NONPARAMETRIC SIGNIFICANCE TESTING IN MEASUREMENT ERROR MODELS

LUKE TAYLOR*

Abstract. This paper 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: income, life satisfaction, health and risk aversion. 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. This is an important policy question when nominal interest rates approach the zero lower bound.

Key words and phrases. Significance test; Measurement error; Nonparametric regression; Deconvolution; $\sqrt{n}$ convergence.

*London School of Economics, Department of Economics. I am greatly indebted to Peter Robinson and Taisuke Otsu for their invaluable help throughout this project. I am also grateful to Karun Adusumilli, Hao Dong, Joachim Freyberger, Javier Hidalgo, Tatiana Komarova, Yukitoshi Matsushita, Jörn-Steffen Pischke, Ricardo Reis, Marcia Schafgans and participants at the Bristol Econometrics Group Conference and the LSE Econometrics Work-in-Progress Seminar for helpful comments. All remaining errors are my own. Financial support from the ESRC is gratefully acknowledged.

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 paper 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 paper 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 paper, 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

[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.

exacerbated in the presence of measurement error we follow a non-smoothing approach in this paper.

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 Otsu and Taylor (2016) proposing tests for this scenario.

Finally, in this paper 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 paper is organised as follows. Section 2 outlines the hypothesis of interest and the test statistic, as well as discussing possible alternatives to our test. Section 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 4 considers the small sample performance of our test through a Monte Carlo study. Section 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 6 concludes. We relegate all mathematical proofs to Appendix A.

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 paper

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 paper to deal with any form of measurement error, we start with the classical measurement error assumption and leave non-classical error for future work.[2] The case of unknown $f_\epsilon(\cdot)$ is considered in Section 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. 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 paper to deal with such a situation.

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

$$E\big[(Y - E[Y|X_{(1)}])|X\big] = 0 \qquad a.s. \tag{2.1}$$

[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.

In the spirit of Bierens (1982, 1990), we can write (2.1) 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, 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. Note that throughout this paper we denote the Fourier transform of any function, $h(x)$, as $h^{\mathrm{ft}}(t) = \int e^{\mathrm{i}t\cdot x} h(x) dx$ where $\mathrm{i} = \sqrt{-1}$.

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 paper 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 $K^{\mathrm{ft}}(t) = \mathcal{I}(-1 \leq t \leq 1)$, 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 $K^{\mathrm{ft}}(t) = (1-t^2)^3\mathcal{I}(-1 \leq t \leq 1)$.

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 paper; we conjecture that very similar results would be found with the Kolmogorov-Smirnov form of the test.

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 $m(w)$,

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

where $\sigma_{\epsilon_j}^2$ 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

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

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

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

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 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. Asymptotic Properties

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. For convenience we also introduce the multi-index notation: for the vector $\vec{l} = (l_1, \ldots, l_d)$ define $|\vec{l}| = l_1 + \cdots + l_d$, $\vec{l}! = l_1! \cdots l_d!$ and for some $x = (x_1, \ldots, x_d)$, $x^{\vec{l}} = x_1^{l_1} \cdots x_d^{l_d}$. 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 D.

(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 an infinite order kernel and 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*

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

(iv): $M(\cdot) \equiv f_X(\cdot)m(\cdot)$ *and* $R(\cdot) \equiv f_{X_{(1)}}(\cdot)r(\cdot)$ *have* q *continuous derivatives and the following Lipschitz conditions hold for some* $g_{(1)}(\cdot)$ *and* $g(\cdot)$

$$\begin{aligned} \left|\nabla^\rho_{l_1,\ldots l_{d_1}} R(x_{(1)} + \eta_{(1)}) - \nabla^\rho_{l_1,\ldots l_{d_1}} R(x_{(1)})\right| &\leq g_{(1)}(x_{(1)})||\eta_{(1)}||_1, \\ \left|\nabla^\rho_{l_1,\ldots l_d} M(x + \eta) - \nabla^\rho_{l_1,\ldots l_d} M(x)\right| &\leq g(x)||\eta||_1, \end{aligned}$$

for all $\rho \leq q$*, where* $\nabla^\rho_{l_1,\ldots l_d} f(s) = \frac{\partial^\rho f(s)}{\partial s_{l_1}\cdots\partial s_{l_d}}$.

(v): *Each of the following are finite:* $E\left[Y^4\right]$, $E\left[m(X)^4\right]$, $E\left[r(X_{(1)})^4\right]$, $E\left[g(X)^4\right]$ *and* $E\left[g_{(1)}(X_{(1)})^4\right]$.

(vi): *As* $n \to \infty$ *it holds that* $nb^{(d+d_1)\frac{5}{4}} \to \infty$ *and* $nb^{2(q+1)} \to 0$.

Assumption D (i) is common in the literature of classical measurement error. The case of unknown $f_\epsilon(\cdot)$ is deferred until Section 3.4. For Assumption D (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 2. Although Assumption D (iii) is fairly restrictive, it is satisfied by the commonly used Sinc kernel. Assumption D (iv) and (v) give smoothness restrictions on the functions $m(\cdot)$, $r(\cdot)$, $g(\cdot)$ and $g_{(1)}(\cdot)$ as well as Y; conditions on q are discussed below. The first condition of Assumption D (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.

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

Assumption O.

(i): *$f^{\mathrm{ft}}_{\epsilon}(t) \neq 0$ 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*

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

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

(ii): *$q \geq k$ for some $C_k \neq 0$ with $k > 0$.*

Assumption O (i) is the ordinary smooth condition. Specifically, it requires that $\tilde{f}^{\mathrm{ft}}_{\epsilon_j}(s)$ 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, $\tilde{f}^{\mathrm{ft}}_{\epsilon_j}(s) \sim C|s|^{-\alpha}$. Assumption O (ii) requires sufficient smoothness from $M(\cdot)$ and $R(\cdot)$ relative to the density of the measurement error, for example, in the case of Laplace error we would require $M(\cdot)$ and $R(\cdot)$ to be twice continuously differentiable.

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

Assumption S.

(i): *$f^{\mathrm{ft}}_{\epsilon}(t) \neq 0$ for all $t \in \mathbb{R}^d$ and there exist positive constants C, μ, γ_0 and γ such that*

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

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

(ii): *$q \geq \gamma_0 + \gamma$.*

Assumption S (i) requires $\tilde{f}^{\mathrm{ft}}_{\epsilon_j}(s)$ 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. Assumption S (ii) requires sufficient smoothness of $M(\cdot)$ and $R(\cdot)$ relative to the measurement error density, for example for the case of Gaussian error we require $M(\cdot)$ and $R(\cdot)$ to be twice continuously differentiable. 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 1.

(i): *Suppose that Assumptions D and O 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 D and S 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 1 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 D (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.

3.2. **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 $c_n \to 0$ and $\Delta(\cdot)$ is a bounded, non-zero function. The local power properties are given by the following theorem.

Theorem 2.

(i): *Suppose that Assumptions D and O 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 D and S 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 2 shows that under both ordinary smooth and supersmooth measurement error our test is able to detect local, linear alternatives drifting at the rate $c_n = n^{-1/2}$.

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. **Dependent Data.** To increase the applicability of the test proposed in this paper, 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 $1 \leq i, j \leq d$, however we concentrate on the i.i.d. case for ease of derivations.

We wish to show that our previous results continue to hold under weak dependence. To do so, we require the following additional assumptions.

Assumption T.

(i): *$\beta(j) = O(j^{-\eta})$ for a particular $\eta > 0$ which is discussed in Appendix A.*

(ii): *For some $\delta,\ \varsigma > 0$, as $n \to \infty$ it holds that $n^{1-\frac{\varsigma}{2}} b^{(d+d_1)\frac{5(2+\delta)}{8}} \to \infty$.*

(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 $R_{X_{(1)j}|W_{(1)i}}(x_{(1)}|w_{(1)}) \equiv r(x_{(1)}) f_{X_{(1)j}|W_{(1)i}}(x_{(1)}|w_{(1)})$ have q continuous derivatives with respect to x and $x_{(1)}$, respectively, and the following Lipschitz conditions hold for some $h_{(1)}(\cdot,\cdot)$ and $h(\cdot,\cdot)$*

$$\begin{aligned} \left|\nabla^{\rho}_{l_1,\ldots l_{d_1}} R_{X_{(1)j}|W_{(1)i}}(x_{(1)} + \eta_{(1)}|x_{(1)}) - \nabla^{\rho}_{l_1,\ldots l_{d_1}} R_{X_{(1)j}|W_{(1)i}}(x_{(1)}|x_{(1)})\right| &\leq h_{(1)}(x_{(1)}, x_{(1)})||\eta_{(1)}||_1, \\ \left|\nabla^{\rho}_{l_1,\ldots l_d} M_{X_j|W_i}(x+\eta|x) - \nabla^{\rho}_{l_1,\ldots l_d} M_{X_j|W_i}(x|x)\right| &\leq h(x,x)||\eta||_1, \end{aligned}$$

for all $\rho \leq q$ and $1 \leq i < j \leq n$, where it is understood that the derivative is taken with respect to the first set of arguments, and $E\left[|h(W_i,x)|^{2+\delta}\right] < \infty$.

Assumption T (i) concerns the degree of dependence between events separated in time; the larger η the more quickly the dependence decays to zero. Assumptions T (ii)-(iv) require a slight strengthening of the corresponding Assumptions D (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 3.

(i): *Suppose that Assumptions D, O and T (i)-(v) hold true, then under* H_0,

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

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 D, S, T (i)-(iv) and (vi) 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.

3.4. **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. With 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

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

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 true 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 $f_\epsilon^{\text{ft}}(\cdot)$ with the test using $\hat{f}_\epsilon^{\text{ft}}(t)$. We first introduce the following additional assumptions.

Assumption U.

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

(ii): $q \geq 2\alpha$.

(iii): $q \geq 2(\gamma_0 + \gamma)$.

Assumption U (i) is satisfied by the estimators of Li and Vuong (1998) (see Lemma 4 in Evdokimov, 2010) and Delaigle, Hall and Meister (2008). Notice, Assumption U (ii) and (iii) require a slight strengthening of Assumption O (ii) and Assumption S (ii), respectively.

Theorem 4.

(i): *Suppose that Assumptions D, U (i) and either O and U (ii) or S and U (iii) hold true, then Theorems 1 and 2 continue to hold if $f_\epsilon^{\text{ft}}(\cdot)$ is replaced with $\hat{f}_\epsilon^{\text{ft}}(t)$.*

(ii): *Suppose that Assumptions D, T, U (i) and either O and U (ii) or S and U (iii) hold true, then Theorem 3 continues to hold if $f_\epsilon^{\text{ft}}(\cdot)$ is replaced with $\hat{f}_\epsilon^{\text{ft}}(t)$.*

3.5. **Bootstrap.** The asymptotic distributions derived in Theorem 1 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

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

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

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

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 $k > T$, $Z_k = Z_{k \bmod T}$ 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 $\{Y_i, W_i\}_{i=1}^n$.*

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 $\mathcal{W}(\cdot;\xi) = \mathcal{I}(\cdot \leq \xi)$, which satisfies Assumption D (ii). For all simulations we use the Sinc kernel

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

which satisfies Assumptions D (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 D (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 $b_{DM} = \left(\frac{1}{n}\right)^{1/3d_2}$; 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 3.5 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 paper, 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	$b_0 + 0.25$	$b_{DM} - 0.25$	b_{DM}	$b_{DM} + 0.25$	
100	**5%**	4.9	4.3	4.5	5.2	5.3	6.7	4.1
	10%	9.0	9.4	9.9	10.8	11.2	15.7	8.1
200	**5%**	4.8	5.1	4.9	5.2	5.7	7.9	4.6
	10%	10.2	11.0	11.0	10.9	10.6	15.5	8.1
Super Smooth								
100	**5%**	5.4	5.9	6.2	4.6	4.1	5.8	4.9
	10%	10.2	9.8	10.0	10.0	10.3	13.7	9.8
200	**5%**	4.7	4.7	4.3	5.0	4.9	7.2	3.9
	10%	10.1	9.7	9.5	10.5	10.8	14.8	8.6

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	$b_0 + 0.25$	$b_{DM} - 0.25$	b_{DM}	$b_{DM} + 0.25$	
100	**5%**	12.2	15.2	16.2	8.2	7.0	6.4	4.5
	10%	20.3	24.0	25.1	14.8	13.3	12.7	10.4
200	**5%**	16.2	24.5	26.7	8.5	7.5	7.9	6.6
	10%	27.5	33.9	36.7	18.4	16.8	15.1	13.5
Super Smooth								
100	**5%**	12.2	14.8	16.2	7.8	6.7	6.1	5.4
	10%	20.6	23.9	25.9	15.1	13.8	11.9	11.3
200	**5%**	18.8	28.0	31.7	6.8	6.6	6.5	9.9
	10%	29.3	39.2	44.3	13.7	13.6	12.6	18.6

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	$b_0 + 0.25$	$b_{DM} - 0.25$	b_{DM}	$b_{DM} + 0.25$	
100	**5%**	7.2	8.0	7.7	5.7	5.3	7.5	4.4
	10%	14.1	12.6	13.0	11.4	11.2	15.7	8.9
200	**5%**	11.0	11.9	11.6	6.2	5.9	8.1	3.8
	10%	17.7	20.5	20.2	13.5	11.4	16.6	8.9
Super Smooth								
100	**5%**	7.8	8.3	8.4	5.0	4.2	6.4	3.8
	10%	16.0	15.1	15.7	11.0	10.5	13.7	8.9
200	**5%**	10.7	11.2	11.1	5.6	5.0	7.5	3.7
	10%	18.9	20.3	20.6	11.6	10.7	15.3	8.5

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	$b_0 + 0.25$	$b_{DM} - 0.25$	b_{DM}	$b_{DM} + 0.25$	
100	**5%**	58.5	63.3	63.0	28.4	20.8	15.3	14.5
	10%	71.7	76.0	75.1	48.3	38.0	32.0	23.9
200	**5%**	89.1	91.4	90.3	69.7	58.5	41.3	44.2
	10%	96.7	99.3	95.6	86.8	78.7	67.4	58.3
Super Smooth								
100	**5%**	56.3	63.8	65.0	19.1	15.0	10.6	20.8
	10%	70.4	76.5	77.4	37.3	30.0	26.2	31.7
200	**5%**	85.7	92.6	92.9	53.6	38.3	26.8	66.7
	10%	94.1	96.4	96.6	77.8	64.5	51.2	78.0

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 good size control despite being invalid. This is likely due to the bootstrap procedure being used in that case. However, this is some size distortion for larger bandwidths, this is also found and discussed in DM. The t-test based on the IV seems slightly undersized, although as the sample size was increased this approached the nominal level[4].

In each of the three alternatives considered the test proposed in this paper 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 an upper benchmark with which to compare our nonparametric test, however, our test clearly dominates. This reflects the arguments given in Section 2.1. As can be seen in Table 4, as we increase the sample size the IV test does gain considerable power but still seems to lag behind the test proposed in this paper. In general, increasing the amount of measurement error or increasing the degree of nonlinearities in the model, increases the gap between our test and the two alternatives. It is encouraging to see little dependence on the bandwidth across all three alternative specifications.

.

[4]The results are available from the author upon request.

5. Applications

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 2 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 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 two IQ tests as repeated noisy measurements of true cognitive ability and use the estimator proposed by Delaigle, Hall and Meister (2008) to estimate the Fourier transform of the measurement error. We use the i.i.d. bootstrap procedure discussed in Section 3.5, using the same parameter settings as in Section 4. We also use the same kernel as used in

Section 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 shows the asymptotic properties of our test do not depend on the choice of bandwidth, providing they satisfy Assumption D (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 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**	**0.25**	**0.45**	**0.65**		
Income	0.000	0.005	0.005	0.111	0.051	0.000	0.002	0.000
Health	0.085	0.055	0.045	0.965	0.673	0.372	0.270	0.055
Life Satisfaction	0.286	0.698	0.703	0.548	0.186	0.191	0.090	0.111
Risk Aversion	0.477	0.563	0.673	0.593	0.467	0.633	0.744	0.821

Our results agree with the previous literature in finding that cognitive ability has a significant impact on income and health. We also find an insignificant 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. Whilst in the Health and Life Satisfaction

regressions there are some clear differences in conclusions depending on which test is used. For example, for the effect on health, it appears that in a nonparametric regression accounting for measurement error has a substantial effect which cannot simply be ignored. 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 B for the interested reader.

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

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

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

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

and, by lagging this difference by one period

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

where

$$v_{t-1} = E_{t-2}[\pi_{t+1}] - E_{t-1}[\pi_{t+1}],$$

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

Hence

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

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 5.1 we choose the bandwidth $b = \left(\frac{1}{n}\right)^{1/2(d+d_1)} \approx 0.85$. 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 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							
	0.65	**0.85**	**1.05**	**0.3**	**0.5**	**0.7**		
Durables	0.004	0.028	0.062	0.554	0.363	0.283	0.090	0.102
Non-Durables	0.082	0.096	0.112	0.737	0.446	0.243	0.014	0.422

Our results indicate that there does appear to be a significant relationship between inflation expectations and current expenditure. 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.

As the theory would suggest there is a stronger relationship for durable goods, over non-durables. This is because durable goods are more likely to be bought on credit and their purchase can be more easily substituted from one period to the next. Appendix B 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. This may help to explain why, in the previous literature, linear specifications have been unable to find a significant relationship, and why many 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.[5] 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.

6. Conclusion

This paper 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 paper. 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

[5]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 paper to tests of general conditional moment equalities, or to add to the growing literature on testing conditional moment inequalities.

Appendix A. Mathematical Appendix

A.1. **Proof of Theorem 1.**

A.1.1. *Proof of (i).* Throughout this and the proceeding proofs we will make use of the following Lemma.

Lemma 1. *Under Assumptions D and O*

$$\int x_j^k \tilde{\mathcal{K}}_b(x_j)\, dx_j \sim \begin{cases} \frac{C_k(k!)}{b(2\pi \mathrm{i} b)^k} & \text{for } k \leq \alpha, \\ 0 & \text{otherwise.} \end{cases}$$

□

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

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

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

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

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

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

where

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

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 $m_2(x) \equiv E\left[Y^2|X = x\right]$,

$$\begin{aligned}
& E\left[|p_n(Z_i, Z_j)|^2\right] \\
\leq\ & 4E\left|(Y_i - Y_j)\int \mathcal{K}_b\left(\frac{x - W_i}{b}\right)\mathcal{K}_b\left(\frac{x_{(1)} - W_{(1)j}}{b}\right)\mathcal{W}(x;\xi)dx\right|^2 \\
\leq\ & 8E\left[m_2(X_i)\left|\int \mathcal{K}_b\left(\frac{x - W_i}{b}\right)\mathcal{K}_b\left(\frac{x_{(1)} - W_{(1)j}}{b}\right)\mathcal{W}(x;\xi)dx\right|^2\right], \qquad \text{(A.1)}
\end{aligned}$$

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

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

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 1 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 O (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 $V_O(\xi,\xi') = \lim_{n\to\infty} E\left[(r_n(Z_i;\xi)-\theta_n(\xi))\,(r_n(Z_i;\xi')-\theta_n(\xi'))\right]$ (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 D and O*

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

□

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

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

For T_1

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

where

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

By similar arguments

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

where

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

So

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

where the final equality follows by four multivariate Taylor expansions and the properties of the infinite order kernel, for example

$$
\begin{aligned}
&\int K(u)M(x-ub)du \\
=\;& M(x)+h_1(x,u)b+\cdots+h_{q-1}(x,u)b^{q-1}+\tilde{h}_q(x,u)b^q
\end{aligned}
$$

where

$$
h_\rho(x,u) = \frac{1}{\rho!}\sum_{|\vec{l}|=\rho}\nabla^\rho_{l_1,\ldots l_d}M(x)\int u_1^{l_1}\cdots u_d^{l_d}K(u)du,
$$

and

$$
\begin{aligned}
\tilde{h}_q(x,u) &= \frac{1}{q!}\sum_{|\vec{l}|=q}\int\nabla^q_{l_1,\ldots l_d}M(x^*)u_1^{l_1}\cdots u_d^{l_d}K(u)du \\
&= \frac{1}{q!}\sum_{|\vec{l}|=q}\int\nabla^q_{l_1,\ldots l_d}\left(M(x^*)-M(x)\right)u_1^{l_1}\cdots u_d^{l_d}K(u)du \\
&\quad +\frac{1}{q!}\sum_{|\vec{l}|=q}\nabla^q_{l_1,\ldots l_d}M(x)\int u_1^{l_1}\cdots u_d^{l_d}K(u)du \\
&= O(b)
\end{aligned}
$$

where the final equality for $\tilde{h}_q(x,u)$ follows by Assumption D (v) and (vi) and the fact that $\int u_1^{l_1}\cdots u_d^{l_d}K(u)du = 0$ under Assumption D (iii). The same reasoning implies $h_\rho(x,u) = 0$. Under the null hypothesis $m(x) = r(x_{(1)})$, hence $\sqrt{n}\theta_n = o(1)$ by Assumption D (vi), 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).

A.1.2. *Proof of (ii).* For the supersmooth case we will make use of the following Lemma.

Lemma 2. *Under Assumptions D and S*

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

for $k = 1, 2, 3, \ldots$.

□

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

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

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

Proposition 3. *Under Assumptions D and S*

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

□

Hence, we conclude $\int Var\left[r_n(Z_i;\xi)\right]d\mu(\xi) < \infty$ 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

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

As in the proof of Theorem 1 (i)

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

since $m(x) = r(x_{(1)})$ under the null hypothesis and by Assumption D (vi) ($nb^{2(q+1)} \to 0$).

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).

A.2. **Proof of Theorem 2.**

A.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

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

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).

A.2.2. *Proof of (ii).* The proof is almost identical to Theorem 2 (i).

A.3. **Proof of Theorem 3.**

A.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 δ, $\varsigma > 0$, assume $\beta(j) = O(j^{\eta}) = O\left(j^{(\varsigma-2)(2+\delta)/\delta}\right)$, then the residual from the Hoeffding projection is bounded as

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

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

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

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

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

Hence

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

and

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

using Assumption T (ii) $(n^{1-\frac{\varsigma}{2}}b^{(d+d_1)\frac{5(2+\delta)}{8}} \to \infty)$.

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 need to 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 D, O and T*

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

□

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

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

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

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

By similar arguments

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

where

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

So

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

where the final equality follows by similar arguments as for the same proof in 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.

A.3.2. *Proof of (ii).* Very similar reasoning as above can be applied to the supersmooth case. We omit the proof for brevity.

A.4. **Proof of Theorem 4.** 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 $f_\epsilon^{\mathrm{ft}}(\cdot)$, denoted $\hat{f}_\epsilon^{\mathrm{ft}}(\cdot)$, we define

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

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

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

For T_1, we use the same reasoning to show

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

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

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

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

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

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

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

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

Notice that the multiple in T_{12} is the same as the U-statistic we dealt with in Theorem 1 but divided by $\frac{1}{f_\epsilon^{\mathrm{ft}}(s_{(1)}/b)}$. We can make use of slightly adapted versions of Lemma 1 and 2 to show that $T_{12}=o_p(1)O_p(n^{-1/2})$ in the ordinary smooth and supersmooth cases, respectively. For

example, in the ordinary smooth case

$$\begin{aligned}
\frac{1}{2\pi b}\int e^{-\mathrm{i}t\cdot a}\frac{\tilde{K}^{\mathrm{ft}}(t)}{f_\epsilon^{\mathrm{ft}}(t/b)^2}dt &\sim \frac{1}{2\pi b}\int e^{-\mathrm{i}t\cdot a}\left(\sum_{v=0}^{\alpha} C_v\left|\frac{t}{b}\right|^v\right)^2\tilde{K}^{\mathrm{ft}}(t)dt \\
&= \frac{1}{2\pi b}\int e^{-\mathrm{i}t\cdot a}\sum_{v=0}^{\alpha}\sum_{z=0}^{\alpha} C_vC_z\left|\frac{t}{b}\right|^{z+v}\tilde{K}^{\mathrm{ft}}(t)dt \\
&= \frac{1}{b}\sum_{v=0}^{\alpha}\sum_{z=0}^{\alpha}\frac{C_vC_z}{(-2\pi\mathrm{i}b)^{z+v}}\tilde{K}^{(z+v)}(a).
\end{aligned}$$

In the same manner as Lemma 1, combining this with the properties of the infinite order kernel, we can write

$$\int x_j^k\left(\frac{1}{2\pi b}\int e^{-\mathrm{i}t\cdot a}\frac{\tilde{K}^{\mathrm{ft}}(t)}{f_\epsilon^{\mathrm{ft}}(t/b)^2}dt\right)dx_j \sim \begin{cases} \frac{(k!)}{b(2\pi\mathrm{i}b)^k}\sum_v\sum_z{}_{z+v=k} C_vC_z & \text{for } k\leq 2\alpha, \\ 0 & \text{otherwise.}\end{cases}$$

We can then use analogous arguments as in the proof of Theorem 1, along with the slight strengthening of Assumption 3.1 (ii) ($q \geq \alpha$) to Assumption 3.4 (ii) ($q \geq 2\alpha$), to show

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

This result makes intuitive sense. The convergence rate of the U-statistic in Theorem 1 does not depend on $f_\epsilon^{\mathrm{ft}}(\cdot)$. We could simply refer to $f_\epsilon^{\mathrm{ft}}(\cdot)^2$ as being ordinary smooth with parameter 2α instead of α. The same reasoning can be applied to the supersmooth case.

We now 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.

A.5. **Proof of Propositions.**

A.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 1. 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 3.

A.5.2. *Proof of Proposition 2.* To study $Var[r_n(Z_i;\xi)]$, we write $r_n(Z_i;\xi)$ as follows

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

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

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

Then by a change of variables and multivariate Taylor expansion, we can write

$$\begin{aligned}
&E\left[K\left(x_{(1)}+\frac{W_{(1)i}-X_{(1)j}}{b}\right)\Big|W_{(1)i}\right] \\
&= b^{d_1}\int K(x_{(1)}+s_{(1)})f_{X_{(1)}}(W_{(1)i}-s_{(1)}b)ds_{(1)} \\
&= b^{d_1}\left(\sum_{\rho=0}^{q-1}\tilde{a}_\rho(x_{(1)},W_{(1)i})+\tilde{a}_q^*(x_{(1)},W_{(1)i})\right) \qquad \text{(A.2)}
\end{aligned}$$

where the first equality follows from a change of variables and the data being i.i.d., and where we use the definition

$$\tilde{a}_\rho(x_{(1)},W_{(1)i}) \equiv \frac{b^\rho}{\rho!}\sum_{|\vec{l}|=\rho}\int K(x_{(1)}+s_{(1)})s_1^{l_1}\cdots s_{d_1}^{l_{d_1}}ds_{(1)}\nabla^\rho_{l_1,\dots l_{d_1}}f_{X_{(1)}}(W_{(1)i})$$

and the remainder term is given by

$$\tilde{a}_q^*(x_{(1)},W_{(1)i}) \equiv \frac{b^q}{q!}\sum_{|\vec{l}|=q}\int K(x_{(1)}+s_{(1)})s_1^{l_1}\cdots s_{d_1}^{l_{d_1}}\nabla^q_{l_1,\dots l_{d_1}}f_{X_{(1)}}(W^*_{(1)if})ds_{(1)}$$

where $W^*_{(1)if}$ lies between $W_{(1)i}$ and $(W_{(1)i}-s_{(1)}b)$. Since $K(\cdot)$ is an infinite-order kernel, we can show

$$\int s^{\vec{l}}K(s+x)ds=(-x)^{\vec{l}}=(-x_1)^{l_1}\cdots(-x_d)^{l_d}, \qquad \text{(A.3)}$$

hence

$$\tilde{a}_\rho(x_{(1)}, W_{(1)i}) \quad = \quad \frac{b^\rho}{\rho!}\sum_{|\vec{l}|=\rho}(-x_1)^{l_1}\cdots(-x_{d_1})^{l_{d_1}}\nabla^\rho_{l_1,\ldots l_{d_1}} f_{X_{(1)}}(W_{(1)i}).$$

Combining this with (A.2), we can write

$$\begin{aligned} r_{1n}(Z_i;\xi) &= b^d\sum_{\rho=0}^{q-1}\int \mathcal{K}_b\left(x\right)a_\rho(x_{(1)}, W_{(1)i})\mathcal{W}(W_i+ab;\xi)dx \\ &\quad +b^d\int \mathcal{K}_b\big(x\big)a^*_q(x_{(1)}, W_{(1)i})\mathcal{W}(W_i+ab;\xi)dx. \\ &\equiv r_{11n}(Z_i;\xi)+r_{12n}(Z_i;\xi), \end{aligned}$$

where

$$a_\rho(x_{(1)}, W_{(1)i}) \quad \equiv \quad \frac{b^\rho}{\rho!}\sum_{|\vec{l}|=\rho}x_1^{l_1}\cdots x_{d_1}^{l_{d_1}}\nabla^\rho_{l_1,\ldots l_{d_1}}\Big(Y_i f_{X_{(1)}}(W_{(1)i})-R(W_{(1)i})\Big),$$

and

$$\begin{aligned} a^*_q(x_{(1)}, W_{(1)i}) &\equiv \frac{b^q}{q!}\sum_{|\vec{l}|=q}\int K(x_{(1)}+s_{(1)})s_1^{l_1}\cdots s_{d_1}^{l_{d_1}} \\ &\quad \times\nabla^q_{l_1,\ldots l_{d_1}}\Big(Y_i f_{X_{(1)}}(W^*_{(1)if})-R(W^*_{(1)iR})\Big)\,ds_{(1)} \end{aligned}$$

with $W^*_{(1)iR}$ lying between $W_{(1)i}$ and $(W_{(1)i}-s_{(1)}b)$.

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

$$\begin{aligned}
Var[r_{11n}(Z_i;\xi)] &\leq E[r_{11n}(Z_i;\xi)^2] \\
&\leq b^{2d}||\mathcal{W}(\cdot;\xi)||_\infty^2 \sum_{\rho=0}^{(q-1)}\sum_{\eta=0}^{(q-1)} E\Big[\int\int \mathcal{K}_b\left(x\right)\mathcal{K}_b\left(z\right) \\
&\quad \times a_\rho(x,W_{(1)i})a_\eta(z,W_{(1)i})dxdz\Big] \\
&= O(b^{2d})\sum_{\rho=0}^{(q-1)}\sum_{\eta=0}^{(q-1)}\frac{b^\rho b^\eta}{\rho!\eta!}\left[\begin{array}{c} \left(\sum_{|\vec{l}|=\rho}\int \mathcal{K}_b(x)x_1^{l_1}\cdots x_{d_1}^{l_\rho}dx\right) \\ \times\left(\sum_{|\vec{l}|=\eta}\int \mathcal{K}_b(z)z_1^{l_1}\cdots z_{d_1}^{l_\eta}dz\right)\end{array}\right] \\
&\quad \times E\left[\begin{array}{c} \nabla^\rho_{l_1,\ldots l_{d_1}}\left(Y_i f_{X_{(1)}}(W_{(1)i})-R(W_{(1)i})\right) \\ \times\nabla^\eta_{l_1,\ldots l_{d_1}}\left(Y_i f_{X_{(1)}}(W_{(1)i})-R(W_{(1)i})\right)\end{array}\right] \\
&\sim O(1)
\end{aligned}$$

where the final equality follows from Assumptions D (ii), (iv) and (v) and using Lemma 1 which shows

$\left(\int \mathcal{K}_b(x)x_1^{l_1}\cdots x_{d_1}^{l_\rho}dx\right)\sim O\left(b^{-|\vec{l}|-d}\right)$. 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 have

$$\begin{aligned}
Var[r_{12n}(Z_i;\xi)] &\leq E\left[r_{12n}(Z_i;\xi)^2\right] \\
&\leq b^{2d}||\mathcal{W}(\cdot;\xi)||_\infty^2 E\left[\left(\int \mathcal{K}_b\left(x\right)a_q^*(x_{(1)},W_{(1)i})dx\right)^2\right].
\end{aligned}$$

Notice that

$$\begin{aligned}
\int \mathcal{K}_b\left(x\right) a_q^*(x_{(1)}, W_{(1)i})dx &= \frac{b^q}{q!}\sum_{|\vec{l}|=q}\int \mathcal{K}_b\left(x\right)\int K(x_{(1)}+s_{(1)})s_1^{l_1}\cdots s_{d_1}^{l_{d_1}} \\
&\quad \times \nabla^q_{l_1,\ldots l_{d_1}}\left(Y_i f_{X_{(1)}}(W^*_{(1)if}) - R(W^*_{(1)iR})\right) ds_{(1)}dx \\
&\sim \frac{b^q}{q!}\sum_{|\vec{l}|=P}\int \mathcal{K}_b(x)\int K(x_{(1)}+s_{(1)})s_1^{l_1}\cdots s_{d_1}^{l_{d_1}} \\
&\quad \times \nabla^q_{l_1,\ldots l_{d_1}}\left[\begin{array}{c} Y_i\left\{f_{X_{(1)}}(W^*_{(1)if}) - f_{X_{(1)}}(W_{(1)i})\right\} \\ -\left\{R(W^*_{(1)iR}) - R(W_{(1)i})\right\}\end{array}\right] ds_{(1)}dx, \\
&\quad + \frac{b^q}{q!}\sum_{|\vec{l}|=q}\int \mathcal{K}_b(x)\int K(x_{(1)}+s_{(1)})s_1^{l_1}\cdots s_{d_1}^{l_{d_1}} ds_{(1)}dx \\
&\quad \times \nabla^q_{l_1,\ldots l_{d_1}}\left[R(W_{(1)i}) - Y_i f_{X_{(1)}}(W_{(1)i})\right] \\
&\equiv r_{121n}(Z_i;\xi) + r_{122n}(Z_i;\xi)
\end{aligned}$$

where the final equality adds and subtracts the last term in that expression. Dealing with $r_{121n}(Z_i;\xi)$ first

$$\begin{aligned}
E[r_{121n}(Z_i;\xi)^2] &\leq b^{2d}||\mathcal{W}(\cdot;\xi)||^2_\infty \frac{b^{2q}}{(q!)^2}E\left[\left\{\sum_{|\vec{l}|=q}\int \mathcal{K}_b\left(x\right)\int K(x_{(1)}+s_{(1)})s_1^{l_1}\cdots s_{d_1}^{l_{d_1}}\right.\right. \\
&\quad \left.\left. \times \nabla^q_{l_1,\ldots l_{d_1}}\left[\begin{array}{c} Y_i\left\{f_{X_{(1)}}(W^*_{(1)if}) - f_{X_{(1)}}(W_{(1)i})\right\} \\ -\left\{R(W^*_{(1)iR}) - R(W_{(1)i})\right\}\end{array}\right] ds_{(1)}dx\right\}^2\right] \\
&\leq b^{2d}||\mathcal{W}(\cdot;\xi)||^2_\infty \frac{b^{2q}}{(q!)^2}b^2 E\left[g(W_{(1)i})^2(1+|Y_i|)^2\right] \\
&\quad \times \left(\sum_{|\vec{l}|=q}\sum_{j=1}^{d_1}\int\int \mathcal{K}_b\left(x\right)\int K(x_{(1)}+s_{(1)})s_1^{l_1}\cdots s_{d_1}^{l_{d_1}} s_j ds_{(1)}dx\right)^2 \\
&= O\left(b^{2d+2(q+1)}\right)\left(\sum_{|\vec{l}|=q}\sum_{j=1}^{d_1}\int \mathcal{K}_b(x)x_1^{l_1}\cdots x_{d_1}^{l_{d_1}}x_j dx\right)^2 \\
&= O(1)
\end{aligned}$$

where the second inequality follows from Assumptions N (ii) and (iii), the penultimate equality follows from (A.3), and the final equality makes use of Lemma 1 which shows $\sum_{|\vec{l}|=q} \int \mathcal{K}(x) x_1^{l_1} \cdots x_{d_1}^{l_{d_1}} x_j dx = O\left(b^{-(q+1)-d}\right)$.

For $r_{122n}(Z_i;\xi)$ we have

$$
\begin{aligned}
E\left[r_{122n}(Z_i;\xi)^2\right] &= O\left(b^{2d}\right) \frac{b^{2q}}{(q!)^2} \left(\sum_{|\vec{l}|=q} \int \mathcal{K}(x) x_1^{l_1} \cdots x_{d_1}^{l_\rho} dx \right)^2 \\
&\quad \times E\left[\left\{ \nabla^q_{l_1,\ldots l_{d_1}} \left(Y_i f_{X_{(1)}}(W_{(1)i}) - R(W_{(1)i}) \right) \right\}^2 \right] \\
&= O(1)
\end{aligned}
$$

where the first equality follows from (A.3) and the second from Lemma 1 and Assumptions D (iv) and (v).

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

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

and define

$$
a_\rho(x, W_i) \equiv \frac{b^\rho}{\rho!} \sum_{|\vec{l}|=\rho} x_1^{l_1} \cdots x_d^{l_d} \nabla^\rho_{l_1,\ldots l_d} \left(Y_i f_X(W_i) - M(W_i) \right),
$$

and

$$
a_q^*(x, W_i) \equiv \frac{b^q}{q!} \sum_{|\vec{l}|=q} \int K(x+s) s_1^{l_1} \cdots s_d^{l_d} \nabla^q_{l_1,\ldots l_d} \left(Y_i f_X(W_{if}^*) - M(W_{iM}^*) \right) ds,
$$

where W_{iM}^* lies between W_i and $(W_i - sb)$ and W_{if}^* also lies between W_i and $(W_i - sb)$. 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

$$r_{21n}(Z_i;\xi) = b^{d_1} \int \mathcal{K}_b\left(x_{(1)}\right) a_\rho(x, W_i) \mathcal{W}(W_i + ab;\xi) dx$$

and

$$r_{21n}(Z_i;\xi) = b^{d_1} \int \mathcal{K}_b\big(x_{(1)}\big) a_q^*(x, W_i) \mathcal{W}(W_i + ab;\xi) dx$$

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

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

Combining these results we have

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

This concludes the proof Proposition 1.

A.5.3. *Proof of Proposition 3.* To study $Var[r_n(Z_i;\xi)]$, as in the proof of Proposition 1, we write $r_n(Z_i;\xi)$ as follows

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

Again, we can write

$$
\begin{aligned}
r_{1n}(Z_i;\xi) &= b^d\sum_{\rho=0}^{q-1}\int \mathcal{K}_b\left(x\right)a_\rho(x,W_{(1)i})\mathcal{W}(W_i+ab;\xi)dx \\
&\quad + b^d\int \mathcal{K}_b\big(x\big)a_q^*(x,W_{(1)i})\mathcal{W}(W_i+ab;\xi)dx. \\
&\equiv r_{11n}(Z_i;\xi) + r_{12n}(Z_i;\xi),
\end{aligned}
$$

where we have

$$
a_\rho(x,W_{(1)i}) \equiv \frac{b^\rho}{\rho!}\sum_{|\vec{l}|=\rho} x_1^{l_1}\cdots x_{d_1}^{l_{d_1}}\nabla^\rho_{l_1,\ldots l_{d_1}}\Big(Y_i f_{X_{(1)}}(W_{(1)i}) - R(W_{(1)i})\Big),
$$

and

$$
\begin{aligned}
a_q^*(x,W_{(1)i}) &\equiv \frac{b^q}{q!}\sum_{|\vec{l}|=q}\int K(x_{(1)}+s_{(1)})s_1^{l_1}\cdots s_{d_1}^{l_{d_1}} \\
&\quad \times\nabla^q_{l_1,\ldots l_{d_1}}\Big(Y_i f_{X_{(1)}}(W^*_{(1)if}) - R(W^*_{(1)iR})\Big)\,ds_{(1)}
\end{aligned}
$$

with $W^*_{(1)iR}$ lying between $W_{(1)i}$ and $(W_{(1)i}-s_{(1)}b)$.

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

$$\begin{aligned} Var[r_{11n}(Z_i;\xi)] &\leq E[r_{11n}(Z_i;\xi)^2] \\ &\leq ||\mathcal{W}(\cdot;\xi)||_\infty^2 \sum_{\rho=0}^{(q-1)} \sum_{\eta=0}^{(q-1)} \frac{1}{\rho!\eta!(-2\pi\mathrm{i})^{\rho+\eta}} \\ &\quad \times E\left[\begin{array}{c} \nabla^\rho_{l_1,\ldots l_{d_1}} \Big(Y_i f_{X_{(1)}}(W_{(1)i}) - R(W_{(1)i}) \Big) \\ \times \nabla^\eta_{l_1,\ldots l_{d_1}} \Big(Y_i f_{X_{(1)}}(W_{(1)i}) - R(W_{(1)i}) \Big) \end{array} \right] \\ &= O(1) \end{aligned}$$

where the first inequality follows from Lemma 2 and the final equality follows from Assumptions D (ii), (iv) and (v). Hence

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

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.

A.5.4. *Proof of Proposition 4.* The proof is very similar to that of Proposition 2, as such we only outline the parts of the proof that differ as a result of the dependence in the data. We write $r_{1n}(Z_i;\xi)$ as follows

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

where the second equality follows from the fact that the data is i.i.d.. Then by a change of variables and multivariate Taylor expansion, we can write

$$\begin{aligned}
& E\left[K\left(x_{(1)}+\frac{W_{(1)i}-X_{(1)j}}{b}\right)\middle|W_{(1)i}\right] \\
= \;& b^{d_1}\int K(x_{(1)}+s_{(1)})f_{X_{(1)j}|W_{(1)i}}\left(W_{(1)i}-s_{(1)}b|W_{(1)i}\right)ds_{(1)} \\
= \;& b^{d_1}\left(\sum_{\rho=0}^{q-1}\tilde{a}_\rho(x_{(1)},W_{(1)i})+\tilde{a}_q^*(x_{(1)},W_{(1)i})\right) \qquad \text{(A.4)}
\end{aligned}$$

where the first equality follows from a change of variables and where we use the definition

$$\tilde{a}_\rho(x_{(1)},W_{(1)i}) \;\equiv\; \frac{b^\rho}{\rho!}\sum_{|\vec{l}|=\rho}\int K(x_{(1)}+s_{(1)})s_1^{l_1}\cdots s_{d_1}^{l_{d_1}}ds_{(1)}\nabla^\rho_{l_1,\ldots l_{d_1}}f_{X_{(1)j}|W_{(1)i}}(W_{(1)i}|W_{(1)i})$$

and the remainder term is given by

$$\tilde{a}_q^*(x_{(1)},W_{(1)i}) \;\equiv\; \frac{b^q}{q!}\sum_{|\vec{l}|=q}\int K(x_{(1)}+s_{(1)})s_1^{l_1}\cdots s_{d_1}^{l_{d_1}}\nabla^q_{l_1,\ldots l_{d_1}}f_{X_{(1)j}|W_{(1)i}}(W^*_{(1)if}|W_{(1)i})ds_{(1)}$$

where $W^*_{(1)if}$ lies between $W_{(1)i}$ and $(W_{(1)i}-s_{(1)}b)$. As in the proof of Proposition 2, we can write

$$\begin{aligned}
r_{1n}(Z_i;\xi) \;&=\; b^d\sum_{\rho=0}^{q-1}\int \mathcal{K}_b\left(x\right)a_\rho(x_{(1)},W_{(1)i})\mathcal{W}(W_i+ab;\xi)dx \\
&\quad +b^d\int \mathcal{K}_b\big(x\big)a_q^*(x_{(1)},W_{(1)i})\mathcal{W}(W_i+ab;\xi)dx. \\
&\equiv\; r_{11n}(Z_i;\xi)+r_{12n}(Z_i;\xi),
\end{aligned}$$

where

$$a_\rho(x_{(1)},W_{(1)i}) \;\equiv\; \frac{b^\rho}{\rho!}\sum_{|\vec{l}|=\rho}x_1^{l_1}\cdots x_{d_1}^{l_{d_1}}\nabla^\rho_{l_1,\ldots l_{d_1}}\Big(Y_if_{X_{(1)j}|W_{(1)i}}(W_{(1)i}|W_{(1)i})-R_{X_{(1)j}|W_{(1)i}}(W_{(1)i}\big|W_{(1)i})\Big),$$

and

$$
\begin{aligned}
a^*_P(x_{(1)}, W_{(1)i}) \quad &\equiv \quad \frac{b^q}{q!}\sum_{|\vec{l}|=q}\int K(x_{(1)}+s_{(1)})s_1^{l_1}\cdots s_{d_1}^{l_{d_1}} \\
&\quad \times\nabla^q_{l_1,\ldots l_{d_1}}\left(Y_i f_{X_{(1)j}|W_{(1)i}}(W^*_{(1)if}|W_{(1)i}) - R_{X_{(1)j}|W_{(1)i}}(W^*_{(1)iR}\big|W_{(1)i})\right) ds_{(1)}
\end{aligned}
$$

with $W^*_{(1)iR}$ lying between $W_{(1)i}$ and $(W_{(1)i} - s_{(1)}b)$.

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

$$
\begin{aligned}
E\left[|r_{11n}(Z_i;\xi)|^{2+\tilde{\delta}}\right] \quad &\leq \quad ||\mathcal{W}(\cdot;\xi)||_\infty^{2+\tilde{\delta}}\, b^{d\left(2+\tilde{\delta}\right)}2^{\left(1+\tilde{\delta}\right)}\sum_{\rho=0}^{(q-1)}\left|\frac{b^\rho}{\rho!}\int \mathcal{K}_b\left(x\right)\sum_{|\vec{l}|=\rho} x_1^{l_1}\cdots x_{d_1}^{l_{d_1}}dx\right|^{2+\tilde{\delta}} \\
&\quad \times E\left|\nabla^\rho_{l_1,\ldots l_{d_1}}\left(Y_i f_{X_{(1)j}|W_{(1)i}}(W_{(1)i}|W_{(1)i}) - R_{X_{(1)j}|W_{(1)i}}(W_{(1)i}\big|W_{(1)i})\right)\right|^{2+\tilde{\delta}} \\
&= \quad O(1)
\end{aligned}
$$

where the final equality follows from Assumptions T (iii) and (iv) and using Lemma 1 which shows $\left(\int \mathcal{K}_b(x)x_1^{l_1}\cdots x_{d_1}^{l_\rho}dx\right) \sim O\left(b^{-|\vec{l}|-d}\right)$.

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.

A.6. **Proof of Lemmas.**

A.6.1. *Proof of Lemma 1.* 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 an infinite-order kernel we have

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

we can write

$$
\int x_j^k \tilde{\mathcal{K}}_b(x_j)\,dx_j \sim \begin{cases} \frac{C_k(k!)}{b(2\pi\mathrm{i}b)^k} & \text{for } k \le \alpha, \\ 0 & \text{otherwise.} \end{cases}
$$

A.6.2. *Proof of Lemma 2.* We can write the deconvolution kernel as follows

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

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 $\gamma_0 + \gamma$ to be an integer for the derivatives of the kernel function to exist. This is satisfied by the majority of common distribution functions. Given this, and the fact that for an infinite-order kernel we have

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

we can write

$$
\int x_j^k \tilde{\mathcal{K}}_b(x_j)\,dx_j \sim \frac{C\mu^{(k-\gamma_0)}}{b(-2\pi\mathrm{i}b)^k}.
$$

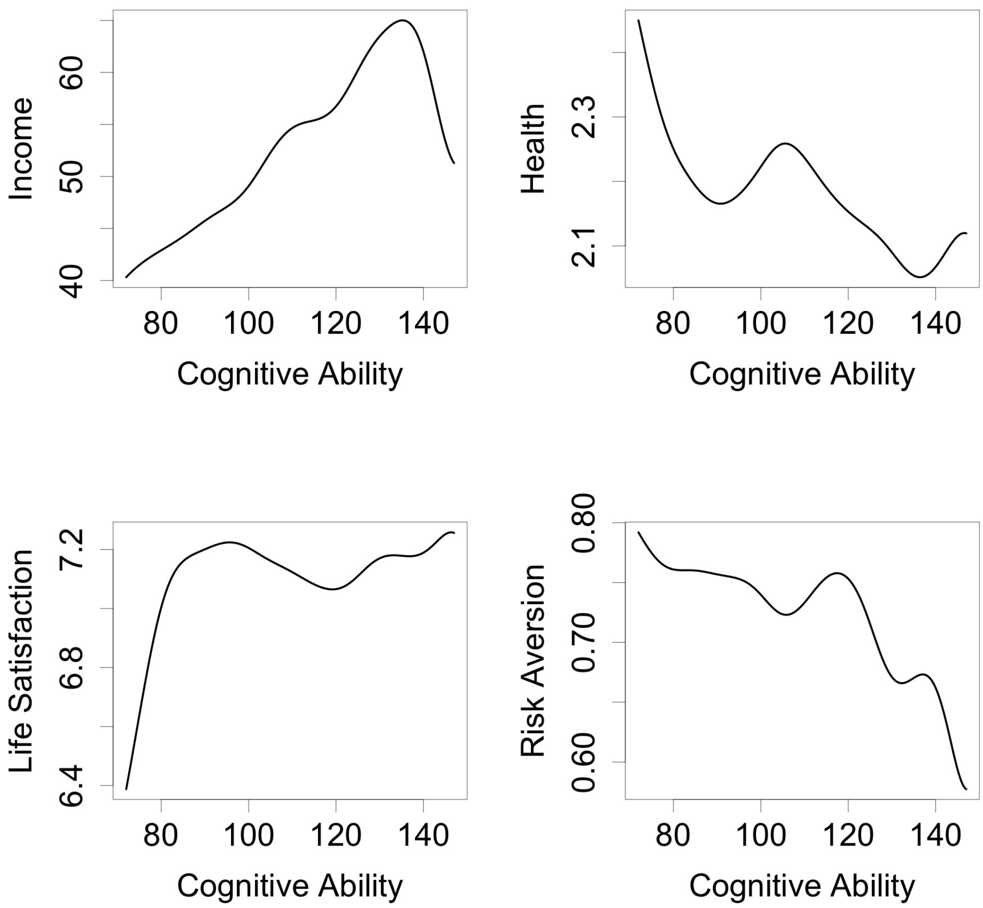

Figure B.1. Nonparametric plots of the relationship between Cognitive Ability and Income, Health, Life Satisfaction and Risk Aversion. All control variables are set at their respective means and $b = 0.65$.

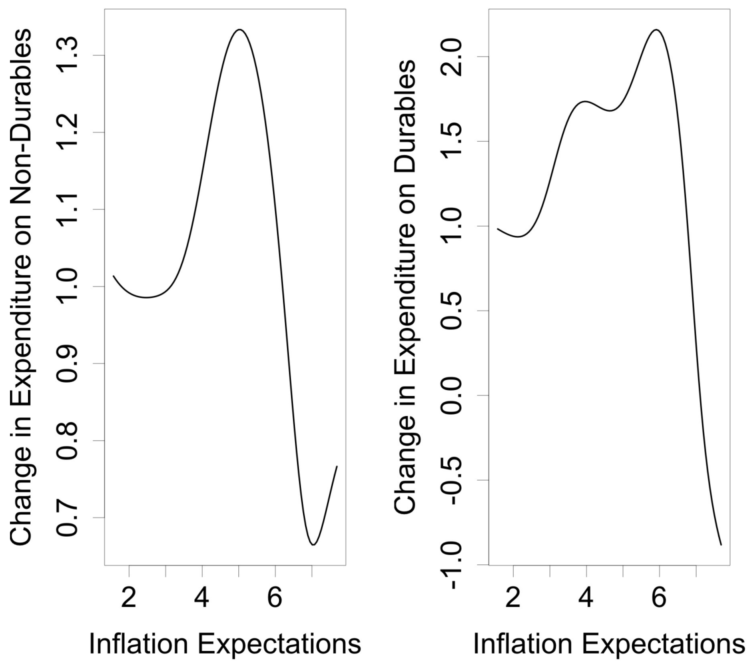

FIGURE B.2. Nonparametric plot of the relationship between Inflation Expectations and Change in Expenditure on Durables and Non-Durables. All control variables are set at their respective means and $b = 0.85$.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[17] Eggertsson, G.B. (2008) Great Expectations and the End of the Depression. *The American Economic Review*, 98(4), pp.1476-1516.

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

[19] Fan, J. (1992) Deconvolution with supersmooth distributions, *Canadian Journal of Statistics*, 20, 155-169.

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

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

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

[23] Gensowski, M., Heckman, J. and Savelyev, P. (2011) The effects of education, personality, and IQ on earnings of high-ability men. *Unpublished manuscript.*

[24] Hall, P. and Ma, Y. (2007) Testing the suitability of polynomial models in errors-in-variables problems. *Annals of Statistics*, 35(6), pp.2620-2638.

[25] Hall, R.E. (2011) The long slump. *The American Economic Review*, 101(2), pp.431-469.

[26] Härdle, W. and Mammen, E. (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics*, pp.1926-1947.

[27] Heckman, J.J., Stixrud, J. and Urzua, S. (2006) The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior (No. w12006). *National Bureau of Economic Research.*

[28] Hernstein, R.J. and Murray, C. (1994) The bell curve. Free Press. New York.

[29] Juster, F.T., Wachtel, P., Hymans, S. and Duesenberry, J. (1972) Inflation and the Consumer. *Brookings Papers on Economic Activity*, 1972(1), pp.71-121.

[30] Krugman, P. (2012) Not enough inflation. *New York Times.*

[31] Krugman, P.R., Dominquez, K.M. and Rogoff, K. (1998) It's baaack: Japan's slump and the return of the liquidity trap. *Brookings Papers on Economic Activity*, 1998(2), pp.137-205.

[32] Lavergne, P., Maistre, S. and Patilea, V. (2015) A significance test for covariates in nonparametric regression. *Electronic Journal of Statistics*, 9, pp.643-678.

[33] Lavergne, P. and Vuong, Q.H. (1996) Nonparametric selection of regressors: The nonnested case. *Econometrica*, pp.207-219.

[34] Lavergne, P. (2001) An equality test across nonparametric regressions. *Journal of Econometrics*, 103(1), pp.307-344.

[35] Li, Q. (1999) Consistent model specification tests for time series econometric models. *Journal of Econometrics*, 92(1), pp.101-147.

[36] Li, T. and Vuong, Q. (1998) Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis*, 65(2), pp.139-165.

[37] 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.

[38] Mavroeidis, S., Plagborg-Møller, M. and Stock, J.H. (2014) Empirical evidence on inflation expectations in the New Keynesian Phillips Curve. *Journal of Economic Literature*, 52(1), pp.124-188.

[39] Meister, A. (2009) *Deconvolution Problems in Nonparametric Statistics*, Springer.

[40] Neumeyer, N. and Dette, H. (2003) Nonparametric comparison of regression curves: an empirical process approach. *Annals of Statistics*, 31(3), pp.880-920.

[41] Otsu, T. and Taylor, L. (2016) Specification testing for errors-in-variables models. *Unpublished Manuscript.*

[42] Politis, D.N. and Romano, J.P. (1994) Limit theorems for weakly dependent Hilbert space valued random variables with applications to the stationary bootstrap. *Statistica Sinica*, 4(2), pp.461-476.

[43] Powell, J.L., Stock, J.H. and Stoker, T.M. (1989) Semiparametric estimation of index coefficients. *Econometrica*, pp.1403-1430.

[44] Racine, J. (1997) Consistent significance testing for nonparametric regression. *Journal of Business & Economic Statistics*, 15(3), pp.369-378.

[45] Robinson, P.M. (1989) Hypothesis testing in semiparametric and nonparametric models for econometric time series. *Review of Economic Studies*, 56(4), pp.511-534.

[46] Romer, C.D. (2011) Dear Ben: it's time for your Volcker moment. *New York Times.*

[47] Romer, C.D. and Romer, D.H. (2013) The missing transmission mechanism in the monetary explanation of the Great Depression. *The American Economic Review*, 103(3), pp.66-72.

[48] Schennach, S.M. (2013) Measurement error in nonlinear models–a review. *Advances in Economics and Econometrics*, 3, pp.296-337.

[49] Shafir, E., Diamond, P. and Tversky, A. (1997) Money illusion. *The Quarterly Journal of Economics*, pp.341-374.

[50] Stinchcombe, M.B. and White, H. (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric theory*, 14(03), pp.295-325.

[51] Stute, W. (1997) Nonparametric model checks for regression. *Annals of Statistics*, pp.613-641.

[52] Van Es, A. J. and Uh H.-W. (2005) Asymptotic normality for kernel type deconvolution estimators, *Scandinavian Journal of Statistics*, 32, 467-483.

[53] Veenhoven, R. and Choi, Y. (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.

[54] Volcker, P.A. (2011) A Little Inflation Can Be A Dangerous Thing. *New York Times.*

[55] 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.

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

E-mail address: l.n.taylor@lse.ac.uk