

SFB 649 Discussion Paper 2016-024

Simultaneous Inference for the Partially Linear Model with a Multivariate Unknown Function when the Covariates are Measured with Errors

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



* Hanyang University, South Korea

*² Purdue University, United States of America

*³ Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

<http://sfb649.wiwi.hu-berlin.de>
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin



SFB 649 ECONOMIC RISK BERLIN

Simultaneous Inference for the Partially Linear Model with a Multivariate Unknown Function when the Covariates are Measured with Errors*

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

July 25, 2016

Abstract

In this paper, we analyze the nonparametric part of a partially linear model when the covariates in parametric and non-parametric parts are subject to measurement errors. Based on a two-stage semi-parametric estimate, we construct a uniform confidence surface of the multivariate function for simultaneous inference. The developed methodology is applied to perform inference for the U.S. gasoline demand where the income and price variables are measured with errors. The empirical results strongly suggest that the linearity of the *U.S.* gasoline demand is rejected.

Key words: Measurement error, Partially linear model, Regression calibration, Non-parametric function, Semi-parametric regression, Uniform confidence surface, Simultaneous inference, U.S. Gasoline demand, Non-linearity

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".

[†]Department of Economics and Finance, Hanyang University, Seoul, South Korea. E-mail: kunhokim8@gmail.com. Tel: +82-2-2220-1022.

[‡]Department of Statistics, Purdue University, West Lafayette, IN 47906, U.S.A. E-mail: skchao74@purdue.edu. Tel: +1 (765) 496-9544. Fax: +1 (765) 494-0558. 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 makes it not only possible to include discrete predictors but also creates a chance of parts of the model to be estimated with high precision. These and other reasons have made this model class very successful, e.g. [Härdle et al. \(2000\)](#). In several applications though, one has a data structure that is not covered by the inferential analysis of the literature so far.

To be more specific, consider the semi-parametric model:

$$Y_i = \mathbf{S}_i^\top \boldsymbol{\beta} + g(\mathbf{T}_i) + \zeta_i, \quad \mathbf{S}_i = \mathbf{Z}_i + \xi_i, \quad \mathbf{T}_i = \mathbf{X}_i + \eta_i \quad (1.1)$$

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 $i = 1, \dots, n$, 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 and is considered an improvement from the literature ([Koul and Song, 2010](#)). The distribution of η_i is assumed to be known. The measurement errors ξ_i and η_i are called the Berkson measurement error ([Berkson \(1950\)](#)) and are used typically in situations where true individual observations are not available but the average for the group where the individuals belong is available. See Section 1.1. for the justification of using Berkson errors in our model framework. The model (1.1) is widely used due to its flexibility to combine the parametric linear part $\mathbf{S}_i^\top \boldsymbol{\beta}$ and the non-parametric non-linear component $g(\mathbf{T}_i)$. See [Härdle et al. \(2000\)](#) for more on the partially linear model framework.

The important aspect of (1.1) is that it allows the covariates to be *measured with errors*. It is well-known that economic variables, such as income and price variables, are often measured with some degree of inaccuracy. For example, (1.1) could represent a demand function for certain goods, where \mathbf{T}_i is a (2×1) vector of income and price variables, respectively. The income and price variables in $g(\cdot, \cdot)$ are likely to be measured with errors because individuals are usually unwilling to reveal their true incomes and because the prices are often estimated rather than observed directly. The framework (1.1) can address this issue effectively. Section 1.1. discusses this in detail and also explains why the Berkson error assumption particularly fits the income and price variables used in our study. Given that economic theory usually provides little guidance for how to model the demand function ([Blundell et al. \(2012\)](#)), one could borrow a partially linear semi-parametric structure in the error-in-variable setting, such as (1.1). For these reasons, (1.1) could be a reasonable approximation for the true underlying demand function, among others.

Under the framework (1.1), our goal is to test the hypothesis on $g(\cdot)$, such that:

$$H_{10} : g(\mathbf{t}) = g_0(\mathbf{t}), \quad \forall \mathbf{t} \in \mathbb{R}^d \quad (1.2)$$

where $g_0(\cdot)$ is some known real-valued function. For instance, $g(\mathbf{t})$ could stand for a true demand function while $g_0(\mathbf{t})$ 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 (1.2) under (1.1). If the null hypothesis is accepted, then the traditional log-linear demand function is justified. Otherwise, one needs to search for another framework that fits the demand structure better than the simple log-linear framework. The log-linear demand function for gasoline, subject to Berkson measurement errors, is discussed in detail in Section 3.1.

In the literature on measurement error models, *regression calibration* (Carroll et al. (1995)) is widely used to deal with the error-contaminated covariates. Under the assumptions,

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

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

$$Y_i = \mathbf{Z}_i^\top \boldsymbol{\beta} + \mu(\mathbf{X}_i) + \epsilon_i \quad (1.3)$$

where $\epsilon_i \stackrel{\text{def}}{=} g(\mathbf{T}_i) - \mu(\mathbf{X}_i) + \boldsymbol{\xi}_i^\top \boldsymbol{\beta} + \zeta_i$. Note here that $\mathbb{E}(\epsilon_i|\mathbf{Z}_i = \mathbf{z}, \mathbf{X}_i = \mathbf{x}) = 0$ due to $\mu(\mathbf{x}) = \mathbb{E}[g(\mathbf{T}_i)|\mathbf{X}_i = \mathbf{x}]$. 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 transformation of (1.1) to (1.3) is called the *regression calibration*, and is an effective way to deal with measurement errors because the covariates in (1.3) are observed.

Since the distribution of η_i is known, a test of H_{10} can be carried out by testing the following hypothesis instead:

$$H_{20} : \mu(\mathbf{x}) = \mu_0(\mathbf{x}), \quad \forall \mathbf{x} \in \mathcal{X} \quad (1.4)$$

where $\mu_0(\mathbf{x}) \stackrel{\text{def}}{=} \mathbb{E}[g_0(\mathbf{T}_i)|\mathbf{X}_i = \mathbf{x}]$ and \mathcal{X} is the compact range of \mathbf{X}_i in (1.3). Under Berkson errors, \mathbf{X}_i typically represents the *observed* average for a group of *unobserved* individual observations, such as individual incomes belonging to a certain income bracket. Hence the assumption of a compact range \mathcal{X} makes sense. Clearly, H_{10} implies H_{20} although the converse is not true in general. However, if the family of densities $f_\eta(\cdot)$ for η_i is complete, then $\mathbb{E}[g(\mathbf{T}_i) - g_0(\mathbf{T}_i)|\mathbf{X}_i = \mathbf{x}] = 0$ for all \mathbf{x} implies $g = g_0$ almost surely. For more on the regression calibration under measurement errors, we refer to Koul and Song (2008), Koul and Song (2010), among others.

To perform inference for H_{20} in (1.4), we construct the *uniform confidence surface* (*UCS*) of $\mu(\cdot)$ in (1.3) and observe whether the *UCS* contains the parametric specification under the null hypothesis. The construction of the *UCS* 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:

$$\lim_{n \rightarrow \infty} \mathbb{P}\{f_n(\mathbf{x}) \leq \mu(\mathbf{x}) \leq g_n(\mathbf{x}) \text{ for all } \mathbf{x} \in \mathcal{X}\} = 1 - \alpha \quad (1.5)$$

Given the *UCS* of $\mu(\cdot)$, one can test H_{20} by checking whether $f_n(\mathbf{x}) \leq \mu_0(\mathbf{x}) \leq g_n(\mathbf{x})$ holds for *all* $\mathbf{x} \in \mathcal{X}$. If the condition does not hold for *some* $\mathbf{x} \in \mathcal{X}$, then we *reject* H_{20} at level α , which leads to the rejection of H_{10} as well. That is, even if the condition holds for all $\mathbf{x} \in \mathcal{X}$ except for only one, the null hypothesis still gets rejected by the test.

1.1 Why Berkson measurement errors?

The innovation of this paper is to perform uniform inference of (1.1) under the presence of Berkson errors in the covariate terms. Hence it is essential to justify the use of Berkson errors in (1.1). To explain why such errors are used here, we consider the two different error structures:

$$\mathbf{T}_i = \mathbf{X}_i + \eta_i \tag{1.6}$$

$$\mathbf{X}_i = \mathbf{T}_i + \eta_i \tag{1.7}$$

where \mathbf{T}_i is the “unobserved” true process and \mathbf{X}_i is the “observed” proxy in (1.1). Here (1.6) is the framework with the Berkson measurement error used by the current study, while (1.7) is the classical measurement error structure. Typically, (1.6) fits better than (1.7) if the true individual observations are not available but the average value for the group where the individuals belong is available. We argue that the framework (1.6) is a good fit for the household income and gasoline price variables used in this study for the following reasons.

In our data set, the observations for the income variable are collected through household survey. Each surveyed individual is asked to choose the category where their true annual income belongs. For example, if the true income is 7,000 *USD* per year for a household, then the respondent chooses the 5,000 *USD*– 9,999 *USD* category, instead of reporting the true income. That is, we only get to observe such categorical data for the household income, not the true numerical values. In this situation, the “midpoint” of each category is typically taken as the value for the income variable, which certainly fits the framework (1.6) better than (1.7). In other words, the household-specific true but unavailable income \mathbf{T} fluctuates randomly around the available group average \mathbf{X} , resulting in Berkson-type errors. A similar example can be found in an epidemiological study in which an individual’s true exposure to some contaminant is not observed, but instead, what is available is the average concentration of the contaminant in the region where the individual lives (Schennach (2013)).

Similarly, (1.6) is also a potentially better structure for the gasoline price variable than (1.7). The data set used in this study does not include the gasoline price data observed directly from gas stations. Instead, each household reports their “approximate” total annual expenditure on gasoline and the total gallons of gasoline purchased each year. Then, the gasoline price is obtained through dividing the total expenditure by the the total gallons purchased. As argued in Hyslop and Imbens (2001) and also emphasized in Schennach (2013), individuals reporting data on their own are typically aware of the uncertainty in their estimates and, as a result, tend to report an “average” over all possible estimates consistent with the information available to them. In our case, each respondent is unlikely to report the exact total expenditure on gasoline, and instead likely to report the midpoint of the plausible range of expenditure, thus leading to Berkson-type errors for the gasoline price variable as well. For these reasons, we believe that (1.6) is a reasonable representation of the error structure for the variables used in gasoline demand, such as the household income and the gasoline price variables.

The organization of the paper is the following: Section 2 introduces the methodology proposed to perform uniform inference of the partially linear model. We first employ a two-step procedure

to estimate the model and then carry out the construction of uniform confidence surface (*UCS*) based on the estimates. Both the asymptotic-based and simulation-based constructions of *UCS* are introduced. Section 3 handles an application of the proposed methodology. We estimate and perform inference of a semi-parametric and partially linear gasoline demand function for the *U.S.* economy. 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 uniform inference are relegated to Appendix.

Notations. For any vector $\mathbf{v} = (v_1, v_2, \dots, v_p) \in \mathbb{R}^p$, we let $|\mathbf{v}| = (\sum_{i=1}^p v_i^2)^{1/2}$. For any random vector \mathbf{V} , we write $\mathbf{V} \in \mathcal{L}^q$ ($q > 0$) if $\|\mathbf{V}\|_q = [\mathbb{E}(|\mathbf{V}|^q)]^{1/q} < \infty$. 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 $\sup_{\mathbf{x}} U(\mathbf{x})$ for $\sup_{\mathbf{x} \in \mathcal{X}} U(\mathbf{x})$.

2 Methodology

Inference for (1.3) depends on the estimates of $\boldsymbol{\beta}$ and $\mu(\cdot)$. Given the semi-parametric structure of (1.3), it is natural to employ a two-step method to estimate the unknowns. That is, the parametric part of the model is first estimated, and then the estimate is used for the estimation of the non-parametric component. For an application of the two-step methodology to economic issues, see Kim (2016) among others. The smoothness condition on $\mu(\cdot)$ ensures that the simultaneous inference for the true underlying process is feasible in this study.

2.1 Estimation of $\boldsymbol{\beta}$ and $\mu(\cdot)$

To estimate the parameter vector $\boldsymbol{\beta}$ in (1.3), we employ the Robinson estimator (Robinson (1988)). The estimator is widely used in estimating the parameters in a semi-parametric framework, such as (1.3). To implement the estimation, we first define:

$$\begin{aligned} Y_i^\# &= Y_i - \mathbb{E}(Y_i | \mathbf{X}_i) \\ \mathbf{Z}_i^\# &= \mathbf{Z}_i - \mathbb{E}(\mathbf{Z}_i | \mathbf{X}_i) \end{aligned}$$

where $Y_i^\# = \mathbf{Z}_i^{\#\top} \boldsymbol{\beta} + \epsilon_i$ due to (1.3) and $\mathbb{E}(\epsilon_i | \mathbf{X}_i) = 0$. Naturally, $Y_i^\#$ and $\mathbf{Z}_i^\#$ can be estimated using the Nadaraya-Watson estimator (Nadaraya (1964); Watson (1964)):

$$\begin{aligned} \tilde{Y}_i^\# &= Y_i - \hat{g}(\mathbf{X}_i) \\ \tilde{\mathbf{Z}}_i^\# &= \mathbf{Z}_i - \hat{f}(\mathbf{X}_i) \end{aligned}$$

where

$$\hat{g}(\mathbf{x}) = \frac{\sum_{i=1}^n Y_i K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right)}, \quad \hat{f}(\mathbf{x}) = \frac{\sum_{i=1}^n \mathbf{Z}_i K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right)}.$$

Here $K\{(\mathbf{x} - \mathbf{X}_i)/h\} = K((x_1 - X_{1i})/h) \times \dots \times K((x_d - X_{di})/h)$, where $K(\cdot)$ is a kernel function and h is a smoothing parameter, called the bandwidth. Then, the Robinson estimate $\hat{\boldsymbol{\beta}}_R$ is defined

by:

$$\widehat{\boldsymbol{\beta}}_R \stackrel{\text{def}}{=} \left(\sum_{i=1}^n \widetilde{\mathbf{Z}}_i^\# \left(\widetilde{\mathbf{Z}}_i^\# \right)^\top \right)^{-1} \cdot \left(\sum_{i=1}^n \widetilde{\mathbf{Z}}_i^\# \widetilde{Y}_i^\# \right). \quad (2.1)$$

Robinson (Robinson (1988)) shows that this estimate is \sqrt{n} -consistent to the true parameter. Given $\widehat{\boldsymbol{\beta}}_R$, we are ready to estimate $\mu(\mathbf{x})$. By (1.3), Then, an estimate of $\mu(\cdot)$ is obtained by the following optimization:

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

which leads to

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

where $\widehat{f}_{\mathbf{X}}(\mathbf{x}) \stackrel{\text{def}}{=} \frac{1}{nh^d} \sum_{i=1}^n K \left(\frac{\mathbf{x} - \mathbf{X}_i}{h} \right)$ is a non-parametric estimate of $f_{\mathbf{X}}$, the joint density of $\mathbf{X}_i^\top = [X_{1i}, \dots, X_{di}]$. A popular choice for the kernel function is the Epanechnikov kernel with a compact support. The bandwidth is frequently selected by some data-driven method, such as the generalized cross-validation (GCV) (Craven and Wahba (1979)), which will be used in this study.

2.2 Uniform confidence surface (UCS) of $\mu(\cdot)$

To construct the UCS 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:

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

We define

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

Then, by a Taylor expansion on $H_n(\mathbf{x})$,

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

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

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

which leads to

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

In sum,

$$\tilde{\mu}(\mathbf{x}) - \mu(\mathbf{x}) = \frac{H_n(\mathbf{x}) - \mathbb{E}H_n(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} + R_n(\mathbf{x}) \quad (2.5)$$

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

$$\frac{\sqrt{nh^d f_{\mathbf{X}}(\mathbf{x})}}{\sigma(\mathbf{x})} |\tilde{\mu}(\mathbf{x}) - \mu(\mathbf{x})| \leq |U_n(\mathbf{x})| + \frac{\sqrt{nh^d f_{\mathbf{X}}(\mathbf{x})}}{\sigma(\mathbf{x})} |R_n(\mathbf{x})| \quad (2.6)$$

where

$$\sigma^2(\mathbf{x}) \stackrel{\text{def}}{=} \int \{y^* - \mu(\mathbf{x})\}^2 f_{Y^*|\mathbf{X}}(y^*|\mathbf{x}) dy^* \quad (2.7)$$

$$U_n(\mathbf{x}) \stackrel{\text{def}}{=} \sqrt{\frac{nh^d}{\sigma^2(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})}} (H_n(\mathbf{x}) - \mathbb{E}H_n(\mathbf{x})) \quad (2.8)$$

Here $y^* \stackrel{\text{def}}{=} y - \mathbf{z}^\top \boldsymbol{\beta}$. Let $\sigma = \sigma(\mathbf{x})$ for simplicity. Then,

$$\begin{aligned} U_n(\mathbf{x}) &= \sqrt{\frac{nh^d}{\sigma^2 f_{\mathbf{X}}(\mathbf{x})}} \left\{ \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right) (Y_i - \mathbf{Z}_i^\top \boldsymbol{\beta} - \mu(\mathbf{x})) \right. \\ &\quad \left. - \frac{1}{h^d} \mathbb{E} \left[K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right) (Y_i - \mathbf{Z}_i^\top \boldsymbol{\beta} - \mu(\mathbf{x})) \right] \right\} \\ &= \sqrt{\frac{nh^d}{\sigma^2 f_{\mathbf{X}}(\mathbf{x})}} \left\{ \int \int \frac{1}{h^d} K\left(\frac{\mathbf{x} - \mathbf{t}}{h}\right) (y^* - \mu(\mathbf{x})) dF_n(\mathbf{t}, y^*) \right. \\ &\quad \left. - \frac{1}{h^d} \int \int K\left(\frac{\mathbf{x} - \mathbf{t}}{h}\right) (y^* - \mu(\mathbf{x})) dF(\mathbf{t}, y^*) \right\} \\ &= \frac{1}{\sqrt{h^d \sigma^2 f_{\mathbf{X}}(\mathbf{x})}} \int \int K\left(\frac{\mathbf{x} - \mathbf{t}}{h}\right) (y^* - \mu(\mathbf{x})) dZ_n(\mathbf{t}, y^*) \end{aligned}$$

where $Z_n(\mathbf{x}, y^*) \stackrel{\text{def}}{=} \sqrt{n} (F_n(\mathbf{x}, y^*) - F(\mathbf{x}, y^*))$. 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:

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

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

$$T(y, \mathbf{u}) = \{F_{X_1|Y}(u_1|y), F_{X_2|Y}(u_2|u_1, y), \dots, F_{X_d|X_{d-1}, \dots, X_1, Y}(u_d|u_{d-1}, \dots, u_1, y), F_Y(y)\}. \quad (2.9)$$

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

$$\mathbb{P}\left\{\frac{\sqrt{2d \log(h^{-1})}}{\lambda_K} \left(\sup_{\mathbf{x} \in \mathcal{X}} |U_{5,n}(\mathbf{x})| - d_n\right) < u\right\} \rightarrow \exp(-2 \exp(-u)) \quad (2.10)$$

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

Proposition 2.1. *By (2.10) and Assumption (A1)-(A6),*

$$\mathbb{P}\left\{\frac{\sqrt{2d \log(h^{-1})}}{\lambda_K} \left(\sup_{\mathbf{x} \in \mathcal{X}} |U_n(\mathbf{x})| - d_n\right) < u\right\} \rightarrow \exp(-2 \exp(-u)) \quad (2.11)$$

Proof of Theorem 2.1. The proof follows directly by Lemma C.3– Lemma C.8. \square

Moreover, by Lemma B.1 in the Appendix,

$$\sup_{\mathbf{x}} |R_n(\mathbf{x})| = o_{\mathbb{P}}\left(\frac{1}{\sqrt{nh^d \log(h^{-1})}}\right) \quad (2.12)$$

Then, by (2.6), (2.11) and (2.12), we obtain the following theorem:

Theorem 2.2. *Let $b \rightarrow 0$ and $nb \rightarrow \infty$. Then,*

$$\mathbb{P}\left\{\frac{\sqrt{2d \log(h^{-1})}}{\lambda_K} \left(\sup_{\mathbf{x} \in \mathcal{X}} \frac{\sqrt{nh^d f_{\mathbf{X}}(\mathbf{x})}}{\sigma(\mathbf{x})} |\tilde{\mu}(\mathbf{x}) - \mu(\mathbf{x})| - d_n\right) < u\right\} \rightarrow \exp(-2 \exp(-u)) \quad (2.13)$$

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

Proof of Theorem 2.2. By (2.6),

$$\begin{aligned} & \frac{\sqrt{2d \log(h^{-1})}}{\lambda_K} \left(\sup_{\mathbf{x}} \frac{\sqrt{nh^d f_{\mathbf{X}}(\mathbf{x})}}{\sigma(\mathbf{x})} |\tilde{\mu}(\mathbf{x}) - \mu(\mathbf{x})| - d_n\right) \\ & \leq \frac{\sqrt{2d \log(h^{-1})}}{\lambda_K} \left(\sup_{\mathbf{x}} |U_n(\mathbf{x})| - d_n\right) + \sup_{\mathbf{x}} \frac{\sqrt{2d f_{\mathbf{X}}(\mathbf{x}) nh^d \log(h^{-1})}}{\lambda_K \sigma(\mathbf{x})} |R_n(\mathbf{x})| \end{aligned}$$

By (2.11) and (2.12), the theorem follows. \square

Since the asymptotic result (2.13) involves the infeasible estimate $\tilde{\mu}(\mathbf{x})$ that depends on the unknown β , Theorem 2.2 cannot be directly applied to construct the UCS of the unknown $\mu(\cdot)$. However, the result can be extended to derive the following theorem that can be utilized to construct the UCS of $\mu(\cdot)$:

Theorem 2.3. Let $b \rightarrow 0$ and $nb \rightarrow \infty$. Then,

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

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

Proof of Theorem 2.3. The proof follows by Theorem 2.2 and by the \sqrt{n} -consistency of the Robinson estimator $\hat{\beta}_R$ (Robinson, 1988). We omit the detail. \square

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

$$\left[\hat{\mu}(\mathbf{x}) \pm \sqrt{\frac{\sigma^2(\mathbf{x}) \lambda_K^2}{nh^d f_{\mathbf{X}}(\mathbf{x})}} \left(d_n + \frac{q_{1-\alpha}}{\sqrt{2d \log(h^{-1})}} \right) \right] \quad (2.15)$$

where $q_{1-\alpha} = -\log[-1/2 \log(1 - \alpha)]$ is the $(1 - \alpha) \times 100\%$ quantile of the Gumbel distribution in (2.13). Here $q_{0.95} = 3.66$, $\det(\Sigma) = 25/16$ and $\lambda_K = 0.6$ for the Epanechnikov kernel. Since $\sigma^2(\mathbf{x})$ and $f_{\mathbf{X}}(\mathbf{x})$ in (2.15) are unknown, the feasible UCS is

$$\left[\hat{\mu}(\mathbf{x}) \pm \sqrt{\frac{\hat{\sigma}^2(\mathbf{x}) \lambda_K^2}{nh^d \hat{f}_{\mathbf{X}}(\mathbf{x})}} \left(d_n + \frac{q_{1-\alpha}}{\sqrt{2d \log(h^{-1})}} \right) \right] \quad (2.16)$$

where the estimate of $\sigma^2(\mathbf{x})$ is given by $\hat{\sigma}^2(\mathbf{x}) \stackrel{\text{def}}{=} \frac{1}{nh^d \hat{f}_{\mathbf{X}}(\mathbf{x})} \sum_{i=1}^n \hat{\epsilon}_i^2 K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right)$ and $\hat{\epsilon}_i$ is the residual from the estimation of (1.3).

From Theorem 2.3, the convergence order to the asymptotic Gumbel distribution is $1/\sqrt{\log(n)}$, which means that the convergence is very slow. To that end, one can alternatively employ the *simulation-based* method to construct the UCS of $\mu(\mathbf{x})$ by using the following property:

$$\mathbb{P}\left(\sup_{1 \leq i \leq m} |Z_i| - d_m \leq \frac{u}{\sqrt{2 \log(m)}}\right) \rightarrow \exp(-2 \exp(-u)) \quad (2.17)$$

where $m \stackrel{\text{def}}{=} \inf\{k \in \mathbb{Z} : k \geq h^{-1}\}$ and Z_i are an IID standard normals and

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

The proof of (2.17) is given on page 98–99 of Ferguson (1996). The result is also used by Zhao and Wu (2008). Note that (2.14) and (2.17) share the same asymptotic Gumbel distribution, which means that one can use the *sampling distribution* of $\sup_{1 \leq i \leq m} |Z_i|$ to replace the quantile from the Gumbel distribution. Thus, the $(1 - \alpha) \times 100\%$ UCS of $\ell(x)$ is approximated by:

$$\left[\hat{\mu}(\mathbf{x}) \pm \sqrt{\frac{\hat{\sigma}^2(\mathbf{x}) \lambda_K^2}{nh^d \hat{f}_{\mathbf{X}}(\mathbf{x})}} \left(d_n + \frac{q_{1-\alpha}^{**}}{\sqrt{2d \log(h^{-1})}} \right) \right] \quad (2.18)$$

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 $\sup_{1 \leq i \leq m} |Z_i|$. In the following application, we employ (2.18) to construct the UCS of $\mu(\mathbf{x})$ in (1.3).

3 Application

3.1 *U.S.* gasoline demand

As an application of the methodology developed in this work, we consider the *U.S.* gasoline demand. The rapid growth in the *U.S.* gasoline consumption has been of interest to many due to its environmental consequences and the role as an economic indicator. 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 the earlier work to the case of Canadian gasoline consumption.

In contrast to Schmalensee and Stoker (1999) and Blundell et al. (2012), we focus on statistical inference of the non-parametric component in the model. Our model framework, which is a special case of (1.1) with $d = 2$, is the following:

$$\begin{aligned} \log(\text{TOTMILES}_i) = & \beta_1 \log(\text{DRVR}_i) + \beta_2 \log(\text{VEHS}_i) + \beta_3 \log(\text{HHSIZE}_i) \\ & + \beta_4 \text{CHILD}_i + \beta_5 \text{SEX}_i + \beta_6 \text{RURAL}_i \\ & + \beta_7' \text{Region} + g(\log(\text{INCOME}_i), \log(\text{PRICE}_i)) + \zeta_i \end{aligned} \quad (3.1)$$

where 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 estimated gasoline price. Here DRVR_i , VEHS_i and HHSIZE are regressors that represent the number of drivers, vehicles and family members for household i , respectively. The other regressors are dummy variables such that $\text{CHILD}_i = 1$ for a household with a child, $\text{SEX}_i = 1$ for a female respondent, and $\text{RURAL}_i = 1$ 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 descriptions and the summary statistics for these variables are given by Table 1.

As discussed in detail in Section 1.1., the household income and gasoline price variables in $g(\cdot, \cdot)$ of (3.1) are likely to be contaminated by Berkson-type 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 the data in practice, the *median* values for each category are taken for estimation, which leads to the unobserved true income randomly fluctuating around the observed median. This clearly represents Berkson-type errors. Similarly, the gasoline price data are likely to be contaminated with Berkson errors as well, 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 from the original source. 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 Berkson-type errors into the gasoline price estimates here. For these reasons, the methodology developed

in this study can be readily used to perform simultaneous inference for the unknown $g(\cdot, \cdot)$ in the gasoline demand equation (3.1) where the covariates are possibly measured with Berkson errors.

3.2 Data

The household survey data for model (3.1) are from the Residential Transportation and Energy Consumption Surveys (*RTECS*), 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 surveys are carried out every three years. The survey used in this paper was conducted in year 2001. That is, the data set used in our study is an updated version of the data used by [Schmalensee and Stoker \(1999\)](#). The total number of observations in our sample is 22,178. In this study, we only consider the observations from the households with non-zero number of miles driven, drivers, and cars owned.

The response variable in the gasoline demand is either total miles traveled or consumption of gasoline each year. Observations on mileage driven in each year were collected directly from odometer readings. These observations were then combined with estimated miles-per-gallon figures for each vehicle owned to construct total gallons of gasoline demanded. For the income variable in (3.1), households report their annual income in one of 9 ranges (in thousands of dollars). We take the *midpoint* of each interval. The lowest income category is “less than \$5,000”. For the highest income category (income over \$75,000), we use the conditional median for national household income above \$75,000, which is approximately \$100,000. The gasoline price is obtained by dividing the total household expenditure on gasoline by the total gallons of gasoline demanded each year.

Number of drivers in each household is categorized into “1”, “2”, “3”, and “4 or more drivers”. Similarly, number of family members *HHSIZE* is categorized into “1”, “2”, “3”, “4”, “5”, and “6 or more persons”. Number of vehicles used by each household is categorized in a similar way. The region dummy variables cover the nine different partitions of the *U.S.* territory. The names for each region are provided by Table 1. For example, if a certain household belongs to the *New England* region, then the first region dummy variable for that household takes on one, while the other region dummies for the same household take on zero.

3.3 Empirical results

The summary statistics for the variables in (3.1) are provided by Table 1. They include the mean and standard deviation of each variable and the *OLS* estimate and its standard error for the corresponding coefficient. 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, which indicates that the variables in (3.1) are statistically very significant with small *p*-values.

The estimation and uniform inference results for the *non-parametric* portion of (3.1) are presented by Figs 1–3. In particular, Figs 2 and 3 show the estimated $\mu(\mathbf{x})$ (i.e. the solid curve) in (1.3) and its corresponding 95 percent uniform confidence surface (*UCS*) (i.e. the surrounding

dotted band). Here x_1 refers to the household income, while x_2 represents the gasoline price. Because the true income and price variables are not observed due to the measurement errors here, the estimation of gasoline demand is based on the calibrated model (1.3) instead of the true model (3.1).

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 (3.1), which is a common assumption in demand analysis, one should be able to insert a straight line into the constructed *USC* in “*all*” of the panels in Figs 2–3. That is, if one cannot insert a straight line into *all* of the constructed *USCs* in Figs 2–3, then the linearity of $\mu(\cdot, \cdot)$ is *rejected* at 5 percent level (and thus the linearity of $g(\cdot, \cdot)$ is rejected as well). Obviously, the *UCSs* presented in Fig.3 cannot contain any straight line in them because of the non-linearity of the estimates and of the corresponding *USCs*. Hence the linearity hypothesis for $g(\cdot, \cdot)$ in (3.1) is clearly *rejected* at 5 percent level for the reasons discussed in the introduction of the paper.

4 Concluding Remarks

The paper illustrates how to conduct the uniform inference of a semi-parametric non-linear model when the covariate terms are contaminated by Berkson measurement errors. To that end, we employ the regression calibration (Carroll et al. (1995)) to transform the 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 uniform confidence surface, 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. The methodology is then applied to perform inference of the linearity of *U.S.* gasoline demand. The study shows that the linearity hypothesis is rejected at 5 percent level, mainly due to the non-linear relationship between the *U.S.* gasoline consumption and the household income.

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. This would require a different framework to handle the issue of measurement error. Another potential extension is to consider the *classical* measurement error structure. Due to the nature of the available data in this study, we assume that the model covariates are contaminated by Berkson-type errors.

Instead, one can assume that the error structure is the classical one and perform uniform inference of the model, accordingly. Further insight can be gained by extending the current work in these and other possible directions.

Appendices

In the appendix, we provide the technical details for the theoretical results given in the main text. Section [A](#) presents the assumptions. Section [B](#) and [C](#) contain the proof for the theorems in the main text. Section [D](#) lists some useful results for proving our theory.

A Assumptions

Let $\Gamma_n = \{y : |y| \leq a_n\}$. $\sigma_n^2(\mathbf{x}) = \mathbb{E}[(Y^* - \mu(\mathbf{x}))^2 \mathbf{1}(Y^* \in \Gamma_n) | \mathbf{X} = \mathbf{x}]$ and $\tilde{\sigma}_n^2 = \mathbb{E}[(Y^* - \mu(\mathbf{x}))^2 \mathbf{1}(Y^* \notin \Gamma_n) | \mathbf{X} = \mathbf{x}]$. Denote the vector of ones and zeros by $\mathbf{1} = (1, \dots, 1)$ and $\mathbf{0} = (0, \dots, 0)$.

The following assumptions are assumed to hold for Theorem [2.2](#).

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

(A2) Given $a_n \asymp (h^{-3d} \log n)^{1/(b_1-2)} \rightarrow \infty$ for some $b_1 > 2$, it holds that

- (i) $0 < \sup_{\mathbf{x} \in \mathcal{X}} \left| \int (y^* - \mu(\mathbf{x}))^{b_1} f_{Y^*|\mathbf{X}}(y^*|\mathbf{x}) dv \right| < \infty$.
- (ii) There exists $C^* > 0$ such that

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

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

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

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

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

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

Assumption [\(A1\)](#) gives constraint on the kernel function, and is satisfied by popular kernels such as Epanechnikov and quartic. 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 similar to those required by [Johnston \(1982\)](#) and [Härdle \(1989\)](#). [\(A6\)](#) implies undersmoothing.

B Bound for R_n

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

$$\|R_n(\mathbf{x})\| = o_{\mathbb{P}} \left\{ \left(nh^d \log(n) \right)^{-1/2} \right\} \quad (\text{B.1})$$

Proof. Recall that $R_n(\mathbf{x}) = \frac{\mathbb{E}H_n(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} + \frac{H_n(\mathbf{x})(f_{\mathbf{X}}(\mathbf{x}) - \hat{f}_{\mathbf{X}}(\mathbf{x}))}{f_{\mathbf{X}}(\mathbf{x})\hat{f}_{\mathbf{X}}(\mathbf{x})}$. Then,

$$\sup_{\mathbf{x}} |R_n(\mathbf{x})| \leq \sup_{\mathbf{x}} \left| \frac{\mathbb{E}H_n(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} \right| + \sup_{\mathbf{x}} \left| \frac{H_n(\mathbf{x})(f_{\mathbf{X}}(\mathbf{x}) - \hat{f}_{\mathbf{X}}(\mathbf{x}))}{f_{\mathbf{X}}(\mathbf{x})\hat{f}_{\mathbf{X}}(\mathbf{x})} \right| \quad (\text{B.2})$$

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

$$\sup_{\mathbf{x}} |f_{\mathbf{X}}(\mathbf{x}) - \hat{f}_{\mathbf{X}}(\mathbf{x})| = O_{\mathbb{P}} \left(h^2 + \sqrt{\frac{\log(n)}{nh^2}} \right) \quad (\text{B.3})$$

Then, by (B.3),

$$\sup_{\mathbf{x}} \left| \frac{\hat{f}_{\mathbf{X}}(\mathbf{x}) - f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})\hat{f}_{\mathbf{X}}(\mathbf{x})} \right| \leq \frac{\sup_{\mathbf{x}} |\hat{f}_{\mathbf{X}}(\mathbf{x}) - f_{\mathbf{X}}(\mathbf{x})|}{\inf_{\mathbf{x}} f_{\mathbf{X}}(\mathbf{x})(f_{\mathbf{X}}(\mathbf{x}) - \epsilon_n)} = O_{\mathbb{P}} \left(h^2 + \sqrt{\frac{\log(n)}{nh^2}} \right) \quad (\text{B.4})$$

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

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

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 inequality is from change of variable, and the final rate is obtained by the summability condition of $K(\cdot)$ from (A1) and the bounded $f_{\mathbf{X}}(\mathbf{x})$ from (A4). Thus,

$$\sup_{\mathbf{x}} |\mathbb{E}H_n(\mathbf{x})| = \mathcal{O}(h^s) \quad (\text{B.5})$$

Furthermore,

$$\sup_{\mathbf{x}} |H_n(\mathbf{x})| \leq \sup_{\mathbf{x}} |H_n(\mathbf{x}) - \mathbb{E}H_n(\mathbf{x})| + \sup_{\mathbf{x}} |\mathbb{E}H_n(\mathbf{x})|, \quad (\text{B.6})$$

where $\sup_{\mathbf{x}} |H_n(\mathbf{x}) - \mathbb{E}H_n(\mathbf{x})| = O_{\mathbb{P}}((\log n/nh^d)^{1/2})$ from an 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 (B.2) and (B.4)–(B.6),

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

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

C Proof of Proposition 2.1

The steps of the proof mainly follow Chao et al. (2016).

$$U_n(\mathbf{x}) = \frac{1}{\sqrt{h^2 \sigma^2 f_{\mathbf{X}}(\mathbf{x})}} \int \int K\left(\frac{\mathbf{x} - \mathbf{u}}{h}\right) (y^* - \mu(\mathbf{x})) dZ_n(\mathbf{u}, y^*) \quad (\text{C.1})$$

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

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

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

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

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

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

$$U_{4,n}(\mathbf{x}) = \frac{1}{\sqrt{h^2 \sigma_n^2(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})}} \int \sqrt{\sigma_n(\mathbf{u})^2 f(\mathbf{u})} K\left(\frac{\mathbf{x} - \mathbf{u}}{h}\right) dW(\mathbf{u}) \quad (\text{C.6})$$

$$U_{5,n}(\mathbf{x}) = \frac{1}{h} \int K\left(\frac{\mathbf{x} - \mathbf{u}}{h}\right) dW(\mathbf{u}) \quad (\text{C.7})$$

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

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

Definition C.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 p th-face of B is $\Pi_{i \neq p}(s_i, t_i]$. Disjoint blocks B and C are p -neighbors if they abut and have the same p th 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 C.2 (Bickel and Wichura (1971) p.1658). *Let $X : \mathbb{R}^d \rightarrow \mathbb{R}$. The increment of X on the block B , denoted $X(B)$, is defined by*

$$X(B) = \sum_{\alpha \in \{0,1\}^d} (-1)^{d-|\alpha|} X\{\mathbf{s} + \alpha \odot (\mathbf{t} - \mathbf{s})\}, \quad (\text{C.8})$$

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

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 C.3. $\|U_n - U_{0,n}\| = \mathcal{O}_p((\log n)^{-1/2})$.

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

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

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

$$|\widehat{U}_{n,0}(\mathbf{x}) - U_{n,0}(\mathbf{x})| = \left| (\sigma(\mathbf{x})/\sigma_n(\mathbf{x}) - 1)U_{n,0}(\mathbf{x}) \right|.$$

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

$$\sup_{\mathbf{x} \in \mathcal{X}} |\sigma(\mathbf{x})/\sigma_n(\mathbf{x}) - 1| = \mathcal{O}\{(\log n)^{-1}\}. \quad (\text{C.9})$$

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

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

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

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

as $n \rightarrow \infty$, hence $E_2 = \mathcal{O}_p((\log n)^{-1/2})$.

We now show that $E_1 = \mathcal{O}_p((\log n)^{-1})$. To do this, it is enough to show the weak convergence of $\log n E_1$, 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

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

where

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

and

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

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

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

Thus,

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

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

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

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

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

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

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

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

Note that

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

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

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

By symmetry,

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

Hence, by (A.1) in (A2),

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

Analogously we obtain the estimate for

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

which finally yields by Hölder's inequality,

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

Applying Lemma D.1 with $\gamma_1 = \gamma_2 = \lambda_1 = \lambda_2 = 1$ yields the tightness. \square

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

Proof of Lemma C.4. In this proof, we adopt the notation that if $\alpha \in \{0,1\}^{d+1}$, then we write $\alpha = (\alpha_1, \alpha_2)$ where $\alpha_1 \in \{0,1\}$ and $\alpha_2 \in \{0,1\}^d$. In the computation below, we focus on $B_{\mathbf{x}} = \prod_{j=1}^d [x_j - Ah, x_j + Ah]$ instead of \mathbb{R}^d since K has compact support. Recall definition C.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,

$$\begin{aligned}
U_{0,n}(\mathbf{x}) &= \frac{1}{\sqrt{h^d f_{\mathbf{X}}(\mathbf{x}) \sigma_n^2(\mathbf{x})}} \left[\int_{B_{\mathbf{x}}} \int_{\Gamma_n} Z_n(y, \mathbf{u}) d\left((y^* - \mu(\mathbf{x})) K\left(\frac{\mathbf{x} - \mathbf{u}}{h}\right)\right) \right. \\
&\quad + \left. \left\{ Z_n(\cdot_1, \cdot_2) (\cdot_1 - \mu(\mathbf{x})) K\left(\frac{\mathbf{x} - \cdot_2}{h}\right) \right\} (\Gamma_n \times B_{\mathbf{x}}) \right. \\
&\quad + \left. \left\{ \sum_{\alpha \in \{0,1\}^{d+1} - \{\mathbf{0}, \mathbf{1}\}} \int \int_{(\Gamma_n \times B_{\mathbf{x}})_{\alpha}} Z_n(\cdot_1, \cdot_2) d^{\alpha_1}(\cdot_1 - \mu(\mathbf{x})) \partial^{\alpha_2} K\left(\frac{\mathbf{x} - \cdot_2}{h}\right) \right\} (\Gamma_n \times B_{\mathbf{x}})_{\mathbf{1} - \alpha} \right]
\end{aligned} \tag{C.11}$$

where $\mathbf{1} = (1, \dots, 1) \in \{0, 1\}^{d+1}$ and $\mathbf{0} = (0, \dots, 0) \in \{0, 1\}^{d+1}$. $(\Gamma_n \times B_{\mathbf{x}})$ 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 (C.11) can be evaluated with the formula (C.8). $(\Gamma_n \times B_{\mathbf{x}})_{\mathbf{1} - \alpha}$ can be viewed as the projection of $\Gamma_n \times B_{\mathbf{x}}$ on to the space spanned by those axes whose numbers correspond to positions of ones of the multi-index $\mathbf{1} - \alpha$. This leaves us with an $|\alpha|$ -fold integral.

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

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

$$\begin{aligned}
&h^{d/2} n^{1/6} (\log n)^{-\epsilon - (2d+4)/3} a_n^{-1} |U_{0,n}(\mathbf{x}) - U_{1,n}(\mathbf{x})| \\
&\leq \mathcal{O}(1) \left| \frac{a_n^{-1}}{\sqrt{f_{\mathbf{X}}(\mathbf{x}) \sigma_n^2(\mathbf{x})}} \right| \left\{ 2a_n \left| \int_{B_{\mathbf{x}}} dK((\mathbf{x} - \mathbf{u})/h) \right| + 2a_n \left| K\left(\frac{\mathbf{x} - \cdot_2}{h}\right) \right| (B_{\mathbf{x}}) \right. \\
&\quad + 2a_n \left| \sum_{\alpha_1=1, \alpha_2 \in \{0,1\}^d - \{\mathbf{1}_2\}} \int_{(B_{\mathbf{x}})_{\alpha_2}} \partial^{\alpha_2} K\left(\frac{\mathbf{x} - \cdot_2}{h}\right) \right| (B_{\mathbf{x}})_{\mathbf{1}_2 - \alpha_2} \\
&\quad \left. + 2a_n \left| \sum_{\alpha_1=0, \alpha_2 \in \{0,1\}^d - \{\mathbf{0}_2\}} \int_{(B_{\mathbf{x}})_{\alpha_2}} \partial^{\alpha_2} K\left(\frac{\mathbf{x} - \cdot_2}{h}\right) \right| (B_{\mathbf{x}})_{\mathbf{1}_2 - \alpha_2} \right\}, \text{ a.s.}
\end{aligned} \tag{C.12}$$

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

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

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

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

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

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

Proof of Lemma C.6. Define

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

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

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

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

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

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

$$\begin{aligned}
\mathbb{E} \left[\left(\frac{V(\mathbf{t}) - V(\mathbf{s})}{h} \right)^2 \right] &\leq \frac{2}{h^{d+2}} \int \int_{\Gamma_n} (\mu(\mathbf{s}) - \mu(\mathbf{t}))^2 K^2\left(\frac{\mathbf{s} - \mathbf{u}}{h}\right) f(y^*, \mathbf{u}) dy^* d\mathbf{u} + \\
&\quad \frac{2}{h^{d+2}} \int \int_{\Gamma_n} (\mu(\mathbf{u}) - \mu(\mathbf{t}))^2 \left[K\left(\frac{\mathbf{t} - \mathbf{u}}{h}\right) - K\left(\frac{\mathbf{s} - \mathbf{u}}{h}\right) \right]^2 f(y^*, \mathbf{u}) dy^* d\mathbf{u} \stackrel{\text{def}}{=} I_1 + I_2.
\end{aligned}$$

We estimate I_1 first,

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

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

For I_2 , by (A3), we obtain

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

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 D.2 we obtain the estimate $\gamma(\epsilon) \leq C \max\{\sqrt{\epsilon/h}, \epsilon/h\} \leq C\sqrt{\epsilon}/h$ for $\epsilon < 1$ and $C > 0$, and

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

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

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

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

Lemma C.7. $U_{3,n}(\mathbf{x}) \stackrel{d}{=} U_{4,n}(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$.

Proof of Lemma C.7. The proof resembles the proof for Lemma A.5 in the supplement material of Chao et al. (2016) and is omitted for brevity. \square

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

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

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

To apply Lemma D.2, we need to estimate

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

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

$$\begin{aligned} \mathbb{E} \left[\left(\frac{\tilde{V}_n(\mathbf{t})}{h} \right)^2 \right] &\leq 2Ch^{-d-2} \left\{ \int \left[\sqrt{f_{\mathbf{X}}(\mathbf{u})} - \sqrt{f_{\mathbf{X}}(\mathbf{t})} \right]^2 K^2 \left(\frac{\mathbf{t} - \mathbf{u}}{h} \right) d\mathbf{u} \right. \\ &\quad \left. + \int \left[\sqrt{\sigma_n^2(\mathbf{u})} - \sqrt{\sigma_n^2(\mathbf{t})} \right]^2 K^2 \left(\frac{\mathbf{t} - \mathbf{u}}{h} \right) d\mathbf{u} \right\}, \end{aligned}$$

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

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

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

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

On the other hand,

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

From

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

we obtain

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

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

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

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

D Auxiliary Results

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

$$\mathbb{E}[|X_n(B)|^{\gamma_1} |X_n(C)|^{\gamma_2}] \leq \lambda(B)^{\lambda_1} \lambda(C)^{\lambda_2}, \quad (\text{D.1})$$

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

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

Lemma D.2 (Meerschaert, M. M., Wang, W. and Xiao, Y. (2013)). *Suppose that $Y = \{Y(\mathbf{t}), \mathbf{t} \in \mathbb{R}^d\}$ is a centered Gaussian random field with values in \mathbb{R} , and denote*

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

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

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

and

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

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

$$P \left\{ \sup_{\mathbf{t} \in \mathcal{S}} |Y(\mathbf{t})| > a \right\} \leq 2^{2d+2} \left(\frac{r}{Q^{-1}(1/a)} + 1 \right)^d \frac{\sigma + a^{-1}}{a} \exp \left\{ -\frac{a^2}{2(\sigma + a^{-1})^2} \right\}, \quad (\text{D.2})$$

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

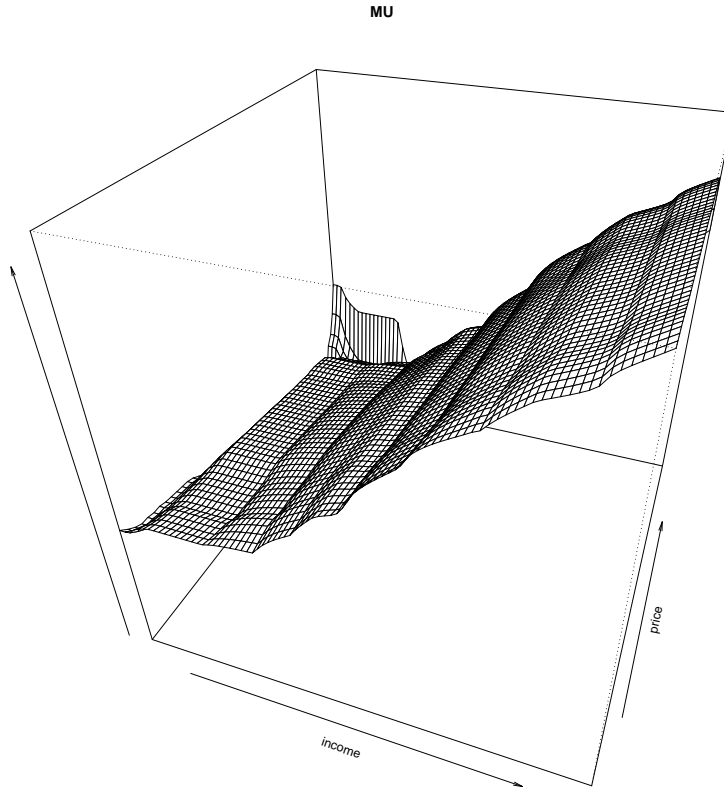



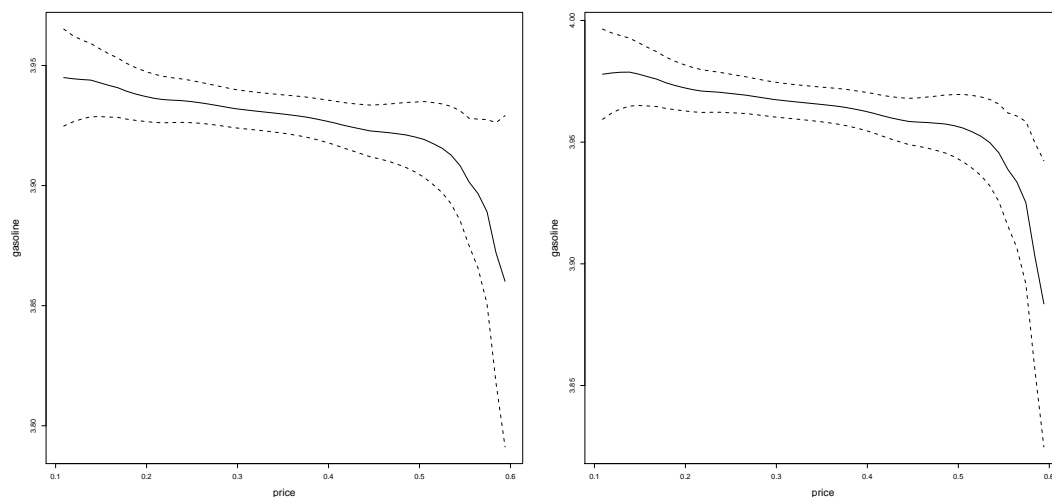
Figure 1: U.S. gasoline demand in income and price.  EIV_UCS

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. (2016+). Confidence corridors for multivariate generalized quantile regression. *Journal of Business & Economic Statistics*.

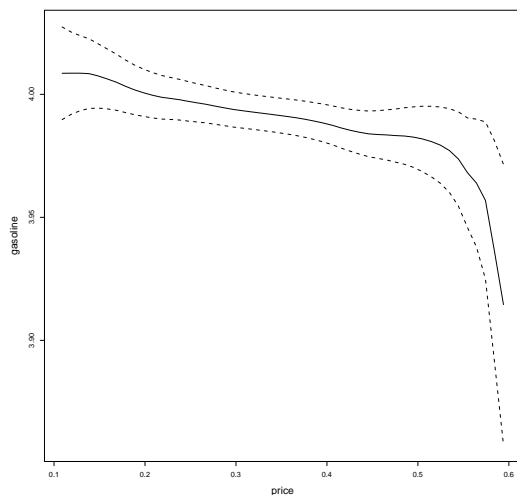
Table 1: Variables in the *RTECS* Surveys from Year 2001

Variable Description	Mean	Std. Dev.	OLS estimate (std. error)
Log of Total Miles Traveled (<i>TOTMILES</i>)	4.214	0.423	-
Log of Income (<i>INCOME</i>)	4.623	0.333	0.180 (0.007)
Log of Number of Drivers (<i>DRVR</i>)	0.236	0.170	0.188 (0.024)
Log of Number of Vehicles (<i>VEHS</i>)	0.248	0.212	0.910 (0.013)
Log of Household Size (<i>HHSIZE</i>)	0.344	0.227	0.037 (0.028)
Child Dummy (<i>CHILD</i>)	0.348	0.476	1.911 (0.037)
Gender Dummy (<i>SEX</i>)	0.584	0.493	-0.035 (0.004)
Urban Residence Dummy (<i>RURAL</i>)	0.254	0.435	0.097 (0.005)
<i>Region Dummy Variables (Region)</i>			
New England	0.091	0.288	0.026 (0.011)
Middle Atlantic	0.187	0.390	0.027 (0.010)
East North Central	0.069	0.254	NA (-)
West North Central	0.026	0.160	1.983 (0.035)
South Atlantic	0.090	0.286	1.884 (0.034)
East South Central	0.102	0.302	1.699 (0.032)
West South Central	0.055	0.227	1.926 (0.035)
Mountain	0.199	0.399	1.883 (0.035)
Pacific	0.181	0.385	1.795 (0.034)



(a) 25th-percentile income

(b) 50th-percentile income



(c) 75th-percentile income

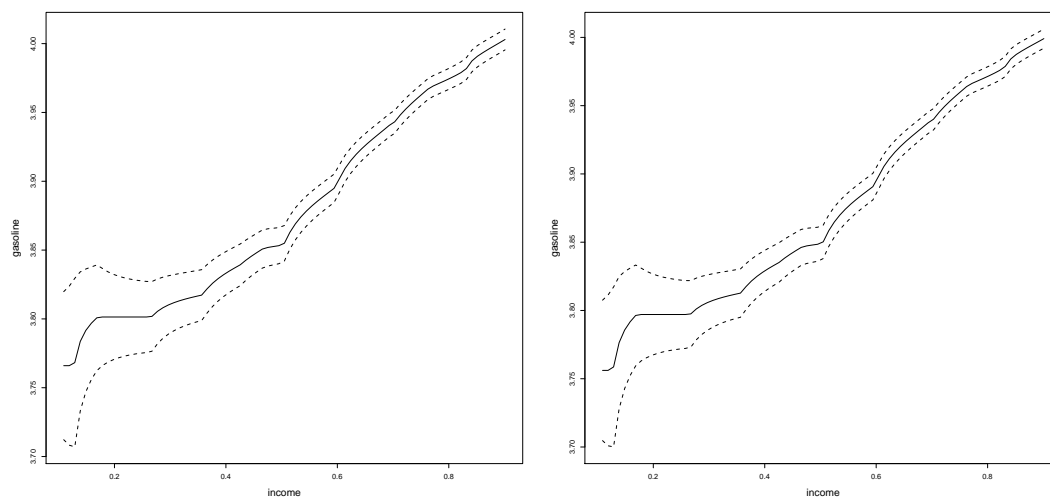
Figure 2: U.S. gasoline demand; The dashed band is 95% uniform confidence surface.  EIV_UCS

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

Dedecker, J., Merlevède, 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.

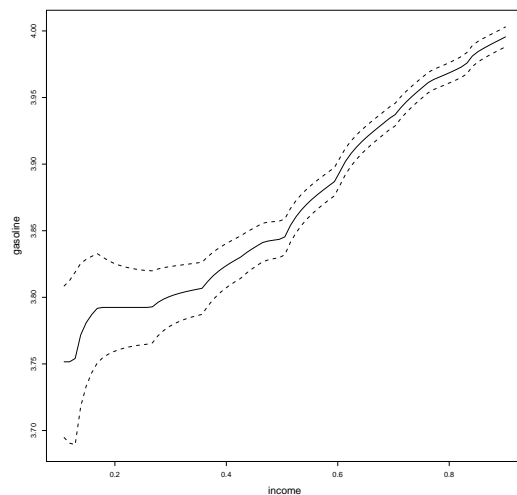
Ferguson, T. S. (1996). *A Course in large sample theory*. Chapman and Hall, CRC.

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



(a) 25th-percentile price

(b) 50th-percentile price



(c) 75th-percentile price

Figure 3: U.S. gasoline demand; The dashed band is 95% uniform confidence surface.  EIV_UCS

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.

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 Reviews*, 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 regression 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.
- Meerschaert, M. M., Wang, W. and Xiao, Y. (2013). Fernique-type inequalities and moduli of continuity for anisotropic Gaussian random fields. *Transactions of the American Mathematical Society*, 365(2):1081–1107.
- Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and Its Applications*, 9:141–142.
- 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.
- Schennach, S. M. (2013). Regressions with berkson errors in covariates – a nonparametric approach. *Annals of Statistics*, 41:1642–1668.
- Schmalensee, R. and Stoker, T. M. (1999). Household gasoline demand in the united states. *Econometrica*, 67:645–662.
- Watson, G. S. (1964). Smooth regression analysis. *Sankya, Series A.*, 26:359–372.
- 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. *Annals of Statistics*, 36:1854–1878.

SFB 649 Discussion Paper Series 2016

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 001 "Downside risk and stock returns: An empirical analysis of the long-run and short-run dynamics from the G-7 Countries" by Cathy Yi-Hsuan Chen, Thomas C. Chiang and Wolfgang Karl Härdle, January 2016.
- 002 "Uncertainty and Employment Dynamics in the Euro Area and the US" by Aleksei Netsunajev and Katharina Glass, January 2016.
- 003 "College Admissions with Entrance Exams: Centralized versus Decentralized" by Isa E. Hafalir, Rustamdjan Hakimov, Dorothea Kübler and Morimitsu Kurino, January 2016.
- 004 "Leveraged ETF options implied volatility paradox: a statistical study" by Wolfgang Karl Härdle, Sergey Nasekin and Zhiwu Hong, February 2016.
- 005 "The German Labor Market Miracle, 2003 -2015: An Assessment" by Michael C. Burda, February 2016.
- 006 "What Derives the Bond Portfolio Value-at-Risk: Information Roles of Macroeconomic and Financial Stress Factors" by Anthony H. Tu and Cathy Yi-Hsuan Chen, February 2016.
- 007 "Budget-neutral fiscal rules targeting inflation differentials" by Maren Brede, February 2016.
- 008 "Measuring the benefit from reducing income inequality in terms of GDP" by Simon Voigts, February 2016.
- 009 "Solving DSGE Portfolio Choice Models with Asymmetric Countries" by Grzegorz R. Dlugoszek, February 2016.
- 010 "No Role for the Hartz Reforms? Demand and Supply Factors in the German Labor Market, 1993-2014" by Michael C. Burda and Stefanie Seele, February 2016.
- 011 "Cognitive Load Increases Risk Aversion" by Holger Gerhardt, Guido P. Biele, Hauke R. Heekeren, and Harald Uhlig, March 2016.
- 012 "Neighborhood Effects in Wind Farm Performance: An Econometric Approach" by Matthias Ritter, Simone Pieralli and Martin Odening, March 2016.
- 013 "The importance of time-varying parameters in new Keynesian models with zero lower bound" by Julien Albertini and Hong Lan, March 2016.
- 014 "Aggregate Employment, Job Polarization and Inequalities: A Transatlantic Perspective" by Julien Albertini and Jean Olivier Hairault, March 2016.
- 015 "The Anchoring of Inflation Expectations in the Short and in the Long Run" by Dieter Nautz, Aleksei Netsunajev and Till Strohsal, March 2016.
- 016 "Irrational Exuberance and Herding in Financial Markets" by Christopher Boortz, March 2016.
- 017 "Calculating Joint Confidence Bands for Impulse Response Functions using Highest Density Regions" by Helmut Lütkepohl, Anna Staszewska-Bystrova and Peter Winker, March 2016.
- 018 "Factorisable Sparse Tail Event Curves with Expectiles" by Wolfgang K. Härdle, Chen Huang and Shih-Kang Chao, March 2016.
- 019 "International dynamics of inflation expectations" by Aleksei Netsunajev and Lars Winkelmann, May 2016.
- 020 "Academic Ranking Scales in Economics: Prediction and Imputation" by Alona Zharova, Andrija Mihoci and Wolfgang Karl Härdle, May 2016.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".



SFB 649 Discussion Paper Series 2016

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 021 "CRIX or evaluating blockchain based currencies" by Simon Trimborn and Wolfgang Karl Härdle, May 2016.
- 022 "Towards a national indicator for urban green space provision and environmental inequalities in Germany: Method and findings" by Henry Wüstemann, Dennis Kalisch, June 2016.
- 023 "A Mortality Model for Multi-populations: A Semi-Parametric Approach" by Lei Fang, Wolfgang K. Härdle and Juhyun Park, June 2016.
- 024 "Simultaneous Inference for the Partially Linear Model with a Multivariate Unknown Function when the Covariates are Measured with Errors" by Kun Ho Kim, Shih-Kang Chao and Wolfgang K. Härdle, August 2016.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

