Essays in Nonparametric Estimation and Inference

Luke Taylor

London School of Economics and Political Science

A thesis submitted to the Department of Economics of the London School of Economics and Political Science for the degree of Doctor of Philosophy

April 2017

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I declare my thesis is 27455 words.

Statement of joint work. I can confirm that Chapters 1 and 2 are joint work with Dr. Taisuke Otsu. I contributed 50% of the work in each case.

Luke Taylor

To my parents and to Leanne.

Abstract

This thesis consists of three chapters which represent my journey as a researcher during this PhD. The uniting theme is nonparametric estimation and inference in the presence of data problems. The first chapter begins with nonparametric estimation in the presence of a censored dependent variable and endogenous regressors. For Chapters 2 and 3 my attention moves to problems of inference in the presence of mismeasured data.

In Chapter 1 we develop a nonparametric estimator for the local average response of a censored dependent variable to endogenous regressors in a nonseparable model where the unobservable error term is not restricted to be scalar and where the nonseparable function need not be monotone in the unobservables. We formalise the identification argument put forward in Altonji, Ichimura and Otsu (2012), construct a nonparametric estimator, characterise its asymptotic property, and conduct a Monte Carlo investigation to study its small sample properties. We show that the estimator is consistent and asymptotically normally distributed.

Chapter 2 considers specification testing for regression models with errors-in-variables. In contrast to the method proposed by Hall and Ma (2007), our test allows general nonlinear regression models. Since our test employs the smoothing approach, it complements the nonsmoothing one by Hall and Ma in terms of local power properties. We establish the asymptotic properties of our test statistic for the ordinary and supersmooth measurement error densities and develop a bootstrap method to approximate the critical value. We apply the test to the specification of Engel curves in the US. Finally, some simulation results endorse our theoretical findings: our test has advantages in detecting high frequency alternatives and dominates the existing tests under certain specifications.

Chapter 3 develops a nonparametric significance test for regression models with measurement error in the regressors. To the best of our knowledge, this is the first test of its kind. We use a 'semi-smoothing' approach with nonparametric deconvolution estimators and show that our test is able to overcome the slow rates of convergence associated with such estimators. In particular,

our test is able to detect local alternatives at the $\sqrt{n}$ rate. We derive the asymptotic distribution under i.i.d. and weakly dependent data, and provide bootstrap procedures for both data types. We also highlight the finite sample performance of the test through a Monte Carlo study. Finally, we discuss two empirical applications. The first considers the effect of cognitive ability on a range of socio-economic variables. The second uses time series data - and a novel approach to estimate the measurement error without repeated measurements - to investigate whether future inflation expectations are able to stimulate current consumption.

Acknowledgements

I will be forever grateful to my supervisors Peter Robinson and Taisuke Otsu. They have been incredibly patient with me and provided me with support throughout my time at LSE. They have given me far more of their time than I ever expected and I'm incredibly lucky to have had them help me on this journey. I know I would not have made it through this PhD without them.

The faculty at LSE have been incredibly helpful and generous with their time and advice. It's such a huge advantage to have so many talented econometricians attend work in progress seminars. In particular, I would like to personally thank Javier Hidalgo, Abhimanyu Gupta, Tatiana Komarova, Yukitoshi Matsushita, Jörn-Steffen Pischke, Ricardo Reis, Senay Sokullu and Marcia Schafgans. I am also grateful to my peers who have provided me with many useful discussions and support: Karun Adusumilli, Hao Dong, Miguel Bandeira Da Silva, William Matcham and Dandan Wang. I would also like to express my thanks for the generous funding from the ESRC.

My family and friends have provided me with constant support throughout this long process. Without my parents' help I would never have been able to start this PhD, let alone finish it. Leanne has been there through all the ups and downs and I'll be eternally grateful for her love and support.

Contents

List of Tables 10

List of Figures 11

Chapter 1. Estimation of Nonseparable Models with Censored Dependent Variables and Endogenous Regressors 12

1.1. Introduction 12

1.2. Main Results 14

1.3. Simulation 25

1.4. Conclusion 27

1.5. Appendix - Mathematical Proofs 28

1.6. Appendix - Simulation results 38

Chapter 2. Specification Testing for Errors-in-Variables Models 41

2.1. Introduction 41

2.2. Setup and Test Statistic 44

2.3. Asymptotic Properties 47

2.4. Simulation 54

2.5. Empirical Example 61

2.6. Appendix - Mathematical Proofs 63

Chapter 3. Nonparametric Significance Testing in Measurement Error Models 72

3.1. Introduction 72

3.2. Methodology 75

3.3. Asymptotic Properties 80

3.4. Simulation 90

3.5. Applications 94

3.6. Conclusion 103

3.7. Appendix - Mathematical Proofs 105

3.8. Appendix - Empirical Applications 128

Bibliography 130

List of Tables

1 Simulations - Chapter 2 55

2 Simulations - Chapter 2 56

3 Simulations - Chapter 2 57

4 Simulations - Chapter 2 58

5 Simulations - Chapter 2 59

6 Engel Curve (P-Values) 62

1 Simulations - Chapter 3 92

2 Simulations - Chapter 3 92

3 Simulations - Chapter 3 92

4 Simulations - Chapter 3 93

5 Cognitive Ability (P-Values) 97

6 Inflation Expectations (P-Values) 102

List of Figures

1 Cognitive Ability Plot 128

2 Inflation Expectation Plot 129

CHAPTER 1

Estimation of Nonseparable Models with Censored Dependent Variables and Endogenous Regressors

1.1. Introduction

One of the greatest contributions of econometrics is the development of estimation and inference methods in the presence of endogenous explanatory variables. The classic literature mostly focuses on linear simultaneous equation systems and has been extended to various contexts. In the case of censored dependent variables, Amemiya (1979) and Smith and Blundell (1986) study estimation and testing of simultaneous equation Tobit models, where the linear regression function and joint normality of the error distribution are maintained. In this chapter, we study how to evaluate nonparametrically the marginal effects of the endogenous explanatory variables to the censored dependent variable when both the regression function and distributional forms are unknown and the error term may not be additively separable.

In particular, we seek to extend the work by Altonji, Ichimura and Otsu (2012), AIO henceforth, by introducing endogeneity into a nonseparable model with a censored dependent variable. AIO (Sections 5.1 and 5.2) describe how to accommodate endogenous regressors into their identification analysis. The aims of this chapter are to formalise their identification argument, develop a nonparametric estimator for the local average response, and derive its asymptotic properties. We also carry out a Monte Carlo investigation to study the small sample properties.

Our estimator can be seen as an extension of the classic Tobit maximum likelihood estimator in several directions. We allow the unobservable error term to enter into the model in a nonseparable manner; this is a far more realistic assumption and the popularity of such models in the recent literature highlights this fact (see, for example Matzkin, 2007, and references therein). We allow the dependent variable to depend on the regressors and error term in a nonlinear way, in the same manner as AIO. We also do not constrain the dependent variable to be monotonic in the error term and allow it to be censored from both above and below, moreover we allow the censoring points to depend on the regressors. Finally, we allow the regressors to be correlated with the error term.

Since endogeneity is an issue that plagues many economic models, the possible applications of the estimator we consider are extensive. Commonly cited examples of nonseparable models with censoring are consumer demand functions at corner solutions. An interesting example is Altonji, Hayashi and Kotlikoff (1997) where a monetary transfer from parents to children only occurs if the marginal utility gained from the additional consumption of their child is greater than the marginal utility lost from the fall in their own consumption. Auctions provide another possible application for this estimator. Different forms of the Tobit estimator are commonly used to analyse auction data because of the various forms of censoring found in these settings, for example Jofre-Bonet and Pesendorfer (2003). In general, the estimator developed in this chapter is applicable in all settings where the Tobit estimator is used. For example, Shishko and Rostker (1976) estimate the supply of labour for second jobs using the Tobit estimator. In this setting it is highly likely that unobservable characteristics such as ability and tastes for spending enter the supply function in a nonseparable way. See McDonald and Moffitt (1980) for further examples. More recently there has been much interest in nonseparable models, however many cases have failed to take into account censoring. For example, several examples of hedonic models considered in Heckman, Matzkin and Nesheim (2010) are likely to suffer from censoring.

The identification strategy used in this chapter follows AIO very closely. However, the strategy must be adapted to take into account endogeneity. In this chapter we use a control function approach. This involves conditioning on the residuals from a first stage regression of the endogenous regressors on instruments to fix the distribution of the unobservable error term. This conditioning is then undone by averaging over the distribution of the residuals (see Blundell and Powell, 2003).

As a parameter of interest, we focus on the local average response conditional on the dependent variable being uncensored. This is in contrast to the local average response across the whole sample, which would be more suited to cases where censoring is due to failures in measurement. AIO focus on the exogenous case and only briefly introduce endogeneity as an extension to the model. Whilst Altonji and Matzkin (2005) discuss identification and estimation of the local average response in a nonseparable model without censoring.

There has been considerable interest in nonseparable models with endogenous regressors over the last 15 years (see, for example, Imbens and Newey, 2009, Chernozhukov, Fernandez-Val and Kowalksi, 2015, and a review by Matzkin, 2007). Schennach, White and Chalak (2012) consider triangular structural systems with nonseparable functions that are not monotonic in the scalar unobservable. They find that local indirect least squares is unable to estimate the local average response, but can be used to test if there is no effect from the regressor in this general case. Hoderlein and Mammen (2007) also drop the assumption of monotonicity and showed that by using regression quantiles identification can be achieved. However, this result was obtained in the absence of endogenous regressors. Censoring in nonseparable models has received little attention; Lewbel and Linton (2002) consider censoring in a separable model and Chen, Dhal and Khan (2005) study a partially separable model.

This chapter is organised as follows. Section 1.2 presents the main results: nonparametric identification of the local average response (Section 1.2.1) and nonparametric estimation of the identified object (Section 1.2.2). In Section 1.3, we assess the small sample properties of the proposed estimator via Monte Carlo simulation. Section 1.4 concludes.

1.2. Main Results

In this section, we consider identification and estimation of the model based on cross-section data. Our notation closely follows that of AIO. The model is set up such that the dependent variable Y is observed only when a latent variable falls within a certain interval,

$$Y = \begin{cases} M(X,U) & \text{if } L(X) < M(X,U) < H(X), \\ C_L & \text{if } M(X,U) \leq L(X), \\ C_H & \text{if } H(X) \leq M(X,U), \end{cases}$$

where X is a d-dimensional vector of observables and $M : \mathbb{R}^d \times \mathbb{U} \mapsto \mathbb{R}$ is a differentiable function with respect to the first argument, indexed by an unobservable random object U. The support $\mathbb{U}$ of U is possibly infinite dimensional. Also $L(X)$ and $H(X)$ are scalar-valued functions of

X,[1] and C_L and C_H are indicators to signify censoring from below and above, respectively. For example, they may be coded as C_L ="censored from below" and C_H ="censored from above". This model represents a generalisation of the Tobit model, where $M(X,U) = X'\beta + U$, $L(X) = 0$, $H(X) = \infty$, and U is normal and independent from X.

Let $I_M(X,U) = I\{L(X) < M(X,U) < H(X)\}$, where $I\{\cdot\}$ is the indicator function. As a parameter of interest, we focus on the local average response given that $X = x$ and Y is not censored, that is

$$\beta(x) = E[\nabla M(X,U) | X = x, I_M(X,U) = 1], \tag{1}$$

where $\nabla M(X,U)$ is the partial derivative of M with respect to X. AIO investigated identification and estimation of $\beta(x)$ when X and U are independent and discussed briefly identification of $\beta(x)$ when X is endogenous and can be correlated with U. Here we formalise their identification argument and develop a nonparametric estimator of $\beta(x)$.

Without censoring, the local average response $\beta_{AM}(x) = E[\nabla M(X,U)|X = x]$ with endogenous X was proposed and studied in Altonji and Matzkin (2005). They go on to discuss several motivations of the local average response. Our object of interest, $\beta(x)$, in (1) shares similar motivations. As a particular example, Aaronson (1998) investigates the effects of average neighbourhood income X on college attendance Y holding the distribution of $U|X = x$ fixed. Thus, the local average response is the parameter of interest in Aaronson's (1998) empirical analysis.

We note that for the linear case $M(X,U) = X'\beta + U$, the object $\beta(x)$ coincides with the slope parameter β in the Tobit model with endogenous X. Also, as briefly mentioned in Section 1.1, Altonji, Hayashi and Kotlikoff (1997) consider altruism based models of money transfers from parents to children, and study the effects of endowments X to money transfers Y. The money transfers are obviously censored from below by 0 and it is reasonable to suspect correlation between the endowments X and unobserved preferences U of the parents and children. Thus, $\beta(x)$ is a parameter of interest in the empirical study of Altonji, Hayashi and Kotlikoff (1997). See, for example, Raut and Tran (2005) and Kaziango (2006) for further examples.

[1] It is possible to allow both $L(\cdot)$ and $H(\cdot)$ to depend on additional observed variables, for example $L(\tilde{X})$ and $H(\tilde{X})$ where $\tilde{X}$ contains X as a subvector, without affecting the proceeding results. We restrict attention only to X for ease of exposition.

1.2.1. Identification. We employ a control function approach to identify the average derivative, $\beta(x)$, in the presence of endogenous X. This is a standard approach in the literature (see, for example, Blundell and Powell, 2003). It is assumed that the researcher observes a vector of random variables W satisfying

$$X = \varphi(W) + V, \quad E[V|W] = 0 \text{ a.s.},$$

$$U \perp W | V,$$

where V is the error term. Under this setup, we wish to identify the local average response $\beta(x)$ in (1) based on the observables (Y, X, W). Note that the function $\varphi(\cdot)$ is identified by the conditional mean $\varphi(w) = E[X|W = w]$. Thus in the identification analysis below, we treat V as observable. Although conditional independence $U \perp W|V$ is a strong assumption, it is hard to avoid unless further restrictions are placed on the functional form of $M(x, u)$, such as monotonicity in a scalar u.

Using the auxiliary variable V, the parameter of interest can be written as

$$\begin{aligned} \beta(x) &= \int_u \nabla M(x,u) dP(u|X = x, I_M(X,U) = 1) \\ &= \int_v \beta(x,v) dP(v|X = x, I_M(X,U) = 1), \end{aligned} \tag{2}$$

where dP is the Lebesgue density of U and $\beta(x, v) = \int_u \nabla M(x, u)dP(u|X = x, I_M(X, U) = 1, V = v)$. Note that we observe X and $I_M(X, U) = I\{Y \neq C_L, C_H\}$, and that V is treated as observable. Thus the conditional distribution of V given $X = x$ and $I_M(X, U) = 1$ is identified. Based on (2), it is sufficient to identify $\beta(x, v)$. Let $G_M(x, v) = \Pr\{I_M(X, U) = 1|X = x, V = v\}$. By using the assumptions on V, the object $\beta(x, v)$ can be written as

$$\begin{aligned} \beta(x,v) &= \int_{u \in \{u: I_M(x,u)=1\}} \nabla M(x,u) dP(u|X = x, V = v)/G_M(x,v) \\ &= \int_{u \in \{u: I_M(x,u)=1\}} \nabla M(x,u) dP(u|\varphi(W) = \varphi(w), V = v)/G_M(x,v) \\ &= \int_{u \in \{u: I_M(x,u)=1\}} \nabla M(x,u) dP(u|V = v)/G_M(x,v). \end{aligned}$$

Similarly, observe that

$$\begin{aligned}\Psi(x,v) &= E[M(X,U)|X=x, I_M(X,U)=1, V=v] \\ &= \int_{u\in\{u:I_M(x,u)=1\}} M(x,u)dP(u|V=v)/G_M(x,v).\end{aligned}$$

Note that $\Psi(x,v)$ is identified as the conditional mean of Y given $X=x$, $V=v$, and $I_M(X,U)=1$ (uncensored). The basic idea for identification is to compare the derivative of the conditional mean $\triangledown\Psi(x,v)$ with the conditional mean of the derivative of $\beta(x,v)$.

For expositional purposes only, we tentatively assume that $M(x,u)$ is continuous and monotonic in scalar u for each x; we show how this assumption can be dropped later. Using the Leibniz rule to differentiate $\Psi(x,v)$ with respect to x while holding v constant gives

$$\begin{aligned}\triangledown[\Psi(x,v)G_M(x,v)] &= \int_{u_L(x)}^{u_H(x)} \triangledown M(x,u)dP(u|V=v) \\ &\quad +M(x,u_H(x))dP(u_H(x)|V=v)\triangledown u_H(x) \\ &\quad -M(x,u_L(x))dP(u_L(x)|V=v)\triangledown u_L(x), \qquad (3)\end{aligned}$$

where $u_H(x)$ and $u_L(x)$ solve $M(x,u)=H(x)$ and $M(x,u)=L(x)$, respectively, so that $M(x,u_H(x))=H(x)$ and $M(x,u_L(x))=L(x)$. Denoting $G_H(x,v)=\Pr\{Y=C_H|X=x,V=v\}$ and $G_L(x,v)=\Pr\{Y=C_L|X=x,V=v\}$, we obtain $\triangledown G_H(x,v)=-dP(u_H(x)|V=v)\triangledown u_H(x)$ and $\triangledown G_L(x,v)=dP(u_L(x)|V=v)\triangledown u_L(x)$. Combining these results, $\beta(x,v)$ can be written as

$$\beta(x,v)=\triangledown\Psi(x,v)+\{\Psi(x,v)\triangledown G_M(x,v)+H(x)\triangledown G_H(x,v)+L(x)\triangledown G_L(x,v)\}/G_M(x,v). \qquad (4)$$

Since each term on the right hand side of this equation is identified, we conclude that the parameter of interest $\beta(x)$ is identified.

It is instructive to give an intuitive outline of why the identification argument of AIO fails in the presence of endogeneity. Notice, under exogeneity of X,

$$\Psi^*(x)G_M^*(x)=\int_{u_L(x)}^{u_H(x)} M(x,u)dP(u|X=x)=\int_{u_L(x)}^{u_H(x)} M(x,u)dP(u), \qquad (5)$$

where $\Psi^*(x) = E[M(X,U)|X = x, I_M(X,U) = 1]$ and $G_M^*(x) = \Pr\{I_M(X,U) = 1|X = x\}$. Identification of $\beta(x)$ in AIO is achieved by differentiating (5) with respect to x and solving for $\beta(x)$. However, when X is endogenous, this argument does not apply. In particular, letting $p(u|x)$ denote the conditional density of $U|X = x$, the Leibniz rule yields

$$\begin{aligned} \triangledown[\Psi^*(x)G_M^*(x)] &= \triangledown\left[\int_{u_L(x)}^{u_H(x)} M(x,u)p(u|x)du\right] \\ &= \beta(x) + \int_{u_L(x)}^{u_H(x)} M(x,u)\triangledown p(u|x)du \\ &\quad +M(x,u_H(x))p(u_H(x)|x)\triangledown u_H(x) \\ &\quad -M(x,u_L(x))p(u_L(x)|x)\triangledown u_L(x). \end{aligned}$$

Note that the second term on the right hand side is not estimable. Therefore, the identification strategy of AIO based on the above equation does not apply to the case of endogenous X.

We now show that the above argument for identification holds under more general conditions. The following assumptions are imposed.

Assumption 1.

(i): *$X = \varphi(W) + V$ with $E[V|W] = 0$ a.s. and $U \perp W|V$.*

(ii): *$L(\cdot)$ and $H(\cdot)$ are continuous at x and satisfy $L(x') < H(x')$ for all x' in a neighbourhood of x, and $\Pr\{M(X,U) = L(X)|X = x\} = \Pr\{M(X,U) = H(X)|X = x\} = 0$.*

(iii): *$G_L(\cdot,V)$, $G_M(\cdot,V)$, and $G_H(\cdot,V)$ are differentiable a.s. at x and $G_M(x,V) > 0$ a.s.*

(iv): *$M(\cdot,U)$ is differentiable a.s. at each x' in a neighbourhood of x, and there exists an integrable function $B : \mathbb{U} \to \mathbb{R}$ such that $|\triangledown M(x',U)| \leq B(U)$ a.s. for all x' in a neighbourhood of x.*

Assumption 1 (i) is a key condition required to use a control function approach. This assumption is considered as an alternative to using instrumental variables, say Z, satisfying $U \perp Z$. As explained in Blundell and Powell (2003, p. 332), the control function assumption is "no more nor less general" than the instrumental variable assumption, and both are implied by the stronger assumption $(U,V)\perp Z$. Assumption 1 (ii)-(iv) are adaptations of those in AIO to allow endogenous X. Assumption 1 (ii) is reasonable given that $H(x)$ and $L(x)$ are defined as the upper and lower bound. Assumption 1 (iii) and (iv) simply reflect that we wish to estimate some form of derivatives. The last condition of (iv) allows the order of integration and differentiation to be changed. Note that we do not need to assume X is continuous, however, for ease of exposition we restrict our attention to kernels for continuous variables. Under these assumptions, we can show that the identification formula for $\beta(x)$ based on (2) and (4) still holds true.

THEOREM 1. *Under Assumption 1, $\beta(x)$ is identified by (2), where $\beta(x,v)$ is identified by (4).*

This theorem formalises the identification argument described in AIO (Section 5.1). It should be noted that for this theorem, the object U can be a scalar, vector, or even an infinite dimensional object, the function $M(x,u)$ need not be monotone in u, and the region of integration for u need not be rectangular. A key insight for this result is that the Leibniz-type identity in (3) holds under weaker conditions (see Lemma 1 in Appendix 1.5).

1.2.2. Estimation. Based on Theorem 1, the local average response is written as

$$\beta(x) = \int_v \left[\nabla\Psi(x,v) + \frac{1}{G_M(x,v)} \left\{ \begin{array}{c} \Psi(x,v)\nabla G_M(x,v) \\ +H(x)\nabla G_H(x,v) \\ +L(x)\nabla G_L(x,v) \end{array} \right\} \right] dP(v|X=x, I_M(X,U)=1). \quad (6)$$

To estimate $\beta(x)$, we estimate each unknown component on the right hand side by a nonparametric estimator. Suppose X and V are absolutely continuous with respect to the Lebesgue measure. Let $f_M(\cdot)$ be generic notation for the joint or conditional density given that $I_M(X,U)=1$ (Y is uncensored). For example, $f_M(y|x)$ means the conditional density of Y given $X=x$ and $I_M(X,U)=1$; $E_M[\cdot]$ and $Var_M(\cdot)$ are defined analogously. For estimation, it is convenient to rewrite $\beta(x)$ in the following form

$$\beta(x) = f_M(x)^{-1}(1,1,H(x),L(x)) \begin{pmatrix} \xi(x) \\ \zeta(x) \\ \eta(x) \\ \theta(x) \end{pmatrix}, \quad (7)$$

where

$$
\begin{aligned}
\xi(x) &= \int_y y \nabla f_M(y,x)dy - \int_v \frac{\int_y y f_M(y,x,v)dy \nabla f_M(x,v)}{f_M(x,v)} dv, \\
\zeta(x) &= \int_v \int_y y f_M(y,x,v)dy \frac{\nabla G_M(x,v)}{G_M(x,v)} dv, \\
\eta(x) &= \int_v f_M(x,v) \frac{\nabla G_H(x,v)}{G_M(x,v)} dv, \qquad \theta(x) = \int_v f_M(x,v) \frac{\nabla G_L(x,v)}{G_M(x,v)} dv.
\end{aligned}
$$

Each component in $\beta(x)$ is estimated as follows. The boundary functions $H(x)$ and $L(x)$ are estimated by the local maximum and minimum, respectively, i.e.,

$$
\begin{aligned}
\hat{H}(x) &= \max_{i:|X_i - x| \leq b_n^H, Y_i \neq C_L, C_H} Y_i, \\
\hat{L}(x) &= \min_{i:|X_i - x| \leq b_n^L, Y_i \neq C_L, C_H} Y_i,
\end{aligned}
$$

where b_n^H and b_n^L are bandwidths. Let $K(a)$ be a $\dim(a)$-variate product kernel function such that $K(a) = \prod_{k=1}^{\dim(a)} \kappa(a^{(k)})$. As a proxy for V_i we use

$$
\hat{V}_i = X_i - \hat{\varphi}(W_i),
$$

where

$$
\hat{\varphi}(W_i) = \tau(\hat{f}(W_i), h_n) \frac{1}{n b_n^d} \sum_{j=1}^{n} X_j K\left(\frac{W_i - W_j}{b_n}\right),
$$

$\hat{f}(w) = \frac{1}{n b_n^d} \sum_{j=1}^{n} K\left(\frac{w - W_j}{b_n}\right)$ is the kernel density estimator for W, and

$$
\tau(t, h_n) = \begin{cases} 1/t & \text{if } t \geq 2h_n, \\ \frac{1}{8}\left\{\frac{49(t-h_n)^3}{h_n^4} - \frac{76(t-h_n)^4}{h_n^5} + \frac{31(t-h_n)^5}{h_n^6}\right\} & \text{if } h_n \leq t < 2h_n, \\ 0 & \text{if } t < h_n \end{cases}
$$

is a trimming function parameterised by h_n. This trimming term, due to Ai (1997), is introduced to deal with the denominator (or small density) problem of kernel estimators. The choice of h_n is briefly discussed in Ai (1997); it seems to be of little importance provided $h_n \to 0$. Integrating out $G_M(x,v)$, our estimator for $\Pr\{Y_i \neq C_L, C_H\}$ is given by $\hat{G}_M = n_M/n$, where $n_M = \sum_{i=1}^n I\{Y_i \neq C_L, C_H\}$ is the number of uncensored observations. Similarly, define $n_H = \sum_{i=1}^n I\{Y_i = C_H\}$, $n_L = \sum_{i=1}^n I\{Y_i = C_L\}$, $\hat{G}_H = n_H/n$, and $\hat{G}_L = n_L/n$. The conditional densities and their derivatives are estimated by

$$\begin{aligned}
\hat{f}_M(y,x,v) &= \frac{1}{n_M b_n^{2d+1}} \sum_{i:Y_i \neq C_L, C_H} K\left(\frac{y-Y_i}{b_n}\right) K\left(\frac{x-X_i}{b_n}\right) K\left(\frac{v-\hat{V}_i}{b_n}\right), \\
\hat{f}_M(x,v) &= \frac{1}{n_M b_n^{2d}} \sum_{i:Y_i \neq C_L, C_H} K\left(\frac{x-X_i}{b_n}\right) K\left(\frac{v-\hat{V}_i}{b_n}\right), \\
\triangledown \hat{f}_M(y,x) &= \frac{1}{n_M b_n^{d+2}} \sum_{i:Y_i \neq C_L, C_H} K\left(\frac{y-Y_i}{b_n}\right) \triangledown K\left(\frac{x-X_i}{b_n}\right), \\
\triangledown \hat{f}_M(x,v) &= \frac{1}{n_M b_n^{2d+1}} \sum_{i:Y_i \neq C_L, C_H} \triangledown K\left(\frac{x-X_i}{b_n}\right) K\left(\frac{v-\hat{V}_i}{b_n}\right), \\
\hat{f}(x,v) &= \frac{1}{n b_n^{2d}} \sum_{i=1}^{n} K\left(\frac{x-X_i}{b_n}\right) K\left(\frac{v-\hat{V}_i}{b_n}\right), \\
\triangledown \hat{f}(x,v) &= \frac{1}{n b_n^{2d+1}} \sum_{i=1}^{n} \triangledown K\left(\frac{x-X_i}{b_n}\right) K\left(\frac{v-\hat{V}_i}{b_n}\right).
\end{aligned}$$

The conditional probability $G_M(x,v)$ and its derivative are estimated by

$$\begin{aligned}
\hat{G}_M(x,v) &= \hat{G}_M \frac{\hat{f}_M(x,v)}{\hat{f}(x,v)}, \\
\triangledown \hat{G}_M(x,v) &= \hat{G}_M \frac{\triangledown \hat{f}_M(x,v)}{\hat{f}(x,v)} - \hat{G}_M \frac{\hat{f}_M(x,v) \triangledown \hat{f}(x,v)}{\hat{f}(x,v)^2}.
\end{aligned}$$

Similarly, $\triangledown G_H(x,v)$ and $\triangledown G_L(x,v)$ are estimated by

$$\begin{aligned}
\triangledown \hat{G}_H(x,v) &= \hat{G}_H \frac{\triangledown \hat{f}_H(x,v)}{\hat{f}(x,v)} - \hat{G}_H \frac{\hat{f}_H(x,v) \triangledown \hat{f}(x,v)}{\hat{f}(x,v)^2}, \\
\triangledown \hat{G}_L(x,v) &= \hat{G}_L \frac{\triangledown \hat{f}_L(x,v)}{\hat{f}(x,v)} - \hat{G}_L \frac{\hat{f}_L(x,v) \triangledown \hat{f}(x,v)}{\hat{f}(x,v)^2},
\end{aligned}$$

respectively, where $\hat{f}_H(x,v), \hat{f}_L(x,v), \triangledown \hat{f}_H(x,v), \triangledown \hat{f}_L(x,v), \hat{G}_H$ and $\hat{G}_L$ are defined analogously to their uncensored counterparts.

Based on the above notation and introducing the trimming terms $\tau(\hat{f}_M(x,v),h_n)$ and $\tau(\hat{f}(x,v),h_n)$, the components in $\beta(x)$ are estimated by

$$
\begin{aligned}
\hat{\xi}(x) &= \int_y y\nabla\hat{f}_M(y,x)dy - \int_v \left\{\int_y y\hat{f}_M(y,x,v)dy\right\}\nabla\hat{f}_M(x,v)\tau(\hat{f}_M(x,v),h_n)dv,\\
\hat{\zeta}(x) &= \int_v \left\{\int_y y\hat{f}_M(y,x,v)dy\right\}\nabla\hat{f}_M(x,v)\tau(\hat{f}_M(x,v),h_n)dv\\
&\quad - \int_v \left\{\int_y y\hat{f}_M(y,x,v)dy\right\}\nabla\hat{f}(x,v)\tau(\hat{f}(x,v),h_n)dv,\\
\hat{\eta}(x) &= \frac{\hat{G}_H}{\hat{G}_M}\int_v \nabla\hat{f}_H(x,v)dv - \frac{\hat{G}_H}{\hat{G}_M}\int_v \hat{f}_H(x,v)\nabla\hat{f}(x,v)\tau(\hat{f}(x,v),h_n)dv,\\
\hat{\theta}(x) &= \frac{\hat{G}_L}{\hat{G}_M}\int_v \nabla\hat{f}_L(x,v)dv - \frac{\hat{G}_L}{\hat{G}_M}\int_v \hat{f}_L(x,v)\nabla\hat{f}(x,v)\tau(\hat{f}(x,v),h_n)dv.
\end{aligned}
$$

The estimator $\hat{\beta}(x)$ is obtained by plugging the above estimators into (7).[2] Note that the integrals are approximated using numerical methods. If there is no censoring from above or below (i.e., $L(X)=-\infty$ or $H(X)=+\infty$, respectively), then we remove the term $\hat{\eta}(x)$ or $\hat{\theta}(x)$, respectively.

To analyse the asymptotic behaviour of $\hat{\beta}(x)$, we introduce the following assumptions. Let $|\cdot|$ be the Euclidean norm and $m_M(x,v)=E[Y|X=x,I_M(X,U)=1,V=v]$.

Assumption 2.

- **(i):** $\{Y_i,X_i,W_i,V_i\}_{i=1}^n$ *is i.i.d.*
- **(ii):** $E[a(W,X)|X]<\infty$ *for* $a(W,X)=E[Y^4|W,X]$, $E\left[\left|\frac{\nabla f_M(X,V)}{f_M(X,V)}\right|^4\middle|W,X\right]$, $E\left[\left|\frac{\nabla f(X,V)}{f(X,V)}\right|^4\middle|W,X\right]$, $E\left[\left|\nabla_{v'}\left(\frac{\nabla f_M(X,V)}{f_M(X,V)}\right)\right|^4\middle|W,X\right]$, *and* $E\left[\left|\nabla_{v'}\left(\frac{\nabla f(X,V)}{f(X,V)}\right)\right|^4\middle|W,X\right]$. *Furthermore,* $E[|\varphi(W)|^4|X]<\infty$, $E[|m_M(X,V)|^{4+\delta}]<\infty$, $E[|G_M(X,V)|^{4+\delta}]<\infty$, $E[|G_H(X,V)|^{2+\delta}]<\infty$, *and* $E[|G_L(X,V)|^{2+\delta}]<\infty$ *for some* $\delta>0$.
- **(iii):** $f_M(x,v)$ *and* $f(w)$ *are continuously differentiable of order* s *with respect to* (x,v) *and* w*, respectively, and all the derivatives are bounded over* (x,v) *and* w*, respectively. Also* $\int_v\int_x f_M(x,v)^{1-a}dxdv<\infty$ *and* $\int_v\int_x f(x,v)^{1-a}dxdv<\infty$ *for some* $0<a\le 1$.

[2]In this chapter, we employ the Nadaraya-Watson kernel estimator to construct $\hat{\beta}(x)$ because it simplifies the theoretical analysis below. It is also possible to use the formula in (6) and estimate the right hand side by local linear or polynomial estimators as in AIO. It is known that local polynomial fitting has some desirable properties, such as an absence of boundary effects and minimax efficiency (see, Section 3.2 of Fan and Gijbels, 1996). On the other hand, to estimate the conditional probabilities G_M, G_H, and G_L, local polynomial estimators are not constrained to lie between 0 and 1 (Hall, Wolff and Yao, 1999). Furthermore, the formula in (6) involves the conditional density $dP(v|X=x,I_M(X,U)=1)$, and its local polynomial fitting may require an additional bandwidth parameter for the dependent variable (Fan, Yao and Tong, 1996). A full comparison of different estimation methods is beyond the scope of this chapter.

(iv): *$E_M[Y|X=x,V=v]f_M(x,v)$ and $E[X|W=w]f(w)$ are continuously first-order differentiable with respect to (x,v) and w, respectively. Also, $\sup_{x,v}|E_M[Y|X=x,V=v]f_M(x,v)|<\infty$ and $\sup_{w}|E[X|W=w]f(w)|<\infty$.*

(v): *K is a product kernel taking the form of $K(a)=\prod_{k=1}^{\dim(a)}\kappa(a^{(k)})$, where κ is bounded and symmetric around zero. K satisfies $\int_a|K(a)|^{2+\delta}da<\infty$ for some $\delta>0$, $\int_a|a\nabla K(a)|da<\infty$, and $|a||K(a)|\to 0$ as $|a|\to\infty$, and the Fourier transform Ψ of K satisfies $\int_u \sup_{b\geq 1}|\Psi(bu)|du<\infty$. In addition,*

$$\int_a a^j K(a)du \begin{cases} =1 & \text{if } j=0,\\ =0 & \text{if } 1\leq j\leq s-1,\\ <\infty & \text{if } j=s. \end{cases}$$

(vi): *As $n\to\infty$, it holds $h_n\to 0$, $b_n\to 0$, $nb_n^{d+2}\to\infty$, $nb_n^{d+2+2s}\to 0$, $nb_n^{d+2}\int_w I\{f(w)<2h_n\}f(w,x)dw\to 0$, $\sqrt{nb_n^{d+2}}\{\hat{H}(x)-H(x)\}\xrightarrow{p}0$, and $\sqrt{nb_n^{d+2}}\{\hat{L}(x)-L(x)\}\xrightarrow{p}0$.*

(vii): *The partial derivatives with respect to x of $f_M(y,x)$, $f(x,v)$, $f_M(x,v)$, $f_H(x,v)$, and $f_L(x,v)$ exist up to the third order and are bounded. The first order partial derivatives with respect to v of $f_M(x,v)$, $f(x,v)$, $\log(\nabla f_M(x,v))$, and $\log(\nabla f(x,v))$ exist and are bounded.*

Assumption 2 (i) is on the sampling of data. This assumption can be weakened to allow for near-epoch dependent random variables (see Andrews, 1995). Assumption 2 (ii) contains boundedness conditions for the moments. Assumption 2 (iii) and (iv) are required to establish uniform convergence results for the kernel estimators in $\hat{\beta}(x)$. In particular, the last condition in (iii) is a restriction on the thickness of the tails of $f_M(x,v)$ and $f(x,v)$, which is required for the uniform convergence of the trimming terms. Assumption 2 (iv) is required for the uniform convergence of the kernel estimators to conditional expectations. Assumption 2 (v) contains standard bias-reducing conditions for a higher order kernel. Assumption 2 (vi) lists conditions on the bandwidth b_n and trimming parameter h_n as well as assumptions on the speed of convergence of the boundary function estimators $\hat{H}(x)$ and $\hat{L}(x)$. Chernozhukov (1998) and Altonji, Ichimura and Otsu (2013) provide primitive conditions for the convergence rates of $\hat{H}(x)$ and

$\hat{L}(x)$. Assumption 2 (vii) is required since we need to estimate the first order derivatives of these functions.

The asymptotic distribution of the nonparametric estimator $\hat{\beta}(x)$ for the local average response $\beta(x)$ is obtained as follows.

Theorem 2. *Under Assumptions 1 and 2,*

$$\sqrt{nb_n^{d+2}}\{\hat{\beta}(x) - \beta(x)\} \xrightarrow{d} N(0, c(x)'V(x)c(x)),$$

where $c(x) = (1, 1, H(x), L(x))'$ *and*

$$V(x) = \begin{pmatrix} \sigma_\xi^2 & 0 & 0 & 0 \\ 0 & \sigma_\zeta^2 & \sigma_{\zeta\eta} & \sigma_{\zeta\theta} \\ 0 & \sigma_{\zeta\eta} & \sigma_\eta^2 & \sigma_{\eta\theta} \\ 0 & \sigma_{\zeta\theta} & \sigma_{\eta\theta} & \sigma_\theta^2 \end{pmatrix} \otimes f_M(x,v)^{-1} G_M^{-2} \int_a \triangledown K(a) \triangledown K(a)' da,$$

$$\begin{aligned}
\sigma_\xi^2 &= \int_v \frac{Var_M(Y|x,v)}{G_M(x,v)} f_M(x,v)dv, \\
\sigma_\zeta^2 &= \int_v m_M(x,v)^2 G_M(x,v)(1 - G_M(x,v)) f_M(x,v)dv, \\
\sigma_\eta^2 &= H(x)^2 \int_v G_H(x,v)(1 - G_H(x,v)) f_M(x,v)dv, \\
\sigma_\theta^2 &= L(x)^2 \int_v G_L(x,v)(1 - G_L(x,v)) f_M(x,v)dv, \\
\sigma_{\zeta\eta} &= -H(x)^2 \int_v m_M(x,v) G_M(x,v) G_H(x,v) f_M(x,v)dv, \\
\sigma_{\zeta\theta} &= -L(x)^2 \int_v m_M(x,v) G_M(x,v) G_L(x,v) f_M(x,v)dv, \\
\sigma_{\eta\theta} &= -H(x)^2 L(x)^2 \int_v G_L(x,v) G_H(x,v) f_M(x,v)dv.
\end{aligned}$$

This theorem says that our nonparametric estimator $\hat{\beta}(x)$ is consistent and asymptotically normal. Note that the $\sqrt{nb_n^{d+2}}$-convergence rate of $\hat{\beta}(x)$ is identical to that of AIO for the case of exogenous X. However, the asymptotic variance is different from that of AIO. Both $c(x)$ and $V(x)$ can be estimated consistently in the same manner as the estimator itself; by replacing each component by the nonparametric estimator.

Here we focus on the estimation of $\beta(x)$ for a given x. As a summary of $\beta(x)$ over some range $\mathbb{X}$, it is also interesting to consider the average estimator

$$\hat{\beta} = \frac{\sum_{i=1}^{n} I\{X_i \in \mathbb{X}\}\hat{\beta}(X_i)}{\sum_{i=1}^{n} I\{X_i \in \mathbb{X}\}}.$$

For the case of exogenous X, the working paper version of AIO (Altonji, Ichimura and Otsu, 2008) studied the asymptotic properties of $\hat{\beta}$ and showed it is $\sqrt{n}$-consistent and asymptotically normal. Although a formal investigation is significantly more complicated and lengthy, we conjecture that $\hat{\beta}$ possesses similar asymptotic properties.

1.3. Simulation

In this section we evaluate the small sample properties of our nonparametric estimator. As a data generating process, we consider the following model:

$$Y = \begin{cases} M(X,U) & \text{if } 1 < M(X,U) < 8, \\ 1 & \text{if } M(X,U) \leq 1, \\ 8 & \text{if } 8 \leq M(X,U), \end{cases}$$

where

$$\begin{aligned} M(X,U) &= \alpha_0 + \alpha_1 X + \alpha_2 XU + U, \\ X &= W + U + \epsilon, \\ W &\sim U[0,6], \quad \epsilon \sim U[-1,1], \quad U \sim N(0,1). \end{aligned}$$

Note that $L(X) = 1$, $H(X) = 8$, $\varphi(W) = W$ and the variable $V = U + \epsilon$ plays the role of the control variable. We consider four parameterisations $(\alpha_0, \alpha_1, \alpha_2) = (1, 0.5, 0.5)$, $(0, 1, 0.5)$, $(2, 0, 1.5)$, and $(1.5, 1, 0)$ (called Models 1-4, respectively). In all cases the censoring points are treated as known. The local average response $\beta(x)$ is evaluated at $x \in \{1, 2, 3, 4, 5\}$. The sample size is set at $n = 1000$.

The simulation results are reported in Appendix 1.6. All results are based on 1000 Monte Carlo replications. In the tables, the rows labeled "Value of x" denote the values of x at which $\beta(x)$ is evaluated, and the rows labeled "True Value" report the true values of $\beta(x)$ (computed by Monte

Carlo integrations). The rows labeled "NPE" report the mean over Monte Carlo replications for the nonparametric estimator developed in this chapter. The rows labeled "No Endogeneity Control" report the mean for the nonparametric estimator without controlling for endogeneity, which is created by excluding the control function from our estimator. This estimator is the nonparametric estimator developed in AIO. Although in this simulation study, we use kernel estimators rather than local polynomial estimators as adopted in AIO. The rows labeled "No Censoring Control" report the mean for the nonparametric estimator without controlling for censoring. This estimator is the nonparametric estimator developed in Altonji and Matzkin (2005). In this chapter this estimator refers to using only $\hat{\xi}(x)$. For all nonparametric estimators, we use Silverman's plug-in bandwidth for b_n and the Gaussian kernel for $K(\cdot)$. Also, in the simulation study, we do not incorporate the trimming term (i.e., set as $\tau(t, h_n) = 1/t$) as there appears to be little effect in the results. This may be due to the uniformly distributed variables used, however, trimming tends to be necessary only insofar as proving theoretical properties and rarely impacts practical performance. To evaluate the integrals in the estimators,we employ adaptive quadratures. The rows labeled "Tobit" report the mean over Monte Carlo replications for the maximum likelihood Tobit estimator using the fourth-order polynomial regression function with no adjustment for endogeneity. The rows labeled "SD" report the standard deviation over Monte Carlo replications for each estimator. Finally, the rows labeled "NPE (Half Bandwidth)" report the mean over Monte Carlo replications for our nonparametric estimator using half of the plug-in bandwidth.

Model 1 is the benchmark case. The proposed estimator "NPE" shows a superb performance. It has small bias across all values of x and reasonably small standard deviations (compared to the Tobit estimator, for example). The half bandwidth estimator also shows reasonable results. Compared to "NPE", as expected, the half bandwidth estimator yields smaller bias but larger standard deviation. The "No Endogeneity Control" estimator proposed in AIO incurs biases for all values of x. It seems there is no noticeable pattern in the bias. It has large upward bias at $x = 2$ and large downward bias at $x = 5$. Also, the "No Censoring Control" estimator proposed in Altonji and Matzkin (2005) shows severe downward biases. These results show that in the current setting, it is crucial to control for both endogeneity and censoring problems at the same time. The "Tobit" estimator also shows considerable bias for most values of x which is not surprising.

Models 2 and 3 consider the case without an intercept and without the linear term in X, respectively. For both cases, we obtained similar results. The "NPE" estimator and the half bandwidth estimator show reasonable performance for most values of x; other estimators are (often significantly) biased. Model 4 considers the linear separable model. However, since X is endogenous, the Tobit estimator is still inconsistent and the simulations confirms the presence of this endogeneity bias.

Our "NPE" estimator works well for most cases. However, when $x = 1$ or 5 (i.e., near the boundaries of the support of X), it may incur non-negligible bias (see, Model 3 with $x = 1$ and Model 4 with $x = 5$). For such cases, we should introduce a trimming term to avoid low density problems or a boundary correction kernel.

1.4. Conclusion

In this chapter we develop a nonparametric estimator for the local average response of a censored dependent variable to an endogenous regressor in a nonseparable model. The unobservable error term is not restricted to be scalar and the nonseparable function need not be monotone in the unobservable. We formalise the identification argument in Altonji, Ichimura and Otsu (2012) in the case of endogenous regressors, and study the asymptotic properties of the nonparametric estimator. Our simulation results suggest that it is important to correct for the effects of both censoring and endogeneity.

Further research is needed in dynamic settings, as well as looking at how measurement error impacts such models and how discrete regressors complicate the identification argument. It could also prove possible to use the results from this chapter along with AIO to create a test for endogeneity in this censored, nonseparable model.

1.5. Appendix - Mathematical Proofs

1.5.1. Proof of Theorem 1. Theorem 1 follows directly from (2), (4), and Lemma 1 below.

LEMMA 1. *Under Assumption 1,*

$$\begin{aligned}\triangledown \int M(x,u)I_M(x,u)dP(u|V=v) &= \int \triangledown M(x,u)I_M(x,u)dP(u|V=v) \\ &\quad -H(x)\triangledown G_H(x,v) - L(x)\triangledown G_L(x,v),\end{aligned}$$

for almost every v.

The proof of Lemma 1 follows trivially from the proof of AIO (2012, Lemma 3.1); the adapted proof is included here for completeness.

It is sufficient to prove Lemma 1 for $\triangledown_1$, the partial derivative with respect to the first element of x:

$$\begin{aligned}&\triangledown_1 \int M(x,u)I_M(x,u)dP(u|V=v) \\ = &\int \triangledown_1 M(x,u)I_M(x,u)dP(u|V=v) - H(x)\triangledown_1 G_H(x,v) - L(x)\triangledown_1 G_L(x,v),\end{aligned}$$

for almost every v. The left hand side is given by

$$\begin{aligned}&\lim_{\epsilon\to 0}\left[\int M(x+\epsilon\mathbf{e}_1,u)I_M(x+\epsilon\mathbf{e}_1,u)dP(u|V=v) - \int M(x,u)I_M(x,u)dP(u|V=v)\right]/\epsilon \\ = &\lim_{\epsilon\to 0}\int [M(x+\epsilon\mathbf{e}_1,u)-M(x,u)]I_M(x+\epsilon\mathbf{e}_1,u)dP(u|V=v)/\epsilon \\ &+\lim_{\epsilon\to 0}\int M(x,u)[I_M(x+\epsilon\mathbf{e}_1,u)-I_M(x,u)]dP(u|V=v)/\epsilon \\ = &T_1+T_2,\end{aligned}$$

where $\mathbf{e}=(1,0,\ldots,0)'$. Assumption 1 (ii) and (iv) imply

$$\lim_{\epsilon\to 0} I_M(x+\epsilon\mathbf{e},U) = I_M(x,U) \text{ a.s.}$$

Thus, the Lebesgue dominated convergence theorem implies

$$T_1 = \int \triangledown_1 M(x,u)I_M(x,u)dP(u|V=v),$$

for almost every v. For T_2, using Assumption 1 (ii),

$$\begin{aligned}
& I_M(x+\epsilon\mathbf{e}, U) - I_M(x, U) \\
= \; & I\{L(x+\epsilon\mathbf{e}) < M(x+\epsilon\mathbf{e}, U)\} - I\{L(x) < M(x, U)\} \\
& + I\{M(x+\epsilon\mathbf{e}, U) < H(x+\epsilon\mathbf{e})\} - I\{M(x, U) < H(x)\} \text{ a.s.},
\end{aligned}$$

for all $\epsilon > 0$ sufficiently close to 0. Therefore,

$$\begin{aligned}
T_2 = \; & \lim_{\epsilon\to 0} \int M(x,u)[I\{L(x+\epsilon\mathbf{e}) < M(x+\epsilon\mathbf{e}, u)\} - I\{L(x) < M(x,u)\}]dP(u|V=v)/\epsilon \\
& + \lim_{\epsilon\to 0} \int M(x,u)[I\{M(x+\epsilon\mathbf{e}, u) < H(x+\epsilon\mathbf{e}_1)\} - I\{M(x,u) < H(x)\}]dP(u|V=v)/\epsilon.
\end{aligned}$$

Noting that $I\{L(x+\epsilon\mathbf{e}) < M(x+\epsilon\mathbf{e}, u)\} = 1 - I\{L(x+\epsilon\mathbf{e}) \geq M(x+\epsilon\mathbf{e}, u)\}$, the proof is completed by the following lemma.

LEMMA 2. *Under Assumption 1,*

$$\lim_{\epsilon\to 0} \int M(x,u)[I\{M(x+\epsilon\mathbf{e}, u) > L(x+\epsilon\mathbf{e})\} - I\{M(x,u) > L(x)\}]dP(u|V=v)/\epsilon = -L(x)\nabla_1 G_L(x,v), \tag{8}$$

for almost every v.

Proof of Lemma 2. Presented here is only the argument for the lower bound. The argument for the upper bound is analogous. To prove this lemma, it is sufficient to show that both an upper bound and a lower bound of the left hand side converge to the right hand side as $\epsilon \to 0$. The left hand side can be written as

$$\begin{aligned}
& \lim_{\epsilon\to 0} \int M(x,u) I\{M(x+\epsilon\mathbf{e}, u) > L(x+\epsilon\mathbf{e})\} I\{M(x,u) \leq L(x)\} dP(u|V=v)/\epsilon \\
& - \lim_{\epsilon\to 0} \int M(x,u) I\{M(x+\epsilon\mathbf{e}, u) \leq L(x+\epsilon\mathbf{e})\} I\{M(x,u) > L(x)\} dP(u|V=v)/\epsilon,
\end{aligned}$$

for almost every v. Assumption 1 (iv) implies that if $M(x+\epsilon\mathbf{e}, u) \leq L(x+\epsilon\mathbf{e})$, then $M(x,u) \leq L(x+\epsilon\mathbf{e}) + \epsilon B(u)$ for all ϵ sufficiently close to 0. Similarly, $M(x+\epsilon\mathbf{e}, u) > L(x+\epsilon\mathbf{e})$ implies $M(x,u) > L(x+\epsilon\mathbf{e}) - \epsilon B(u)$ for all ϵ sufficiently close to 0. Consequently, the left hand side of

(8) can be bounded from below by

$$
\begin{aligned}
&\lim_{\epsilon\to 0}\int L(x+\epsilon\mathbf{e})I\{M(x+\epsilon\mathbf{e},u)>L(x+\epsilon\mathbf{e})\}I\{M(x,u)\le L(x)\}dP(u|V=v)/\epsilon\\
&-\lim_{\epsilon\to 0}\int B(u)I\{M(x+\epsilon\mathbf{e},u)>L(x+\epsilon\mathbf{e})\}I\{M(x,u)\le L(x)\}dP(u|V=v)\\
&-\lim_{\epsilon\to 0}\int L(x+\epsilon\mathbf{e})I\{M(x+\epsilon\mathbf{e},u)\le L(x+\epsilon\mathbf{e})\}I\{M(x,u)> L(x)\}dP(u|V=v)/\epsilon\\
&-\lim_{\epsilon\to 0}\int B(u)I\{M(x+\epsilon\mathbf{e},u)\le L(x+\epsilon\mathbf{e})\}I\{M(x,u)> L(x)\}dP(u|V=v),
\end{aligned}
$$

for almost every v. By Assumption 1 (ii) and (iv), the Lebesgue dominated convergence theorem implies that the second and fourth terms converge to 0. The first and third terms can be combined to give

$$
\lim_{\epsilon\to 0}L(x+\epsilon\mathbf{e})\int[I\{M(x+\epsilon\mathbf{e},u)>L(x+\epsilon\mathbf{e})\}-I\{M(x,u)>L(x)\}]dP(u|V=v)/\epsilon=-L(x)\triangledown_1 G_L(x,v),
$$

for almost every v. The same reasoning obtains an equivalent result for $-H(x)\triangledown_1 G_H(x,v)$. Therefore, the conclusion follows.

1.5.2. Proof of Theorem 2. Note that the convergence rates of $\hat{f}_M(x)$, $\hat{H}(x)$, and $\hat{L}(x)$ are faster than the derivative estimators contained in $(\hat{\xi}(x),\hat{\zeta}(x),\hat{\eta}(x),\hat{\theta}(x))$. Thus, under Assumption 2 (i), (ii), (v), and (vi),

$$
\sqrt{nb_n^{d+2}}\{\hat{\beta}(x)-\beta(x)\} \;=\; c(x)'\sqrt{nb_n^{d+2}}\begin{pmatrix}\hat{\xi}(x)-\xi(x)\\ \hat{\zeta}(x)-\zeta(x)\\ \hat{\eta}(x)-\eta(x)\\ \hat{\theta}(x)-\theta(x)\end{pmatrix}+o_p(1),
$$

where $c(x)'=f_M(x)^{-1}(1,1,H(x),L(x))$.

In the following lemma, we derive the asymptotic linear form of $\hat{\xi}(x)-\xi(x)$. Let $\tilde{f}_M(a)$ be the object defined by replacing $\hat{V}_i$ in $\hat{f}_M(a)$ with V_i.

Lemma 3. *Under Assumption 2,*

$$
\begin{aligned}
\hat{\xi}(x)-\xi(x) &= \left\{\frac{1}{n_M b_n^{d+1}} \sum_{i:Y_i\neq C_L,C_H} Y_i \nabla K\left(\frac{x-X_i}{b_n}\right) - \int_y y\nabla f_M(y,x)dy\right\} \\
&\quad -\left\{\frac{1}{n_M b_n^{d+1}} \sum_{i:Y_i\neq C_L,C_H} m_M(x,V_i) \nabla K\left(\frac{x-X_i}{b_n}\right) - \int_v m_M(x,v)\nabla f_M(x,v)dv\right\} \\
&\quad +o_p((nb_n^{d+2})^{-1/2}).
\end{aligned}
$$

Proof of Lemma 3. Decompose

$$
\begin{aligned}
\hat{\xi}(x)-\xi(x) &= \int_y y\{\nabla \hat{f}_M(y,x)-\nabla f_M(y,x)\}dy \\
&\quad -\int_v \left\{\int_y y\hat{f}_M(y,x,v)dy - \int_y y f_M(y,x,v)dy\right\} \nabla \hat{f}_M(x,v)\tau(\hat{f}_M(x,v),h_n)dv \\
&\quad -\int_v \left\{\int_y y f_M(y,x,v)dy\right\} \{\nabla \hat{f}_M(x,v)-\nabla f_M(x,v)\}\tau(\hat{f}_M(x,v),h_n)dv \\
&\quad -\int_v \left\{\int_y y f_M(y,x,v)dy\right\} \nabla f_M(x,v)\{\tau(\hat{f}_M(x,v),h_n)-\tau(f_M(x,v),0)\}dv \\
&\equiv T_1-T_2-T_3-T_4.
\end{aligned}
$$

For T_2, decompose

$$
\begin{aligned}
T_2 &= \int_v \left\{\int_y y\hat{f}_M(y,x,v)dy - \int_y y f_M(y,x,v)dy\right\} \{\nabla \hat{f}_M(x,v)-\nabla f_M(x,v)\}\tau(\hat{f}_M(x,v),h_n)dv \\
&\quad +\int_v \left\{\int_y y\hat{f}_M(y,x,v)dy - \int_y y f_M(y,x,v)dy\right\} \nabla f_M(x,v)\{\tau(\hat{f}_M(x,v),h_n)-\tau(f_M(x,v),0)\}dv \\
&\quad +\int_v \left\{\int_y y\hat{f}_M(y,x,v)dy - \int_y y f_M(y,x,v)dy\right\} \nabla f_M(x,v)\tau(f_M(x,v),0)dv \\
&\equiv T_{21}+T_{22}+T_{23}.
\end{aligned}
$$

For T_{23},

$$
\begin{aligned}
T_{23} &= \int_v \left\{\int_y y\hat{f}_M(y,x,v)dy - \int_y y\tilde{f}_M(y,x,v)dy\right\} \nabla f_M(x,v) f_M(x,v)^{-1}dv \\
&\quad +\int_v \left\{\int_y y\tilde{f}_M(y,x,v)dy - \int_y y f_M(y,x,v)dy\right\} \nabla f_M(x,v) f_M(x,v)^{-1}dv \\
&\equiv T_{231}+T_{232}.
\end{aligned}
$$

For T_{232},

$$\begin{aligned}
T_{232} &= \int_v \left\{ \frac{1}{n_M b_n^{2d+1}} \sum_{i:Y_i \neq C_L, C_H} \int_y yK\left(\frac{y-Y_i}{b_n}\right) dy K\left(\frac{x-X_i}{b_n}\right) K\left(\frac{v-V_i}{b_n}\right) \right. \\
&\quad \left. - \int_y y f_M(y,x,v) dy \right\} \frac{\nabla f_M(x,v)}{f_M(x,v)} dv \\
&= \int_v \left\{ \frac{1}{n_M b_n^{2d}} \sum_{i:Y_i \neq C_L, C_H} Y_i K\left(\frac{x-X_i}{b_n}\right) K\left(\frac{v-V_i}{b_n}\right) - \int_y y f_M(y,x,v) dy \right\} \frac{\nabla f_M(x,v)}{f_M(x,v)} dv \\
&= \frac{1}{n_M b_n^d} \sum_{i:Y_i \neq C_L, C_H} Y_i \frac{\nabla f_M(x,V_i)}{f_M(x,V_i)} K\left(\frac{x-X_i}{b_n}\right) - \int_v \int_y y \frac{\nabla f_M(x,v)}{f_M(x,v)} f_M(y,x,v) dy dv + O_p(b_n^s) \\
&= O_p((nb_n^d)^{-1/2}) + O_p(b_n^s),
\end{aligned}$$

where the second equality follows from the change of variables $a = \frac{y-Y_i}{b_n}$ and Assumption 2 (v), the third equality also follows from the change of variables $a = \frac{v-V_i}{b_n}$ and Assumption 2 (v), and the last equality follows from a central limit theorem for the kernel estimator in the form of $\frac{1}{n_M b_n^d} \sum_{i:Y_i \neq C_L, C_H} g_1(Y_i, V_i) K\left(\frac{x-X_i}{b_n}\right)$ with $g_1(Y_i, V_i) \equiv Y_i \frac{\nabla f_M(x,V_i)}{f_M(x,V_i)}$.

For T_{231},

$$\begin{aligned}
T_{231} &= \int_v \frac{1}{n_M b_n^{2d}} \sum_{i:Y_i \neq C_L, C_H} Y_i K\left(\frac{x-X_i}{b_n}\right) \left\{ K\left(\frac{v-V_i+\hat{e}_i}{b_n}\right) - K\left(\frac{v-V_i}{b_n}\right) \right\} \frac{\nabla f_M(x,v)}{f_M(x,v)} dv \\
&= \int_v \frac{1}{n_M b_n^{2d}} \sum_{i:Y_i \neq C_L, C_H} Y_i K\left(\frac{x-X_i}{b_n}\right) K'\left(\frac{v-V_i}{b_n}\right) \frac{\hat{e}_i}{b_n} \frac{\nabla f_M(x,v)}{f_M(x,v)} dv + o_p(n^{-1/2}) \\
&= \frac{1}{n_M b_n^d} \sum_{i:Y_i \neq C_L, C_H} g_2(Y_i, V_i) \hat{e}_i K\left(\frac{x-X_i}{b_n}\right) (1+o(1)) + o_p(n^{-1/2}),
\end{aligned}$$

where the first equality follows from the change of variables $a = \frac{y-Y_i}{b_n}$ and the definition $\hat{e}_i \equiv \hat{\varphi}(W_i) - \varphi(W_i)$, the second equality follows from an expansion around $\hat{e}_i = 0$ and $\max_{1 \leq i \leq n} |\hat{e}_i| = o_p(n^{-1/4})$ (by applying the uniform convergence result in Andrews, 1995, Theorem 1, based on Assumption 2), and the third equality follows from the change of variables $a = \frac{v-V_i}{b_n}$ with

$\int_a K'(a)\,da = 0$ and the definition $g_2(Y_i, V_i) \equiv Y_i \nabla_{v'} \left(\frac{\nabla f_M(x,V_i)}{f_M(x,V_i)} \right) \int_a K'(a)\,a da$ based on Assumption 2 (ii) and (v). We can break down T_{231} further as follows

$$
\begin{aligned}
& \frac{1}{n_M b_n^d} \sum_{i:Y_i \neq C_L, C_H} \hat{e}_i g_2(Y_i, V_i) K\left(\frac{x - X_i}{b_n}\right) \\
= & \frac{1}{n_M b_n^d} \sum_{i:Y_i \neq C_L, C_H} \left\{ \tau(\hat{f}_W(W_i), h_n) \frac{1}{n b_n^d} \sum_{j=1}^{n} X_j K\left(\frac{W_i - W_j}{b_n}\right) - \varphi(W_i) \right\} g_2(Y_i, V_i) K\left(\frac{x - X_i}{b_n}\right) \\
= & \frac{1}{n_M b_n^d} \sum_{i:Y_i \neq C_L, C_H} \left\{ \{\tau(\hat{f}_W(W_i), h_n) - \tau(f(W_i), 0)\} \frac{1}{n b_n^d} \sum_{j=1}^{n} X_j K\left(\frac{W_i - W_j}{b_n}\right) \right\} g_2(Y_i, V_i) K\left(\frac{x - X_i}{b_n}\right) \\
& + \frac{1}{n_M b_n^d} \sum_{i:Y_i \neq C_L, C_H} \left\{ f(W_i)^{-1} \frac{1}{n b_n^d} \sum_{j=1}^{n} X_j K\left(\frac{W_i - W_j}{b_n}\right) - \varphi(W_i) \right\} g_2(Y_i, V_i) K\left(\frac{x - X_i}{b_n}\right) \\
= & T_{2311} + T_{2312}.
\end{aligned}
$$

We denote $T_{2312} = \frac{1}{n n_M b_n^d} \sum_{i:Y_i \neq C_L, C_H} \sum_{j=1}^{n} C_{ij}$. Using the definition of $\varphi(W_i)$, the mean of C_{ij} is

$$
\begin{aligned}
& E[C_{ij}] \\
= & E\left[\frac{g_2(Y_i, V_i)}{f(W_i)} \left\{ X_j \frac{1}{b_n^d} K\left(\frac{W_i - W_j}{b_n}\right) - \int \tilde{x} f(\tilde{x}, W_i) d\tilde{x} \right\} K\left(\frac{x - X_i}{b_n}\right) \right] \\
= & E\left[\left\{ E\left[X_j \frac{1}{b_n^d} K\left(\frac{W_i - W_j}{b_n}\right) \middle| Y_i, V_i, X_i, W_i \right] - \int \tilde{x} f(\tilde{x}, W_i) d\tilde{x} \right\} \frac{g_2(Y_i, V_i)}{f(W_i)} K\left(\frac{x - X_i}{b_n}\right) \right].
\end{aligned}
$$

Note that by the change of variables $a = \frac{W_i - w}{b_n}$ and Assumption 2 (v),

$$
E\left[X_j \frac{1}{b_n^d} K\left(\frac{W_i - W_j}{b_n}\right) \middle| Y_i, V_i, X_i, W_i \right] = \int \tilde{x} f(\tilde{x}, W_i) d\tilde{x} + O(b_n^s),
$$

and therefore $E[T_{2312}] = O_p(b_n^{s-d})$. Similarly, we obtain $E[C_{ij}^2] = O_p(b_n)$ by using Assumption 2 (ii), (v), and (vi), which implies $Var(T_{2312}) = O_p(n^{-2} b_n^{-d+1})$. Combining these results, we obtain $T_{2312} = o_p((n b_n^{d+2})^{-1/2})$.

For T_{2311}, an expansion of $\tau(\hat{f}(W_i), h_n)$ around $\hat{f}(W_i) = f(W_i)$ yields

$$
\begin{aligned}
& T_{2311} \\
= & \frac{1}{n_M b_n^d} \sum_{i:Y_i \neq C_L, C_H} \left\{ \{\tau(f(W_i), h_n) - \tau(f(W_i), 0)\} \frac{1}{n b_n^d} \sum_{j=1}^n X_j K\left(\frac{W_i - W_j}{b_n}\right) \right\} g(Y_i, V_i) K\left(\frac{x - X_i}{b_n}\right) \\
& + \frac{1}{n_M b_n^d} \sum_{i:Y_i \neq C_L, C_H} \left\{ \{\tau'(f(W_i), h_n) \left\{\hat{f}(W_i) - f(W_i)\right\} \frac{1}{n b_n^d} \sum_{j=1}^n X_j K\left(\frac{W_i - W_j}{b_n}\right) \right\} g(Y_i, V_i) K\left(\frac{x - X_i}{b_n}\right) \\
& + \frac{1}{n_M b_n^d} \sum_{i:Y_i \neq C_L, C_H} \left\{ O_p\left(\max_{1 \le i \le n} |\hat{f}(W_i) - f(W_i)|^2\right) \frac{1}{n b_n^d} \sum_{j=1}^n X_j K\left(\frac{W_i - W_j}{b_n}\right) \right\} g(Y_i, V_i) K\left(\frac{x - X_i}{b_n}\right) \\
\equiv & T_{23111} + T_{23112} + T_{23113}.
\end{aligned}
$$

By applying the uniform convergence result of Andrews (1995, Theorem 1), we obtain $\max_{1 \le i \le n} |\hat{f}(W_i) - f(W_i)| = o_p(n^{-1/4})$, which implies $T_{23113} = o_p(n^{-1/2})$. For T_{23111}, using two change of variable arguments, Taylor expansions, the Cauchy-Scwharz inequality, and noting that $\{\tau(f(w), h_n) f(w) - 1\}$ is bounded, we can write the mean of T_{23111} as

$$
\begin{aligned}
E[T_{23111}] &= E\left[\{\tau(f(W_i), h_n) - \tau(f(W_i), 0)\} \frac{1}{b_n^d} X_j K\left(\frac{W_i - W_j}{b_n}\right) \frac{1}{b_n^d} g_2(Y_i, V_i) K\left(\frac{x - X_i}{b_n}\right)\right] \\
&= E\left[\{\tau(f(W_i), h_n) - \tau(f(W_i), 0)\} \varphi(W_i) f(W_i) \frac{1}{b_n^d} g_2(Y_i, V_i) K\left(\frac{x - X_i}{b_n}\right)\right] + O(b_n^s) \\
&= \int I\{f(w) < 2h_n\} \{\tau(f(w), h_n) f(w) - 1\} \varphi(w) E[g_2(y, v)|w, x] f(w, x) dw + O(b_n^s) \\
&\le \sqrt{\int I\{f(w) < 2h_n\} f(w, x) dw} \sqrt{\int |\varphi(w) E[g_2(y, v)|w, x]|^2 f(w, x) dw} + O(b_n^s),
\end{aligned}
$$

where $\int |\varphi(w) E[g_2(y, v)|w, x]|^2 f(w, x) dw < \infty$ by Assumption 2 (ii). Thus $\sqrt{n b_n^{d+2}} E[T_{23111}] \to 0$ by Assumption 2 (vi). Using similar arguments, we have

$$
\begin{aligned}
& E[T_{23111}^2] \\
= & \frac{1}{n n_M} E\left[\{\tau(f(W_i), h_n) - \tau(f(W_i), 0)\}^2 \frac{1}{b_n^{2d}} X_j^2 K\left(\frac{W_i - W_j}{b_n}\right)^2 \frac{1}{b_n^{2d}} g_2(Y_i, V_i)^2 K\left(\frac{x - X_i}{b_n}\right)^2\right] \\
\le & \sqrt{\int I\{f(w) < 2h_n\} f(w, x) dw} \sqrt{\int |E[g_2(y, v)^2|w, x]|^2 f(w, x) dw} O(n^{-2} b_n^{-2d+1}),
\end{aligned}
$$

which implies $\sqrt{n b_n^{d+2}} Var(T_{23111}) \to 0$. Combining these results, we obtain $\sqrt{n b_n^{d+2}} T_{23111} \xrightarrow{p} 0$. For T_{23112}, a similar argument to T_{2312} implies that $T_{23112} = o_p((n b_n^{d+2})^{-1/2})$.

For T_{22}, it holds

$$
\begin{aligned}
T_{22} &= \int_v \left\{ \int_y y\hat{f}_M(y,x,v)dy - \int_y y f_M(y,x,v)dy \right\} \nabla f_M(x,v)\{\tau(\hat{f}_M(x,v),h_n) - \tau(f_M(x,v),0)\}dv \\
&\leq C \sup_{x,v} \left| \int_y y\hat{f}_M(y,x,v)dy - \int_y y f_M(y,x,v)dy \right| \sup_{x,v} |\tau(\hat{f}_M(x,v),h_n) - \tau(f_M(x,v),0)| \\
&= o_p(n^{-1/2}),
\end{aligned}
$$

where the last equality follows from

$$
\begin{aligned}
\sup_{x,v} \left| \int_y y\hat{f}_M(y,x,v)dy - \int_y y f_M(y,x,v)dy \right| &= O_p(n^{-1/2}b_n^{-2d}), \\
\sup_{x,v} \left| \tau(\hat{f}_M(x,v),h_n) - \tau(f_M(x,v),0) \right| &= O_p(n^{-1/2}b_n^{-2d}),
\end{aligned}
$$

again, using Andrews (1995, Theorem 1). Thus we obtain $\sqrt{nb_n^{d+2}}T_{22} \xrightarrow{p} 0$. Similarly, we can show that $\sqrt{nb_n^{d+2}}T_{21} \xrightarrow{p} 0$. Combining these results, we obtain $\sqrt{nb_n^{d+2}}T_2 \xrightarrow{p} 0$. By a similar approach to T_2, we can show that $\sqrt{nb_n^{d+2}}T_4 \xrightarrow{p} 0$. For T_3, following a similar argument to T_{22} and T_{231},

$$
\begin{aligned}
T_3 &= \int_v \left\{ \int_y y f_M(y,x,v)dy \right\} \{\nabla\hat{f}_M(x,v) - \nabla f_M(x,v)\}\{\tau(\hat{f}_M(x,v),h_n) - \tau(f_M(x,v),0)\}dv \\
&\quad + \int_v \left\{ \int_y y f_M(y,x,v)dy \right\} \{\nabla\hat{f}_M(x,v) - \nabla\tilde{f}_M(x,v)\} f_M(x,v)^{-1}dv \\
&\quad + \int_v \left\{ \int_y y f_M(y,x,v)dy \right\} \{\nabla\tilde{f}_M(x,v) - \nabla f_M(x,v)\} f_M(x,v)^{-1}dv \\
&= \int_v \left\{ \int_y y f_M(y,x,v)dy \right\} \{\nabla\tilde{f}_M(x,v) - \nabla f_M(x,v)\} f_M(x,v)^{-1}dv + o_p((nb_n^{d+2})^{-1/2}).
\end{aligned}
$$

For T_1, again in a similar way to T_{231}, we can show

$$
\begin{aligned}
T_1 &= \int_y y\{\nabla\hat{f}_M(y,x) - \nabla\tilde{f}_M(y,x)\}dy + \int_y y\{\nabla\tilde{f}_M(y,x) - \nabla f_M(y,x)\}dy \\
&= \int_y y\{\nabla\tilde{f}_M(y,x) - \nabla f_M(y,x)\}dy + o_p((nb_n^{d+2})^{-1/2}).
\end{aligned}
$$

Combining these results, the conclusion follows.

By repeating these steps, we can obtain the asymptotic linear forms for $\hat{\zeta}(x)$, $\hat{\eta}(x)$, and $\hat{\theta}(x)$ (the proofs are omitted).

Lemma 4. *Under Assumption 2,*

$$\begin{aligned}
\hat{\zeta}(x) - \zeta(x) &= \frac{1}{n_M b_n^{d+1}} \sum_{i:Y_i \neq C_L, C_H} m_M(x, V_i) \nabla K\left(\frac{x - X_i}{b_n}\right) \\
&\quad - \frac{1}{n_M b_n^{d+1}} \sum_{i=1}^{n} m_M(x, V_i) G_M(x, V_i) \nabla K\left(\frac{x - X_i}{b_n}\right) \\
&\quad + \int_v m_M(x, v) \frac{f_M(x, v)}{f(x, v)} \nabla f(x, v) dv - \int_v m_M(x, v) \nabla f_M(x, v) dv + o_p((nb_n^{d+2})^{-1/2}), \\
\hat{\eta}(x) - \eta(x) &= \frac{1}{n_M b_n^{d+1}} \sum_{i:Y_i = C_H} \nabla K\left(\frac{x - X_i}{b_n}\right) - \frac{1}{n_M b_n^{d+1}} \sum_{i=1}^{n} G_H(x, V_i) \nabla K\left(\frac{x - X_i}{b_n}\right) \\
&\quad + \frac{G_H}{G_M} \int_v \frac{f_H(x, v)}{f(x, v)} \nabla f(x, v) dv - \frac{G_H}{G_M} \int_v \nabla f_H(x, v) dv + o_p((nb_n^{d+2})^{-1/2}), \\
\hat{\theta}(x) - \theta(x) &= \frac{1}{n_M b_n^{d+1}} \sum_{i:Y_i = C_L} \nabla K\left(\frac{x - X_i}{b_n}\right) - \frac{1}{n_M b_n^{d+1}} \sum_{i=1}^{n} G_L(x, V_i) \nabla K\left(\frac{x - X_i}{b_n}\right) \\
&\quad + \frac{G_L}{G_M} \int_v \frac{f_L(x, v)}{f(x, v)} \nabla f(x, v) dv - \frac{G_L}{G_M} \int_v \nabla f_L(x, v) dy + o_p((nb_n^{d+2})^{-1/2}).
\end{aligned}$$

It remains to derive the asymptotic variance for our estimator. By Lemma 3, the asymptotic variance of $\hat{\xi}(x)$ is

$$\begin{aligned}
Var\left(\sqrt{nb_n^{d+2}}\{\hat{\xi}(x) - \xi(x)\}\right) &\to \lim_{n\to\infty} \frac{n^2}{n_M^2 b_n^d} E\left[I\{Y_i \neq C_L, C_H\}(Y_i - m_M(x, V_i))^2 \nabla K\left(\frac{x - X_i}{b_n}\right)^2\right] \\
&= G_M^{-2} \int_v \frac{Var_M(Y|x, v)}{G_M(x, v)} f(x, v) dv \int_a \nabla K(a)^2 da,
\end{aligned}$$

where the equality follows from the change of variables. Also, by Lemma 4,

$$\begin{aligned}
Var\left(\sqrt{nb_n^{d+2}}\{\hat{\zeta}(x) - \zeta(x)\}\right) &\to G_M^{-2} \int_v m_M(x, v)^2 G_M(x, v)(1 - G_M(x, v)) f(x, v) dv \int_a \nabla K(a)^2 da, \\
Var\left(\sqrt{nb_n^{d+2}}\{\hat{\eta}(x) - \eta(x)\}\right) &\to G_M^{-2} \int_v G_H(x, v)(1 - G_H(x, v)) f(x, v) dv \int_a \nabla K(a)^2 da, \\
Var\left(\sqrt{nb_n^{d+2}}\{\hat{\theta}(x) - \theta(x)\}\right) &\to G_M^{-2} \int_v G_L(x, v)(1 - G_L(x, v)) f(x, v) dv \int_a \nabla K(a)^2 da.
\end{aligned}$$

For the asymptotic covariance terms, we have

$$
\begin{aligned}
Cov\left(\sqrt{nb_n^{d+2}}\{\hat{\xi}(x)-\xi(x)\},\sqrt{nb_n^{d+2}}\{\hat{\zeta}(x)-\zeta(x)\}\right) &\rightarrow 0,\\
Cov\left(\sqrt{nb_n^{d+2}}\{\hat{\xi}(x)-\xi(x)\},\sqrt{nb_n^{d+2}}\{\hat{\eta}(x)-\eta(x)\}\right) &\rightarrow 0,\\
Cov\left(\sqrt{nb_n^{d+2}}\{\hat{\xi}(x)-\xi(x)\},\sqrt{nb_n^{d+2}}\{\hat{\theta}(x)-\theta(x)\}\right) &\rightarrow 0.
\end{aligned}
$$

Also note that

$$
\begin{aligned}
& Cov\left(\sqrt{nb_n^{d+2}}\{\hat{\zeta}(x)-\zeta(x)\},\sqrt{nb_n^{d+2}}\{\hat{\eta}(x)-\eta(x)\}\right)\\
= & \lim_{n\rightarrow\infty}\frac{n^2}{n_M^2 b_n^d}\left\{\begin{array}{r} E\left[m_M(x,V_i)G_M(x,V_i)G_H(x,V_i)\nabla K\left(\frac{x-X_i}{b_n}\right)^2\right]\\ -E\left[I\{Y_i=C_H\}m_M(x,V_i)G_M(x,V_i)\nabla K\left(\frac{x-X_i}{b_n}\right)^2\right]\\ -E\left[I\{Y_i\neq C_H,C_L\}m_M(x,V_i)G_H(x,V_i)\nabla K\left(\frac{x-X_i}{b_n}\right)^2\right]\end{array}\right\}\\
= & -G_M^{-2}\int_v m_M(x,v)G_M(x,v)G_H(x,v)f(x,v)dv\int_a \nabla K(a)^2 da.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
& Cov\left(\sqrt{nb_n^{d+2}}\{\hat{\zeta}(x)-\zeta(x)\},\sqrt{nb_n^{d+2}}\{\hat{\theta}(x)-\theta(x)\}\right)\\
\rightarrow & -G_M^{-2}\int_v m_M(x,v)G_M(x,v)G_L(x,v)f(x,v)dv\int_a \nabla K(a)^2 da,
\end{aligned}
$$

and

$$
\begin{aligned}
& Cov\left(\sqrt{nb_n^{d+2}}\{\hat{\eta}(x)-\eta(x)\},\sqrt{nb_n^{d+2}}\{\hat{\theta}(x)-\theta(x)\}\right)\\
\rightarrow & -G_M^{-2}\int_v G_L(x,v)G_H(x,v)f(x,v)dv\int_a \nabla K(a)^2 da.
\end{aligned}
$$

Under Assumption 2, the proof is completed by applying a central limit theorem to the linear form of $(\hat{\xi}(x),\hat{\zeta}(x),\hat{\eta}(x),\hat{\theta}(x))$ obtained in Lemmas 3 and 4.

Model 1	$Y = 1 + 0.5X + 0.5XU + U$, 58.5% uncensored				
Value of x	1	2	3	4	5
True Value	0.799	0.752	0.709	0.657	0.601
NPE	0.735	0.678	0.623	0.619	0.666
SD	(0.119)	(0.119)	(0.130)	(0.155)	(0.208)
NPE (Half Bandwidth)	0.781	0.754	0.675	0.634	0.611
SD	(0.280)	(0.316)	(0.367)	(0.446)	(0.554)
No Endogeneity Control	1.086	1.231	0.808	0.553	0.194
SD	(0.170)	(0.251)	(0.304)	(0.341)	(0.454)
No Censoring Control	0.392	0.529	0.509	0.414	0.341
SD	(0.074)	(0.088)	(0.093)	(0.104)	(0.112)
Tobit	1.639	0.925	0.675	0.890	1.554
SD	(0.152)	(0.163)	(0.109)	(0.127)	(0.182)

Model 2	$Y = X + 0.5XU + U$, 60.4% uncensored				
Value of x	1	2	3	4	5
True Value	1.399	1.252	1.154	1.052	0.949
NPE	1.336	1.119	0.986	1.051	1.024
SD	(0.276)	(0.234)	(0.267)	(0.340)	(0.471)
NPE (Half Bandwidth)	1.415	1.264	1.083	1.015	0.892
SD	(0.513)	(0.500)	(0.619)	(0.756)	(1.016)
No Endogeneity Control	1.667	1.695	1.180	0.913	0.522
SD	(0.245)	(0.319)	(0.378)	(0.477)	(0.643)
No Censoring Control	0.496	0.809	0.765	0.611	0.489
SD	(0.102)	(0.114)	(0.118)	(0.124)	(0.142)
Tobit	2.924	1.535	1.081	1.338	2.101
SD	(0.285)	(0.156)	(0.120)	(0.137)	(0.174)

Model 3	$Y = 2 + 1.5XU + U$, 53.8% uncensored				
Value of x	1	2	3	4	5
True Value	0.802	0.725	0.595	0.493	0.417
NPE	0.166	0.516	0.641	0.565	0.622
SD	(0.121)	(0.186)	(0.252)	(0.317)	(0.441)
NPE (Half Bandwidth)	0.259	0.601	0.610	0.504	0.412
SD	(0.309)	(0.472)	(0.689)	(0.968)	(1.234)
No Endogeneity Control	0.850	1.520	1.014	0.779	1.282
SD	(0.180)	(0.288)	(0.382)	(0.500)	(0.680)
No Censoring Control	0.349	0.368	0.282	0.192	0.171
SD	(0.072)	(0.091)	(0.108)	(0.123)	(0.138)
Tobit	0.597	0.830	0.768	0.930	1.873
SD	(0.179)	(0.136)	(0.163)	(0.215)	(0.237)

Model 4	$Y = 1.5 + X + U$, 79.5% uncensored				
Value of x	1	2	3	4	5
True Value	1	1	1	1	1
NPE	1.024	0.984	0.939	0.721	0.448
SD	(0.126)	(0.134)	(0.168)	(0.202)	(0.360)
NPE (Half Bandwidth)	1.091	0.990	1.016	0.866	0.477
SD	(0.319)	(0.365)	(0.462)	(0.589)	(1.028)
No Endogeneity Control	1.221	1.442	1.033	0.286	-1.047
SD	(0.123)	(0.309)	(0.360)	(0.442)	(0.696)
No Censoring Control	0.738	0.936	0.931	0.683	0.312
SD	(0.069)	(0.070)	(0.070)	(0.063)	(0.086)
Tobit	1.350	1.115	1.039	1.116	1.352
SD	(0.052)	(0.050)	(0.035)	(0.050)	(0.053)

CHAPTER 2

Specification Testing for Errors-in-Variables Models

2.1. Introduction

As is the case with most decisions, the choice to employ nonparametric techniques over parametric ones is not always obvious, and making the wrong decision can be costly. If we are able to confirm that a parametric model is correctly specified, we can gain considerably by using parametric estimators. Meanwhile, if we are not fully convinced of this, we should appeal to nonparametric estimation. A popular solution to this problem of estimation choice involves comparing the distance between some parametric and nonparametric estimator; this has been studied in detail by Härdle and Mammen (1993). Other tests for the suitability of parametric models have been studied by Azzalini, Bowman and Härdle (1989), Eubank and Spiegelman (1990), Horowitz and Spokoiny (2001), and Fan and Huang (2001) among many others.

Measurement error is a problem that is rife in datasets from many disciplines. Examples from biology, economics, geography, medicine, and physics are abundant (see, for example, Fuller, 1987, and Meister, 2009). Determining the validity of a parametric model becomes even more important in the presence of measurement error because in this setting nonparametric estimators have even slower convergence properties whilst in many cases parametric estimators retain their $\sqrt{n}$-consistency. However, when the data are contaminated by measurement error, conventional specification tests have, in general, incorrect size and may also suffer from low power properties.

In this chapter, we propose a specification, or goodness-of-fit test, for (possibly nonlinear) regression models with errors-in-variables by comparing the distance between the parametric and nonparametric fits based on deconvolution techniques. We establish asymptotic properties of the test statistic and propose a bootstrap critical value. As we discuss below, in contrast to existing methods, our test allows nonlinear regression models and possesses desirable power properties.

In the enormous literature on specification testing, relatively little attention has been given to the issue of measurement error despite its obvious need. Papers such as Zhu, Song and

Cui (2003), Zhu and Cui (2005), and Cheng and Kukush (2004) propose χ^2 statistics based on moment conditions of observables implied from errors-in-variables regression models. However, as is the case without measurement error, these tests are generally inconsistent for some fixed alternatives. Song (2008) proposes a consistent specification test for linear errors-in-variables regression models by comparing nonparametric and model-based estimators on the conditional mean function of the dependent variable Y given the *mismeasured* observable covariates W, that is $E[Y|W]$. As we clarify at the end of Section 2.2, this approach may not have sensible local power for the original hypothesis on $E[Y|X]$, where X is a vector of error-free unobservable covariates. Hall and Ma (2007) propose a nonsmoothing specification test for regression models with errors-in-variables which is able to detect local alternatives at the $\sqrt{n}$-rate. We propose a smoothing specification test that complements Hall and Ma's (2007) nonsmoothing approach (see further discussion below).[1]

Consistent specification tests can be broadly split into those that use a nonparametric estimator (called smoothing tests) and those that do not (called nonsmoothing or integral-transform tests). In contrast to Hall and Ma (2007) who adopt the nonsmoothing approach, we propose a kernel-based smoothing test for the goodness-of-fit of parametric regression models with errors-in-variables. There are two important features of our test. First, our smoothing test is not restricted to polynomial models; allowing testing of general nonlinear regression models. Second, analogous to the literature on conventional specification testing, our smoothing test complements Hall and Ma's (2007) test (if applied to polynomial models) due to its distinct power properties. Rosenblatt (1975) explains that although local power properties of nonsmoothing tests suggest they are more powerful than smoothing tests, 'there are other types of local alternatives for which tests based on density estimates are more powerful'. Fan and Li (2000) show that in the non-measurement error case, smoothing tests are generally more powerful for high frequency alternatives and nonsmoothing tests are more powerful for low frequency alternatives. Thus,

[1] Other papers that study specification testing under measurement error includes Butucea (2007), Holzmann and Boysen (2006), Holzmann, Bissantz and Munk (2007), and Ma *et al.* (2011) (for testing probability densities), Koul and Song (2009, 2010) (for Berkson measurement error models), and Song (2009) and Xu and Zhu (2015) (for errors-in-variables models with validation data).

smoothing tests 'should be viewed as complements to, rather than substitutes for, [nonsmoothing tests].' Our simulation results suggest that this phenomenon extends to errors-in-variables models.

In contrast to the above papers and our own, Ma *et al.* (2011) moves away from Wald-type tests where restricted and unrestricted estimates are compared. They propose a local test that is more analogous to the score test where only the model under the null hypothesis must be estimated. They extend this idea to an omnibus test that is able to detect departures from the null in virtually all directions using a system of different basis functions with which to test against.

To determine critical values for our smoothing test, we propose a bootstrap procedure. Measurement error can cause difficulties in applying conventional bootstrap procedures because the true regressor, regression error, and measurement error are all unobserved. Moreover, in order to estimate the distributions of test statistics, deconvolution techniques are typically required which converge at a much slower rate than $\sqrt{n}$. Hall and Ma (2007) discuss this issue and note, 'the bootstrap is seldom used in the context of errors-in-variables'. They outline a procedure which involves estimating the distribution of the unobservable regressor using a kernel deconvolution estimator, and obtain bootstrap counterparts for the regression error using a wild bootstrap method. We propose a much simpler procedure involving a perturbation of each summand of our test statistic.

This chapter is organized as follows. Section 2.2 describes the setup in detail and introduces the test statistic and its motivation. Section 2.3 outlines the main asymptotic properties of the test statistic and discusses how to implement the test in the case where the distribution of the measurement error is unknown but repeated measurements on the contaminated covariates are available. Section 2.4 analyses the small sample properties of the test through a Monte Carlo experiment and Section 2.5 applies the test to the specification of Engel curves. All mathematical proofs are deferred to Appendix 2.6.

2.2. Setup and Test Statistic

Consider the nonparametric regression model

$$Y = m(X) + U \quad \text{with } E[U|X] = 0,$$

where $Y \in \mathbb{R}$ is a response variable, $X \in \mathbb{R}^d$ is a vector of covariates, and $U \in \mathbb{R}$ is the error term. In this chapter, we focus on the situation where X is not directly observable due to the measurement mechanism or nature of the environment. Instead a vector of variables W is observed through

$$W = X + \epsilon,$$

where $\epsilon \in \mathbb{R}^d$ is a vector of measurement errors that has a known density $f_\epsilon(\cdot)$ and is independent of (Y, X). The case of unknown density $f_\epsilon(\cdot)$ will be discussed in Section 2.3.1. We are interested in specification, or goodness-of-fit, testing of a parametric functional form of the regression function $m(\cdot)$. More precisely, for a parametric model $m_\theta(\cdot)$, we wish to test the hypothesis

$$\begin{aligned} \mathrm{H}_0 &: \quad m(x) = m_\theta(x) \text{ for almost every } x \in \mathbb{R}^d, \\ \mathrm{H}_1 &: \quad \mathrm{H}_0 \text{ is false,} \end{aligned}$$

based on the random sample $\{Y_i, W_i\}_{i=1}^n$ of observables (whilst X_i is unobservable).

To test the null H_0, we adapt the approach of Härdle and Mammen (1993), which compares nonparametric and parametric regression fits, to the errors-in-variables model. As a nonparametric estimator of $m(\cdot)$, we use the deconvolution kernel estimator (see, for example, Fan and Truong, 1993, and Meister, 2009, for a review)

$$\hat{m}(x) = \frac{\sum_{i=1}^n Y_i \mathcal{K}_b(x - W_i)}{\sum_{i=1}^n \mathcal{K}_b(x - W_i)},$$

where

$$\mathcal{K}_b(a) = \frac{1}{(2\pi)^d} \int e^{-\mathrm{i}t \cdot a} \frac{K^{\mathrm{ft}}(tb)}{f_\epsilon^{\mathrm{ft}}(t)} dt,$$

is the so-called deconvolution kernel, $\mathrm{i} = \sqrt{-1}$, b is a bandwidth, and $K^{\mathrm{ft}}(\cdot)$ and $f_\epsilon^{\mathrm{ft}}(\cdot)$ are the Fourier transforms of a kernel function $K(\cdot)$ and the measurement error density $f_\epsilon(\cdot)$, respectively. [2] $K^{\mathrm{ft}}(\cdot)$ acts as a regularisation factor to deal with the ill-posed inverse problem that measurement error causes. Note that throughout this thesis, we denote the Fourier transform of any function, $h(x)$, as $h^{\mathrm{ft}}(t) = \int e^{\mathrm{i}t\cdot x} h(x)dx$. Also, throughout this chapter we assume $f_\epsilon^{\mathrm{ft}}(t) \neq 0$ for all $t \in \mathbb{R}^d$ and $K^{\mathrm{ft}}(\cdot)$ has compact support so that the deconvolution kernel is well-defined. On the other hand, if one imposes a parametric functional form $m_\theta(\cdot)$ on the regression function, several methods are available to estimate θ under certain regularity conditions. For example, based on Butucea and Taupin (2008), we can estimate the parameter θ by the (weighted) least squares regression of Y on the implied conditional mean function $E[m_\theta(X)|W]$. In this chapter, we do not specify the construction of the estimator $\hat{\theta}$ for θ except for a mild assumption on the convergence rate (see Section 2.3 for details).

In order to construct a test statistic for H_0, as in Härdle and Mammen (1993), we compare the nonparametric and parametric estimators of the regression function based on the L_2-distance,

$$D_n = n \int \left| \hat{m}(x)\hat{f}(x) - [K_b * m_{\hat{\theta}}\hat{f}](x) \right|^2 dx,$$

where $|\cdot|$ is the Euclidean norm, $\hat{f}(x) = \frac{1}{n}\sum_{i=1}^n \mathcal{K}_b(x - W_i)$ is the deconvolution kernel density estimator for X, $K_b(x) = \frac{1}{b^d}K\left(\frac{x}{b}\right)$, and $[K_b * m_{\hat{\theta}}\hat{f}](x) = \int K_b(x-a)m_{\hat{\theta}}(a)\hat{f}(a)da$ is a convolution. The convolution by the (original) kernel function $K_b(\cdot)$ plays an analogous role to the smoothing operator in Härdle and Mammen (1993) and removes the bias term from the nonparametric estimator. Note that the Fourier transform of a convolution is given by the product of the Fourier transforms. Thus by Parseval's identity, the distance D_n is alternatively written as

$$D_n = \frac{n}{(2\pi)^d} \int \frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2} \left| \frac{1}{n}\sum_{i=1}^n Y_i e^{\mathrm{i}t\cdot W_i} - [m_{\hat{\theta}}\hat{f}]^{\mathrm{ft}}(t) f_\epsilon^{\mathrm{ft}}(t) \right|^2 dt.$$

Based on this expression, the distance D_n can be interpreted as a contrast of the nonparametric and model-based estimators for $E[Ye^{\mathrm{i}t\cdot W}]$. To define the test statistic for H_0, we further

[2]To simplify the exposition, we concentrate on the case where all elements of X are mismeasured. If X contains both correctly measured and mismeasured covariates (denoted by X_1 and X_2, respectively), then the kernel estimator is modified as $\hat{m}(x) = \frac{\sum_{i=1}^n Y_i K_{1b}(x_1 - X_{1i})\mathcal{K}_b(x_2 - W_i)}{\sum_{i=1}^n K_{1b}(x_1 - X_{1i})\mathcal{K}_b(x_2 - W_i)}$, where $K_{1b}(a) = \frac{1}{b^{d_1}}K_1\left(\frac{a}{b}\right)$ and $K_1(\cdot)$ is a conventional kernel function for X_1, and analogous results can be established.

decompose D_n as

$$\begin{aligned} D_n &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{(2\pi)^d}\int \frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}|\zeta_i(t)|^2 dt + \frac{1}{n}\sum_{i\neq j}\frac{1}{(2\pi)^d}\int \frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\zeta_i(t)\overline{\zeta_j(t)}dt \\ &\equiv B_n + T_n, \end{aligned} \tag{9}$$

where $\overline{\zeta_i(t)}$ is the complex conjugate of $\zeta_i(t) \equiv Y_i e^{\mathrm{i}t\cdot W_i} - \int e^{\mathrm{i}s\cdot W_i} m_{\hat{\theta}}^{\mathrm{ft}}(t-s)\frac{K^{\mathrm{ft}}(sb)}{f_\epsilon^{\mathrm{ft}}(s)}ds f_\epsilon^{\mathrm{ft}}(t)$. The second term T_n plays a dominant role in the limiting behavior of D_n and the first term B_n is considered a bias term. That is to say, following similar arguments to those presented in Appendix 2.6, the appropriately normalised B_n is asymptotically negligible. Therefore, we neglect B_n and employ T_n as our test statistic for H_0. In the next section, we study the asymptotic behaviour of T_n.

We close this section by a remark on an alternative testing approach. To test the null hypothesis H_0, one may consider testing some implication of H_0 on the conditional mean $E[Y|W]$ of observables, i.e., consider $H_0' : f_W(w)E[Y\big|W = w] = \int m_\theta(w-u)f_X(w-u)f_\epsilon(u)du$ for almost every w, and test H_0' by a conventional method, such as Härdle and Mammen (1993). This approach was employed by Song (2008). To clarify the rationale of our testing approach based on T_n against the conventional approach for H_0', consider the following local alternative hypothesis for the regression function

$$m_n(x) = m_\theta(x) + 2a_n\cos(A_n x)\left(\frac{\sin x}{x}\right),$$

where $a_n \to 0$ and $A_n \to \infty$ as $n \to \infty$. In this case, $m_n(\cdot)$ converges to $m_\theta(\cdot)$ at the rate of a_n under the L_2-norm, and the test based on T_n will have non-trivial power for a certain rate of a_n. On the other hand, local power of the test based on the implied null H_0' is determined by the L_2-norm of the convolution $\{(m_n - m_\theta)f_X\} * f_\epsilon$. By Parseval's identity and the Fourier shift formula, we have

$$\|\{(m_n - m_\theta)f_X\} * f_\epsilon\|^2 = a_n^2\left\|\{q^{\mathrm{ft}}(\cdot - A_n) + q^{\mathrm{ft}}(\cdot + A_n)\}f_\epsilon^{\mathrm{ft}}\right\|^2,$$

where $q(x) = \left(\frac{\sin x}{x}\right)f_X(x)$. For example, if $f_\epsilon(\cdot)$ is Laplace with $f_\epsilon^{\mathrm{ft}}(t) = 1/(1+t^2)$, then we can see that the L_2-norm $\|\{(m_n - m_\theta)f_X\} * f_\epsilon\|$ is of order a_n/A_n^2. By letting A_n diverge at an arbitrarily fast rate, the rate a_n/A_n^2 becomes arbitrarily fast so that any conventional test for H_0' fails to detect deviations from this null. Therefore, as far as the researcher is concerned with testing the functional form of the regression function $m(\cdot)$, we argue that our statistic T_n tests directly the null hypothesis H_0 and possesses desirable local power properties compared to the conventional tests on H_0'.

2.3. Asymptotic Properties

In this section, we present asymptotic properties of the test statistic T_n defined in (9). We first derive the limiting distribution of T_n under the null hypothesis H_0. To this end, we impose the following assumptions.

Assumption 3.

- **(i):** $\{Y_i, X_i, \epsilon_i\}_{i=1}^n$ *are i.i.d.* ϵ *is independent of* (Y, X) *and has a known density* $f_\epsilon(\cdot)$.
- **(ii):** $f^{\mathrm{ft}}, m^{\mathrm{ft}}, \frac{\partial}{\partial\theta}(m_\theta^{\mathrm{ft}}) \in L_1(\mathbb{R}^d) \cap L_2(\mathbb{R}^d)$.
- **(iii):** $K^{\mathrm{ft}}(t)$ *is compactly supported on* $[-1, 1]^d$*, is symmetric around zero (i.e.,* $K^{\mathrm{ft}}(t) = K^{\mathrm{ft}}(-t)$*), and is bounded.*
- **(iv):** *As* $n \to \infty$*, it holds that* $b \to 0$ *and* $nb^d \to \infty$.

Assumption 3 (i) is common in the literature of classical measurement error. Extensions to the case of unknown $f_\epsilon(\cdot)$ will be discussed in Section 2.3.1. Assumption 3 (ii) contains boundedness conditions on the Fourier transforms of the density $f(\cdot)$ of X and the regression function $m(\cdot)$, as well as the derivative, with respect to θ, of the Fourier transform of $m_\theta(\cdot)$. Assumption 3 (iii) and (iv) contain standard conditions on the kernel function $K(\cdot)$ and bandwidth b, respectively. A popular choice for the kernel function in the context of deconvolution methods is the sinc kernel $K(x) = \frac{\sin x}{\pi x}$ whose Fourier transform is equal to $K^{\mathrm{ft}}(t) = \mathbb{I}\{-1 \leq t \leq 1\}$.

For additional assumptions, we consider two cases characterised by bounds on the decay rate of the tail of the characteristic function of the measurement error, $f_\epsilon^{\mathrm{ft}}(\cdot)$. Let $\sigma^2(x) = E[U^2|X = x]$ be the conditional variance of the error term. The first case, known as ordinary smooth measurement error (or, in the statistics literature, the mildly ill-posed case), contains the following assumptions.

ASSUMPTION 4.

(i): $f_\epsilon^{\mathrm{ft}}(t) \neq 0$ *for all* $t \in \mathbb{R}^d$ *and there exist positive constants* c, C, *and* α *such that*

$$c|t|^{-d\alpha} \leq |f_\epsilon^{\mathrm{ft}}(t)| \leq C|t|^{-d\alpha},$$

as $|t| \to \infty$.

(ii): $\int |[mf]^{\mathrm{ft}}(t)|^2 dt < \infty$, $\int |[m^2 f]^{\mathrm{ft}}(t)|^2 dt < \infty$, *and* $\int |[\sigma^2 f]^{\mathrm{ft}}(t)|^2 dt < \infty$.

(iii): $\hat{\theta} - \theta = o_p(n^{-1/2} b^{-d(\frac{1}{4}+\alpha)})$ *under* H_0.

Assumption 4 (i) requires that the Fourier transform $f_\epsilon^{\mathrm{ft}}(\cdot)$ decays in some finite power. A popular example of an ordinary smooth density is the Laplace density. Assumption 4 (ii) contains boundedness conditions on the Fourier transforms of the density $f(\cdot)$ of X, regression function $m(\cdot)$, and conditional error variance $\sigma^2(\cdot)$. Assumption 4 (iii) is on the convergence rate of the estimator $\hat{\theta}$ for θ when the parametric model is correctly specified. Note that this assumption is satisfied if $\hat{\theta}$ is $\sqrt{n}$-consistent for θ. When the regression model under the null hypothesis is linear (i.e., $m_\theta(x) = x'\theta$), we can employ the methods in, for example Gleser (1981), Bickel and Ritov (1987), or van der Vaart (1988). For nonlinear regression, we may choose the estimators by, for example, Taupin (2001) or Butucea and Taupin (2008) under certain regularity conditions. It is interesting to note that in contrast to the no measurement error case as in Härdle and Mammen (1993), the limiting distribution of the estimation error $\sqrt{n}(\hat{\theta} - \theta)$ does not influence the first-order asymptotic properties of the test statistic T_n. This is because the measurement error slows down the convergence rate of the dominant term of T_n.

For the second case, known as supersmooth measurement error (or, in the statistics literature, the severely ill-posed case), we concentrate on the case of scalar X (i.e., $d = 1$), and impose the following assumptions.

Assumption 5. *Suppose $d = 1$.*

(i): $f_\epsilon^{\mathrm{ft}}(t) \neq 0$ *for all* $t \in \mathbb{R}$ *and there exist positive constants* C_ϵ, μ, γ_0, *and* $\gamma > 1$ *such that*

$$f_\epsilon^{\mathrm{ft}}(t) \sim C_\epsilon |t|^{\gamma_0} e^{-|t|^\gamma/\mu},$$

as $|t| \to \infty$. *Also, there exist constants* $A > 0$ *and* $\beta \geq 0$ *such that*

$$K^{\mathrm{ft}}(1-t) = At^\beta + o(t^\beta),$$

as $t \to 0$.

(ii): $E[Y^4] < \infty$, $E[W^4] < \infty$, $\int |t|^{2\beta} \left|\frac{\partial}{\partial\theta} m_\theta^{\mathrm{ft}}(t)\right|^2 dt < \infty$, *and* $\int |t|^{2\beta} |m^{\mathrm{ft}}(t)|^2 dt < \infty$.

(iii): $\hat{\theta} - \theta = o_p(n^{-1/2} b^{(\gamma-1)/2+\gamma\beta+\gamma_0} e^{1/(\mu b^\gamma)})$.

Assumption 5 (i) is adopted from Holzmann and Boysen (2006). This assumption requires that the Fourier transform $f_\epsilon^{\mathrm{ft}}(\cdot)$ decays at an exponential rate. An example of a supersmooth density satisfying this assumption is the normal density, where $C_\epsilon = 1$, $\gamma_0 = 0$, $\gamma = 2$, and $\mu = 2$. However, due to the requirement $\gamma > 1$, the Cauchy density is excluded. As is clarified in the proof of Theorem 3 (iii) below, the condition $\gamma > 1$ is imposed to make a bias term negligible. Assumption 5 (i) also contains an additional condition on the kernel function. For example, the sinc kernel $K(x) = \frac{\sin x}{\pi x}$ satisfies this assumption with $A = 1$ and $\beta = 0$. Similarly to the ordinary smooth case, Assumption 5 (ii) contains boundedness conditions on some Fourier transforms, and Assumption 5 (iii) regards the convergence rate of the estimator $\hat{\theta}$. Again, the $\sqrt{n}$-consistency of $\hat{\theta}$ is sufficient.

Under these assumptions, the null distribution of T_n is obtained as follows.

Theorem 3.

(i): *Suppose that Assumptions 3 and 4 hold true. Then under* H_0*,*

$$C_{V,b}^{-1/2} T_n \xrightarrow{d} N\left(0, \frac{2}{(2\pi)^{2d}}\right),$$

where $C_{V,b} = O(b^{-d(1+4\alpha)})$ *is defined in (11) in Appendix 2.6.*

(ii): *Suppose that Assumptions 3 and 5 hold true with* $d = 1$ *and* $\epsilon \sim N(0,1)$*. Then under* H_0*,*

$$\varphi(b) T_n \xrightarrow{d} \sum_{k=1}^{\infty} \lambda_k (Z_k^2 - 1),$$

where $\varphi(b) = \frac{(2\pi)2^{2\beta}}{b^{1+4\beta} e^{1/b^2} A^2 \Gamma(1+2\beta)}$ *with the gamma function* Γ*,* $\{Z_k\}$ *is an independent sequence of standard normal random variables and* $\{\lambda_k\}$ *is defined in (21) in Appendix 2.6.*

(iii): *Suppose that Assumptions 3 and 5 hold true with* $d = 1$*. Then under* H_0*,*

$$\varphi(b) T_n \xrightarrow{d} \sum_{k=1}^{\infty} \lambda_k (Z_k^2 - 1),$$

where $\varphi(b) = \frac{(2\pi)2^{2\beta}\gamma^{1+2\beta}C_\epsilon^2}{\mu^{1+2\beta} b^{\gamma-1+2\gamma\beta+2\gamma_0} e^{2/(\mu b^\gamma)} A^2 \Gamma(1+2\beta)}$ *with the gamma function* Γ*,* $\{Z_k\}$ *is an independent sequence of standard normal random variables and* $\{\lambda_k\}$ *is defined in (21) in Appendix 2.6.*

Theorem 3 (i) says that for the ordinary smooth case, the test statistic T_n is asymptotically normal. The normalising term $C_{V,b}$ comes from the variance of the U-statistic of the leading term in T_n. Note that the convergence rate $C_{V,b}^{-1/2} = O\left(b^{d(\frac{1}{2}+2\alpha)}\right)$ of the statistic T_n is slower than the rate $O(b^{d/2})$ of Härdle and Mammen's (1993) statistic for the no measurement error case. As the dimension d of X or the decay rate α of $f_\epsilon^{\mathrm{ft}}(\cdot)$ increases, the convergence rate of T_n becomes slower.

Theorem 3 (ii) focuses on the case of normal measurement error, and shows that the test statistic converges to a weighted sum of chi-squared random variables. The normalising term $\varphi(b)$ is characterised by the shape of the kernel function specified in Assumption 5 (i). For example, if we employ the sinc kernel (i.e., $A = 1$ and $\beta = 0$), the normalisation becomes $\varphi(b) = \frac{2\pi}{be^{1/b^2}\Gamma(1)}$. In this supersmooth case, the non-normal limiting distribution emerges because the leading term of the statistic T_n is characterised by a degenerate U-statistic with a fixed kernel (see, for example, Serfling, 1980, Theorem 5.5.2). In contrast, for the ordinary smooth case in Part (i) of this theorem, the leading term is characterised by a U-statistic with a varying kernel so that the central limit theorem in Hall (1984) applies. An analogous result is obtained in Holzmann and Boysen (2006) for the integrated squared error of the deconvolution density estimator.

Theorem 3 (iii) presents the limiting null distribution of the test statistic for the case of general supersmooth measurement error. In this case, after normalisation by $\phi(b)$, the test statistic obeys the same limiting distribution as the normal case in Part (ii) of this theorem. Thus, similar comments to Part (ii) apply. The normalisation term $\phi(b)$ is characterised by the shapes of the kernel function and the Fourier transform $f_\epsilon^{\mathrm{ft}}(t)$ of the measurement error specified in Assumption 5 (i).

Although Theorem 3 (ii) and (iii) focus on the case of scalar ϵ, our technical argument may be extended to the vector case. For example, if we assume that the elements of the d-dimensional vector ϵ are mutually independent, then the Fourier transform $f_\epsilon(\cdot)$ becomes the product of the Fourier transforms of the marginals. We may impose Assumption 5 (i) for each marginal density. To keep things simple we can choose the multivariate kernel function to be a product kernel. With these assumptions in place, the deconvolution kernel analogously becomes a product deconvolution kernel. The proofs of the theorem remain very similar using inner products and terms defined as products over the d dimensions.

Theorem 3 can be applied to obtain critical values for testing the null H_0 based on T_n. Alternatively, we can compute the critical values by bootstrap methods. A bootstrap counterpart of T_n is given by perturbing each summand in T_n as follows

$$T_n^* = \frac{1}{n} \sum_{i \neq j} \nu_i^* \nu_j^* \frac{1}{(2\pi)^d} \int \frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2} \zeta_i(t) \overline{\zeta_j(t)} dt, \tag{10}$$

where $\{\nu_i^*\}_{i=1}^n$ is an i.i.d. sequence which is mean zero with unit variance and independent of $\{Y_i, W_i\}_{i=1}^n$. The asymptotic validity of this bootstrap procedure follows by a similar argument to Delgado, Dominguez and Lavergne (2006, Theorem 6).

In order to investigate the power properties of the test based on T_n, we consider a local alternative hypothesis of the form

$$\mathrm{H}_{1n} : m(x) = m_\theta(x) + c_n \Delta(x), \text{ for almost every } x \in \mathbb{R}^d$$

where $c_n \to 0$ and $\Delta(x)$ is a non-zero function such that the limits $\lim_{n\to\infty} \Delta_n$ and $\lim_{n\to\infty} \Upsilon_n$ defined in (23) and (24), respectively, in Appendix 2.6 exist. The local power properties are obtained as follows.

Theorem 4.

(i): *Suppose that Assumptions 3 and 4 hold true. Then under* H_{1n} *with* $c_n = n^{-1/2} b^{-d(\frac{1}{4}+\alpha)}$,

$$C_{V,b}^{-1/2} T_n \xrightarrow{d} N\left(\lim_{n\to\infty} \Delta_n, \frac{2}{(2\pi)^{2d}} \right).$$

(ii): *Suppose that Assumptions 3 and 5 hold true with* $d = 1$ *and* $\epsilon \sim N(0,1)$. *Then under* H_{1n} *with* $c_n = n^{-1/2} b^{1/2+2\beta} e^{1/(2b^2)}$,

$$\varphi(b) T_n \xrightarrow{d} \lim_{n\to\infty} \Upsilon_n + \sum_{k=1}^{\infty} \lambda_k (Z_k^2 - 1).$$

(iii): *Suppose that Assumptions 3 and 5 hold true with* $d = 1$. *Then under* H_{1n} *with* $c_n = b^{(\lambda-1)/2+\lambda\beta+\lambda_0} e^{1/(\mu b^\lambda)}$,

$$\varphi(b) T_n \xrightarrow{d} \lim_{n\to\infty} \Upsilon_n + \sum_{k=1}^{\infty} \lambda_k (Z_k^2 - 1),$$

Theorem 4 (i) says that under the ordinary smooth case, our test has non-trivial power against local alternatives drifting with the rate of $c_n = n^{-1/2} b^{-d(\frac{1}{4}+\alpha)}$. This is a nonparametric rate; the test based on T_n becomes less powerful as the dimension d of X or the decay rate α of $f_\epsilon^{\mathrm{ft}}(\cdot)$ increases. For the non-measurement error case, Härdle and Mammen's (1993) statistic has non-trivial power for local alternatives with the rate of $n^{-1/2} b^{-d/4}$. Therefore, the test is less powerful as a result of the measurement error. Theorem 4 (ii) and (iii) present local power properties of our test for the normal and general supersmooth measurement error cases, respectively. Except for the normalising constants, the test statistic has the same limiting distribution. Also, for $c_n \to 0$, the bandwidth b should decay at a log rate. As an example, consider the case of

$\epsilon \sim N(0,1)$. In this case, if we choose $b \sim (\log n)^{-1/2}$, then the rate for the local alternative will be $c_n \sim (\log n)^{-1/4-\beta}$. Therefore, for the supersmooth case, the rate for the local alternative is typically a log rate.

2.3.1. Case of Unknown $f_\epsilon(\cdot)$. In practical applications, it is sometimes unrealistic to assume that the density of the measurement error, $f_\epsilon(\cdot)$, is known to the researcher. In the literature on nonparametric deconvolution several estimation methods for $f_\epsilon(\cdot)$ are available, these are typically based on additional data (see, for example, Section 2.6 of Meister (2009) for a review). Although the analysis of the asymptotic properties is different, we can modify the test statistic T_n by inserting the estimated Fourier transform of the measurement error density, $\hat{f}_\epsilon^{\text{ft}}(\cdot)$.

For example, suppose the researcher has access to repeated measurements on X in the form of $W = X + \epsilon$ and $W^r = X + \epsilon^r$, where ϵ and ϵ^r are identically distributed and $(X, \epsilon, \epsilon^r)$ are mutually independent, see Delaigle, Hall and Meister (2008) for a list of examples of such repeated measurements. If we further assume that the Fourier transform $f_\epsilon^{\text{ft}}(\cdot)$ is real-valued (that is the density $f_\epsilon(\cdot)$ is symmetric around zero), then we can employ the estimator proposed by Delaigle, Hall and Meister (2008)

$$\hat{f}_\epsilon^{\text{ft}}(t) = \left| \frac{1}{n} \sum_{i=1}^{n} \cos\{t(W_i - W_i^r)\} \right|^{1/2}.$$

Delaigle, Hall and Meister (2008) studied the asymptotic properties of the deconvolution density and regression estimators using $\hat{f}_\epsilon^{\text{ft}}(\cdot)$ and found conditions to guarantee that the difference between the estimators with known $f_\epsilon(\cdot)$ and those with unknown $f_\epsilon(\cdot)$ are asymptotically negligible. Under similar conditions, we can expect that the asymptotic distributions of the test statistic T_n obtained above remain unchanged when we replace $f_\epsilon^{\text{ft}}(\cdot)$ with $\hat{f}_\epsilon^{\text{ft}}(\cdot)$. If the researcher wishes to remove the assumption that $f_\epsilon^{\text{ft}}(\cdot)$ is real-valued, it may be possible to employ the estimator by Li and Vuong (1998) based on Kotlarski's identity.

2.4. Simulation

We evaluate the small sample performance of our test through a Monte Carlo experiment. To begin, we consider the same data generating process as Hall and Ma (2007) for ease of comparison. We also compare our test to Song (2008). Recall that although Song's (2008) and Hall and Ma's (2007) tests are confined to polynomial regression models, our test allows nonlinear models. Specifically we take the true unobservable regressor $\{X_i\}_{i=1}^n$ to be distributed as uniformly on $[-3, 4]$ and $Y_i = 1 + 1.5X_i + C\cos(X_i) + U_i$, where $U_i \sim N(0, 1)$ and C is a constant to be varied. The contaminated regressor is given by $W_i = X_i + \epsilon_i$. We consider two distributions for ϵ_i to be drawn from. For the ordinary smooth case, we use the Laplace distribution with variance of 0.5. For the supersmooth case, we use $N(0, 1)$. We use the following kernel for our simulations (Fan, 1992)

$$K(x) = \frac{48\cos(x)}{\pi x^4}\left(1 - \frac{15}{x^2}\right) - \frac{144\sin(x)}{\pi x^5}\left(2 - \frac{5}{x^2}\right).$$

We report results for a range of sample sizes, bandwidths, and nominal levels of the test. Specifically, for the ordinary and supersmooth cases, we choose the bandwidths according to the rules of thumb $b = c\left(\frac{5\sigma^4}{n}\right)^{1/9}$ and $b = c\left(\frac{2\sigma^2}{\log(n)}\right)^{1/2}$, respectively, where σ is the standard deviation of the measurement error and c varies in the grid $\{0.01, 0.05, 0.1, 0.5, 1, 1.5\}$ so we can analyse the sensitivity of our test to the bandwidth. For the parametric estimator we use the polynomial estimator of degree 2 proposed by Cheng and Schneeweiss (1998) so as to remain consistent with the experiment conducted by Hall and Ma (2007). For the test of Song (2008) we use the same kernel as for our test and choose bandwidths by cross-validation (there was little dependence on the bandwidths so we report only for these cross-validated values). All results are based on 1000 Monte Carlo replications.

Table 1 takes $C = 0$ so as to asses the level accuracy of our test. To study the power properties of the test, we take $C = 1.5$ in Table 2. The critical values for all tests are based on 99 replications of the bootstrap procedure (results were very similar for 199 replications and hence are not reported). The perturbation random variable ν^* for the bootstrap is drawn from the Rademacher distribution.

Finally, to highlight the power advantages of our test under high frequency alternatives we consider the slightly altered data generating process $Y_i = 1 + 1.5X_i + \cos(\pi\delta X_i) + U_i$, where δ is a constant to be varied; larger values corresponding to higher frequency alternatives. All other parameter settings remain unchanged. Results for these experiments are shown in Tables 3-5. The columns labeled 'HM' correspond to the power of the test proposed in Hall and Ma (2007) and the columns labeled 'S' correspond to the power of the test proposed by Song (2008).

Table 1: $Y = 1 + 1.5X + U$

Ordinary Smooth		**Bandwidth**					
n	**Level**	**0.01**	**0.05**	**0.1**	**0.5**	**1**	**1.5**
50	**1%**	3.0%	2.6%	2.3%	3.1%	2.4%	0.6%
	5%	7.6%	7.1%	6.3%	6.5%	6.4%	2.4%
	10%	12.7%	11.9%	10.7%	9.8%	10.1%	5.7%
100	**1%**	2.2%	2.0%	2.8%	2.3%	2.2%	0.6%
	5%	4.9%	5.6%	6.8%	6.3%	6.9%	3.4%
	10%	11.6%	10.5%	11.9%	12.4%	11.1%	7.3%
Super Smooth							
50	**1%**	2.1%	1.9%	1.3%	1.4%	1.9%	1.2%
	5%	5.3%	5.2%	4.9%	5.4%	5.7%	4.3%
	10%	11.1%	9.3%	9.6%	10.8%	10.4%	7.7%
100	**1%**	2.9%	2.4%	2.7%	2.0%	1.7%	1.9%
	5%	6.9%	6.5%	6.0%	6.6%	5.3%	5.8%
	10%	12.3%	10.7%	10.3%	10.6%	10.8%	10.5%

Table 2: $Y = 1 + 1.5X + 1.5\cos(X) + U$

Ordinary Smooth		**Bandwidth**						HM	S
n	**Level**	**0.01**	**0.05**	**0.1**	**0.5**	**1**	**1.5**		
50	**1%**	43.2%	33.7%	26.5%	40.6%	81.8%	63.3%	71.0%	86.1%
	5%	63.1%	52.7%	45.5%	61.6%	92.1%	78.3%	76.9%	92.8%
	10%	73.5%	63.1%	57.8%	72.2%	94.8%	86.2%	85.3%	95.3%
100	**1%**	67.4%	50.1%	38.5%	66.5%	99.3%	97.6%	95.8%	88.6%
	5%	84.2%	71.3%	61.4%	85.1%	99.8%	99.2%	97.1%	99.1%
	10%	91.0%	80.1%	73.0%	92.3%	99.9%	99.9%	99.1%	99.9%
Super Smooth									
50	**1%**	28.1%	22.3%	19.7%	12.2%	24.5%	36.1%	66.2%	67.2%
	5%	46.6%	38.7%	33.5%	24.1%	44.5%	55.2%	72.0%	84.2%
	10%	57.4%	48.4%	43.2%	34.9%	56.9%	67.3%	80.4%	88.6%
100	**1%**	54.0%	35.9%	23.5%	15.4%	47.2%	69.0%	94.3%	89.3%
	5%	70.7%	52.5%	41.5%	28.8%	63.8%	84.6%	95.9%	94.5%
	10%	78.1%	64.5%	53.9%	42.1%	73.7%	89.8%	97.7|%	97.8%

Table 3: $Y = 1 + 1.5X + \cos(\pi X) + U$

Ordinary Smooth		**Bandwidth**						HM	S
n	**Level**	**0.01**	**0.05**	**0.1**	**0.5**	**1**	**1.5**		
100	**1%**	21.3%	15.9%	13.7%	12.9%	6.6%	2.4%	13.2%	8.9%
	5%	40.3%	30.3%	26.5%	27.5%	17.6%	6.8%	27.5%	19.6%
	10%	51.4%	45.0%	38.2%	38.0%	26.5%	13.6%	38.7%	26.8%
200	**1%**	35.6%	19.7%	12.7%	19.4%	11.0%	2.6%	20.3%	14.4%
	5%	54.5%	37.2%	27.9%	36.2%	22.2%	8.7%	37.4%	24.4%
	10%	66.4%	50.0%	39.8%	49.0%	32.2%	17.0%	50.0%	39.0%
Super Smooth									
100	**1%**	15.3%	12.4%	7.3%	4.3%	4.0%	2.6%	4.7%	11.2%
	5%	29.0%	22.3%	18.1%	11.5%	9.6%	7.1%	11.0%	17.4%
	10%	38.7%	31.7%	27.3%	19.5%	17.1%	14.8%	20.3%	30.3%
200	**1%**	23.8%	16.7%	10.6%	4.3%	5.3%	3.0%	5.7%	14.7%
	5%	40.9%	29.1%	22.0%	13.7%	11.0%	8.2%	14.2%	26.0%
	10%	52.9%	38.4%	32.3%	21.3%	18.6%	12.7%	24.3%	34.9%

Table 4: $Y = 1 + 1.5X + \cos(2\pi X) + U$

Ordinary Smooth		**Bandwidth**						HM	S
n	**Level**	**0.01**	**0.05**	**0.1**	**0.5**	**1**	**1.5**		
100	**1%**	20.9%	16.7%	13.8%	6.9%	5.6%	1.8%	9.2%	7.3%
	5%	38.7%	31.5%	28.0%	17.1%	12.5%	5.6%	20.6%	13.6%
	10%	49.3%	43.2%	37.8%	25.4%	19.5%	9.9%	29.7%	23.7%
200	**1%**	35.9%	20.8%	15.3%	7.8%	4.8%	1.3%	9.3%	13.6%
	5%	55.9%	37.4%	28.9%	18.4%	11.2%	4.6%	21.7%	23.8%
	10%	66.8%	49.8%	40.4%	28.6%	17.6%	10.1%	31.4%	31.9%
Super Smooth									
100	**1%**	16.1%	11.2%	9.1%	5.3%	4.0%	3.1%	5.2%	7.6%
	5%	30.4%	22.4%	17.9%	12.6%	11.0%	7.3%	11.6%	17.3%
	10%	41.3%	35.0%	26.1%	20.6%	17.4%	13.7%	19.3%	27.1%
200	**1%**	23.6%	13.4%	9.4%	5.1%	4.7%	3.7%	5.3%	8.9%
	5%	39.4%	25.0%	20.0%	11.8%	12.0%	8.4%	13.1%	25.0%
	10%	50.8%	35.8%	30.5%	20.5%	19.4%	13.9%	20.3%	32.2%

Table 5: $Y = 1 + 1.5X + \cos(3\pi X) + U$

Ordinary Smooth		**Bandwidth**						HM	S
n	**Level**	**0.01**	**0.05**	**0.1**	**0.5**	**1**	**1.5**		
100	**1%**	22.8%	17.0%	11.8%	7.9%	5.4%	1.8%	9.4%	7.3%
	5%	39.3%	32.6%	28.2%	17.7%	12.9%	5.9%	20.8%	19.0%
	10%	50.6%	46.1%	38.1%	27.2%	19.8%	10.4%	31.8%	27.1%
200	**1%**	36.4%	20.1%	14.0%	8.1%	4.2%	1.9%	9.9%	7.9%
	5%	54.1%	36.3%	29.2%	19.4%	10.7%	5.0%	21.8%	21.8%
	10%	67.1%	48.9%	40.0%	27.6%	18.0%	10.3%	32.0%	27.7%
Super Smooth									
100	**1%**	17.4%	10.9%	7.5%	4.7%	4.6%	3.3%	5.6%	8.5%
	5%	31.4%	22.8%	19.6%	12.8%	10.1%	9.1%	12.0%	17.4%
	10%	42.0%	33.2%	28.3%	20.8%	16.9%	14.3%	21.0%	26.1%
200	**1%**	21.5%	15.6%	9.3%	5.5%	4.6%	2.9%	4.8%	14.0%
	5%	39.1%	29.0%	21.6%	12.3%	11.6%	8.1%	14.1%	22.5%
	10%	50.9%	37.5%	30.2%	21.1%	17.9%	13.7%	22.6%	27.8%

The results are encouraging and seem to be consistent with the theory. Table 1 indicates that our test tracks the nominal level relatively closely. There does appear to be some dependence on the bandwidth; smaller bandwidths tending to lead to an over-rejection and larger bandwidths leading to an under-rejection of the null hypothesis.

Table 2 gives a direct comparison to the tests proposed in Hall and Ma (2007) and Song (2008). As we expected, in this low frequency alternative setting, our test is generally slightly less powerful than the other tests. Having said this, in the ordinary smooth case for several choices of bandwidth our test does display the highest power. Hall and Ma's (2007) test is able to detect local alternatives at the $\sqrt{n}$-rate for both ordinary and supersmooth measurement error distributions, and the test of Song (2008) is able to detect local alternatives at the rate $\sqrt{nb^{d/2}}$ in both cases. However, our test achieves a slower polynomial rate in the ordinary smooth case and only a $\log(n)$-rate in the super smooth case. Thus it is not surprising to see our test underperform when faced with Gaussian measurement error. However, the test is still able to enjoy considerable power in this case especially for larger sample sizes.

On the other hand, as mentioned earlier, we suspect that our test is better suited to detecting high frequency alternatives than Hall and Ma (2007). This is confirmed in Tables 3-5. We find that for smaller bandwidths our test is more powerful across the range of δ. Unfortunately, the power of our test shows considerable variation across the bandwidth choices. For smaller bandwidths the power is generally much higher. This is intuitive and is explained in Fan and Li (2000). Nonsmoothing tests can be thought of as smoothing tests but with a fixed bandwidth. Thus it is the asymptotically vanishing nature of the bandwidth in smoothing tests that allows for the superior detection of high frequency alternatives. When smaller bandwidths are employed, the test is better able to pick up on these rapid changes.

As discussed at the end of Section 2.2, the test of Song (2008) will have poor power properties for some high frequency alternatives due to testing the hypothesis based on $E[Y|W]$ rather than $E[Y|X]$. This fact is also reflected in the Monte Carlo simulations where the power falls as we move to higher frequency alternatives and is inferior to the test we propose. Interestingly the test of Song (2008) appears to dominate the test of Hall and Ma (2007) for the supersmooth case but not for the ordinary smooth case.

We can learn from these simulations that for reasonably small samples with supersmooth measurement error, perhaps the tests proposed by Hall and Ma (2007) or Song (2008) would be a wiser choice if one suspects deviations from the null of a low frequency type, otherwise our test appears to be superior. However, we suggest that to avoid any risk of very low power the test proposed in this chapter may be the best option.

In order to account for the dependence of our test on the bandwidth we may look to employ the ideas of Horowitz and Spokoiny (2001) to construct a test that is adaptive to the smoothness of the regression function. In order to do this we could construct a test statistic of the form

$$T_{A,n} = \max_{b_n \in \mathbb{B}_n} T_n$$

where $\mathbb{B}_n$ is a finite set of bandwidths. To obtain valid critical values we can use a bootstrap procedure similar to the one proposed in Section 2.3. Specifically we construct a bootstrap counterpart as

$$T^*_{A,n} = \max_{b_n \in \mathbb{B}_n} T^*_n$$

where T^*_n is defined in (10). It is beyond the scope of this chapter to determine the asymptotic properties of such a test but this could prove to be a fruitful area for future research. Alternatively, another interesting line of future work might consider obtaining an analytic expression for the power function of the test which could be maximised with respect to the bandwidth to obtain an optimal bandwidth choice.

2.5. Empirical Example

We apply our test to the specification of Engel curves for food, clothing and transport. An Engel curve describes the relationship between an individual's purchases of a particular good and their total resources and hence provides an estimate of a good's income elasticity. Much work has been carried out on the estimation of Engel curves. In particular, the functional form has been shown to significantly affect estimates of income elasticity (see for example Leser, 1963). Hausmann, Newey and Powell (1995) highlighted the problem that measurement error plays in the estimation of Engel curves. To the best of our knowledge no previous work has tested the parametric specification of Engel curves whilst accounting for the inherent measurement error in the data.

We concentrate on the Working-Leser specification put forward by Leser (1963)

$$Y_i \quad = \quad a_0 + a_1\big(X_i \log(X_i)\big) + a_2 X_i + U_i$$

where Y_i is the expenditure on a given good of consumer i and X_i is the *true* total expenditure of consumer i. It is commonly believed that the measurement error in total expenditure is multiplicative, hence we take $\tilde{X}_i = \log(X_i)$ as our true regressor and adjust the specification accordingly, as in Schennach (2004). We use data from the Consumer Expenditure Survey where we take the third quarter of 2014 as our sample, giving 4312 observations. To account for the measurement error we make use of repeated measurements of X. Specifically, we use total expenditure from the current quarter as one measurement and total expenditure from the previous quarter as the other. To estimate the parametric form we employ the estimator proposed by Schennach (2004). For the nonparametric estimator we use Delaigle, Hall and Meister (2008) and select the bandwidth using the cross-validation approach also proposed in that paper. To analyse the sensitivity of our test to the choice of bandwidth we report results for various bandwidths around the cross-validated choice $b \approx 0.15$. We use the same kernel and bootstrap procedure as implemented in the Monte Carlo simulations.

Table 6 reports the p-value for our specification test on food, clothing and transport for a range of bandwidths.

Table 6: Engel Curve (P-Values)

Good	**Bandwidth**					
	0.05	**0.10**	**0.15**	**0.25**	**0.35**	**0.5**
Food	0.00	0.00	0.00	0.00	0.00	0.07
Clothing	0.00	0.00	0.00	0.00	0.00	0.10
Transport	0.00	0.00	0.00	0.00	0.00	0.00

We can see that the test is reasonably robust to the choice of bandwidth. The parametric specification is rejected for all bandwidths apart from 0.5 where we fail to reject the null hypothesis at the 5% level for food and at the 10% level for clothing. Interestingly Härdle and Mammen (1993) obtained similar findings in the case of transport but tended to fail to reject the Working-Leser specification for food. Thus, it appears that accounting for measurement error is indeed very important to draw the correct conclusions and must not simply be ignored.

2.6. Appendix - Mathematical Proofs

Hereafter, $f(b) \sim g(b)$ means $f(b)/g(b) \to 1$ as $b \to 0$.

2.6.1. Proof of Theorem 3.

2.6.1.1. *Proof of (i).* First, we define the normalisation term $C_{V,b}$ and characterise its asymptotic order. Let

$$\begin{aligned}
\xi_i(t) &= Y_i e^{\mathrm{i}t\cdot W_i} - \int e^{\mathrm{i}s\cdot W_i} m^{\mathrm{ft}}(t-s)\frac{K^{\mathrm{ft}}(sb)}{f_\epsilon^{\mathrm{ft}}(s)} ds f_\epsilon^{\mathrm{ft}}(t), \\
H_{i,j} &= \int \frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\xi_i(t)\overline{\xi_j(t)}dt.
\end{aligned}$$

Then $C_{V,b}$ is defined as

$$\begin{aligned}
C_{V,b} &= E[H_{1,2}^2] \qquad (11)\\
&= \int\int \frac{|K^{\mathrm{ft}}(t_1b)|^2}{|f_\epsilon^{\mathrm{ft}}(t_1)|^2}\frac{|K^{\mathrm{ft}}(t_2b)|^2}{|f_\epsilon^{\mathrm{ft}}(t_2)|^2}\Bigg|\left\{[m^2f]^{\mathrm{ft}}(t_1+t_2)+[\sigma^2 f]^{\mathrm{ft}}(t_1+t_2)\right\} f_\epsilon^{\mathrm{ft}}(t_1+t_2) \\
&\quad + \int\int f_W^{\mathrm{ft}}(s_1+s_2)m^{\mathrm{ft}}(t_1-s_1)m^{\mathrm{ft}}(t_2-s_2)\frac{K^{\mathrm{ft}}(s_1b)}{f_\epsilon^{\mathrm{ft}}(s_1)}\frac{K^{\mathrm{ft}}(s_2b)}{f_\epsilon^{\mathrm{ft}}(s_2)}ds_1ds_2 f_\epsilon^{\mathrm{ft}}(t_1)f_\epsilon^{\mathrm{ft}}(t_2) \\
&\quad - \int [mf]^{\mathrm{ft}}(t_2+s_1)f_\epsilon^{\mathrm{ft}}(t_2+s_1)m^{\mathrm{ft}}(t_1-s_1)\frac{K^{\mathrm{ft}}(s_1b)}{f_\epsilon^{\mathrm{ft}}(s_1)}ds_1 f_\epsilon^{\mathrm{ft}}(t_1) \\
&\quad - \int [mf]^{\mathrm{ft}}(t_1+s_1)f_\epsilon^{\mathrm{ft}}(t_1+s_1)m^{\mathrm{ft}}(t_2-s_1)\frac{K^{\mathrm{ft}}(s_1b)}{f_\epsilon^{\mathrm{ft}}(s_1)}ds_1 f_\epsilon^{\mathrm{ft}}(t_2)\Bigg|^2 dt_1dt_2.
\end{aligned}$$

To find the order of $C_{V,b}$, we consider the square of each of these four terms and all of their cross products. For example,

$$\begin{aligned}
&\int\int \frac{|K^{\mathrm{ft}}(t_1b)|^2}{|f_\epsilon^{\mathrm{ft}}(t_1)|^2}\frac{|K^{\mathrm{ft}}(t_2b)|^2}{|f_\epsilon^{\mathrm{ft}}(t_2)|^2}\left|\left\{[m^2f]^{\mathrm{ft}}(t_1+t_2)+[\sigma^2 f]^{\mathrm{ft}}(t_1+t_2)\right\} f_\epsilon^{\mathrm{ft}}(t_1+t_2)\right|^2 dt_1dt_2 \\
\sim\;& b^{-2d-4d\alpha}\int\int |K^{\mathrm{ft}}(a_1)|^2|K^{\mathrm{ft}}(a_2)|^2|a_1|^{2d\alpha}|a_2|^{2d\alpha}\left|1+((a_1+a_2)/b)^2\right|^{-d\alpha} \\
&\times\left|[(m^2+\sigma^2)f]^{\mathrm{ft}}((a_1+a_2)/b)\right|^2 da_1da_2 \\
\sim\;& b^{-d-4d\alpha}\int \left|1+a^2\right|^{-d\alpha}\left|[(m^2+\sigma^2)f]^{\mathrm{ft}}(a)\right|^2 da\int |K^{\mathrm{ft}}(a_2)|^4|a_2|^{4d\alpha}da_2 \\
=\;& O(b^{-d(1+4\alpha)}), \qquad (12)
\end{aligned}$$

where the first wave relation follows from the change of variables $(a_1,a_2)=(t_1b,t_2b)$ and Assumption 4 (i), the second wave relation follows from the change of variables $a=(a_1+a_2)/b$, and the equality follows from Assumption 3 (iii) and 4 (ii). All other squared and cross terms can be bounded in the same manner, hence we obtain $C_{V,b} = O(b^{-d(1+4\alpha)})$.

Second, we show that the estimation error of θ is negligible for the limiting distribution of T_n. Decompose $\zeta_i(t) = \xi_i(t) + \rho_i(t)$, where

$$\rho_i(t) = \int e^{\mathrm{i}s\cdot W_i}\{m_\theta^{\mathrm{ft}}(t-s) - m_{\hat{\theta}}^{\mathrm{ft}}(t-s)\}\frac{K^{\mathrm{ft}}(sb)}{f_\epsilon^{\mathrm{ft}}(s)}ds f_\epsilon^{\mathrm{ft}}(t).$$

Then the test statistic T_n is written as

$$\begin{aligned}
T_n &= \frac{1}{n}\sum_{i\neq j}\frac{1}{(2\pi)^d}\int \frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\xi_i(t)\overline{\xi_j(t)}dt + \frac{1}{n}\sum_{i\neq j}\frac{1}{(2\pi)^d}\int \frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\rho_i(t)\overline{\rho_j(t)}dt \\
&\quad + \frac{1}{n}\sum_{i\neq j}\frac{1}{(2\pi)^d}\int \frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\rho_i(t)\overline{\xi_j(t)}dt + \frac{1}{n}\sum_{i\neq j}\frac{1}{(2\pi)^d}\int \frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\xi_i(t)\overline{\rho_j(t)}dt \\
&\equiv \tilde{T}_n + R_{1n} + R_{2n} + R_{3n}.
\end{aligned}$$

By an expansion around $\hat{\theta} = \theta$ and Assumption 4 (iii), the term R_{1n} satisfies

$$R_{1n} = o_p(b^{-d/2-2\alpha})\left|\frac{1}{n^2}\sum_{i\neq j}\frac{1}{(2\pi)^d}\int \frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\rho_{1i}(t)\overline{\rho_{1j}(t)}dt\right|, \tag{13}$$

where $\rho_{1i}(t) = \int e^{\mathrm{i}s\cdot W_i}\frac{\partial m_\theta^{\mathrm{ft}}(t-s)}{\partial\theta}\frac{K^{\mathrm{ft}}(sb)}{f_\epsilon^{\mathrm{ft}}(s)}ds f_\epsilon^{\mathrm{ft}}(t)$. By the Cauchy-Schwarz inequality and Assumption 3 (ii),

$$\begin{aligned}
E\left[\int \frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\rho_{1i}(t)\overline{\rho_{1j}(t)}dt\right] &= \int |K^{\mathrm{ft}}(tb)|^2\left|\int f^{\mathrm{ft}}(s)\frac{\partial}{\partial\theta}\big(m_\theta^{\mathrm{ft}}(t-s)\big)K^{\mathrm{ft}}(sb)ds\right|^2 dt \\
&= O(1).
\end{aligned} \tag{14}$$

Also, by applying the same argument to (12) under Assumption 4 (ii), we have

$$E\left[\left(\int \frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\rho_{1i}(t)\overline{\rho_{1j}(t)}dt\right)^2\right] = O(b^{-d(1+4\alpha)}). \tag{15}$$

Combining (13)-(15) and $C_{V,b} = O(b^{-d(1+4\alpha)})$, we obtain $C_{V,b}^{-1/2}R_{1n} = o_p(1)$. In the same manner we can show $C_{V,b}^{-1/2}R_{2n} = o_p(1)$ and $C_{V,b}^{-1/2}R_{3n} = o_p(1)$ under Assumption 4 (ii) and (iii) and thus $C_{V,b}^{-1/2}T_n = C_{V,b}^{-1/2}\tilde{T}_n + o_p(1)$.

Second, we derive the limiting distribution of $C_{V,b}^{-1/2}\tilde{T}_n$. Note that $\tilde{T}_n$ is written as $\tilde{T}_n = \frac{1}{n}\sum_{i\neq j}\frac{1}{(2\pi)^d}H_{i,j}$ and is a U-statistic with zero mean (because $E[Y\exp(\mathrm{i}t\cdot W)] = [m_\theta f]^{\mathrm{ft}}(t)f_\epsilon^{\mathrm{ft}}(t)$ under H_0). To prove the asymptotic normality of $\tilde{T}_n$, we apply the central limit theorem in Hall (1984, Theorem 1). To this end, it is enough to show

$$\frac{E[H_{1,2}^4]}{n(E[H_{1,2}^2])^2} \to 0, \quad \text{and} \quad \frac{E[G_{1,2}^2]}{(E[H_{1,2}^2])^2} \to 0, \tag{16}$$

where $G_{i,j} = E[H_{1,i}H_{1,j}|Y_1, W_1]$. Recall that $C_{V,b} = E[H_{1,2}^2]$ defined in (11) satisfies $C_{V,b} = O\left(b^{-d(1+4\alpha)}\right)$. By a similar argument to bound $E[H_{1,2}^2]$ in (12), we can show

$$E[H_{1,2}^4] = E\left[\int\cdots\int\prod_{k=1}^4\frac{|K^{\mathrm{ft}}(t_k b)|^2}{|f_\epsilon^{\mathrm{ft}}(t_k)|^2}\xi_1(t_k)\overline{\xi_2(t_k)}dt_1\cdots dt_4\right] = O(b^{-3d(1+8\alpha)}).$$

For $E[G_{1,2}^2]$, we can equivalently look at

$$\begin{aligned}
& E[H_{1,3}H_{1,4}H_{2,3}H_{2,4}] \\
= & \int\cdots\int\prod_{k=1}^4\frac{|K^{\mathrm{ft}}(t_k b)|^2}{|f_\epsilon^{\mathrm{ft}}(t_k)|^2}\xi_1(t_1)\overline{\xi_3(t_1)}\xi_1(t_2)\overline{\xi_4(t_2)}\xi_2(t_3)\overline{\xi_3(t_3)}\xi_2(t_4)\overline{\xi_4(t_4)}dt_1\cdots dt_4 \\
= & O(b^{-d(1+8\alpha)}).
\end{aligned}$$

These results combined with Assumption 3 (iv) guarantee the conditions in (16). Thus, Hall (1984, Theorem 1) implies

$$C_{V,b}^{-1/2}\tilde{T}_n \xrightarrow{d} N\left(0, \frac{2}{(2\pi)^{2d}}\right),$$

and the conclusion follows.

2.6.1.2. *Proof of (ii).* A similar argument to the proof of Part (i) guarantees $\varphi(b)T_n = \varphi(b)\tilde{T}_n + o_p(1)$. Thus we hereafter derive the limiting distribution of $\tilde{T}_n$. Decompose $\tilde{T}_n = \bar{T}_n + r_{1n} + r_{2n} + r_{3n}$, where

$$\begin{aligned}
\bar{T}_n &= \frac{1}{n}\sum_{i\neq j}\frac{1}{2\pi}\int\frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}Y_ie^{\mathrm{i}tW_i}\overline{Y_je^{\mathrm{i}tW_j}}dt, \qquad (17)\\
r_{1n} &= \frac{1}{n}\sum_{i\neq j}\frac{1}{2\pi}\int\frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\left(\int e^{\mathrm{i}sW_i}m^{\mathrm{ft}}(t-s)\frac{K^{\mathrm{ft}}(sb)}{f_\epsilon^{\mathrm{ft}}(s)}dsf_\epsilon^{\mathrm{ft}}(t)\right)\overline{\left(\int e^{\mathrm{i}sW_j}m^{\mathrm{ft}}(t-s)\frac{K^{\mathrm{ft}}(sb)}{f_\epsilon^{\mathrm{ft}}(s)}dsf_\epsilon^{\mathrm{ft}}(t)\right)}dt,\\
r_{2n} &= \frac{1}{n}\sum_{i\neq j}\frac{1}{2\pi}\int\frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}Y_ie^{\mathrm{i}tW_i}\overline{\left(\int e^{\mathrm{i}sW_j}m^{\mathrm{ft}}(t-s)\frac{K^{\mathrm{ft}}(sb)}{f_\epsilon^{\mathrm{ft}}(s)}dsf_\epsilon^{\mathrm{ft}}(t)\right)}dt,\\
r_{3n} &= \frac{1}{n}\sum_{i\neq j}\frac{1}{2\pi}\int\frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\left(\int e^{\mathrm{i}sW_i}m^{\mathrm{ft}}(t-s)\frac{K^{\mathrm{ft}}(sb)}{f_\epsilon^{\mathrm{ft}}(s)}dsf_\epsilon^{\mathrm{ft}}(t)\right)\overline{Y_je^{\mathrm{i}tW_j}}dt.
\end{aligned}$$

First, we derive the limiting distribution of $\bar{T}_n$. Observe that

$$\begin{aligned}
\bar{T}_n &= \frac{1}{n}\sum_{i\neq j}\frac{1}{2\pi}\int|K^{\mathrm{ft}}(tb)|^2e^{t^2}Y_iY_j\{\cos(tW_i)\cos(tW_j)+\sin(tW_i)\sin(tW_j)\}dt\\
&= \frac{1}{nb}\sum_{i\neq j}\frac{1}{2\pi}\int|K^{\mathrm{ft}}(t)|^2e^{(t/b)^2}Y_iY_j\left\{\cos\left(\frac{tW_i}{b}\right)\cos\left(\frac{tW_j}{b}\right)+\sin\left(\frac{tW_i}{b}\right)\sin\left(\frac{tW_j}{b}\right)\right\}dt\\
&= \left(\frac{1}{b}\frac{1}{2\pi}\int|K^{\mathrm{ft}}(t)|^2e^{(t/b)^2}dt\right)\frac{1}{n}\sum_{i\neq j}Y_iY_j\left\{\cos\left(\frac{W_i}{b}\right)\cos\left(\frac{W_j}{b}\right)+\sin\left(\frac{W_i}{b}\right)\sin\left(\frac{W_j}{b}\right)\right\}\\
&\quad +O_p(b^{2+4\beta}e^{1/b^2})\\
&\equiv \left(\frac{1}{b}\frac{1}{2\pi}\int|K^{\mathrm{ft}}(t)|^2e^{(t/b)^2}dt\right)\tilde{T}_n+O_p(b^{2+4\beta}e^{1/b^2}), \qquad (18)
\end{aligned}$$

where the first equality follows from $f_\epsilon^{\mathrm{ft}}(t)=e^{-t^2/2}$ and $e^{\mathrm{i}tW_i}=\cos(tW_i)+\mathrm{i}\sin(tW_i)$, the second equality follows from a change of variables, and the third equality follows from Holzmann and Boysen (2006, Theorem 1) based on Assumption 5 (ii). Note that

$$\tilde{T}_n=\frac{1}{n}\sum_{i\neq j}Y_iY_j\left[\begin{array}{c}\left\{\cos\left(\frac{X_i}{b}\right)\cos\left(\frac{\epsilon_i}{b}\right)-\sin\left(\frac{X_i}{b}\right)\sin\left(\frac{\epsilon_i}{b}\right)\right\}\left\{\cos\left(\frac{X_j}{b}\right)\cos\left(\frac{\epsilon_j}{b}\right)-\sin\left(\frac{X_j}{b}\right)\sin\left(\frac{\epsilon_j}{b}\right)\right\}\\ +\left\{\sin\left(\frac{X_i}{b}\right)\cos\left(\frac{\epsilon_i}{b}\right)+\cos\left(\frac{X_i}{b}\right)\sin\left(\frac{\epsilon_i}{b}\right)\right\}\left\{\sin\left(\frac{X_j}{b}\right)\cos\left(\frac{\epsilon_j}{b}\right)+\cos\left(\frac{X_j}{b}\right)\sin\left(\frac{\epsilon_j}{b}\right)\right\}\end{array}\right]. \qquad (19)$$

From van Es and Uh (2005, proof of Lemma 6), it holds $\left(\frac{X_i}{b}\bmod 2\pi\right)\xrightarrow{d}V_i^X\sim U[0,2\pi]$ and $\left(\frac{\epsilon_i}{b}\bmod 2\pi\right)\xrightarrow{d}V_i^\epsilon\sim U[0,2\pi]$ as $b\to 0$ for each i, where V_i^ϵ is independent from (Y_i,V_i^X). Thus by applying Holzmann and Boysen (2006, Lemma 1), $\tilde{T}_n$ has the same limiting distribution with

$\tilde{T}_n^V = \frac{1}{n}\sum_{i\neq j} h(Q_i, Q_j)$, where $Q_i = (Y_i, V_i^X, V_i^\epsilon)$ and

$$h(Q_i, Q_j) = Y_iY_j \left[\begin{array}{c} \left\{\cos(V_i^X)\cos(V_i^\epsilon) - \sin(V_i^X)\sin(V_i^\epsilon)\right\}\left\{\cos(V_j^X)\cos(V_j^\epsilon) - \sin(V_j^X)\sin(V_j^\epsilon)\right\} \\ +\left\{\sin(V_i^X)\cos(V_i^\epsilon) + \cos(V_i^X)\sin(V_i^\epsilon)\right\}\left\{\sin(V_j^X)\cos(V_j^\epsilon) + \cos(V_j^X)\sin(V_j^\epsilon)\right\} \end{array} \right].$$

Observe that $\operatorname{Cov}(h(Q_1, Q_2), h(Q_1, Q_3)) = 0$ because $E[\cos(V_i^\epsilon)] = 0$. Therefore, by applying the limit theorem for degenerate U-statistics with a fixed kernel h (Serfling, 1980, Theorem 5.5.2), we obtain

$$\tilde{T}_n^V \xrightarrow{d} \sum_{k=1}^{\infty} \lambda_k (Z_k^2 - 1), \tag{20}$$

where $\{Z_k\}$ is an independent sequence of standard normal random variables and $\{\lambda_k\}$ are the eigenvalues of the integral operator

$$(\Lambda g)(Q_1) = \lambda g(Q_1). \tag{21}$$

where $(\Lambda g)(Q_1) = E[h(Q_1, Q_2)g(Q_2)|Q_1]$. Also, van Es and Uh (2005, Lemma 5) gives

$$\frac{1}{2\pi}\int |K^{\mathrm{ft}}(t)|^2 e^{(t/b)^2} dt \sim \frac{b}{\varphi(b)}, \tag{22}$$

where $\Gamma(\cdot)$ is the gamma function. Combining (18)-(22),

$$\varphi(b)\bar{T}_n \xrightarrow{d} \sum_{k=1}^{\infty} \lambda_i (Z_i^2 - 1).$$

Next, we show that r_{1n} is negligible. Observe that

$$\begin{aligned} r_{1n} &= \frac{1}{nb^3}\sum_{i\neq j}\frac{1}{2\pi}\int |K^{\mathrm{ft}}(t)|^2 \left(\int e^{\mathrm{i}sW_i/b} m^{\mathrm{ft}}\left(\frac{t-s}{b}\right)\frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/b)} ds\right)\overline{\left(\int e^{\mathrm{i}sW_j/b} m^{\mathrm{ft}}\left(\frac{t-s}{b}\right)\frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/b)} ds\right)} dt \\ &= \frac{1}{nb^3}\sum_{i\neq j}\frac{1}{2\pi}\int |K^{\mathrm{ft}}(t)|^2 \left[\begin{array}{c} \left\{\int \cos\left(\frac{s_1W_i}{b}\right) m^{\mathrm{ft}}\left(\frac{t-s_1}{b}\right)\frac{K^{\mathrm{ft}}(s_1)}{f_\epsilon^{\mathrm{ft}}(s_1/b)} ds_1\right\} \\ \times\left\{\int \cos\left(\frac{s_2W_j}{b}\right) m^{\mathrm{ft}}\left(\frac{s_2-t}{b}\right)\frac{K^{\mathrm{ft}}(s_2)}{f_\epsilon^{\mathrm{ft}}(-s_2/b)} ds_2\right\} \\ +\left\{\int \sin\left(\frac{s_1W_i}{b}\right) m^{\mathrm{ft}}\left(\frac{t-s_1}{b}\right)\frac{K^{\mathrm{ft}}(s_1)}{f_\epsilon^{\mathrm{ft}}(s_1/b)} ds_1\right\} \\ \times\left\{\int \sin\left(\frac{s_2W_j}{b}\right) m^{\mathrm{ft}}\left(\frac{s_2-t}{b}\right)\frac{K^{\mathrm{ft}}(s_2)}{f_\epsilon^{\mathrm{ft}}(-s_2/b)} ds_2\right\} \end{array} \right] dt \\ &= \left(\frac{1}{2\pi}\int\int\int \frac{|K^{\mathrm{ft}}(t)|^2 K^{\mathrm{ft}}(s_1)K^{\mathrm{ft}}(s_2)}{f_\epsilon^{\mathrm{ft}}(s_1/b)f_\epsilon^{\mathrm{ft}}(-s_2/b)} m^{\mathrm{ft}}\left(\frac{t-s_1}{b}\right) m^{\mathrm{ft}}\left(\frac{s_2-t}{b}\right) ds_1 ds_2 dt\right) \\ &\quad \times \frac{1}{nb^3}\sum_{i\neq j}\left\{\cos\left(\frac{W_i}{b}\right)\cos\left(\frac{W_j}{b}\right) + \sin\left(\frac{W_i}{b}\right)\sin\left(\frac{W_j}{b}\right)\right\} + O_p(b^{2+4\beta}e^{1/b^2}), \end{aligned}$$

where the first equality follows from a change of variables, the second equality follows from a direct calculation using $e^{\mathrm{i}sW_i} = \cos(sW_i) + \mathrm{i}\sin(sW_i)$, the third equality follows from Holzmann and Boysen (2006, Theorem 1) based on Assumption 5 (ii). By a similar argument used to show (20), it holds

$$\frac{1}{n}\sum_{i\neq j}\left\{\cos\left(\frac{W_i}{b}\right)\cos\left(\frac{W_j}{b}\right)+\sin\left(\frac{W_i}{b}\right)\sin\left(\frac{W_j}{b}\right)\right\} = O_p(1).$$

Also, we obtain

$$\begin{aligned}
&\frac{1}{2\pi}\int\int\int \frac{|K^{\mathrm{ft}}(t)|^2 K^{\mathrm{ft}}(s_1)K^{\mathrm{ft}}(s_2)}{f_\epsilon^{\mathrm{ft}}(s_1/b)f_\epsilon^{\mathrm{ft}}(-s_2/b)} m^{\mathrm{ft}}\left(\frac{t-s_1}{b}\right) m^{\mathrm{ft}}\left(\frac{s_2-t}{b}\right) ds_1 ds_2 dt \\
= &\frac{b^4 e^{1/b^2}}{2\pi}\int\int\int \left[\begin{array}{c} |K^{\mathrm{ft}}(t)|^2 K^{\mathrm{ft}}(1-b^2v_1)K^{\mathrm{ft}}(1-b^2v_2) \\ \times e^{\frac{(1-b^2v_1)^2-1}{2b^2}} e^{\frac{(1-b^2v_2)^2-1}{2b^2}} m^{\mathrm{ft}}\left(\frac{t-1+b^2v_1}{b}\right) m^{\mathrm{ft}}\left(\frac{1-b^2v_2-t}{b}\right)\end{array}\right] dv_1 dv_2 dt \\
\sim &\frac{A^2 b^{4+4\beta} e^{1/b^2}}{2\pi}\left(\int |K^{\mathrm{ft}}(t)|^2 m^{\mathrm{ft}}\left(\frac{t-1}{b}\right) m^{\mathrm{ft}}\left(\frac{1-t}{b}\right) dt\right)\left(\int v_1^\beta e^{-v_1} dv_1\right)\left(\int v_2^\beta e^{-v_2} dv_2\right) \\
\sim &\frac{A^2\Gamma(\beta+1)^2 b^{5+6\beta} e^{1/b^2}}{2\pi}\int |t|^{2\beta}|m^{\mathrm{ft}}(t)|^2 dt \\
= &O(b^{5+6\beta}e^{1/b^2}),
\end{aligned}$$

where the first equality follows from changes of variables $s_1 = 1 - b^2 v_1$ and $s_2 = 1 - b^2 v_2$, the wave relations follow from Assumption 5 (i), and the last equality follows from Assumption 5 (ii). Combining these results,

$$\varphi(b) r_{1n} = O_p(b^{1+2\beta}),$$

and thus r_{1n} is negligible. Similar arguments imply that the terms r_{2n} and r_{3n} are also asymptotically negligible and the conclusion follows.

2.6.1.3. *Proof of (iii).* The proof for the general supersmooth case follows the same steps as in the proof of Part (ii) for the normal case. As the proof is similar, we omit the most part. Hereafter we show why the condition $\gamma > 1$ is imposed in this case. The dominant term $\bar{T}_n$ defined in (17) satisfies

$$\bar{T}_n \sim \frac{1}{nb}\sum_{i\neq j}\frac{1}{2\pi C_\epsilon^2}\int |K^{\mathrm{ft}}(t)|^2\left|\frac{t}{b}\right|^{-2\gamma_0} e^{\frac{2|t|^\gamma}{\mu b^\gamma}} Y_i Y_j\left\{\cos\left(\frac{tW_i}{b}\right)\cos\left(\frac{tW_j}{b}\right)+\sin\left(\frac{tW_i}{b}\right)\sin\left(\frac{tW_j}{b}\right)\right\} dt.$$

We now show that

$$
\begin{aligned}
D_{\cos} \quad \equiv \quad & \frac{1}{nb}\sum_{i\neq j}\frac{1}{2\pi C_\epsilon^2}\int |K^{\mathrm{ft}}(t)|^2\left|\frac{t}{b}\right|^{-2\gamma_0} e^{\frac{2|t|^\gamma}{\mu b^\gamma}} Y_i Y_j\left\{\cos\left(\frac{tW_i}{b}\right)\cos\left(\frac{tW_j}{b}\right)\right\}dt \\
& -\left(\frac{1}{2\pi C_\epsilon^2}\int |K^{\mathrm{ft}}(t)|^2\left|\frac{t}{b}\right|^{-2\gamma_0} e^{\frac{2|t|^\gamma}{\mu b^\gamma}}dt\right)\frac{1}{nb}\sum_{i\neq j}Y_iY_j\left\{\cos\left(\frac{W_i}{b}\right)\cos\left(\frac{W_j}{b}\right)\right\}
\end{aligned}
$$

is asymptotically negligible, as well as the correspondingly defined $D_{\sin}$. We have seen that each term is zero mean. Following the proof of Holzmann and Boysen (2006, Theorem 1), we obtain

$$
\left|\cos\left(\frac{tW_i}{b}\right)\cos\left(\frac{tW_j}{b}\right)-\cos\left(\frac{W_i}{b}\right)\cos\left(\frac{W_j}{b}\right)\right|\leq (1-t)\frac{(|W_i|+|W_k|)}{b}.
$$

Thus, similar arguments to van Es and Uh (2005, Lemmas 1 and 5) using Assumption 5 (ii) imply

$$
\begin{aligned}
\mathrm{Var}(D_{\cos}) \quad \leq \quad & O(n^{-2}b^{4\gamma_0-4})\left(\int (1-t)|K^{\mathrm{ft}}(t)|^2\,|t|^{-2\gamma_0}\,e^{\frac{2|t|^\gamma}{\mu b^\gamma}}dt\right)^2\sum_{i\neq j}E\left[|Y_i|^2|Y_j|^2(|W_i|+|W_k|)^2\right] \\
= \quad & O\left(b^{4\gamma_0-4}\left(b^{\gamma(2+2\beta)}e^{\frac{2}{\mu b^\gamma}}\right)^2\right),
\end{aligned}
$$

and we obtain $D_{\cos}=O_p\left(b^{2(\gamma-1)+2\gamma\beta+2\gamma_0}e^{\frac{2}{\mu b^\gamma}}\right)$. The same argument applies to $D_{\sin}$. Note that

$$
\begin{aligned}
\bar{T}_n \quad = \quad & \left(\frac{1}{b}\frac{1}{2\pi C_\epsilon^2}\int |K^{\mathrm{ft}}(t)|^2\left|\frac{t}{b}\right|^{-2\gamma_0} e^{\frac{2|t|^\gamma}{\mu b^\gamma}}dt\right)\frac{1}{n}\sum_{i\neq j}Y_iY_j\left\{\cos\left(\frac{W_i}{b}\right)\cos\left(\frac{W_j}{b}\right)+\sin\left(\frac{W_i}{b}\right)\sin\left(\frac{W_j}{b}\right)\right\} \\
& +O\left(b^{2(\gamma-1)+2\gamma\beta+2\gamma_0}e^{\frac{2}{\mu b^\gamma}}\right) \\
= \quad & \frac{A^2\mu^{1+2\beta}b^{\gamma-1+2\gamma\beta+2\gamma_0}e^{\frac{2}{\mu b^\gamma}}\Gamma(2\beta+1)}{(2\lambda)^{1+2\beta}\pi C_\epsilon^2}\tilde{T}_n+O\left(b^{2(\gamma-1)+2\gamma\beta+2\gamma_0}e^{\frac{2}{\mu b^\gamma}}\right) \\
\equiv \quad & \frac{\tilde{T}_n}{\phi(b)}+O\left(b^{2(\gamma-1)+2\gamma\beta+2\gamma_0}e^{\frac{2}{\mu b^\gamma}}\right),
\end{aligned}
$$

where the second equality follows from the definition of $\tilde{T}_n$ in (19) and a modification of van Es and Uh (2005, Lemma 5). Therefore, we obtain

$$
\phi(b)T_n=\tilde{T}_n+O(b^{\gamma-1}).
$$

The limiting distribution of $\tilde{T}_n$ is obtained in the proof of Part (ii). The remainder term becomes negligible if we impose $\gamma>1$.

2.6.2. Proof of Theorem 4.

2.6.2.1. *Proof of (i).* By a similar argument to the proof of Theorem 3 (i), the estimation error $\hat{\theta} - \theta$ is negligible for the asymptotic properties of T_n and thus it is written as

$$\begin{aligned} T_n &= \frac{1}{n}\sum_{i \neq j} \frac{1}{(2\pi)^d} \int \frac{|K^{\text{ft}}(tb)|^2}{|f_\epsilon^{\text{ft}}(t)|^2} \xi_i(t)\overline{\xi_j(t)}dt + \frac{1}{n}\sum_{i \neq j} \frac{1}{(2\pi)^d} \int \frac{|K^{\text{ft}}(tb)|^2}{|f_\epsilon^{\text{ft}}(t)|^2} \eta_i(t)\overline{\eta_j(t)}dt \\ &\quad + \frac{1}{n}\sum_{i \neq j} \frac{1}{(2\pi)^d} \int \frac{|K^{\text{ft}}(tb)|^2}{|f_\epsilon^{\text{ft}}(t)|^2} \xi_i(t)\overline{\eta_j(t)}dt + \frac{1}{n}\sum_{i \neq j} \frac{1}{(2\pi)^d} \int \frac{|K^{\text{ft}}(tb)|^2}{|f_\epsilon^{\text{ft}}(t)|^2} \eta_i(t)\overline{\xi_j(t)}dt + o_p(C_{V,b}^{1/2}) \\ &\equiv \tilde{T}_n + R_{1n}^* + R_{2n}^* + R_{3n}^* + o_p(C_{V,b}^{1/2}), \end{aligned}$$

where

$$\begin{aligned} \eta_i(t) &= \int e^{\text{i}s \cdot W_i} \{m^{\text{ft}}(t-s) - m_\theta^{\text{ft}}(t-s)\} \frac{K^{\text{ft}}(sb)}{f_\epsilon^{\text{ft}}(s)} ds f_\epsilon^{\text{ft}}(t) \\ &= c_n \int e^{\text{i}s \cdot W_i} \Delta^{\text{ft}}(t-s) \frac{K^{\text{ft}}(sb)}{f_\epsilon^{\text{ft}}(s)} ds f_\epsilon^{\text{ft}}(t), \end{aligned}$$

under H_{1n}. By Theorem 3 (i), it holds $C_{V,b}^{-1/2}\tilde{T}_n \xrightarrow{d} N\left(0, \frac{2}{(2\pi)^{2d}}\right)$. For R_{1n}^*, observe that

$$\begin{aligned} E[C_{V,b}^{-1/2}R_{1n}^*] &= \frac{(n-1)c_n^2}{(2\pi)^d C_{V,b}^{1/2}} \int\int\int |K^{\text{ft}}(tb)|^2 K^{\text{ft}}(s_1 b) K^{\text{ft}}(s_2 b) \Delta^{\text{ft}}(t-s_1)\Delta^{\text{ft}}(s_2-t) f^{\text{ft}}(s_1) f^{\text{ft}}(-s_2) ds_1 ds_2 dt \\ &\equiv \Delta_n. \end{aligned} \tag{23}$$

By the definition of c_n, $C_{V,b} = O(b^{-d(1+4\alpha)})$ (obtained in the proof of Theorem 3 (i)), and Assumption 3 (ii), it holds $E[C_{V,b}^{-1/2}R_{1n}^*] = O(1)$ and the limit of Δ_n exists. Also, a similar argument to (12) yields

$$\begin{aligned} E[R_{1n}^{*2}] &= c_n^4 \int \cdots \int \frac{|K^{\text{ft}}(t_1 b)|^2 |K^{\text{ft}}(t_2 b)|^2 K^{\text{ft}}(s_1 b) K^{\text{ft}}(s_2 b) K^{\text{ft}}(s_3 b) K^{\text{ft}}(s_4 b)}{f_\epsilon^{\text{ft}}(s_1) f_\epsilon^{\text{ft}}(-s_2) f_\epsilon^{\text{ft}}(s_3) f_\epsilon^{\text{ft}}(-s_4)} f_W^{\text{ft}}(s_1+s_3) f_W^{\text{ft}}(-s_2-s_4) \\ &\quad \times \Delta^{\text{ft}}(t_1-s_1)\Delta^{\text{ft}}(s_2-t_1)\Delta^{\text{ft}}(t_2-s_3)\Delta^{\text{ft}}(s_4-t_2) ds_1 \cdots ds_4 dt_1 dt_2 \\ &= O(b^{-d(1+4\alpha)}). \end{aligned}$$

Therefore, $\text{Var}(C_{V,b}^{-1/2}R_{1n}^*) \to 0$ and we obtain $C_{V,b}^{-1/2}R_{1n}^* \xrightarrow{p} \lim_{n\to\infty}\Delta_n$. Finally, using similar arguments combined with $E[\xi_i(t)] = 0$, we can show that $C_{V,b}^{-1/2}R_{2n}^* \xrightarrow{p} 0$ and $C_{V,b}^{-1/2}R_{3n}^* \xrightarrow{p} 0$. Combining these results, the conclusion follows.

2.6.2.2. *Proof of (ii).* Similarly to the proof of Part (i), we can decompose

$$T_n = \tilde{T}_n + R_{1n}^* + R_{2n}^* + R_{3n}^* + o_p(\varphi(b)^{-1}).$$

Theorem 3 (ii) implies the limiting distribution of $\varphi(b)\tilde{T}_n$. For R_{1n}^*, note that

$$\begin{aligned} E[\varphi(b)R_{1n}^*] &= \varphi(b)(n-1)c_n^2 \int\int\int |K^{\mathrm{ft}}(tb)|^2 K^{\mathrm{ft}}(s_1 b)K^{\mathrm{ft}}(s_2 b) \\ &\quad \times \Delta^{\mathrm{ft}}(t-s_1)\Delta^{\mathrm{ft}}(s_2-t)f^{\mathrm{ft}}(s_1)f^{\mathrm{ft}}(-s_2)ds_1 ds_2 dt \\ &\equiv \Upsilon_n, \end{aligned} \tag{24}$$

and the limit of Υ_n exists from the definition of c_n. Also, by similar treatment to r_{1n} in the proof of Theorem 3 (ii), we can show $\mathrm{Var}(\varphi(b)R_{1n}^*) \to 0$ and thus $\varphi(b)R_{1n}^* \to \lim_{n\to\infty}\Upsilon_n$. Using similar arguments, combined with $E[\xi_{1i}(t)] = 0$, we can again show that R_{2n}^* and R_{3n}^* are asymptotically negligible. Therefore, the conclusion follows.

2.6.2.3. *Proof of (iii).* The proof is identical to that of Part (ii) with $\varphi(b)$ replaced by $\phi(b)$ and setting $c_n = b^{(\lambda-1)/2+\lambda\beta+\lambda_0} e^{\frac{1}{\mu b^\lambda}}$.

CHAPTER 3

Nonparametric Significance Testing in Measurement Error Models

3.1. Introduction

The importance of significance testing hardly needs stating; it is arguably the most widely used of any statistical test. Significance tests are used to determine the validity of our economic theories as well as to justify model simplification. This latter application is particularly relevant in nonparametric estimation which suffers from convergence rates whose speed falls as the number of regressors increase - ‘the curse of dimensionality’. Given the growing popularity of nonparametric estimation, significance tests have never been more important. Moreover, measurement error in nonparametric estimation amplifies the curse of dimensionality. Consequently, the benefit from model parsimony, and hence the importance of significance testing, is even greater when working with contaminated data.

Such contaminated data is a well known source of inconsistency in estimators, and correspondingly, invalidity of test statistics. This is a problem that plagues economic, medical, social and physical data sets; in fact, measurement error can be found in nearly every type of data. One possible cause for such noisy data is an imperfect measurement instrument, for example survey data is commonly held to be contaminated with error. However, we argue that measurement error is a far more general phenomenon - whenever the variables in our theory do not exactly match the variables in our data, measurement error is present. Given its prevalence and undesirable consequences, contaminated data is a problem that cannot be ignored.

The main contribution of this chapter can be highlighted in the following way. In a linear model, where some variables are mismeasured, to test the significance of a subset of regressors (correctly or incorrectly measured) a Wald test can be used based on an IV regression. However, if we wish to move beyond a simplistic linear specification, to allow the relationship between the outcome and the full set of regressors to be left undetermined, there is currently no method to conduct such a test. This chapter solves this problem. It should be emphasised that the situation with which we concern ourselves is very general - the measurement error need not affect the variables whose significance we are testing, it may be that only one of the controlling variables suffers from measurement error.

Theory often provides little guidance on model specification; in the majority of cases model choice - linearity in particular - is determined according to simplicity rather than adequacy. Under model misspecification, estimators are generally inconsistent, and consequently, statistical tests which use such estimators have incorrect size. Hence, tests based on parametric choices are likely to be invalid. To overcome this problem, many tests, including the one proposed in this chapter, are conducted using nonparametric methods which impose less stringent conditions on functional form. Moreover, significance testing in a nonparametric framework is likely to be more intuitive. We ask, does variable X affect variable Y? Rather than, for example, does X affect Y in a linear way?

Unfortunately, the relaxation of assumptions when using any nonparametric estimator comes at the cost of slower convergence rates. This results in a reduction in power for tests based on such estimators. To remedy this problem, in the specification testing literature, so called 'non-smoothing' tests were developed which only require estimation of the regression function under the null hypothesis. For a specification test, this negates the need for nonparametric estimation (as the model under the null is parametric) and allows the detection of local, linear alternatives at the $\sqrt{n}$ rate (see for example Bierens, 1990, and Stute, 1997). This approach was extended to the problem of significance testing and, despite the model under the null now being nonparametric, these tests also resulted in $\sqrt{n}$ rates of detection (see for example Chen and Fan, 1999, and Delgado and Manteiga, 2001). This is in contrast to 'smoothing' tests, which estimate the model under the null and alternative, and typically attain slower then $\sqrt{n}$ convergence in both testing problems (see for example Härdle and Mammen, 1993, and Fan and Li, 1996).[1] Hence, the key benefit of non-smoothing tests is intrinsically linked to the curse of dimensionality. Since this is exacerbated in the presence of measurement error we follow a non-smoothing approach in this chapter.

[1]Of course, this is not to say that smoothing tests do not have benefits - in general they achieve greater power in detecting high-frequency alternatives.

However, there is a key difference between a conventional non-smoothing test and an analogous one with measurement error. Non-smoothing tests are conducted by first converting a finite number of conditional moment restrictions into uncountably many unconditional moment restrictions. A simple empirical average can then be taken to estimate these moment restrictions. Unfortunately, when the data is contaminated we are not able to take this empirical average since the regressors are unobservable. Instead, we must multiply by the estimated density of the full set of regressors and integrate over their range. In this sense we refer to this approach as a 'semi-smoothing' approach as we require nonparametric estimation using the full set of regressors. Nonetheless, we show we are still able to retain $\sqrt{n}$ convergence rates for this test.

There is a plethora of research on nonparametric significance testing when the data is uncontaminated. Fan and Li (1996), Aït-Sahalia, Bickel and Stoker (2001) and Lavergne and Vuong (1996) among many others develop smoothing techniques for this problem, all of which suffer from the curse of dimensionality. Whilst Delgado and Manteiga (2001) propose a consistent test able to detect alternatives converging to the null hypothesis at the $\sqrt{n}$ rate using the non-smoothing approach. Neumeyer and Dette (2003) develop a general non-smoothing test for the equality of two nonparametric regression curves, whilst Lavergne (2001) provides an analogous result using smoothing techniques. Chen and Fan (1999) propose a non-smoothing significance test in a time series context by extending the work of Robinson (1989), whilst Li (1999) develops an analogous test using smoothing methods. There is also a recent line of research which cleverly combines the two approaches. Lavergne, Maistre and Patilea (2015) consider a hybrid approach, creating a consistent test that has rates of convergence that do not depend on the dimension of the regressor but are equivalent to those achieved by smoothing tests with a single covariate. Finally, Racine (1997) follows a different method, testing whether the partial derivatives of the regression function with respect to the variables being tested are zero.

As of yet there appears to be no results on significance testing in the presence of measurement error. However, there has been some work carried out on other testing problems in this setting. Most notably, specification testing has received some attention with Hall and Ma (2007) and Chapter 2 of this thesis proposing tests for this scenario.

Finally, in this chapter we will use deconvolution techniques to estimate the regression function. The literature on nonparametric estimation and inference in the presence of measurement error has used deconvolution methods rather extensively, the interested reader is referred to the comprehensive survey of Schennach (2013). However, only in exceptional circumstances have $\sqrt{n}$ rates of convergence been obtained in these nonparametric settings (see for example Fan, 1995).

This chapter is organised as follows. Section 3.2 outlines the hypothesis of interest and the test statistic, as well as discussing possible alternatives to our test. Section 3.3 presents the asymptotic properties of our statistic, including theory for when the density of the measurement error must be estimated. This section also extends our results to weakly dependent data and provides bootstrap procedures to obtain critical values. Section 3.4 considers the small sample performance of our test through a Monte Carlo study. Section 3.5 considers two empirical applications of the test. The first uses cross-sectional data to test the effect of cognitive ability on income, life satisfaction, health and risk aversion. The second answers the important policy question of whether future inflation expectations are able to stimulate current consumption. Finally, Section 3.6 concludes. We relegate all mathematical proofs to Appendix 3.7.

3.2. Methodology

Consider the nonparametric regression model

$$Y = m(X) + U \quad \text{with } E[U|X] = 0.$$

$Y \in \mathbb{R}$ is a response variable, $X = (X_{(1}', X_{(2)}')' \in \mathbb{R}^d$ is a vector of regressors, where $X_{(1)} \in \mathbb{R}^{d_1}$, $X_{(2)} \in \mathbb{R}^{d_2}$ with $d = d_1 + d_2$, and $U \in \mathbb{R}$ is the regression error term. Throughout this chapter we denote the first d_1 elements of any vector $z \in \mathbb{R}^d$ by $z_{(1)}$. Similarly, $z_{(2)}$ denotes the final d_2 elements, whilst z_j denotes the j^{th} element of z.

We assume that X is not directly observable due to measurement error. Instead the variable W is observed through the relation

$$W = X + \epsilon,$$

where $\epsilon \in \mathbb{R}^d$ is a vector of measurement errors with independent components, has a known density $f_\epsilon(\cdot)$ and is independent of (Y, X). Since this is the first attempt to deal with any form of measurement error, we start with the classical measurement error assumption and leave nonclassical error for future work.[2] The case of unknown $f_\epsilon(\cdot)$ is considered in Section 3.3.4.

We are interested in testing the significance of the set of variables in $X_{(2)}$. More precisely, define $r(x_{(1)}) \equiv E[Y|X_{(1)} = x_{(1)}]$, we wish to test the hypothesis

$$\begin{aligned} \mathrm{H}_0 &: \; m(x_{(1)}, x_{(2)}) = r(x_{(1)}) \text{ for almost every } (x_{(1)}, x_{(2)}) \in \mathbb{R}^d, \\ \mathrm{H}_1 &: \; \mathrm{H}_0 \text{ is false,} \end{aligned}$$

based on the random sample $\{Y_i, W_i\}_{i=1}^n$ of observables; the case of dependent data is deferred until Section 3.3.3. It is important to emphasise that although we focus on the case in which all variables are affected by measurement error, this test is applicable to any combination of error-free and contaminated regressors. For example, $X_{(1)}$ may contain a single mismeasured regressor and $X_{(2)}$ may be perfectly measured - to the best of our knowledge, this is the first attempt to deal with such a situation.

Notice that the null hypothesis is equivalent to the conditional moment restriction

$$E\big[(Y - E[Y|X_{(1)}])|X\big] = 0 \qquad a.s. \tag{25}$$

In the spirit of Bierens (1982, 1990), we can write (25) as an unconditional moment restriction of the form

$$T(\xi) \;\; \equiv \;\; E\big[(Y - E[Y|X_{(1)}])f_{X_{(1)}}(X_{(1)})\mathcal{W}(X;\xi)\big] = 0 \qquad \text{for all } \xi \in \Xi,$$

where $f_{X_{(1)}}(\cdot)$ is the density function of $X_{(1)}$ and $\mathcal{W}(X;\xi) = \mathcal{W}(X_{(1)}, X_{(2)};\xi)$ is a 'generically totally revealing' function (see Stinchcombe and White, 1998) indexed by $\xi \in \Xi$ with $\Xi \subseteq \mathbb{R}^d$ a compact set with non-empty interior. As in Bierens (1990), to simplify our analysis, without loss of generality, we can define $\mathcal{W}(X;\xi) = \bar{\mathcal{W}}(\Phi(X);\xi)$, where $\Phi(\cdot)$ is a one-to-one mapping from $\mathbb{R}^d$ to a compact set. Common choices for $\bar{\mathcal{W}}(X;\xi)$ include $e^{\mathrm{i}\xi' X}$ and $e^{\xi' X}$ used in Bierens (1982,

[2]In many situations of nonclassical measurement error, it is only the variance of the error which depends on the true regressor. For example, the variance of the measurement error for reported income is likely to be larger for larger values of the true income. In such a situation we could use a multiplicative error, $W = X\epsilon$, where ϵ is still independent of X, yet the variance of the error now depends on X. We can then convert this into an additive structure by simply taking the natural logarithm.

1990), respectively, the logistic function, $1/[1+\exp(c-\xi' X)]$, with $c \neq 0$, as used in White (1989), as well as $\mathcal{I}(X \leq \xi)$ proposed by Stute (1997), where $\mathcal{I}(\cdot)$ is the indicator function. The multiplication by $f_{X_{(1)}}(X_{(1)})$ in the definition of $T(\xi)$ is used only to remove the random denominator in $E[Y|X_{(1)}]$ and hence simplify analysis.

We propose to estimate $T(\xi)$ as

$$\hat{T}_n(\xi) = \int\int \left(y - \hat{r}\left(x_{(1)}\right)\right) \hat{f}_{X_{(1)}}\left(x_{(1)}\right) \hat{f}_{Y,X}\left(y,x\right) \mathcal{W}\left(x;\xi\right) dxdy.$$

Notice that our statistic is quite different from a conventional non-smoothing statistic which would take the form

$$\sum_{i=1}^{n} \left(Y - \hat{r}\left(X_{(1)i}\right)\right) \hat{f}_{X_{(1)}}\left(X_{(1)i}\right) \mathcal{W}\left(X_i;\xi\right).$$

By introducing measurement error, the true regressors become unobservable and an empirical average is unable to be taken. Instead we must multiply by the estimated joint density of the data and integrate over their range.

As a nonparametric estimator of $r(\cdot)$, we use the deconvolution kernel estimator

$$\hat{r}(x_{(1)}) = \frac{\sum_{i=1}^{n} Y_i \mathcal{K}_b\left(\frac{x_{(1)}-W_{(1)i}}{b}\right)}{\sum_{i=1}^{n} \mathcal{K}_b\left(\frac{x_{(1)}-W_{(1)i}}{b}\right)},$$

where

$$\mathcal{K}_b(a) = \frac{1}{(2\pi b)^{\dim(a)}} \int e^{-\mathrm{i}t\cdot a} \frac{K^{\mathrm{ft}}(t)}{f_\epsilon^{\mathrm{ft}}(t/b)} dt,$$

$\dim(a)$ is the dimension of the argument a, b is a bandwidth and $K^{\mathrm{ft}}(\cdot)$ and $f_\epsilon^{\mathrm{ft}}(\cdot)$ are the Fourier transforms of a kernel function $K(\cdot)$ and the measurement error density $f_\epsilon(\cdot)$, respectively.

To estimate each of the densities we employ

$$\begin{aligned}
\hat{f}_{X_{(1)}}(x_{(1)}) &= \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_b\left(\frac{x_{(1)}-W_{(1)i}}{b}\right), \\
\hat{f}_{Y,X}(y,x) &= \frac{1}{n}\sum_{i=1}^{n} K_b\left(\frac{y-Y_i}{b}\right) \mathcal{K}_b\left(\frac{x-W_i}{b}\right).
\end{aligned}$$

Since Y is observable we use a combination of a standard kernel and deconvolution kernel function in the estimator for $f_{Y,X}(\cdot,\cdot)$. Throughout the chapter we assume $f_\epsilon^{\mathrm{ft}}(t) \neq 0$ for all $t \in \mathbb{R}^d$ and $K^{\mathrm{ft}}(\cdot)$ has compact support so that $\mathcal{K}_b(\cdot)$ is well-defined. Both of these assumptions are common in the literature. The former assumption is satisfied by most conventional distributions, however, a notable exception is the uniform distribution. If this assumption is unlikely to hold, a possible solution involves using a ridge parameter approach (see for example Meister, 2009). The latter assumption is merely a restriction on the choice of kernel; common choices which satisfy this restriction include the Sinc kernel, $K(x) = \frac{\sin x}{\pi x}$, where $K^{\text{ft}}(t) = \mathcal{I}(-1 \leq t \leq 1)$, and the kernel due to Fan (1992), $K(x) = \frac{48\cos(x)}{\pi x^4}\left(1 - \frac{15}{x^2}\right) - \frac{144\sin(x)}{\pi x^5}\left(2 - \frac{5}{x^2}\right)$, where $K^{\text{ft}}(t) = (1-t^2)^3\mathcal{I}(-1 \leq t \leq 1)$.

Given these estimators we can write

$$\begin{aligned}\hat{T}_n(\xi) &= \int\int \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} yK_b\left(\frac{y-Y_i}{b}\right)\mathcal{K}_b\left(\frac{x-W_i}{b}\right)\mathcal{K}_b\left(\frac{x_{(1)}-W_{(1)j}}{b}\right)\mathcal{W}(x;\xi)dxdy \\ &\quad - \int \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Y_j\mathcal{K}_b\left(\frac{x-W_i}{b}\right)\mathcal{K}_b\left(\frac{x_{(1)}-W_{(1)j}}{b}\right)\mathcal{W}(x;\xi)dx \\ &= \frac{1}{n^2}\sum_{i\neq j}^{n}(Y_i - Y_j)\int \mathcal{K}_b\left(\frac{x-W_i}{b}\right)\mathcal{K}_b\left(\frac{x_{(1)}-W_{(1)j}}{b}\right)\mathcal{W}(x;\xi)dx,\end{aligned}$$

where the second equality follows from $\int yK_b\left(\frac{y-Y_i}{b}\right)dy = Y_i$, using a change of variables.

In order to construct a test statistic based on $\hat{T}_n(\xi)$ we propose a Cramer-von Mises type test

$$CM_n = \int_{\Xi}\left|\hat{T}_n(\xi)\right|^2 d\mu(\xi),$$

where $\mu(\cdot)$ is an absolutely continuous probability measure on Ξ and $|\cdot|$ is the norm for complex numbers. An alternative approach would be to use a Kolmogorov-Smirnov test of the form

$$K_n = \sup_{\xi\in\Xi}\left|\hat{T}_n(\xi)\right|.$$

It has been found that the Cramer-von Mises test tends to outperform the Kolmogorov-Smirnov test when testing the equality of distributions and, as discussed in Chen and Fan (1999), the Cramer-von Mises test can be better directed towards different alternatives by the choice of $\mu(\cdot)$. As such, we concentrate on the Cramer-von Mises test in this chapter; we conjecture that very similar results would be found with the Kolmogorov-Smirnov form of the test.

3.2.1. Alternative Approaches. Before we proceed to the properties of the test, it is instructive to consider the available alternatives. First, we discuss why the naive approach of conducting a conventional nonparametric significance test using the mismeasured regressors is a poor choice. In this case we would test

$$H_0 : E\left[Y|W_{(1)} = w_{(1)}, W_{(2)} = w_{(2)}\right] = E\left[Y|W_{(1)} = w_{(1)}\right] \text{ for almost every } (w_{(1)}, w_{(2)}) \in \mathbb{R}^d.$$

Chesher (1991) provides a simple relation between $E[Y|W = w]$ and $m(w)$,

$$E[Y|W = w] \quad = \quad m(w) + \sum_{j=1}^{d} \frac{\sigma^2_{\epsilon_j}}{2} \left(m^{(j)(j)}(w) + 2m^{(j)}(w)(\ln(f_X(w))^{(j)} \right) + o(\sigma^2_{\epsilon_j}),$$

where $\sigma^2_{\epsilon_j}$ is the variance of the measurement error associated with the j^{th} regressor and for any function, $g(\cdot)$, $g^{(j)}(\cdot)$ and $g^{(j)(j)}(\cdot)$ denote the first and second derivative with respect to the j^{th} argument. Although this expression is derived under small measurement error asymptotics, it gives a good insight into the problem of conducting a nonparametric significance test using mismeasured variables. Instead of investigating the distance $m(w) - r(w_{(1)})$, the naive approach would consider

$$\begin{aligned}\left(m(w) - r(w_{(1)})\right) + \sum_{j=1}^{d} \frac{\sigma^2_{\epsilon_j}}{2} \left(m^{(j)(j)}(w) - r^{(j)(j)}(w_{(1)}) \right) \\ + \sigma^2_{\epsilon_j} \left((\ln(f_X(w))^{(j)} m^{(j)}(w) - (\ln(f_{X_1}(w_{(1)}))^{(j)} r^{(j)}(w_{(1)}) \right) + o(\sigma^2_{\epsilon_j}),\end{aligned}$$

where $r^{(j)}(\cdot) \equiv 0$ and $r^{(j)(j)}(\cdot) \equiv 0$ for $j > d_1$. Notice that even if $m(w) = r(w_{(1)})$, in general

$$\sum_{j=1}^{d} \sigma^2_{\epsilon_j} \left((\ln(f_X(w))^{(j)} m^{(j)}(w) - (\ln(f_{X_1}(w_{(1)}))^{(j)} r^{(j)}(w_{(1)}) \right) \neq 0$$

and the test has incorrect size and is inconsistent.

A second potential alternative to our test is to linearise the model by taking, for example, a finite polynomial in the regressors, estimate the regression coefficients using an IV approach, and conduct a Wald test. Note that for each transformation of each variable, the instrument, typically a repeated measurement, is the analogous transformation. Of course, this is a simplification of the model we present and would not lead to a consistent test. However, for practical purposes we may hope that this approach fares reasonably well. Unfortunately, as our simulation results indicate in Section 3.4, this is not the case. The reason is twofold: nonlinear transformations of mismeasured regressors generally exacerbate the measurement error problem, additionally the strength of the instruments typically deteriorates with nonlinear transformations.

Finally, it may be tempting to appeal to the common textbook wisdom that measurement error causes attenuation bias, hence any significance test in the presence of measurement error is simply overly cautious. Unfortunately such an argument fails to hold once any additional variables are included or any nonlinearities are added.

3.3. Asymptotic Properties

3.3.1. Distribution Under H_0. In this section we derive the asymptotic distribution of CM_n under both the null hypothesis and a Pitman local alternative. We proceed by first clarifying the asymptotic distribution of $\hat{T}_n(\xi)$ before using the continuous mapping theorem to derive the asymptotic distribution of CM_n.

To simplify our analysis we will use product kernels of the following form. As in Masry (1993), let $\tilde{K}(\cdot)$ be a univariate kernel and $\tilde{K}^{\text{ft}}(\cdot)$ denote its Fourier transform. Define the univariate deconvolution kernel as

$$\tilde{\mathcal{K}}_b(x_j) = \frac{1}{2\pi b} \int e^{-ita} \frac{\tilde{K}^{\text{ft}}(t)}{\tilde{f}^{\text{ft}}_{\epsilon_j}(t/b)} dt.$$

Finally, set $K(x) = \prod_{j=1}^{\dim(x)} \tilde{K}(x_j)$ and $\mathcal{K}_b(x) = \prod_{j=1}^{\dim(x)} \tilde{\mathcal{K}}_b(x_j)$. Since we assume that ϵ is vector valued with independent elements, we can write $f^{\text{ft}}_{\epsilon}(t) = \prod_{j=1}^{\dim(t)} \tilde{f}^{\text{ft}}_{\epsilon_j}(t_j)$, where $\tilde{f}^{\text{ft}}_{\epsilon_j}(\cdot)$ is the Fourier transform of ϵ_j. Together these imply $K^{\text{ft}}(x) = \prod_{j=1}^{\dim(x)} \tilde{K}^{\text{ft}}(x_j)$. [3]

Throughout this paper we use the notation $f(b) \sim g(b)$ to mean $f(b)/g(b) \to 1$ as $b \to 0$ and $||\cdot||_1$, and $||\cdot||_\infty$ to denote the L_1 norm and the supremum norm, respectively. We impose the following assumptions.

[3] Notice that it would be straightforward to adjust our estimator to allow for a combination of correctly measured and mismeasured regressors by replacing the deconvolution kernel within the product with a standard kernel; analogous results could be obtained. Furthermore, these correctly measured regressors may be discrete. However, our theory does not allow for discrete mismeasured regressors as this constitutes a form of nonclassical measurement error.

Assumption 6.

(i): $\{Y_i, W_i\}_{i=1}^n$ *and* $\{\epsilon_i\}_{i=1}^n$ *are i.i.d. where* ϵ *is mutually independent as well as independent of* (Y, X) *and has a known density* $f_\epsilon(\cdot)$.

(ii): $\sup_{\xi\in\Xi} ||\mathcal{W}(\cdot;\xi)||_\infty^2 < \infty$ *with* $\Xi \subseteq \mathbb{R}^d$ *a compact set with non-empty interior.*

(iii): $\tilde{K}(\cdot)$ *is a second-order kernel and is infinitely differentiable. In particular,* $\tilde{K}^{\mathrm{ft}}(\cdot)$ *is compactly supported on* $[-1, 1]$, *symmetric around zero (i.e.,* $\tilde{K}^{\mathrm{ft}}(t) = \tilde{K}^{\mathrm{ft}}(-t)$*), bounded, and satisfies*

$$\int y^k \tilde{K}(y)dy \begin{cases} 1 & \textit{for } k = 0, \\ 0 & \textit{for } k = 1, \end{cases}$$

and $\int y^2 \tilde{K}(y)dy \neq 0$.

(iv): $M(\cdot) \equiv f_X(\cdot)m(\cdot)$ *and* $R(\cdot) \equiv f_{X_{(1)}}(\cdot)r(\cdot)$ *have continuous and bounded second derivatives.*

(v): *Each of the following are finite:* $E\left[Y^4\right]$, $E\left[M(X)^2\right]$, $E\left[R(X_{(1)})^2\right]$.

(vi): *As* $n \to \infty$ *it holds that* $nb^{(d+d_1)\frac{5}{4}} \to \infty$ *and* $nb^4 \to 0$.

Assumption 6 (i) is common in the literature of classical measurement error. The case of unknown $f_\epsilon(\cdot)$ is deferred until Section 3.3.4. For Assumption 6 (ii), since we have defined $\mathcal{W}(\cdot;\xi) = \bar{\mathcal{W}}(\Phi(\cdot);\xi)$, this condition is satisfied by all of the commonly used weight functions given in Section 3.2. Assumption 6 (iii) is fairly standard in deconvolution problems and is satisfied by the commonly used Sinc kernel or the kernel proposed in Fan (1992) which was briefly mentioned in 3.2. Assumption 6 (iv) gives smoothness restrictions on the functions $M(\cdot)$ and $R(\cdot)$, whilst Assumption 6 (v) contains standard assumptions on Y and the underlying regression functions. The first condition of Assumption 6 (vi) is required to ensure the error of the Hoeffding projection is asymptotically negligible, the second is required to remove the bias from the nonparametric estimators. The second rate can be generalised to $nb^{2q} \to 0$ where q is the order of the kernel (we have used a simple second-order kernel), i.e. higher order kernels can be used to remove the bias more quickly. This will prove particularly relevant when working with high dimensional data since the first condition in (vi) must hold simultaneously.

As is typical in the nonparametric measurement error literature, we consider two separate cases characterised by bounds on the decay rate of the tail of the characteristic function of the measurement error, $\tilde{f}^{\text{ft}}_{\epsilon_j}(\cdot)$. In each case we introduce some additional assumptions. For the ordinary smooth case we impose the following.

Assumption 7.

(i): *$f^{\text{ft}}_{\epsilon}(t) \neq 0$ for all $t \in \mathbb{R}^d$ and there exist finite constants $C_0, \ldots, C_\alpha$ with $C_0 \neq 0$ and $\alpha > 0$ such that*

$$\tilde{f}^{\text{ft}}_{\epsilon_j}(s) \sim \frac{1}{\sum_{v=0}^{\alpha} C_v |s|^v},$$

for all $1 \leq j \leq d$ as $|s| \to \infty$.

Assumption 7 (i) is the ordinary smooth condition. Specifically, it requires that $\tilde{f}^{\text{ft}}_{\epsilon_j}(s)$ decays to zero at a polynomial rate as $|s| \to \infty$. Examples of densities that are ordinary smooth are Laplace and gamma. Notice that this is slightly more general than the typical assumption that is seen in the literature, $\tilde{f}^{\text{ft}}_{\epsilon_j}(s) \sim C|s|^{-\alpha}$.

For the second case, known as supersmooth measurement error, we impose the following assumption.

Assumption 8.

(i): *$f^{\text{ft}}_{\epsilon}(t) \neq 0$ for all $t \in \mathbb{R}^d$ and there exist positive constants C, μ, $\gamma_0 \leq 0$ and $\gamma > 0$ such that*

$$\tilde{f}^{\text{ft}}_{\epsilon_j}(s) \sim C|s|^{\gamma_0} e^{-\mu|s|^\gamma},$$

for all $1 \leq j \leq d$ as $|s| \to \infty$ with $\gamma - \gamma_0$ an integer.

Assumption 8 (i) requires $\tilde{f}^{\text{ft}}_{\epsilon_j}(s)$ to decay to zero at an exponential rate as $|s| \to \infty$. The most common example of a density satisfying this supersmooth assumption is the normal density, where $C = 1$, $\gamma_0 = 0$, $\gamma = 2$, and $\mu = \frac{\sigma^2}{2}$. The majority of conventional distributions satisfy the integer constraint. Also, as opposed to some settings with supersmooth measurement error (see for example Van Es and Uh, 2005), the Cauchy distribution is not excluded from our analysis.

The asymptotic distributions of CM_n under the null hypothesis for both the ordinary smooth and the supersmooth cases are given by the following theorem.

THEOREM 5.

(i): *Suppose that Assumptions 6 and 7 hold true, then under* H_0,

$$nCM_n \to_d \int |Z_{O,\infty}(\xi)|^2 \, d\mu(\xi) \sim \sum_{i=1}^{\infty} \lambda_{O,i}\nu_i^2.$$

(ii): *Suppose that Assumptions 6 and 8 hold true, then under* H_0,

$$nCM_n \to_d \int |Z_{S,\infty}(\xi)|^2 \, d\mu(\xi) \sim \sum_{i=1}^{\infty} \lambda_{S,i}\nu_i^2.$$

Where $Z_{O,\infty}(\cdot)$ and $Z_{S,\infty}(\cdot)$ are zero mean Gaussian processes on $L_2(\Xi, \mu)$, with covariance functions $V_O : \Xi \times \Xi \to \mathbb{R}^+$ and $V_S : \Xi \times \Xi \to \mathbb{R}^+$, respectively, where V_O and V_S are defined in Appendix A. ν_i are i.i.d. $N(0,1)$ random variables and $\lambda_{O,i}$ and $\lambda_{S,i}$ are the solutions to the eigenvalue problems

$$\begin{aligned} \int V_O(\xi, \xi')\psi_{O,i}(\xi')d\mu(\xi') &= \lambda_{O,i}\psi_{O,i}(\xi), \\ \int V_S(\xi, \xi')\psi_{S,i}(\xi')d\mu(\xi') &= \lambda_{S,i}\psi_{S,i}(\xi), \end{aligned}$$

respectively. The eigenvalues $\lambda_{O,i}$ and $\lambda_{S,i}$ are real valued, non-negative and satisfy $\sum_{i=1}^{\infty} \lambda_{O,i} < \infty$ and $\sum_{i=1}^{\infty} \lambda_{S,i} < \infty$, respectively.

Theorem 5 shows that in each case the test statistic converges to a weighted sum of chi-squared random variables. Unlike typical results in the deconvolution estimation and inference literature we are able to achieve $\sqrt{n}$ rates of convergence. To the best of our knowledge, this is the first case in which parametric rates of convergence have been obtained when using non-parametric estimation in the presence of supersmooth measurement error. However, two notable papers deserve mention here. Hall and Ma (2007) develop a non-smoothing specification test which is able to achieve $\sqrt{n}$ rates of convergence but does not involve any nonparametric estimation. Fan (1995) obtains $\sqrt{n}$ convergence for average derivative estimators for ordinary smooth measurement error, although does not extend this result to the supersmooth case.

There is a strong link between average derivative estimators - of the type studied by Powell, Stock and Stoker (1989), for example - and non-smoothing tests. We exploit this connection and use a similar approach to Fan (1995) in the derivation of the asymptotic properties of our test.

We extend this approach to the supersmooth case by noticing that supersmooth error can be thought of as an ordinary smooth problem with $\alpha = \infty$.

It is instructive to understand why we are able to achieve parametric rates in such nonparametric problems. Heuristically, nonparametric estimators achieve slower rates of convergence, typically $\sqrt{nb}$ for univariate problems, because they effectively only use a window of nb observations at each point of the estimation. Non-smoothing tests are able to regain $\sqrt{n}$ rates by averaging these nonparametric estimators over the full range of the data and so use all n observations in the final test statistic. The same reasoning explains why nonparametric average derivative estimators are also able to achieve a $\sqrt{n}$ rate of convergence. However, the semi-smoothing approach of this paper is slightly different to each of these cases since we are unable to average over all observations as they are unobservable. Nonetheless, by integrating over the range of all possible values the regressors may take, we implicitly draw all observations of the mismeasured variable into the final test statistic and so recover the $\sqrt{n}$ convergence rate. When viewed in this light, it is perhaps not so surprising that even in the supersmooth case we can escape the curse of dimensionality.

This theorem also shows that the asymptotic distribution of the test does not depend on the bandwidth. As such, providing Assumption 6 (vi) is satisfied, we hope the test will show little dependence on the bandwidth in finite samples and negate the need for an 'optimal' choice.

3.3.2. Distribution Under a Sequence of Local Alternatives. To study the power properties of the test, we assume a local, linear alternative of the form

$$\mathrm{H}_{1n} : m(x) = r\left(x_{(1)}\right) + \frac{1}{\sqrt{n}}\Delta(x), \text{ for almost every } x \in \mathbb{R}^d$$

where $\Delta(\cdot)$ is a bounded, non-zero function. The local power properties are given by the following theorem.

THEOREM 6.

(i): *Suppose that Assumptions 6 and 7 hold true, then under* H_{1n}*,*

$$nCM_n \to_d \int \left|\tilde{Z}_{O,\infty}(\xi)\right|^2 d\mu(\xi) \sim \sum_{i=1}^{\infty} \left(\bar{\Delta}_{O,i} + \sqrt{\tilde{\lambda}_{O,i}}\nu_i\right)^2.$$

(ii): *Suppose that Assumptions 6 and 8 hold true, then under* H_{1n}*,*

$$nCM_n \to_d \int \left|\tilde{Z}_{S,\infty}(\xi)\right|^2 d\mu(\xi) \sim \sum_{i=1}^{\infty} \left(\bar{\Delta}_{S,i} + \sqrt{\tilde{\lambda}_{S,i}}\nu_i\right)^2.$$

Where $\tilde{Z}_{O,\infty}(\xi)$ and $\tilde{Z}_{S,\infty}(\xi)$ are Gaussian processes with mean functions $\bar{\Delta}_{O,}(\cdot)$ and $\bar{\Delta}_{S,}(\cdot)$, respectively, each defined in Appendix A, and covariance functions $\tilde{V}_O : \Xi \times \Xi \to \mathbb{R}^+$ and $\tilde{V}_S : \Xi \times \Xi \to \mathbb{R}^+$, respectively, again with each defined in Appendix A. $\bar{\Delta}_{O,i} \equiv \int \bar{\Delta}_{O,}(\xi)\tilde{\psi}_{O,i}(\xi)d\mu(\xi)$ where $\tilde{\psi}_{O,i}(\cdot)$ are the eigenfunctions of the equation

$$\int \tilde{V}_O(\xi,\xi')\tilde{\psi}_{O,i}(\xi')d\mu(\xi') = \tilde{\lambda}_{O,i}\tilde{\psi}_{O,i}(\xi),$$

and $\bar{\Delta}_{S,i}$ is defined analogously. As before, the eigenvalues $\tilde{\lambda}_{O,i}$ and $\tilde{\lambda}_{S,i}$ are real valued, non-negative and satisfy $\sum_{i=1}^{\infty} \tilde{\lambda}_{O,i} < \infty$ and $\sum_{i=1}^{\infty} \tilde{\lambda}_{S,i} < \infty$ respectively.

Theorem 6 shows that under both ordinary smooth and supersmooth measurement error our test is able to detect local, linear alternatives drifting at the rate $n^{-1/2}$.

It is our belief that nonparametric measurement error techniques are underused in applied work, in part, because of the very slow rates of convergence that are typically attained. In particular, in perhaps the most likely setting of Gaussian measurement error, convergence is usually at the rate $\ln(n)$. We hope that the results presented here encourage the use of this test in future applied work.

3.3.3. Dependent Data. To increase the applicability of our test, it is important to allow for applications involving time series data. In this section we extend our asymptotic results to permit weakly dependent data.

To be precise, we assume that the data, in particular the correctly measured regressors and the dependent variable, come from a strictly stationary, absolutely regular process. We borrow the notation from Robinson (1989) to define the degree of dependence. Let M_a^b denote the σ-

algebra of events generated by $V_a, \ldots, V_b$, for $-\infty \leq a \leq b \leq \infty$, where $V = (Y, X)$. We assume

$$\beta(j) \equiv E\left\{\sup_{A \in M_j^{\infty}} \left|Pr\left(A \middle| M_{-\infty}^{0}\right) - Pr(A)\right|\right\} \to 0$$

as $j \to \infty$. Absolutely regular processes can be seen as lying somewhere between uniformly and strongly mixing processes in terms of dependence.

Notice that the dependence is in the true regressor and not the measurement error and we continue to impose the assumption of classical measurement error. It is possible to also allow for dependence within the measurement error in a similar manner to the treatment of dependence in the regressors, in this case we would need to assume a smoothness condition on the joint distribution of ϵ_i and ϵ_j for $1 \leq i, j \leq n$, however we concentrate on the i.i.d. case for ease of derivations.

To show that our previous results continue to hold under weak dependence we require the following additional assumptions.

Assumption 9.

- **(i):** *$\beta(j) = O(j^{-\eta})$ for a particular $\eta > 0$ which is discussed in Appendix A.*
- **(ii):** *For some $\varsigma > 0$, as $n \to \infty$ it holds that $n^{1-\frac{\varsigma}{2}} b^{(d+d_1)\frac{5}{4}} \to \infty$.*
- **(iii):** $\sup_{\xi \in \Xi} ||\mathcal{W}(\cdot;\xi)||_{\infty}^{2+\delta} < \infty$ *for some $\delta > 0$ with $\Xi \subseteq \mathbb{R}^d$ a compact set with non-empty interior.*
- **(iv):** *$M_{X_j|W_i}(x|w) \equiv m(x) f_{X_j|W_i}(x|w)$ and $R_{X_{(1)j}|W_{(1)i}}(x_{(1)}|w_{(1)}) \equiv r(x_{(1)}) f_{X_{(1)j}|W_{(1)i}}\left(x_{(1)}|w_{(1)}\right)$ have continuous and bounded second derivatives.*

Assumption 9 (i) concerns the degree of dependence between events separated in time; the larger η the more quickly the dependence decays to zero. Assumptions 9 (ii)-(iv) require a slight strengthening of the corresponding Assumptions 6 (ii), (iv) and (vi). Notice that since we retain the assumption of classical measurement error we retain the product form of our deconvolution kernel which acts to simplify our theoretical analysis.

THEOREM 7.

(i): *Suppose that Assumptions 6, 7 and 9 hold true, then under* H_0*,*

$$nCM_n \to_d \int |Z_{OT,\infty}(\xi)|^2 \, d\mu(\xi) \sim \sum_{i=1}^{\infty} \lambda_{OT,i}\nu_i^2,$$

and under H_1

$$nCM_n \to_d \int \left|\tilde{Z}_{OT,\infty}(\xi)\right|^2 d\mu(\xi) \sim \sum_{i=1}^{\infty} \left(\bar{\Delta}_{OT,i} + \sqrt{\tilde{\lambda}_{OT,i}}\nu_i\right)^2.$$

(ii): *Suppose that Assumptions 6, 8 and 9 hold true, then under* H_0*,*

$$nCM_n \to_d \int |Z_{ST,\infty}(\xi)|^2 \, d\mu(\xi) \sim \sum_{i=1}^{\infty} \lambda_{ST,i}\nu_i^2,$$

and under H_1

$$nCM_n \to_d \int \left|\tilde{Z}_{ST,\infty}(\xi)\right|^2 d\mu(\xi) \sim \sum_{i=1}^{\infty} \left(\bar{\Delta}_{ST,i} + \sqrt{\tilde{\lambda}_{ST,i}}\nu_i\right)^2.$$

Each object is defined analogously to Theorems 1 and 2, the only substantial differences are the limiting covariance functions as defined in Appendix A.

3.3.4. Unknown f_ϵ. In this section we discuss how, by estimating $f_\epsilon^{\mathrm{ft}}(\cdot)$, it can be possible to drop the rather restrictive assumption of a known measurement error density. Unsurprisingly, we need additional information. Typically this comes in the form of two repeated measurements, say W and W^r, or may alternatively come from a validation data set. We abstract from the estimation of $f_\epsilon^{\mathrm{ft}}(\cdot)$ and simply require some consistent estimator, $\hat{f}_\epsilon^{\mathrm{ft}}(\cdot)$. There are two leading cases. The first uses repeated measurements of the form

$$\begin{aligned} W &= X + \epsilon \\ W^r &= X + \epsilon^r \end{aligned}$$

where ϵ and ϵ^r are identically distributed with zero mean and $(X, \epsilon, \epsilon^r)$ are mutually independent. Under the assumption that the density $f_\epsilon(\cdot)$ is symmetric around zero, Delaigle, Hall and Meister

(2008) propose the following estimator

$$\hat{f}_{\epsilon}^{\mathrm{ft}}(t) = \left| \frac{1}{n} \sum_{i=1}^{n} \cos\{t(W_i - W_i^r)\} \right|^{1/2}.$$

Common examples of such data are found in medical studies where measurements and tests on patients are repeated at different points in time, for example systolic blood pressure measurements. It is also becoming increasingly popular to ask the same questions multiple times in social and economic surveys to obtain repeated measurements. Other examples of repeated data include different IQ tests (or other aptitude tests) which can be used as repeated measurements of true intelligence, as well as GDP and GDI acting as two mismeasured versions of economic activity (see Delaigle, Hall and Meister, 2008, for further examples).

The second estimator is due to Li and Vuong (1998) and requires weaker assumptions. Specifically, the repeated measurements can take the same form but where ϵ and ϵ^r need not be identically distributed nor have densities that are symmetric around zero, however, they must still have zero mean and $(X, \epsilon, \epsilon^r)$ must still be mutually independent. Naturally, the estimation procedure is more complex than the case of Delaigle, Hall and Meister (2008), the interested reader is referred to Li and Vuong (1998) for further details.

The following theorem shows the asymptotic equivalence of the test using $f_{\epsilon}^{\mathrm{ft}}(\cdot)$ with the test using $\hat{f}_{\epsilon}^{\mathrm{ft}}(t)$. We first introduce the following additional assumptions.

Assumption 10.

(i): $\max_{t \leq \frac{1}{b}} \left| \hat{f}_{\epsilon}^{\mathrm{ft}}(t) - f_{\epsilon}^{\mathrm{ft}}(t) \right| = o_p(1)$.

Assumption 10 (i) is satisfied by the estimators of Li and Vuong (1998) (see Lemma 4 in Evdokimov, 2010) and Delaigle, Hall and Meister (2008).

Theorem 8.

(i): *Suppose that Assumptions 6, 10 and either 7 or 8 hold true, then Theorems 5 and 6 continue to hold if $f_{\epsilon}^{\mathrm{ft}}(\cdot)$ is replaced with $\hat{f}_{\epsilon}^{\mathrm{ft}}(t)$.*

(ii): *Suppose that Assumptions 6, 9, 10 and either 7 or 8 hold true, then Theorem 7 continues to hold if $f_{\epsilon}^{\mathrm{ft}}(\cdot)$ is replaced with $\hat{f}_{\epsilon}^{\mathrm{ft}}(t)$.*

3.3.5. Bootstrap. The asymptotic distributions derived in Theorem 5 can be used to obtain critical values. However, as explained in Bierens and Ploberger (1997), the eigenvalues depend on the covariance function which in turn depends on the underlying distribution of the data. As such, the asymptotic distributions are case dependent and challenging to estimate in practice. Given this difficulty, it may be wiser to implement a bootstrap procedure.

Measurement error models provide quite a challenge for bootstrap procedures because neither the true regressor nor the measurement error is observable. Any residual based bootstrap approach is infeasible in a measurement error context since the true regressors are needed to construct the residuals. It would be possible to follow an approach similar to Hall and Ma (2007): estimating the density of the true regressor using deconvolution techniques, applying a wild bootstrap approach for the measurement error, and sampling from these respective densities. However, the estimated density will suffer from the slow rates of convergence associated with deconvolution estimation and the approach is very computationally expensive. In addition, the choice of several tuning parameters are needed. Instead, we suggest a simple alternative based on a pairs bootstrap.

Recall, we write our statistic as

$$nCM_n \quad = \quad \int_{\Xi} \left| \sqrt{n}\hat{T}_n(\xi) \right|^2 d\mu(\xi).$$

We can construct a bootstrap sample, $\{Y_i^*, W_i^*\}_{i=1}^n$, by resampling with replacement from $\{Y_i, W_i\}_{i=1}^n$. To impose the null hypothesis, construct $\tilde{T}_n^*(\xi) \equiv \hat{T}_n^*(\xi) - \hat{T}_n(\xi)$, where $\hat{T}_n^*(\xi)$ is defined in the same manner as $\hat{T}_n(\xi)$ but using $\{Y_i^*, W_i^*\}_{i=1}^n$. Finally, the bootstrap test statistic is given by

$$nCM_n^* \quad = \quad \int_{\Xi} \left| \sqrt{n}\tilde{T}_n^*(\xi) \right|^2 d\mu(\xi).$$

When working with dependent data we must adapt the above procedure. We use the stationary bootstrap of Politis and Romano (1994) to obtain our bootstrap sample and proceed as above. We briefly outline the stationary bootstrap procedure here for ease of reference. The data, $\{Z_t\}_{t=1}^T = \{Y_t, W_t\}_{t=1}^T$, is strictly stationary and absolutely regular. Let

$B_{t,s} = \{Z_t, Z_{t+1}, \ldots, Z_{t+s-1}\}$ be a block of data of length s. For Z_k with $k > T$, $Z_k = Z_{k \mod T}$ and $Z_0 = Z_T$. Let $L_1, L_2, \ldots$ be a sequence of i.i.d. geometric random variables independent of Z_t, such that $Pr\{L_i = m\} = (1 - p_T)^{(m-1)} p_T$ for $m = 1, 2, \ldots$, where $p_T \in [0, 1]$ depends on the sample size. Finally, let $I_1, I_2, \ldots$ be a sequence of i.i.d. random variables with a discrete uniform distribution on $\{1, \ldots, T\}$ independent of Z_t and L_t. To generate the bootstrap sample, $\{Z_t^*\}_{t=1}^T$, sample a sequence of blocks of random length, $B_{I_1,L_1}, B_{I_2,L_2}, \ldots$. The first T observations from this sequence of blocks creates the bootstrap sample.

Proposition 1.

(i): *Suppose that Assumptions D, and either Assumption O or S hold true, then the asymptotic distribution of CM_n under the the null hypothesis is the same as the asymptotic distribution of CM_n^* conditional on $\{Y_i, W_i\}_{i=1}^n$.*

(ii): *Suppose that Assumptions D, T, and either Assumption O or S hold true, then, using the stationary bootstrap, the asymptotic distribution of CM_n under the the null hypothesis is the same as the asymptotic distribution of CM_n^* conditional on $\{Y_i, W_i\}_{i=1}^n$.*

3.4. Simulation

To study the small sample properties of our test we conduct a Monte Carlo experiment. Since this is the first nonparametric significance test designed to account for measurement error, it is difficult to give a direct comparison to any existing tests. However, we report results for the test of Delgado and Manteiga (2001) (DM henceforth) as well as a Wald test based on an IV regression with functional form: $\beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_2 + \beta_4 X_2^2$. A repeated measurement, as well as its square, are used as the instruments. We should make clear that the test of DM is not designed for a measurement error setting whilst the IV test, although able to accommodate measurement error, is a parametric test.

We concentrate on a regression with two regressors. The true, unobservable regressors (X_1, X_2) are each distributed independently $U[0, 1]$. The contaminated regressors are given by $W_k = X_k + \epsilon_k$, for $k = 1, 2$. We generate a second independent measurement of X_k given by $W_k^r = X_k + \epsilon_k^r$ where ϵ_k^r is distributed independently and identically to ϵ_k. For the ordinary smooth case, we take ϵ to be drawn from the Laplace distribution with variance equal to half the variance of X. For the supersmooth case, we use a zero mean Gaussian error with variance also equal to half that of X. Hence, the signal to noise ratio in both cases is $\frac{2}{3}$. Since both distributions are symmetric around 0 we can use the repeated data to estimate $f_{\epsilon}^{ft}(\cdot)$ using the estimator of Delaigle, Hall and Meister (2008).

We consider several data generating processes

$$Y = 1 + X_1 + U \qquad DGP(1)$$

$$Y = 1 + X_1 + 10\sin(2\pi X_2)^2 + U \qquad DGP(2)$$

$$Y = 1 + X_1 + 10(X_1 - X_1^2)(X_2 - X_2^2) + U \qquad DGP(3)$$

$$Y = 1 + X_1 + 10(X_2 - X_2^2) + U \qquad DGP(4)$$

where $U \sim N(0,1)$. Clearly, DGP(1) corresponds to the null model, whilst DGP (2)-(4) represent a range of possible deviations from the null.

For our weighting function we choose $\mathcal{W}(\cdot;\xi) = \exp(\xi'\cdot)$, which satisfies Assumption 6 (ii); results were similar for other commonly used weighting functions. For all simulations we use the Sinc kernel

$$K(x) = \frac{\sin(x)}{x}$$

which satisfies Assumptions 6 (iii). We report results for a small ($n = 100$) and a moderate ($n = 200$) sample size as well as a range of bandwidths. Specifically, for the ordinary and supersmooth cases, we use $b_0 = \left(\frac{1}{n}\right)^{1/2(d+d_1)}$ which satisfies Assumption 6 (vi), but consider a range of bandwidths around this choice, allowing us to analyse the sensitivity of our test to the bandwidth. For the test of DM we use a similar set of bandwidths based on the rule-of-thumb $b_{DM} = \left(\frac{1}{n}\right)^{1/3d_1}$; this is taken from the simulations carried out in DM. The critical values for our test are constructed using the i.i.d. bootstrap procedure outlined in Section 3.3.5 with 499 replications. For DM we use the bootstrap procedure denoted as C_n^{**} in their paper. The perturbation random variable ν^* for their bootstrap is the Mammen two-point distribution. All results are based on 1000 Monte Carlo replications.

Table 1 shows results for the level accuracy of the three tests. The column labeled 'My Test' reports results for the test proposed in this chapter, the column labeled 'DM' refers to the test of Delgado and Manteiga (2001), and 'IV' displays results for the Wald test based on the IV quadratic regression using the repeated measurement. Tables 2 - 4 display the power results for DGP(2) - (4), respectively.

Table 1: $Y = 1 + X_1 + U$

		My Test			**DM**			**IV**
Ordinary Smooth		**Bandwidth**						
n	**Level**	$b_0 - 0.25$	b_0	$b_0 + 0.25$	$b_{DM} - 0.1$	b_{DM}	$b_{DM} + 0.1$	
100	**5%**	4.4	4.6	4.8	5.6	6.0	7.8	4.5
	10%	9.3	8.8	9.5	11.3	11.2	16.2	9.9
200	**5%**	4.3	4.8	5.2	5.8	5.2	6.8	5.6
	10%	9.9	11.6	10.9	11.5	10.2	14.7	9.6
Super Smooth								
100	**5%**	4.9	3.8	3.9	5.3	4.5	6.4	4.6
	10%	10.4	9.9	10.1	9.8	9.6	13.4	9.2
200	**5%**	4.5	5.6	5.5	5.3	5.7	8.2	5.7
	10%	9.7	10.4	10.0	11.6	11.9	15.7	10.8

Table 2: $Y = 1 + X_1 + 10 \sin(2\pi X_2)^2 + U$

		My Test			**DM**			**IV**
Ordinary Smooth		**Bandwidth**						
n	**Level**	$b_0 - 0.25$	b_0	$b_0 + 0.25$	$b_{DM} - 0.1$	b_{DM}	$b_{DM} + 0.1$	
100	**5%**	13.5	18.0	17.8	6.5	6.0	5.8	4.3
	10%	22.7	29.2	29.8	14.4	12.5	11.4	8.1
200	**5%**	16.6	30.0	30.0	10.3	9.2	8.2	5.8
	10%	27.4	41.3	42.2	19.0	18.1	17.1	12.9
Super Smooth								
100	**5%**	11.3	17.6	18.8	7.2	6.4	5.8	4.8
	10%	21.0	28.6	29.0	13.7	11.9	12.1	8.9
200	**5%**	15.7	30.0	31.4	7.5	6.1	6.0	8.8
	10%	26.1	43.2	44.8	12.7	12.0	12.1	19.0

Table 3: $Y = 1 + X_1 + 10 \left(X_1 - X_1^2\right) \left(X_2 - X_2^2\right) + U$

		My Test			**DM**			**IV**
Ordinary Smooth		**Bandwidth**						
n	**Level**	$b_0 - 0.25$	b_0	$b_0 + 0.25$	$b_{DM} - 0.1$	b_{DM}	$b_{DM} + 0.1$	
100	**5%**	8.5	8.5	8.0	6.5	5.9	8.6	4.4
	10%	17.0	16.0	15.3	11.4	12.1	16.9	10.1
200	**5%**	13.3	13.9	12.9	7.5	5.6	8.4	5.0
	10%	22.8	21.5	21.2	14.4	11.9	16.4	9.2
Super Smooth								
100	**5%**	7.5	9.2	9.0	5.2	4.1	6.2	4.4
	10%	16.4	15.5	14.9	9.2	8.9	13.6	8.6
200	**5%**	9.2	11.9	12.0	5.8	5.3	9.2	5.5
	10%	17.4	22.1	21.1	11.8	12.4	17.7	10.6

Table 4: $Y = 1 + X_1 + 10\left(X_2 - X_2^2\right) + U$

		My Test			**DM**			**IV**
Ordinary Smooth		**Bandwidth**						
n	**Level**	$b_0 - 0.25$	b_0	$b_0 + 0.25$	$b_{DM} - 0.1$	b_{DM}	$b_{DM} + 0.1$	
100	**5%**	81.7	81.3	79.8	29.4	21.1	16.1	13.7
	10%	89.4	89.6	88.7	51.8	42.1	34.6	23.7
200	**5%**	96.5	97.8	96.3	73.1	59.5	43.5	44.3
	10%	98.1	99.2	98.4	89.6	81.8	72.1	58.8
Super Smooth								
100	**5%**	78.5	84.4	84.1	17.6	12.5	10.4	19.8
	10%	88.3	91.1	91.1	35.5	27.0	22.1	30.9
200	**5%**	91.8	98.5	98.5	52.1	40.6	29.1	71.0
	10%	96.4	99.1	99.1	76.1	63.1	50.3	80.6

The results appear to reflect the theoretical findings and look encouraging. The bootstrap procedure controls the size of the test well for both Laplacian and Gaussian measurement error with only a slight dependence on the bandwidth. The DM test also has relatively good size control despite being invalid. This is likely due to the bootstrap procedure being used. However, there is some size distortion for larger bandwidths which is also found and discussed in DM. The t-test based on the IV seems slightly undersized, although as the sample size was increased it did approach the nominal level.

In each of the three alternatives considered the test proposed in this chapter dominates the two alternatives. Whilst this is unsurprising for DGP(2) and DGP(3), given the alternative tests are not designed for these situations, for DGP(4) the IV test is based on the correct parametric specification. This should act as a benchmark with which to compare our nonparametric test, however, our test clearly dominates. This reflects the arguments given in Section 3.2.1. As can be seen in Table 4, as we increase the sample size the IV test does gain considerable power but still lags behind the test proposed in this chapter. In general, increasing the amount of measurement error or increasing the degree of nonlinearity in the model, increases the gap between our test and the two alternatives. Our test shows a slight dependence on the bandwidth despite the theoretical results showing there should be none. Of course, this dependence may simply be a result of using a finite sample. However, there appears to be a consistent pattern across all models and measurement error specifications where the power drops relatively sharply for the smallest bandwidth. This may be a result of a violation of Assumption 6 (vi) ($nb^{(d+d_1)\frac{5}{4}} \to \infty$)

which requires the bandwidth to be large enough such that the approximation from the Hoeffding projection is asymptotically negligible.

3.5. Applications

3.5.1. Cognitive Ability. In this section we use our test to determine whether cognitive ability has a significant effect on a series of key socio-economic variables: income, life satisfaction, health and risk aversion. Each of these relationships have received varying degrees of attention in the past. However, little consideration has been given to either the effect of measurement error, caused by using proxies for true cognitive ability, or to allowing a nonparametric relationship. Notice that if we were instead interested in the effect of education on these variables holding constant cognitive ability, our test would be equally applicable. It should be stated at the outset that this section acts merely to give a flavour of the potential uses of our test and does not attempt to give an in-depth analysis of such questions; this would require an entire paper in and of itself.

To tackle these questions we use the novel data set known as the 'Brabant survey'. The data consists of information on nearly 3000 individuals from the Dutch province of North Brabant. In 1952 a survey was taken of nearly 6000 12 year old children. Their names and addresses were kept and 30 years later Joop Hartog tracked down and reinterviewed almost 3000 of the original individuals. The data covers family background and three measures of IQ taken when the participants were 12 years old as well as information on their education, income, marital status, number of children, health, life satisfaction and a measure of their risk aversion taken in follow up surveys in 1983 and 1993. Education is the highest level of education achieved measured on a 4 point scale, whilst family background is based on the father's occupation measured on a 3 point scale. The first IQ test is the Raven Progressive Matrices test designed to measure general intelligence, the second is a verbal intelligence test, the third is an abstract thinking test. The health and life satisfaction variables are self-reported ratings on a scale of 1-10. Finally, the measure of risk aversion is the Arrow-Pratt absolute measure calculated from prices given for a simple lottery.

The effect of cognitive ability on income has been studied extensively in the past. In general, results have shown that cognitive ability has a positive impact on earnings, see for example

Hernstein and Murray (1994) and Cawley, Heckman and Vytlacil (2001). One of the few papers to tackle the problem of measurement error in this setting is Heckman, Stixrud and Urzua (2006) who investigate the effects of cognitive and noncognitive ability on a range of social and economic outcomes and conclude that cognitive ability has a significant, positive effect on wages.

There is a plethora of research which looks at the relationship between education and health but there has been far less which has considered the role of cognitive ability in determining health outcomes. A notable exception is Conti, Heckman and Urzua (2011) who provide a thorough investigation of this topic, including allowing for measurement error in cognitive ability, however, their focus is on estimating the treatment effect of education. They find that cognitive ability, developed as early as age 10, is an important determinant of health at age 30, but these effects differ between men and women and between mental and physical health.

Research into the effect of cognitive ability on happiness, or life satisfaction, has predominantly been confined to the field of psychology and can be broadly split into two categories. The first investigates the effect at an individual level, the second looks at the aggregate level across nations, see Veenhoven and Choi (2012) for an aggregation of these results. Findings have been very mixed with both positive, negative and no effects being found.

Although there has been some work in the psychology literature, little attention has been given to the effect of cognitive ability on risk aversion by economists, despite its importance. Dohmen et al. (2010) is one of the few papers in the economics literature to look at this question and highlight its important implications. They collect and analyse their own data to find that more intelligent individuals are significantly less risk averse; this has important theoretical and empirical implications in, for example, contract designs and screening. However, their analysis does not account for measurement error and assumes a linear functional form for the regression.

In our study, we use the three IQ tests as repeated noisy measurements of true cognitive ability and use a factor model approach. Specifically, we assume

$$\begin{aligned} T_1 &= \alpha_1 C + \epsilon_1 \\ T_2 &= \alpha_2 C + \epsilon_2 \\ T_3 &= \alpha_3 C + \epsilon_3 \end{aligned}$$

where $T_1, \ldots, T_3$ denote IQ tests 1 - 3, C is latent cognitive ability, $\epsilon_1, \ldots, \epsilon_3$ denote the measurement errors for each test and $\alpha_1, \ldots, \alpha_3$ are the factor loadings. Without loss of generality we can normalise $\alpha_1 = 1$. To estimate α_2 and α_3, notice

$$\frac{Cov(T_2, T_3)}{Cov(T_1, T_2)} = \frac{\alpha_2 \alpha_3 Var(C)}{\alpha_2 Var(C)} = \alpha_3$$

and similarly for α_2. Sample counterparts can then be used to construct

$$\begin{aligned} T_1 &= C + \epsilon_1 \\ \frac{T_2}{\hat{\alpha}_2} &= C + \frac{\epsilon_2}{\hat{\alpha}_2} \\ \frac{T_3}{\hat{\alpha}_3} &= C + \frac{\epsilon_3}{\hat{\alpha}_3}. \end{aligned}$$

Notice, for example, if C and ϵ_2 are independent so too are C and $\frac{\epsilon_2}{\hat{\alpha}_2}$. To use the estimator proposed by Delaigle, Hall and Meister (2008) for the Fourier transform of the measurement error requires the error terms to be identically distributed. Since this is unlikely to hold in this case we use the estimator proposed in Li and Vuong (1998).

We use the i.i.d. bootstrap procedure discussed in Section 3.3.5, using the same parameter settings as in Section 3.4. We also use the same kernel as used in Section 3.4. For the other regressors we use a conventional product Gaussian kernel. For each dependent variable, in addition to cognitive ability we control for education, marital status, number of children, gender and, for all except the regression on income, income. All variables are standardised to have zero mean and unit variance.

Unfortunately, there is currently no theory to guide the choice of a data driven bandwidth for nonparametric testing with measurement error. However, Section 3.3 shows the asymptotic properties of our test do not depend on the choice of bandwidth, providing they satisfy Assumption 6 (vi). As such, we choose $b = \left(\frac{1}{n}\right)^{1/2(d+d_1)} \approx 0.65$ which satisfies this assumption. However, we consider a range of values around 0.65 to analyse the sensitivity of our results to the bandwidth. The p-values of our test are displayed in Table 5, along with the test of Delgado and Manteiga (2001) and t-tests based on IV and OLS quadratic regressions using IQ and IQ^2.

Table 5: Cognitive Ability (P-Values)

Dependent Variable	My Test			DM			IV	OLS
	Bandwidth							
	0.45	**0.65**	**0.85**	**0.35**	**0.55**	**0.75**		
Income	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000
Health	0.060	0.038	0.034	0.134	0.020	0.012	0.624	0.706
Life Satisfaction	0.286	0.022	0.021	0.236	0.096	0.082	0.161	0.738
Risk Aversion	0.477	0.310	0.273	0.910	0.874	0.770	0.908	0.855

Our results agree with the previous literature in finding that cognitive ability has a significant impact on income and health. We also find a significant relationship for life satisfaction which may help to add some more convincing evidence to this side of the debate. However, our findings on risk aversion disagree with the results of Dohmen et al (2010). Given the agreement in our results from the test of Delgado and Manteiga (2001), the IV and the OLS regressions, it would appear that this difference is driven by the data. It is interesting to see in the Health and Life Satisfaction regressions there are some clear differences in conclusions depending on whether a nonparametric test is used or not. Furthermore, for the effect on life satisfaction, it appears that in a nonparametric regression accounting for measurement error can lead to a change in conclusions when testing at the 5% significance level. Unfortunately, the dependence on the bandwidth rears its ugly head again. Similarly to the results in Section 3.4, it appears that the power of the test falls for a small bandwidth. In the case of risk aversion, although the same conclusion is reached with all tests, the difference in p-values is quite stark. It should also be emphasised that knowing the result from one, or a combination, of the alternative tests would not shed any light on the likely outcome of our test.

We provide regression plots of each of these relationships in Appendix 3.8 for the interested reader.

3.5.2. Inflation Expectations. What policy options does a central banker have when their hands are tied by the zero lower bound on nominal interest rates? This is currently a very important policy question in many developed economies. Several prominent commentators have suggested that future inflation expectations provide an alternative route for monetary policy to stimulate the current economy. For example, Paul Krugman has been a frequent advocate of this 'unconventional monetary policy' (see for example Krugman, 1998 and 2013), as well as Romer (2011) and Hall (2011) among many others. It has even been proposed, by Eggertsson (2008), that increases in inflation expectations were a key contributing factor to the end of the Great Depression, whilst Romer and Romer (2013) suggest that deflationary expectations were part of the cause. Correia, Farhi, Nicolini, and Teles (2013) formalise this idea and construct a framework to study the theoretical underpinnings of a relationship between inflation expectations and consumption at the zero lower bound.

The classic Euler Equation relating current and future consumption is

$$U'(C_t) = \beta U'(C_{t+1}) \frac{i_{t+1}}{\pi_{t+1}},$$

where $U'(\cdot)$ is the partial derivative of the utility function with respect to consumption, C_t is consumption in period t, β is the discount factor, and i_t and π_t are the nominal interest rate and inflation rate, respectively, in period t. In theory, higher expected future inflation should cause a tilting of consumption towards the present and away from the future through a relative cheapening of current consumption. However, empirical findings on this intertemporal substitution have been conflicting.

Using repeated cross sectional data from the Michigan Survey of Consumers, Bachman, Berg and Sims (2015) find a small negative effect of inflation expectations on readiness to spend on durable goods; Burke and Ozdagli (2013) find similar results using the New York Fed Survey data. There are several suggested explanations for these findings. High inflation expectations may indicate a loss in faith of policy makers and may suggest uncertain times ahead. This is an often quoted argument against using unconventional monetary policy to stimulate the economy (see for example Volcker, 2011). We aim to control for this channel by including the standard deviation of inflation forecasts as a measure of uncertainty. Inflation can also be seen as a tax on cash or other liquid assets, as well as generally reducing real total wealth, each of which are likely to reduce consumption in all time periods. Finally, Bachman, Berg and Sims (2015) point to money illusion as a possible cause. It has been shown on numerous occasions that the public struggle to understand the difference between nominal and real rates (see for example Shafir, Diamond and Tversky, 1997).

In a much earlier work, Juster and Wachtel (1972) used aggregate time series data and found a negative relationship between inflation expectations and current consumption of durable goods. Finally, D'Acunto, Hoang and Weber (2016) take a different approach and exploit an unexpected announcement of a future increase in VAT in Germany to construct a natural experiment which suggests households do increase consumption in response to an increase in expected future inflation. We hope that our analysis will add robustness to this somewhat contradictory literature.

We use aggregate quarterly time series data from the USA for the period 1981-2016. The dependent variables are the percentage change in expenditure on consumer durables and non-durables, respectively, taken from The Bureau of Economic Analysis. We test the significance of expected future inflation and control for the expected change in nominal interest rates, unemployment and GDP as well as the standard deviation of expected future inflation across all forecasters. The expectations data is taken from the Survey of Professional Forecasters. Our choice to use aggregate time series data as well as the Survey of Professional Forecasters, rather than individual level cross-sectional data, is motivated by the desire to avoid any effect that asking a survey respondent to think about future inflation may have on their consumption decisions. We also believe that at the aggregate level, expectations by professional forecasters are likely to be more representative of the entire population than a random subsample of that population. The reason being that many people base their expectations of future economic conditions on the advice of these professional forecasters.

Our baseline model is

$$C_t = m\left(E_t[\pi_{t+2}], X_t\right) + u_t,$$

where C_t is expenditure at time t, X_t are the set of control variables and $E_t[\pi_{t+2}]$ denotes expected inflation over the next two quarters, formed at time t. Given our use of survey data, our measurement error is given by

$$\pi^s_{(t+2)|t} \quad = \quad E_t[\pi_{t+2}] + \epsilon_t,$$

where $\pi^s_{(t+2)|t}$ is the survey forecast at time t for the annualised inflation rate in 6 months' time - the first subscript denoting the forecast period and the second subscript denoting the period the survey was taken - and ϵ_t denotes the measurement error. We should make it clear that this error is not the forecast error, $E_t[\pi_{t+2}] - \pi_{t+2}$, but simply the error made from using survey data and the fact that we are using a subsample of experts to proxy for the population-wide expectation.

In the literature on New Keynesian Philips curve estimation, it has been suggested by Mavroeidis, Plagborg-Møller and Stock (2014) that expectations formed at time t are likely to cause endogeneity issues on top of any problems of measurement error. To mitigate any possibility of endogeneity, we use the predetermined variable $\pi^s_{(t+2)|(t-1)}$, i.e. the expected inflation rate over the next 6 months, but formed in the previous quarter, in place of $\pi^s_{(t+2)|t}$. However, this adds another layer of measurement error to the problem. Notice

$$\pi^s_{(t+2)|(t-1)} \quad = \quad E_t[\pi_{t+2}] + v_t + \epsilon_{t-1}$$

where

$$\begin{aligned} v_t &= E_{t-1}[\pi_{t+2}] - E_t[\pi_{t+2}], \\ \epsilon_{t-1} &= \pi^s_{(t+2)|t-1} - E_{t-1}[\pi_{t+2}]. \end{aligned}$$

v_t can be thought of as a news shock, that is, how the true expectation changes as you move forward one period. ϵ_{t-1} is the same measurement error we had previously, but lagged by one period. If ϵ_t is considered to be classical measurement error, and the news shock is assumed to be white noise, $(v_t + \epsilon_{t-1})$ can be seen as classical measurement error also.

The Survey of Professional Forecasters allows us access to repeated measurements since it surveys several forecasters in each period. However, if we use the estimator proposed in Delaigle, Hall and Meister (2008), although ϵ_{t-1} is different for each forecaster, v_t is constant across all forecasters, hence will be cancelled out in our estimate of the characteristic function of $v+\epsilon$. We need a different approach. Notice

$$\pi^s_{(t+2)|(t-1)} - \pi^s_{(t+2)|t} \quad = \quad v_t + \epsilon_{t-1} - \epsilon_t$$

and, by lagging this difference by one period

$$\pi^s_{(t+1)|(t-2)} - \pi^s_{(t+1)|(t-1)} \quad = \quad v_{t-1} + \epsilon_{t-2} - \epsilon_{t-1}$$

where

$$\begin{aligned} v_{t-1} &= E_{t-2}[\pi_{t+1}] - E_{t-1}[\pi_{t+1}], \\ \epsilon_{t-2} &= \pi^s_{(t+1)|t-2} - E_{t-2}[\pi_{t+1}]. \end{aligned}$$

Hence

$$\left(\pi^s_{(t+2)|(t-1)} - \pi^s_{(t+2)|t}\right) + \left(\pi^s_{(t+1)|(t-2)} - \pi^s_{(t+1)|(t-1)}\right) \quad = \quad v_t + v_{t-1} + \epsilon_{t-2} - \epsilon_t$$

and we can use the following estimator for the characteristic function of $v+\epsilon$

$$\hat{f}^{\text{ft}}_{\epsilon+v}(u) = \left| \frac{1}{n} \sum_{t=2}^{n} \cos\left\{ u\left(\pi^s_{(t+2)|(t-1)} - \pi^s_{(t+2)|(t)} + \pi^s_{(t+1)|(t-2)} - \pi^s_{(t+1)|(t-1)}\right)\right\} \right|^{1/2}.$$

Interestingly, this novel estimator requires no repeated measurements, but instead utilises the dynamics within the model. Notice that we must assume v and ϵ to be i.i.d., independent and strictly stationary for the validity of this estimator, as well as the usual assumption that $f_\epsilon(\cdot)$ is symmetric around zero. Alternatively, we could use the estimator of Li and Vuong (1998). Again, for this estimator we do not require repeated observations from our survey data. Instead we can use $\pi^s_{(t+2)|t}$ and $\pi^s_{(t+2)|t-1}$ as repeated measurements since the errors need not be identically distributed for the validity of this estimator. For each of the other control variables involving expectations, we follow the same approach.

As in Section 3.5.1 we choose the bandwidth $b = \left(\frac{1}{n}\right)^{1/2(d+d_1)} \approx 0.75$. Again, we consider a range of bandwidths around this value to analyse the robustness of our results to the choice of bandwidth. All other parameter settings are as in Section 3.5.1 and all variables are standardised to have zero mean and unit variance.

Table 6 displays the p-values for our test, the test of Delgado and Manteiga (2001) and t-tests from an IV and OLS quadratic regression with $\pi^s_{(t+2)|(t-1)}$ and its square.

Table 6: Inflation Expectations (P-Values)

Dependent Variable	**My Test**			**DM**			**IV**	**OLS**
	Bandwidth							
	0.5	**0.75**	**1.00**	**0.45**	**0.65**	**0.85**		
Durables	0.088	0.040	0.044	0.294	0.300	0.266	0.190	0.855
Non-Durables	0.202	0.204	0.200	0.284	0.260	0.265	0.214	0.771

Our results indicate that there does appear to be a significant relationship between inflation expectations and current expenditure in the case of durable goods. The differences in p-values between our test and the test of Delgado and Manteiga (2001) indicates that measurement error again has a considerable impact on our conclusions.

It is also interesting to see that for the case of non-durable goods we do not find a significance effect of inflation expectations. This seems sensible given economic theory; durable goods are more likely to be bought on credit and their purchase can be more easily substituted from one period to the next in comparison to non-durable goods. Appendix 3.8 provides regression plots of the two relationships. In each case there is a similar relationship; it appears that at low levels of expected inflation there is the predicted positive effect on current consumption. However, as inflation forecasts become very large this relationship becomes negative, most likely a result of public anxiety about future economic conditions and/or a loss in faith in the ability of their central bank. This may help to explain why, in the previous literature, linear specifications have been unable to find a significant relationship, and why many actually report a negative relationship.

As a result of these findings, it is useful to investigate whether the significance that is found by our test is being driven by the positive relationship at lower inflation expectations or by the negative relationship at the other end. Unfortunately, we do not have enough data for the high inflation subset of the data, however, we can still test the hypothesis on the low inflation subset. In particular, we use observations for which the survey forecast was less than or equal to 5%. We find little difference in our results for this subset as compared with the results from the full dataset.[4] This is perhaps not too surprising given that more then 90% of our data falls within this low inflation range.

These findings have important implications for policy makers, suggesting that there is scope to utilise inflation expectations to stimulate current consumption. However, we should proceed with caution when inflation expectations are high since the relationship may reverse in this case. In addition, these findings again highlight the need to account for measurement error when conducting nonparametric testing.

3.6. Conclusion

This chapter develops, to the best of our knowledge, the first nonparametric significance test for regression models with mismeasured regressors. In particular, the measurement error need not enter the model through the regressors of interest and may only impact the controlling variables. Our test is able to overcome the slow rates of convergence associated with kernel deconvolution estimation and detect local alternatives at the $\sqrt{n}$ rate. The asymptotic distribution is shown to be case dependent and difficult to estimate in practice, as such we provide bootstrap procedures to obtain critical values. We extend our results from the i.i.d. setting to the case of weakly dependent data and outline the properties of the test when the density of the measurement error is unobserved. Finally we consider two empirical applications to highlight the wide applicability of the test. The first tests the significance of cognitive ability on income, life satisfaction, health, and risk aversion. The second shows that future inflation expectations are a viable channel for policy makers to stimulate current consumption. In this example we also showed a novel approach to estimating the measurement error density without the need for repeated measurements.

There are a number of natural avenues for future work stemming from this chapter. We have focussed solely on the case of classical measurement error, however in many situations this is unlikely to hold, as such an equivalent test able to accommodate nonclassical error would be extremely valuable. Also, there is currently no theory for the selection of a data dependent

[4]Detailed results can be obtained from the author upon request.

bandwidth in testing problems when measurement error is present, furthermore bandwidth choice when a mixture of error free and contaminated regressors are present is a very practical and worthwhile problem to solve. Finally, it would not be difficult to extend the ideas and results in this chapter to tests of general conditional moment equalities, or to add to the growing literature on testing conditional moment inequalities.

3.7. Appendix - Mathematical Proofs

3.7.1. Proof of Theorem 5.

3.7.1.1. *Proof of (i).* Throughout this and the proceeding proofs we will make use of the following Lemma.

LEMMA 5. *Under Assumptions 6 and 7, for* $k = \{0, 1\}$

$$\int x_j^k \tilde{\mathcal{K}}_b(x_j) dx_j \quad \sim \quad \frac{C_k}{(2\pi \mathrm{i})^k \, b^{(k+1)}}.$$

□

Define $Z_i \equiv (Y_i, W_i')'$. We write $\hat{T}_n(\xi)$ as a second-order U-statistic

$$\begin{aligned} \hat{T}_n(\xi) &= \frac{1}{2}\frac{(n-1)}{n}\binom{n}{2}^{-1} \sum_{1 \le i < j \le n} p_n(Z_i, Z_j; \xi) \\ &\equiv \frac{1}{2}\frac{(n-1)}{n} U_n(\xi), \end{aligned}$$

where $p_n(Z_i, Z_j; \xi)$ is a symmetric kernel defined as

$$\begin{aligned} p_n(Z_i, Z_j; \xi) \equiv \; & (Y_i - Y_j) \int \mathcal{K}_b\left(\frac{x - W_i}{b}\right) \mathcal{K}_b\left(\frac{x_{(1)} - W_{(1)j}}{b}\right) \mathcal{W}(x; \xi) dx \\ & + (Y_j - Y_i) \int \mathcal{K}_b\left(\frac{x - W_j}{b}\right) \mathcal{K}_b\left(\frac{x_{(1)} - W_{(1)i}}{b}\right) \mathcal{W}(x; \xi) dx. \end{aligned}$$

For the time being we shall drop the notational dependence on ξ only to minimise excess notation, however it should not be forgotten that all objects in the proceeding analysis depend on ξ. The Hoeffding projection of U_n, $\hat{U}_n$, is given by

$$\hat{U}_n \quad = \quad \theta_n + \frac{2}{n}\sum_{i=1}^{n}[r_n(Z_i) - \theta_n],$$

where

$$r_n(Z_i) \quad \equiv \quad E[p_n(Z_i, Z_j)|Z_i]$$

and

$$\theta_n \equiv E[r_n(Z_i)] = E[p_n(Z_i, Z_j)].$$

First, we need to show that the difference between $\hat{U}_n$ and U_n is asymptotically negligible. To this end we appeal to Lemma 3.1 in Powell, Stock and Stoker (1989) which states that $\hat{U}_n - U_n = o_p(1)$ if $E\left[|p_n(Z_i, Z_j)|^2\right] = o(n)$ for $i \neq j$. In our case, we must show $E\left[|p_n(Z_i, Z_j; \xi)|^2\right] = o(n)$ uniformly over ξ. Define $m_2(x) \equiv E\left[Y^2 | X = x\right]$,

$$\begin{aligned}
& E\left[|p_n(Z_i, Z_j)|^2\right] \\
\leq \quad & 4E\left|(Y_i - Y_j)\int \mathcal{K}_b\left(\frac{x - W_i}{b}\right)\mathcal{K}_b\left(\frac{x_{(1)} - W_{(1)j}}{b}\right)\mathcal{W}(x;\xi)dx\right|^2
\end{aligned}$$

where we have used the C_r inequality. Using Hölder's inequality, and assumption $\mathrm{E}|Y_i|^4 \leq \infty$, we can bound this in the following manner

$$\begin{aligned}
E\left[|p_n(Z_i, Z_j)|^2\right] &= O(1)E\left|(Y_i - Y_j)\int \mathcal{K}_b\left(\frac{x - W_i}{b}\right)\mathcal{K}_b\left(\frac{x_{(1)} - W_{(1)j}}{b}\right)\mathcal{W}(x;\xi)dx\right|^2 \\
&= O(1)\left(E\left|Y_i\right|^4\right)^{\frac{1}{2}} \\
&\quad \times \int \left(E\left[\left|\mathcal{K}_b\left(\frac{x - W_i}{b}\right)\right|^{\frac{8}{3}}\right]\right)^{\frac{3}{4}} \left(E\left[\left|\mathcal{K}_b\left(\frac{x_{(1)} - W_{(1)j}}{b}\right)\right|^{\frac{8}{3}}\right]\right)^{\frac{3}{4}} dx.
\end{aligned}$$

Notice, for an arbitrary v

$$\begin{aligned}
E\left|\mathcal{K}_b\left(\frac{x - W_i}{b}\right)\right|^v &= \int \left|\mathcal{K}_b\left(\frac{x - w}{b}\right)\right|^v f_W(w)dw \\
&= b^d \int \left|\mathcal{K}_b(z)\right|^v f_W(x - bz)dz \\
&= O(b^d)\int \left|\mathcal{K}_b(z)\right|^v dz \\
&= O\left(b^{d - dv}\right)
\end{aligned}$$

using a change of variables, the boundedness of $f_W(\cdot)$ and a simple extension of Lemma 5 in the final equality. Hence

$$\begin{aligned}
E\left[|p_n(Z_i, Z_j)|^2\right] &= O\left(b^{-(d+d_1)\frac{5}{4}}\right) \\
&= o(n),
\end{aligned}$$

using Assumption 6 (ii) ($nb^{(d+d_1)\frac{5}{4}} \to \infty$).

The next step is to apply a central limit theorem to $\big(r_n(Z_i;\xi) - \theta_n(\xi)\big)$. Since $\big(r_n(Z_i;\xi) - \theta_n(\xi)\big)$ is i.i.d. and zero mean, if we can show $E\big[\int (r_n(Z_i;\xi) - \theta_n(\xi))^2 d\mu(\xi)\big] < \infty$ (a sufficient condition for tightness of the process), then the central limit theorem for Hilbert space-valued random variables can be applied. This result shows that $n^{-1/2}\sum_{i=1}^n [r_n(Z_i;\cdot) - \theta_n(\cdot)]$ converges weakly to a zero mean Gaussian process, say $Z_{O,\infty}(\cdot)$, on $L_2(\Xi,\mu)$, with covariance function $V_O : \Xi \times \Xi \to \mathbb{R}^+$ defined by $V_O(\xi,\xi') = \lim_{n\to\infty} E\left[(r_n(Z_i;\xi) - \theta_n(\xi))\left(r_n(Z_i;\xi') - \theta_n(\xi')\right)\right]$ (see Politis and Romano, 1994).

To show $E\left[\int (r_n(Z_i;\xi) - \theta_n(\xi))^2 d\mu(\xi)\right] = \int Var\big[r_n(Z_i;\xi)\big] d\mu(\xi) < \infty$ we calculate the bound of $Var\big[r_n(Z_i;\xi)\big]$, uniformly in ξ, in the following proposition.

Proposition 2. *Under Assumptions 6 and 7*

$$Var\left(r_n(Z_i;\xi)\right) = O(1).$$

□

Hence, we conclude $\int Var\big[r_n(Z_i;\xi)\big] d\mu(\xi) < \infty$ allowing us to apply the central limit theorem for Hilbert-space valued random variables. Combining these results we have shown that $\sqrt{n}\hat{T}_n(\xi)$ converges weakly to a Gaussian process with mean θ_n and covariance function

$$V_O(\xi,\xi') = \lim_{n\to\infty} E\big[\big(r_n(Z_i;\xi) - \theta_n(\xi)\big)\big(r_n(Z_i;\xi') - \theta_n(\xi')\big)\big].$$

We must now consider the value of θ_n under the null hypothesis

$$
\begin{aligned}
\theta_n &= 2E\left[\left(Y_i - Y_j\right)\int \mathcal{K}_b\left(\frac{x-W_i}{b}\right)\mathcal{K}_b\left(\frac{x_{(1)}-W_{(1)j}}{b}\right)\mathcal{W}\left(x;\xi\right)dx\right] \\
&= 2E\left[\left(m\left(X_i\right) - r\left(X_{(1)j}\right)\right)\int \mathcal{K}_b\left(\frac{x-W_i}{b}\right)\mathcal{K}_b\left(\frac{x_{(1)}-W_{(1)j}}{b}\right)\mathcal{W}\left(x;\xi\right)dx\right] \\
&= \frac{2}{\left(2\pi b\right)^{d+d_1}}E\left[\left(m\left(X_i\right) - r\left(X_{(1)j}\right)\right)\int\left\{\begin{array}{c} \int e^{-\mathrm{i}t\left(\frac{x-W_i}{b}\right)}\frac{K^{\mathrm{ft}}(t)}{f_\epsilon^{\mathrm{ft}}(t/b)}dt \\ \times\int e^{-\mathrm{i}s_{(1)}\left(\frac{x_{(1)}-W_{(1)j}}{b}\right)}\frac{K^{\mathrm{ft}}\left(s_{(1)}\right)}{f_\epsilon^{\mathrm{ft}}\left(s_{(1)}/b\right)}ds_{(1)}\end{array}\right\}\mathcal{W}\left(x;\xi\right)dx\right] \\
&= \frac{2}{\left(2\pi b\right)^{d+d_1}}E\left[m\left(X_i\right)\int\left\{\begin{array}{c} \int e^{-\mathrm{i}t\left(\frac{x-X_i}{b}\right)}K^{\mathrm{ft}}\left(t\right)dt \\ \times\int e^{-\mathrm{i}s_{(1)}\left(\frac{x_{(1)}-X_{(1)j}}{b}\right)}K^{\mathrm{ft}}\left(s_{(1)}\right)ds_{(1)}\end{array}\right\}\mathcal{W}\left(x;\xi\right)dx\right] \\
&\quad -\frac{2}{\left(2\pi b\right)^{d+d_1}}E\left[r\left(X_{(1)j}\right)\int\left\{\begin{array}{c} \int_t e^{-\mathrm{i}t\left(\frac{x-X_i}{b}\right)}K^{\mathrm{ft}}\left(t\right)dt \\ \times\int e^{-\mathrm{i}s_{(1)}\left(\frac{x_{(1)}-X_{(1)j}}{b}\right)}K^{\mathrm{ft}}\left(s_{(1)}\right)ds_{(1)}\end{array}\right\}\mathcal{W}\left(x;\xi\right)dx\right] \\
&= T_1 - T_2.
\end{aligned}
$$

For T_1

$$
\begin{aligned}
T_1 &= \frac{2}{\left(2\pi b\right)^{d+d_1}}\int\int\int E\left[m\left(X_i\right)e^{-\mathrm{i}t\left(\frac{x-X_i}{b}\right)}\right]E\left[e^{-\mathrm{i}s_{(1)}\left(\frac{x_{(1)}-X_{(1)j}}{b}\right)}\right]K^{\mathrm{ft}}\left(t\right)K^{\mathrm{ft}}\left(s_{(1)}\right)\mathcal{W}\left(x;\xi\right)dtds_{(1)}dx \\
&= \frac{2}{\left(2\pi b\right)^{d+d_1}}\int\int\int\left[mf_X\right]^{\mathrm{ft}}\left(\frac{t}{b}\right)f_{X_{(1)}}^{\mathrm{ft}}\left(\frac{s_{(1)}}{b}\right)e^{-\mathrm{i}t\left(\frac{x}{b}\right)}e^{-\mathrm{i}s_{(1)}\left(\frac{x_{(1)}}{b}\right)}K^{\mathrm{ft}}\left(t\right)K^{\mathrm{ft}}\left(s_{(1)}\right)\mathcal{W}\left(x;\xi\right)dtds_{(1)}dx \\
&= \frac{2}{\left(2\pi b\right)^{d+d_1}}\int\int\left[mf_X\right]^{\mathrm{ft}}\left(\frac{t}{b}\right)K^{\mathrm{ft}}\left(t\right)e^{-\mathrm{i}t\frac{x}{b}}dt\int f_{X_{(1)}}^{\mathrm{ft}}\left(\frac{s_{(1)}}{b}\right)K^{\mathrm{ft}}\left(s_{(1)}\right)e^{-\mathrm{i}s_{(1)}\frac{x_{(1)}}{b}}ds_{(1)}\mathcal{W}\left(x;\xi\right)dx \\
&= 2\int\left[mf_X * K\left(\frac{\cdot}{b}\right)\right]\left(x\right)\left[f_{X_{(1)}} * K\left(\frac{\cdot}{b}\right)\right]\left(x_{(1)}\right)\mathcal{W}\left(x;\xi\right)dx
\end{aligned}
$$

where

$$
\begin{aligned}
\left[mf_X * K\left(\frac{\cdot}{b}\right)\right]\left(x\right) &= \int m\left(a\right)f_X\left(a\right)K\left(\frac{x-a}{b}\right)da \\
\left[f_{X_{(1)}} * K\left(\frac{\cdot}{b}\right)\right]\left(x_{(1)}\right) &= \int f_{X_{(1)}}\left(c\right)K\left(\frac{x_{(1)}-c}{b}\right)dc.
\end{aligned}
$$

By similar arguments

$$
T_2 = 2\int\left[f_X * K\left(\frac{\cdot}{b}\right)\right]\left(x\right)\left[rf_{X_{(1)}} * K\left(\frac{\cdot}{b}\right)\right]\left(x_{(1)}\right)\mathcal{W}\left(x;\xi\right)dx
$$

where

$$\left[f_X * K\left(\frac{\cdot}{b}\right)\right](x) = \int f_X(a)\, K\left(\frac{x-a}{b}\right) da$$
$$\left[r f_{X_{(1)}} * K\left(\frac{\cdot}{b}\right)\right]\left(x_{(1)}\right) = \int r(c)\, f_{X_{(1)}}(c)\, K\left(\frac{x_{(1)}-c}{b}\right) dc.$$

So

$$\begin{aligned}
\theta_n &= 2\int \left\{ \begin{array}{c} \left[m f_X * K\left(\frac{\cdot}{b}\right)\right](x) \left[f_{X_{(1)}} * K\left(\frac{\cdot}{b}\right)\right]\left(x_{(1)}\right) \\ -\left[r f_{X_{(1)}} * K\left(\frac{\cdot}{b}\right)\right]\left(x_{(1)}\right)\left[f_X * K\left(\frac{\cdot}{b}\right)\right](x) \end{array} \right\} \mathcal{W}(x;\xi)\, dx \\
&= 2\int\int\int \{m(a) - r(c)\}\, f_X(a)\, f_{X_{(1)}}(c)\, K\left(\frac{x-a}{b}\right) K\left(\frac{x_{(1)}-c}{b}\right) dadc\mathcal{W}(x;\xi)\, dx \\
&= 2\int\int M(x-ub)\, K(u)\, du \int_{\tilde{u}} f_{X_{(1)}}\left(x_{(1)} - \tilde{u}b\right) K(\tilde{u})\, d\tilde{u}\mathcal{W}(x;\xi)\, dx \\
&\quad -2\int\int R\left(x_{(1)} - ub\right) K(u)\, du \int_u f_X(x-\tilde{u}b)\, K(\tilde{u})\, d\tilde{u}\mathcal{W}(x;\xi)\, dx \\
&= 2\int\int \left\{m(x) - r(x_{(1)})\right\} f_X(x)\, f_{X_{(1)}}\left(x_{(1)}\right) \mathcal{W}(x;\xi)\, dx + O\left(b^2\right)
\end{aligned}$$

where the final equality results from the use of a second-order kernel, as in Powell, Stock and Stoker (1989). Under the null hypothesis $m(x) = r(x_{(1)})$, hence $\sqrt{n}\theta_n = o(1)$ by Assumption 6 (vi) ($nb^4 \to 0$), and $\sqrt{n}\hat{T}_n(\cdot)$ converges weakly to a zero mean Gaussian process, say $Z_{O,\infty}(\cdot)$, on $L_2(\Xi, \mu)$, with covariance function $V_O : \Xi \times \Xi \to \mathbb{R}^+$.

Finally, we apply the continuous mapping theorem to show

$$nCM_n(\xi) \to_d \int |Z_{O,\infty}(\xi)|^2\, d\mu(\xi).$$

To characterise this asymptotic distribution we appeal to Bierens and Ploberger (1997) Theorem 3. This allows us to write $\int |Z_{O,\infty}(\xi)|^2\, d\mu(\xi) \sim \sum_{i=1}^{\infty} \lambda_{O,i}\nu_i^2$ where ν_i are i.i.d. $N(0,1)$ random variables and $\lambda_{O,i}$ are the solutions to the eigenvalue problem

$$\int V_O(\xi,\xi')\psi_{O,i}(\xi')d\mu(\xi') = \lambda_{O,i}\psi_{O,i}(\xi).$$

The eigenvalues $\lambda_{O,i}$ are real valued, non-negative and satisfy $\sum_{i=1}^{\infty} \lambda_{O,i} < \infty$. This completes the proof of Theorem 1 (i).

3.7.1.2. *Proof of (ii).* For the supersmooth case we will make use of the following Lemma.

LEMMA 6. *Under Assumptions 6 and 8*

$$\int \tilde{\mathcal{K}}_b(x_j)dx_j = \frac{C\mu^{\left(\frac{\gamma_0}{\gamma}\right)}}{b\left(\frac{\gamma_0}{\gamma}\right)!}$$

and

$$\int x_j\tilde{\mathcal{K}}_b(x_j)dx_j = \frac{C\mu^{\left(\frac{1-\gamma_0}{\gamma}\right)}}{b^2 2\pi \mathrm{i}\left(\frac{1-\gamma_0}{\gamma}\right)!}.$$

□

The method of proof is the same as for part (i). The first task is to show $E\big[|p_n(Z_i,Z_j;\xi)|^2\big] = o(n)$ uniformly in ξ. In the same way that Lemma 5 was used in Theorem 1 (i), Lemma 6 can be used to show

$$E\left|\mathcal{K}_b\left(\frac{x-W_i}{b}\right)\right|^v = O(b^{d-dv}),$$

for an arbitrary v. Hence, in the same manner, we have

$$\begin{aligned} E\big[|p_n(Z_i,Z_j)|^2\big] &= O\left(b^{-(d+d_1)\frac{5}{4}}\right) \\ &= o(n). \end{aligned}$$

To apply the central limit theorem for Hilbert-space valued random variables we must show $\int Var\big[r_n(Z_i;\xi)\big]d\mu(\xi) < \infty$. We appeal to the following proposition.

PROPOSITION 3. *Under Assumptions 6 and 8*

$$Var\left(r_n(Z_i;\xi)\right) = O(1).$$

□

Hence, we conclude $\int Var\big[r_n(Z_i;\xi)\big]d\mu(\xi) < \infty$ and can again use the central limit theorem for Hilbert-space valued random variables. Thus, $\sqrt{n}\hat{T}_n(\xi)$ converges weakly to a Gaussian process with mean θ_n and covariance function

$$V_S(\xi,\xi') = \lim_{n\to\infty} E\big[\big(r_n(Z_i;\xi)-\theta_n(\xi)\big)\big(r_n(Z_i;\xi')-\theta_n(\xi')\big)\big].$$

As in the proof of Theorem 1 (i)

$$\begin{aligned}\sqrt{n}\theta_n &= \sqrt{n}2\int \left(m(x)-r(x_{(1)})\right)\mathcal{W}(x;\xi)f_X(x)f_{X_{(1)}}(x_{(1)})dx + O\left(\sqrt{n}b^2\right)\\ &= o(1)\end{aligned}$$

since $m(x)=r(x_{(1)})$ under the null hypothesis and by Assumption 6 (vi) ($nb^4\to 0$).

Again, we can apply the continuous mapping theorem and Theorem 3 in Bierens and Ploberger (1997) to show

$$nCM_n(\xi)\to_d \int |Z_{O,\infty}(\xi)|^2\, d\mu(\xi) \sim \sum_{i=1}^{\infty}\lambda_{S,i}\nu_i^2$$

where ν_i are i.i.d. $N(0,1)$ random variables and $\lambda_{S,i}$ are the solutions to the eigenvalue problem

$$\int V_S(\xi,\xi')\psi_{S,i}(\xi')d\mu(\xi') = \lambda_{S,i}\psi_{S,i}(\xi).$$

The eigenvalues $\lambda_{S,i}$ are real valued, non-negative and satisfy $\sum_{i=1}^{\infty}\lambda_{S,i}<\infty$. This concludes the proof of Theorem 1 (ii).

3.7.2. Proof of Theorem 2.

3.7.2.1. *Proof of (i).* The proof follows along the same lines as Theorem 1 (i). However, notice that under the alternative hypothesis

$$\begin{aligned}
\sqrt{n}\theta_n &= \sqrt{n}2\int \left(m(x) - r(x_{(1)})\right)\mathcal{W}(x;\xi)f_X(x)f_{X_{(1)}}(x_{(1)})dx + o(1) \\
&= \sqrt{n}2\int c_n\Delta(x)\mathcal{W}(x;\xi)f_X(x)f_{X_{(1)}}(x_{(1)})dx + o(1) \\
&= 2\int \Delta(x)\mathcal{W}(x;\xi)f_X(x)f_{X_{(1)}}(x_{(1)})dx + o(1) \\
&\equiv \bar{\Delta}_O(\xi) + o(1).
\end{aligned}$$

Combining this with the results in Theorem 1 (i), $\sqrt{n}\hat{T}_n(\cdot)$ converges weakly to a Gaussian process, say $\tilde{Z}_{O,\infty}(\cdot)$, on $L_2(\Xi,\mu)$, with mean function $\bar{\Delta}_O(\cdot)$ and covariance function

$$\tilde{V}_O(\xi,\xi') = \lim_{n\to\infty} E\big[\big(r_n(Z_i;\xi) - \theta_n(\xi)\big)\big(r_n(Z_i;\xi') - \theta_n(\xi')\big)\big].$$

Finally we apply the continuous mapping theorem to show

$$nCM_n(\xi) \to_d \int \left|\tilde{Z}_{O,\infty}(\xi)\right|^2 d\mu(\xi).$$

To characterise this asymptotic distribution we again appeal to Theorem 3 of Bierens and Ploberger (1997). This allows us to write $\int |Z_{O,\infty}(\xi)|^2\, d\mu(\xi) \sim \sum_{i=1}^{\infty}\left(\bar{\Delta}_{O,i} + \sqrt{\tilde{\lambda}_{O,i}}\nu_i\right)^2$ where ν_i are i.i.d. $N(0,1)$ random variables, $\tilde{\lambda}_{O,i}$ are the solutions to the eigenvalue problem

$$\int \tilde{V}_O(\xi,\xi')\tilde{\psi}_{O,i}(\xi')d\mu(\xi') = \tilde{\lambda}_{O,i}\tilde{\psi}_{O,i}(\xi)$$

and $\bar{\Delta}_{O,i} = \int \bar{\Delta}_{O,}(\xi)\psi_{O,i}(\xi)d\mu(\xi)$. As before, the eigenvalues $\tilde{\lambda}_{O,i}$ are real valued, non-negative and satisfy $\sum_{i=1}^{\infty}\tilde{\lambda}_{O,i} < \infty$. This completes the proof of Theorem 2 (i).

3.7.2.2. *Proof of (ii).* The proof is almost identical to Theorem 2 (i).

3.7.3. Proof of Theorem 7.

3.7.3.1. *Proof of (i).* To show the residual from the Hoeffding projection is asymptotically negligible we follow the arguments in Robinson (1989) which extends Proposition 2 of Denker and Keller (1983) to allow the kernel of the U-statistic to depend on the sample size. For some $\delta,\ \varsigma > 0$, assume $\beta(j) = O(j^{\eta}) = O\left(j^{(\varsigma-2)(2+\delta)/\delta}\right)$, then the residual from the Hoeffding projection can be bounded as

$$\sup_{\xi\in\Xi}\left(\sqrt{n}\left\{\hat{U}(\xi) - U(\xi)\right\}\right) \quad = \quad O_p\left(n^{\frac{-1+\varsigma}{2}} s_{\delta}^{\frac{1}{2+\delta}}\right)$$

where $s_\delta = \sup_{\xi\in\Xi}\max_{i\neq j} E\left[|p_n(Z_i, Z_j)|^{2+\delta}\right]$. In particular choose $\frac{1}{\eta} < \frac{\delta}{2+\delta} < \frac{2}{\eta}$ (see Robinson, 1989). Using Hölder's inequality, and since we have assumed $\mathrm{E}|Y_i|^4 \leq \infty$, we can write

$$\begin{aligned} E\left[|p_n(Z_i, Z_j)|^{2+\delta}\right] &= O(1)E\left|(Y_i - Y_j)\int \mathcal{K}_b\left(\frac{x - W_i}{b}\right)\mathcal{K}_b\left(\frac{x_{(1)} - W_{(1)j}}{b}\right)\mathcal{W}(x;\xi)dx\right|^{2+\delta} \\ &= O(1)\left(E\left|Y_i\right|^4\right)^{\frac{2+\delta}{4}} \\ &\quad \times\int\left(E\left[\left|\mathcal{K}_b\left(\frac{x-W_i}{b}\right)\right|^{\frac{8}{3}}\right]\right)^{\frac{3(2+\delta)}{8}}\left(E\left[\left|\mathcal{K}_b\left(\frac{x_{(1)}-W_{(1)j}}{b}\right)\right|^{\frac{8}{3}}\right]\right)^{\frac{3(2+\delta)}{8}} dx. \end{aligned}$$

As in the proof of Theorem 1 (i), for an arbitrary v

$$E\left|\mathcal{K}_b\left(\frac{x - W_i}{b}\right)\right|^{v} \quad = \quad O\left(b^{d-dv}\right).$$

Hence

$$s_\delta = O\left(b^{-(d+d_1)\frac{5(2+\delta)}{8}}\right)$$

and

$$\begin{aligned} \sup_{\xi\in\Xi}\left(\sqrt{n}\left\{\hat{U}(\xi) - U(\xi)\right\}\right) &= O_p\left(n^{\frac{-1+\varsigma}{2}} b^{-(d+d_1)\frac{5}{8}}\right) \\ &= o_p(1) \end{aligned}$$

using Assumption 9 (ii) ($n^{1-\frac{\varsigma}{2}} b^{(d+d_1)\frac{5}{4}} \to \infty$).

Next we make use of the central limit theorem for Hilbert-space valued, absolutely regular, stationary random variables from Politis and Romano (1994) (Theorem 2.3, i). To use this result we must show $\sup_{\xi\in\Xi}\max_{i\neq j} E\left[|r_n(Z_i;\xi)|^{2+\tilde{\delta}}\right] < \infty$ for some $\tilde{\delta} > 0$. We appeal to the following proposition.

PROPOSITION 4. *Under Assumptions 6, 7 and 9*

$$E\left[|r_n(Z_i;\xi)|^{2+\tilde{\delta}}\right] \quad = \quad O(1).$$

□

The final step is to show $\theta_n = o_p(1)$.

$$\begin{aligned}
\theta_n &= 2E\left[(Y_i - Y_j)\int \mathcal{K}_b\left(\frac{x-W_i}{b}\right)\mathcal{K}_b\left(\frac{x_{(1)}-W_{(1)j}}{b}\right)\mathcal{W}(x;\xi)\,dx\right] \\
&= 2E\left[\left(m(X_i) - r\left(X_{(1)j}\right)\right)\int \mathcal{K}_b\left(\frac{x-W_i}{b}\right)\mathcal{K}_b\left(\frac{x_{(1)}-W_{(1)j}}{b}\right)\mathcal{W}(x;\xi)\,dx\right] \\
&= \frac{2}{(2\pi b)^{d+d_1}}E\left[m(X_i)\int\left\{\begin{array}{c}\int e^{-\mathrm{i}t\left(\frac{x-X_i}{b}\right)}K^{\mathrm{ft}}(t)\,dt \\ \times\int e^{-\mathrm{i}s_{(1)}\left(\frac{x_{(1)}-X_{(1)j}}{b}\right)}K^{\mathrm{ft}}\left(s_{(1)}\right)ds_{(1)}\end{array}\right\}\mathcal{W}(x;\xi)\,dx\right] \\
&\quad -\frac{2}{(2\pi b)^{d+d_1}}E\left[r\left(X_{(1)j}\right)\int\left\{\begin{array}{c}\int_t e^{-\mathrm{i}t\left(\frac{x-X_i}{b}\right)}K^{\mathrm{ft}}(t)\,dt \\ \times\int e^{-\mathrm{i}s_{(1)}\left(\frac{x_{(1)}-X_{(1)j}}{b}\right)}K^{\mathrm{ft}}\left(s_{(1)}\right)ds_{(1)}\end{array}\right\}\mathcal{W}(x;\xi)\,dx\right] \\
&= T_1 - T_2.
\end{aligned}$$

For T_1

$$\begin{aligned}
T_1 &= \frac{2}{(2\pi b)^{d+d_1}}\int\int\int E\left[m(X_i)\,e^{-\mathrm{i}t\left(\frac{x-X_i}{b}\right)}e^{-\mathrm{i}s_{(1)}\left(\frac{x_{(1)}-X_{(1)j}}{b}\right)}\right]K^{\mathrm{ft}}(t)\,K^{\mathrm{ft}}\left(s_{(1)}\right)\mathcal{W}(x;\xi)\,dtds_{(1)}dx \\
&= \frac{2}{(2\pi b)^{d+d_1}}\int\int\int\left[mf_{X_iX_{(1)j}}\right]^{\mathrm{ft}}\left(\frac{t}{b},\frac{s_{(1)}}{b}\right)e^{-\mathrm{i}t\left(\frac{x}{b}\right)}e^{-\mathrm{i}s_{(1)}\left(\frac{x_{(1)}}{b}\right)}K^{\mathrm{ft}}(t)\,K^{\mathrm{ft}}\left(s_{(1)}\right)\mathcal{W}(x;\xi)\,dtds_{(1)}dx \\
&= \frac{2}{(2\pi b)^{d+d_1}}\int\int\int\left[mf_{X_iX_{(1)j}}\right]^{\mathrm{ft}}\left(\frac{t}{b},\frac{s_{(1)}}{b}\right)K^{\mathrm{ft}}\left(t,s_{(1)}\right)e^{-\mathrm{i}\left\{t\left(\frac{x}{b}\right)+s_{(1)}\left(\frac{x_{(1)}}{b}\right)\right\}}dtds_{(1)}\mathcal{W}(x;\xi)\,dx \\
&= \frac{2}{2\pi}\int\left[mf_{X_iX_{(1)j}} * K\left(\frac{\cdot}{b}\right)\right]\left(x,x_{(1)}\right)\mathcal{W}(x;\xi)\,dx
\end{aligned}$$

where the penultimate equality follows from the product form of $K^{\mathrm{ft}}(\cdot)$ and where

$$\left[mf_{X_iX_{(1)j}} * K\left(\frac{\cdot}{b}\right)\right](x,x_1) \quad = \quad \int\int m(a)\,f_{X_iX_{(1)j}}(a,c)K\left(\frac{x-a}{b}\right)K\left(\frac{x_{(1)}-c}{b}\right)dadc.$$

By similar arguments

$$T_2 = \frac{2}{2\pi} \int \left[r f_{X_i X_{(1)j}} * K\left(\frac{\cdot}{b}\right) \right] \left(x, x_{(1)}\right) \mathcal{W}\left(x; \xi\right) dx$$

where

$$\left[r f_{X_i X_{(1)j}} * K\left(\frac{\cdot}{b}\right) \right] \left(x, x_1\right) = \int \int r\left(c\right) f_{X_i X_{(1)j}}(a, c) K\left(\frac{x-a}{b}\right) K\left(\frac{x_{(1)}-c}{b}\right) dadc.$$

So

$$\begin{aligned}
\theta_n &= \frac{2}{2\pi} \int \left\{ \begin{array}{l} \left[m f_{X_i X_{(1)j}} * K\left(\frac{\cdot}{b}\right) \right] \left(x, x_{(1)}\right) \\ - \left[r f_{X_i X_{(1)j}} * K\left(\frac{\cdot}{b}\right) \right] \left(x, x_1\right) \end{array} \right\} \mathcal{W}\left(x; \xi\right) dx \\
&= \frac{2}{2\pi} \int \int \int \left\{ m\left(a\right) - r\left(c\right) \right\} f_{X_i X_{(1)j}}(a, c) K\left(\frac{x-a}{b}\right) K\left(\frac{x_{(1)}-c}{b}\right) dadc \mathcal{W}\left(x; \xi\right) dx \\
&= \frac{2}{2\pi} \int \int \int m\left(x-ub\right) f_{X_i X_{(1)j}}(x-ub, x_{(1)}-\tilde{u}b) K\left(u\right) K\left(\tilde{u}\right) dud\tilde{u} \mathcal{W}\left(x; \xi\right) dx \\
&\quad + \frac{2}{2\pi} \int \int \int r\left(x_{(1)}-\tilde{u}b\right) f_{X_i X_{(1)j}}(x-ub, x_{(1)}-\tilde{u}b) K\left(u\right) K\left(\tilde{u}\right) dud\tilde{u} \mathcal{W}\left(x; \xi\right) dx \\
&= \frac{2}{2\pi} \int \int \left\{ m\left(x\right) - r(x_{(1)}) \right\} f_{X_i X_{(1)j}}(x, x_{(1)}) \mathcal{W}\left(x; \xi\right) dx + O\left(b^2\right)
\end{aligned}$$

where the final equality follows by similar arguments as for the proof of Theorem 1. Hence, we conclude $\sqrt{n}\hat{T}_n(\cdot)$ converges weakly to a zero mean Gaussian process on $L_2(\Xi, \mu)$ with covariance function

$$V_{OT}(\xi, \xi') = \lim_{n\to\infty} \sum_{j=-\infty}^{\infty} E\big[\big(r_n(Z_i; \xi) - \theta_n(\xi)\big)\big(r_n(Z_{i+j}; \xi') - \theta_n(\xi')\big)\big].$$

and the rest of Theorem 1 (i) applies. The part of the theorem related to Theorem 2 (i) is proved in an almost identical manner.

3.7.3.2. *Proof of (ii).* Very similar reasoning as above can be applied to the supersmooth case. We omit the proof for brevity.

3.7.4. Proof of Theorem 8. The proofs of part (i), (ii) and (iii) follow in very similar ways. For brevity we show only the proof for (i).

For some consistent estimator of $f_\epsilon^{\text{ft}}(\cdot)$, denoted $\hat{f}_\epsilon^{\text{ft}}(\cdot)$, we define

$$\hat{\mathcal{K}}_b(a) \equiv \frac{1}{(2\pi b)^{\dim(a)}} \int e^{-\mathrm{i}t\cdot a} \frac{K^{\text{ft}}(t)}{\hat{f}_\epsilon^{\text{ft}}(t/b)} dt$$

and

$$\hat{\hat{T}}_n(\xi) \equiv \frac{1}{n^2} \sum_{i \neq j}^{n} (Y_i - Y_j) \int \hat{\mathcal{K}}_b\left(\frac{x - W_i}{b}\right) \hat{\mathcal{K}}_b\left(\frac{x_{(1)} - W_{(1)j}}{b}\right) \mathcal{W}(x;\xi) dx.$$

Using the identity $\frac{1}{\hat{a}} = \frac{1}{a} - \frac{\frac{\hat{a}-a}{a^2}}{1+\frac{\hat{a}-a}{a}}$, we can write

$$\begin{aligned}
\hat{\hat{T}}_n(\xi) &= \frac{1}{n^2}\frac{1}{(2\pi b)^{d+d_1}} \sum_{i \neq j}^{n} (Y_i - Y_j) \int\int\int \left\{ \begin{array}{c} e^{-\mathrm{i}t\cdot\left(\frac{x-W_i}{b}\right)} e^{-\mathrm{i}s_{(1)}\cdot\left(\frac{x_{(1)}-W_{(1)i}}{b}\right)} \\ \times \frac{K^{\text{ft}}(t)}{\hat{f}_\epsilon^{\text{ft}}(t/b)} \frac{K^{\text{ft}}\left(s_{(1)}\right)}{\hat{f}_\epsilon^{\text{ft}}\left(s_{(1)}/b\right)} \end{array} \right\} \mathcal{W}(x;\xi) dt ds_{(1)} dx \\
&= \frac{1}{n^2}\frac{1}{(2\pi b)^{d+d_1}} \sum_{i \neq j}^{n} (Y_i - Y_j) \int\int\int \left\{ \begin{array}{c} e^{-\mathrm{i}t\cdot\left(\frac{x-W_i}{b}\right)} e^{-\mathrm{i}s_{(1)}\cdot\left(\frac{x_{(1)}-W_{(1)i}}{b}\right)} \\ \times \frac{K^{\text{ft}}(t)}{f_\epsilon^{\text{ft}}(t/b)} \frac{K^{\text{ft}}\left(s_{(1)}\right)}{\hat{f}_\epsilon^{\text{ft}}\left(s_{(1)}/b\right)} \end{array} \right\} \mathcal{W}(x;\xi) dt ds_{(1)} dx \\
&\quad -\frac{1}{n^2}\frac{1}{(2\pi b)^{d+d_1}} \sum_{i \neq j}^{n} (Y_i - Y_j) \int\int\int e^{-\mathrm{i}t\cdot\left(\frac{x-W_i}{b}\right)} e^{-\mathrm{i}s_{(1)}\cdot\left(\frac{x_{(1)}-W_{(1)i}}{b}\right)} K^{\text{ft}}(t) \\
&\quad \times \left(\frac{\hat{f}_\epsilon^{\text{ft}}(t/b) - f_\epsilon^{\text{ft}}(t/b)}{f_\epsilon^{\text{ft}}(t/b)^2}\right) \left(\frac{1}{1 + \frac{\hat{f}_\epsilon^{\text{ft}}(t/b) - f_\epsilon^{\text{ft}}(t/b)}{f_\epsilon^{\text{ft}}(t/b)}}\right) \frac{K^{\text{ft}}\left(s_{(1)}\right)}{\hat{f}_\epsilon^{\text{ft}}\left(s_{(1)}/b\right)} \mathcal{W}(x;\xi) dt ds_{(1)} dx \\
&\equiv T_1 - T_2.
\end{aligned}$$

For T_1, we use the same reasoning to show

$$\begin{aligned}
T_1 &= \frac{1}{n^2}\frac{1}{(2\pi b)^{d+d_1}} \sum_{i \neq j}^{n} (Y_i - Y_j) \int\int\int \left\{ \begin{array}{c} e^{-\mathrm{i}t\cdot\left(\frac{x-W_i}{b}\right)} e^{-\mathrm{i}s_{(1)}\cdot\left(\frac{x_{(1)}-W_{(1)i}}{b}\right)} \\ \times \frac{K^{\text{ft}}(t)}{f_\epsilon^{\text{ft}}(t/b)} \frac{K^{\text{ft}}\left(s_{(1)}\right)}{f_\epsilon^{\text{ft}}\left(s_{(1)}/b\right)} \end{array} \right\} \mathcal{W}(x;\xi) dt ds_{(1)} dx \\
&\quad -\frac{1}{n^2}\frac{1}{(2\pi b)^{d+d_1}} \sum_{i \neq j}^{n} (Y_i - Y_j) \int\int\int e^{-\mathrm{i}t\cdot\left(\frac{x-W_i}{b}\right)} e^{-\mathrm{i}s_{(1)}\cdot\left(\frac{x_{(1)}-W_{(1)i}}{b}\right)} \frac{K^{\text{ft}}(t)}{f_\epsilon^{\text{ft}}(t/b)} K^{\text{ft}}(s_{(1)}) \\
&\quad \times \left(\frac{\hat{f}_\epsilon^{\text{ft}}(s_{(1)}/b) - f_\epsilon^{\text{ft}}(s_{(1)}/b)}{f_\epsilon^{\text{ft}}(s_{(1)}/b)^2}\right) \left(\frac{1}{1 + \frac{\hat{f}_\epsilon^{\text{ft}}\left(s_{(1)}/b\right) - f_\epsilon^{\text{ft}}\left(s_{(1)}/b\right)}{f_\epsilon^{\text{ft}}\left(s_{(1)}/b\right)}}\right) \mathcal{W}(x;\xi) dt ds_{(1)} dx \\
&\equiv T_{11} - T_{12}.
\end{aligned}$$

Notice that $T_{11} = \hat{T}_n(\xi)$ is the original statistic when $f_\epsilon^{\text{ft}}(\cdot)$ is known.

We now show T_{12} is asymptotically negligible. First, notice that depending on the speed of convergence of $\hat{f}_\epsilon^{\text{ft}}\left(s_{(1)}/b\right) \to_p f_\epsilon^{\text{ft}}\left(s_{(1)}/b\right)$ and the order of $f_\epsilon^{\text{ft}}\left(s_{(1)}/b\right)$

$$\left(\frac{1}{1+\frac{\hat{f}_\epsilon^{\text{ft}}\left(s_{(1)}/b\right)-f_\epsilon^{\text{ft}}\left(s_{(1)}/b\right)}{f_\epsilon^{\text{ft}}\left(s_{(1)}/b\right)}}\right)$$

will either be $o_p(1)$ or $O_p(1)$. In which case

$$\left(\frac{\hat{f}_\epsilon^{\text{ft}}\left(s_{(1)}/b\right)-f_\epsilon^{\text{ft}}\left(s_{(1)}/b\right)}{1+\frac{\hat{f}_\epsilon^{\text{ft}}\left(s_{(1)}/b\right)-f_\epsilon^{\text{ft}}\left(s_{(1)}/b\right)}{f_\epsilon^{\text{ft}}\left(s_{(1)}/b\right)}}\right) = o_p(1)$$

by the consistency of $\hat{f}_\epsilon^{\text{ft}}\left(s_{(1)}/b\right)$. Hence

$$T_{12} = o_p(1)\left(\frac{1}{n^2}\frac{1}{(2\pi b)^{d+d_1}}\sum_{i\neq j}^{n}(Y_i-Y_j)\int\int\int\left\{\begin{array}{c} e^{-\mathrm{i}t\cdot\left(\frac{x-W_i}{b}\right)}e^{-\mathrm{i}s_{(1)}\cdot\left(\frac{x_{(1)}-W_{(1)i}}{b}\right)} \\ \times\frac{K^{\text{ft}}(t)}{f_\epsilon^{\text{ft}}(t/b)}\frac{K^{\text{ft}}\left(s_{(1)}\right)}{f_\epsilon^{\text{ft}}\left(s_{(1)}/b\right)^2}\end{array}\right\}\mathcal{W}(x;\xi)dtds_{(1)}dx\right).$$

Notice that the multiple in T_{12} is the same as the U-statistic we dealt with in Theorem 1 but divided by $\frac{1}{f_\epsilon^{\text{ft}}(s_{(1)}/b)}$. So we have to deal with $\frac{1}{f_\epsilon^{\text{ft}}(s_{(1)}/b)^2}$ rather than $\frac{1}{f_\epsilon^{\text{ft}}(s_{(1)}/b)}$. However, recall that the convergence rate of the U-statistic in Theorem 1 did not depend on $f_\epsilon^{\text{ft}}(\cdot)$. We could simply refer to $f_\epsilon^{\text{ft}}(\cdot)^2$ as being ordinary smooth with parameter 2α instead of α. For the supersmooth case we can apply the same reasoning. Thus, we can use the same arguments as in the proofs of Theorem 1 and 2 to show that $T_{12} = o_p(1)O_p(n^{-1/2})$ in both the ordinary smooth and supersmooth cases.

We return to T_2. As for T_1, we write

$$\begin{aligned}
T_2 &= \frac{1}{n^2}\frac{1}{(2\pi b)^{d+d_1}}\sum_{i\neq j}^{n}(Y_i-Y_j)\int\int\int e^{-\mathrm{i}t\cdot\left(\frac{x-W_i}{b}\right)}e^{-\mathrm{i}s_{(1)}\cdot\left(\frac{x_{(1)}-W_{(1)i}}{b}\right)}K^{\mathrm{ft}}(t)\\
&\quad\times\left(\frac{\hat{f}_\epsilon^{\mathrm{ft}}(t/b)-f_\epsilon^{\mathrm{ft}}(t/b)}{f_\epsilon^{\mathrm{ft}}(t/b)}\right)\left(\frac{1}{1+\frac{\hat{f}_\epsilon^{\mathrm{ft}}(t/b)-f_\epsilon^{\mathrm{ft}}(t/b)}{f_\epsilon^{\mathrm{ft}}(t/b)}}\right)\frac{K^{\mathrm{ft}}\left(s_{(1)}\right)}{\hat{f}_\epsilon^{\mathrm{ft}}\left(s_{(1)}/b\right)}\mathcal{W}(x;\xi)\,dtds_{(1)}dx\\
&= \frac{1}{n^2}\frac{1}{(2\pi b)^{d+d_1}}\sum_{i\neq j}^{n}(Y_i-Y_j)\int\int\int e^{-\mathrm{i}t\cdot\left(\frac{x-W_i}{b}\right)}e^{-\mathrm{i}s_{(1)}\cdot\left(\frac{x_{(1)}-W_{(1)i}}{b}\right)}K^{\mathrm{ft}}(t)\\
&\quad\times\left(\frac{\hat{f}_\epsilon^{\mathrm{ft}}(t/b)-f_\epsilon^{\mathrm{ft}}(t/b)}{f_\epsilon^{\mathrm{ft}}(t/b)}\right)\left(\frac{1}{1+\frac{\hat{f}_\epsilon^{\mathrm{ft}}(t/b)-f_\epsilon^{\mathrm{ft}}(t/b)}{f_\epsilon^{\mathrm{ft}}(t/b)}}\right)\frac{K^{\mathrm{ft}}\left(s_{(1)}\right)}{f_\epsilon^{\mathrm{ft}}\left(s_{(1)}/b\right)}\mathcal{W}(x;\xi)\,dtds_{(1)}dx\\
&\quad-\frac{1}{n^2}\frac{1}{(2\pi b)^{d+d_1}}\sum_{i\neq j}^{n}(Y_i-Y_j)\int\int\int e^{-\mathrm{i}t\cdot\left(\frac{x-W_i}{b}\right)}e^{-\mathrm{i}s_{(1)}\cdot\left(\frac{x_{(1)}-W_{(1)i}}{b}\right)}\frac{K^{\mathrm{ft}}(t)}{f_\epsilon^{\mathrm{ft}}(t/b)^2}\frac{K^{\mathrm{ft}}\left(s_{(1)}\right)}{f_\epsilon^{\mathrm{ft}}\left(s_{(1)}/b\right)^2}\\
&\quad\times\left(\frac{\hat{f}_\epsilon^{\mathrm{ft}}(t/b)-f_\epsilon^{\mathrm{ft}}(t/b)}{1+\frac{\hat{f}_\epsilon^{\mathrm{ft}}(t/b)-f_\epsilon^{\mathrm{ft}}(t/b)}{f_\epsilon^{\mathrm{ft}}(t/b)}}\right)\left(\frac{\hat{f}_\epsilon^{\mathrm{ft}}\left(s_{(1)}/b\right)-f_\epsilon^{\mathrm{ft}}\left(s_{(1)}/b\right)}{1+\frac{\hat{f}_\epsilon^{\mathrm{ft}}\left(s_{(1)}/b\right)-f_\epsilon^{\mathrm{ft}}\left(s_{(1)}/b\right)}{f_\epsilon^{\mathrm{ft}}\left(s_{(1)}/b\right)}}\right)\mathcal{W}(x;\xi)\,dtds_{(1)}dx\\
&\equiv T_{21}-T_{22}.
\end{aligned}$$

As for T_{12} we have

$$\begin{aligned}
T_{21} &= o_p(1)\frac{1}{n^2}\frac{1}{(2\pi b)^{d+d_1}}\sum_{i\neq j}^{n}(Y_i-Y_j)\int\int\int\left\{\begin{array}{c} e^{-\mathrm{i}t\cdot\left(\frac{x-W_i}{b}\right)}e^{-\mathrm{i}s_{(1)}\cdot\left(\frac{x_{(1)}-W_{(1)i}}{b}\right)}\\ \times\frac{K^{\mathrm{ft}}(t)}{f_\epsilon^{\mathrm{ft}}(t/b)}\frac{K^{\mathrm{ft}}\left(s_{(1)}\right)}{f_\epsilon^{\mathrm{ft}}\left(s_{(1)}/b\right)}\end{array}\right\}\mathcal{W}(x;\xi)\,dtds_{(1)}dx\\
&= o_p(1)O_p\left(n^{-1/2}\right).
\end{aligned}$$

For T_{22} we can write

$$\begin{aligned}
T_{22} &= o_p(1)\frac{1}{n^2}\frac{1}{(2\pi b)^{d+d_1}}\sum_{i\neq j}^{n}(Y_i-Y_j)\int\int\int\left\{\begin{array}{c} e^{-\mathrm{i}t\cdot\left(\frac{x-W_i}{b}\right)}e^{-\mathrm{i}s_{(1)}\cdot\left(\frac{x_{(1)}-W_{(1)i}}{b}\right)}\\ \times\frac{K^{\mathrm{ft}}(t)}{f_\epsilon^{\mathrm{ft}}(t/b)^2}\frac{K^{\mathrm{ft}}\left(s_{(1)}\right)}{f_\epsilon^{\mathrm{ft}}\left(s_{(1)}/b\right)^2}\end{array}\right\}\mathcal{W}(x;\xi)\,dtds_{(1)}dx\\
&= o_p(1)O_p\left(n^{-1/2}\right)
\end{aligned}$$

using similar arguments as for T_{12}.

Hence, we have shown $\sqrt{n}\hat{\hat{T}}_n(\xi)=\sqrt{n}\hat{T}_n(\xi)+o_p(1)$. The rest of the proofs of Theorem 1, 2 and 3 can be applied to obtain the result.

3.7.5. Proof of Propositions.

3.7.5.1. *Proof of Proposition 1.* For the proof of part (i) it is straightforward to show that $\tilde{T}_n^*(\xi)$ converges weakly to a zero mean Gaussian process conditional on the data; this follows directly from the simple pairs resampling approach and the proof of Theorem 5. Hence, CM_n^* has the same limiting distribution as the limiting distribution of CM_n under the null, conditional on the data.

For part (ii), we first make use of Proposition 3.1 in Politis and Romano (1994) which shows $\{Z_t\}_{t=1}^T = \{Y_t, W_t\}_{t=1}^T$ is strictly stationary and absolutely regular. The proof then follows as in the proof of Theorem 7.

3.7.5.2. *Proof of Proposition 2.* To study $Var[r_n(Z_i;\xi)]$, we write $r_n(Z_i;\xi)$ as follows

$$\begin{aligned} r_n(Z_i;\xi) &= \int \mathcal{K}_b\left(\frac{x-W_i}{b}\right) \left\{ \begin{array}{c} Y_i E\left[\mathcal{K}_b\left(\frac{x_{(1)}-W_{(1)j}}{b}\right)\middle| Y_i, W_i\right] \\ -E\left[Y_j\mathcal{K}_b\left(\frac{x_{(1)}-W_{(1)j}}{b}\right)\middle| Y_i, W_i\right] \end{array} \right\} \mathcal{W}(x;\xi)dx \\ &\quad + \int \mathcal{K}_b\left(\frac{x_{(1)}-W_{(1)i}}{b}\right) \left\{ \begin{array}{c} E\left[Y_j\mathcal{K}_b\left(\frac{x-X_j}{b}\right)\middle| Y_i, W_i\right] \\ -Y_i E\left[\mathcal{K}_b\left(\frac{x-W_j}{b}\right)\middle| Y_i, W_i\right] \end{array} \right\} \mathcal{W}(x;\xi)dx \\ &\equiv r_{1n}(Z_i;\xi) + r_{2n}(Z_i;\xi). \end{aligned}$$

We deal with each term separately; consider first $r_{1n}(Z_i;\xi)$. Using the fact that
$E\left[Y_j\mathcal{K}_b\left(\frac{x_{(1)}-W_{(1)j}}{b}\right)\middle| Y_i, W_i\right] = E\left[\frac{1}{b^{d_1}}Y_j K\left(\frac{x_{(1)}-X_{(1)j}}{b}\right)\middle| Y_i, W_i\right]$

$$\begin{aligned} r_{1n}(Z_i;\xi) &= b^{-d_1}\int \mathcal{K}_b\left(\frac{x-W_i}{b}\right) \left\{ \begin{array}{c} Y_i E\left[K\left(\frac{x_{(1)}-X_{(1)j}}{b}\right)\middle| Y_i, W_i\right] \\ -E\left[Y_j K\left(\frac{x_{(1)}-X_{(1)j}}{b}\right)\middle| Y_i, W_i\right] \end{array} \right\} \mathcal{W}(x;\xi)dx \\ &= b^{d-d_1}\int \mathcal{K}_b(a) \left\{ \begin{array}{c} Y_i E\left[K\left(a_{(1)} + \frac{W_{(1)i}-X_{(1)j}}{b}\right)\middle| Y_i, W_i\right] \\ -E\left[Y_j K\left(a_{(1)} + \frac{W_{(1)i}-X_{(1)j}}{b}\right)\middle| Y_i, W_i\right] \end{array} \right\} \mathcal{W}(W_i+ab;\xi)da. \end{aligned}$$

Then we can write

$$
\begin{aligned}
& E\left[K\left(a_{(1)}+\frac{W_{(1)i}-X_{(1)j}}{b}\right)\middle|Y_i, W_i\right] \\
= \;& b^{d_1}\int K(a_{(1)}+s_{(1)})f_{X_{(1)}}(W_{(1)i}-s_{(1)}b)ds_{(1)} \\
= \;& b^{d_1}\int K(a_{(1)}+s_{(1)})ds_{(1)}f_{X_{(1)}}(W_{(1)i})+b^{d_1}\int K(a_{(1)}+s_{(1)})\sum_{j=1}^{d_1}s_j bY_i\nabla_j f_{X_{(1)}}\left(W^*_{(1)i}\right)ds_{(1)} \\
= \;& b^{d_1}f_{X_{(1)}}(W_{(1)i})+b^{d_1}\int K(a_{(1)}+s_{(1)})\left(f_{X_{(1)}}(W_{(1)i}-s_{(1)}b)-f_{X_{(1)}}(W_{(1)i})\right)ds_{(1)}
\end{aligned}
$$

where the first equality follows from a change of variables and the fact that the data is i.i.d., the second equality uses a multivariate Taylor expansion where $W^*_{(1)i}$ lies between $W_{(1)i}$ and $(W_{(1)i}-s_{(1)}b)$, and the third equality uses the properties of the kernel stated in Assumption 6 (iii). Hence, we can write

$$
\begin{aligned}
r_{1n}(Z_i;\xi) \;=\;& b^d\left(Y_i f_{X_{(1)}}(W_{(1)i})-R(W_{(1)i})\right)\int \mathcal{K}_b\,(a)\,\mathcal{W}(W_i+ab;\xi)da \\
& +b^{d+1}\sum_{j=1}^{d_1}\int \mathcal{K}_b(a)\left\{\int K(a_{(1)}+s_{(1)})s_j\left(Y_i\nabla_j f_{X_{(1)}}\left(W^*_{(1)i}\right)-\nabla_j R\left(W^*_{(1)i}\right)\right)ds_{(1)}\right\} \\
& \times\mathcal{W}(W_i+ab;\xi)da. \\
\equiv\;& r_{11n}(Z_i;\xi)+r_{12n}(Z_i;\xi).
\end{aligned}
$$

We bound the variance of $r_{11n}(Z_i;\xi)$ as follows

$$
\begin{aligned}
Var[r_{11n}(Z_i;\xi)] \;\le\;& E[r_{11n}(Z_i;\xi)^2] \\
\le\;& b^{2d}||\mathcal{W}(\cdot;\xi)||^2_\infty\left(\int \mathcal{K}_b\,(a)\,da\right)^2 E\left[\left(Y_i f_{X_{(1)}}(W_{(1)i})-R(W_{(1)i})\right)^2\right] \\
\sim\;& b^{2d}\left(\frac{C_0}{b}\right)^{2d}||\mathcal{W}(\cdot;\xi)||^2_\infty E\left[\left(Y_i f_{X_{(1)}}(W_{(1)i})-R(W_{(1)i})\right)^2\right] \\
=\;& O(1)
\end{aligned}
$$

where the wave relation follows from Lemma 5 which shows $\int \mathcal{K}_b\,(a)\,da \sim \left(\frac{C_0}{b}\right)^{\dim(a)}$, and the final equality follows from Assumptions 6 (ii) and (v). Hence

$$
n^{-1/2}\sum_{i=1}^{n}\left(r_{11n}(Z_i;\xi)-E[r_{11n}(Z_i;\xi)]\right) \;=\; O(1).
$$

For $r_{12n}(Z_i;\xi)$ we can bound the variance as

$$
\begin{aligned}
Var\left[r_{12n}(Z_i;\xi)\right] &\leq E\left[r_{12n}(Z_i;\xi)^2\right] \\
&= O\left(b^{2d+2}\right) E\Bigg[\sum_{j=1}^{d_1}\int \mathcal{K}_b(a)\int K(a_{(1)}+s_{(1)})s_j \\
&\quad \times \left(Y_i\nabla_j f_{X_{(1)}}\left(W^*_{(1)i}\right)-\nabla_j R\left(W^*_{(1)i}\right)\right) ds_{(1)}da\Bigg]^2 \\
&= O\left(b^{2d+2}\right) E\left[(1+|Y_i|)^2\right]\left(\int \mathcal{K}_b(a)\left\{\sum_{j=1}^{d_1}\int K(a_{(1)}+s_{(1)})s_j ds_{(1)}\right\} da\right)^2 \\
&= O\left(b^{2(d+1)}\right)\left(\sum_{j=1}^{d_1}\int \mathcal{K}_b(a)a_j da\right)^2 \\
&= O(1)
\end{aligned}
$$

where the first equality follows from Assumption 6 (ii), the second uses Assumption 6 (iv), the penultimate equality follows from the change of variables $a_{(1)}+s_{(1)}=v_{(1)}$ and the properties of a second-order kernel to give $\int K(a_{(1)}+s_{(1)})s_j ds_{(1)}=a_j$, and the final equality makes use of Lemma 5 which can be used to show $\int \mathcal{K}_b\left(a\right)a_1 da \sim \frac{C_1}{2\pi i b^2}\left(\frac{C_0}{b}\right)^{\dim(a)-1}$.

Hence

$$
n^{-1/2}\sum_{i=1}^{n}\left(r_{12n}(Z_i;\xi)-E[r_{12n}(Z_i;\xi)]\right) = O_p(1).
$$

For $r_{2n}(Z_i;\xi)$ we follow a similar approach. Notice

$$
E\left[K\left(x+\frac{W_i-X_j}{b}\right)\middle|Y_i,W_i\right] = b^d\int K(x+s)f_X(W_i-sb)ds.
$$

Then, similarly to $r_{1n}(Z_i;\xi)$, we can split $r_{2n}(Z_i;\xi)$ as

$$
r_{2n}(Z_i;\xi) \equiv r_{21n}(Z_i;\xi)+r_{22n}(Z_i;\xi),
$$

where

$$
r_{21n}(Z_i;\xi) = b^{d_1}\left(Y_i f_X(W_i)-M(W_i)\right)\int \mathcal{K}_b\left(a_{(1)}\right)\mathcal{W}(W_i+ab;\xi)da
$$

and

$$r_{22n}(Z_i;\xi) \quad = \quad b^{d_1+1}\sum_{j=1}^{d}\int \mathcal{K}_b\big(a_{(1)}\big)\left\{\int K(a+s)s_j\left(Y_i\nabla_j f_X\left(W_i^*\right)-\nabla_j M\left(W_i^*\right)\right)ds\right\}\mathcal{W}(W_i+ab;\xi)da.$$

Using arguments very similar to those used to bound $Var[r_{1n}(Z_i;\xi)]$, we have

$$Var\left(r_{2n}(Z_i;\xi)\right) \quad = \quad O(1).$$

Combining these results we have

$$Var\left(r_n(Z_i;\xi)\right) \quad = \quad O(1).$$

This concludes the proof Proposition 2.

3.7.5.3. *Proof of Proposition 3.* To study $Var[r_n(Z_i;\xi)]$, as in the proof of Proposition 2, we write $r_n(Z_i;\xi)$ as follows

$$\begin{aligned} r_n(Z_i;\xi) \quad &= \quad \int \mathcal{K}_b\bigg(\frac{x-W_i}{b}\bigg)\left\{\begin{array}{c} Y_iE\Big[\mathcal{K}_b\Big(\frac{x_{(1)}-W_{(1)j}}{b}\Big)\Big|Y_i,W_i\Big] \\ -E\Big[Y_j\mathcal{K}_b\Big(\frac{x_{(1)}-W_{(1)j}}{b}\Big)\Big|Y_i,W_i\Big]\end{array}\right\}\mathcal{W}(x;\xi)dx \\ &\quad +\int \mathcal{K}_b\bigg(\frac{x_{(1)}-W_{(1)i}}{b}\bigg)\left\{\begin{array}{c} E\Big[Y_j\mathcal{K}_b\Big(\frac{x-X_j}{b}\Big)\Big|Y_i,W_i\Big] \\ -Y_iE\Big[\mathcal{K}_b\Big(\frac{x-W_j}{b}\Big)\Big|Y_i,W_i\Big]\end{array}\right\}\mathcal{W}(x;\xi)dx \\ &\equiv \quad r_{1n}(Z_i;\xi)+r_{2n}(Z_i;\xi). \end{aligned}$$

Again, we can write

$$r_{1n}(Z_i;\xi) \quad \equiv \quad r_{11n}(Z_i;\xi)+r_{12n}(Z_i;\xi),$$

where

$$r_{11n}(Z_i;\xi) \quad \equiv \quad b^d\left(Y_if_{X_{(1)}}(W_{(1)i})-R(W_{(1)i})\right)\int \mathcal{K}_b\left(a\right)\mathcal{W}(W_i+ab;\xi)da,$$

and

$$
\begin{aligned}
r_{12n}(Z_i;\xi) \quad \equiv \quad & b^{d+1}\sum_{j=1}^{d_1}\int \mathcal{K}_b(a)\left\{\int K(a_{(1)}+s_{(1)})s_j\left(Y_i\nabla_j f_{X_{(1)}}\left(W^*_{(1)i}\right)-\nabla_j R\left(W^*_{(1)i}\right)\right)ds_{(1)}\right\} \\
& \times\mathcal{W}(W_i+ab;\xi)da.
\end{aligned}
$$

In the same manner as the proof of Proposition 2, we bound the variance of $r_{12n}(Z_i;\xi)$ as follows

$$
\begin{aligned}
Var\left[r_{12n}(Z_i;\xi)\right] \quad \leq \quad & E\left[r_{12n}(Z_i;\xi)^2\right] \\
= \quad & O\left(b^{2d+2}\right)E\Bigg[\sum_{j=1}^{d_1}\int \mathcal{K}_b(a)\int K(a_{(1)}+s_{(1)})s_j \\
& \times\left(Y_i\nabla_j f_{X_{(1)}}\left(W^*_{(1)i}\right)-\nabla_j R\left(W^*_{(1)i}\right)\right)ds_{(1)}da\Bigg]^2 \\
= \quad & O\left(b^{2d+2}\right)E\left[(1+|Y_i|)^2\right]\left(\int \mathcal{K}_b(a)\left\{\sum_{j=1}^{d_1}\int K(a_{(1)}+s_{(1)})s_j ds_{(1)}\right\}da\right)^2 \\
= \quad & O\left(b^{2(d+1)}\right)\left(\sum_{j=1}^{d_1}\int \mathcal{K}_b(a)a_j da\right)^2 \\
= \quad & O\left(b^{2(d+1)}\right)O\left(b^{-2(d+1)}\right) \\
= \quad & O(1)
\end{aligned}
$$

where the penultimate equality follows from Lemma 6. Hence

$$
n^{-1/2}\sum_{i=1}^{n}\left(r_{12n}(Z_i;\xi)-E[r_{12n}(Z_i;\xi)]\right) \quad = \quad O_p(1).
$$

The remainder of the proof is almost identical to the proof of Proposition 2, hence is omitted for brevity. This concludes the proof Proposition 3.

3.7.5.4. *Proof of Proposition 4.* The proof is very similar to that of Propositions 2 and 3, as such we only outline the parts of the proof that differ as a result of the dependence in the data and concentrate on the ordinary smooth case. By a change of variables, we write $r_{1n}(Z_i;\xi)$ as follows

$$\begin{aligned} r_{1n}(Z_i;\xi) &= b^{-d_1}\int \mathcal{K}_b\left(\frac{x-W_i}{b}\right)\left\{\begin{array}{c} Y_i E\left[K\left(\frac{x_{(1)}-X_{(1)j}}{b}\right)\Big| Y_i, W_i\right] \\ -E\left[Y_j K\left(\frac{x_{(1)}-X_{(1)j}}{b}\right)\Big| Y_i, W_i\right]\end{array}\right\}\mathcal{W}(x;\xi)dx \\ &= b^{d-d_1}\int \mathcal{K}_b(a)\left\{\begin{array}{c} Y_i E\left[K\left(a_{(1)}+\frac{W_{(1)i}-X_{(1)j}}{b}\right)\Big| Y_i, W_i\right] \\ -E\left[Y_j K\left(a_{(1)}+\frac{W_{(1)i}-X_{(1)j}}{b}\right)\Big| Y_i, W_i\right]\end{array}\right\}\mathcal{W}(W_i+ab;\xi)da, \end{aligned}$$

We can also write the following expectation as

$$\begin{aligned} & E\left[K\left(a_{(1)}+\frac{W_{(1)i}-X_{(1)j}}{b}\right)\Big| Y_i, W_i\right] \\ = & \int K\left(a_{(1)}+\frac{W_{(1)i}-v_{(1)}}{b}\right) f_{X_{(1)j}|Y_i,W_i}\left(v_{(1)}\big|Y_i,W_i\right)dv_{(1)} \\ = & \; b^{d_1}\int K\left(a_{(1)}+s_{(1)}\right) f_{X_{(1)j}|Y_i,W_i}\left(W_{(1)i}-s_{(1)}b\big|Y_i,W_i\right)ds_{(1)} \\ = & \; b^{d_1} f_{X_{(1)j}|Y_i,W_i}\left(W_{(1)i}\big|Y_i,W_i\right) + b^{d_1+1}\sum_{j=1}^{d_1}\int K\left(a_{(1)}+s_{(1)}\right) s_j Y_i \nabla_j f_{X_{(1)}}\left(W^*_{(1)i}\right)ds_{(1)} \end{aligned}$$

using similar arguments as in the proof of Proposition 2. Thus, we can split $r_{1n}(Z_i;\xi)$ as

$$r_{1n}(Z_i;\xi) \equiv r_{11n}(Z_i;\xi) + r_{12n}(Z_i;\xi),$$

where

$$r_{11n}(Z_i;\xi) \equiv b^d\left\{Y_i f_{X_{(1)j}|Y_i,W_i}\left(W_{(1)i}\big|Y_i,W_i\right) - R\left(W_{(1)i}\right)\right\}\int \mathcal{K}_b\left(a\right)\mathcal{W}(W_i+ab;\xi)da,$$

and

$$\begin{aligned} r_{12n}(Z_i;\xi) \equiv \; & b^{d+1}\sum_{j=1}^{d_1}\int \mathcal{K}_b(a)\left\{\int K(a_{(1)}+s_{(1)})s_j\left(Y_i\nabla_j f_{X_{(1)j}|Y_i,W_i}\left(W^*_{(1)i}\big|Y_i,W_i\right) - \nabla_j R\left(W^*_{(1)i}\right)\right)ds_{(1)}\right\} \\ & \times\mathcal{W}(W_i+ab;\xi)da. \end{aligned}$$

We bound $E\left[|r_{11n}(Z_i;\xi)|^{2+\tilde{\delta}}\right]$ as follows

$$\begin{aligned}
E\left[|r_{11n}(Z_i;\xi)|^{2+\delta}\right] &\leq b^{d(2+\tilde{\delta})}||\mathcal{W}(\cdot;\xi)||_{\infty}^{2+\tilde{\delta}}\left(\int \mathcal{K}_b\left(a\right)da\right)^{2+\tilde{\delta}} \\
&\quad \times E\left[\left(Y_i f_{X_{(1)j}|Y_i,W_i}\left(W_{(1)i}\middle|Y_i,W_i\right)-R\left(W_{(1)i}\right)\right)^{2+\tilde{\delta}}\right] \\
&\sim O\left(b^{d(2+\tilde{\delta})}\right)\left(\frac{C_0}{b}\right)^{2d}E\left[\left(Y_i f_{X_{(1)j}|Y_i,W_i}\left(W_{(1)i}\middle|Y_i,W_i\right)-R\left(W_{(1)i}\right)\right)^{2+\tilde{\delta}}\right] \\
&= O(1)
\end{aligned}$$

where the wave relation follows from Lemma 5.

We omit the rest of the proof for brevity since it is straightforward to extend the remainder of Proposition 2 to the dependent case in the same manner as we have just shown. This concludes the proof Proposition 4.

3.7.6. Proof of Lemmas.

3.7.6.1. *Proof of Lemma 5.* We can write the deconvolution kernel as follows

$$
\begin{aligned}
\tilde{\mathcal{K}}_b(a) &\sim \frac{1}{2\pi b}\int_{-\infty}^{\infty} e^{-\mathrm{i}t\cdot a}\sum_{v=0}^{\alpha} C_v \left|\frac{t}{b}\right|^v \tilde{K}^{\mathrm{ft}}(t)dt \\
&= \frac{1}{2\pi b}\int_{-\infty}^{\infty} e^{-\mathrm{i}t\cdot a}\sum_{v=0}^{\alpha} C_v \left|\frac{t}{b}\right|^v \tilde{K}^{\mathrm{ft}}(|t|)dt \\
&= \frac{1}{2\pi b}\left(\sum_{v=0}^{\alpha}\frac{C_v}{(2\pi \mathrm{i}b)^v}\int_{-\infty}^{\infty} e^{-\mathrm{i}t\cdot a}\tilde{K}^{(v)\mathrm{ft}}(|t|)dt\right) \\
&= \frac{1}{2\pi b}\left(\sum_{v=0}^{\alpha}\frac{C_v}{(-2\pi \mathrm{i}b)^v}\int_{-\infty}^{\infty} e^{-\mathrm{i}t\cdot a}\tilde{K}^{(v)\mathrm{ft}}(t)dt\right) \\
&= \frac{1}{b}\sum_{v=0}^{\alpha}\frac{C_v}{(-2\pi \mathrm{i}b)^v}\tilde{K}^{(v)}(a),
\end{aligned}
$$

where we have used the fact $\tilde{K}^{ft}(t) = \tilde{K}^{ft}(-t)$ in the first equality, in the second we use the result that the Fourier transform of the p^{th} derivative, $f^{(p)\mathrm{ft}}(t)$, is equal to $(2\pi \mathrm{i}t)^p f^{\mathrm{ft}}(t)$, and in the third we have used the result that for a symmetric kernel its derivative is anti-symmetric. Given this, and the fact that for a second-order kernel we have: for any $p \geq 0$ and $k = \{0,1\}$

$$
\int x_j^k \tilde{K}^{(p)}(x_j)\,dx_j = \begin{cases} (-1)^k k! & \text{for } k = p \\ 0 & \text{for } k \neq p. \end{cases}
$$

Combinig these two results we can write

$$
\int \tilde{\mathcal{K}}_b(x_j)dx_j \sim \frac{C_0}{b}
$$

and

$$
\int x_j \tilde{\mathcal{K}}_b(x_j)dx_j \sim \frac{C_1}{2\pi \mathrm{i}b^2}.
$$

Notice that for the gamma distribution $C_1 = 0$, hence $\int x_j \tilde{\mathcal{K}}_b(x_j)dx_j = 0$.

3.7.6.2. *Proof of Lemma 6.* We can write the deconvolution kernel as follows

$$\begin{aligned}
\tilde{\mathcal{K}}_b(a) &\sim \frac{C}{b(2\pi)} \int_{-\infty}^{\infty} e^{-\mathrm{i}t\cdot a} \left|\frac{t}{b}\right|^{-\gamma_0} e^{\mu|\frac{t}{b}|^{\gamma}} \tilde{K}^{\mathrm{ft}}(t)dt \\
&= \frac{C}{b(2\pi)} \int_{-\infty}^{\infty} e^{-\mathrm{i}t\cdot a} \left|\frac{t}{b}\right|^{-\gamma_0} e^{\mu|\frac{t}{b}|^{\gamma}} \tilde{K}^{\mathrm{ft}}(|t|)dt \\
&= \frac{Ce^1}{b(2\pi)} \int_{-\infty}^{\infty} e^{-\mathrm{i}t\cdot a} \left|\frac{t}{b}\right|^{-\gamma_0} \sum_{v=0}^{\infty} \mu^v \left|\frac{t}{b}\right|^{\gamma v} \frac{1}{v!} \tilde{K}^{\mathrm{ft}}(|t|)dt \\
&= \frac{C}{b(2\pi)} \left(\sum_{v=0}^{\infty} \frac{\mu^v}{(2\pi \mathrm{i} b)^{(\gamma v - \gamma_0)}\, v!} \int_{-\infty}^{\infty} e^{-\mathrm{i}t\cdot a} \tilde{K}^{(\gamma v-\gamma_0)\mathrm{ft}}(|t|)dt \right) \\
&= \frac{C}{b(2\pi)} \left(\sum_{v=0}^{\infty} \frac{\mu^v}{(-2\pi \mathrm{i} b)^{(\gamma v - \gamma_0)}\, v!} \int_{-\infty}^{\infty} e^{-\mathrm{i}t\cdot a} \tilde{K}^{(\gamma v-\gamma_0)\mathrm{ft}}(t)dt \right) \\
&= \frac{1}{b} \sum_{v=0}^{\infty} \frac{C\mu^v}{(-2\pi \mathrm{i} b)^{(\gamma v - \gamma_0)}\, v!} \tilde{K}^{(\gamma v-\gamma_0)}(a)
\end{aligned}$$

where we have used the fact $\tilde{K}^{\mathrm{ft}}(t) = \tilde{K}^{\mathrm{ft}}(-t)$ in the first equality, in the second we use a Taylor expansion around $|t| = 0$ to show $e^{a^{\gamma}} = \sum_{k=0}^{\infty} \frac{a^{\gamma k}}{k!}$, in the third we use the result that the Fourier transform of the p^{th} derivative, $f^{(p)\mathrm{ft}}(t)$, is equal to $(2\pi \mathrm{i} t)^p f^{\mathrm{ft}}(t)$, and the penultimate equality follows from the fact that the derivative of a symmetric function is anti-symmetric. Notice that we require $\gamma - \gamma_0$ to be a natural number for the derivatives of the kernel function to exist. This is satisfied by the majority of common distribution functions.

We also have the following property for a second-order kernel: for any $p \geq 0$ and $k = \{0, 1\}$

$$\int x_j^k \tilde{K}^{(p)}(x_j)\, dx_j = \begin{cases} (-1)^k k! & \text{for } k = p \\ 0 & \text{for } k \neq p. \end{cases}$$

Combinig these two results we can write

$$\int \tilde{\mathcal{K}}_b(x_j) dx_j = \frac{C\mu^{\left(\frac{\gamma_0}{\gamma}\right)}}{b\left(\frac{\gamma_0}{\gamma}\right)!}$$

and

$$\int x_j \tilde{\mathcal{K}}_b(x_j) dx_j = \frac{C\mu^{\left(\frac{1-\gamma_0}{\gamma}\right)}}{b^2 2\pi \mathrm{i} \left(\frac{1-\gamma_0}{\gamma}\right)!}.$$

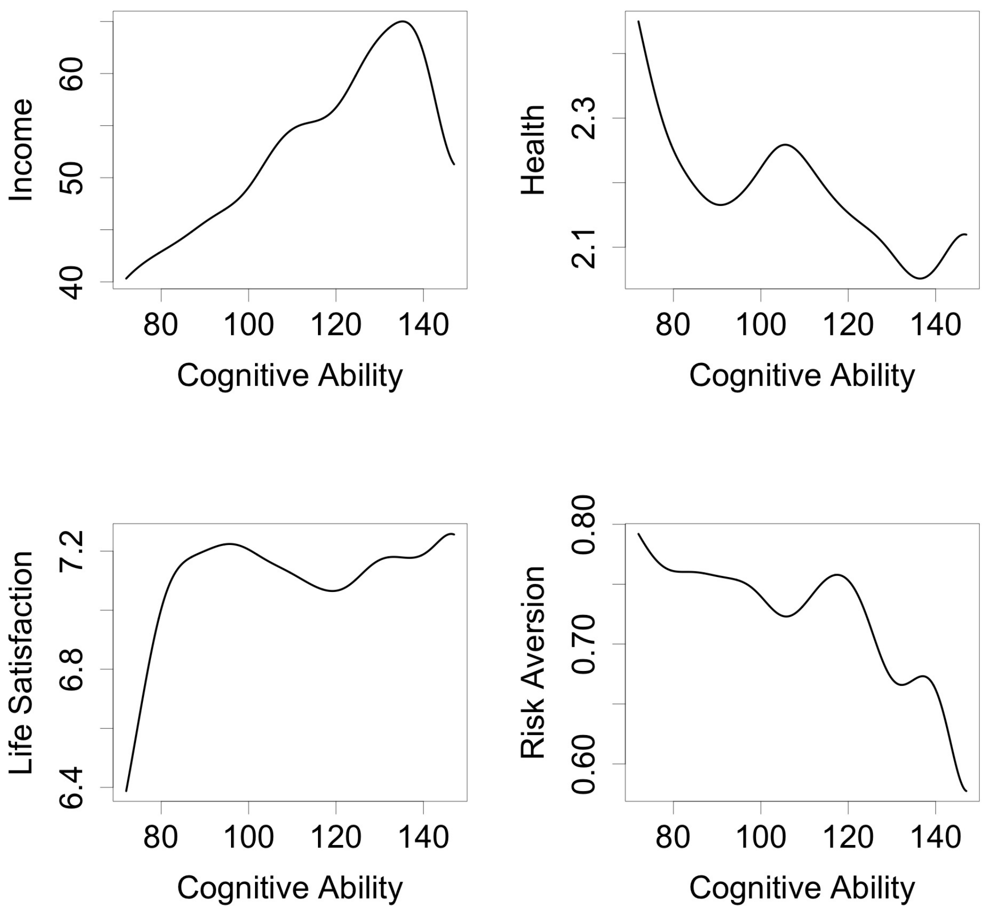

FIGURE 1. Nonparametric plots of the relationship between Cognitive Ability and Income, Health, Life Satisfaction and Risk Aversion. All control variables are set at their respective means and $b = 0.65$.

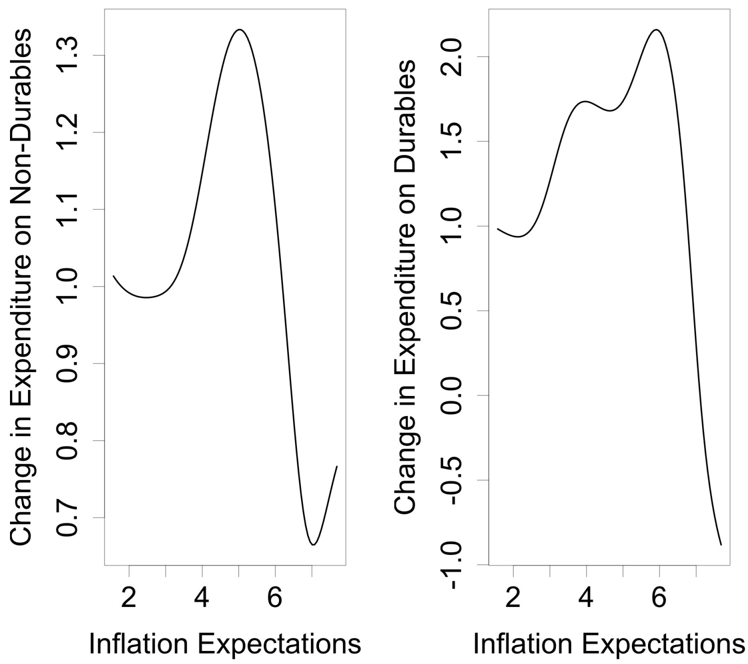

FIGURE 2. Nonparametric plot of the relationship between Inflation Expectations and Change in Expenditure on Durables and Non-Durables. All control variables are set at their respective means and $b = 0.85$.

Bibliography

[1] Aaronson, D. (1998) Using sibling data to estimate the impact of neighborhoods on children's educational outcomes, *Journal of Human Resources*, 33, 915–946.

[2] Ai, C. (1997) A semiparametric maximum likelihood estimator, *Econometrica*, 65, 933-963.

[3] Aït-Sahalia, Y., Bickel, P.J. and T.M. Stoker (2001) Goodness-of-fit tests for kernel regression with an application to option implied volatilities. *Journal of Econometrics*, 105(2), pp.363-412.

[4] Altonji, J. G., Hayashi, F. and L. J. Kotlikoff (1997) Parental altruism and inter vivos transfers: theory and evidence, *Journal of Political Economy*, 105, 1121-1166.

[5] Altonji, J. G., Ichimura, H. and T. Otsu (2008) Estimating derivatives in nonseparable models with limited dependent variables, Cowles Foundation Discussion Paper #1668.

[6] Altonji, J. G., Ichimura, H. and T. Otsu (2012) Estimating derivatives in nonseparable models with limited dependent variables, *Econometrica*, 80, 1701-1719.

[7] Altonji, J. G. and R. L. Matzkin (2005) Cross section and panel data estimators for nonseparable models with endogenous regressors, *Econometrica*, 73, 1053-1102.

[8] Amemiya, T. (1979) The estimation of a simultaneous equation Tobit model, *International Economic Review*, 20, 169-181.

[9] Andrews, D. W. K. (1995) Nonparametric kernel estimation for semiparametric models, *Econometric Theory*, 11, 560-586.

[10] Bachmann, R., Berg, T.O. and E.R. Sims (2015) Inflation expectations and readiness to spend: cross-sectional evidence. *American Economic Journal: Economic Policy*, 7(1), pp.1-35.

[11] Bickel, P. J. and A. J. C. Ritov (1987) Efficient estimation in the errors-in-variables model, *Annals of Statistics*, 15, 513-540.

[12] Bierens, H.J. (1982) Consistent model specification tests. *Journal of Econometrics*, 20(1), pp.105-134.

[13] Bierens, H.J. (1990) A consistent conditional moment test of functional form. *Econometrica,* pp.1443-1458.

[14] Bierens, H.J. and W. Ploberger (1997) Asymptotic theory of integrated conditional moment tests. *Econometrica*, pp.1129-1151.

[15] Blundell, R. and J. L. Powell (2003) Endogeneity in nonparametric and semiparametric regression models, in Dewatripont, M., Hansen, L. P. and S. J. Turnovsky (eds.), *Advances in Economics and Econometrics: Eighth World Congress*, vol. II, Cambridge University Press.

[16] Burke, M.A. and A. Ozdagli (2013) Household inflation expectations and consumer spending: evidence from panel data. *Federal Reserve Bank of Boston* (No. 13-25).

[17] Butucea, C. (2007) Goodness-of-fit testing an quadratic functional estimation from indirect observations, *Annals of Statistics*, 35, 1907-1930.

[18] Butucea, C. and M.-L. Taupin (2008) New M-estimators in semi-parametric regression with errors in variables, *Annales de l'Institut Henri Poincaré-Probabilités et Statistiques*, 44, 393-421.

[19] Carroll, R. J., Spiegelman, C. H., Lan, K. K. G., Bailey, K. T. and R. D. Abbott (1984) On errors-in-variables for binary regression models, *Biometrika*, 71, 19-25.

[20] Cawley, J., Heckman, J. and E. Vytlacil (2001) Three observations on wages and measured cognitive ability. *Labour Economics*, 8(4), pp.419-442.

[21] Chen, S., Dahl, G. B. and S. Khan (2005) Nonparametric identification and estimation of a censored location-scale regression model, *Journal of the American Statistical Association*, 100, 212-221.

[22] Chen, X. and Y. Fan (1999) Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series. *Journal of Econometrics* 91(2), pp.373-401.

[23] Cheng, C.-L. and A. G. Kukush (2004) A goodness-of-fit test for a polynomial errors-in-variables model, *Ukrainian Mathematical Journal,* 56.4 641-661.

[24] Cheng, C.-L. and H. Schneeweiss (1998) Polynomial regression with errors in the variables, *Journal of the Royal Statistical Society: Series B,* 60, 189–199.

[25] Chernozhukov, V. (1998) Nonparametric extreme regression quantiles, Working paper.

[26] Chernozhukov, V., Fernández-Val, I. and A. Kowalski. (2015) Quantile regression with censoring and endogeneity, *Journal of Econometrics* 186.1, 201-221.

[27] Chesher, A. (1991) The effect of measurement error. *Biometrika,* pp.451-462.

[28] Chesher, A. (2003) Local identification in nonseparable models, *Econometrica*, 71, 1405-1441.

[29] Chesher, A. (2010) Instrumental variable models for discrete outcomes, *Econometrica*, 78, 575-601.

[30] Conti, G., Heckman, J.J. and S. Urzua (2011) Early endowments, education, and health. *Unpublished manuscript.*

[31] Correia, I., Farhi, E., Nicolini, J.P. and P. Teles (2013) Unconventional fiscal policy at the zero bound. *The American Economic Review*, 103(4), pp.1172-1211.

[32] D'Acunto, F., Hoang, D. and M. Weber (2016) Unconventional fiscal policy, inflation expectations, and consumption expenditure. *Unpublished manuscript.*

[33] Delaigle, A., Hall, P. and A. Meister (2008) On deconvolution with repeated measurements. *Annals of Statistics*, pp.665-685.

[34] Delgado, M. A., Dominguez, M. A. and P. Lavergne (2006) Consistent tests of conditional moment restrictions, *Annales D'Economie et de Statistique*, 81, 33-67.

[35] Delgado, M.A. and W.G. Manteiga (2001) Significance testing in nonparametric regression based on the bootstrap. *Annals of Statistics*, pp.1469-1507.

[36] Denker, M. and G. Keller (1983) On U-statistics and v. Mises statistics for weakly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 64(4), pp.505-522.

[37] Dohmen, T.J., Falk, A., Huffman, D. and U. Sunde (2010) Are risk aversion and impatience related to cognitive ability? *The American Economic Review*, 100 (3), 1238-1260.

[38] Eggertsson, G.B. (2008) Great expectations and the end of the depression. *The American Economic Review*, 98(4), pp.1476-1516.

[39] Eubank, R. L. and C. H. Spiegelman (1990) Testing the goodness of fit of a linear model via nonparametric regression techniques, *Journal of the American Statistical Association*, 85, 387-392.

[40] Evdokimov, K. (2010) Identification and estimation of a nonparametric panel data model with unobserved heterogeneity. Department of Economics, Princeton University.

[41] Fan, J. (1992) Deconvolution with Supersmooth Distributions, *Canadian Journal of Statistics*, 20, 155-169.

[42] Fan, J. and I. Gijbels (1996) *Local Polynomial Modelling and Its Applications*, CRC Press.

[43] Fan, J. and L. S. Huang (2001) Goodness-of-fit tests for parametric regression models, *Journal of the American Statistical Association*, 96, 640-652.

[44] Fan, J. and E. Masry (1992) Multivariate regression estimation with errors-in-variables: asymptotic normality for mixing processes. *Journal of multivariate analysis*, 43(2), pp.237-271.

[45] Fan, J. and Y. K. Truong (1993) Nonparametric regression with errors in variables, *Annals of Statistics*, 4, 1900-1925.

[46] Fan, J., Yao, Q. and H. Tong (1996) Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems, *Biometrika*, 83, 189-206.

[47] Fan, Y. (1995) Average derivative estimation with errors-in-variables. *Journal of Nonparametric Statistics*, 4(4), pp.395-407.

[48] Fan, Y. and Q. Li (1996) Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica*, pp.865-890.

[49] Fan, Y. and Q. Li (2000) Consistent model specification tests: Kernel-based versus Bierens' ICM tests, *Econometric Theory*, 16, 1016-1041.

[50] Fuller, W. A. (1987) *Measurement Error Models*, Wiley.

[51] Gensowski, M., Heckman, J. and P. Savelyev (2011) The effects of education, personality, and IQ on earnings of high-ability men. *Unpublished manuscript.*

[52] Gleser, L. J. (1981) Estimation in a multivariate "errors in variables" regression model: large sample results, *Annals of Statistics*, 9, 24-44.

[53] Hall, P. (1984) Central limit theorem for integrated square error of multivariate nonparametric density estimators, *Journal of Multivariate Analysis*, 14, 1-16.

[54] Hall, P. and Y. Ma (2007) Testing the suitability of polynomial models in error-in-variables problems, *Annals of Statistics*, 35, 2620–2638.

[55] Hall, P., Wolff, R. C. L. and Q. Yao (1999) Methods for estimating a conditional distribution function, *Journal of the American Statistical Association*, 94, 154-163.

[56] Hall, R.E. (2011) The long slump. *The American Economic Review*, 101(2), pp.431-469.

[57] Härdle, W. and E. Mammen (1993) Comparing nonparametric versus parametric regression fits, *Annals of Statistics*, 21, 1926-1947.

[58] Hausman, J., Newey, W. and J. Powell (1995) Nonlinear Errors in Variables. Estimation of Some Engel Curves, *Journal of Econometrics*, 65, 205–233.

[59] Heckman, J. J., Matzkin, R. and L. Nesheim (2010) Nonparametric identification and estimation of nonadditive hedonic models, *Econometrica*, 78, 1569-1591.

[60] Heckman, J.J., Stixrud, J. and S. Urzua (2006) The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior (No. w12006). *National Bureau of Economic Research.*

[61] Hernstein, R.J. and C. Murray (1994) The bell curve. Free Press. New York.

[62] Hoderlein, S. and E. Mammen (2007) Identification of marginal effects in nonseparable models without monotonicity, *Econometrica*, 75, 1513-1518.

[63] Holzmann, H., Bissantz, N. and A. Munk (2007) Density testing in a contaminated sample, *Journal of multivariate analysis*, 98, 57-75.

[64] Holzmann, H. and L. Boysen (2006) Integrated square error asymptotics for supersmooth deconvolution, *Scandinavian Journal of Statistics*, 33, 849-860.

[65] Horowitz, J. L. and V. G. Spokoiny (2001) An adaptive, rate optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, 69, 599-631.

[66] Ichimura, H., Otsu, T. and J. G. Altonji (2013) Nonparametric intermediate order regression quantiles, Working paper.

[67] Imbens, G. W. and W. K. Newey (2009) Identification and estimation of triangular simultaneous equations models without additivity, *Econometrica*, 77, 1481-1512.

[68] Jofre-Bonet, M. and M. Pesendorfer (2003) Estimation of a dynamic auction game, *Econometrica*, 71, 1443-1489.

[69] Juster, F.T., Wachtel, P., Hymans, S. and J. Duesenberry (1972) Inflation and the Consumer. *Brookings Papers on Economic Activity*, 1972(1), pp.71-121.

[70] Kaziango, H. (2006) Motives for household private transfers in Burkina Faso, *Journal of Development Economics*, 79, 73-117.

[71] Koul, H. L. and W.-X. Song (2009) Minimum distance regression model checking with Berkson measurement errors, *Annals of Statistics*, 37, 132-156.

[72] Koul, H. L. and W.-X. Song (2010) Model checking in partial linear regression models with Berkson measurement errors, *Statistica Sinica*, 20, 1551-1579.

[73] Krugman, P. (2012) Not enough inflation. *New York Times.*

[74] Krugman, P.R., Dominquez, K.M. and K. Rogoff (1998) It's baaack: Japan's slump and the return of the liquidity trap. *Brookings Papers on Economic Activity*, 1998(2), pp.137-205.

[75] Lavergne, P., Maistre, S. and V. Patilea (2015) A significance test for covariates in nonparametric regression. *Electronic Journal of Statistics*, 9, pp.643-678.

[76] Lavergne, P. and Q.H. Vuong (1996) Nonparametric selection of regressors: The nonnested case. *Econometrica*, pp.207-219.

[77] Lavergne, P. (2001) An equality test across nonparametric regressions. *Journal of Econometrics*, 103(1), pp.307-344.

[78] Leser, C. (1963) Forms of Engel Functions, *Econometrica*, 31, 694–703.

[79] Lewbel, A. and O. Linton (2002) Nonparametric censored and truncated regression, *Econometrica*, 70, 765-779.

[80] Li, Q. (1999) Consistent model specification tests for time series econometric models. *Journal of Econometrics*, 92(1), pp.101-147.

[81] Li, T. and Q.H. Vuong (1998) Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis*, 65(2), pp.139-165.

[82] Ma, Y., Hart, J. D., Janicki, R. and R. J. Carroll (2011) Local and omnibus goodness-of-fit tests in classical measurement error models, *Journal of the Royal Statistical Society: Series B,* 73, 81–98

[83] Masry, E. (1993) Strong consistency and rates for deconvolution of multivariate densities of stationary processes. *Stochastic processes and their applications*, 47(1), pp.53-74.

[84] Matzkin, R. L. (2007) Nonparametric identification, in Heckman, J. J. and E. E. Leamer (eds.), Handbook of Econometrics, vol. 6B, ch. 73.

[85] Mavroeidis, S., Plagborg-Møller, M. and J.H. Stock (2014) Empirical evidence on inflation expectations in the New Keynesian Phillips Curve. *Journal of Economic Literature*, 52(1), pp.124-188.

[86] Meister, A. (2009) *Deconvolution Problems in Nonparametric Statistics*, Springer.

[87] McDonald, J. F. and R. A. Moffitt (1980) The uses of Tobit analysis, *Review of Economics and Statistics*, 62, 318-321.

[88] Neumeyer, N. and H. Dette (2003) Nonparametric comparison of regression curves: an empirical process approach. *Annals of Statistics*, 31(3), pp.880-920.

[89] Otsu, T. and L. Taylor (2016) Specification testing for errors-in-variables models. *Unpublished Manuscript.*

[90] Politis, D.N. and J.P. Romano (1994) Limit theorems for weakly dependent Hilbert space valued random variables with applications to the stationary bootstrap. *Statistica Sinica*, 4(2), pp.461-476.

[91] Powell, J.L., Stock, J.H. and T.M. Stoker (1989) Semiparametric estimation of index coefficients. *Econometrica*, pp.1403-1430.

[92] Racine, J. (1997) Consistent significance testing for nonparametric regression. *Journal of Business & Economic Statistics*, 15(3), pp.369-378.

[93] Raut, L. K. and L. H. Tran (2005) Parental human capital investment and old-age transfers from children: Is it a loan contract?, *Journal of Development Economics*, 77, 2, 389-414.

[95] Robinson, P.M. (1989) Hypothesis testing in semiparametric and nonparametric models for econometric time series. *Review of Economic Studies*, 56(4), pp.511-534.

[96] Romer, C.D. (2011) Dear Ben: it's time for your Volcker moment. *New York Times.*

[97] Romer, C.D. and D.H. Romer (2013) The missing transmission mechanism in the monetary explanation of the Great Depression. *The American Economic Review*, 103(3), pp.66-72.

[98] Rosenblatt, M. (1975) A quadratic measure of deviation of two-dimensional density estimates and a test of independence, *Annals of Statistics*, 1-14.

[99] Schennach, S. (2004) Estimation of nonlinear models with measurement error, *Econometrica* 72, 33–75.

[100] Schennach, S.M. (2013) Measurement error in nonlinear models–a review. *Advances in Economics and Econometrics*, 3, pp.296-337.

[101] Schennach, S. M., White, H. and K. Chalak (2012) Local indirect least squares and average marginal effects in nonseparable structural systems, *Journal of Econometrics*, 166, 282-302.

[102] Serfling, R. J. (1980) *Approximation Theorems of Mathematical Statistics*, Wiley.

[103] Shafir, E., Diamond, P. and A. Tversky (1997) Money illusion. *The Quarterly Journal of Economics*, pp.341-374.

[104] Stinchcombe, M.B. and H. White (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric theory*, 14(03), pp.295-325.

[105] Stute, W. (1997) Nonparametric model checks for regression. *Annals of Statistics*, pp.613-641.

[106] Shishko, R. and B. Rostker (1976) The economics of multiple job holding, *American Economic Review*, 66, 298-308.

[107] Smith, R. J. and R.W. Blundell (1986) An exogeneity test for a simultaneous equation Tobit model with an application to labor supply, *Econometrica*, 1986, 679-685. Azzalini, A., Bowman, A. W. and W. Härdle (1989) On the use of nonparametric regression for model checking, *Biometrika*, 76, 1-11.

[108] Song, W.-X. (2008) Model checking in errors-in-variables regression, *Journal of Multivariate Analysis*, 99, 2406-2443.

[109] Song, W.-X. (2009) Lack-of-fit testing in errors-in-variables regression model with validation data, *Statistics & Probability Letters*, 79, 765-773.

[110] Taupin, M.-L. (2001) Semi-parametric estimation in the nonlinear structural errors-in-variables model, *Annals of Statistics*, 29, 66-93.

[111] van der Vaart, A. W. (1988) Estimating a real parameter in a class of semiparametric models, *Annals of Statistics*, 16, 1450-1474.

[112] van Es, A. J. and H.-W. Uh (2005) Asymptotic normality for kernel type deconvolution estimators, *Scandinavian Journal of Statistics*, 32, 467-483.

[113] Veenhoven, R. and Y. Choi (2012) Does intelligence boost happiness? Smartness of all pays more than being smarter than others. *International Journal of Happiness and Development*, 1(1), pp.5-27.

[114] Volcker, P.A. (2011) A Little Inflation Can Be A Dangerous Thing. *New York Times.*

[115] White, H. (1989) An additional hidden unit test for neglected nonlinearity in multilayer feedforward networks. *In International Joint Conference on Neural Networks* (pp. 451-455). IEEE.

[116] Xu, W.-L. and L.-X. Zhu (2015) Nonparametric check for partial linear errors-in-covariables models with validation data, *Annals of Institute of Statistical Mathematics*, 67, 793-815.

[117] Zhu, L.-X. and H.-J. Cui (2005) Testing the adequacy for a general linear errors-in-variables model, *Statistica Sinica*, 15, 1049-1068.

[118] Zhu, L.-X. Song, W.-X. and H.-J. Cui (2003) Testing lack-of-fit for a polynomial errors-in-variables model, *Acta Mathematicae Applicatae Sinic*a, 19, 353-362.