Supplemental Material for:

Anti-Proportional Bandwidth Selection for Smoothing (Non-)Sparse Functional Data with Covariate Adjustments

Dominik Liebl

Statistische Abteilung

University of Bonn

Adenauerallee 24-42

53113 Bonn, Germany

Outline: Section 1 contains the proofs of the Theorems 3.1 and 3.2. Further discussions and supplementary results on the bandwidth results and the rule-of-thumb procedure used in the application are found in Section 2. The data sources are described in detail in Section 3.

1. Proofs

1.1. Proof of Theorem 3.1

Technical assumptions:

A-RD (Random Design) The triple (Y_{it}, X_{it}, Z_t) has the same distribution as (Y, X, Z) with p.d.f. f_{YXZ}. The conditional r.v. $X|Z$ is i.i.d. with p.d.f. $f_{X|Z}(x) > 0$ for all $x \in S_{X|Z}$ and zero else, where $S_{X|Z} = [a(Z), b(Z)] \subset \mathbb{R}$, and $a(z) < b(z)$ for all $z \in [z_{\min}, z_{\max}]$. The marginal p.d.f. of X is $f_X(x) > 0$ for all $x \in S_X$, where $S_X = [\min_z a(z), \max_z b(z)]$. The p.d.f. of Z is $f_Z(z) > 0$ for all $z \in S_Z$ and zero else, where $S_Z = [z_{\min}, z_{\max}] \subset \mathbb{R}$. The joint p.d.f. is then given by $f_{XZ}(x, z) = f_{X|Z}(x) f_Z(z) > 0$ for all $(x, z) \in S_{XZ}$, where $S_{XZ} = S_{X|Z} \times S_Z$. This implies that the conditional p.d.f. of $(X, X)|Z$ is $f_{XX|Z}(x_1, x_2) > 0$ for all $(x_1, x_2) \in S^2_{X|Z}$ and zero else. The joint p.d.f. is given by $f_{XXZ}(x_1, x_2, z) = f_{XX|Z}(x_1, x_2) f_Z(z) > 0$ for all $(x_1, x_2, z) \in S_{XXZ}$, where $S_{XXZ} = S^2_{X|Z} \times S_Z$. All p.d.f.'s in are continuously differentiable for all points within their supports.

A-SM (Smoothness) All second-order derivatives of μ (and γ) are continuous for all points within its support. The autocovariance functions γ (and $\tilde{\gamma}$) and γ_u (and $\tilde{\gamma}_u$) are continuously differentiable for all points within their supports.

A-BW (Bandwidths) For "small-n" ($n \sim T^\theta$ with $0 \leq \theta < 1/5$) it is not restrictive to assume that $h_{\mu,X} \sim h_{\mu,Z} \sim h_\mu$ and $h_{\gamma,X} \sim h_{\gamma,Z} \sim h_\gamma$ as this is implied by the well known optimal bandwidth results.

Bandwidths for $\hat{\mu}$: $h_\mu \to 0$, $Tn\, h_\mu^4 \to \infty$ as $Tn \to \infty$.

Bandwidths for $\hat{\gamma}$: $h_\gamma \to 0$, $TN\, h_\gamma^5 \to \infty$ as $TN \to \infty$.

For "large-n" ($n \sim T^\theta$ with $1/5 < \theta < \infty$) the two bandwidth parameters h_X and h_Z cannot be assumed to be of the same order, which makes the set of bandwidth assumptions less compact, but essentially equivalent.

Bandwidths for $\hat{\mu}$: $h_{\mu,X} \to 0$, $h_{\mu,Z} \to 0$, $Th_{\mu,Z} \to \infty$, $nh_{\mu,X} \to \infty$, $Tnh_{\mu,X}^3 h_{\mu,Z} \to \infty$, $Tnh_{\mu,X} h_{\mu,Z}^3 \to \infty$, $Tnh_{\mu,X}^{-3} h_{\mu,Z}^5 \to \infty$, $Tnh_{\mu,X}^5 h_{\mu,Z}^{-3} \to \infty$ as $Tn \to \infty$.

Bandwidths for $\hat{\gamma}$: $h_{\gamma,X} \to 0$, $h_{\gamma,Z} \to 0$, $Th_{\gamma,Z} \to \infty$, $Nh_{\gamma,X} \to \infty$, $TNh_{\gamma,X}^4 h_{\gamma,Z} \to \infty$, $TNh_{\gamma,X}^2 h_{\gamma,Z}^3 \to \infty$, $TNh_{\gamma,X}^{-2} h_{\gamma,Z}^5 \to \infty$, $TNh_{\gamma,X}^6 h_{\gamma,Z}^{-3} \to \infty$ as $TN \to \infty$.

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

Proof of Theorem 3.1, part *(i)*: Let x and z be interior points of supp(f_{XZ}) and define $\mathbf{H}_\mu = \mathrm{diag}(h_{\mu,X}^2, h_{\mu,Z}^2)$, $\mathbf{X} = (X_{11}, \ldots, X_{nT})^\top$, and $\mathbf{Z} = (Z_1, \ldots, Z_T)^\top$. Taylor-expansion of μ around (x, z), the conditional bias of the estimator $\hat{\mu}(x, z; \mathbf{H})$ can be written as

$$\mathbb{E}(\hat{\mu}(x,z;\mathbf{H}_\mu) - \mu(x,z)|\mathbf{X},\mathbf{Z}) = \frac{1}{2}u_1^\top \left((Tn)^{-1}[\mathbf{1},\mathbf{X}_x,\mathbf{Z}_z]^\top \mathbf{W}_{\mu,xz}[\mathbf{1},\mathbf{X}_x,\mathbf{Z}_z]\right)^{-1} \times \\ \times (Tn)^{-1}[\mathbf{1},\mathbf{X}_x,\mathbf{Z}_z]^\top \mathbf{W}_{\mu,xz}\left(\boldsymbol{\mathcal{Q}}_\mu(x,z) + \mathbf{R}_\mu(x,z)\right), \tag{24}$$

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

$$(X_{it} - x, Z_t - z)\boldsymbol{\mathcal{H}}_\mu(x,z)(X_{it} - x, Z_t - z)^\top \in \mathbb{R}$$

with $\boldsymbol{\mathcal{H}}_\mu(x, z)$ being the Hessian matrix of the regression function $\mu(x, z)$. The $Tn \times 1$ vector $\mathbf{R}_\mu(x, z)$ holds the remainder terms

$$o\left((X_{it} - x, Z_t - z)(X_{it} - x, Z_t - z)^\top\right) \in \mathbb{R}.$$

Next we derive asymptotic approximations for the 3×3 matrix $\left((Tn)^{-1}[\mathbf{1},\mathbf{X}_x,\mathbf{Z}_z]^\top \mathbf{W}_{\mu,xz}[\mathbf{1},\mathbf{X}_x,\mathbf{Z}_z]\right)^{-1}$ and the 3×1 matrix $(Tn)^{-1}[\mathbf{1},\mathbf{X}_x,\mathbf{Z}_z]^\top \mathbf{W}_{\mu,xz}\boldsymbol{\mathcal{Q}}_\mu(x,z)$ of the right hand side of Eq. (24). Using standard procedures from kernel density estimation it is easy to derive that

$$(Tn)^{-1}[\mathbf{1},\mathbf{X}_x,\mathbf{Z}_z]^\top \mathbf{W}_{\mu,xz}[\mathbf{1},\mathbf{X}_x,\mathbf{Z}_z] = \\ \begin{pmatrix} f_{XZ}(x,z) + o_p(1) & \nu_2(K_\mu)\mathbf{D}_{f_{XZ}}(x,z)^\top \mathbf{H}_\mu + o_p(\mathbf{1}^\top \mathbf{H}_\mu) \\ \nu_2(K_\mu)\mathbf{H}_\mu \mathbf{D}_{f_{XZ}}(x,z) + o_p(\mathbf{H}_\mu \mathbf{1}) & \nu_2(K_\mu)\mathbf{H}_\mu f_{XZ}(x,z) + o_p(\mathbf{H}_\mu) \end{pmatrix},$$

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

$$\left((Tn)^{-1}[\mathbf{1},\mathbf{X}_x,\mathbf{Z}_z]^\top \mathbf{W}_{\mu,xz}[\mathbf{1},\mathbf{X}_x,\mathbf{Z}_z]\right)^{-1} = \tag{25}$$

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

The 3×1 matrix $(Tn)^{-1}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,xz} \boldsymbol{\mathcal{Q}}_\mu(x,z)$ can be partitioned as following:

$$(Tn)^{-1}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z]^\top \mathbf{W}_{xz} \boldsymbol{\mathcal{Q}}_\mu(x,z) \quad = \quad \begin{pmatrix} \texttt{upper element} \\ \texttt{lower bloc} \end{pmatrix},$$

where the 1×1 dimensional `upper element` can be approximated by a scalar variable equal to

$$\begin{aligned} &(Tn)^{-1} \sum_{it} K_{\mu,h}(X_{it} - x, Z_t - z)(X_{it} - x, Z_t - z) \boldsymbol{\mathcal{H}}_\mu(x,z)(X_{it} - x, Z_t - z)^\top \\ &= (\nu_2(\kappa))^2\, tr\left\{\mathbf{H}_\mu \boldsymbol{\mathcal{H}}_\mu(x,z)\right\} f_{XZ}(x,z) + o_p(tr(\mathbf{H}_\mu)) \end{aligned} \qquad (26)$$

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

$$\begin{aligned} &(Tn)^{-1} \sum_{it} \left\{ K_{\mu,h}(X_{it} - x, Z_t - z)(X_{it} - x, Z_t - z) \boldsymbol{\mathcal{H}}_\mu(x,z)(X_{it} - x, Z_t - z)^\top \right\} \times \\ &\times (X_{it} - x, Z_t - z)^\top = O_p(\mathbf{H}_\mu^{3/2} \mathbf{1}). \end{aligned} \qquad (27)$$

Plugging in the approximations of Eqs. (25)-(27) into the first summand of the conditional bias expression in Eq. (24) leads to the following expression

$$\begin{aligned} &\frac{1}{2} u_1^\top ((Tn)^{-1}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,xz}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z])^{-1} (Tn)^{-1}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,xz} \boldsymbol{\mathcal{Q}}_\mu(x,z) = \\ &= \frac{1}{2} (\nu_2(\kappa))^2\, tr\left\{\mathbf{H}_\mu \boldsymbol{\mathcal{H}}_\mu(x,z)\right\} + o_p(tr(\mathbf{H}_\mu)). \end{aligned}$$

Furthermore, it is easily seen that the second summand of the conditional bias expression in Eq. (24), which holds the remainder term, is given by

$$\frac{1}{2} u_1^\top ((Tn)^{-1}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,xz}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z])^{-1} (Tn)^{-1}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,xz} \mathbf{R}_\mu(x,z) = o_p(tr(\mathbf{H}_\mu)).$$

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

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

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

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

$$\begin{aligned} =& u_1^\top ([\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,xz}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z])^{-1} [\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,xz}\, \mathrm{Cov}(\mathbf{Y}|\mathbf{X}, \mathbf{Z})\, \mathbf{W}_{\mu,xz}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z] \\ & ([\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,xz}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z])^{-1} u_1 \\ =& u_1^\top ((Tn)^{-1}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,xz}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z])^{-1} ((Tn)^{-2}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,xz}\, \mathrm{Cov}(\mathbf{Y}|\mathbf{X}, \mathbf{Z})\, \mathbf{W}_{\mu,xz}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z]) \\ & ((Tn)^{-1}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,xz}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z])^{-1} u_1, \end{aligned} \qquad (28)$$

where $\mathrm{Cov}(\mathbf{Y}|\mathbf{X},\mathbf{Z})$ is the $Tn \times Tn$ matrix with typical elements

$$\mathrm{Cov}(Y_{it}, Y_{js}|X_{it}, X_{js}, Z_t, Z_s) = \gamma_{|t-s|}((X_{it}, Z_t), (X_{js}, Z_s)) + \sigma_\epsilon^2 \mathbb{1}(i = j \text{ and } t = s)$$

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

We begin with analyzing the 3×3 matrix

$$(Tn)^{-2}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,xz} \,\mathrm{Cov}(\mathbf{Y}|\mathbf{X},\mathbf{Z})\, \mathbf{W}_{\mu,xz}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z]$$

using the following three Lemmas 1.1-1.3.

Lemma 1.1. *The upper-left scalar (block) of the matrix* $(Tn)^{-2}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,xz}\mathrm{Cov}(\mathbf{Y}|\mathbf{X},\mathbf{Z})\mathbf{W}_{\mu,xz}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z]$ *is given by*

$$\begin{aligned}
& (Tn)^{-2}\mathbf{1}^\top \mathbf{W}_{\mu,xz}\mathrm{Cov}(\mathbf{Y}|\mathbf{X},\mathbf{Z})\mathbf{W}_{\mu,xz}\mathbf{1} \\
= \;& (Tn)^{-1} f_{XZ}(x,z)|\mathbf{H}_\mu|^{-1/2} R(K_\mu)\left(\gamma(x,x,z) + \sigma_\epsilon^2\right)(1 + O_p(tr(\mathbf{H}_\mu^{1/2}))) \\
+ \;& T^{-1}(f_{XZ}(x,z))^2 \left[\left(\frac{n-1}{n}\right) h_{\mu,Z}^{-1} R(\kappa)\frac{\gamma(x,x,z)}{f_Z(z)} + c_r\right](1 + O_p(tr(H^{1/2}))) \\
= \;& O_p((Tn)^{-1}|\mathbf{H}_\mu|^{-1/2}) + O_p(T^{-1}h_{\mu,Z}^{-1}).
\end{aligned}$$

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

$$\begin{aligned}
& (Tn)^{-2}\mathbf{1}^\top \mathbf{W}_{\mu,xz}\mathrm{Cov}(\mathbf{Y}|\mathbf{X},\mathbf{Z})\mathbf{W}_{\mu,xz}\begin{pmatrix} (X_{11} - x, Z_1 - z) \\ \vdots \\ (X_{nT} - x, Z_T - z) \end{pmatrix} \\
= \;& (Tn)^{-1} f_{XZ}(x,z)|\mathbf{H}_\mu|^{-1/2}(\mathbf{1}^\top \mathbf{H}_\mu^{1/2}) R(K_\mu)\left(\gamma(x,x,z) + \sigma_\epsilon^2\right)(1 + O_p(tr(\mathbf{H}_\mu^{1/2}))) \\
+ \;& T^{-1}(f_{XZ}(x,z))^2(\mathbf{1}^\top \mathbf{H}_\mu^{1/2}) \left[\left(\frac{n-1}{n}\right) h_{\mu,Z}^{-1} R(\kappa)\frac{\gamma(x,x,z)}{f_Z(z)} + c_r\right](1 + O_p(tr(\mathbf{H}_\mu^{1/2}))) \\
= \;& O_p((Tn)^{-1}|\mathbf{H}_\mu|^{-1/2}(\mathbf{1}^\top \mathbf{H}_\mu^{1/2})) + O_p(T^{-1}(\mathbf{1}^\top \mathbf{H}_\mu^{1/2})h_{\mu,Z}^{-1}).
\end{aligned}$$

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

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

$$\begin{aligned}
& (Tn)^{-2}\left(((X_{11} - x), (Z_1 - z))^\top, \ldots, ((X_{nT} - x), (Z_T - z)^\top)\right) \times \\
& \times \mathbf{W}_{\mu,xz}\mathrm{Cov}(\mathbf{Y}|\mathbf{X},\mathbf{Z})\mathbf{W}_{\mu,xz}\begin{pmatrix} (X_{11} - x, Z_1 - z) \\ \vdots \\ (X_{nT} - x, Z_T - z) \end{pmatrix} \\
= \;& (Tn)^{-1} f_{XZ}(x,z)|\mathbf{H}_\mu|^{-1/2}\mathbf{H}_\mu R(K_\mu)\left(\gamma(x,x,z) + \sigma_\epsilon^2\right)(1 + O_p(tr(\mathbf{H}_\mu^{1/2}))) \\
+ \;& T^{-1}(f_{XZ}(x,z))^2\mathbf{H}_\mu \left[\left(\frac{n-1}{n}\right) h_{\mu,Z}^{-1} R(\kappa)\frac{\gamma(x,x,z)}{f_Z(z)} + c_r\right](1 + O_p(tr(\mathbf{H}_\mu^{1/2}))) \\
= \;& O_p((Tn)^{-1}|\mathbf{H}_\mu|^{-1/2}\mathbf{H}_\mu) + O_p(T^{-1}\mathbf{H}_\mu h_{\mu,Z}^{-1}).
\end{aligned}$$

Using the approximations for the bloc-elements of the matrix $(Tn)^{-2}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,xz}\text{Cov}(\mathbf{Y}|\mathbf{X},\mathbf{Z})\mathbf{W}_{\mu,xz}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z]$, given by the Lemmas 1.1-1.3, and the approximation for the matrix $\left((Tn)^{-1}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z]^\top \mathbf{W}_{\mu,xz}[\mathbf{1}, \mathbf{X}_x, \mathbf{Z}_z]\right)^{-1}$, given in (25), we can approximate the conditional variance of the bivariate local linear estimator, given in (28). Some tedious yet straightforward matrix algebra leads to $\mathbb{V}(\hat{\mu}(x, z; \mathbf{H}_\mu)|\mathbf{X}, \mathbf{Z}) =$

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

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

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

$$\begin{aligned}
s_1 &= (Tn)^{-2}\sum_{it}(K_{\mu,h}(X_{it}-x, Z_t-z))^2\,\mathbb{V}(Y_{it}|\mathbf{X},\mathbf{Z}) && (29)\\
&= (Tn)^{-1}|\mathbf{H}_\mu|^{-1/2} f_{XZ}(x,z)R(K_\mu)\left(\gamma(x,x,z)+\sigma_\epsilon^2\right) + O((Tn)^{-1}|\mathbf{H}_\mu|^{-1/2}\,tr(\mathbf{H}_\mu^{1/2})),\\
s_2 &= (Tn)^{-2}\sum_{ij}\sum_{\substack{ts\\ t\neq s}} K_{\mu,h}(X_{it}-x, Z_t-z)\,\text{Cov}(Y_{it}, Y_{js}|\mathbf{X},\mathbf{Z})\,K_{\mu,h}(X_{js}-x, Z_s-z) && (30)\\
&= T^{-1}(f_{XZ}(x,z))^2 c_r + O_p(T^{-1}tr(\mathbf{H}_\mu^{1/2}))\\
s_3 &= (Tn)^{-2}\sum_{\substack{ij\\ i\neq j}}\sum_{t} h_{\mu,X}^{-1}\kappa(h_{\mu,X}^{-1}(X_{it}-x))(h_{\mu,Z}^{-1}\kappa(h_{\mu,Z}^{-1}(Z_t-z)))^2\,\text{Cov}(Y_{it}, Y_{jt}|\mathbf{X},\mathbf{Z})\times && (31)\\
&\times h_{\mu,X}^{-1}\kappa(h_{\mu,X}^{-1}(X_{jt}-x))\\
&= T^{-1}(f_{XZ}(x,z))^2\left[\left(\frac{n-1}{n}\right)h_{\mu,Z}^{-1}R(\kappa)\frac{\gamma(x,x,z)}{f_Z(z)}\right] + O_p(T^{-1}tr(\mathbf{H}_\mu^{1/2})),
\end{aligned}$$

where $|c_r| \leq \frac{2c_\gamma r}{1-r} < \infty$. Summing up (29)-(30) leads to the result in Lemma 1.1. Lemmas 1.2 and 1.3 differ from Lemma 1.1 only with respect to the additional factors $\mathbf{1}^\top\mathbf{H}_\mu^{1/2}$ and $\mathbf{H}_\mu$. These come in due to the usual substitution step for the additional data parts $(X_{it}-x, Z_t-z)$.

1.2. Proof of Theorem 3.2

The proof of Theorem 3.2 follows the same arguments as in the proof of Theorem 3.1. The only additional issue we need to consider, is that we do not directly observe the dependent variables, namely, the true theoretical raw covariances C_{ijt} as defined in Eq. (4), but only their empirical versions $\hat{C}_{ijt}$ as defined in Eq. (9). In the following we show that this additional estimation error is asymptotically negligible in comparison to the approximation error of $\hat{\gamma}$ derived under the usage of the theoretical raw covariates C_{ijt} as given in Eq. (20) of Theorem 3.3.

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

$$\begin{aligned}\hat{C}_{ijt} = C_{ijt} &+ (Y_{it} - \mu(X_{it}, Z_t))(\mu(X_{jt}, Z_t) - \hat{\mu}(X_{jt}, Z_t)) \\ &+ (Y_{jt} - \mu(X_{jt}, Z_t))(\mu(X_{it}, Z_t) - \hat{\mu}(X_{it}, Z_t)) \\ &+ (\mu(X_{it}, Z_t) - \hat{\mu}(X_{it}, Z_t))(\mu(X_{jt}, Z_t) - \hat{\mu}(X_{jt}, Z_t)).\end{aligned}$$

Further, it follows from Eq. (19) in Theorem 3.3 that

$$\left(\hat{\mu}(X_{it}, Z_t; h_{\mu,X,\mathrm{opt}}, h_{\mu,Z,\mathrm{opt}}) - \mu(X_{it}, Z_t)\right)\big|\mathbf{X}, \mathbf{Z} = \begin{cases} O_p\left((Tn)^{-2/3}\right) & \text{if } 0 \leq \theta \leq 1/5 \\ O_p\left(T^{-4/5}\right) & \text{if } \theta > 1/5, \end{cases}$$

for all i and t, where $h_{\mu,X,\mathrm{opt}}$ and $h_{\mu,Z,\mathrm{opt}}$ denote the θ-specific AMISE optimal bandwidth choices as defined in Eqs. (13), (15), (34), and (35). Therefore, under our setup we have that

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

which is of an order of magnitude smaller than the approximation error of $\hat{\gamma}$ derived under the usage of the theoretical raw covariates C_{ijt} as given in Eq. (20) of Theorem 3.3.

2. Further discussions

2.1. AMISE.I *optimal bandwidth selection*

The following expressions are derived under the hypothesis that the first variance terms in parts *(ii)* of Theorems 3.1 and 3.2 are dominating. The AMISE.I expression for the local linear estimator $\hat{\mu}$ is given by

$$\begin{aligned}\mathrm{AMISE.I}_{\hat{\mu}}\left(h_{\mu,X}, h_{\mu,Z}\right) &= (Tn)^{-1}\, h_{\mu,X}^{-1}\, h_{\mu,Z}^{-1}\, R(K_\mu)\, Q_{\mu,1} + \\ &+ \frac{1}{4}\left(\nu_2(K_\mu)\right)^2 \left[h_{\mu,X}^4\, \mathcal{I}_{\mu,XX} + 2\, h_{\mu,X}^2\, h_{\mu,Z}^2\, \mathcal{I}_{\mu,XZ} + h_{\mu,Z}^4\, \mathcal{I}_{\mu,ZZ}\right],\end{aligned} \tag{32}$$

$$\begin{aligned}\text{where}\qquad Q_{\mu,1} &= \textstyle\int_{S_{XZ}} \left(\gamma(x,x,z) + \sigma_\epsilon^2\right)\, d(x,z), \\ \mathcal{I}_{\mu,XX} &= \textstyle\int_{S_{XZ}} \left(\mu^{(2,0)}(x,z)\right)^2\, w_\mu\, d(x,z), \\ \mathcal{I}_{\mu,ZZ} &= \textstyle\int_{S_{XZ}} \left(\mu^{(0,2)}(x,z)\right)^2\, w_\mu\, d(x,z), \quad \text{and} \\ \mathcal{I}_{\mu,XZ} &= \textstyle\int_{S_{XZ}} \mu^{(2,0)}(x,z)\mu^{(0,2)}(x,z)\, w_\mu\, d(x,z).\end{aligned}$$

This is a known expression for the AMISE of a two-dimensional local linear estimator with a diagonal bandwidth matrix (see, e.g., Herrmann et al. (1995)) and can be derived using the formulas in Wand & Jones (1994).

The AMISE.I expression for the local linear estimator $\hat{\gamma}$ is less common as it is based on two bandwidths, which leads to

$$\begin{aligned}\mathrm{AMISE.I}_{\hat{\gamma}}\left(h_{\gamma,X}, h_{\gamma,Z}\right) &= (TN)^{-1}\, h_{\gamma,X}^{-2}\, h_{\gamma,Z}^{-1}\, R(K_\gamma)\, Q_{\gamma,1} + \\ + \frac{1}{4}\left(\nu_2(K_\gamma)\right)^2 &\left[2\, h_{\gamma,X}^4 \left(\mathcal{I}_{\gamma,X_1X_1} + \mathcal{I}_{\gamma,X_1X_2}\right) + 4\, h_{\gamma,X}^2\, h_{\gamma,Z}^2\, \mathcal{I}_{\gamma,X_1Z} + h_{\gamma,Z}^4\, \mathcal{I}_{\gamma,ZZ}\right],\end{aligned} \tag{33}$$

where

$$\begin{aligned}
Q_{\gamma,1} &= \int_{S_{XXZ}} \left(\tilde{\gamma}((x_1,x_2),(x_1,x_2),z) + \sigma_\varepsilon^2(x_1,x_2,z)\right) d(x_1,x_2,z) \\
\mathcal{I}_{\gamma,X_1X_1} &= \int_{S_{XXZ}} \left(\gamma^{(2,0,0)}(x_1,x_2,z)\right)^2 w_\gamma\, d(x_1,x_2,z), \\
\mathcal{I}_{\gamma,X_1X_2} &= \int_{S_{XXZ}} \left(\gamma^{(2,0,0)}(x_1,x_2,z)\gamma^{(0,2,0)}(x_1,x_2,z)\right) w_\gamma\, d(x_1,x_2,z), \\
\mathcal{I}_{\gamma,X_1Z} &= \int_{S_{XXZ}} \gamma^{(2,0,0)}(x_1,x_2,z)\gamma^{(0,0,2)}(x_1,x_2,z)\, w_\gamma\, d(x_1,x_2,z), \quad \text{and} \\
\mathcal{I}_{\gamma,ZZ} &= \int_{S_{XXZ}} \left(\gamma^{(0,0,2)}(x_1,x_2,z)\right)^2 w_\gamma\, d(x_1,x_2,z).
\end{aligned}$$

Equation (33) can be derived again using the formulas in Wand & Jones (1994) with the further simplifications that $\mathcal{I}_{\gamma,X_1X_1} = \mathcal{I}_{\gamma,X_2X_2}$, $\mathcal{I}_{\gamma,X_1X_2} = \mathcal{I}_{\gamma,X_2X_1}$, and $\mathcal{I}_{\gamma,X_1Z} = \mathcal{I}_{\gamma,X_2Z}$ due to the symmetry of the covariance function, where the expressions $\mathcal{I}_{\gamma,X_2X_2}$, $\mathcal{I}_{\gamma,X_2X_1}$, and $\mathcal{I}_{\gamma,X_2Z}$ are defined in correspondence with those above.

The AMISE. I expressions (32) and (33) allow us to analytically derive the AMISE. I optimal bandwidth pairs. For the mean function we have

$$h_{\mu,X,\text{AMISE.I}} = \left(\frac{R(K_\mu)\, Q_{\mu,1}\, \mathcal{I}_{\mu,ZZ}^{3/4}}{Tn\ (\nu_2(K_\mu))^2\ \left[\mathcal{I}_{\mu,XX}^{1/2}\,\mathcal{I}_{\mu,ZZ}^{1/2} + \mathcal{I}_{\mu,XZ}\right]\mathcal{I}_{\mu,XX}^{3/4}}\right)^{1/6} \tag{34}$$

$$h_{\mu,Z,\text{AMISE.I}} = \left(\frac{\mathcal{I}_{\mu,XX}}{\mathcal{I}_{\mu,ZZ}}\right)^{1/4} h_{\mu,X,\text{AMISE.I}} \tag{35}$$

which corresponds to the result in Herrmann et al. (1995). The AMISE. I optimal bandwidths for the covariance function are given by

$$h_{\gamma,X,\text{AMISE.I}} = \left(\frac{R(K_\gamma)\ Q_{\gamma,1}\ 4\,\sqrt{2}\,\mathcal{I}_{\gamma,ZZ}^{3/2}}{TN\ (\nu_2(K_\gamma))^2\ \left(2\ (\nu_2(K_\gamma))^2\ \mathcal{I}_{\gamma,X_1Z} + C_{\mathcal{I}}\right)\ (C_{\mathcal{I}} - \mathcal{I}_{\gamma,X_1Z})^{3/2}}\right)^{1/7} \tag{36}$$

$$h_{\gamma,Z,\text{AMISE.I}} = \left(\frac{C_{\mathcal{I}} - \mathcal{I}_{\gamma,X_1Z}}{2\,\mathcal{I}_{\gamma,ZZ}}\right)^{1/2} h_{\gamma,X,\text{AMISE.I}}, \tag{37}$$

where $C_{\mathcal{I}} = \left(\mathcal{I}_{\gamma,X_1Z}^2 + 4\,(\mathcal{I}_{\gamma,X_1X_1} + \mathcal{I}_{\gamma,X_1X_2})\,\mathcal{I}_{\gamma,ZZ}\right)^{1/2}$. Obviously, the expressions in (36) and (37) are much less readable than those in (34) and (35). This is the burden of a two times higher dimension when estimating γ instead of μ.

2.2. AMISE. II *optimal bandwidth selection*

The AMISE. II expression of the three-dimensional local linear estimator $\hat{\gamma}$ is given by

$$\begin{aligned}
\text{AMISE. II}_{\hat{\gamma}}\,(h_{\gamma,X}, h_{\gamma,Z}) &= \overbrace{(TN)^{-1}\, h_{\gamma,X}^{-2}\, h_{\gamma,Z}^{-1}\, R(K_\gamma)\, Q_{\gamma,1}}^{\text{2nd Order}} + \overbrace{T^{-1}\, h_{\gamma,Z}^{-1}\, R(\kappa)\, Q_{\gamma,2}}^{\text{1st Order}} + \qquad (38) \\
&+ \frac{1}{4}\,(\nu_2(K_\gamma))^2 \left[\underbrace{2\, h_{\gamma,X}^4\,(\mathcal{I}_{\gamma,X_1X_1} + \mathcal{I}_{X_1X_2})}_{\text{3rd Order}} + \underbrace{4\, h_{\gamma,X}^2\, h_{\gamma,Z}^2\, \mathcal{I}_{\gamma,X_1Z}}_{\text{2nd Order}} + \underbrace{h_{\gamma,Z}^4\, \mathcal{I}_{\gamma,ZZ}}_{\text{1st Order}}\right],
\end{aligned}$$

where $Q_{\gamma,2} = \int_{S_{XXZ}} \tilde{\gamma}((x_1,x_2),(x_1,x_2),z)\, f_{XX}(x_1,x_2)\, d(x_1,x_2,z)$ and all other quantities are defined in the preceding section below of Eq. (33).

The lowest possible AMISE value under the AMISE. II scenario can be achieved if there exists a X-bandwidth which, first, allows us to profit from the (partial) annulment of the classical bias-variance tradeoff, but, second, assures that the AMISE. II scenario remains maintained. The first requirement is achieved if the X-bandwidth is of a smaller order of magnitude than the Z-bandwidth, i.e., if $h_{\gamma,X} = o(h_{\gamma,Z})$. This restriction makes those bias components that depend on $h_{\gamma,X}$ asymptotically negligible, since it implies that $h^2_{\gamma,X}\,h^2_{\gamma,Z} = o(h^4_{\gamma,Z})$ and therefore that $h^4_{\gamma,X} = o(h^2_{\gamma,X}\,h^2_{\gamma,Z})$. The latter leads to the order relations between the third, fourth, and fifth AMISE. II term as indicated in Eq. (38). The second requirement is achieved if the X-bandwidth does not converge to zero too fast, namely if $(Nh^2_{\gamma,X})^{-1} = o(1)$, which implies the order relation between the first two AMISE. II terms as indicated in Eq. (38).

2.3. Global polynomial fits

We suggest approximating the unknown quantities of the AMISE. I optimal bandwidth expressions in Eqs. (34), (35), (36), and (37) and those of the AMISE. II optimal bandwidth expressions in Eqs. (15), (16), (13), and (14) using five different global polynomial models of order four referred to as: μ_{poly}, γ_{poly}, $\tilde{\gamma}_{\text{poly}}$, $[\gamma(x,x,z)+\sigma^2_\epsilon]_{\text{poly}}$, and $[\tilde{\gamma}((x_1,x_2),(x_1,x_2),z)+\sigma^2_\varepsilon(x_1,x_2,z)]_{\text{poly}}$. Given estimates of these polynomial models, allows us to approximate the unknown quantities $\mathcal{I}_{\mu,XX}$, $\mathcal{I}_{\mu,XZ}$, $\mathcal{I}_{\mu,ZZ}$, $Q_{\mu,1}$, $Q_{\mu,2}$, $\mathcal{I}_{\gamma,X_1X_1}$, $\mathcal{I}_{\gamma,X_1Z}$, $\mathcal{I}_{\gamma,ZZ}$, $Q_{\gamma,1}$, and $Q_{\gamma,2}$ by the empirical versions of

$$
\begin{aligned}
\mathcal{I}_{\mu_{\text{poly}},XX} &= \int_{S_{XZ}} (\mu^{(2,0)}_{\text{poly}}(x,z))^2\, w_\mu\, d(x,z),\\
\mathcal{I}_{\mu_{\text{poly}},XZ} &= \int_{S_{XZ}} \mu^{(2,0)}_{\text{poly}}(x,z)\mu^{(0,2)}_{\text{poly}}(x,z)\, w_\mu\, d(x,z),\\
\mathcal{I}_{\mu_{\text{poly}},ZZ} &= \int_{S_{XZ}} (\mu^{(0,2)}_{\text{poly}}(x,z))^2\, w_\mu\, d(x,z),\\
Q_{\mu_{\text{poly}},1} &= \int_{S_{XZ}} [\gamma(x,x,z)+\sigma^2_\epsilon]_{\text{poly}}\, d(x,z),\\
Q_{\mu_{\text{poly}},2} &= \int_{S_{XZ}} \gamma_{\text{poly}}(x,x,z)\, f_X(x)\, d(x,z),\\
\mathcal{I}_{\gamma_{\text{poly}},X_1X_1} &= \int_{S_{XZ}} (\gamma^{(2,0,0)}_{\text{poly}}(x,x,z))^2\, w_\mu\, d(x,z),\\
\mathcal{I}_{\gamma_{\text{poly}},X_1Z} &= \int_{S_{XXZ}} \gamma^{(2,0,0)}_{\text{poly}}(x,x,z)\gamma^{(0,0,2)}_{\text{poly}}(x,x,z)\, w_\gamma\, d(x,x,z),\\
\mathcal{I}_{\gamma_{\text{poly}},ZZ} &= \int_{S_{XXZ}} (\gamma^{(0,0,2)}_{\text{poly}}(x,x,z))^2\, w_\gamma\, d(x,x,z),\\
Q_{\gamma_{\text{poly}},1} &= \int_{S_{XXZ}} [\tilde{\gamma}((x_1,x_2),(x_1,x_2),z)+\sigma^2_\varepsilon]_{\text{poly}}\, d(x_1,x_2,z),\\
Q_{\gamma_{\text{poly}},2} &= \int_{S_{XXZ}} \tilde{\gamma}_{\text{poly}}((x_1,x_2),(x_1,x_2),z)\; f_{XX}(x_1,x_2)\, d(x_1,x_2,z).
\end{aligned}
$$

Remember that the weight functions $w_\mu = w_\mu(x,z)$ and $w_\gamma = w_\gamma(x_1,x_2,z)$ are defined as $w_\mu(x,z) = f_{XZ}(x,z)$ and $w_\gamma(x_1,x_2,z) = f_{XXZ}(x_1,x_2,z)$. Estimates for these densities are constructed from kernel density estimates of the single pdfs f_X and f_Z, where we use the Epanechnikov kernel and the bandwidth selection procedure of Sheather & Jones (1991). Note that it is necessary to estimate the models μ_{poly} and γ_{poly} with interactions, since otherwise their partial derivatives would degenerate – all necessary further details are found in the following list:

- The model μ_{poly} is fitted via regressing Y_{it} on powers (each up to the fourth power) of X_{it}, Z_t, and $X_{it}\cdot Z_t$ for all i and t.

- The model γ_{poly} is fitted via regressing $C^{\text{poly}}_{ijt} = (Y_{it} - \mu_{\text{poly}}(X_{it}, Z_t))(Y_{jt} - \mu_{\text{poly}}(X_{jt}, Z_t))$ on powers (each up to the fourth power) of X_{it}, X_{jt}, Z_t, $X_{it} \cdot Z_t$, and $X_{jt} \cdot Z_t$ for all t and all i, j with $i \neq j$.
- The model $\tilde{\gamma}_{\text{poly}}$ is fitted via regressing $\mathbb{C}^{\text{poly}}_{(ij),(kl),t} = (C^{\text{poly}}_{ijt} - \gamma_{\text{poly}}(X_{it}, X_{jt}, Z_t))(C^{\text{poly}}_{klt} - \gamma_{\text{poly}}(X_{kt}, X_{lt}, Z_t))$ on powers (each up to the fourth power) of X_{it}, X_{jt}, X_{kt}, X_{lt}, and Z_t for all t and all i, j, k, l with $(i, j) \neq (k, l)$.
- The model $[\gamma(x, x, z) + \sigma^2_\epsilon]_{\text{poly}}$ is fitted via regressing the noise contaminated diagonal values C^{poly}_{iit} on powers (each up to the fourth power) of X_{it}, and Z_t for all i and t.
- The model $[\tilde{\gamma}((x_1, x_2), (x_1, x_2), z) + \sigma^2_\varepsilon(x_1, x_2, z)]_{\text{poly}}$ is fitted via regressing the noise contaminated diagonal values $\mathbb{C}^{\text{poly}}_{(ij),(ij),t}$ on powers (each up to the fourth power) of X_{it}, X_{jt}, and Z_t for all i, j, and t.

For the computation of the Bonferroni type confidence intervals we also need to approximate the error variance σ^2_ϵ. This is done via the empirical version of:

$$\sigma^2_{\text{poly},\epsilon} = \frac{1}{\int_{S_{XZ}} 1 d(x, z)} \int_{S_{XZ}} \left([\gamma(x, x, z) + \sigma^2_\epsilon]_{\text{poly}} - \gamma_{\text{poly}}(x, x, z) \right) d(x, z). \tag{39}$$

Of course, it is always possible (and often necessary) to use more complex polynomial models that include more interaction terms and higher order polynomials. Though, due to the relatively simple structured data this rule-of-thumb method works very well. The global polynomial fits of the mean functions are shown in Figure 6. Their general shapes are plausible and we can expect them to be very useful in approximating the above unknown quantities, though these pilot estimates are not perfect substitutes for the final local linear estimates. Particularly, the temperature effects in the second time period show some implausible wave shapes, which do not show up in the nonparametric fit shown in Figure 5.

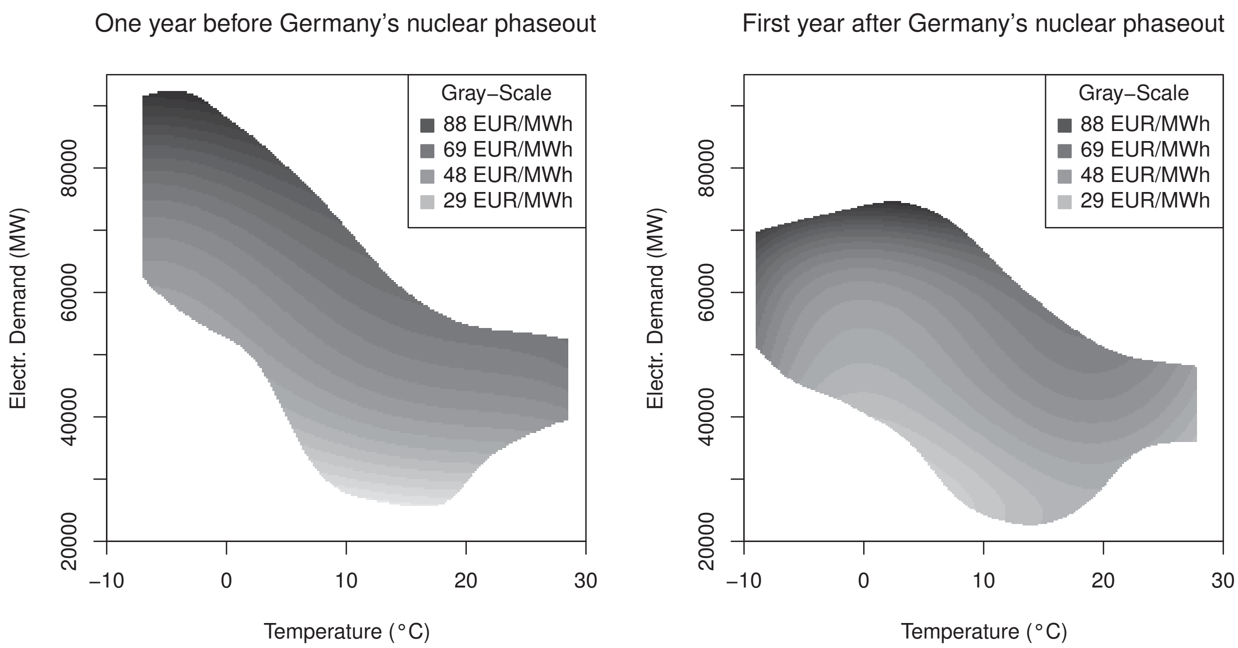

Fig. 6. Approximated mean functions using global polynomial regression models.

3. Data Sources

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

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