $U_{ij} \in [a(Z_i), b(Z_i)]$ and that the discretization points (Y_{ij}, U_{ij}) suggest steeper functions $X_i(\cdot, Z_i)$ for cold days than for warm days. These observations motivate our modeling assumption that $X_i(\cdot, Z_i) \in L^2[a(Z_i), b(Z_i)]$; see Horváth and Kokoszka (2012), Chapter 2, for fundamental properties of random functions in the space L^2.

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 in the data for very cold days (see Figure 2), though not obvious on other days. Therefore, the objective of this article is a two-sample test of the pointwise null hypothesis of equal means against the alternative of larger mean values after Germany's nuclear phaseout, that is,

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

where $\mu_A(u, z) = \mathbb{E}(X_i^A(u, z))$ and $\mu_B(u, z) = \mathbb{E}(X_i^B(u, z))$ are the mean functions of the random price functions *A*fter, $X_i^A(\cdot, z)$, and *B*efore, $X_i^B(\cdot, z)$, 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 for sparse functional data suggested by Jiang and Wang (2010). Recently, it has been demonstrated that the asymptotic results of Jiang and Wang (2010) neglect an additional functional-data-specific variance term which is asymptotically negligible, but typically not negligible in practice. Neglecting this additional variance term leads to a too small variance component resulting in size distortions of the test statistic and in invalid test decisions (Liebl (2019a)). Therefore, we take into account the small sample correction proposed by Liebl (2019a) and propose a two-sample test statistic, which guarantees valid test decisions in practical finite samples as well (see also the in-depth simulation study in Liebl (2019a)).

In order to control for the effects of the remaining fundamental market factors, that is, the price of natural gas, CO_2 emission allowances, and coal (see, e.g., Maciejowska and Weron (2016)) we use what is called an event study approach. Event studies are essentially two-sample test problems comparing a "control" sample, from the period just before the event, with a "treatment" sample, from the period just after the event. The time period before the event is called "estimation window" and the time period after the event is called "event window". The idea is to choose small enough time windows for which the noncontrolled, but less important, market factors do not have confounding effects (McWilliams and Siegel (1997)). The event study method used today was introduced by Ball and Brown (1968) and Fama et al. (1969). A well-known introductory survey article is written by MacKinlay (1997). We are not the first to use the event study approach for analyzing effects that are due to Germany's unexpected nuclear phaseout after the Fukushima Daiichi nuclear disaster. Ferstl, Utz and Wimmer (2012) analyze the stock prices of energy companies using an event study. Betzer, Doumet and Rinne (2013) use an event study to test for differences in the shareholder wealth of German nuclear energy companies. Thoenes (2014) considers the effect on futures prices.

In terms of the research question, our paper is closely related to the recent work of Grossi, Heim and Waterson (2017), who consider differences in the electricity spot prices before and after Germany's nuclear phaseout. Therefore, we use the approach of Grossi, Heim and Waterson (2017) as a benchmark for comparing our nonparametric test approach. While Grossi, Heim and Waterson (2017) estimate the price differences conditionally on demand, we additionally allow for an interaction effect with temperature. This way our nonparametric test result demonstrates that a Simpson's paradox (Wagner (1982)) can explain the unexpected structure of the parametric price differences reported by Grossi, Heim and Waterson (2017).

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 pointwise 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 (see, for instance, Benko, Härdle and Kneip (2009), Cao, Yang and Todem (2012), Fogarty and Small (2014), Gromenko and Kokoszka (2012), Hall and Van Keilegom (2007), Horváth, Kokoszka and Reeder (2013), Vsevolozhskaya, Greenwood and Holodov (2014), Zhang and Chen (2007), Zhang and Wang (2016)). 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; however, they do not focus on a sparse functional data context. Two recent works on modeling and forecasting electricity data using methods from functional data analysis are Shah and Lisi (2018) and Lisi and Shah (2018). Readers with a general interest in functional data analysis are referred to the textbooks of Ferraty and Vieu (2006), Horváth and Kokoszka (2012), Ramsay and Silverman (2005), and Hsing and Eubank (2015).

The rest of the paper is organized as follows. The next section introduces the LLK estimator and our two-sample test statistic. Section 3 introduces approximations to the unknown bias, variance and bandwidth components. Section 4 contains the real data study. The paper concludes with a discussion in Section 5. The proofs of our theoretical results are based on standard arguments in nonparametric statistics and can be found in the Supplementary Material (Liebl (2019b)).

2. Nonparametric two-sample inference. 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 $X_i(\cdot, Z_i) \in L^2([0,1])$. The standardization can be achieved as $U_{ij}^{\text{new}} = (U_{ij}^{\text{orig}} - a(Z_i^{\text{orig}}))/(b(Z_i^{\text{orig}}) - a(Z_i^{\text{orig}}))$ and $Z_i^{\text{new}} = (Z_i^{\text{orig}} - \min_{1\le i\le n}(Z_i^{\text{orig}}))/(\max_{1\le i\le n}(Z_i^{\text{orig}}) - \min_{1\le i\le n}(Z_i^{\text{orig}}))$. The functional interval borders $a(\cdot)$ and $b(\cdot)$ 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 rewritten 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 $\mathbf{Z} = (Z_1, \dots, Z_n)^\top$, that is,

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

where $X_i^c(\cdot, z)$, U_{ij}, and Z_i are assumed to be a stationary weakly dependent functional and univariate time series. The error term ϵ_{ij} is a classical i.i.d. error term with mean zero, finite variance $\mathrm{V}(\epsilon_{ij}) = \sigma_\epsilon^2$, 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

$$\begin{aligned}
&\operatorname{Corr}\big(X_i^c(U_{ij}, Z_i), X_i^c(U_{ik}, Z_i)|U_{ij} = u_1, U_{ik} = u_2, Z_i = z\big) \\
&\quad = \operatorname{Corr}\big(X_i^c(u_1, z), X_i^c(u_2, z)\big) \approx 1 \qquad \text{for } u_1 \approx u_2,
\end{aligned}$$

which leads to the above mentioned functional-data-specific variance term that makes it necessary to use the finite sample correction proposed by (Liebl (2019a)).

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:

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

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, for example, the Epanechnikov or the Gaussian kernel, $h_{\mu,U}$ denotes the bandwidth in U direction, and $h_{\mu,Z}$ the bandwidth in Z direction. 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 $\mathbf{Y} = (Y_{11}, Y_{12}, \dots, Y_{n,m-1}, Y_{nm})^\top$.

The following theorem is an adjusted version of Corollary 3.1, part (b) in Liebl (2019a) and takes into account the finite sample correction for the variance component:

THEOREM 2.1 (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 $h_{\mu,U} \asymp h_{\mu,Z} \asymp$*

$(nm)^{-1/6}$. *Then the LLK estimator* $\hat{\mu}(u,z)$ *in equation* (2.2) *is asymptotically normal*, *that is*,

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

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

$$B_\mu(u,z;h_{\mu,U},h_{\mu,Z}) = \frac{1}{2}\nu_2(K_\mu)\big(h_{\mu,U}^2\mu^{(2,0)}(u,z) + h_{\mu,Z}^2\mu^{(0,2)}(u,z)\big),$$

$$V_\mu^{\mathrm{I}}(u,z;h_{\mu,U},h_{\mu,Z}) = \frac{1}{nm}\left[\frac{R(K_\mu)}{h_{\mu,U}h_{\mu,Z}}\frac{\gamma(u,u,z)+\sigma_\epsilon^2}{f_{UZ}(u,z)}\right],$$

$$V_\mu^{\mathrm{II}}(u,z;h_{\mu,Z}) = \frac{1}{n}\left[\left(\frac{m-1}{m}\right)\frac{R(\kappa)}{h_{\mu,Z}}\frac{\gamma(u,u,z)}{f_Z(z)}\right], \quad \textit{and}$$

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

This theorem is valid under Assumptions A1–A5, *which are listed in the Supplementary Material* (*Liebl* (2019b)).

Theorem 2.1 generalizes the corresponding result in Liebl (2019a) by additionally allowing for a time series context with weakly dependent auto-correlation structure; a proof can be found in the Supplementary Material (Liebl (2019b)). The theorem implies the standard optimal convergence rate $(nm^{-1/3})$ for bivariate LLK estimators. The finite sample correction is accomplished by the additional second variance term $V_\mu^{\mathrm{II}}(u,z;h_{\mu,Z})$ which could be dropped from a pure asymptotic view. However, this second variance term is typically not negligible in practice and serves as a very effective finite sample correction (Liebl (2019a)). 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 decorrelation effect is referred to as the "whitening window" property (see, for instance, Fan and Yao (2003), Chapter 5.3).

Theorem 2.1 without the variance term $V_\mu^{\mathrm{II}}(u,z;h_{\mu,Z})$, that is, without finite sample correction, is essentially equivalent to Theorem 3.2 in Jiang and Wang (2010), who, however, consider the case where m is bounded, that is, $1 < m \le c$ for some small $c < \infty$ (e.g., $c = 4$ or $c = 5$), as typically assumed in the literature on sparse functional data analysis (cf. Yao, Müller and Wang (2005), Zhang and Wang (2016), and many others). In contrast, our asymptotic normality result allows for bounded $m < \infty$ as well as nonbounded $m \to \infty$, such that $m/n^{1/5} \to 0$. This explains the empirical finding in Liebl (2019a) that our asymptotic normality result with finite sample correction provides very good finite sample approximations for practical small-m (e.g., $m = 5$) as well as moderate-m cases (e.g., $m = 15$)—by contrast to the corresponding results in Jiang and Wang (2010).

The following corollary follows directly from Theorem 2.1 and contains the asymptotic normality result for our two-sample test statistic:

COROLLARY 2.1 (Two-sample test statistic). *Under the same conditions as in Theorem 2.1 and under the null hypothesis H_0: $\mu_A(u,z) = \mu_B(u,z)$, the following two two-sample test statistic is asymptotically normal*:

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

where the dependencies on the bandwidth parameters $h_{\mu,U}$ and $h_{\mu,Z}$ are suppressed for readability reasons.

The test statistic $Z_{u,z}$ is infeasible as it depends on the unknown bias, variance, and bandwidth expressions, B_μ, V^{I}_μ, V^{II}_μ, $h_{\mu,U,}$ and $h_{\mu,Z}$. In our application, we use the practical rule-of-thumb bandwidth, bias and variance approximations as described in the following section.

REMARK. A common number m (m: number of discretization points per function X_i) may be unrealistic for some applications. However, our estimators can be directly applied to data-scenarios with function-specific numbers of discretization points m_i with $i = 1, \dots, n$. To adjust our asymptotic analysis for this situation, one can consider the case where $m \to \infty$ with $m \le m_i$ for all $i = 1, \dots, n$ (cf. Zhang and Chen (2007)). As Hall, Müller and Wang (2006), Zhang and Chen (2007), and Zhang and Wang (2016) we do not consider random numbers m_i, but if m_i are random, our theory can be considered as conditional on m_i.

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

$$\begin{aligned}&\hat{B}_\mu(u,z;h_{\mu,U},h_{\mu,Z})\\ &\quad = \frac{\nu_2(K_\mu)}{2}\big(h^2_{\mu,U}\hat{\mu}^{(2,0)}(u,z;g_{\mu,U},g_{\mu,Z}) + h^2_{\mu,Z}\hat{\mu}^{(0,2)}(u,z;g_{\mu,U},g_{\mu,Z})\big),\end{aligned} \tag{3.1}$$

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,

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

with $e_3^\top = (0,0,1,0,0,0,0)$, $\mathbf{U}^{1:3}_u = [\mathbf{U}_u, \mathbf{U}^2_u, \mathbf{U}^3_u]$, $\mathbf{Z}^{1:3}_z = [\mathbf{Z}_z, \mathbf{Z}^2_z, \mathbf{Z}^3_z]$, and diagonal matrix $\mathbf{W}_{\mu,uz}$ with weights $g^{-1}_{\mu,U}\kappa(g^{-1}_{\mu,U}(U_{ij}-u))g^{-1}_{\mu,Z}\kappa(g^{-1}_{\mu,Z}(Z_i - z))$ on its diagonal, where $g_{\mu,U}$ and $g_{\mu,Z}$ are the bandwidths in U and Z direction. 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

$$\hat{V}_\mu^{\mathrm{I}}(u,z;h_{\mu,U},h_{\mu,Z},h_{\gamma,U},h_{\gamma,Z}) = \frac{1}{nm}\left[\frac{R(K_\mu)}{h_{\mu,U}h_{\mu,Z}}\frac{\hat{\gamma}^{\mathrm{ND}}(u,u,z;h_{\gamma,U},h_{\gamma,Z})}{\hat{f}_{UZ}(u,z)}\right]. \tag{3.2}$$

The Noisy Diagonal (ND) 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 follows:

$$\hat{\gamma}^{\mathrm{ND}}(u,u,z;h_{\gamma,U},h_{\gamma,Z}) = e_1^\top([\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top\mathbf{W}_{\gamma,uz}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z])^{-1}[\mathbf{1},\mathbf{U}_u,\mathbf{Z}_z]^\top\mathbf{W}_{\gamma,uz}\hat{\mathbf{C}}^{\mathrm{ND}}, \tag{3.3}$$

with $h_{\gamma,U}$ and $h_{\gamma,Z}$ denoting the bandwidths in U and Z direction and with $\hat{\mathbf{C}}^{\mathrm{ND}} = (\hat{C}_{111},\dots,\hat{C}_{ijj},\dots,\hat{C}_{nmm})^\top$ consisting only of the noisy diagonal raw-covariances $\hat{C}_{ijj}^{\mathrm{ND}} = (Y_{ij} - \hat{\mu}(U_{ij},Z_i;h_{\mu,U},h_{\mu,Z}))^2$ for which $\mathbb{E}(\hat{C}_{ijj}|\mathbf{U},\mathbf{Z}) \approx \gamma(U_{ij},U_{ik},Z_i)+\sigma_\epsilon^2$. Note that $\hat{\gamma}^{\mathrm{ND}}$ is equivalent to the LLK estimator "$\hat{V}$" in Jiang and Wang (2010). The estimate $\hat{f}_{UZ}(u,z)$ is computed using the bivariate kernel density estimation function `kde2d()` of the R-package `MASS` (Venables and Ripley (2002)), where the involved bandwidths are selected using the R-function `width.SJ()` of the R-package `MASS` containing an implementation of the method of Sheather and Jones (1991).

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

$$\hat{V}_\mu^{\mathrm{II}}(u,z;h_{\mu,Z},h_{\gamma,U},h_{\gamma,Z}) = \frac{1}{n}\left[\left(\frac{m-1}{m}\right)\frac{R(\kappa)}{h_{\mu,Z}}\frac{\hat{\gamma}(u,u,z;h_{\gamma,U},h_{\gamma,Z})}{\hat{f}_Z(z)}\right]. \tag{3.4}$$

The estimate $\hat{f}_Z(z)$ is computed using the R-function `density()` for univariate kernel density estimation, where the involved bandwidth is selected using the R-function `width.SJ()` of the R-package MASS. The LLK estimator $\hat{\gamma}(u_1,u_2,z;h_{\gamma,U},h_{\gamma,Z})$ of $\gamma(u_1,u_2,z)$ is defined as follows:

$$\begin{aligned}\hat{\gamma}(u_1,u_2,z;h_{\gamma,U},h_{\gamma,Z}) &= e_1^\top([\mathbf{1},\mathbf{U}_{u_1},\mathbf{U}_{u_2},\mathbf{Z}_z]^\top\mathbf{W}_{\gamma,u_1u_2z}[\mathbf{1},\mathbf{U}_{u_1},\mathbf{U}_{u_2},\mathbf{Z}_z])^{-1}\\ &\quad\times[\mathbf{1},\mathbf{U}_{u_1},\mathbf{U}_{u_2},\mathbf{Z}_z]^\top\mathbf{W}_{\gamma,u_1u_2z}\hat{\mathbf{C}}.\end{aligned} \tag{3.5}$$

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 $\hat{\mathbf{C}} = (\hat{C}_{112}, \dots, \hat{C}_{ijk}, \dots, \hat{C}_{nm,m-1})^\top$ consisting only of the off-diagonal raw-covariances

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

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

4. 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. Energy economists typically use Monte Carlo simulations in order to approximate the price effect of Germany's nuclear phaseout (see, for instance, Bruninx et al. (2013)). Empirical data-based evidence, however, is scarce. Thoenes (2014) uses an event study approach to estimate the effect of Germany's nuclear phaseout on electricity futures prices. The work of Grossi, Heim and Waterson (2017) considers the less speculative electricity spot prices and can be seen as a parametric counterpart to our work. Therefore, we use the approach of Grossi, Heim and Waterson (2017) as a benchmark for our nonparametric approach.

In Section 4.1 we introduce the benchmark models. In Section 4.2 we compare the benchmark results with the results of our nonparametric two-sample test statistic and demonstrate that a Simpson's paradox can explain the unexpected finding in Grossi, Heim and Waterson (2017).

Data. The data for our analysis come from different sources that are described in detail in the Supplementary Material (Liebl (2019b)). The German electricity market, like many others, provides purchase guarantees for renewable energy sources. Therefore, the relevant variable for pricing is electricity demand (or "load") minus electricity infeeds from RES and an additional correction for the net imports of electricity from neighboring countries (see, e.g., Paraschiv, Erni and Pietsch (2014)). Correspondingly, in our application U_{ij} refers to *residual* electricity demand defined as $U_{ij} = D_{ij} - R_{ij} + N_{ij}$, with $R_{ij} = W_{ij} + S_{ij}$ and $N_{ij} = I_{ij} - E_{ij}$, where D_{ij} denotes electricity demand, R_{ij} denotes infeeds from renewable energy sources, W_{ij} denotes wind-power, S_{ij} denotes solar-power, N_{ij} denotes net-imports, I_{ij} denotes electricity imports, and

E_{ij} denotes electricity exports. The effect of further renewable energy sources 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"; they are caused by market speculations involving potential capacity scarcities and need to be modeled using different approaches (see, for instance, Burger et al. (2004), Chapter 4). Our data set consists of the peak-hour prices ($m = 12$; from 9 am to 8 pm) of the working days from one year before ($n_B = 242$) and one year after ($n_A = 239$) Germany's partial nuclear phaseout on March 15, 2011. We consider only working days, since for weekends there are different compositions of the power plant portfolio. The same reasoning applies to holidays and so-called Brückentage, which are extra days off that bridge single working days between a bank holiday and the weekend. Therefore, we remove also all holidays and Brückentage from the data. The temporal gaps in the data due to weekend days, holidays, and Brückentage do not violate our theoretical assumptions on the auto-covariance structure, since such gaps do not increase the auto-correlations. The main fundamental market factors (i.e., temperature, gas, CO_2 allowance, and coal prices) are available at a daily sampling scheme. Legal issues do not allow us to publish the original data sets, however, simulated data sets that closely resemble the original data can be found in the online Supplementary Material (Liebl (2019b)).

4.1. *Parametric benchmark models.* As a benchmark case study for our non-parametric approach, we use the following two increasingly complex parametric regression models:

$$Y_i = \alpha_1 + \alpha_2 d_i + \alpha_3 Z_i + \sum_{k=1}^{K} \beta_k \mathfrak{X}_{ik} + \epsilon_i, \tag{4.1}$$

$$\begin{aligned} Y_i = {} & \alpha_1 + \alpha_2 U_i + \alpha_3 U_i^2 + d_i\left(\alpha_4 + \alpha_5 U_i + \alpha_6 U_i^2\right) \\ & + \alpha_7 Z_i + \sum_{k=1}^{K} \beta_k \mathfrak{X}_{ik} + \epsilon_i, \end{aligned} \tag{4.2}$$

where $Y_i = 12^{-1} \sum_{j=1}^{12} Y_{ij}$ denotes the daily mean (peak-hours) electricity spot price, $U_i = 12^{-1} \sum_{j=1}^{12} U_{ij}$ denotes the daily mean (peak-hours) Residual Demand (RD), d_i is a dummy variable which equals zero for all time points before Germany's nuclear phaseout and equals one for all other time points, Z_i denotes air temperature, $\mathfrak{X}_{ik}$ contains further control variables, and ϵ_i is a classical Gaussian statistical error term. As control variables $\mathfrak{X}_{ik}$ we use the same fundamental market factors as in the fundamental electricity market model of Maciejowska and Weron (2016), that is, temperature, CO_2 emission allowance prices, coal prices, and natu-

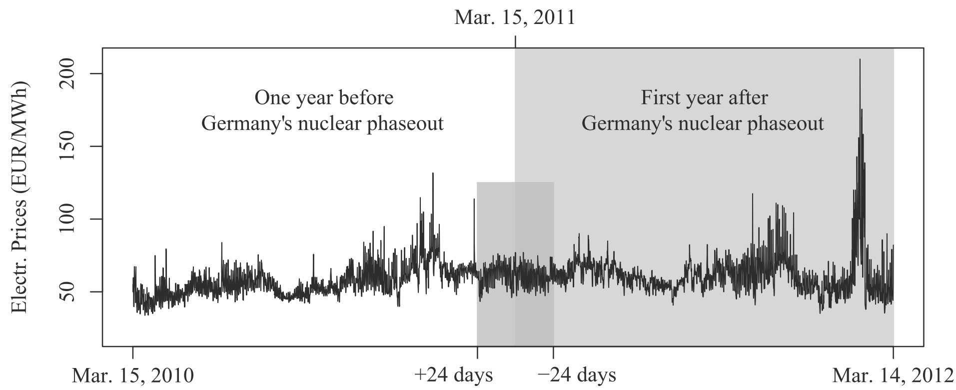

FIG. 3. *Time series of Germany's hourly electricity spot prices.*

ral gas prices. For lignite and nuclear energy resources there are no relevant market prices (Grossi, Heim and Waterson (2017)).

Model (4.1) estimates a possible price effect using a simple dummy variable approach. Model (4.2) corresponds to the regression model of Grossi, Heim and Waterson (2017), who estimate the price effect of Germany's nuclear phaseout using the quadratic dummy-interaction function $d_i(\alpha_4 + \alpha_5 x_i + \alpha_6 x_i^2)$; see Table 4 in Grossi, Heim and Waterson (2017). We cannot consider all of the control variables proposed by Grossi, Heim and Waterson (2017), since we do not have access to the data used in their case study (missing variables: export congestion index, residual supply index (i.e., a market power index), and low/high river level). In order to rectify this shortcoming, we use the event study approach with short estimation and event windows each containing 24 days (see Figure 3). Our window size is one day smaller than the window size in the related event study of Thoenes (2014), since we want to exclude the strong price jump one day before our estimation window. Our estimation results are robust to smaller window sizes.

The F statistics in the lower panel of Table 1 show that the simple benchmark Model (4.1) is insufficient for describing the price effect of Germany's nuclear phaseout, expect if the temperature-component is added to the control variables (III and IV). In contrast to this, Model (4.2), which contains the quadratic dummy-interaction function, describes the price effect with a significantly strong explanatory power in all model-specifications I-IV. Comparing the different model-specifications reveals that the resource prices (CO_2, coal, and gas) are jointly insignificant control variables: neither adding them to the smallest model-specification (I vs II), nor removing them from the larges model-specification (III vs IV) has a significant impact on the fit of the models. That is, our event study approach successfully minimizes the confounding effects of the control variables CO_2, coal, and gas, which are known to be less important in the short run. Furthermore, it demonstrates the importance of including the temperature-component.

TABLE 1
Event study estimation results for the parametric benchmark Models (4.1) and (4.2)

	Model (4.1)				Model (4.2)			
	I	II	III	IV	I	II	III	IV
RD					0.0	0.0	0.0	0.0
RD^2					−0.0	−0.0	−0.0	−0.0
Dummy	−1.1	1.9	5.7*	2.6*	63.3	88.2	61.3	42.1
Dummy × RD					−0.0	−0.0	−0.0	−0.0
Dummy × RD^2					0.0	0.0	0.0	0.0
Temperature			−1.3*	−1.2*			−0.8*	−0.8*
$Temperature^2$			0.1*	0.0			0.0	0.0*
CO_2 price		0.4	−1.4			2.2	0.7	
Coal price		−0.7	−0.2			0.2	0.4	
Gas price		0.8	0.2			−0.9	−1.3	
R^2	0.0	0.1	0.4	0.3	0.5	0.6	0.6	0.6
Adj. R^2	0.0	0.0	0.3	0.3	0.5	0.5	0.5	0.5
F statistics	1.1	1.5	4.2*	7.6*	9.0*	6.0*	6.4*	8.6*
I vs II	1.6				1.0			
II vs III		8.8*				4.1*		
III vs IV			0.9				1.1	

Note: $^* p < 0.05$.

The graph of the estimated quadratic dummy-interaction function of benchmark Model (4.2-IV) together with its 95% confidence interval is shown in Figure 4. The confidence interval covers the dummy-interaction function re

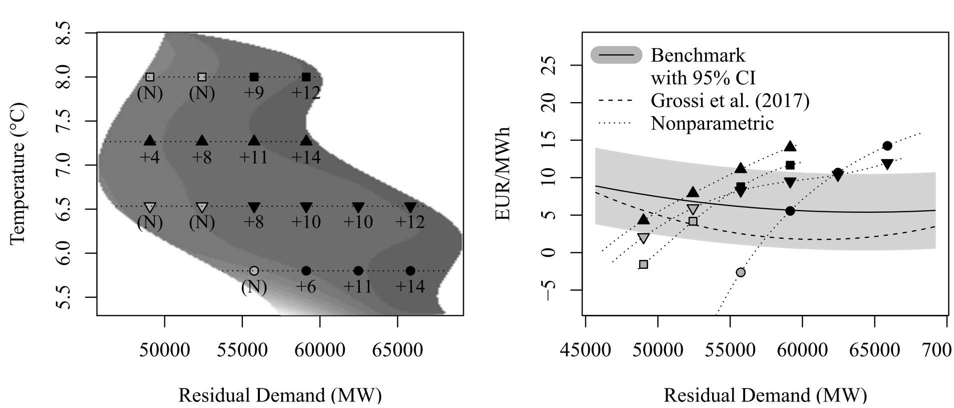

FIG. 4. LEFT: *Contour plot of the price difference surface* $\hat{\mu}_A(u, z) - \hat{\mu}_B(u, z)$; *(N) denotes insignificant results*. RIGHT: *Comparisons of the results from benchmark Model* (4.2-IV), *the original result of Grossi, Heim and Waterson* (2017), *and our nonparametric result*.

ported[1] by Grossi, Heim and Waterson (2017). The dummy-interaction function of our benchmark Model shows slightly larger values than reported by Grossi, Heim and Waterson (2017), since the market reactions to Germany's unexpected nuclear phaseout are less moderated in the short time frame of our event study than in the long time frame (±2 years) considered by Grossi, Heim and Waterson (2017).

4.2. *Nonparametric testing*. In contrast to the parametric benchmark models, we do not assume any specific functional form for the price effect of Germany's nuclear phaseout. Additionally, we allow for interactions between residual demand and temperature. As for any nonparametric model, however, we need to deal with the curse of dimensionality which prevents us from considering all control variables of potential relevance. Therefore, we focus on the two most important control variables (residual demand and temperature) and use the event study approach in order to minimize the confounding effects of the remaining less important control variables as in our benchmark study.

Residual demand and temperature are the two most important market fundamental control variables within the short term of an event study. First, because the interplay of residual demand with the merit-order curve determines electricity spot prices (see Figure 1). Second, because the merit-order curve itself depends on temperature. For the latter dependency, there are multiple reasons, both fundamental and speculative. The most important fundamental reason is that the merit-order curve is determined by the generation costs of the conventional (i.e., nuclear, lignite, coal, and gas) power plants. These conventional power plants are thermal-based electricity generators using river water as their primary cooling resource. The resulting thermal pollution is substantial and environmentally hazardous and, therefore, strictly regulated by public institutions. This regulation affects the cost structure, that is, the merit-order curve (McDermott and Nilsen (2014)). For given values of electricity demand, the merit-order curve is higher (lower) on warm (cold) days, when the river water has a smaller (larger) cooling capacity[2] (see Figure 2, for middle to high temperatures). However, there is also a reverse interaction effect between residual demand and temperature which can revert the fundamental price component described above. Due to the use of heating and cooling devices, cold and hot days are associated with the large amounts of electricity demand resulting in market situations where the electricity production capacities become scarce.[3] In these situations of market stress, one observes

[1]The data for this graph are extracted from Figure 6 in Grossi, Heim and Waterson (2017) using the WebPlotDigitizer of Rohatgi (2018).

[2]McDermott and Nilsen (2014) propose to use river temperature as a control variable. However, we do not have access to their data and therefore use air temperature as a proxy which is known to correlate strongly with river temperature (see Rabi, Hadzima-Nyarko and Šperac (2015)).

[3]The use of air conditioning systems is, however, less extensive in Germany than, for instance, in the US.

the high electricity prices (see Figure 2 for cold temperatures), since the electricity producers demand additional scarcity premiums (see Burger, Graeber and Schindlmayr (2008), Chapter 4).

To conclude, the effects of residual demand and temperature and their interaction effects are quite complex and it is easy to end up with a misspecified model, particularly when using a parametric approach. Despite these complexities, the parametric benchmark Model (4.2), proposed by Grossi, Heim and Waterson (2017), does not consider interaction effects between electricity demand and temperature. In the following, we compare our nonparametric test results with our parametric benchmarks and demonstrate the importance of this interaction effect.

The left plot in Figure 4 shows a contour plot of the price difference surface $\hat{\mu}_B - \hat{\mu}_A$ of the mean estimates before and after Germany's nuclear phaseout. The support of the difference surface equals the intersection $\mathrm{supp}(\hat{\mu}_B) \cap \mathrm{supp}(\hat{\mu}_A)$ of the supports of the mean functions $\hat{\mu}_B$ and $\hat{\mu}_A$,

$$\mathrm{supp}(\hat{\mu}_B) = \{(u, z) : \hat{a}_B(z) \le u \le \hat{b}_B(z) \text{ with } \hat{z}_{\min,B} \le z \le \hat{z}_{\max,B}\}, \quad \text{and}$$

$$\mathrm{supp}(\hat{\mu}_A) = \{(u, z) : \hat{a}_A(z) \le u \le \hat{b}_A(z) \text{ with } \hat{z}_{\min,A} \le z \le \hat{z}_{\max,A}\},$$

where the empirical boundary functions $\hat{a}_P(\cdot)$ and $\hat{b}_P(\cdot)$, $P \in \{A, B\}$, are computed using the LLK boundary estimator of Martins-Filho and Yao (2007).

In order to test the pointwise 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 $Z_{u,z}$ described in Section 2 with plugged-in empirical bias and variance expressions and GCV-optimal bandwidths as described in Section 3. The test statistic is evaluated at $G = 18$ regular grid-points (u_j, z_j), $j = 1, \dots, G$, within the intersection $\mathrm{supp}(\hat{\mu}_B) \cap \mathrm{supp}(\hat{\mu}_A)$. These test-points are shown by the points in the plots of Figure 4, where the different point shapes correspond to different temperature values. In order to account for the multiple testing we use a Bonferroni-adjusted significance level α/G where we set $\alpha = 0.05$. Significant differences at the chosen test-points are depicted by the numerical values in the left plot of Figure 4; nonsignificant differences are marked by "(N)". We are interested in pointwise hypotheses as we would like to identify significant and nonsignificant test-points. The Bonferroni adjustment is known to be conservative, but works well for our application. Practitioners seeking for an alternative to the Bonferroni adjustment are referred, for instance, to the work of Cox and Lee (2008).

The plots in Figure 4 show that the price differences are large (moderate) for large (moderate) values of electricity demand. This is in line with our expectations, since the merit-order curve is known to be steep (relatively flat) for large (moderate) values of electricity demand resulting in large (moderate) price differences (see Figure 1). Furthermore, there is a clear interaction effect in the price differences. If temperature increases, the thermal power plans loose cooling capacities which increases the production costs—particularly for the less efficient power plants. The latter results in a steeper merit-order curve and, therefore, larger price differences.

5. Discussion. On March 15, 2011, the German government showed a drastic reaction to the nuclear disaster in Fukushima Daiichi, Japan, and permanently shut down 40% of its nuclear power plants. This political decision raised concerns about increases in electricity prices and subsequent problems for industry and households. Empirical studies on possible price effects, however, are scarce and existing studies are based on restrictive parametric model assumptions.

In this work we add a functional data perspective based on the merit-order model, the most important model for explaining electricity spot prices (see Bublitz, Keles and Fichtner (2017), Burger et al. (2004), Cludius et al. (2014), Grossi, Heim and Waterson (2017), Hirth (2013), Liebl (2013), Sensfuß, Ragwitz and Genoese (2008)). We extend the work of Liebl (2013) and additionally control for nonfunctional covariate adjustments.

In order to test for a possible price effect, we compare the multivariate nonparametric local linear kernel estimates of the mean price functions before and after Germany's nuclear phase out on March 15, 2011, using a pointwise test statistic. Nonparametric smoothing of the pooled data is used, since the underlying daily price functions are only observed at 12 noisy discretization points ("sparse functional data"). The existing asymptotic results on this nonstandard smoothing problem only consider the leading variance term and neglect an additional functional data specific variance term, which is asymptotically negligible, but typically not practically negligible. Ignoring this additional variance term can result in serious size distortions and invalid test decisions (see Liebl (2019a)). Therefore, we propose a finite sample correction that considers also the second functional data specific variance term (see Theorem 2.1 and Corollary 2.1). Theorem 2.1 generalizes the main result in Liebl (2019a) by allowing for a time series context with weak dependency structure.

We compare our nonparametric test results with parametric benchmark results replicating the results recently reported by Grossi, Heim and Waterson (2017). Our results confirm the existence of a price effect due to Germany's abrupt nuclear phase out, but our price effect is structured quite differently to the results in Grossi, Heim and Waterson (2017). While our nonparametric price differences are highest for large values of residual demand, the parametric benchmark models estimate the highest price differences for small values of residual demand (see right panel in Figure 4). One of the fundamental differences between our approach and the approach of Grossi, Heim and Waterson (2017) is that we take into account interactions with the important temperature factor, while Grossi, Heim and Waterson (2017) do not allow for this kind of interaction effect. Such a reversal of effects that is due to introducing an additional conditioning variable is known as Simpson's paradox (see, for instance, Wagner (1982)). Grossi, Heim and Waterson (2017) concede that their result is "unexpected" and present different market explanations for their unexpected result; however, a possible model-misspecification is not taken into account. Our nonparametric case study points at such a possible model-misspecification and demonstrates that a Simpson's paradox can explain their unexpected finding.

Acknowledgements. I would like to thank the referees and the Editors for their constructive feedback which helped to improve this research work.

SUPPLEMENTARY MATERIAL

Supplement A: R-codes and data (DOI: 10.1214/18-AOAS1230SUPPA; .pdf). This supplementary material contains the R codes of the real data application and simulated data which closely resembles the original data set.

Supplement B: Supplementary paper (DOI: 10.1214/18-AOAS1230SUPPB; .zip). This supplementary paper contains the proofs of our theoretical results and a detailed description of the data sources.

REFERENCES

BALL, R. and BROWN, P. (1968). An empirical evaluation of accounting income numbers. *J. Acc. Res.* 159–178.

BENKO, M., HÄRDLE, W. and KNEIP, A. (2009). Common functional principal components. *Ann. Statist.* **37** 1–34. MR2488343

BENTH, F. E., KHOLODNYI, V. A. and LAURENCE, P., eds. (2014). *Quantitative Energy Finance: Modeling, Pricing, and Hedging in Energy and Commodity Markets*. Springer, New York. MR3185034

BETZER, A., DOUMET, M. and RINNE, U. (2013). How policy changes affect shareholder wealth: The case of the Fukushima Daiichi nuclear disaster. *Appl. Econ. Lett.* **20** 799–803.

BRUNINX, K., MADZHAROV, D., DELARUE, E. and D'HAESELEER, W. (2013). Impact of the German nuclear phase-out on Europe's electricity generation—a comprehensive study. *Energy Policy* **60** 251–261.

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*, 1st ed. Wiley, New York.

BURGER, M., KLAR, B., MÜLLER, A. and SCHINDLMAYR, G. (2004). A spot market model for pricing derivatives in electricity markets. *Quant. Finance* **4** 109–122.

CAO, G., YANG, L. and TODEM, D. (2012). Simultaneous inference for the mean function based on dense functional data. *J. Nonparametr. Stat.* **24** 359–377. MR2921141

CARDOT, H. (2007). Conditional functional principal components analysis. *Scand. J. Stat.* **34** 317–335. MR2346642

CHARNIGO, R. and SRINIVASAN, C. (2015). A multivariate generalized C_p and surface estimation. *Biostatistics* **16** 311–325. MR3365430

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 Econ.* **44** 302–313.

COX, D. D. and LEE, J. S. (2008). Pointwise testing with functional data using the Westfall–Young randomization method. *Biometrika* **95** 621–634. MR2443179

FAMA, E. F., FISHER, L., JENSEN, M. C. and ROLL, R. (1969). The adjustment of stock prices to new information. *Internat. Econom. Rev.* **10** 1–21.

FAN, J. and YAO, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. *Springer Series in Statistics*. Springer, New York. MR1964455

FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. *Springer Series in Statistics*. Springer, New York. MR2229687

FERSTL, R., UTZ, S. and WIMMER, M. (2012). The effect of the Japan 2011 disaster on nuclear and alternative energy stocks worldwide: An event study. *Bus. Res.* **5** 25–41.

FOGARTY, C. B. and SMALL, D. S. (2014). Equivalence testing for functional data with an application to comparing pulmonary function devices. *Ann. Appl. Stat.* **8** 2002–2026. MR3292487

GROMENKO, O. and KOKOSZKA, P. (2012). Testing the equality of mean functions of ionospheric critical frequency curves. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **61** 715–731. MR2993506

GROMENKO, O., KOKOSZKA, P. and SOJKA, J. (2017). Evaluation of the cooling trend in the ionosphere using functional regression with incomplete curves. *Ann. Appl. Stat.* **11** 898–918. MR3693551

GROSSI, L., HEIM, S. and WATERSON, M. (2017). The impact of the German response to the Fukushima earthquake. *Energy Econ.* **66** 450–465.

HALL, P., MÜLLER, H.-G. and WANG, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34** 1493–1517. MR2278365

HALL, P. and VAN KEILEGOM, I. (2007). Two-sample tests in functional data analysis starting from discrete data. *Statist. Sinica* **17** 1511–1531. MR2413533

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

HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications. Springer Series in Statistics*. Springer, New York. MR2920735

HORVÁTH, L., KOKOSZKA, P. and REEDER, R. (2013). Estimation of the mean of functional time series and a two-sample problem. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 103–122. MR3008273

HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators. Wiley Series in Probability and Statistics*. Wiley, Chichester. MR3379106

JIANG, C.-R. and WANG, J.-L. (2010). Covariate adjusted functional principal components analysis for longitudinal data. *Ann. Statist.* **38** 1194–1226. MR2604710

LI, M., STAICU, A.-M. and BONDELL, H. D. (2015). Incorporating covariates in skewed functional data models. *Biostatistics* **16** 413–426. MR3365436

LIEBL, D. (2013). Modeling and forecasting electricity spot prices: A functional data perspective. *Ann. Appl. Stat.* **7** 1562–1592. MR3127959

LIEBL, D. (2019a). Inference for sparse and dense functional data with covariate adjustments. *J. Multivariate Anal.* **170** 315–335.

LIEBL, D. (2019b). Supplement to "Nonparametric testing for differences in electricity prices: The case of the Fukushima nuclear accident." DOI:10.1214/18-AOAS1230SUPPA, DOI:10.1214/18-AOAS1230SUPPB.

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

LISI, F. and SHAH, I. (2018). Forecasting next-day electricity demand and prices based on functional models. Working paper.

MACIEJOWSKA, K. and WERON, R. (2016). Short-and mid-term forecasting of baseload electricity prices in the UK: The impact of intra-day price relationships and market fundamentals. *IEEE Trans. Power Syst.* **31** 994–1005.

MACKINLAY, A. C. (1997). Event studies in economics and finance. *J. Econ. Lit.* **35** 13–39.

MARTINS-FILHO, C. and YAO, F. (2007). Nonparametric frontier estimation via local linear regression. *J. Econometrics* **141** 283–319. MR2411745

MCDERMOTT, G. R. and NILSEN, Ø. A. (2014). Electricity prices, river temperatures, and cooling water scarcity. *Land Econ.* **90** 131–148.

MCWILLIAMS, A. and SIEGEL, D. (1997). Event studies in management research: Theoretical and empirical issues. *Acad. Manage. J.* **40** 626–657.

PARASCHIV, F., ERNI, D. and PIETSCH, R. (2014). The impact of renewable energies on EEX day-ahead electricity prices. *Energy Policy* **73** 196–210.

RABI, A., HADZIMA-NYARKO, M. and ŠPERAC, M. (2015). Modelling river temperature from air temperature: Case of the river Drava (Croatia). *Hydrol. Sci. J.* **60** 1490–1507.

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2168993

ROHATGI, A. (2018). WebPlotDigitizer, version: 4.1. Austin, TX.

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. *Ann. Appl. Stat.* **5** 2024–2051. MR2884930

SHAH, I. and LISI, F. (2018). Forecasting of electricity price through a functional prediction of supply and demand curves. Working paper.

SHEATHER, S. J. and JONES, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. Roy. Statist. Soc. Ser. B* **53** 683–690. MR1125725

THOENES, S. (2014). Understanding the determinants of electricity prices and the impact of the German nuclear moratorium in 2011. *Energy J.* **35** 61–78.

VENABLES, W. N. and RIPLEY, B. D. (2002). *Modern Applied Statistics with S*, 4th ed. Springer, New York.

VSEVOLOZHSKAYA, O., GREENWOOD, M. and HOLODOV, D. (2014). Pairwise comparison of treatment levels in functional analysis of variance with application to erythrocyte hemolysis. *Ann. Appl. Stat.* **8** 905–925. MR3262539

WAGNER, C. H. (1982). Simpson's paradox in real life. *Amer. Statist.* **36** 46–48.

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561

ZHANG, J.-T. and CHEN, J. (2007). Statistical inferences for functional data. *Ann. Statist.* **35** 1052–1079. MR2341698

ZHANG, X. and WANG, J.-L. (2016). From sparse to dense functional data and beyond. *Ann. Statist.* **44** 2281–2321. MR3546451

ZHANG, W. and WEI, Y. (2015). Regression based principal component analysis for sparse functional data with applications to screening growth paths. *Ann. Appl. Stat.* **9** 597–620. MR3371327

STATISTISCHE ABTEILUNG
UNIVERSITY OF BONN
ADENAUERALLEE 24-26
53113 BONN
GERMANY
E-MAIL: dliebl@uni-bonn.de