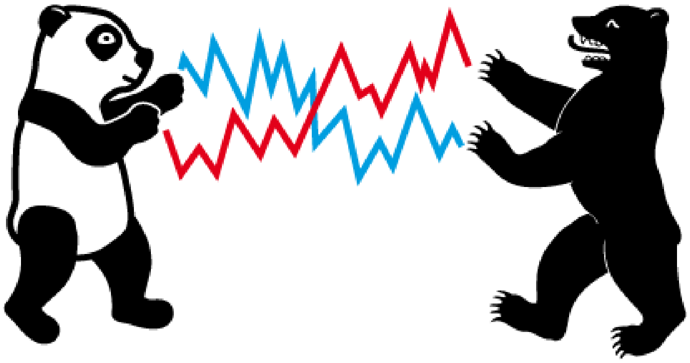
IRTG 1792 Discussion Paper 2020-008

Simultaneous Inference of the Partially Linear Model with a Multivariate Unknown Function

Kun Ho Kim *
Shih-Kang Chao *2
Wolfgang K. Härdle *3 *4



* Yeshiva University, USA
*2 University of Missouri, USA
*3 Humboldt-Universität zu Berlin, Germany
*4 Singapore Management University, Singapore

This research was supported by the Deutsche Forschungsgesellschaft through the International Research Training Group 1792 "High Dimensional Nonstationary Time Series".

http://irtg1792.hu-berlin.de
ISSN 2568-5619

International Research Training Group 1792

Simultaneous Inference of the Partially Linear Model with a Multivariate Unknown Function*

Kun Ho Kim† Shih-Kang Chao‡ Wolfgang K. Härdle§

May 5, 2020

Abstract

In this paper, we conduct simultaneous inference of the non-parametric part of a partially linear model when the non-parametric component is a multivariate unknown function. Based on semi-parametric estimates of the model, we construct a simultaneous confidence region of the multivariate function for simultaneous inference. The developed methodology is applied to perform simultaneous inference for the U.S. gasoline demand where the income and price variables are contaminated by Berkson errors. The empirical results strongly suggest that the linearity of the *U.S.* gasoline demand is rejected. The results are also used to propose an alternative form for the demand.

Key words: Simultaneous inference, Multivariate function, Simultaneous confidence region, Berkson error, Regression calibration.
JEL Classifications: C12, C13, C14

*The authors gratefully acknowledge financial support from the Deutsche Forschungsgemeinschaft through the International Research Training Group IRTG 1792 "High Dimensional Non Stationary Time Series" and the Collaborative Research Center CRC 649 "Economic Risk".

†Katz School, Yeshiva University, New York, NY 10016, U.S.A. E-mail: kunho.kim@yu.edu. Tel: +1 646 592 4717.

‡Department of Statistics, University of Missouri, MO 65211, U.S.A. E-mail: chaosh@missouri.edu. Tel: +1 573 882-1577. Partially supported by Office of Naval Research (ONR N00014-15-1-2331).

§Ladislaus von Bortkiewicz Chair of Statistics, C.A.S.E. - Center for applied Statistics and Economics, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany. Email: haerdle@wiwi.hu-berlin.de. Sim Kee Boon Institute for Financial Economics, Singapore Management University, 50 Stamford Road, Singapore 178899, Singapore.

1 Introduction

Partially linear models are welcome compromise between a pure nonparametric and a sometimes too restricted parametric specification. The semi-parametric structure not only makes it possible to include discrete predictors, but also to estimate part of the model with high precision. These and other reasons have made this model class very successful, e.g. Härdle et al. (2000). A typical assumption in the existing partially linear model literature is that the non-parametric part is univariate. In several applications though, one has a data structure described by:

$$Y_i = \mathbf{Z}_i^\top \boldsymbol{\beta} + \mu(\mathbf{X}_i) + \epsilon_i \tag{1.1}$$

where Y_i is a scalar random variable, $\mathbf{Z}_i$ is a $(\ell \times 1)$ random vector, and $\mathbf{X}_i$ is a $(d \times 1)$ random vector for $i = 1, \cdots, n$, respectively. In addition, ϵ_i is a mean zero IID random error that is uncorrelated with $\mathbf{Z}_i$ and $\mathbf{X}_i$. Here $\boldsymbol{\beta}$ and $\mu(\cdot)$ are a $(\ell \times 1)$ vector of unknown parameters and an unknown smooth function, respectively. Inference of the unknown function $\mu(\cdot)$ can be conducted even when the covariate terms on the RHS of (1.1) are not fully observed, as illustrated in Secion 3. The model (1.1) is widely used due to its flexibility to combine the parametric linear part $\mathbf{Z}_i^\top \boldsymbol{\beta}$ and the non-parametric non-linear component $\mu(\mathbf{X}_i)$. See Härdle et al. (2000) for more on the partially linear model framework.

The primary contribution of this paper is to introduce a methodology for simultaneous inference of the *multivariate* function $\mu(\cdot)$ in (1.1) when $d \geq 2$. The majority of the literature on (1.1) and on its variants has focused on simultaneous inference for a "univariate" function (i.e. $d = 1$): Johnston (1982) conducts simultaneous inference for an univariate mean regression function. Härdle (1989) derives simultaneous confidence bands (SCB) for one-dimensional kernel M-estimators. Fan and Zhang (2000) and Zhou and Wu (2010) show how to perform simultaneous inference of linear models with varying coefficients. Wu and Zhao (2007) and Kim (2016) work on inference of univariate time trend in mean regression. Zhao and Wu (2008) and Liu and Wu (2010) conduct simultaneous inference of the univariate mean and univariate volatility functions of a discretized version of the stochastic diffusion model. Moreover, Härdle and Song (2010) and Guo and Härdle (2012) construct uniform confidence bands for conditional quantile and expectile functions, respectively. Song et al. (2012) employ bootstrap procedures for local constant quantile estimators to overcome the slow convergence of asymptotic confidence bands. Although they contribute to the literature, all of the mentioned papers deal with the univariate function case. Chao et al. (2017) extend it to the case of multivariate quantile regression functions. However, to the best of our knowledge, no paper has worked on simultaneous inference of $\mu(\cdot)$ with $d \geq 2$ in the partially linear model (1.1). This paper attempts to undertake the task. The main results that we obtain in this paper have some resemblance to those in Chao et al. (2017). However, due to the partially linear structure in (1.1), the main assumptions of this paper are based on the conditional distribution of the filtered response $Y - \mathbf{Z}^\top \boldsymbol{\beta}$ on $\mathbf{X}$, rather than on the conditional distribution of Y on all the covariate terms, which is technically more challenging to handle than in Chao et al. (2017).

Simultaneous inference of (1.1) is conducted through the construction of simultaneous confidence region (SCR). Consider testing the following hypotheses for (1.1):

$$H_0 : \mu(\cdot) = \mu_\theta(\cdot) \qquad \text{versus} \qquad H_1 : H_0 \text{ is not true.} \tag{1.2}$$

where $\mu_\theta(\cdot)$ is a multivariate "parametric" function suggested by related economic theory. To test the hypothesis in (1.2), we construct the SCR of $\mu(\cdot)$ and observe whether the SCR contains the parametric specification "entirely". The construction of the SCR with confidence level $100(1-\alpha)\%$, $\alpha \in (0,1)$, requires us to find two functions $f_n(\cdot)$ and $g_n(\cdot)$ based on data, such that:

$$\lim_{n\to\infty} \mathbb{P}\{f_n(\boldsymbol{x}) \leq \mu(\boldsymbol{x}) \leq g_n(\boldsymbol{x}) \text{ for all } \boldsymbol{x} \in \mathcal{X}\} = 1 - \alpha \tag{1.3}$$

where $\boldsymbol{x}$ is a $(d \times 1)$ vector in a compact set $\mathcal{X}$. Given the SCR of $\mu(\cdot)$, one can test (1.2) by checking whether or not the condition $f_n(\boldsymbol{x}) \leq \mu_\theta(\boldsymbol{x}) \leq g_n(\boldsymbol{x})$ holds for *all* $\boldsymbol{x} \in \mathcal{X}$. If the condition does not hold for *some* $\boldsymbol{x} \in \mathcal{X}$, then we *reject* the null hypothesis at level α. That is, even if the condition holds for all $\boldsymbol{x} \in \mathcal{X}$ except for only one, the null hypothesis still gets rejected by the test.

The relative advantage of the SCR-based inference over other standard inferential procedures utilizing some integrated-squared-difference type statistic and its associated p-value such as those in (Härdle and Mammen (1993)) is in its effectiveness in suggesting the right function form of $\mu(\cdot)$ in (1.1). In case the null hypothesis in (1.2) is somehow rejected, it would be rather difficult to figure out the reason for rejection if the inference is based on such standard test statistic and its associated p-value. Thus, it would not be straightforward to suggest an alternative to the parametric function under the null hypothesis in such as case. However, if the inference is based on the proposed SCR, one can easily figure out the reason for the rejection by "locating graphically" where the SCR is violated by the parametric null $\mu_\theta(\cdot)$. Hence it would be relatively straightforward to propose an alternative parametric form. For more on this, see the last paragraph of Section 3.4.

The proposed methodology is applied to perform simultaneous inference of the U.S. gasoline demand. The gasoline demand is of interest to many, including policy makers, due to its environmental consequences and the role as an economic indicator. With little guidance on the form of demand function provided by economic theory (Blundell et al. (2012)), however, we refer to the semi-parametric demand structure (Schmalensee and Stoker (1999); Yatchew and No (2001); Blundell et al. (2012)), which is a specification of the general one provided by (1.1). A popular candidate for the function $\mu(\cdot)$ in (1.1) is the *log-linear* structure. The linear structure is widely used because it provides a simple but useful analytical framework and because the coefficients under the framework represent important structural parameters such as income/price elasticities. Despite these advantages, any parametric forms including the linear one are essentially arbitrary and may be misspecified in ways that produce seriously erroneous results. Hence we propose to test its validity as $\mu(\cdot)$ in (1.1) through the simultaneous inference proposed in this work.

The organization of the paper is the following: Section 2 introduces the methodology proposed to perform simultaneous inference of the partially linear model (1.1) with $d \geq 2$. We estimate the partially linear model and carry out the construction of simultaneous confidence region (SCR) based on the estimate. Both the asymptotic-based and simulation-based constructions of SCR are introduced. Section 3 handles an application of the proposed methodology. We estimate and perform simultaneous inference of a semi-parametric and partially linear U.S. gasoline demand under the Berkson errors. The data are explained and the empirical results are discussed in detail as well. Section 4 concludes the paper and discusses related future research. The mathematical proofs regarding the simultaneous inference for a multivariate function are relegated to Appendix.

Notations. For any vector $\boldsymbol{v} = (v_1, v_2, \ldots, v_p) \in \mathbb{R}^p$, we let $|\boldsymbol{v}| = \left(\sum_{i=1}^p v_i^2\right)^{1/2}$. For any random vector $\mathbf{V}$, we write $\mathbf{V} \in \mathcal{L}^q$ $(q > 0)$ if $\|\mathbf{V}\|_q = \left[\mathbb{E}\left(|\mathbf{V}|^q\right)\right]^{1/q} < \infty$. In particular, $\|\mathbf{V}\| = \|\mathbf{V}\|_2$. In addition, we write $a_n \asymp b_n$ if $|a_n/b_n|$ is bounded away from 0 and ∞ for all large n. For brevity, we sometimes write $\sup_{\boldsymbol{x}} U(\boldsymbol{x})$ for $\sup_{\boldsymbol{x} \in \mathcal{X}} U(\boldsymbol{x})$.

2 Methodology

2.1 Assumptions

Let $Y_i^* \stackrel{\text{def}}{=} Y_i - \mathbf{Z}_i^\top \boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is the true coefficient in the model (1.1). Denote h as the bandwidth. Without loss of generality, we assume equal bandwidth for all directions of $\boldsymbol{x}$. The assumptions for the theoretic results of this study are the following:

(A1) K is a kernel function of order $s-1$ (see (A3)) with bounded support $[-A, A]^d$, and is continuously differentiable up to order d with bounded derivatives, i.e. $\partial^{\boldsymbol{\alpha}} K = \partial^{\alpha_1} \partial^{\alpha_2} \ldots \partial^{\alpha_d} K \in L^1(\mathbb{R}^d)$ exists and is continuous for all multi-indices $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)^\top \in \{0, 1\}^d$.

(A2) Given $a_n \asymp (h^{-3d}\log n)^{1/(b_1-2)} \to \infty$ for some $b_1 > 2$, assume that the conditional density $f_{Y^*|\boldsymbol{X}}(y^*|\boldsymbol{x})$ satisfies

(i) $0 < \sup_{\boldsymbol{x}\in\mathcal{X}} \left|\int \left|y^* - \mu(\boldsymbol{x})\right|^{b_1} f_{Y^*|\boldsymbol{X}}(y^*|\boldsymbol{x})dv\right| < \infty$.

(ii) There exists $C^* > 0$ such that

$$\left(h^{-3d}\log n \sup_{\boldsymbol{x}\in\mathcal{X}} \int_{\{|y^*|>a_n\}} y^{*2} f_{Y^*|\boldsymbol{X}}(y^*|\boldsymbol{x})dy^*\right)^{1/2} \leq C^*. \tag{2.1}$$

(iii) $n^{-1/6}h^{-d/2}a_n = \mathcal{O}(n^{-\nu})$, for some constant $\nu > 0$.

(A3) The function $\mu(\boldsymbol{x})$ is in Hölder class with order $s > d$.

(A4) The density $f_{\mathbf{X}}(\boldsymbol{x})$ of $\mathbf{X}$ is bounded, continuously differentiable and its gradient is uniformly bounded. Moreover, $\inf_{\boldsymbol{x}\in\mathcal{X}} f_{\boldsymbol{X}}(\boldsymbol{x}) > 0$ for domain $\mathcal{X}$.

(A5) The joint probability density function $f(y^*, \boldsymbol{x})$ is bounded and continuously differentiable up to sth order (needed for the Rosenblatt transform). The conditional density $f_{Y^*|\boldsymbol{X}}(y^*|\boldsymbol{x})$ is bounded and continuouly differentiable with respect to $\boldsymbol{x}$.

(A6) h satisfies $\sqrt{nh^d}h^s\sqrt{\log n} \to 0$ (undersmoothing), and $nh^{3d}(\log n)^{-2} \to \infty$.

Assumption (A1) gives constraints on the kernel function, and is satisfied by popular kernels such as the Epanechnikov and quartic kernels. The moment condition of the model error (A2)(i), and the tail moment condition of the response variable (A2)(ii) are similar to those in the simultaneous confidence band literature such as Johnston (1982) and Härdle (1989). The condition (A2)(iii) is necessary for bounding the strong approximation error, which is given by Theorem 3.2 of Dedecker et al. (2014). (A3)-(A5) are adaptations of those in Johnston (1982) and Härdle (1989) to multivariate covariates. (A6) implies undersmoothing.

2.2 Simultaneous confidence region (SCR) of $\mu(\cdot)$

Estimation of $\mu(\cdot)$ in (1.1) is done by carrying out the following optimization:

$$\widehat{\mu}(\boldsymbol{x}) = \underset{\theta}{\operatorname{argmin}} \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{\boldsymbol{x} - \mathbf{X}_i}{h}\right) \left(Y_i - \mathbf{Z}_i^\top \widehat{\boldsymbol{\beta}}_R - \theta\right)^2 \tag{2.2}$$

where $\widehat{\boldsymbol{\beta}}_R$ is the Robinson estimate (Robinson, 1988) of $\boldsymbol{\beta}$ in (1.1). Here $K(\cdot)$ is a kernel function and h is a smoothing parameter, typically called the bandwidth. This leads to:

$$\widehat{\mu}(\boldsymbol{x}) \stackrel{\text{def}}{=} \frac{1}{nh^d \widehat{f}_{\boldsymbol{X}}(\boldsymbol{x})} \sum_{i=1}^{n} K\left(\frac{\boldsymbol{x} - \mathbf{X}_i}{h}\right) \left(Y_i - \mathbf{Z}_i^\top \widehat{\boldsymbol{\beta}}_R\right), \tag{2.3}$$

where $\widehat{f}_{\boldsymbol{X}}(\boldsymbol{x}) \stackrel{\text{def}}{=} \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{\boldsymbol{x} - \mathbf{X}_i}{h}\right)$ is a non-parametric estimate of $f_{\boldsymbol{X}}$, the joint density of $\mathbf{X}_i^\top = [X_{1i}, \cdots, X_{di}]$. A popular choice for the kernel function $K(\cdot)$ is the Epanechnikov kernel with a compact support. The bandwidth h is frequently selected by some data-driven method, such as the generalized cross-validation (GCV) (Craven and Wahba, 1979). In this study, the bandwidth chosen by GCV is adjusted downward to conduct the undersmoothing in (A6). To construct the SCR of $\mu(\cdot)$, we adopt the methodology in Härdle (1989), Johnston (1982) and extend it to the multi-dimensional case. First, consider the optimization (2.2) when $\boldsymbol{\beta}$ is *known*:

$$\widetilde{\mu}(\boldsymbol{x}) = \underset{\theta}{\operatorname{argmin}} \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{\boldsymbol{x} - \mathbf{X}_i}{h}\right) \left(Y_i - \mathbf{Z}_i^\top \boldsymbol{\beta} - \theta\right)^2. \tag{2.4}$$

We define

$$\begin{aligned}
\widetilde{H}_n(\theta, \boldsymbol{x}) &\stackrel{\text{def}}{=} \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{\boldsymbol{x} - \mathbf{X}_i}{h}\right) \left(Y_i - \mathbf{Z}_i^\top \boldsymbol{\beta} - \theta\right) \\
H_n(\boldsymbol{x}) &\stackrel{\text{def}}{=} \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{\boldsymbol{x} - \mathbf{X}_i}{h}\right) \left(Y_i - \mathbf{Z}_i^\top \boldsymbol{\beta} - \mu(\boldsymbol{x})\right)
\end{aligned}$$

Then,

$$\widetilde{H}_n(\widetilde{\mu}, \boldsymbol{x}) = H_n(\boldsymbol{x}) + \{\mu(\boldsymbol{x}) - \widetilde{\mu}(\boldsymbol{x})\} \widehat{f}_{\boldsymbol{X}}(\boldsymbol{x}).$$

Since $\widetilde{H}_n(\widetilde{\mu}, \boldsymbol{x}) = 0$ by the first-order condition $(F.O.C.)$ of (2.4), we have

$$\widetilde{\mu}(\boldsymbol{x}) - \mu(\boldsymbol{x}) = \frac{H_n(\boldsymbol{x})}{\widehat{f}_{\boldsymbol{X}}(\boldsymbol{x})}$$

which leads to

$$\widetilde{\mu}(\boldsymbol{x}) - \mu(\boldsymbol{x}) = \frac{H_n(\boldsymbol{x})}{f_{\boldsymbol{X}}(\boldsymbol{x})} + \frac{H_n(\boldsymbol{x})\{f_{\boldsymbol{X}}(\boldsymbol{x}) - \widehat{f}_{\boldsymbol{X}}(\boldsymbol{x})\}}{f_{\boldsymbol{X}}(\boldsymbol{x})\widehat{f}_{\boldsymbol{X}}(\boldsymbol{x})}.$$

In sum,

$$\widetilde{\mu}(\boldsymbol{x}) - \mu(\boldsymbol{x}) = \frac{H_n(\boldsymbol{x}) - \mathbb{E}H_n(\boldsymbol{x})}{f_{\boldsymbol{X}}(\boldsymbol{x})} + R_n(\boldsymbol{x}) \tag{2.5}$$

where $R_n(\boldsymbol{x}) \overset{\text{def}}{=} \frac{\mathbb{E}H_n(\boldsymbol{x})}{f_{\boldsymbol{X}}(\boldsymbol{x})} + \frac{H_n(\boldsymbol{x})\left\{f_{\boldsymbol{X}}(\boldsymbol{x}) - \widehat{f}_{\boldsymbol{X}}(\boldsymbol{x})\right\}}{f_{\boldsymbol{X}}(\boldsymbol{x})\widehat{f}_{\boldsymbol{X}}(\boldsymbol{x})}$. Then, by (2.5),

$$\frac{\sqrt{nh^d f_{\boldsymbol{X}}(x)}}{\sigma(\boldsymbol{x})} |\widetilde{\mu}(\boldsymbol{x}) - \mu(\boldsymbol{x})| = |U_n(\boldsymbol{x})| + \widetilde{R}_n(\boldsymbol{x}) \tag{2.6}$$

where

$$\begin{aligned}
\widetilde{R}_n(\boldsymbol{x}) &= \left| U_n(\boldsymbol{x}) + \frac{\sqrt{nh^d f_{\boldsymbol{X}}(\boldsymbol{x})}}{\sigma(\boldsymbol{x})} R_n(\boldsymbol{x}) \right| - |U_n(\boldsymbol{x})| && (2.7)\\
\sigma^2(\boldsymbol{x}) &\overset{\text{def}}{=} \int \{y^* - \mu(\boldsymbol{x})\}^2 f_{Y^*|\mathbf{X}}(y^*|\boldsymbol{x})\, dy^* && (2.8)\\
U_n(\boldsymbol{x}) &\overset{\text{def}}{=} \sqrt{\frac{nh^d}{\sigma^2(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})}} (H_n(\boldsymbol{x}) - \mathbb{E}H_n(\boldsymbol{x})) && (2.9)
\end{aligned}$$

Here $y^* \stackrel{\text{def}}{=} y - \boldsymbol{z}^\top \boldsymbol{\beta}$. Let $\sigma = \sigma(\boldsymbol{x})$ for simplicity. Then, after some elementary calculations, we obtain:

$$U_n(\boldsymbol{x}) = \frac{1}{\sqrt{h^d \sigma^2 f_{\boldsymbol{X}}(\boldsymbol{x})}} \int\int K\left(\frac{\boldsymbol{x}-\boldsymbol{t}}{h}\right)(y^* - \mu(\boldsymbol{x}))\, dZ_n(\boldsymbol{t}, y^*)$$

where $Z_n(\boldsymbol{x}, y^*) \stackrel{\text{def}}{=} \sqrt{n}\,(F_n(\boldsymbol{x}, y^*) - F(\boldsymbol{x}, y^*))$. Here $F_n(\cdot)$ is the empirical cumulative distribution function (c.d.f.) while $F(\cdot)$ is the true c.d.f. Moreover, we can define the following processes:

$$\begin{aligned}
U_{1,n}(\boldsymbol{x}) &= \frac{1}{\sqrt{h^d \sigma^2 f_{\boldsymbol{X}}(\boldsymbol{x})}} \int\int K\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{h}\right)(y^* - \mu(\boldsymbol{x}))\, dB_n\left\{T(y^*, \boldsymbol{u})\right\};\\
U_{2,n}(\boldsymbol{x}) &= \frac{1}{\sqrt{h^d \sigma^2 f_{\boldsymbol{X}}(\boldsymbol{x})}} \int\int K\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{h}\right)(y^* - \mu(\boldsymbol{x}))\, dW_n\left\{T(y^*, \boldsymbol{u})\right\};\\
U_{3,n}(\boldsymbol{x}) &= \frac{1}{\sqrt{h^d \sigma^2 f_{\boldsymbol{X}}(\boldsymbol{x})}} \int\int K\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{h}\right)(y^* - \mu(\boldsymbol{u}))\, dW_n\left\{T(y^*, \boldsymbol{u})\right\};\\
U_{4,n}(\boldsymbol{x}) &= \frac{1}{\sqrt{h^d \sigma^2 f_{\boldsymbol{X}}(\boldsymbol{x})}} \int \sigma f_{\boldsymbol{X}}^{1/2}(\boldsymbol{u}) K\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{h}\right) dW(\boldsymbol{u});\\
U_{5,n}(\boldsymbol{x}) &= h^{-d/2} \int K\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{h}\right) dW(\boldsymbol{u}),
\end{aligned}$$

where $\{B_n\}$ are a sequence of Brownian bridges and $\{W_n\}$ are a sequence of Wiener processes satisfying $B_n\{T(y, \boldsymbol{u})\} = W_n\{T(y, \boldsymbol{u})\} - F(y, \boldsymbol{u}) W_n(1, ..., 1)$. Here $W(\cdot)$ is the Wiener process and $T(y, \boldsymbol{x})$ is the d dimensional Rosenblatt transformation (Rosenblatt, 1976):

$$T(y, \boldsymbol{u}) = \left\{F_{X_1|Y}(u_1|y), F_{X_2|Y}(u_2|u_1, y), ..., F_{X_d|X_{d-1},...,X_1,Y}(u_d|u_{d-1}, ..., u_1, y), F_Y(y)\right\} \tag{2.10}$$

Then, from Theorem 2 of Rosenblatt (1976), suppose the volume of $\mathcal{X}$ is 1

$$\mathbb{P}\left\{\frac{\sqrt{2d\log(h^{-1})}}{\lambda_K}\left(\sup_{\boldsymbol{x}\in\mathcal{X}} |U_{5,n}(\boldsymbol{x})| - d_n\right) < u\right\} \rightarrow \exp\left(-2\exp(-u)\right) \tag{2.11}$$

where $\lambda_K = (\int_{\mathbb{R}} K^2(u)du)^{1/2}$ and $d_n = \sqrt{2d\log(h^{-1})} + \frac{1}{\sqrt{2d\log(h^{-1})}}\Big(\frac{1}{2}(d-1)\log\log(h^{-1}) + \log\Big(\big(\frac{2d}{\pi}\big)^{d/2}\sqrt{\frac{\det(\Sigma)}{4d\pi}}\Big)\Big)$. This leads to the following proposition:

Proposition 2.1. *Under Assumptions (A1)-(A6), suppose that the volume of $\mathcal{X}$ is 1,*

$$\mathbb{P}\left\{\frac{\sqrt{2d\log(h^{-1})}}{\lambda_K}\left(\sup_{\boldsymbol{x}\in\mathcal{X}}|U_n(\boldsymbol{x})| - d_n\right) < u\right\} \quad\rightarrow\quad \exp\left(-2\exp(-u)\right) \tag{2.12}$$

Proof of Proposition 2.1. The proof follows directly by Lemma B.3– Lemma B.8 in Appendix. □

Moreover, by Lemma A.1 in the Appendix,

$$\sup_{\boldsymbol{x}\in\mathcal{X}}|R_n(\boldsymbol{x})| = o_{\mathbb{P}}\left(\frac{1}{\sqrt{nh^d\log(h^{-1})}}\right) \tag{2.13}$$

Then, by (2.6), (2.12) and (2.13), we obain the following theorem:

Theorem 2.2. *Under Assumptions (A1)-(A6), suppose that the volume of $\mathcal{X}$ is 1,*

$$\mathbb{P}\left\{\frac{\sqrt{2d\log(h^{-1})}}{\lambda_K}\left(\sup_{\boldsymbol{x}\in\mathcal{X}}\frac{\sqrt{nh^d f_{\boldsymbol{X}}(\boldsymbol{x})}}{\sigma(\boldsymbol{x})}|\widetilde{\mu}(\boldsymbol{x}) - \mu(\boldsymbol{x})| - d_n\right) < u\right\} \\ \rightarrow\quad \exp\left(-2\exp(-u)\right) \tag{2.14}$$

where $\widetilde{\mu}(\boldsymbol{x})$ is the infeasible estimate of $\mu(\boldsymbol{x})$ defined by (2.4).

Proof of Theorem 2.2. By (2.6),

$$\begin{aligned}&\frac{\sqrt{2d\log(h^{-1})}}{\lambda_K}\left(\sup_{\boldsymbol{x}\in\mathcal{X}}\frac{\sqrt{nh^d f_{\boldsymbol{X}}(\boldsymbol{x})}}{\sigma(\boldsymbol{x})}|\widetilde{\mu}(\boldsymbol{x}) - \mu(\boldsymbol{x})| - d_n\right)\\ =\ &\frac{\sqrt{2d\log(h^{-1})}}{\lambda_K}\left(\sup_{\boldsymbol{x}\in\mathcal{X}}|U_n(\boldsymbol{x})| - d_n\right) + \bar{R}_n,\end{aligned}$$

where by using triangle inequality,

$$
\begin{aligned}
|\bar{R}_n| &= \frac{\sqrt{2d\log(h^{-1})}}{\lambda_K}\Big| \sup_{\boldsymbol{x}\in\mathcal{X}} \big||U_n(\boldsymbol{x})| + \widetilde{R}_n(\boldsymbol{x})\big| - \sup_{\boldsymbol{x}\in\mathcal{X}} |U_n(\boldsymbol{x})|\Big| \\
&\leq \sup_{\boldsymbol{x}\in\mathcal{X}} \frac{\sqrt{2d\log(h^{-1})}}{\lambda_K} \sup_{\boldsymbol{x}\in\mathcal{X}} |\widetilde{R}_n(\boldsymbol{x})| \\
&\leq \sup_{\boldsymbol{x}\in\mathcal{X}} \frac{\sqrt{2df_{\boldsymbol{X}}(\boldsymbol{x})nh^d\log(h^{-1})}}{\lambda_K\sigma(\boldsymbol{x})} |R_n(\boldsymbol{x})|
\end{aligned}
$$

By (2.12) and (2.13), $\bar{R}_n = o_{\mathbb{P}}(1)$ and the theorem follows. □

Since the asymptotic result (2.14) involves the infeasible estimate $\widetilde{\mu}(\boldsymbol{x})$ that depends on the unknown $\boldsymbol{\beta}$, Theorem 2.2 cannot be directly applied to construct the *SCR* of the unknown $\mu(\cdot)$. However, the result can be extended to derive the following theorem that can be utilized to construct the *SCR* of $\mu(\cdot)$:

Theorem 2.3. *Suppose that the volume of $\mathcal{X}$ is 1. Under Assumptions (A1)-(A6),*

$$
\mathbb{P}\left\{\frac{\sqrt{2d\log(h^{-1})}}{\lambda_K}\left(\sup_{\boldsymbol{x}\in\mathcal{X}} \frac{\sqrt{nh^d f_{\boldsymbol{X}}(\boldsymbol{x})}}{\sigma(\boldsymbol{x})} |\widehat{\mu}(\boldsymbol{x}) - \mu(\boldsymbol{x})| - d_n\right) < u\right\} \\
\to \exp\left(-2\exp(-u)\right) \qquad (2.15)
$$

where $\widehat{\mu}(\boldsymbol{x})$ is the feasible estimate of $\mu(\boldsymbol{x})$ defined by (2.3).

Proof of Theorem 2.3. The proof follows by Theorem 2.2 and by the $\sqrt{n}$-consistency of $\widehat{\boldsymbol{\beta}}_R$ (Robinson, 1988). We omit the details. □

Remark 2.4. *If the volume $vol(\mathcal{X})$ is not 1, then replacing the $\log(h^{-1})$ in both Theorem 2.2 and 2.3 by $\log(vol(\mathcal{X})h^{-1})$.*

2.3 Implementation

By Theorem 2.3, the $(1-\alpha)\times 100\%$ SCR of $\mu(\boldsymbol{x})$ is

$$\left[\widehat{\mu}(\boldsymbol{x}) \pm \sqrt{\frac{\sigma^2(\boldsymbol{x})\lambda_K^2}{nh^d f_{\boldsymbol{X}}(\boldsymbol{x})}}\left(d_n + \frac{q_{1-\alpha}}{\sqrt{2d\log(h^{-1})}}\right)\right] \tag{2.16}$$

where $q_{1-\alpha} = -\log[-1/2\log(1-\alpha)]$ is the $(1-\alpha)\times 100\%$ quantile of the Gumbel distribution in (2.14). Here $q_{0.95} = 3.66$, $\det(\Sigma) = 25/16$ and $\lambda_K = 0.6$ for the Epanechnikov kernel. Since $\sigma^2(\boldsymbol{x})$ and $f_{\boldsymbol{X}}(\boldsymbol{x})$ in (2.16) are unknown, the feasible SCR is

$$\left[\widehat{\mu}(\boldsymbol{x}) \pm \sqrt{\frac{\widehat{\sigma}^2(\boldsymbol{x})\lambda_K^2}{nh^d \widehat{f}_{\boldsymbol{X}}(\boldsymbol{x})}}\left(d_n + \frac{q_{1-\alpha}}{\sqrt{2d\log(h^{-1})}}\right)\right] \tag{2.17}$$

where the estimate of $\sigma^2(\boldsymbol{x})$ is given by $\widehat{\sigma}^2(\boldsymbol{x}) \stackrel{\text{def}}{=} \frac{1}{nh^d\widehat{f}_{\boldsymbol{X}}(\boldsymbol{x})}\sum_{i=1}^n \widehat{\epsilon}_i^2 K\left(\frac{\boldsymbol{x}-\mathbf{X}_i}{h}\right)$ and $\widehat{\epsilon}_i$ is the residual from the estimation of (1.1).

The SCR (2.17) is based on the asymptotic distribution of the maximum of Gaussian processes. It is shown in Hall (1979) that the rate of convergence of the extreme of Gaussian processes in Kolmogorov distance is only of order $(\log n)^{-1}$, so the coverage error of the asymptotic SCR decays only logarithmically.

In order to obtain SCR with more accurate coverage probabilities, we employ the *simulation-based* method to construct the SCR of $\mu(\boldsymbol{x})$. First, we call from page 98–99 of Ferguson (1996):

$$\mathbb{P}\left(\sup_{1\le j\le m}|Z_j| - d_m \le \frac{u}{\sqrt{2\log(m)}}\right) \quad\rightarrow\quad \exp\left(-2\exp(-u)\right) \tag{2.18}$$

where $m \stackrel{\text{def}}{=} \inf\{k\in\mathbb{Z} : k \ge h^{-d}\}$ and Z_j are an IID standard normals and

$$d_m = \sqrt{2d\log(m)} - \frac{1}{\sqrt{2d\log(m)}}\left[\frac{1}{2}\log\{d\log(m)\} + \log\left(2\sqrt{\pi}\right)\right]$$

Note that (2.15) and (2.18) share the same asymptotic Gumbel distribution. The quantile of the distribution (2.18) can better approximate the quantile of the scaled $\widehat{\mu}(\boldsymbol{x})$ (as (2.15)) than the asymptotic Gumbel distribution. This is based on the fact that $\widehat{\mu}(\boldsymbol{x})$ and $\widehat{\mu}(\boldsymbol{x}')$ are *asymptotically independent* for any pairs $\boldsymbol{x} \neq \boldsymbol{x}'$ since $h^{-d} \to \infty$, so that the IID $\{Z_j\}_{j=1}^m$ has the same asymptotic distribution as the properly scaled $\{\widehat{\mu}(\boldsymbol{x})\}_{\boldsymbol{x}\in\mathcal{X}_h}$, where $\mathcal{X}_h$ is a grid with the grid size depending on h and $m = |\mathcal{X}_h|$.

We can approximate the quantile of $\sup_{1\le j\le m}|Z_j|$ to arbitrary accuracy by sampling $\{|Z_j|\}_{j=1}^m$ sufficiently many times. Thus, the $(1-\alpha)\times 100\%$ SCR of $\ell(x)$ is approximated by:

$$\left[\widehat{\mu}(\boldsymbol{x}) \pm \sqrt{\frac{\widehat{\sigma}^2(\boldsymbol{x})\lambda_K^2}{nh^d\widehat{f}_{\boldsymbol{X}}(\boldsymbol{x})}}\left(d_n + \frac{q_{1-\alpha}^{**}}{\sqrt{2d\log(h^{-1})}}\right)\right] \tag{2.19}$$

where $q_{1-\alpha}^{**} = (q^{**} - d_m)\sqrt{2\log(m)}$ and q^{**} is the $(1-\alpha)\times 100\%$ quantile of the sampling distribution of $\sup_{1\le j\le m}|Z_j|$. The same method has been applied in Zhao and Wu (2008) to obtain the SCR of an univariate function with weakly dependent data. In the following application, we employ (2.19) to construct the SCR of the multivariate function $\mu(\boldsymbol{x})$ in (1.1).

3 Application: Gasoline demand function

It is well-known that the key variables of gasoline demand, household income and gasoline price variables, are typically contaminated by errors since the exact amount of household income is rarely reported and the gasoline price is typically estimated as well. Specifically, the Berkson-error framework (Berkson (1950)) fits our data better than the classical measurement error because a "mid-point" of the income/price range applying to each household is used for the income/price variable. To that end, we consider the following error-in-variable (EIV) framework, a specification of (1.1):

$$Y_i = \mathbf{S}_i^\top\boldsymbol{\beta} + g(\mathbf{T}_i) + \zeta_i, \qquad \mathbf{S}_i = \mathbf{Z}_i + \xi_i, \quad \mathbf{T}_i = \mathbf{X}_i + \eta_i \tag{3.1}$$

where Y_i is a scalar random variable, $\mathbf{S}_i$ is a $(\ell \times 1)$ random vector, and $\mathbf{T}_i$ is a $(d \times 1)$ random vector for $i = 1, \cdots, n$, respectively. In addition, ζ_i is a mean zero IID random error for each i. Here $\boldsymbol{\beta}$ and $g(\cdot)$ are a $(\ell \times 1)$ vector of unknown parameters and an unknown smooth function, respectively. Let $\mathbf{S}$ and $\mathbf{T}$ be *unobserved* due to measurement errors ξ_i and η_i. However, $\mathbf{Z}_i$, $\mathbf{X}_i$ and Y_i are *observed*. The observed covariates $\mathbf{Z}_i$ and $\mathbf{X}_i$, error and measurement errors are mutually independent. Here ζ_i and η_i have zero means and finite variances, and ξ_i has zero mean and covariance matrix Σ_ξ, which does not have to be known.

The distribution of η_i in (3.1) is *not* needed for testing a *linear* hypothesis on $g(\cdot)$ although its distribution is needed in general for a *nonlinear* null function. See Section 3.2 for more on this. The covariates $(\mathbf{S}_i, \mathbf{T}_i)$ in (3.1) is said to be contaminated by the Berkson error, because $(\mathbf{Z}_i, \mathbf{X}_i)$ are unbiased estimators for $(\mathbf{S}_i, \mathbf{T}_i)$ as $\mathsf{E}[(\mathbf{Z}_i, \mathbf{X}_i) - (\mathbf{S}_i, \mathbf{T}_i)] = (0, 0)$. The Berkson error models are appropriate when the true individual observations are not available, but the "average" (unbiased estimators for the individual observations) for the group where the individuals belong is available. Typically, survey data are subject to the Berkson errors.

In (3.1), our primary goal is to test the hypothesis on $g(\cdot)$, such that:

$$H_{10} : g(\boldsymbol{x}) = g_0(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in \mathbb{R}^d \tag{3.2}$$

where $g_0(\cdot)$ is some known real-valued function. For instance, $g(\boldsymbol{x})$ could stand for a true demand function while $g_0(\boldsymbol{x})$ is a parametric *log-linear* demand function that is commonly used in the economics literature. That is, the validity of the widely used log-linear demand structure can be checked by testing the hypothesis (3.2) for the *EIV* model (3.1).

3.1 Regression calibration

In the literature on *EIV* models, *regression calibration* (Carroll et al., 1995) is widely used to deal with the error-contaminated covariate terms. Under the assumptions of (3.1),

$$\mathsf{E}(Y_i|\mathbf{Z}_i = \boldsymbol{z}, \mathbf{X}_i = \boldsymbol{x}) = \boldsymbol{z}^\top\boldsymbol{\beta} + \mu(\boldsymbol{x})$$

where $\mu(\boldsymbol{x}) \stackrel{\text{def}}{=} \mathsf{E}[g(\mathbf{T}_i)|\mathbf{X}_i = \boldsymbol{x}]$. Thus, we are led to the following *calibrated* partially linear regression model:

$$Y_i = \mathbf{Z}_i^\top\boldsymbol{\beta} + \mu(\mathbf{X}_i) + \epsilon_i \tag{3.3}$$

where $\epsilon_i \stackrel{\text{def}}{=} g(\mathbf{T}_i) - \mu(\mathbf{X}_i) + \xi_i^\top\boldsymbol{\beta} + \zeta_i$. Note here that $\mathsf{E}(\epsilon_i|\mathbf{Z}_i = \boldsymbol{z}, \mathbf{X}_i = \boldsymbol{x}) = 0$ due to $\mu(\boldsymbol{x}) = \mathsf{E}[g(\mathbf{T}_i)|\mathbf{X}_i = \boldsymbol{x}]$. That is, ϵ_i is *uncorrelated* with $\mathbf{Z}_i$ and $\mathbf{X}_i$. However, ϵ_i is still dependent on $\mathbf{X}_i$. The key difference between the original *EIV* model (3.1) and the transformed model (3.3) is that the covariate terms in (3.3) are observed while those in (3.1) are not. The transformation of (3.1) to (3.3) is called the *regression calibration*, and is an effective way to deal with the Berkson errors because the covariates in (3.3) are observed.

Moreover, under the Berkson errors in (3.1), the calibrated model (3.3) is free of the *endogeneity* issue because ϵ_i is uncorrelated with $\mathbf{Z}_i$ and $\mathbf{X}_i$. Otherwise, one needs to find instrument variables (*IV*) to conduct the estimation. Interestingly, the convenient uncorrelatedness between the error and covariate terms in (3.3) breaks down under the traditional classical-measurement-error setting because $\mathsf{E}(\xi_i|\mathbf{Z}_i = \boldsymbol{z}, \mathbf{X}_i = \boldsymbol{x}) \neq 0$ under the classical-measurement-error framework $\mathbf{Z}_i = \mathbf{S}_i + \xi_i$. Hence one would still need to find *IV*s to estimate the calibrated one in (3.3) under the classical measurement errors. That is, the regression calibration under the Berkson errors provides a convenient way to conduct inference of 3.3 (and eventually 3.1) without using any *IV*s.

As mentioned in Section 3.1, the distribution of η_i is assumed to be known in (3.1). Hence a test of H_{10} in (3.2) can be carried out by testing the following hypothesis for (3.3) instead:

$$H_{20} : \mu(\boldsymbol{x}) = \mu_0(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in \mathcal{X}, \tag{3.4}$$

where $\mu_0(\boldsymbol{x}) \overset{\text{def}}{=} \mathsf{E}[g_0(\mathbf{T}_i)|\mathbf{X}_i = \boldsymbol{x}]$ and $\mathcal{X}$ is the compact range of $\mathbf{X}_i$. The hypothesis H_{10} in (3.2) is, however, not equivalent to H_{20} in (3.4) in general situation, because $\mu(\boldsymbol{x}) = \mu_0(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathcal{X}$ only implies

$$\mathsf{E}[g(\mathbf{T}_i) - g_0(\mathbf{T}_i)|\mathbf{X}_i = \boldsymbol{x}] = 0, \quad \forall \boldsymbol{x} \in \mathcal{X},$$

which does not necessarily imply $g(\boldsymbol{x}) = g_0(\boldsymbol{x})$, $\forall \boldsymbol{x} \in \mathcal{X}$. To ensure the equivalence between (3.2) and (3.4), we require the family of densities $\{f_\eta(\cdot - \boldsymbol{x}) : \boldsymbol{x} \in \mathcal{X}\}$ for η_i to be *complete* in the following sense:

(C) [Completeness] For any measurable functions $h_0, h_1 : \mathbb{R}^d \to \mathbb{R}$, $\int h_0(\boldsymbol{v}) f_\eta(\boldsymbol{v} - \boldsymbol{x}) d\boldsymbol{v} = \int h_1(\boldsymbol{v}) f_\eta(\boldsymbol{v} - \boldsymbol{x}) dy$ for all $\boldsymbol{x} \in \mathcal{X}$ implies $h_0(\boldsymbol{x}) = h_1(\boldsymbol{x})$ almost everywhere (in Lebesgue measure) for all $\boldsymbol{x} \in \mathcal{X}$.

Condition (C) is satisfied, for example, when the density $f_\eta(\cdot)$ of η_i in (1.1) is continuous with mean 0. Similar discussion can be found in Koul and Song (2008, 2010), among others.

In our application, $g_0(\cdot)$ in (3.2) is a log-linear function. In such a linear case, we do *not* require the knowledge of the distribution of the measurement error η_i in 3.1. To see this, if $g_0(\boldsymbol{x}) = \boldsymbol{\theta}_0^\top \boldsymbol{x}$ in 3.2,

$$\mu_0(\boldsymbol{x}) = \mathsf{E}[\boldsymbol{\theta}_0^\top \mathbf{T}_i|\mathbf{X}_i = \boldsymbol{x}] = \boldsymbol{\theta}_0^\top \boldsymbol{x}$$

Hence the distribution of η_i is not needed to perform simultaneous inference when the null hypothesis (3.2) involves a linear function. However, with some "non-linear" $g_0(\cdot)$,

$$\mu_0(\boldsymbol{x}) = \mathsf{E}[g_0(\mathbf{T}_i)|\mathbf{X}_i = \boldsymbol{x}] = \int g_0(\boldsymbol{x} + y) f_\eta(y) dy$$

where $f_\eta(\cdot)$ is the density function of η_i. Thus, in general, $f_\eta(\cdot)$ is required to derive $\mu_0(\cdot)$ in (3.4) such that one can conduct simultaneous inference of $\mu(\cdot)$. With the known distribution of η_i, the methodology proposed in this study applies to *both* linear *and* nonlinear functions in the hypothesis (3.2). The linear function for (3.2) is used here only because it is the most widely used form of demand. Given the distribution of η_i, our methodology allows us to test a non-linear function in (3.2) as well.

3.2 *U.S.* gasoline demand

Several recent studies analyze demand for gasoline in the *U.S.*/Canadian economy (Hausman and Newey (1995); Yatchew and No (2001); Blundell et al. (2012)). Schmalensee and Stoker (1999) employ the *U.S.* household level data and analyze the *U.S.* gasoline consumption. They estimate their partially linear model using 1988 and 1991 data of approximately 5,000 observations and report a positive relationship between household income and gasoline consumption. Yatchew and No (2001) extends for the Canadian gasoline consumption.

In contrast to the earlier works, we focus on statistical inference of the non-parametric component in the gasoline demand. Our gasoline demand function, which is a special case of (3.1) with $d = 2$, is the following:

$$\begin{aligned}\log(TOTMILES_i) &= \beta_1 \log(DRVR_i) + \beta_2 \log(VEHS_i) + \beta_3 \log(HHSIZE_i) \\ &\quad + \beta_4\, CHILD_i + \beta_5\, SEX_i + \beta_6\, RURAL_i \\ &\quad + \beta_7'\, Region + g\left(\log(INCOME_i), \log(PRICE_i)\right) + \zeta_i \qquad (3.5)\end{aligned}$$

where $\mathrm{TOTMILES}_i$ is total miles traveled by household i, INCOME_i is annual household income in *U.S.* dollars for household i and PRICE_i is gasoline price. Here DRVR_i, VEHS_i and HHSIZE_i are regressors that represent the number of drivers, vehicles and family members for household i, respectively. The other regressors are dummy variables such that $\mathrm{CHILD}_i = 1$ for a household with a child, $\mathrm{SEX}_i = 1$ for a female respondent, and $\mathrm{RURAL}_i = 1$ for a household residing in a rural area. The region dummy Region is a vector of dummy variables that represents different regions of the *U.S*. In total, there are nine different region dummy variables. The dummy variables take either 1 or 0. The data for (3.5) are obtained from the Residential Transportation and Energy Consumption Surveys, which are a series of detailed household surveys on driving behavior and vehicle ownership collected by the *U.S*. Department of Energy, beginning 1979. The survey used in this paper was conducted in year 2001. The total number of observations in our sample is $22,178$. The descriptions and the summary statistics for the variables in (3.5) are given by Table 1.

The hypothesis we consider for (3.5) is the following:

$$H_{10} : g \text{ is a linear function in both argument.} \tag{3.6}$$

As mentioned in (3.2), the log-linear structure is widely used in demand analysis because it provides a simple and useful analytical framework and because its coefficients represent important structural parameters. However, such a parametric form is essentially arbitrary and potentially misspecified (Schmalensee and Stoker (1999), Blundell et al. (2012)). Thus, it is crucial to be able to check its validity through statistical inference. The SCR-based methodology in this study can be readily employed to perform the desired inference.

As mentioned in Section 3.1, the household income and gasoline price variables in (3.5) are likely to be contaminated by Berkson errors. When households are surveyed regarding their annual incomes, they are typically asked to choose the right *categories* for their incomes, rather than to report the exact amounts. When using such data in practice, the *median* values for each category are taken, which leads to the unobserved true income randomly fluctuating around the observed median. This clearly represents the Berkson error defined by (3.1). Similarly, the gasoline price data are likely to be contaminated with the Berkson error since they are based on *self-reported* household expenditure data. In this study, the gasoline price data are obtained through dividing a household's total annual expenditure on gasoline by the total gallons of gasoline purchased each year because the directly observed data on price are not available. Since the self-reported expenditure is likely to be an average of all possible estimates over the plausible range (Hyslop and Imbens (2001)), we naturally introduce the Berkson error into the gasoline price estimates. Through the regression calibration, we conduct the SCR-based inference for (3.5) which is affected by Berkson errors.

The calibrated regression model (cf. (3.3)) states:

$$\begin{aligned}\log(TOTMILES_i) &= \beta_1 \log(DRVR_i) + \beta_2 \log(VEHS_i) + \beta_3 \log(HHSIZE_i) \\ &\quad + \beta_4\, CHILD_i + \beta_5\, SEX_i + \beta_6\, RURAL_i + \beta_7'\, Region \\ &\quad + \mu\left(\log(\mathrm{med}(INCOME_i)),\, \log(\widehat{PRICE}_i)\right) + \epsilon_i \end{aligned} \tag{3.7}$$

where $\mathrm{med}(INCOME_i)$ is the median of the interval that household i belongs, and $\widehat{PRICE}_i$ is defined to be the annual gasoline expenditure divided by the annual gallons of gasoline purchased, which are reported by the ith household in the sample.

3.3 Empirical results

The summary statistics for the variables in (3.5) are provided by Table 1. They include the mean and standard deviation of each variable in the gasoline demand equation, the OLS estimate under the fully parametric linear model, and the standard error for the corresponding OLS estimate. Except for the gender dummy and the gasoline price variable, the OLS estimates are positive. The negative coefficients for the gasoline price variable and the gender dummy indicate that the rising gasoline price makes consumers to switch to other energy sources and that male consumers tend to consume more gasoline than their female counterparts. The corresponding t-statistics are very high, except for $HHSIZE$, which indicates that the variables in (3.5) are statistically very significant with small p-values.

The estimation and simultaneous inference results for the *non-parametric* portion of (3.5) are presented by Figs 1–3. The code that produces these figures can be found in the website `www.quantlet.de`. In particular, Figs 2 and 3 show the estimated $\mu(\boldsymbol{x})$ (i.e. the solid curve) in the calibrated model (3.7) and its corresponding 95 percent simultaneous confidence region (SCR) (i.e. the surrounding dotted band). Here x_1 refers to the household income, while

x_2 represents the gasoline price.

Fig.2 represents the two-dimensional relationship between gasoline demand and gasoline price when the household income is *fixed* at a certain percentile, while Fig.3 represents the relationship between the gasoline demand and household income when the gasoline price is fixed instead. As the traditional demand theory in economics predicts it, Fig.2 shows that the gasoline demand decreases in its price for all three percentiles of income. In contrast, Fig.3 illustrates that the gasoline demand rises in household income regardless of the percentile of price. Interestingly, we can see that the slope of gasoline demand in household income is generally steeper than that in gasoline price. The slope in demand also depends on the percentile of income or that of price, as we observe it from Figs 2 and 3. However, the general trend appears to hold true regardless of the corresponding percentile.

In order to accept the null hypothesis of linearity for $g(\cdot,\cdot)$ in (3.5), which is a common assumption in demand analysis, one should be able to insert a straight line into the constructed SCRs in "*all*" of the panels in Figs 2–3. That is, if one cannot insert a straight line into *all* of the constructed SCRs in Figs 2–3, then the linearity of $\mu(\cdot,\cdot)$ is *rejected* at 5 percent level. Thus, the linearity of $g(\cdot,\cdot)$ is also rejected by the argument in Section 3.2. Obviously, the SCRs presented in Fig.3 cannot contain any straight line in them because of the non-linearity of the estimates and of the corresponding SCRs. Hence the linearity hypothesis for $g(\cdot,\cdot)$ in (3.5) is clearly *rejected* at 5 percent level for the reasons discussed in Section 3.2.

As mentioned in Introduction, the main advantage of SCR-based inference over the usual integrated-squared-difference type statistics (Härdle and Mammen (1993)) is that it is relatively straightforward to suggest an alternative based on the SCR when the null model gets rejected. Given Figs 2–3, the linearity between the household income and gasoline consumption is rejected while that between the gasoline price and gasoline consumption is not. Hence $\beta * \text{price} + g(\text{income})$ is suggested instead of $g(\text{income, price})$ in (3.5). This would be rather difficult to suggest if the original null is rejected based on the integrated-squared-difference type statistics and its associated p-value.

4 Concluding Remarks

The paper illustrates how to conduct the simultaneous inference of a semi-parametric partially non-linear model when the number of covariate terms in the non-linear part is *two or higher*. To that end, we illustrate how to construct the simultaneous confidence region (SCR) for the multivariate unknown function. The developed methodology is applied to perform inference of the gasoline demand function when the model covariates are possibly contaminated by the Berkson-type measurement errors. Through the regression calibration (Carroll et al. (1995)), we transform the original model into the one with *observable* covariate terms and base the inference on the transformed one instead. The inference of the transformed model is conducted through the construction of SCR, which is a multi-dimensional extension of the two-dimensional uniform confidence band (Härdle (1989), Johnston (1982); Kim (2016)). The relating asymptotic properties of the introduced methodology are investigated. In addition, a simulation-based construction of SCR is discussed in comparison to the asymptotic-based approach. The empirical analysis shows that the linearity hypothesis for the U.S. gasoline demand is rejected at 5 percent level, mainly due to the non-linear relationship between the $U.S.$ gasoline consumption and the $U.S.$ household income (see Fig.3). Based on the result, an alternative form for the demand function is also suggested.

Regarding future research, this project suggests a couple of interesting topics for consideration. One of them is to extend the current work to the case of *time series*. Unlike the cross section one considered in this study, time series data inherently possess *temporal dependence* among them. Furthermore, the data might suggest that the underlying process is non-stationary. One possibility is to model the covariate terms as locally stationary processes as in Kim (2016), among others. These features would require a different framework to handle the issue more properly. Another potential extension is to consider the *classical* measurement error structure. Due to the nature of the available income and price data in this study, we assume that the model covariates are contaminated by the Berkson errors. One can instead assume that the error structure is the classical one and perform simultaneous inference of the model, accordingly. Further insight can be gained by extending the current work in these and other possible directions.

References

Berkson, J. (1950). Are there two regressions? *Journal of the American Statistical Association*, 45:164–180.

Bickel, P. J. and Wichura, M. J. (1971). Convergence criteria for multivariate stochastic processes and some applications. *The Annals of Mathematical Statistics*, 42(5):1656–1670.

Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley & Sons.

Blundell, R., Horowitz, J. L., and Parey, M. (2012). Measuring the price responsiveness of gasoline demand: Economic shape restrictions and nonparametric demand estimation. *Quantitative Economics*, 3:29–51.

Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995). *Measurement error in nonlinear models*. Chapman and Hall, Boca Raton.

Chao, S.-K., Proksch, K., Dette, H., and Härdle, W. K. (2017). Confidence corridors for multivariate generalized quantile regression. *Journal of Business & Economic Statistics*, 35:70–85.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerische Mathematik*, 31:377–403.

Dedecker, J., Merlevéde, F., and Rio, E. (2014). Strong approximation of the empirical distribution function for absolutely regular sequences in $\mathbb{R}^d$. *Electronic Journal of Probability*, 19(9):1–56.

Fan, J. and Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scandinavian Journal of Statistics*, 27:715–731.

Ferguson, T. S. (1996). *A Course in large sample theory.* Chapman and Hall, CRC.

Guo, M. and Härdle, W. (2012). Simultaneous confidence bands for expectile functions. *AStA Advances in Statistical Analysis*, 96(4):517–541.

Hall, P. (1979). On the rate of convergence of normal extremes. *Journal of Applied Probability*, 16(2):433–439.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24(3):726–748.

Härdle, W. (1989). Asymptotic maximal deviation of M-smoothers. *Journal of Multivariate Analysis*, 29(2):163–179.

Härdle, W., Liang, H., and Gao, J. (2000). *Partially linear models.* Heidelberg, Physica-Verlag.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, 21(4):1926–1947.

Härdle, W. and Song, S. (2010). Confidence bands in quantile regression. *Econometric Theory*, 26(4):1180–1200.

Hausman, J. and Newey, W. (1995). Nonparametric estimation of exact consumer surplus and deadweight loss. *Econometrica*, 63:1445–1476.

Hyslop, D. R. and Imbens, G. W. (2001). Bias from classical and other forms of measurement error. *Journal of Business & Economic Statistics*, 19:475–481.

Johnston, G. J. (1982). Probabilities of maximal deviations for nonparametric regression function estimates. *Journal of Multivariate Analysis*, 12(3):402–414.

Kim, K. H. (2016). Inference of the trend in a partially linear model with locally stationary regressors. *Econometric Reveiws*, 35(7):1194–1220.

Koul, H. L. and Song, W. (2008). Model checking with berkson measurement errors. *Journal of Statistical Planning and Inference*, 138:1615–1628.

Koul, H. L. and Song, W. (2010). Model checking in partial linear regresion models with berkson measurement errors. *Statistica Sinica*, 20:1551–1579.

Li, Q. and Racine, J. S. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton university press, New Jersey.

Liu, W. and Wu, W. B. (2010). Simultaneous nonparametric inference of time series. *The Annals of Statistics*, 38(4):2388–2421.

Meerschaert, M. M., Wang, W. and Xiao, Y. (2013). Fernique-type inequalities and moduli of continuity for anisotropic Gaussian random fields. *Transations of the American Mathematical Society*, 365(2):1081–1107.

Owen, A. B. (2005). *Multidimensional variation for quasi-Monte Carlo*, volume 2 of *Ser. Biostat.*, pages 49–74. World Sci. Publi., Hackensack, NJ.

Proksch, K. (2012). *Gleichmäßige Konfidenzbereiche für multivariate inverse Regressionsmodelle mit Faltungsoperatoren*. PhD thesis, Ruhr-Universität Bochum.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica*, 56:931–954.

Rosenblatt, M. (1976). On the maximal deviation of k-dimensional density estimates. *The Annals of Probability*, 4(6):1009–1015.

Schmalensee, R. and Stoker, T. M. (1999). Household gasoline demand in the united states. *Econometrica*, 67:645–662.

Song, S., Ritov, Y., and Härdle, W. (2012). Partial linear quantile regression and bootstrap confidence bands. *Journal of Multivariate Analysis*, 107:244–262.

Wu, W. B. and Zhao, Z. (2007). Inference of trends in time series. *Journal of the Royal Statistical Society: Series B*, 69:391–410.

Yatchew, A. and No, J. A. (2001). Household gasoline demand in canada. *Econometrica*, 69:1697–1709.

Zhao, Z. and Wu, W. B. (2008). Confidence bands in nonparametric time series regression. *The Annals of Statistics*, 36:1854–1878.

Zhou, Z. and Wu, W. B. (2010). Simultaneous inference of linear models with time varying coefficients. *Journal of the Royal Statistical Society: Series B*, 72:513–531.

Appendices

In the appendix, we provide the technical details for the theoretical results given in the main text. Section A and B contain the proof for the theorems in the main text. Section C lists some useful results for proving our theory.

We introduce some additional notations. Let $\Gamma_n = \{y : |y| \le a_n\}$. $\sigma_n^2(\boldsymbol{x}) = \mathsf{E}\big[(Y^* - \mu(\boldsymbol{x}))^2 \mathbf{1}(Y^* \in \Gamma_n)|\mathbf{X} = \boldsymbol{x}\big]$ and $\widetilde{\sigma}_n^2 = \mathsf{E}\big[(Y^* - \mu(\boldsymbol{x}))^2 \mathbf{1}(Y^* \notin \Gamma_n)|\mathbf{X} = \boldsymbol{x}\big]$. Denote the vector of ones and zeros by $\mathbf{l} = (1, ..., 1)$ and $\mathbf{0} = (0, ..., 0)$.

A Bound for R_n

Lemma A.1. *Under assumptions (A1)-(A6), suppose ϵ_i is bounded almost surely, then*

$$\|R_n(\boldsymbol{x})\| = o_{\mathbb{P}}\left\{\left(nh^d \log(n)\right)^{-1/2}\right\} \tag{A.1}$$

Proof. *Recall that* $R_n(\boldsymbol{x}) = \frac{\mathbb{E}H_n(\boldsymbol{x})}{f_{\mathbf{X}}(\boldsymbol{x})} + \frac{H_n(\boldsymbol{x})\left(f_{\mathbf{X}}(\boldsymbol{x}) - \widehat{f}(\boldsymbol{x})\right)}{f_{\mathbf{X}}(\boldsymbol{x})\widehat{f}(\boldsymbol{x})}$*. Then,*

$$\sup_{\boldsymbol{x}} |R_n(\boldsymbol{x})| \leq \sup_{\boldsymbol{x}} \left| \frac{\mathbb{E}H_n(\boldsymbol{x})}{f_{\mathbf{X}}(\boldsymbol{x})} \right| + \sup_{\boldsymbol{x}} \left| \frac{H_n(\boldsymbol{x}) \left(f_{\mathbf{X}} - \widehat{f}_{\mathbf{X}}(\boldsymbol{x}) \right)}{f_{\mathbf{X}}(\boldsymbol{x}) \widehat{f}_{\mathbf{X}}(\boldsymbol{x})} \right| \tag{A.2}$$

By Theorem 1.4 in Li and Racine (2007) and (A4),

$$\sup_{\boldsymbol{x}} \left| f_{\mathbf{X}}(\boldsymbol{x}) - \widehat{f}_{\mathbf{X}}(\boldsymbol{x}) \right| = O_{\mathbb{P}} \left(h^2 + \sqrt{\frac{\log(n)}{nh^d}} \right) \tag{A.3}$$

Then, by (A.3),

$$\sup_{\boldsymbol{x}} \left| \frac{\widehat{f}_{\mathbf{X}}(\boldsymbol{x}) - f_{\mathbf{X}}(\boldsymbol{x})}{f_{\mathbf{X}}(\boldsymbol{x}) \widehat{f}_{\mathbf{X}}(\boldsymbol{x})} \right| \leq \frac{\sup_{\boldsymbol{x}} \left| \widehat{f}_{\mathbf{X}}(\boldsymbol{x}) - f_{\mathbf{X}}(\boldsymbol{x}) \right|}{\inf_{\boldsymbol{x}} f_{\mathbf{X}}(\boldsymbol{x})(f_{\mathbf{X}}(\boldsymbol{x}) - \epsilon_n)} = O_{\mathbb{P}} \left(h^2 + \sqrt{\frac{\log(n)}{nh^d}} \right) \tag{A.4}$$

where $\left| \widehat{f}_{\mathbf{X}}(\boldsymbol{x}) - f_{\mathbf{X}}(\boldsymbol{x}) \right| \leq \epsilon_n$ *and* $\epsilon_n \to 0$ *in probability from (A.3). This leads to* $f_{\mathbf{X}}(\boldsymbol{x}) - \epsilon_n \leq \widehat{f}_{\mathbf{X}}(\boldsymbol{x})$*, where* $f_{\mathbf{X}}(\boldsymbol{x}) - \epsilon_n > C$ *for some constant* $C > 0$ *for sufficiently large* n*. Moreover,*

$$\begin{aligned}
\mathbb{E}H_n(\boldsymbol{x}) &= h^{-d}\mathbb{E}\left[K\left(\frac{\boldsymbol{x} - \mathbf{X}_i}{h} \right) (\mu(\mathbf{X}) + \epsilon_i - \mu(\boldsymbol{x})) \right] \\
&= h^{-d}\mathbb{E}\left[K\left(\frac{\boldsymbol{x} - \mathbf{X}_i}{h} \right) (\mu(\mathbf{X}) - \mu(\boldsymbol{x})) \right] \\
&= h^{-d} \int K\left(\frac{\boldsymbol{x} - \boldsymbol{t}}{h} \right) (\mu(\boldsymbol{t}) - \mu(\boldsymbol{x})) f(\boldsymbol{t}) d\boldsymbol{t} \\
&\lesssim h^{-d} \int K\left(\frac{\boldsymbol{x} - \boldsymbol{t}}{h} \right) \|\boldsymbol{x} - \boldsymbol{t}\|_\infty^s f(\boldsymbol{t}) d\boldsymbol{t} \\
&\lesssim h^{-d+s} \int K\left(\frac{\boldsymbol{x} - \boldsymbol{t}}{h} \right) f(\boldsymbol{t}) d\boldsymbol{t} \\
&= h^s \int K(\boldsymbol{u}) f(\boldsymbol{x} - \boldsymbol{u}) d\boldsymbol{u} \\
&= \mathcal{O}(h^s)
\end{aligned}$$

where the fourth inequality follows by assumption (A1) and (A3); the fifth inequality is from the bounded support of $K(\cdot)$ assumed in (A1); the sixth equality is from the change-of-variables, and the final rate is obtained by the summability condition of $K(\cdot)$ from (A1) and the bounded $f_{\boldsymbol{X}}(\boldsymbol{x})$ from (A4). Thus,

$$\sup_{\boldsymbol{x}} |\mathbb{E}H_n(\boldsymbol{x})| = \mathcal{O}(h^s) \tag{A.5}$$

Furthermore,

$$\sup_{\boldsymbol{x}} |H_n(\boldsymbol{x})| \leq \sup_{\boldsymbol{x}} |H_n(\boldsymbol{x}) - \mathbb{E}H_n(\boldsymbol{x})| + \sup_{\boldsymbol{x}} |\mathbb{E}H_n(\boldsymbol{x})|, \tag{A.6}$$

where $\sup_{\boldsymbol{x}} |H_n(\boldsymbol{x}) - \mathbb{E}H_n(\boldsymbol{x})| = O_{\mathbb{P}}((\log n/nh^d)^{1/2})$ *from the application of Bernstein inequality, and a truncation argument to ϵ_i as in the proof for Theorem 2 of Hansen (2008). Then in view of (A.2) and (A.4)–(A.6),*

$$\sup_{\boldsymbol{x}} |R_n(\boldsymbol{x})| = O_{\mathbb{P}}\left(h^s + \sqrt{\frac{\log(n)}{nh^d}}\left(h^2 + \sqrt{\frac{\log(n)}{nh^d}}\right)\right),$$

note that by (A6) we have $h^s\sqrt{nh^d \log n} \to 0$. Hence the lemma follows.

B Proof of Proposition 2.1

The steps of the proof mainly follow Chao et al. (2017).

$$U_n(\boldsymbol{x}) = \frac{1}{\sqrt{h^d\sigma^2 f_{\boldsymbol{X}}(\boldsymbol{x})}} \int\int K\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{h}\right)(y^* - \mu(\boldsymbol{x}))dZ_n(\boldsymbol{u}, y^*) \tag{B.1}$$

$$U_{0,n}(\boldsymbol{x}) = \frac{1}{\sqrt{h^d\sigma_n^2(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})}} \int\int_{\Gamma_n} K\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{h}\right)(y^* - \mu(\boldsymbol{x}))dZ_n(\boldsymbol{u}, y^*) \tag{B.2}$$

$$U_{1,n}(\boldsymbol{x}) = \frac{1}{\sqrt{h^d\sigma_n^2(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})}} \int\int_{\Gamma_n} K\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{h}\right)(y^* - \mu(\boldsymbol{x}))dB_n(T(y^*, \boldsymbol{u})) \tag{B.3}$$

where $B_n\{T(y,\boldsymbol{u})\} = W_n\{T(y,\boldsymbol{u})\} - F(y,\boldsymbol{u})W_n(1,...,1)$ and $T(y,\boldsymbol{u})$ is the d dimensional Rosenblatt transformation:

$$T(y,\boldsymbol{u}) = \left\{F_{X_1|Y}(u_1|y), F_{X_2|Y}(u_2|u_1,y), ..., F_{X_d|X_{d-1},...,X_1,Y}(u_d|u_{d-1},...,u_1,y), F_Y(y)\right\}.$$

$$U_{2,n}(\boldsymbol{x}) = \frac{1}{\sqrt{h^d\sigma_n^2(\boldsymbol{x})f_{\boldsymbol{X}}(\boldsymbol{x})}}\int\int_{\Gamma_n} K\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{h}\right)(y^* - \mu(\boldsymbol{x}))dW_n(T(y^*,\boldsymbol{u})) \tag{B.4}$$

$$U_{3,n}(\boldsymbol{x}) = \frac{1}{\sqrt{h^d\sigma_n^2(\boldsymbol{x})f_{\boldsymbol{X}}(\boldsymbol{x})}}\int\int_{\Gamma_n} K\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{h}\right)(y^* - \mu(\boldsymbol{u}))dW_n(T(y^*,\boldsymbol{u})) \tag{B.5}$$

$$U_{4,n}(\boldsymbol{x}) = \frac{1}{\sqrt{h^d\sigma_n^2(\boldsymbol{x})f_{\boldsymbol{X}}(\boldsymbol{x})}}\int \sqrt{\sigma_n(\boldsymbol{u})^2 f(\boldsymbol{u})}K\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{h}\right)dW(\boldsymbol{u}) \tag{B.6}$$

$$U_{5,n}(\boldsymbol{x}) = \frac{1}{\sqrt{h^d}}\int K\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{h}\right)dW(\boldsymbol{u}) \tag{B.7}$$

Theorem 2.2 follows from applying Theorem 2 of Rosenblatt (1976) on $U_{5,n}(\boldsymbol{x})$.

Next we introduce some notations which are used repeatedly in the following proofs.

Definition B.1 (Neighboring Block in $\mathcal{X} \subset \mathbb{R}^d$, Bickel and Wichura (1971) p.1658)**.** *A* block $B \subset \mathcal{X}$ *is a subset of* $\mathcal{X}$ *of the form* $B = \Pi_i(s_i, t_i]$ *with* s *and* t *in* $\mathcal{X}$*; the* pth-face *of* B *is* $\Pi_{i\neq p}(s_i, t_i]$*. Disjoint blocks* B *and* C *are* p-neigbbors *if they abut and have the same pth face; they are* neighbors *if they are p-neighbors for some* $p \geq 1$*.*

To illustrate the idea of neighboring block, take $d = 3$ for example, the blocks $(s,t] \times (a,b] \times (c,d]$ and $(t,u] \times (a,b] \times (c,d]$ are 1-neighbors for $s \leq t \leq u$.

Definition B.2 (Bickel and Wichura (1971) p.1658)**.** *Let* $X : \mathbb{R}^d \to \mathbb{R}$*. The* increment of X on the block B*, denoted* $X(B)$*, is defined by*

$$X(B) = \sum_{\boldsymbol{\alpha}\in\{0,1\}^d} (-1)^{d-|\boldsymbol{\alpha}|} X\left\{\boldsymbol{s} + \boldsymbol{\alpha} \odot (\boldsymbol{t} - \boldsymbol{s})\right\}, \tag{B.8}$$

where $|\boldsymbol{\alpha}| = \alpha_1 + \alpha_2 + ... + \alpha_d$*, "$\odot$" denotes the* componentwise product*; that is, for any vectors* $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^d$*,* $\boldsymbol{u} \odot \boldsymbol{v} = (u_1v_1, u_2v_2, ..., u_dv_d)$*.*

Below we give some examples of the increment of a multivariate function X on a block:

- $d = 1$: $B = (s, t]$, $X(B) = X(t) - X(s)$;
- $d = 2$: $B = (s_1, t_1] \times (s_2, t_2]$. $X(B) = X(t_1, t_2) - X(t_1, s_2) + X(s_1, s_2) - X(s_1, t_2)$.

Lemma B.3. $\|U_n - U_{0,n}\| = o_p\big((\log n)^{-1/2}\big)$.

Proof of Lemma B.3. By the triangle inequality we have

$$\|U_n - U_{n,0}\| \le \|U_n - \widehat{U}_{n,0}\| + \|\widehat{U}_{n,0} - U_{n,0}\| \stackrel{\text{def}}{=} E_1 + E_2,$$

where $\widehat{U}_{n,0} = \sigma^2(\boldsymbol{x})/\sigma_n(\boldsymbol{x})U_{n,0}(\boldsymbol{x})$ and the terms E_1 and E_2 are defined in an obvious manner. We now show that $E_j = o_p\big\{(\log n)^{-1/2}\big\}$, $j = 1, 2$. Note that

$$|\widehat{U}_{n,0}(\boldsymbol{x}) - U_{n,0}(\boldsymbol{x})| = \Big|\big(\sigma(\boldsymbol{x})/\sigma_n(\boldsymbol{x}) - 1\big)U_{n,0}(\boldsymbol{x})\Big|.$$

It is shown later that $\|U_{n,0}\| = \mathcal{O}_p\big(\sqrt{\log n}\big)$, hence it remains to prove that

$$\sup_{\boldsymbol{x}\in\mathcal{X}} \big|\sigma(\boldsymbol{x})/\sigma_n(\boldsymbol{x}) - 1\big| = o\big\{(\log n)^{-1}\big\}. \tag{B.9}$$

Under $a_n \asymp (h^{-3d}\log n)^{1/(b_1-2)} \to \infty$ as $n \to \infty$, $\sigma_n^2(\boldsymbol{x}) \to \sigma^2(\boldsymbol{x}) > 0$ uniformly in $\boldsymbol{x}$; moreover, by (A2)(ii), we have

$$h^{-3d}\log n \sup_{\boldsymbol{x}\in\mathcal{X}} |\widetilde{\sigma}^2(\boldsymbol{x})| = h^{-3d}\log n \sup_{\boldsymbol{x}\in\mathcal{X}} \left|\int_{|y^*|>a_n} (y^* - \mu(\boldsymbol{x}))^2 f_{Y^*|\boldsymbol{X}}(y^*|\boldsymbol{x})dy^*\right| = \mathcal{O}(1), \tag{B.10}$$

which implies $\sup_{\boldsymbol{x}\in\mathcal{X}} |(\log n)^2\widetilde{\sigma}_n^2(\boldsymbol{x})/\sigma_n^2(\boldsymbol{x})| \le \big|(\log n)h^{3d}\mathcal{O}(1)\big| = o(1)$. Therefore,

$$(\log n)\sup_{\boldsymbol{x}\in\mathcal{X}} \left|\sqrt{\frac{\sigma^2(\boldsymbol{x})}{\sigma_n^2(\boldsymbol{x})}} - 1\right| = (\log n)\sup_{\boldsymbol{x}\in\mathcal{X}} \left|\sqrt{\frac{\widetilde{\sigma}_n^2(\boldsymbol{x}) + \sigma_n^2(\boldsymbol{x})}{\sigma_n^2(\boldsymbol{x})}} - 1\right| \le \sup_{\boldsymbol{x}\in\mathcal{X}} \left|\sqrt{\frac{(\log n)^2\widetilde{\sigma}_n^2(\boldsymbol{x})}{\sigma_n^2(\boldsymbol{x})}}\right| \to 0,$$

as $n \to \infty$, hence $E_2 = o_p\big((\log n)^{-1/2}\big)$.

We now show that $E_1 = o_p((\log n)^{-1})$. To do this, it is enough to show the weak convergence of $\log nE_1$, and it requires to show the finite dimensional convergence to 0 with rate $(\log n)^{-1}$ and the tightness of the process inducing E_1 in $D(\mathcal{X})$ in the sense of Chapter 3 of Billingsley (1968). First, we observe that

$$\begin{aligned}
(\log n)^{1/2}E_1 &= (\log n)^{1/2}\sup_{\boldsymbol{x}\in\mathcal{X}}|U_n(\boldsymbol{x})-\widehat{U}_{n,0}(\boldsymbol{x})| \\
&= (\log n)^{1/2}\sup_{\boldsymbol{x}\in\mathcal{X}}\left|\frac{1}{\sqrt{h^d f_{\mathbf{X}}(\boldsymbol{x})\sigma^2(\boldsymbol{x})}}\int\int_{\{|y|>a_n\}}K\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{h}\right)(y^*-\mu(\boldsymbol{x}))dZ_n(y^*,\boldsymbol{u})\right| \\
&= \sup_{\boldsymbol{x}\in\mathcal{X}}\left|\frac{1}{\sqrt{f_{\mathbf{X}}(\boldsymbol{x})\sigma^2(\boldsymbol{x})}}V_n(\boldsymbol{x})\right|,
\end{aligned}$$

where

$$V_n(\boldsymbol{x}) = \sum_{i=1}^n W_{n,i}(\boldsymbol{x}),$$

and

$$\begin{aligned}
W_{n,i}(\boldsymbol{x}) = (\log n)^{1/2}(nh^d)^{-1/2}\Big\{&\psi(Y_i^*-\mu(\boldsymbol{x}))\mathbf{1}(|Y_i^*|>a_n)K\Big(\frac{\boldsymbol{x}-\mathbf{X}_i}{h}\Big) \\
&-\mathsf{E}\Big[\psi(Y_i^*-\mu(\boldsymbol{x}))\mathbf{1}(|Y_i^*|>a_n)K\Big(\frac{\boldsymbol{x}-\mathbf{X}_i}{h}\Big)\Big]\Big\}.
\end{aligned}$$

Note that $f_{\mathbf{X}}(\boldsymbol{x})\sigma^2(\boldsymbol{x})>0$ for all $\boldsymbol{x}\in\mathcal{X}$ by Assumption (A2) and (A4). By (B.10),

$$\begin{aligned}
\mathsf{E}[W_{n,i}(\boldsymbol{x})^2] &\le (\log n)(nh^d)^{-1}\mathsf{E}\Big[\big(Y_i^*-\mu(\boldsymbol{x})\big)^2\mathbf{1}(|Y_i^*|>a_n)K^2\Big(\frac{\boldsymbol{x}-\mathbf{X}_i}{h}\Big)\Big] \\
&\le (\log n)(nh^d)^{-1}C_K\widetilde{\sigma}_n^2(\boldsymbol{x}) \\
&= \mathcal{O}(h^{2d}n^{-1}).
\end{aligned}$$

Thus,

$$\mathsf{E}\left[\left(\sum_{i=1}^n W_{n,i}(\boldsymbol{x})\right)^2\right] \le n\mathsf{E}\left[(W_{n,i}(\boldsymbol{x}))^2\right] = \mathcal{O}(h^{2d}) = o\left((\log n)^{-1}\right),$$

as $n\to\infty$. From Markov's inequality, $V_n(\boldsymbol{x}) = o_p(1)$ for each fixed $\boldsymbol{x}\in\mathcal{X}$. With this result, finite convergence follows by Cramér-Wold theorem and the detail is omitted.

We now show the tightness of $V_n(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathcal{X}$. To simplify the expression, define

$$g(\boldsymbol{x}) \stackrel{\text{def}}{=} \{Y^* - \mu(\boldsymbol{x})\} K\left(\frac{\boldsymbol{x} - \mathbf{X}}{h}\right).$$

Take arbitrary neighboring blocks $B, C \subset \mathcal{X}$ (see Definition B.1) and suppose $B = \Pi_{i=1}^d (s_i, t_i]$,

$$\begin{aligned}
\mathsf{E}[V_n(B)^2]^{1/2} \leq{}& (\log n)^{1/2} h^{-d/2} \Big\{ \mathsf{E}\Big[\mathbf{1}(Y_i^* > a_n)\Big(\sum_{\boldsymbol{\alpha} \in \{0,1\}^d} (-1)^{d-|\boldsymbol{\alpha}|} g\big(\boldsymbol{s} + \boldsymbol{\alpha} \odot (\boldsymbol{t} - \boldsymbol{s})\big)\Big)^2\Big] \\
&+ \mathsf{E}\Big[\mathbf{1}(Y_i^* < -a_n)\Big(\sum_{\boldsymbol{\alpha} \in \{0,1\}^d} (-1)^{d-|\boldsymbol{\alpha}|} g\big(\boldsymbol{s} + \boldsymbol{\alpha} \odot (\boldsymbol{t} - \boldsymbol{s})\big)\Big)^2\Big]\Big\}^{1/2} \\
\stackrel{\text{def}}{=}{}& (\log n)^{1/2} h^{-d/2} (I_1 + I_2)^{1/2},
\end{aligned}$$

where I_1 and I_2 are defined in an obvious manner. Hence, I_1 can be estimated as

$$\begin{aligned}
I_1 \leq{}& 2 \int\int \mathbf{1}(y^* > a_n) y^{*2} \Big(\sum_{\boldsymbol{\alpha} \in \{0,1\}^d} (-1)^{d-|\boldsymbol{\alpha}|} K\Big[\big(\boldsymbol{s} + \boldsymbol{\alpha} \odot (\boldsymbol{t} - \boldsymbol{s}) - \boldsymbol{u}\big)/h\Big]\Big)^2 f(y^*, \boldsymbol{u}) dy d\boldsymbol{u}. \\
&+ 2 \int\int \mathbf{1}(y^* > a_n) \Big(\sum_{\boldsymbol{\alpha} \in \{0,1\}^d} (-1)^{d-|\boldsymbol{\alpha}|} \mu(\big(\boldsymbol{s} + \boldsymbol{\alpha} \odot (\boldsymbol{t} - \boldsymbol{s})) K\Big[\big(\boldsymbol{s} + \boldsymbol{\alpha} \odot (\boldsymbol{t} - \boldsymbol{s}) - \boldsymbol{u}\big)/h\Big]\Big)^2 f(y^*, \boldsymbol{u}) dy d\boldsymbol{u}
\end{aligned}$$

Note that

$$\begin{aligned}
&\sum_{\boldsymbol{\alpha} \in \{0,1\}^d} (-1)^{d-|\boldsymbol{\alpha}|} K\Big[\big(\boldsymbol{s} + \boldsymbol{\alpha} \odot (\boldsymbol{t} - \boldsymbol{s}) - \boldsymbol{u}\big)/h\Big] = \int_B \partial^{(1,\ldots,1)} K\Big(\frac{\boldsymbol{v} - \boldsymbol{u}}{h}\Big) d\boldsymbol{v} \leq h^{-d} C_{K'} \lambda(B), \\
&\sum_{\boldsymbol{\alpha} \in \{0,1\}^d} (-1)^{d-|\boldsymbol{\alpha}|} \mu(\big(\boldsymbol{s} + \boldsymbol{\alpha} \odot (\boldsymbol{t} - \boldsymbol{s})) K\Big[\big(\boldsymbol{s} + \boldsymbol{\alpha} \odot (\boldsymbol{t} - \boldsymbol{s}) - \boldsymbol{u}\big)/h\Big] \\
&\qquad\qquad\qquad = \int_B \partial^{(1,\ldots,1)} \Big\{\mu(\boldsymbol{v}) K\Big(\frac{\boldsymbol{v} - \boldsymbol{u}}{h}\Big)\Big\} d\boldsymbol{v} \leq h^{-d} C_{K',\mu'} \lambda(B),
\end{aligned}$$

where by (A1) and (A3) the constant $C_{K'}, C_{K',\mu'} > 0$ satisfies $\sup_{\boldsymbol{u} \in \mathcal{D}} |\partial^{\boldsymbol{\alpha}} K(\boldsymbol{u})| \leq C_{K'}$ and $\sup_{\boldsymbol{u} \in \mathcal{D}} |\partial^{\boldsymbol{\alpha}} \{\mu(\boldsymbol{u}) K(\boldsymbol{u})\}| \leq C_{K',\mu'}$. $\lambda(\cdot)$ is the Lebesgue measure.

Taking $C_1 = \max\{C_{K'}, C_{K',\mu'}\} > 0$. Under the constraint that $y^* > a_n > 1$ for large enough n, we have

$$I_1 \le 4 \int \int_{a_n}^{\infty} y^{*2} \big(C_1 h^{-d} \lambda(B)\big)^2 f(y^*, \boldsymbol{u}) dy d\boldsymbol{u} = 4h^{-2d} C_1^2 \lambda(B)^2 \int_{\{y^* > a_n\}} y^{*2} f_{Y^*}(y^*) dy^*.$$

By symmetry,

$$I_2 \le 4h^{-2d} C_1^2 \lambda(B)^2 \int_{\{y < -a_n\}} y^{*2} f_{Y^*}(y^*) dy^*.$$

Hence, by (2.1) in (A2),

$$\mathsf{E}[V_n(B)^2]^{1/2} \le 2C_1 \lambda(B) \left(h^{-3d} \log n \int_{\{|y^*| > a_n\}} y^{*2} f_{Y^*}(y^*) dy^* \right)^{1/2} \le 2C_1 C^* \lambda(B).$$

Analogously we obtain the estimate for

$$\mathsf{E}[V_n(C)^2]^{1/2} \le 2C_1 C^* \lambda(C),$$

which finally yields by Hölder's inequality,

$$\mathsf{E}[|V_n(B)||V_n(C)|] \le \mathsf{E}[|V_n(B)|^2]^{1/2} \mathsf{E}[|V_n(C)|^2]^{1/2} \le 4C_1^2 C^{*2} \lambda(B)\lambda(C).$$

Applying Lemma C.1 with $\gamma_1 = \gamma_2 = \lambda_1 = \lambda_2 = 1$ yields the tightness. □

Lemma B.4. $\|U_{0,n} - U_{1,n}\| = \mathcal{O}_p\big(n^{-1/6} h^{-d/2} (\log n)^{\epsilon + (2d+4)/3} a_n\big)$ *for any* $\epsilon > 0$.

Proof of Lemma B.4. In this proof, we adopt the notation that if $\boldsymbol{\alpha} \in \{0,1\}^{d+1}$, then we write $\boldsymbol{\alpha} = (\alpha_1, \boldsymbol{\alpha}_2)$ where $\alpha_1 \in \{0,1\}$ and $\boldsymbol{\alpha}_2 \in \{0,1\}^d$. In the computation below, we focus on $B_{\boldsymbol{x}} = \Pi_{j=1}^d \big[x_j - Ah, x_j + Ah\big]$ instead of $\mathbb{R}^d$ since K has compact support. Recall definition B.1 of an increment of a function X over a block B. Integration by parts for multivariate integral (see, for example, Theorem 3.4 on p.64 of Proksch (2012)) gives,

$$
\begin{aligned}
U_{0,n}(\boldsymbol{x}) = \frac{1}{\sqrt{h^d f_{\boldsymbol{X}}(\boldsymbol{x})\sigma_n^2(\boldsymbol{x})}} \Bigg[& \int_{B_{\boldsymbol{x}}} \int_{\Gamma_n} Z_n(y,\boldsymbol{u}) \; d\Bigg(\Big(y^* - \mu(\boldsymbol{x})\Big) K\Big(\frac{\boldsymbol{x}-\boldsymbol{u}}{h}\Big)\Bigg) \\
& + \Big\{ Z_n\big(\cdot_1,\cdot_2\big)\big(\cdot_1 - \mu(\boldsymbol{x})\big) K\Big(\frac{\boldsymbol{x}-\cdot_2}{h}\Big)\Big\}\big(\Gamma_n \times B_{\boldsymbol{x}}\big) \\
& + \Bigg\{ \sum_{\boldsymbol{\alpha}\in\{0,1\}^{d+1}-\{\mathbf{0},\mathbf{l}\}} \int\int_{(\Gamma_n\times B_{\boldsymbol{x}})_{\boldsymbol{\alpha}}} Z_n(\cdot_1,\cdot_2)\; d^{\alpha_1}\big(\cdot_1 - \mu(\boldsymbol{x})\big)\partial^{\boldsymbol{\alpha}_2} K\Big(\frac{\boldsymbol{x}-\cdot_2}{h}\Big)\Bigg\}\big(\Gamma_n \times B_{\boldsymbol{x}}\big)_{\mathbf{l}-\boldsymbol{\alpha}}\Bigg]
\end{aligned}
\tag{B.11}
$$

where $\mathbf{l} = (1,...,1) \in \{0,1\}^{d+1}$ and $\mathbf{0} = (0,...,0) \in \{0,1\}^{d+1}$. $\big(\Gamma_n \times B_{\boldsymbol{x}}\big)$ is a $d+1$ dimensional cube. $\cdot_1$ corresponds to the one-dimensional variable y and $\cdot_2$ corresponds to the two-dimensional variable u. The second term in (B.11) can be evaluated with the formula (B.8). $\big(\Gamma_n \times B_{\boldsymbol{x}}\big)_{\mathbf{l}-\boldsymbol{\alpha}}$ can be viewed as the projection of $\Gamma_n \times B_{\boldsymbol{x}}$ on to the space spanned by those axes whose numbers correspond to positions of ones of the multi-index $\mathbf{l}-\boldsymbol{\alpha}$. This leaves us with an $|\boldsymbol{\alpha}|$-fold integral.

Moreover, $d\big\{\big(y^* - \mu(\boldsymbol{x})\big)K((\boldsymbol{x}-\boldsymbol{u})/h)\big\} = \partial_{y^*}\big(y^* - \mu(\boldsymbol{x})\big)\partial_{\boldsymbol{u}}^{\mathbf{l}_2} K\big((\boldsymbol{x}-\boldsymbol{u})/h\big)$, where $\mathbf{l}_2 = (1,...,1) \in \{0,1\}^d$ and $d\big(y^* - \mu(\boldsymbol{x})\big) = 1$.

By applying integration by parts for Brownian integral (Theorem 3.5 on p.70 of Proksch (2012)) to $U_{1,n}(\boldsymbol{x})$, and by Theorem 3.2 in Dedecker et al. (2014), we obtain for every $\epsilon > 0$,

$$
\begin{aligned}
& h^{d/2} n^{1/6} (\log n)^{-\epsilon-(2d+4)/3} a_n^{-1} |U_{0,n}(\boldsymbol{x}) - U_{1,n}(\boldsymbol{x})| \\
& \le \mathcal{O}(1) \Bigg| \frac{a_n^{-1}}{\sqrt{f_{\boldsymbol{X}}(\boldsymbol{x})\sigma_n^2(\boldsymbol{x})}} \Bigg| \Bigg\{ 2a_n \Bigg| \int_{B_{\boldsymbol{x}}} dK((\boldsymbol{x}-\boldsymbol{u})/h) \Bigg| + 2a_n \Bigg| K\Big(\frac{\boldsymbol{x}-\cdot_2}{h}\Big)\Bigg| (B_{\boldsymbol{x}}) \\
& \qquad + 2a_n \Bigg| \sum_{\alpha_1=1,\boldsymbol{\alpha}_2\in\{0,1\}^d-\{\mathbf{l}_2\}} \int_{(B_{\boldsymbol{x}})_{\boldsymbol{\alpha}_2}} \partial^{\boldsymbol{\alpha}_2} K\Big(\frac{\boldsymbol{x}-\cdot_2}{h}\Big)\Bigg| (B_{\boldsymbol{x}})_{\mathbf{l}_2-\boldsymbol{\alpha}_2} \\
& \qquad + 2a_n \Bigg| \sum_{\alpha_1=0,\boldsymbol{\alpha}_2\in\{0,1\}^d-\{\mathbf{0}_2\}} \int_{(B_{\boldsymbol{x}})_{\boldsymbol{\alpha}_2}} \partial^{\boldsymbol{\alpha}_2} K\Big(\frac{\boldsymbol{x}-\cdot_2}{h}\Big)\Bigg| (B_{\boldsymbol{x}})_{\mathbf{l}_2-\boldsymbol{\alpha}_2} \Bigg\}, \; a.s.
\end{aligned}
\tag{B.12}
$$

By (A1), K is of bounded variation in the sense of Hardy and Krause (Owen (2005) definition 2), (B.12) is almost surely bounded. □

Lemma B.5. $\|U_{1,n} - U_{2,n}\| = \mathcal{O}_p\big(h^{d/2}\big)$.

Proof of Lemma B.5. Since $B_n\big(T(y,\boldsymbol{u})\big) = W_n\big(T(y,\boldsymbol{u})\big) - F(y,\boldsymbol{u})W_n(1,...,1)$, we obtain by a change of variables and a first order approximation to $f(y, \boldsymbol{x} - h\boldsymbol{v})$:

$$
\begin{aligned}
&\|U_{1,n} - U_{2,n}\| \\
&\lesssim h^{d/2}\Big|\int K(\boldsymbol{v})\,d\boldsymbol{v}\Big| \left\| \frac{1}{\sqrt{f_{\boldsymbol{X}}(\boldsymbol{x})\sigma_n^2(\boldsymbol{x})}} \int_{\Gamma_n} \big|y^* - \mu(\boldsymbol{x})\big| f(y^*, \boldsymbol{x}) dy^* + \mathcal{O}(h) \right\| |W(1,...,1)|
\end{aligned}
$$

Note that $|W(1,...,1)| = \mathcal{O}_p(1)$, $Y_i^* - \mu(\boldsymbol{x})$ has a finite second moment by (A2)(i) is uniformly bounded on $\mathcal{X}$. $\square$

Lemma B.6. $\|U_{2,n} - U_{3,n}\| = \mathcal{O}_p\big(h^{1-\delta}\big)$, *where* $0 < \delta < 1$.

Proof of Lemma B.6. Define

$$
\begin{aligned}
V_n(\boldsymbol{x}) &\stackrel{\text{def}}{=} U_{2,n}(\boldsymbol{x}) - U_{3,n}(\boldsymbol{x}) \\
&= \frac{1}{\sqrt{h^d f_{\boldsymbol{X}}(\boldsymbol{x})\sigma_n^2(\boldsymbol{x})}} \int\int_{\Gamma_n} \big\{\big(y^* - \mu(\boldsymbol{x})\big) - \big(y^* - \mu(\boldsymbol{u})\big)\big\} K\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{h}\right) dW\big(T(y^*, \boldsymbol{u})\big) \\
&= \frac{1}{\sqrt{h^d f_{\boldsymbol{X}}(\boldsymbol{x})\sigma_n^2(\boldsymbol{x})}} \int\int_{\Gamma_n} \big(\mu(\boldsymbol{u}) - \mu(\boldsymbol{x})\big) K\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{h}\right) dW\big(T(y^*, \boldsymbol{u})\big).
\end{aligned} \tag{B.13}
$$

Via applying mean value theorem to μ and (A3), and the fact that $F_{Y^*|\boldsymbol{X}}(y^*|\boldsymbol{u}) \le 1$ for all y^* and $\boldsymbol{u}$,

$$
\begin{aligned}
\mathsf{E}\left[\left(\frac{V(\boldsymbol{x})}{h}\right)^2\right] &= \frac{1}{h^{d+2} f_{\boldsymbol{X}}(\boldsymbol{x})\sigma_n^2(\boldsymbol{x})} \int\int_{\Gamma_n} \big(\mu(\boldsymbol{u}) - \mu(\boldsymbol{x})\big)^2 K^2\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{h}\right) f(y^*, \boldsymbol{u}) dy^* d\boldsymbol{u} \\
&\le \frac{C_{\mu'}}{h^{d+2} f_{\boldsymbol{X}}(\boldsymbol{x})\sigma_n^2(\boldsymbol{x})} \int \big(F_{Y^*|\boldsymbol{X}}(a_n|\boldsymbol{u}) - F_{Y^*|\boldsymbol{X}}(-a_n|\boldsymbol{u})\big) |\boldsymbol{x} - \boldsymbol{u}|^2 K^2\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{h}\right) f_{\boldsymbol{X}}(\boldsymbol{u}) d\boldsymbol{u} \\
&\le \frac{C^2}{h^2 f_{\boldsymbol{X}}(\boldsymbol{x})\sigma_n^2(\boldsymbol{x})} \int K^2(\boldsymbol{z}) |h\boldsymbol{z}|^2 f_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{z} + \mathcal{O}(h) \le \frac{2C^2 dA^2}{\sigma_n^2(\boldsymbol{x})} \|K\|_2^2 + \mathcal{O}(h),
\end{aligned}
$$

where the last inequality follows from Assumption (A1) that K has bounded support. Therefore,

$$\sigma^2 \stackrel{\text{def}}{=} \sup_{\boldsymbol{x}\in\mathcal{X}} \mathsf{E}\left[\left(\frac{V_n(\boldsymbol{x})}{h}\right)^2\right] \le C + \mathcal{O}(h), \tag{B.14}$$

Now we compute $d(\boldsymbol{s},\boldsymbol{t})$ defined in Lemma C.2.

$$\mathsf{E}\left[\left(\frac{V(\boldsymbol{t})-V(\boldsymbol{s})}{h}\right)^2\right] \le \frac{2}{h^{d+2}}\int\int_{\Gamma_n} (\mu(\boldsymbol{s})-\mu(\boldsymbol{t}))^2 K^2\left(\frac{\boldsymbol{s}-\boldsymbol{u}}{h}\right) f(y^*,\boldsymbol{u})dy^*d\boldsymbol{u}+$$
$$\frac{2}{h^{d+2}}\int\int_{\Gamma_n} (\mu(\boldsymbol{u})-\mu(\boldsymbol{t}))^2 \left[K\left(\frac{\boldsymbol{t}-\boldsymbol{u}}{h}\right)-K\left(\frac{\boldsymbol{s}-\boldsymbol{u}}{h}\right)\right]^2 f(y^*,\boldsymbol{u})dy^*d\boldsymbol{u} \stackrel{\text{def}}{=} I_1+I_2.$$

We estimate I_1 first,

$$\begin{aligned} I_1 &\le \frac{2C_{\mu'}}{h^{d+2}}\int \|\boldsymbol{t}-\boldsymbol{s}\|_\infty^2 K^2\left(\frac{\boldsymbol{s}-\boldsymbol{u}}{h}\right) f_{\boldsymbol{X}}(\boldsymbol{u})d\boldsymbol{u} \\ &\le \frac{2C_{\mu'}}{h^{d+2}}\|\boldsymbol{s}-\boldsymbol{t}\|_\infty^2\int K^2\left(\frac{\boldsymbol{s}-\boldsymbol{u}}{h}\right) f_{\boldsymbol{X}}(\boldsymbol{u})d\boldsymbol{u} \lesssim \frac{\|\boldsymbol{s}-\boldsymbol{t}\|_\infty^2}{h^2}, \end{aligned}$$

where by (A3) there exists constant $C_{\mu'}>0$ such that $\sum_{j=1}^d \sup_{\boldsymbol{x}\in\mathcal{X}}|\partial^j\mu(\boldsymbol{x})| \le C_{\mu'}$, and the last inequality uses (A1) that K is bounded.

For I_2, by (A3), we obtain

$$\begin{aligned} I_2 &\le \frac{2C_{\mu'}}{h^{d+2}}\int \|\boldsymbol{t}-\boldsymbol{u}\|^2\left[K\left(\frac{\boldsymbol{t}-\boldsymbol{u}}{h}\right)-K\left(\frac{\boldsymbol{s}-\boldsymbol{u}}{h}\right)\right]^2 f_{\boldsymbol{X}}(\boldsymbol{u})d\boldsymbol{u} \\ &\le \frac{4C_{\mu'}}{h^{d+2}}\frac{\|\boldsymbol{s}-\boldsymbol{t}\|_\infty}{h}\int \|h\boldsymbol{z}\|^2\left|K(\boldsymbol{z})-K\left(\boldsymbol{z}+\frac{\boldsymbol{s}-\boldsymbol{t}}{h}\right)\right| f_{\boldsymbol{X}}(\boldsymbol{x}+h\boldsymbol{z})d\boldsymbol{z} \\ &\le 4C_{\mu'}\frac{\|\boldsymbol{s}-\boldsymbol{t}\|_\infty}{h}\left[\int_{[-A,A]^d}\|\boldsymbol{z}\|^2|K(\boldsymbol{z})|d\boldsymbol{z}+\int_{[-A,A]^d-\frac{\boldsymbol{s}-\boldsymbol{t}}{h}}\|\boldsymbol{z}\|^2\left|K\left(\boldsymbol{z}+\frac{\boldsymbol{s}-\boldsymbol{t}}{h}\right)\right|d\boldsymbol{z}\right] \lesssim \frac{\|\boldsymbol{s}-\boldsymbol{t}\|_\infty}{h}, \end{aligned}$$

where in the last inequality we again uses the bounded support property of $K(\cdot)$ in (A1). Thus, for the function γ defined in Lemma C.2, we obtain the estimate $\gamma(\epsilon) \le C\max\{\sqrt{\epsilon/h}, \epsilon/h\} \le C\sqrt{\epsilon}/h$ for $\epsilon<1$ and $C>0$, and

$$Q(m) \le C'(2+\sqrt{2})\int_1^\infty \frac{\sqrt{m2^{-y^2}}}{h}dy \le C'\frac{\sqrt{m}}{h},$$

where $C' > 0$ and $m < 2$. Observe that the graph of the inverse of a univariate, injective function $Q(m)$ is its reflection about the diagonal line, so the inverse of an upper bound for Q would be a lower bound for Q^{-1}. Given the upper bound above, we can therefore bound Q^{-1} from below by

$$2 > Q^{-1}(a) \geq (C')^{-2} h^2 a^2.$$

Let $a = \eta h^{-\delta}$ for $0 < \delta < 1$ and arbitrary $\eta > 0$. $Q^{-1}(a^{-1}) = Q^{-1}\big(\eta^{-1} h^{\delta}\big) \geq (C')^{-2} \eta^{-2} h^{2+2\delta}$. $\sigma > a^{-1}$ when n is large. Applying Lemma C.2 yields

$$\mathrm{P}\left\{\sup_{\boldsymbol{x}\in\mathcal{X}} \left|\frac{V_n(\boldsymbol{x})}{\sqrt{h}}\right| > \eta h^{-\delta}\right\} \leq 2^{2d+2}(2r)^d \eta^{2d} h^{-2d(1+\delta)} \frac{2\sigma}{\eta h^{-\delta}} \exp\left\{-\frac{\eta^2 h^{-2\delta}}{8\sigma^2}\right\},$$

for large enough n and all $\eta > 0$, where r is a constant depending on $\mathcal{X}$. □

Lemma B.7. $U_{3,n}(\boldsymbol{x}) \overset{d}{=} U_{4,n}(\boldsymbol{x})$ *for all* $\boldsymbol{x} \in \mathcal{X}$.

Proof of Lemma B.7. The proof resembles the proof for Lemma A.5 in the supplement material of Chao et al. (2017) and is omitted for brevity. □

Lemma B.8. $\|U_{4,n} - U_{5,n}\| = \mathcal{O}_p\big(h^{1-\delta}\big)$, *where* $0 < \delta < 1$.

Proof of Lemma B.8. We will proceed as in Lemma B.6 and apply Lemma C.2. Set

$$\begin{aligned}\widetilde{V}_n(\boldsymbol{x}) &\overset{\text{def}}{=} U_{4,n}(\boldsymbol{x}) - U_{5,n}(\boldsymbol{x}) \\ &= \frac{1}{\sqrt{h^d f_{\boldsymbol{X}}(\boldsymbol{x})\sigma_n^2(\boldsymbol{x})}} \int \left(\sqrt{\sigma_n^2(\boldsymbol{u}) f_{\boldsymbol{X}}(\boldsymbol{u})} - \sqrt{\sigma_n^2(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})}\right) K\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{h}\right) dW(\boldsymbol{u}).\end{aligned}$$

To apply Lemma C.2, we need to estimate

$$\begin{aligned}\mathsf{E}\left[\left(\frac{\widetilde{V}_n(\boldsymbol{t})}{h}\right)^2\right] &= \frac{1}{h^{d+2} f_{\boldsymbol{X}}(\boldsymbol{t})\sigma_n^2(\boldsymbol{t})} \int \left(\sqrt{\sigma_n^2(\boldsymbol{u}) f_{\boldsymbol{X}}(\boldsymbol{u})} - \sqrt{\sigma_n^2(\boldsymbol{t}) f_{\boldsymbol{X}}(\boldsymbol{t})}\right)^2 K^2\left(\frac{\boldsymbol{t}-\boldsymbol{u}}{h}\right) d\boldsymbol{u} \\ &= \frac{1}{h^{d+2} f_{\boldsymbol{X}}(\boldsymbol{t})\sigma_n^2(\boldsymbol{t})} \int \Big\{\sqrt{\sigma_n^2(\boldsymbol{u})}\Big[\sqrt{f_{\boldsymbol{X}}(\boldsymbol{u})} - \sqrt{f_{\boldsymbol{X}}(\boldsymbol{t})}\Big] \\ &\qquad + \sqrt{f_{\boldsymbol{X}}(\boldsymbol{x})}\Big[\sqrt{\sigma_n^2(\boldsymbol{u})} - \sqrt{\sigma_n^2(\boldsymbol{t})}\Big]\Big\}^2 K^2\left(\frac{\boldsymbol{t}-\boldsymbol{u}}{h}\right) d\boldsymbol{u}.\end{aligned}$$

Note that $0 < \inf_{\boldsymbol{x}\in\mathcal{X}} f_{\boldsymbol{X}}(\boldsymbol{x}) < \sup_{\boldsymbol{x}\in\mathcal{X}} f_{\boldsymbol{X}}(\boldsymbol{x}) < \infty$ by (A4), $\sigma_n(\boldsymbol{x})^2 \le \sigma(\boldsymbol{x})^2 < \infty$ for all n and $\boldsymbol{x} \in \mathcal{X}$ by (A2)(i), and for sufficiently large n, $\sigma_n(\boldsymbol{x})^2 > 0$. Hence,

$$\mathsf{E}\left[\left(\frac{\widetilde{V}_n(\boldsymbol{t})}{h}\right)^2\right] \le 2Ch^{-d-2}\left\{\int \left[\sqrt{f_{\boldsymbol{X}}(\boldsymbol{u})} - \sqrt{f_{\boldsymbol{X}}(\boldsymbol{t})}\right]^2 K^2\left(\frac{\boldsymbol{t}-\boldsymbol{u}}{h}\right) d\boldsymbol{u} + \int \left[\sqrt{\sigma_n^2(\boldsymbol{u})} - \sqrt{\sigma_n^2(\boldsymbol{t})}\right]^2 K^2\left(\frac{\boldsymbol{t}-\boldsymbol{u}}{h}\right) d\boldsymbol{u}\right\},$$

We have $\sigma_n^2(\boldsymbol{x}) = \sigma^2(\boldsymbol{x}) - \widetilde{\sigma}_n^2(\boldsymbol{x})$. By (B.10) in Lemma B.3, $\sup_{\boldsymbol{x}\in\mathcal{X}} \widetilde{\sigma}_n^2(\boldsymbol{x}) = \mathcal{O}(h^{3d}\log^{-1} n)$. For large enough n such that $\sigma_n^2(\boldsymbol{x}) > 0$,

$$\left[\sqrt{\sigma_n^2(\boldsymbol{u})} - \sqrt{\sigma_n^2(\boldsymbol{t})}\right]^2 = \left[\frac{\sigma_n^2(\boldsymbol{u}) - \sigma_n^2(\boldsymbol{t})}{\sqrt{\sigma_n^2(\boldsymbol{u})} + \sqrt{\sigma_n^2(\boldsymbol{t})}}\right]^2 \le C\left[\widetilde{\sigma}_n^2(\boldsymbol{t}) - \widetilde{\sigma}_n^2(\boldsymbol{u})\right]^2 = \mathcal{O}(h^{6d}\log^{-2} n).$$

Moreover, $\sqrt{f_{\boldsymbol{X}}(\boldsymbol{x})}$ is continuously differentiable on $\mathcal{X}$ by assumption (A4). Along with $\int |z|^2 K(z) < \infty$ by (A1), we have

$$\sup_{\boldsymbol{t}\in\mathcal{X}} \mathsf{E}\left[\left(\frac{\widetilde{V}_n(\boldsymbol{t})}{h}\right)^2\right] = \mathcal{O}(1).$$

On the other hand,

$$\begin{aligned}
&\mathsf{E}\left[\left(\frac{\widetilde{V}_n(\boldsymbol{t}) - \widetilde{V}_n(\boldsymbol{s})}{h}\right)^2\right] \\
&\le Ch^{-d-2}\int \left\{\left[\sqrt{\sigma_n^2(\boldsymbol{u})f_{\boldsymbol{X}}(\boldsymbol{u})} - \sqrt{\sigma_n^2(\boldsymbol{t})f_{\boldsymbol{X}}(\boldsymbol{t})}\right] K\left(\frac{\boldsymbol{t}-\boldsymbol{u}}{h}\right) \right.\\
&\qquad\left. - \left[\sqrt{\sigma_n^2(\boldsymbol{u})f_{\boldsymbol{X}}(\boldsymbol{u})} - \sqrt{\sigma_n^2(\boldsymbol{s})f_{\boldsymbol{X}}(\boldsymbol{s})}\right] K\left(\frac{\boldsymbol{s}-\boldsymbol{u}}{h}\right)\right\}^2 d\boldsymbol{u} \\
&= Ch^{-d-2}\int \left\{\left[\sqrt{\sigma_n^2(\boldsymbol{u})f_{\boldsymbol{X}}(\boldsymbol{u})} - \sqrt{\sigma_n^2(\boldsymbol{t})f_{\boldsymbol{X}}(\boldsymbol{t})}\right]\left[K\left(\frac{\boldsymbol{t}-\boldsymbol{u}}{h}\right) - K\left(\frac{\boldsymbol{s}-\boldsymbol{u}}{h}\right)\right] \right.\\
&\qquad\left. + \left[\sqrt{\sigma_n^2(\boldsymbol{t})f_{\boldsymbol{X}}(\boldsymbol{t})} - \sqrt{\sigma_n^2(\boldsymbol{s})f_{\boldsymbol{X}}(\boldsymbol{s})}\right] K\left(\frac{\boldsymbol{s}-\boldsymbol{u}}{h}\right)\right\}^2 d\boldsymbol{u} \\
&\le 2Ch^{-d-2}\int \left[\sqrt{\sigma_n^2(\boldsymbol{u})f_{\boldsymbol{X}}(\boldsymbol{u})} - \sqrt{\sigma_n^2(\boldsymbol{t})f_{\boldsymbol{X}}(\boldsymbol{t})}\right]^2\left[K\left(\frac{\boldsymbol{t}-\boldsymbol{u}}{h}\right) - K\left(\frac{\boldsymbol{s}-\boldsymbol{u}}{h}\right)\right]^2 d\boldsymbol{u} \\
&\qquad + 2Ch^{-d-2}\int \left[\sqrt{\sigma_n^2(\boldsymbol{t})f_{\boldsymbol{X}}(\boldsymbol{t})} - \sqrt{\sigma_n^2(\boldsymbol{s})f_{\boldsymbol{X}}(\boldsymbol{s})}\right]^2 K^2\left(\frac{\boldsymbol{s}-\boldsymbol{u}}{h}\right) d\boldsymbol{u} \stackrel{\text{def}}{=} I_1 + I_2.
\end{aligned}$$

From

$$\left[\sqrt{\sigma_n^2(\boldsymbol{t})f_{\boldsymbol{X}}(\boldsymbol{t})}-\sqrt{\sigma_n^2(\boldsymbol{s})f_{\boldsymbol{X}}(\boldsymbol{s})}\right]^2=\left[\frac{\sigma_n^2(\boldsymbol{t})f_{\boldsymbol{X}}(\boldsymbol{t})-\sigma_n^2(\boldsymbol{s})f_{\boldsymbol{X}}(\boldsymbol{s})}{\sqrt{\sigma_n^2(\boldsymbol{t})f_{\boldsymbol{X}}(\boldsymbol{t})}+\sqrt{\sigma_n^2(\boldsymbol{s})f_{\boldsymbol{X}}(\boldsymbol{s})}}\right]^2\leq C\|\boldsymbol{t}-\boldsymbol{s}\|_\infty^2,$$

we obtain

$$I_2=C\frac{\|\boldsymbol{t}-\boldsymbol{s}\|_\infty^2}{h^2}.$$

By change of variables and a similar argument as to bound I_2 in the proof of Lemma B.6, it follows

$$I_1\leq C\frac{\|\boldsymbol{s}-\boldsymbol{t}\|_\infty}{h}.$$

Computing $\gamma(\epsilon)$, $Q(m)$, $Q^{-1}(a)$ as in Lemma B.6. Setting $a=\eta h^{-\delta}$ for $0<\delta<1$ and arbitrary $\eta>0$, and applying Lemma C.2 as in Lemma B.6 give the desired result. $\square$

C Auxiliary Results

Lemma C.1 (Bickel and Wichura (1971): Tightness of processes on a multidimensional cube)**.** *If $\{X_n\}_{n=1}^\infty$ is a sequence in $D[0,1]^d$, $\mathrm{P}(X\in[0,1]^d)=1$. For neighboring blocks B,C in $[0,1]^d$ (see Definition B.1) constants $\lambda_1+\lambda_2>1$, $\gamma_1+\gamma_2>0$, $\{X_n\}_{n=1}^\infty$ is tight if*

$$\mathsf{E}[|X_n(B)|^{\gamma_1}|X_n(C)|^{\gamma_2}]\leq\lambda(B)^{\lambda_1}\lambda(C)^{\lambda_2}, \tag{C.1}$$

where $\mu(\cdot)$ is a finite nonnegative measure on $[0,1]^d$ (for example, Lebesgue measure), where the increment of X_n on the block B is defined by

$$X_n(B)=\sum_{\boldsymbol{\alpha}\in\{0,1\}^d}(-1)^{d-|\boldsymbol{\alpha}|}X_n\big(\boldsymbol{s}+\boldsymbol{\alpha}\odot(\boldsymbol{t}-\boldsymbol{s})\big).$$

Lemma C.2 (Meerschaert, M. M., Wang, W. and Xiao, Y. (2013))**.** *Suppose that $Y=$*

$\{Y(\boldsymbol{t}), \boldsymbol{t} \in \mathbb{R}^d\}$ *is a centered Gaussian random field with values in* $\mathbb{R}$*, and denote*

$$d(\boldsymbol{s}, \boldsymbol{t}) \stackrel{\text{def}}{=} d_Y(\boldsymbol{s}, \boldsymbol{t}) = \left(\mathsf{E}|Y(\boldsymbol{t}) - Y(\boldsymbol{s})|^2\right)^{1/2}, \quad \boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}^d.$$

Let $\mathcal{X}$ *be a compact set contained in a cube with length* r *in* $\mathbb{R}^d$ *and let* $\sigma^2 = \sup_{\boldsymbol{t}\in\mathcal{X}} \mathsf{E}[Y(\boldsymbol{t})^2]$*. For any* $m > 0$*,* $\epsilon > 0$*, define*

$$\gamma(\epsilon) = \sup_{\boldsymbol{s},\boldsymbol{t}\in\mathcal{X}, \|\boldsymbol{s}-\boldsymbol{t}\|\le\epsilon} d(\boldsymbol{s}, \boldsymbol{t})$$

and

$$Q(m) = (2 + \sqrt{2}) \int_1^\infty \gamma(m2^{-y^2}) dy.$$

Then, for all $a > 0$ *which satisfy* $a \ge (1 + 4d \log 2)^{1/2}(\sigma + a^{-1})$*,*

$$\mathrm{P}\left\{\sup_{\boldsymbol{t}\in S} |Y(\boldsymbol{t})| > a\right\} \le 2^{2d+2} \left(\frac{r}{Q^{-1}(1/a)} + 1\right)^d \frac{\sigma + a^{-1}}{a} \exp\left\{-\frac{a^2}{2(\sigma + a^{-1})^2}\right\}, \qquad \text{(C.2)}$$

where $Q^{-1}(a) = \sup\{m : Q(m) \le a\}$*.*

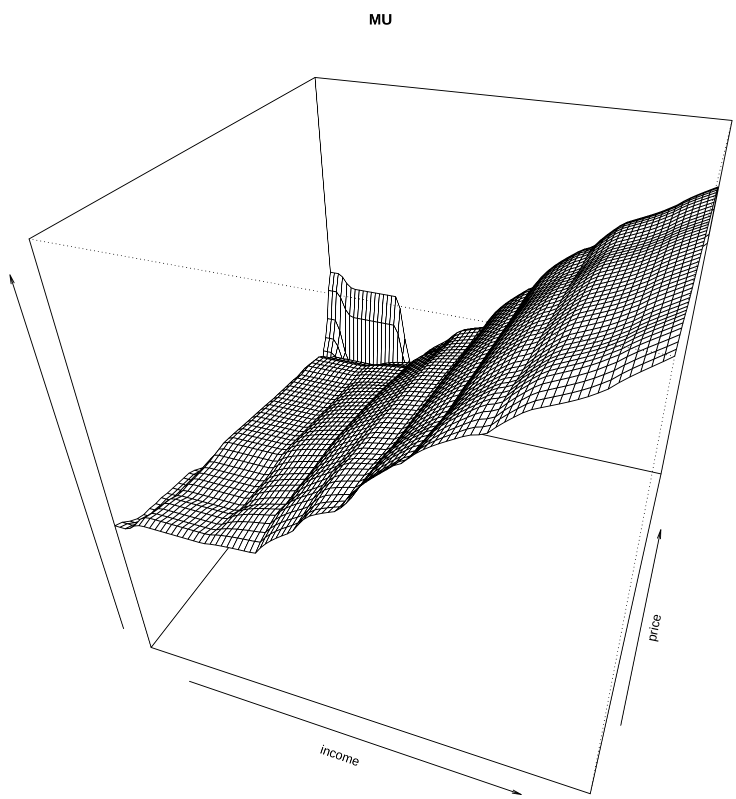
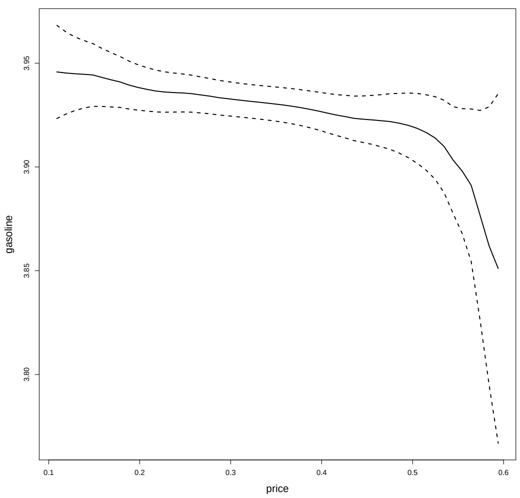


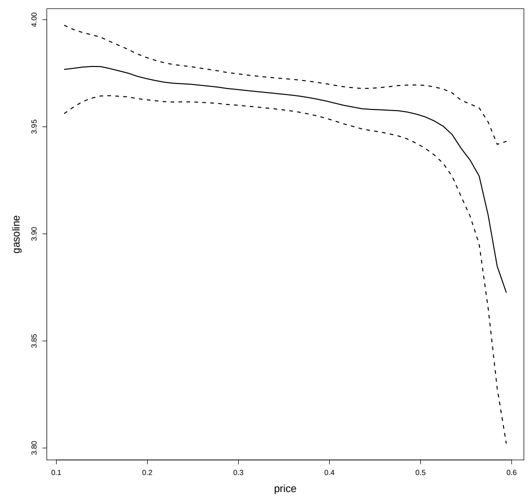
Figure 1: *U.S.* gasoline demand in income and price. The unit is log scale of *U.S.* dollar. The bandwidth is obtained through undersmoothing of the *GCV*-chosen one. EIV_UCS

Table 1: Variables in the *RTECS* Surveys from Year 2001. The unit is in *U.S.* dollar.

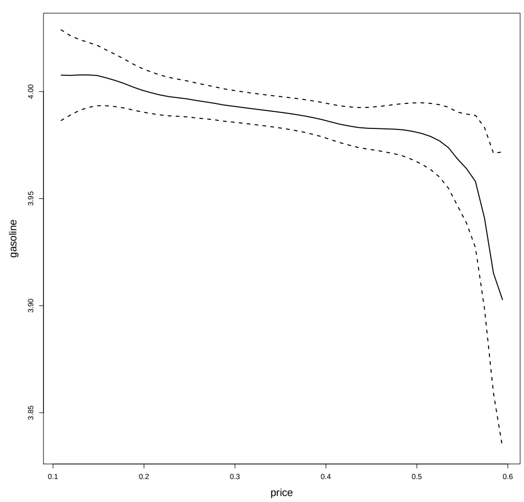
Variable Description	Mean	Std. Dev.	OLS estimate (std. error)
Log of Total Miles Traveled ($TOTMILES$)	4.214	0.423	-
Log of Income ($INCOME$)	4.623	0.333	0.190 *(0.008)*
Log of Gasoline Price ($PRICE$)	0.125	0.026	−0.516 *(0.086)*
Log of Number of Drivers ($DRVR$)	0.236	0.170	0.266 *(0.026)*
Log of Number of Vehicles ($VEHS$)	0.248	0.212	0.825 *(0.014)*
Log of Household Size ($HHSIZE$)	0.344	0.227	0.008 *(0.031)*
Child Dummy ($CHILD$)	0.348	0.476	2.397 *(0.030)*
Gender Dummy (SEX)	0.584	0.493	−0.029 *(0.005)*
Urban Residence Dummy ($RURAL$)	0.254	0.435	0.086 *(0.005)*
Region Dummy Variables ($Region$)			
New England	0.091	0.288	0.812 *(0.011)*
Middle Atlantic	0.187	0.390	0.802 *(0.011)*
East North Central	0.069	0.254	0.783 *(0.012)*
West North Central	0.026	0.160	3.270 *(0.038)*
South Atlantic	0.090	0.286	3.146 *(0.036)*
East South Central	0.102	0.302	2.926 *(0.035)*
West South Central	0.055	0.227	3.214 *(0.038)*
Mountain	0.199	0.399	3.139 *(0.038)*
Pacific	0.181	0.385	3.032 *(0.037)*



(a) 25th-percentile income

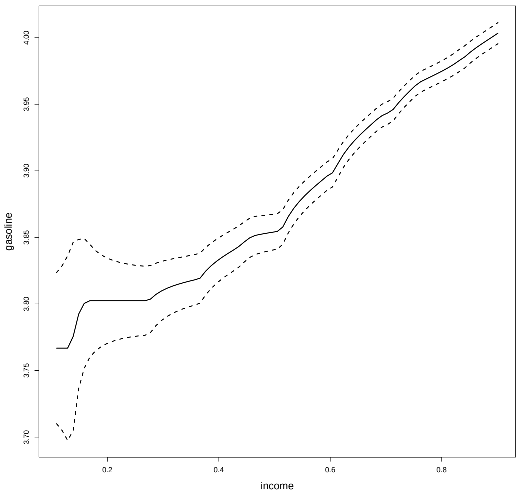


(b) 50th-percentile income

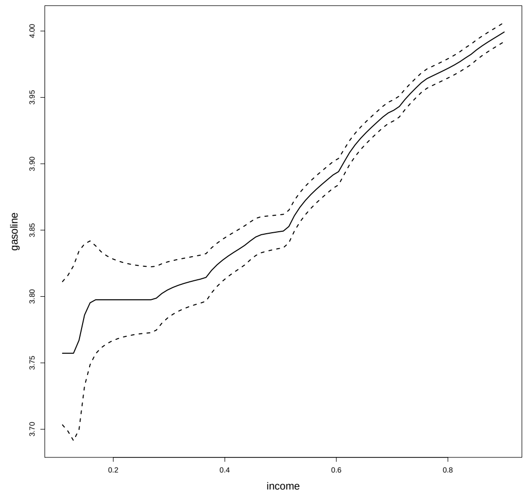


(c) 75th-percentile income

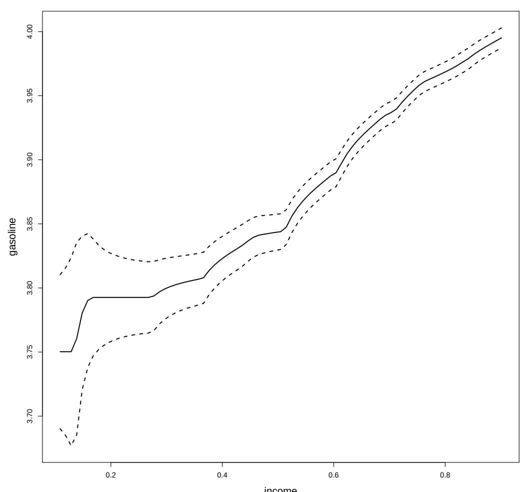
Figure 2: *U.S.* gasoline demand; The dashed band is 95% simultaneous confidence region. EIV_UCS



(a) 25th-percentile price



(b) 50th-percentile price



(c) 75th-percentile price

Figure 3: *U.S.* gasoline demand; The dashed band is 95% simultaneous confidence region. EIV_UCS

IRTG 1792 Discussion Paper Series 2020

For a complete list of Discussion Papers published, please visit http://irtg1792.hu-berlin.de.

001 "Estimation and Determinants of Chinese Banks' Total Factor Efficiency: A New Vision Based on Unbalanced Development of Chinese Banks and Their Overall Risk" by Shiyi Chen, Wolfgang K. Härdle, Li Wang, January 2020.

002 "Service Data Analytics and Business Intelligence" by Desheng Dang Wu, Wolfgang Karl Härdle, January 2020.

003 "Structured climate financing: valuation of CDOs on inhomogeneous asset pools" by Natalie Packham, February 2020.

004 "Factorisable Multitask Quantile Regression" by Shih-Kang Chao, Wolfgang K. Härdle, Ming Yuan, February 2020.

005 "Targeting Cutsomers Under Response-Dependent Costs" by Johannes Haupt, Stefan Lessmann, March 2020.

006 "Forex exchange rate forecasting using deep recurrent neural networks" by Alexander Jakob Dautel, Wolfgang Karl Härdle, Stefan Lessmann, Hsin-Vonn Seow, March 2020.

007 "Deep Learning application for fraud detection in financial statements" by Patricia Craja, Alisa Kim, Stefan Lessmann, May 2020.

008 "Simultaneous Inference of the Partially Linear Model with a Multivariate Unknown Function" by Kun Ho Kim, Shih-Kang Chao, Wolfgang K. Härdle, May 2020.

IRTG 1792, Spandauer Strasse 1, D-10178 Berlin
http://irtg1792.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the IRTG 1792.