# KIER DISCUSSION PAPER SERIES

# KYOTO INSTITUTE
# OF
# ECONOMIC RESEARCH

KYOTO UNIVERSITY

KYOTO, JAPAN

# Optimal Minimax Rates against Non-smooth Alternatives

Kohtaro Hitomi[*]      Masamune Iwasawa[†]      Yoshihiko Nishiyama[‡]

December 30, 2020

**Abstract**

This study investigates optimal minimax rates for specification testing when the alternative hypothesis is built on a set of non-smooth functions. The set consists of bounded functions that are not necessarily differentiable with no smoothness constraints imposed on their derivatives. In the instrumental variable regression set up with an unknown error variance structure, we find that the optimal minimax rate is $n^{-1/4}$, where $n$ is the sample size. The rate is achieved by a simple test based on the difference between non-parametric and parametric variance estimators.

**Keywords:** optimal minimax rate; specification test; instrumental variable regression model; nearest neighbor method

**JEL Classification:** C12; C14

## 1    Introduction

This study investigates uniform power of specification testing for a regression function using the minimax approach. In the minimax approach, the alternative hypothesis is a

---

[*]Kyoto Institute of Technology, hitomi@kit.ac.jp

[†]Corresponding author: Masamune Iwasawa, Department of Economics, Otaru University of Commerce, 3-5-21 Midori, Otaru, Hokkaido 047-8501, Japan Tel: +81(0)134275324, masamune-iwasawa@res.otaru-uc.ac.jp

[‡]Institute of Economic Research, Kyoto University, nishiyama@kier.kyoto-u.ac.jp

set of functions with certain smoothness. This approach reveals how the smoothness and model dimensionality are related to the achievable uniform power (optimal minimax rate) of tests. The optimal minimax rate against smooth alternatives is investigated by Guerre and Lavergne (2002) and subsequent works, such as Horowitz and Spokoiny (2001) and H. Li, Li, and Liu (2016). While these tests achieve the fastest possible uniform power against smooth alternatives, they may not perform well when the alternative includes non-smooth functions.The optimal minimax rate against such non-smooth alternatives is an open question.

In economics, there are several important applications in which a smooth model is applied to a possibly non-smooth function. For example, in empirical estimations of aggregate demand curves (see, e.g., Ball, Mankiw, & Romer, 1988), price level may be modeled as a smooth function of output (real GDP), although the true curve might be non-differentiable in the presence of a liquidity trap or if investment is not elastic to changes in the interest rate. Another example is kinked demand curves in an oligopolistic market with price rigidity (Sweezy, 1939). In the estimation of demand curves, how-ever, the price of a good may be modeled as a smooth function of its quantity. Kinks in demand curves are also caused by consumer behavior, such as asymmetric consumer reactions to price increases and decreases (see Dossche, Heylen, & Van den Poel, 2010 for empirical evidence). Another important example is Engel curve estimation. If pref-erences are structured in a hierarchical manner, the resulting Engel curve will exhibit a kink at points that reflect changes in the capacity of a good to satisfy needs as income increases (Drakopoulos, 1994). For example, basic food expenditure increases as income increases, but after the basic need is satisfied, a higher proportion of income may be spent on less-necessary goods. In this case, the commonly used Working–Leser Engel curve specification may not fit the data well.

Alternative hypotheses in this study include non-differentiable functions and func-tions such that no smoothness constraints are imposed on their derivatives. We find that

the optimal minimax rate against such non-smooth alternatives is $n^{-1/4}$ in the framework of instrumental variable (IV) regression with an unknown error variance structure, where $n$ is the sample size. The rate is achieved by a simple test based on the difference between non-parametric and parametric variance estimators. Simulation studies illustrate that the test has reasonable power against various non-smooth alternatives. The power performance is compared with that of a kernel smoothing test that is rate optimal against smooth alternatives (Zheng, 1996, Hitomi, Iwasawa, & Nishiyama, 2020). The empirical application to Engel curves for food emphasizes the good applicability of the test. The alternative hypothesis in this study is called a non-smooth alternative because it includes less-smooth functions compared with the existing alternative.

The literature on the minimax approach is focused on testing signals in a Gaussian white noise model.[1] Ermakov (1991), Ingster (1993), and Lepski and Tsybakov (2000) show the optimal minimax rates against alternatives within a Hölder class, while Spokoiny (1996) and Lepski and Spokoiny (1999) do so against alternatives within a Besov class. Ingster and Sapatinas (2009) extend this approach to testing a multivariate non-parametric regression model with Gaussian noise against alternatives within an ellipsoid in the Hilbert space with respect to the tensor product of the Fourier basis. Other studies, such as Abramovich, Feis, Italia, and Theofanis (2009), show the optimal minimax rate of testing the additivity assumption of a response function against alternatives within a Besov class.

The optimal minimax rates of specification testing for a non-linear regression model against alternatives within a Hölder class is provided by Guerre and Lavergne (2002). Let us denote the dimension of the regressor by $l$. The alternative consists of $k$-times differentiable functions with its $k$th derivative being Hölder continuous with index $s$. Then, the optimal minimax rate against the smooth alternative ($s + k > l/4$) is $n^{-2(s+k)/[l+4(s+k)]}$. In addition to the test of Guerre and Lavergne (2002), this rate is achieved by the

---

[1] For a recent review of specification tests for regression models, see González-Manteiga and Crujeiras (2013).

3

$K$-nearest neighbor tests proposed by H. Li et al. (2016).

When the alternative is not smooth relative to the model dimensionality ($s+k \leq l/4$), no tests have non-trivial uniform power against the alternative that approaches the null faster than $n^{-1/4}$ (lower bound). Guerre and Lavergne (2002) point out the difficulty of dealing with such irregular non-smooth alternatives. A test that achieves the lower bound may exist if the structure of the error variance conditional on the regressor is known. Without this additional structure, however, it is not known whether any test exists that has non-trivial uniform power against the non-smooth alternative. This study contributes to the literature by showing the set of non-smooth functions against which the optimal minimax rate is $n^{-1/4}$. The non-smooth alternative in this study differs from those considered previously.

The rest of this paper is organized as follows. Section 2 describes the model and testing framework. Assumptions are given in Section 3. The main results for the optimal minimax rate are summarized in Section 4. Section 5 presents Monte Carlo experiments. Section 6 describes the application of the proposed test to Engel curve specifications. Section 7 concludes the text. The proofs of the primary results are provided in the Appendix.

## 2 Framework

Let $\{Y_i, X_i, Z_i\}_{i=1}^n$ be a sample from a random variable $(Y, X, Z) \in \mathbb{R} \times \mathbb{R}^{l_x} \times \mathbb{R}^l$. Consider a parametric model

$$Y_i = g(X_i, \theta) + u_i, \tag{1}$$

where $g(X_i, \theta)$ is a known function up to parameters $\theta \in \Theta$ and $\Theta$ is a compact subset of $\mathbb{R}^{l_\theta}$. A vector of regressors $X_i$ may include endogenous variables that are correlated with $u_i$, where $u_i$ is an error term.

Let $m(Z_i) = E(Y_i|Z_i)$ and $\omega_i = Y_i - m(Z_i)$, where $E(\omega_i|Z_i) = 0$ by definition. We

set the following assumptions on observations at hand.

**Assumption 1.** $\{Y_i, X_i, Z_i\}_{i=1}^n$ *are a random sample on* $(Y, X, Z) \in \mathbb{R} \times \mathbb{R}^{l_x} \times \mathbb{R}^l$*, where* $l_x$ *and* $l$ *are finite. A positive constant* $M < \infty$ *exists such that* $E\left(|\omega_i|^4 | Z_i\right) < M$ *with probability 1 (w.p.1).*

**Assumption 2.** *The density of* $Z$*, denoted as* $f(\cdot) : \mathbb{R}^l \to \mathbb{R}$*, has compact support (without loss of generality* $[0,1]^l$*) and at least one element of* $Z$ *is continuous on* $[0,1]$*. For any* $z \in [0,1]^l$*, bounds exist such that* $0 < \underline{f} \leq f(z) \leq \overline{f} < \infty$ *and* $\|\frac{\partial}{\partial z} f(z)\| < \overline{f}'$*.*

As in Assumption 2, we focus on the case in which at least one element of $Z$ is continuous. The primary reason to assume continuity comes from the minimax approach, in which alternatives consist of functions with some smoothness. When a model is discrete, its support is concentrated on some points, implying that it makes less sense to restrict the shape of the model outside its support. Thus, investigating power properties using the minimax approach is inappropriate for models with only discrete variables. The compactness of the instruments in Assumption 2 is not restrictive at all because it can be achieved by an appropriate monotone transformation.

The fourth moment restriction of the error term in Assumption 1 is required to guarantee the consistency of estimators for the asymptotic variance of the non-standardized test statistic. This is a standard assumption, which corresponds to, for example, the finite fourth moments condition for the estimation of asymptotic variance of generalized method of moments (GMM) estimators.

Let $\delta_\theta(\cdot) \equiv m(\cdot) - E[g(X_i, \theta)|\cdot]$. The null hypotheses is

$$H_0 : \theta_0 \in \Theta \text{ exists such that } \delta_{\theta_0}(Z_i) = 0 \text{ w.p.1.}$$

Note that the test considered in this study is the specification of $E[g(X_i, \theta)|\cdot]$ rather than that of $g(\cdot, \theta)$.

The alternative hypothesis is defined as a set of functions belonging to a smoothness class. Let $\mathcal{M}_\epsilon$ be a class of bounded functions $f(\cdot) : \mathbb{R}^l \to \mathbb{R}$ such that:

$$\mathcal{M}_\epsilon = \left\{ f(\cdot) : \text{ for } \epsilon \text{ such that } \|x - y\| < \epsilon, |f(x) - f(y)| = o(\epsilon^{\frac{l}{4} - 1} \|x - y\|) \text{ as } \epsilon \to 0 \right\},$$

where we suppress the subscript and denote $\mathcal{M}_\epsilon = \mathcal{M}$ if no confusions occur. Then, the alternative is

$$H_{n,1} : \mathcal{M}(\rho_n) = \left\{ m(\cdot) \in \mathcal{M}_\epsilon : \inf_{\theta \in \Theta} E\left[ \delta_\theta(Z_i)^2 \right] \geq \rho_n^2 \right\}.$$

Functions in $\mathcal{M}(\rho_n)$ are separated from the parametric model by $L^2$-distance but the distance can approach 0 at a rate $\rho_n^2$. The alternative hypothesis $H_{n,1}$ enables us to investigate the uniform power of testing, which is called the minimax approach (Ingster, 1993). The minimax approach finds the fastest rate at which $\rho_n$ approaches 0 while assuring the test uniformly detects alternatives. A test is called rate optimal when it has prescribed minimax power uniformly against a set of alternatives that approaches the null hypothesis at a rate faster than any other tests can detect. This rate is then called the optimal minimax rate.[2]

The class $\mathcal{M}$ restricts the smoothness of functions depending on the dimension $l$ of instruments. To see this, let us consider a function $f \in \mathcal{M} : \mathbb{R}^l \to \mathbb{R}$. This function satisfies that $|f(x) - f(y)|/\|x - y\| = o(\epsilon^{\frac{l}{4} - 1})$. Thus, $f$ is differentiable when $l \geq 4$ and it it not necessarily differentiable when $l < 4$. Indeed, without changing the results below, the smoothness condition in the class $\mathcal{M}$ can be replaced with $|f(x) - f(y)| = o(\|x - y\|^{l/4})$ as $x \to y$ for the case of $l \geq 4$.

To the best of our knowledge, this is the first study that investigates the optimal minimax rate against such non-smooth functions. Let $[s]$ be the greatest integer less

---

[2]A formal definition of the optimal minimax rate is given in Definition 1 of Guerre and Lavergne (2002).

than or equal to $s$. The optimal minimax rate for the non-linear regression model is close to $n^{-1/2}$ when alternative consists of functions that are at least $[l/4]$ times differentiable and $[l/4]$th derivative of these functions are Lipschitz of order $l/4 - [l/4]$, where $l$, now, indicates the dimension of the regression model (see, Guerre & Lavergne, 2002). In literature, however, the optimal minimax rate is not investigated when alternative includes any functions whose $[l/4]$th derivatives are not Lipschitz of order $l/4 - [l/4]$. Furthermore, although the lower bound against set of functions that are not $[l/4]$ times differentiable is shown to be $n^{-1/4}$, it is not known whether any test exists that has non-trivial uniform power against such non-smooth alternative. Not all but some of non-smooth functions that are not considered in literature are covered in this study, such as non-differentiable functions with $l < 4$ and functions whose first derivatives may not Lipschitz. To emphasize the difference, $H_{n,1}$ in this study is called non-smooth alternative.

## 3 Assumptions

Assumptions are imposed on the parametric model $g(x, \theta)$. The following Assumptions are standard in literature (Guerre & Lavergne, 2002, Hitomi et al., 2020).

**Assumption 3.** *For all $x$, $g(x, \theta)$ is twice continuously differentiable with respect to $\theta \in \Theta$, where $\Theta$ is a compact subset of $\mathbb{R}^{l_\theta}$.*

**Assumption 4.** $E[\sup_{\theta \in \Theta} \|\frac{\partial}{\partial \theta} g(X_i, \theta)\|^2] < \infty.$

**Assumption 5.** $E[\sup_{\theta \in \Theta} \|\frac{\partial}{\partial \theta \partial \theta'} g(X_i, \theta)\|^2] < \infty.$

**Assumption 6.** *For each $\theta \in \Theta$, $E[\|\frac{\partial}{\partial \theta} g(X_i, \theta)\|^2 | Z_i] < \infty$ w.p.1.*

**Assumption 7.** *For each $\theta \in \Theta$, $E[g(X_i, \theta)^2 | Z_i] < \infty$ w.p.1.*

**Assumption 8.** *For each $\theta \in \Theta$, $E[g(X_i, \theta) | Z_i = z] \in \mathcal{M}_\epsilon$ for some constant $L_g \leq L$.*

**Assumption 9.** *For each $\theta \in \Theta$, $G_\theta \equiv E[\frac{\partial}{\partial\theta}g(X_i, \theta)|Z_i = z]$ is continuous with respect to $z$ on the support of $Z$ and $E(G_\theta G_\theta')$ is non-singular.*

Next, assumptions on parameter estimators are imposed.

**Assumption 10.** *Under the null hypothesis, we have an $\sqrt{n}$-consistent estimator $\hat{\theta}_n \equiv \hat{\theta}$ of $\theta_0$.*

**Assumption 11.** *For each $m(\cdot) \in \mathcal{M}$, a unique pseudo-true value $\theta_m^* \equiv \theta^*$ with respect to the estimator $\hat{\theta}$ exists such that*

*(i) $\sqrt{n}(\hat{\theta} - \theta^*) = O_p(1)$ uniformly with respect to $m(\cdot) \in \mathcal{M}$.*

*For all $m(\cdot) \in \mathcal{M}$, a positive constant $c$ exists such that*

*(ii) $\|\theta_m^* - \theta_0\| \leq c|E\{m(Z_i) - E[g(X_i, \theta_0)|Z_i]\}|$.*

Assumptions 10 and 11 restrict the behavior of estimator $\hat{\theta}$ under the null and alternative hypotheses. Uniform consistency in Assumption 11 (i) is essential for developing the minimax approach that considers uniform power of testing. When all regressors are exogenous (regression model), non-linear least squares estimator may satisfy these assumptions (see, Guerre & Lavergne, 2002 for a sufficient condition). When $X$ includes endogenous variables (IV regression model), identification assumptions as well as other sufficient conditions for GMM estimator and estimator using continuum of unconditional moment restrictions (Carrasco & Florens, 2000; Dominguez & Lobato, 2004) to satisfy Assumptions 10 and 11 are considered in Hitomi et al., 2020.

Let $u_i^* \equiv Y_i - g(X_i, \theta^*)$. Assumption 7 along with the boundedness of the error term in Assumption 1 guarantees $E(u_i^2|Z_i) < \infty$ as well as $E(u_i^{*2}|Z_i) < \infty$.

## 4    Optimal Minimax Rate

The proposition below shows a lower bound $\tilde{\rho}_n$ against which no specification tests demonstrate non-trivial uniform power.

**Proposition 1** (*Lower Bound*). *Suppose Assumptions 1, 2, 3, 4, 8, and 9 hold. Let $\tilde{\rho}_n = n^{-1/4}$, each $\omega_i$ is $N(0,1)$ conditional on $Z_i$, and $Z_i$ is uniformly distributed. For any test $t_n$ with $\sup_{m(\cdot) \in H_0} P(t_n > z_\alpha) \leq \alpha + o(1)$,*

$$\sup_{m(\cdot) \in \mathcal{M}(\rho_n)} P(t_n \leq z_\alpha) \geq 1 - \alpha + o(1), \quad whenever \quad \rho_n = o(\tilde{\rho}_n), \tag{2}$$

*where $z_\alpha$ indicate the $\alpha$ level critical value of test $t_n$.*

Proposition 1 shows that no test has non-trivial uniform power when alternative includes any functions in $\mathcal{M}$ that satisfy $\inf_{\theta \in \Theta} E[\delta_\theta(Z_i)^2] = o(n^{-1/4})$.

Proposition 1 is proved by replacing the minimax problem with a Bayesian problem, which the conventional technique to show the lower bound of the optimal minimax rate (e.g., Ingster, 1993; Spokoiny, 1996; Lepski & Spokoiny, 1999; Lepski & Tsybakov, 2000; Guerre & Lavergne, 2002; Abramovich et al., 2009; Ingster & Sapatinas, 2009). The proof is given in Appendix A.

We propose a test statistic that uses the feature of the nearest neighbor observations. Let $K_{i,j}$ denote the indicator function, which takes 1 if the observation $j$ is the nearest neighbor of observation $i$ and 0 otherwise. Formally, $K_{i,j} = \mathbb{1}(||Z_i - Z_j|| \leq ||Z_i - Z_k||, \forall k \neq i)$ if $i \neq j$ and $K_{i,j} = 0$ if $i = j$, where $|| \cdot ||$ is the Euclidean norm. The nearest neighbor $i^*$ of $i$ is the observation that satisfies $K_{i,i*} = 1$. Let $Y_{i^*}$ and $X_{i^*}$ be the observations of individual $i^*$ that satisfies $K_{i,i*} = 1$. We also define $u_{i^*} = Y_{i^*} - g(X_{i^*}, \theta)$.

Let $\hat{u}_i \equiv Y_i - g(X_i, \hat{\theta})$ be the residuals from the parametric estimation. Then, the test statistic is

$$T_n = \frac{1}{\hat{\mu}\sqrt{n}} \sum_{i=1}^{n} \hat{u}_i \hat{u}_{i^*}, \tag{3}$$

where $\hat{\mu}^2 \equiv n^{-1} \sum_{i=2}^{n} \sum_{j<i} W_{i,j}^2 \hat{u}_i^2 \hat{u}_j^2$ appears for the standardization. The weighting term $W_{i,j}$ is defined as $W_{i,j} = K_{i,j} + K_{j,i}$.

The test statistic is the sample analogue of $E(u_i u_{i^*})$ with normalization. Since

nearest neighbors are determined by instruments, $u_i$ and $u_{i^*}$ are independent conditional on $Z$ for all $i$. This implies $E(u_i u_{i^*}) = E[E(u_i|Z_i)E(u_{i^*}|Z_{i^*})]$, which is zero under $H_0$. Under $H_1$, however, $E[E(u_i|Z_i)E(u_{i^*}|Z_{i^*})] = E[\delta_\theta(Z_i)\delta_\theta(Z_{i^*})] \approx E[\delta_\theta(Z_i)^2] > 0$, which represents the source of testing power. The proposed test is simple and easily implementable with small calculation cost.

The proposed test is based on the difference between non-parametric and parametric variance estimators with bias correction. To observe this, we decompose the test statistic: $\frac{\hat{\mu}}{\sqrt{n}}T_n = \hat{\sigma}_p^2 - \hat{\sigma}_d^2 + \hat{B}$, where $\hat{\sigma}_p^2 \equiv \frac{1}{2n}\sum_{i=1}^n (\hat{u}_i^2 + \hat{u}_{i^*}^2)$ estimates the variance of error term in the parametric model (1) under the null hypothesis and $\hat{\sigma}_d^2 \equiv \frac{1}{2n}\sum_{i=1}^n (Y_i - Y_{i^*})^2$ is the non-parametric difference-based estimator for the variance of $Y_i - E(Y_i|Z_i)$, denoted as $\sigma^2$. The third term $\hat{B}$ is the bias correction for the non-parametric difference-based estimator: $\hat{\sigma}_d^2 - \sigma^2 = \frac{1}{2n}\sum_{i=1}^n (Y_i - Y_{i^*})^2 - \sigma^2 = \left[\frac{1}{2n}\sum_{i=1}^n (u_i - u_{i^*})^2 - \sigma^2\right] + B$, where $B = \frac{1}{n}\sum_{i=1}^n (u_i - u_{i^*})[g(X_i, \theta) - g(X_{i^*}, \theta)] + \frac{1}{2n}\sum_{i=1}^n [g(X_i, \theta) - g(X_{i^*}, \theta)]^2$. $\frac{1}{2n}\sum_{i=1}^n (u_i - u_{i^*})^2$ can be shown to converge to the true variance $\sigma^2$ when the model is true.[3] Thus, $T_n$ is modification of Yatchew's (1988) test with bias correction. This feature is not shared with the $K$-nearest neighbor specification test with increasing $K$, which is known to be rate optimal for the regression function against smooth alternative (H. Li et al., 2016).

Since at least one instrument is assumed continuous, ties do not exist theoretically. However, in practice, ties often exist, for example, when observed variables are rounded. In this case, the number of nearest neighbors may be greater than one and bounded from above, that is, $1 \leq \sum_{j \neq i}^n K_{i,j} \leq \infty$. When ties exist, the proposed test can be modified as follows

$$T_n^{tie} = \frac{1}{\hat{\mu}^{tie}\sqrt{\sum_{i=1}^n \sum_{j \neq i} K_{i,j}}} \sum_{i=1}^n \sum_{j \neq i}^n K_{i,j} \hat{u}_i \hat{u}_j, \qquad (4)$$

where $(\hat{\mu}^{tie})^2 \equiv (\sum_{i=1}^n \sum_{j \neq i} K_{i,j})^{-1} \sum_{i=2}^n \sum_{j<i} W_{i,j}^2 \hat{u}_i^2 \hat{u}_j^2$. Since the number of ties are

---

[3]See, e.g., the proof of Theorem 1 of Yatchew (1988). The difference-based estimator is first provided by Von Neumann, Kent, Bellinson, and Hart (1941) and developed by Gasser, Sroka, and Jennen-Steinmetz (1986), Hall, Kay, and Titterinton (1990), and Munk, Bissantz, Wagner, and Freitag (2005), among others.

asymptotically negligible, the asymptotic distribution, size, and power of the modified test are the same as those of $T_n$. Thus, in the following, discussion are focused on $T_n$.

The following proposition shows that the proposed test converges to the standard normal distribution under the null hypothesis.

**Proposition 2.** *Suppose Assumptions 1, 2, 3, 4, 5, 6, and 10 hold. Then, under the null hypothesis,*

$$T_n \xrightarrow{d} N(0,1). \tag{5}$$

The test is asymptotically one-sided because under the alternative, the source of testing power comes from $\delta_\theta(Z_i)\delta_\theta(Z_{i^*}) = \delta_\theta(Z_i)^2 + [\delta_\theta(Z_{i^*}) - \delta_\theta(Z_i)]\delta_\theta(Z_i)$, which takes positive values when $Z_i = Z_{i^*}$.[4] Thus, the null is rejected when the test statistic lies above the $(1-\alpha)$ quantile of the normal distribution, where $\alpha$ is a significance level.

Proposition 3 below shows that the proposed test has non-trivial uniform power against $H_{n,1}$ that approach the null hypothesis at the rate $\kappa n^{-1/4}$ for a constant $\kappa$.

**Proposition 3.** *Suppose Assumptions 1, 2, 3, 4, 5, 6, 7, 8, and 11 hold. Let $\rho_n = n^{-1/4}$. For any prescribed bound $\beta \in (0, 1-\alpha)$ and any $\theta \in \Theta$, a constant $\kappa$ exists such that*

$$\sup_{m(\cdot)\in\mathcal{M}(\kappa\rho_n)} P(T_n \leq z_\alpha) \leq \beta + o(1). \tag{6}$$

Together with the lower bound in Proposition 1, this result indicates that the proposed test is rate optimal and the optimal minimax rate of the specification tests for IV models against non-smooth alternative $H_{n,1}$ is $n^{-1/4}$.

# 5    Monte Carlo Experiments

Simulations are used to investigate the finite-sample performance of $T_n$ against various non-smooth alternatives. We compare the power performance of $T_n$ with that of a kernel

---

[4]The continuity of $Z$ in Assumption 2 is one of the sufficient conditions to make $Z_{i^*}$ approach $Z_i$ as the sample size increases (see, e.g., Lemma 14.1 of Q. Li & Racine, 2007).

smoothing test (Zheng, 1996). Hitomi et al. (2020) shows that the kernel test for the IV regression model is rate optimal against the set of smooth alternatives and the power can outperform that of other existing tests for IV regression models, including integrated conditional moment tests of Bierens (Bierens, 1982 and Bierens & Ploberger, 1997) and Horowitz (2006) and the exponential tilting test of Donald, Imbens, and Newey (2003).

We test null hypotheses that

$$g(x) = \theta_0 + \theta_1 x, \tag{7}$$

$$g(x, w_1, w_2) = \theta_0 + \theta_1 x + \theta_2 w_1 + \theta_3 w_2, \tag{8}$$

and

$$g(x, w_1, \ldots, w_{10}) = \theta_0 + \theta_1 x + \theta_2 w_1 + \cdots + \theta_{10} w_9. \tag{9}$$

Following Horowitz (2006), we use the following data generating processes (DGP): $X = \Phi\left(\rho v_1 + (1 - \rho^2)^{1/2} v_2\right)$, $Z = \Phi(v_1)$, and $U = 0.2\Phi\left(\eta v_2 + (1 - \eta^2)^{1/2} v_3\right)$, where $\Phi(\cdot)$ denotes the standard normal distribution function. Realizations of $(X, Z, U)$ are obtained by sampling $v_1$, $v_2$, and $v_3$ randomly from $N(0, 1)$. Included exogenous variables $W_1, W_2, \ldots, W_9$ are drawn randomly from $U[0, 1]$.

We consider three sets of DGPs, called DGP 1, DGP 2, and DGP 3: DGP 1: $\rho = 0.8$ and $\eta = 0.1$; DGP 2: $\rho = 0.8$ and $\eta = 0.5$; DGP 3: $\rho = 0.7$ and $\eta = 0.1$.

The outcome is generated by

$$Y = g(\cdot) + \beta_z h(Z) + U, \tag{10}$$

where $g(\cdot)$ represents one of (7), (8), and (9). Misspecification is introduced by the term $\beta_z h(Z)$, where the function $h(\cdot)$ is a Haar wavelet function: $h(z) = \delta$ if $z \in (0, \alpha]$, $h(z) = -\delta$ if $z \in (\alpha, 2\alpha]$, and $h(z) = 0$, otherwise. Because the form of misspecification is determined by $\alpha$ and $\beta$, we run simulations for multiple values of $\alpha$ in the range $(0, 0.5]$.

Then, $\delta$ is given such that $\delta\alpha = 0.05$. For illustrative purposes, Figure 1 shows $h(\cdot)$ for $\alpha = \{0.035, 0.260, 0.485\}$. In all experiments, $\theta_0 = \theta_1 = \cdots = \theta_{10} = 1$, $\beta_z = 0.5$.

The parameters are estimated by an efficient GMM (two-stage least squares). Instruments for $X$ are $Z$ and included exogenous variable(s) under the null model. Sample sizes are set to $n = \{200, 500, 1000\}$ and the results are based on $M = 1,000$ simulation runs.

Table 1 shows the size of $T_n$ at the 5% significance level. The test tends to under-reject the null hypothesis in most cases, indicating that the test is conservative. Although under-rejection may be remarkable when the model is (9), the estimated sizes become closer to the nominal sizes for all DGPs as the sample size increases.

Figure 2 displays the probability of rejecting the null hypothesis for several values of $\alpha$. For all DGPs, the power performance improves as the sample size grows. Overall, power is high for all $\alpha$ except its boundaries. This indicates that the test has less power when the misspecification of a function is concentrated on a narrow range ($\alpha$ small) or is distributed on a wide range ($\alpha$ large) on its support. When $\alpha$ is not at its boundaries, power is close to one even when the dimension of instruments is 10 when $n = 1000$ (third row).

Figure 3 displays the power performance of the kernel smoothing test. The kernel smoothing test and the proposed test (Figure 2) share similar features such that power is lower at the boundary of $\alpha$. A remarkable difference is that the kernel smoothing test does not have reasonable power when the dimension of instruments is 10 (third row). Simulation results support the theoretical results that the rate optimality of the proposed test does not depend on the dimension of instruments, while that of the kernel smoothing test depend on the dimension of instruments. The kernel smoothing test is rate optimal when the alternative is smooth relative to the dimension of instrument. The alternative models employed in this simulation may not smooth enough for the kernel type test to have reasonable power.

Overall, simulation results indicate that the proposed test offers an alternative choice when true regression function might be non-smooth relative to its dimension.

# 6 Application

This section tests the specification of Engel curves (consumer expansion paths). Empirical analyses of Engel curves are important in understanding consumer behavior, because the shape of it illustrates the elasticity of commodities toward total expenditure (inferior, normal, or luxury).

The true Engel curve for foods may be non-smooth (non-differentiable). The expenditure for basic foods is likely to increases as income increases but after the basic need is satisfied, higher proportion of income may spend on less necessary goods. When households' preferences are structured in a hierarchical manner like this, the resulting Engel curve is known to exhibit kinks at points that reflects the change in the capacity of the good as income increases (Drakopoulos, 1994). In the following application, we use the modified test statistic (4) when ties exist.

The baseline model is the Working–Leser specification of Engel curves, a structural model originating in consumer theory (Muellbauer, 1976, Deaton & Muellbauer, 1980, and Jorgenson, Lau, & Stoker, 1982). Let $y_{i,j}$ be individual $i$'s expenditure on good $j$, $X_i \equiv \sum_j y_{i,j}$ be total expenditures, and $W$ be a vector of exogenous variables. Then, the Working–Leser specification of Engel curves is

$$\frac{y_{i,j}}{X_i} = \alpha_{0,j} + \alpha'_{1,j} W_{i,j} + \beta_j \log X_i + \epsilon_{i,j}, \tag{11}$$

where $\epsilon$ is unobserved regression error and $\alpha_0$, $\alpha_1$ and $\beta$ are unknown parameters to be estimated.

We adapts the data and application strategy of Battistin and De Nadai (2015).[5] The

_____

[5]When expenditure data mare measured with errors, the conventional IV approach fails to obtain

data are the 2010 wave of the Bank of Italy's *Survey on Households' Income and Wealth* (SHIW). To estimate Engel curves for food, the model is estimated for each demographic group (couples without children, couples with one child, and couples with more than one child). Exogenous variables $W$ are the household regional variation represented by macro area dummies (North, Center, and South). Instruments for the total expenditure are the average of male logged wages across micro areas (over 100). Mean wages are likely to be uncorrelated with household unobserved characteristics and strongly correlated with total expenditures. The detailed explanation for data sets, estimation results, and sensitivity of the choice of instruments are given in Battistin and De Nadai (2015).

Table 2 presents $T_n$ values. The null hypothesis is that the specification of the model is true. The null is rejected at the 1% significance level for households with one child and more than one children.

# 7    Conclusion

This study investigated the optimal minimax rate when an alternative hypothesis is defined on the set of non-smooth functions $\mathcal{M}$. The set consists of bounded functions that are not necessarily differentiable when $l < 4$ with no smoothness constraints imposed on their first derivative when $l \geq 4$. The optimal minimax rate against such non-smooth alternatives is $n^{-1/4}$ for any model dimension $l$. The rate is lower than the optimal minimax rate reported in literature because the set of alternatives in this study consists of non-smooth functions.

The simple nearest neighbor test $T_n$ is rate optimal. Simulation results show that power of the proposed test can be higher than that of the existing kernel smoothing test when the alternative is non-smooth relative to the model dimensionality. Empirical

---

consistent parameter estimates because measurement errors are non-linear in Engel curves. A consistent estimator under the measurement error is proposed in Battistin and De Nadai (2015). In Appendix C, the model specification of their consistent estimator is also tested.

applications of Engel curves illustrate the good applicability of the proposed test.

A possible extension of this study, on which we are currently working, is rate optimal specification testing under many weak instruments. The body of literature on the estimation and inference of parameters in linear IV regression models with many or many weak instruments is growing (e.g., Andrews & Stock, 2007, Newey & Windmeijer, 2009, Anatolyev & Gospodinov, 2011, Lee & Okui, 2012, Chao, Hausman, Newey, Swanson, & Woutersen, 2014, and references therein). For specification testing, this study shows that the optimal minimax rate for IV regression models is $n^{-1/4}$ for any dimension $l$ of instruments. Although the results of this study hold for any fixed $l$, optimal minimax rates are not obvious under many instrument setups, in which $l$ grows with $n$. Asymptotic properties of specification testing and rate optimality under many weak setups have not been investigated sufficiently.

# References

Abramovich, F., Feis, D., Italia, S., & Theofanis. (2009). Optimal testing for additivity in multiple nonparametric regression. *Annals of the Institute of Statistical Mathematics*, *61(3)*, 691–714.

Anatolyev, S., & Gospodinov, N. (2011). Specification testing in models with many instruments. *Econometric Theory*, *27*, 427–441.

Andrews, D. W., & Stock, J. H. (2007). Testing with many weak instruments. *Journal of Econometrics*, *138*, 24–46.

Ball, L., Mankiw, N. G., & Romer, D. (1988). The new keynesian economics and the output-inflation trade-off. *Brookings papers on economic activity*, *1988*(1), 1–82.

Battistin, E., & De Nadai, M. (2015). Identification and estimation of engel curves with endogenous and unobserved expenditures. *Journal of Applied Econometrics*, *30(3)*, 487–508.

Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometrics*, *20(1)*, 105–134.

Bierens, H. J., & Ploberger, W. (1997). Asymptotic theory of integrated conditional

moment tests. *Econometrica*, *65(5)*, 1129–1151.

Billingsley, P. (2012). *Probability and measure, anniversary edition*. Wiley.

Carrasco, M., & Florens, J.-P. (2000). Generalization of GMM to a continuum of moment conditions. *Econometric Theory*, *16(6)*, 797–834.

Chao, J. C., Hausman, J. A., Newey, W. K., Swanson, N. R., & Woutersen, T. (2014). Testing overidentifying restrictions with many instruments and heteroskedasticity. *Journal of Econometrics*, *178*, 15–21.

Cohen, A., Daubechies, I., & Vial, P. (1993). Wavelets and fast wavelet transform on the interval. *Applied and Computational Harmonic Analysis*, *1*, 54–81.

Daubechies, I. (1992). *Ten lectures on wavelets*. Society for Industrial and Applied Mathematics.

Deaton, A., & Muellbauer, J. (1980). An almost ideal demand system. *The American economic review*, *70(3)*, 312–326.

Dominguez, M. A., & Lobato, I. N. (2004). Consistent estimation of models defined by conditional moment restrictions. *Econometrica*, *72(5)*, 1601–1615.

Donald, S. G., Imbens, G. W., & Newey, W. K. (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics*, *117(1)*, 55–93.

Dossche, M., Heylen, F., & Van den Poel, D. (2010). The kinked demand curve and price rigidity: Evidence from scanner data. *Scandinavian Journal of Economics*, *112*(4), 723–752.

Drakopoulos, S. A. (1994). Hierarchical choice in economics. *Journal of Economic Surveys*, *8*(2), 133–153.

Ermakov, M. S. (1991). Minimax detection of a signal in a Gaussian white noise. *Theory of Probability & Its Applications*, *35(4)*, 667–679.

Gasser, T., Sroka, L., & Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika*, *73(3)*, 625–633.

González-Manteiga, W., & Crujeiras, R. M. (2013). An updated review of goodness-of-fit tests for regression models. *Test*, *22*(3), 361–411.

Guerre, E., & Lavergne, P. (2002). Optimal minimax rates for nonparametric specification testing in regression models. *Econometric Theory*, *18(5)*, 1139–1171.

Hall, P., & Heyde, C. C. (1980). *Martingale limit theory and its application.* Academic press.

Hall, P., Kay, J. W., & Titterinton, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika*, *77(3)*, 521–528.

Hitomi, K., Iwasawa, M., & Nishiyama, Y. (2020). *Optimal minimax rates of specification testing with data-driven bandwidth.* available at: https://sites.google.com/view/masamune-iwasawa/research?authuser=0.

Horowitz, J. L. (2006). Testing a parametric model against a nonparametric alternative with identification through instrumental variables. *Econometrica*, *74(2)*, 521–538.

Horowitz, J. L., & Spokoiny, V. G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, *69(3)*, 599–631.

Ingster, Y. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I, II, III. *Mathematical Methods of Statistics*, *2(2)*, 85–114.

Ingster, Y. I., & Sapatinas, T. (2009). Minimax goodness-of-fit testing in multivariate nonparametric regression. *Mathematical Methods of Statistics*, *18(3)*, 241–269.

Jorgenson, D. W., Lau, L. J., & Stoker, T. M. (1982). The transcendental logarithmic model of aggregate consumer behavior. In R. Basmann & G. Rhodes (Eds.), *Advances in econometrics vol. 1* (pp. 97–238). Greenwich: JAI Press.

Jun, S. J., & Pinkse, J. (2012). Testing under weak identification with conditional moment restrictions. *Econometric Theory*, *28(6)*, 1229–1282.

Lee, Y., & Okui, R. (2012). Hahn–hausman test as a specification test. *Journal of*

*Econometrics*, *167*, 133–139.

Lepski, O. V., & Spokoiny, V. G. (1999). Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative. *Bernoulli*, *5(2)*, 333–358.

Lepski, O. V., & Tsybakov, A. (2000). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probability Theory and Related Fields*, *117(1)*, 17–48.

Lewbel, A. (1996). Demand estimation with expenditure measurement errors on the left and right hand side. *The Review of Economics and Statistics*, *78(4)*, 718–725.

Li, H., Li, Q., & Liu, R. (2016). Consistent model specification tests based on $k$-nearest-neighbor estimation method. *Journal of Econometrics*, *194(1)*, 187–202.

Li, Q., & Racine, J. S. (2007). *Nonparametric econometrics: Theory and practice.* Princeton University Press.

Muellbauer, J. (1976). Community preferences and the representative consumer. *Econometrica*, *44(5)*, 979–999.

Munk, A., Bissantz, N., Wagner, T., & Freitag, G. (2005). On difference-based variance estimation in nonparametric regression when the covariate is high dimensional. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67(1)*, 19–41.

Newey, W. K., & Windmeijer, F. (2009). Generalized method of moments with many weak moment conditions. *Econometrica*, *77*, 687–719.

Spokoiny, V. G. (1996). Adaptive hypothesis testing using wavelets. *The Annals of Statistics*, *24(6)*, 2477–2498.

Sweezy, P. M. (1939). Demand under conditions of oligopoly. *Journal of Political Economy*, *47*(4), 568–573.

Von Neumann, J., Kent, R. H., Bellinson, H. R., & Hart, B. I. (1941). The mean square successive difference. *The Annals of Mathematical Statistics*, *12(2)*, 153–162.

Yatchew, A. J. (1988). Dynamic econometric modeling, proceedings of the third in-

ternational symposium in economic theory and econometrics. In W. A. Barnett, E. R. Berndt, & H. White (Eds.), (pp. 121–135).

Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, *75(2)*, 263–289.

# Tables

Table 1: Size of the proposed test with 5% significance level.

| $n$ | DGP1 | | | DGP2 | | | DGP3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | (7) | (8) | (9) | (7) | (8) | (9) | (7) | (8) | (9) |
| 200 | 0.041 | 0.040 | 0.015 | 0.034 | 0.030 | 0.019 | 0.034 | 0.031 | 0.018 |
| 500 | 0.042 | 0.030 | 0.010 | 0.042 | 0.037 | 0.023 | 0.043 | 0.031 | 0.024 |
| 1000 | 0.041 | 0.037 | 0.025 | 0.054 | 0.040 | 0.028 | 0.049 | 0.040 | 0.033 |

*Note:* DGP 1: $\{\rho, \eta\} = \{0.8, 0.1\}$; DGP 2: $\{\rho, \eta\} = \{0.8, 0.5\}$; DGP 3: $\{\rho, \eta\} = \{0.7, 0.1\}$.

Table 2: Test for Engel curve specification (11) using SHIW 2010 data.

| | No children | One child | More than one child |
|---|---|---|---|
| $T_n$ | 0.414 | 2.765 | 3.193 |
| | (0.339) | (0.003) | (0.001) |
| Sample size | 345 | 709 | 1257 |

*Note:* Presented are the test statistics $T_n$ in equation (3). P-values are given in parentheses.

# Figures
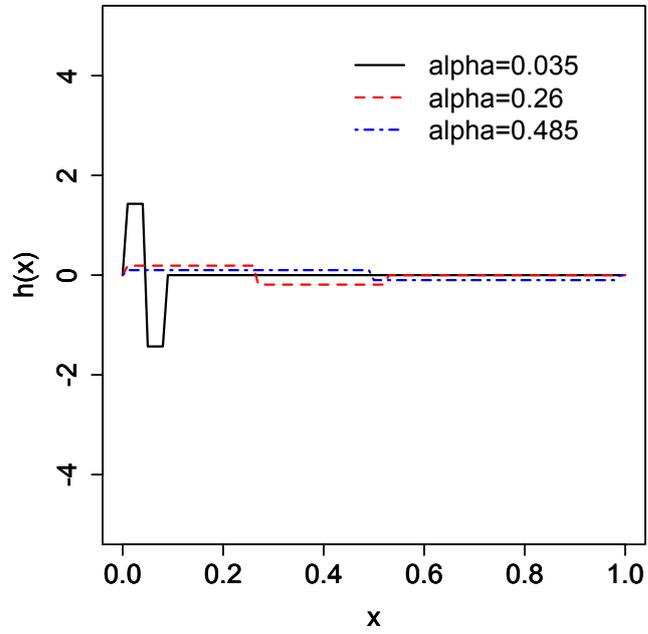


Figure 1: Haar wavelet functions that introduces misspecification.
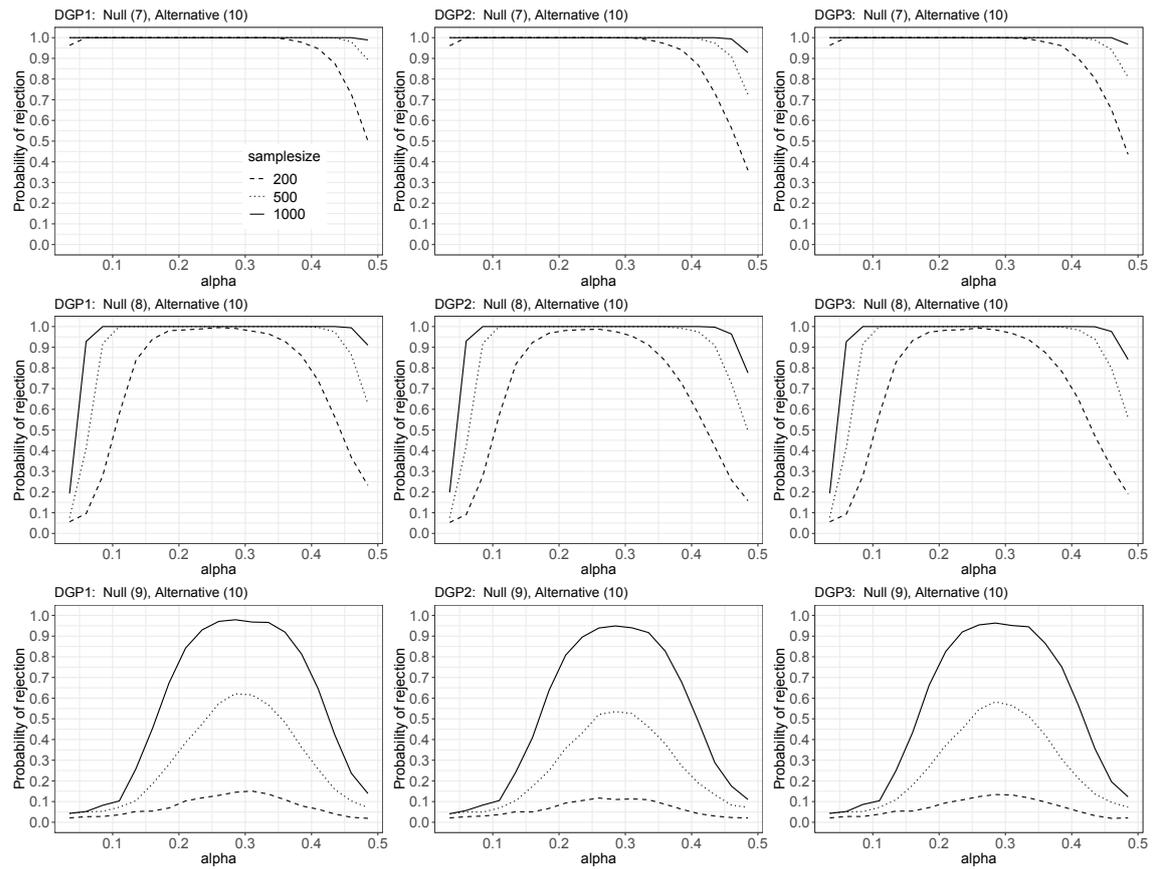
Figure 2: Power functions of proposed test: Probabilities of rejecting the null hypothesis are shown. Significance level is 5%. DGP 1: $\{\rho, \eta\} = \{0.8, 0.1\}$; DGP 2: $\{\rho, \eta\} = \{0.8, 0.5\}$; DGP 3: $\{\rho, \eta\} = \{0.7, 0.1\}$.
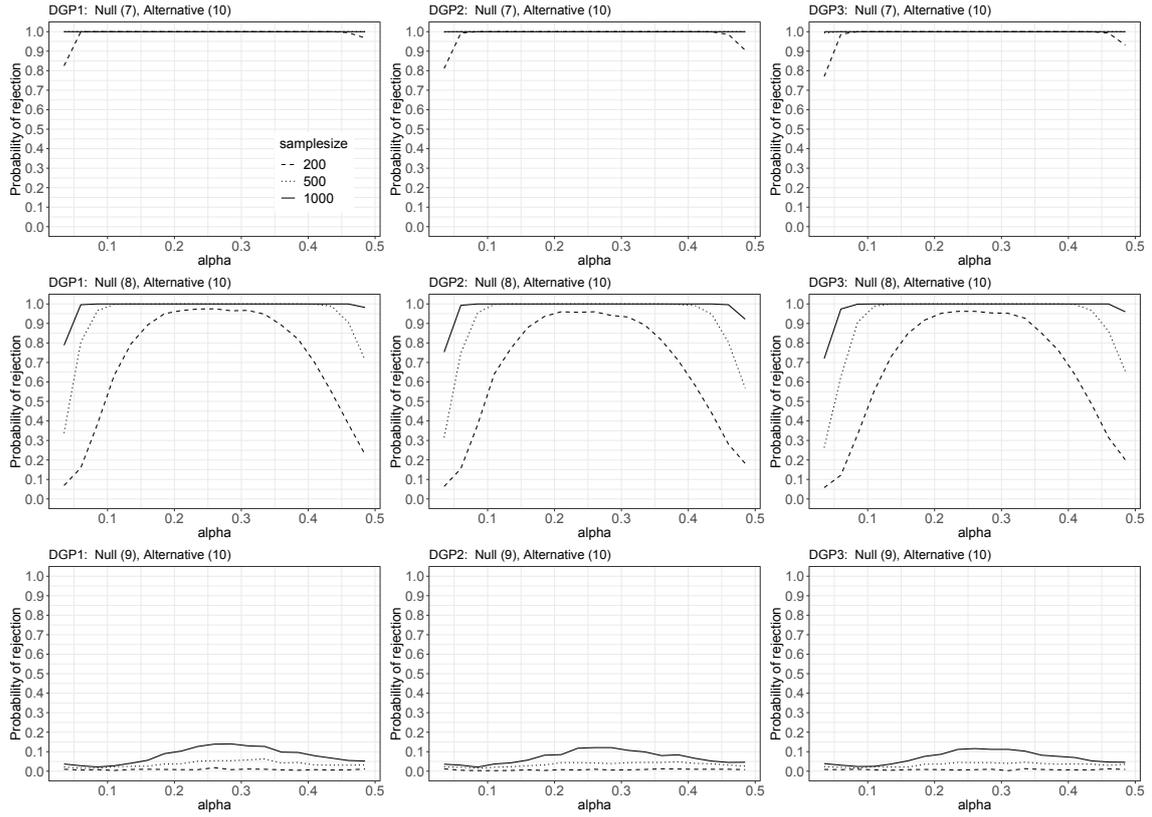
Figure 3: Power functions of kernel smoothing test: Probabilities of rejecting the null hypothesis are shown. Significance level is 5%. DGP 1: $\{\rho, \eta\} = \{0.8, 0.1\}$; DGP 2: $\{\rho, \eta\} = \{0.8, 0.5\}$; DGP 3: $\{\rho, \eta\} = \{0.7, 0.1\}$.

## APPENDIX: Proofs

Appendix gives proofs of propositions. Proofs for lemmas are given in the supplemental material A. Let the variance of $u_i$ conditioned on $Z_i$ be denoted by $\sigma^2(z) \equiv E(u_i^2|Z_i = z)$.

## A-1 Proofs of Proposition 1

*Proof of Proposition 1.*

Let $\psi_{j,\kappa}(z) = 2^{lj/2}\Psi(2^j z - \kappa) = 2^{lj/2}\psi(2^j z_1 - \kappa_1)\cdots\psi(2^j z_l - \kappa_l)$ for some $j \in \mathbb{Z}$ and $\kappa \equiv (\kappa_1,\ldots,\kappa_l)'$ be a $k$-times continuously differentiable orthonormal wavelet function defined on $[0, 2p-1]^l$ for some integer $p$ that satisfies $|\psi_{j,\kappa}(z)| \leq 2^{lj/2}C$ for some constant $C$. The orthonormality implies that $2^{-lj}E[\psi_{j,\kappa}(Z)\psi_{j',\kappa'}(Z)] = \mathbb{1}\{j = j'\}\mathbb{1}\{\kappa = \kappa'\}$, when random variable $Z$ is assumed to be uniformly distributed, where $\mathbb{1}\{j = j'\}$ is an indicator function taking 1 if $j = j'$ and zero otherwise.

The wavelet series $\psi_{j,\kappa}(z)$ defined above can be constructed by using, for example, Daubechies's orthonormal wavelets (Daubechies, 1992).[6] Let $\psi_D(\cdot)$ be Daubechies's orthonormal wavelet with support on $[-p+1, p]$ for some integer $p \geq 1$. The wavelet function becomes $\psi_{j,\kappa}(z) = 2^{lj/2}\psi_D(2^j z_1 - \kappa_1)\cdots\psi_D(2^j z_l - \kappa_l)$. By defining an appropriate collection of $\kappa$ for each $j$, the support of $\psi_{j,\kappa}(z)$ becomes $[0, 2p-1]^l$. Let $\mathcal{K}_j$ denote the collection of all possible distinct values for $\kappa$ such that $\mathcal{K}_j = \{\kappa \in \mathbb{Z}^l : \kappa_\iota = (p-1)+c(2p-1), c = 0, 1, \ldots, 2^j - 1, \iota = 1, 2, \ldots, l\}$. Then, $\mathcal{K}_j$ includes $2^{jl}$ elements for each $j$. Then dyadic cubes, $I_{j,\kappa} \equiv \prod_{\iota=1}^l ((-p+1+\kappa_\iota)2^{-j}, (\kappa_\iota+p)2^{-j}]$, satisfy $\cup_{\kappa \in \mathcal{K}_j} I_{j,\kappa} \subset [0, 2p-1]^l$. Since $\psi_D(2^j z_\iota - \kappa_\iota)$ is zero if $z_\iota$ lies outside of $((-p+1+\kappa_\iota)2^{-j}, (\kappa_\iota+p)2^{-j}]$, $\psi_{j,\kappa}(z)$ is zero if $z \notin I_{j,\kappa}$. Thus, the support of $\psi_{j,\kappa}(z)$ with $\kappa \in \mathcal{K}_j$ is $[0, 2p-1]^l$. Any intersection of two different cubes is always empty; that is., $I_{j,\kappa} \cap I_{j,\kappa'} = \varnothing$ for any $\kappa, \kappa' \in \mathcal{K}_j$ ($\kappa \neq \kappa'$), which implies $\psi_{j,\kappa}(z)\psi_{j,\kappa'}(z) = 0$. Our wavelet function is orthonormal, because it is the tensor product of Daubechies's wavelets. Furthermore, Daubechies's wavelets are

---

[6]Construction of a wavelet function with support $[0, 1]$ is also possible by using, for example, the method proposed by Cohen, Daubechies, and Vial (1993).

known to be $\nu p$ times continuously differentiable, where $\nu \approx 0.2$. Thus, $\psi_{j,\kappa}(\cdot)$ can be constructed to be $k$-times continuously differentiable by taking $p$ to be sufficiently large and satisfying $|\psi_{j,\kappa}(z)| \leq 2^{lj/2}C$ for some constant $C$, which implies $\Psi(\cdot) \leq C$ by definition.

Let $B_\kappa$ be any random sequence with $|B_\kappa| = 1$. We define for a positive constant $\lambda$ such that

$$\delta_{n,\theta_0}(\cdot) \equiv m_n(\cdot) - E[g(X, \theta_0)|\cdot], \qquad m_n(\cdot) = E[g(X, \theta_0)|\cdot] + \lambda\rho_n 2^{-jl/2} \sum_{\kappa \in \mathcal{K}_j} B_\kappa \psi_{j,\kappa}(\cdot),$$

where $E[g(X, \theta_0)|\cdot] \in \mathcal{M}$. The resolution level of wavelets is chosen by $j = \lfloor \log(n)/l \log(2) \rfloor$, where $\lfloor \cdot \rfloor$ is the floor function such that $\lfloor x \rfloor = \max\{z \in \mathbb{Z}|z \leq x\}$, implying $2^{jl} = O(n)$.

Let $\check{\theta}_{m_n}$ satisfy $\inf_{\theta \in \Theta} E[\delta_\theta(Z_i)^2] = E[\delta_{\check{\theta}_{m_n}}(Z_i)^2]$. A constant $C$ exists such that $\|\check{\theta}_{m_n} - \theta_0\| \leq CE|\delta_n(Z_i)|$ under Assumptions 2, 3, and 9 (see Hitomi et al., 2020 for the derivation). Then, we obtain the following result.

**Lemma 1.** *Under Assumptions 1, 2, 3, 4, 8, and 9, $m_n(\cdot)$ belongs to the class of alternatives $\mathcal{M}(\rho_n)$ when $n$ is sufficiently large.*

We construct a Bayesian a priori measure by using the result of Lemma 1 and show that even the optimal Bayesian test with the smallest errors does not have non-trivial power. Replacing the minimax problem by a Bayesian problem is a standard argument to show the lower bound of testing power (see, e.g., Ingster, 1993; Spokoiny, 1996; Lepski & Spokoiny, 1999; Lepski & Tsybakov, 2000; Guerre & Lavergne, 2002; Abramovich et al., 2009; Ingster & Sapatinas, 2009).

A sufficient condition of Proposition 1 is $\overline{L}_n \equiv \sum_{\kappa \in \mathcal{K}_j} \left[ \lambda^2 \rho_n^2 2^{-jl} \sum_{i=1}^n \psi_{j,\kappa}(Z_i)^2 \right]^2 \xrightarrow{p} 0$, which can be derived through straightforward calculations (for the derivation, see Guerre & Lavergne, 2002).[7]

---

[7]An a priori Bayesian measure over $H_0 \cup H_{n,1}$ can be constructed as follows. The a priori distribution $\Pi_0$ defined on $H_0$ has Dirac mass: $\Pi_0[\delta_{\theta_0}(\cdot) = 0] = \Pi_0\{m(\cdot) = E[g(X, \theta_0)|\cdot]\} = 1$. Let $B_\kappa$ be an i.i.d. Rademacher random variable independent of the observations with $P(B_\kappa = 1) = P(B_\kappa = -1) =$

Using the orthonormality and boundedness of the wavelet function, we obtain

$$E(\overline{L}_n)$$

$$= \lambda^4 \rho_n^4 2^{-2jl} \sum_{i=1}^{n} \sum_{\kappa \in \mathcal{K}_j} E\left[\psi_{j,\kappa}(Z_i)^4\right] + \lambda^4 \rho_n^4 2^{-2jl} \sum_{i_1=1}^{n} \sum_{i_2 \neq i_1} \sum_{\kappa \in \mathcal{K}_j} E\left[\psi_{j,\kappa}(Z_{i_1})^2\right] E\left[\psi_{j,\kappa}(Z_{i_2})^2\right]$$

$$\leq \lambda^4 \rho_n^4 2^{-jl} \sum_{i=1}^{n} \sum_{\kappa \in \mathcal{K}_j} E\left[\psi_{j,\kappa}(Z_i)^2\right] + \lambda^4 \rho_n^4 n(n-1) 2^{-2jl} 2^{jl}$$

$$= \lambda^4 \rho_n^4 n 2^{-jl} 2^{jl} + \lambda^4 \rho_n^4 n(n-1) 2^{-2jl} 2^{jl} = O(\rho_n^4 n).$$

Because $\rho_n = o(\tilde{\rho}_n) = o(n^{-1/4})$ by assumption, $\overline{L}_n$ converges to zero in probability. $\quad\square$

## A-2    Proof of Proposition 2

*Proof of Proposition 2.* The frequency that an observation is assigned to be the nearest neighbor of other observations is finite because of the boundedness of the kissing number, that is, $\sum_{i=1}^{n} K_{i,j} < \infty$. Under $H_0$, $\hat{\mu}T_n$ can be decomposed as follows.

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{u}_i \hat{u}_{i^*} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [g(X_{i^*}, \theta_0) - g(X_{i^*}, \hat{\theta})] u_i + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i u_{i^*} + A_n, \qquad \text{(A.1)}$$

where $A_n$ is $o_p(1)$, which comes from the $\sqrt{n}$-consistency of $\hat{\theta}$, the smoothness and finite moment assumption imposed on $g(\cdot)$, and the boundedness of the number of nearest neighbors, $\sum_{i \neq j}^{n} K_{i,j} \leq \infty$.

The following lemmas show the asymptotic behaviors of the first and second terms of equation (A.1). This result cannot be deduced from Lemma B6 of Jun and Pinkse (2012) because the boundedness of the conditional expectation of parametric function is not assumed.

---

1/2.  The a priori distribution $\Pi_{n,1}$ defined on $H_{n,1}$ is $\Pi_{n,1}[\delta_{\theta_0}(\cdot) = \lambda \rho_n 2^{-jl/2} \sum_{\kappa \in \mathcal{K}_j} b_\kappa \psi_{j,\kappa}(\cdot)] = \prod_{\kappa \in \mathcal{K}_j} P(B_\kappa = b_\kappa)$, $b_\kappa \in \{-1, 1\}$, where Lemma 1 guarantees that $\Pi_{n,1}$ is an a priori measure over $H_{n,1}$. Then, $\Pi_n = \Pi_0 + \Pi_{n,1}$ is the a priori Bayesian measure over $H_0 \cup H_{n,1}$.

**Lemma 2.** *Under Assumptions 1, 2, 3, 4, 5, and 10, we have $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [g(X_{i^*}, \theta_0) - g(X_{i^*}, \hat{\theta})] u_i = o_p(1)$.*

**Lemma 3.** *Under Assumption 1, 2, we have $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i u_{i^*} \xrightarrow{d} N(0, \mu^2)$, where $\mu^2$ is asymptotic variance of $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i u_{i^*}$.*

We show $\hat{\mu}^2 \xrightarrow{p} \mu^2$, where $\hat{\mu}^2 = n^{-1} \sum_{i=2}^{n} \sum_{j<i}^{n} W_{i,j}^2 \hat{u}_i^2 \hat{u}_{i^*}^2$, in the following lemma.

**Lemma 4.** *Under Assumptions 1, 2, 3, 4, 5, 6, and 10, we have $\hat{\mu}^2 \xrightarrow{p} \mu^2$ under the null hypothesis.*

$\square$

## A-3 Proof of Proposition 3

*Proof of Proposition 3.* The following lemma holds for the asymptotic behavior of $\hat{\mu}$ under $H_{n,1}$.

**Lemma 5.** *Let Assumptions 1, 2, 3, 4, 5, 6, 7, and 11 hold. Let*

$$\bar{\mu} \equiv \lim_{n \to \infty} \frac{1}{n} \sum_{i=2}^{n} \sum_{j=1}^{i-1} W_{i,j}^2 E\left( u_i^{*2} \big| Z_i \right) E\left( u_j^{*2} \big| Z_j \right), \tag{A.2}$$

*where $u_i^* \equiv Y_i - g(X_i, \theta^*)$. Then, under $H_{n,1}$, $\hat{\mu}^2 = \bar{\mu} + o_p(1)$ and $\bar{\mu}$ is bounded from above uniformly in $m(\cdot) \in \mathcal{M}_\epsilon$.*

Next, we consider the asymptotic behavior of the test statistics under $H_{n,1}$. We decompose $\hat{\mu} T_n$ as follows:

$$\hat{\mu} T_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{u}_i \hat{u}_{i^*} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [Y_i - g(X_i, \hat{\theta})][Y_{i^*} - g(X_{i^*}, \hat{\theta})] = T_n^* - C_1 + C_2, \tag{A.3}$$

where $T_n^* \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [Y_i - g(X_i, \theta^*)][Y_{i^*} - g(X_{i^*}, \theta^*)]$, $C_1 \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_{\theta^*}(Z_i)[g(X_{i^*}, \hat{\theta}) - g(X_{i^*}, \theta^*)]$, and $C_2$ includes terms consisting of the vanishing term $g(X_{i^*}, \hat{\theta}) - g(X_{i^*}, \theta^*)$ times a random variable whose expectation conditioned on instruments is zero. It is

4

straightforward to show that $C_2 = o_p(1)$ uniformly in $m(\cdot) \in \mathcal{M}(\kappa n^{-1/4})$. Lemma 6 below shows that $C_1 = O_p(1)$.

**Lemma 6.** *Under Assumptions 1, 3, 5, 6, 8, and 11,* $\sup_{m(\cdot) \in \mathcal{M}(\kappa n^{-1/4})} C_1 = O_p(1)$.

There is a constant $C > 0$ such that $P(T_n \leq z_\alpha) \leq P\left(T_n^* \leq z_\alpha' + C\right) + o(1)$, where $z_\alpha' \equiv \bar{\mu} z_\alpha$ is bounded uniformly by Lemma 5. Further,

$$P\left(T_n^* \leq z_\alpha' + C\right) = P\left(-[T_n^* - E(T_n^*)] \geq E(T_n^*) - z_\alpha' - C\right) \ \leq \frac{var(T_n^*)}{\{E(T_n^*) - z_\alpha' - C\}^2},$$

if $E(T_n^*) - z_\alpha' - C > 0$. It is then sufficient to show that $\kappa$ can be chosen so that

$$E(T_n^*) - z_\alpha' - C > 0, \tag{A.4}$$

$$\frac{var(T_n^{*\prime})}{\{E(T_n^*) - z_\alpha' - C\}^2} \leq \beta, \tag{A.5}$$

uniformly in $m(\cdot) \in \mathcal{M}(\kappa n^{-1/4})$. Since $E[\delta_{\theta^*}^2(Z_i)] \geq \inf_{\theta \in \Theta} E[\delta_\theta^2(Z_i)] \geq \rho_n^2$, we obtain,

$$
\begin{aligned}
E(T_n^*) = \sqrt{n}E\left[\delta_{\theta^*}(Z_i)\delta_{\theta^*}(Z_{i^*})\right] &\geq \sqrt{n}E\left[\delta_{\theta^*}(Z_i)^2\right] - \sqrt{n}E[|\delta_{\theta^*}(Z_i) - \delta_{\theta^*}(Z_{i^*})|\,|\delta_{\theta^*}(Z_i)|] \\
&\geq \sqrt{n}E\left[\delta_{\theta^*}(Z_i)^2\right]\left[1 - \frac{E[|\delta_{\theta^*}(Z_i) - \delta_{\theta^*}(Z_{i^*})|^2]^{1/2}}{E\left[|\delta_{\theta^*}(Z_i)|^2\right]^{1/2}}\right] \\
&\geq \sqrt{n}\kappa^2\rho_n^2\left[1 - \frac{o(n^{\frac{1}{l} - \frac{1}{4}})E[\|Z_i - Z_{i^*}\|^2]^{1/2}}{\kappa\rho_n}\right] \\
&= \kappa^2\left[1 - o(1)\right],
\end{aligned}
$$

where the last equation is derived from $E[\|Z_i - Z_{i^*}\|] = O(n^{-1/l})$ under Assumption 2 (see, for example, Lemma 14.1 of Q. Li & Racine, 2007). Thus, $E(T_{n,1}^*)$ is positive for a sufficiently large $n$. Then, we obtain

$$\frac{E\left(T_n^*\right) - z_\alpha' - C}{|\sqrt{n}E\left[\delta_{\theta^*}(Z_i)\delta_{\theta^*}(Z_{i^*})\right]|} \geq 1 - \frac{|z_\alpha'| + C}{\kappa^2\{1 - o(1)\}}.$$

5

A large value of $\kappa$ makes the last term in the above equation arbitrarily close to one. Therefore, (A.4) holds by taking a sufficiently large $\kappa$.

To prove (A.5), we define $T_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_i$, where $\eta_i \equiv \sum_{j \neq i} K_{i,j}[Y_i - g(X_i, \theta^*)][Y_j - g(X_j, \theta^*)]$. Let $Z = \{Z_1, Z_2, \ldots, Z_n\}$. From the law of total variance, we obtain

$$var\left(T_n^*\right) = \frac{1}{n} var\left(\sum_{i=1}^n \eta_i\right) = \frac{1}{n} \sum_{i=1}^n E\left[var\left(\eta_i | Z\right)\right] + \frac{1}{n} var\left[\sum_{i=1}^n E\left(\eta_i | Z\right)\right],$$

where the last equality holds because $\eta_i$'s are uncorrelated given $Z$; that is, $\eta_i$'s are i.i.d. conditional on $Z$. Let $\bar{\eta}_i = E[g(X_i, \theta^*)|Z_i] - g(X_i, \theta^*) + \omega_i$. Then, it is obvious that $E(\bar{\eta}_i | Z_i) = 0$ and $E(\bar{\eta}_i^2 | Z_i) \leq E[g(X_i, \theta^*)|Z_i]^2 + E[g(X_i, \theta^*)^2 | Z_i] + \sigma^2(Z_i)$ is bounded by Assumptions 1 and 7. By using these and the boundedness of $\delta_{\theta^*}(Z_j)$, we can show that $E(\eta_i^2 | Z_i) = \sum_{j \neq i} K_{i,j} E\left\{ [\delta_{\theta^*}(Z_i) + \bar{\eta}_i]^2 \big| Z_i \right\} E\left\{ [\delta_{\theta^*}(Z_j) + \bar{\eta}_j]^2 \big| Z_j \right\}$ is bounded from above by a constant $\Lambda$. Similarly, there is a constant $\bar{\Lambda}$ such that $\frac{1}{n} var[\sum_{i=1}^n E(\eta_i | Z_i)] \leq \bar{\Lambda}^2$. This yields $var(T_n^*) \leq \Lambda + \bar{\Lambda}^2$. For a sufficiently large $n$ that satisfies $\left[1 - O(n^{-1/l})\right] > 0$, we obtain

$$\frac{var(T_n^*)}{\left| \sqrt{n} E\left[\delta_{\theta^*}(Z_i)\delta_{\theta^*}(Z_{i^*})\right]\right|^2} \leq \frac{\Lambda + \bar{\Lambda}^2}{\left| \kappa^2 \left\{ 1 - O(n^{-1/l}) \right\} \right|^2}. \tag{A.6}$$

Because this upper bound is bounded and decreasing in $\kappa$, (A.5) holds uniformly in $m(\cdot) \in \mathcal{M}(\kappa n^{-1/4})$.

$\square$

## Supplemental material for the paper entitled
## "Optimal Minimax Rates against Non-smooth Alternatives"

## Supplemental Material A: Proofs

### S.A-1 Proof of Lemma 1

*Proof.* Since $E[g(X, \theta_0)|\cdot] \in \mathcal{M}$, it is enough to show (i) $|\delta_{n,\theta_0}(\cdot)| < \infty$, (ii) $|\delta_{n,\theta_0}(x) - \delta_{n,\theta_0}(y)|^2 = o(\epsilon^{\frac{l}{2}-2}\|x-y\|^2)$ as $\epsilon \to 0$, and (iii) $\inf_{\theta \in \Theta} E[\delta_{n,\theta}(Z_i)^2] \geq \rho_n^2$.

(i) Let $z$ belongs to $I_{j,\kappa}$ without loss of generality. Then,

$$|\delta_{n,\theta_0}(z)| = |\lambda \rho_n 2^{-jl/2} B_\kappa \psi_{j,\kappa}(z)| \leq \lambda \rho_n C < \infty,$$

where $|\psi_{j,\kappa}(z)| \leq 2^{lj/2}C$ for some constant $C$ is used for the derivation.

(ii) Let $x$ belongs to $I_{j,\kappa'}$ and $y$ belongs to $I_{j,\kappa''}$ without loss of generality ($\kappa'$ may be equivalent to $\kappa''$). Then,

$$|\delta_{n,\theta_0}(x) - \delta_{n,\theta_0}(y)|^2 = \left|\lambda \rho_n 2^{-jl/2} \sum_{\kappa \in \mathcal{K}_j} B_\kappa[\psi_{j,\kappa}(x) - \psi_{j,\kappa}(y)]\right|^2$$

$$= \lambda^2 \rho_n^2 |B_{\kappa'}[\Psi_{j,\kappa'}(2^j x - \kappa') - \Psi_{j,\kappa'}(2^j y - \kappa')] - B_{\kappa''}[\Psi_{j,\kappa''}(2^j y - \kappa'') - \Psi_{j,\kappa''}(2^j x - \kappa'')]|^2$$

$$\leq 2\lambda^2 \rho_n^2 2^{2j} \|x-y\|^2 O(1) = O(n^{\frac{2}{l}} \rho_n^2 \|x-y\|^2) = o(n^{\frac{2}{l}-\frac{1}{2}} \|x-y\|^2),$$

where $2^j = O(n^{1/l})$ and $\rho_n = o(n^{-1/4})$. Thus, setting $n = \epsilon^{-l}$ yields (ii).

(iii) We have $\inf_{\theta \in \Theta} E[\delta_\theta(Z_i)^2] = E[\delta_{\check{\theta}_{mn}}(Z_i)^2]$. Then, Minkowski's inequality and Taylor expansion of $g(X_i, \check{\theta}_{mn})$ at $\theta_0$ under Assumptions 3 and 4 yields

$$\{E[\delta_{\check{\theta}_{mn}}(Z_i)^2]\}^{1/2} \geq \{E[\delta_{n,\theta_0}(Z_i)^2]\}^{1/2} - [E(\{E[g(X_i, \theta_0)|Z_i] - E[g(X_i, \check{\theta}_{mn})|Z_i]\}^2)]^{1/2}$$

$$\geq \{E[\delta_{n,\theta_0}(Z_i)^2]\}^{1/2} - C\|\check{\theta}_{mn} - \theta_0\|$$

$$\geq \{E[\delta_{n,\theta_0}(Z_i)^2]\}^{1/2} - C'E[|\delta_{n,\theta_0}(Z_i)|]. \tag{S.A.1}$$

for some positive constant $C$ and $C'$. With respect the the second term,

$$\begin{aligned}
E[|\delta_{n,\theta_0}(Z_i)|] &\leq \lambda\rho_n 2^{-jl/2} \sum_{\kappa\in\mathcal{K}_j} E[|\psi_{j,\kappa}(Z_i)|] \\
&= \lambda\rho_n \sum_{\kappa\in\mathcal{K}_j} \int_0^1 |\Psi(2^j z - \kappa)| f(z) dz \\
&= \lambda\rho_n 2^{-jl} \sum_{\kappa\in\mathcal{K}_j} \int_{-\kappa}^{2^j-\kappa} |\Psi(u)| f\left(\frac{u+\kappa}{2^j}\right) du \\
&\leq \lambda\rho_n 2^{-jl} \int_{-p+1}^{p} |\Psi(u)| \sum_{\kappa\in\mathcal{K}_j} f\left(\frac{u+\kappa}{2^j}\right) du. \tag{S.A.2}
\end{aligned}$$

Let $\mathcal{K}_{j,\iota}$ be the set of $\iota$th element in $\mathcal{K}_j$ for $\iota = 1,\ldots,l$. Let $f'_l(z) = \frac{\partial}{\partial z_l} f(z)$ be the partial differentiation of $f(z)$ with respect to the $l$th element of $z$, which is bounded from above by Assumption 2. Then, we obtain for some $u = (u_1,\ldots,u_l)'$,

$$\begin{aligned}
\sum_{\kappa\in\mathcal{K}_j} f\left(\frac{u+\kappa}{2^j}\right) &= 2^{jl} f(0) + 2^{-j} \sum_{\kappa_1\in\mathcal{K}_{j,1}} \cdots \sum_{\kappa_l\in\mathcal{K}_{j,l}} \left[\sum_{\iota=1}^{l}(u_\iota + \kappa_\iota) f'_\iota(\tilde{u})\right] \\
&\leq 2^{jl} f(0) + 2^{-j} \max_{\iota\in\{1,\ldots,l\}} |f'_\iota(\tilde{u})| \sum_{\kappa_1\in\mathcal{K}_{j,1}} \cdots \sum_{\kappa_l\in\mathcal{K}_{j,l}} \left[\sum_{\iota=1}^{l}|u_\iota + \kappa_\iota|\right] \\
&\leq 2^{jl} f(0) + 2^{-j} \bar{f}' \left[2^{jl} \sum_{\iota=1}^{l}|u_\iota| + \sum_{\kappa_1\in\mathcal{K}_{j,1}} \cdots \sum_{\kappa_l\in\mathcal{K}_{j,l}} \sum_{\iota=1}^{l}|\kappa_\iota|\right] \\
&= 2^{jl} f(0) + 2^{-j} \bar{f}' \left[2^{jl} \sum_{\iota=1}^{l}|u_\iota| + l2^{j(l-1)} \sum_{c=0}^{2^j-1}|(p-1) + c(2p-1)|\right] \\
&= 2^{jl} f(0) + 2^{j(l-1)} \bar{f}' \sum_{\iota=1}^{l}|u_\iota| + l2^{j(l-2)} \bar{f}'(2p-1)\frac{(2^j-1)2^j}{2} + l2^{j(l-2)} \bar{f}' 2^j (p-1) \\
&= 2^{jl} f(0) + 2^{j(l-1)} \bar{f}' \sum_{\iota=1}^{l}|u_\iota| + \frac{l(2p-1)\bar{f}'}{2}(2^{jl} - 2^{j(l-1)}) + l\bar{f}'(p-1)2^{j(l-1)}
\end{aligned}$$

$$= 2^{jl} \left[ f(0) + \frac{l(2p-1)\bar{f}'}{2} \right] + 2^{j(l-1)} \left[ \bar{f}' \sum_{\iota=1}^{l} |u_\iota| - \frac{l(2p-1)\bar{f}'}{2} + l\bar{f}'(p-1) \right].$$

Substituting this into equation (S.A.2) yields

$$
\begin{aligned}
E[|\delta_{n,\theta_0}(Z_i)|] &\leq \lambda \rho_n 2^{-jl} \int_{-p+1}^{p} |\Psi(u)| \sum_{\kappa \in \mathcal{K}_j} f\left( \frac{u+\kappa}{2^j} \right) du \\
&\leq \lambda \rho_n \left[ f(0) + \frac{l(2p-1)\bar{f}'}{2} \right] \int_{-p+1}^{p} |\Psi(u)| du \\
&\quad + \lambda \rho_n 2^{-j} \bar{f}' \int_{-p+1}^{p} |\Psi(u)| \sum_{\iota=1}^{l} |u_\iota| du \\
&\quad - \lambda \rho_n 2^{-j} \left[ \frac{l(2p-1)\bar{f}'}{2} + l\bar{f}'(p-1) \right] \int_{-p+1}^{p} |\Psi(u)| du \\
&\leq \lambda \rho_n C'',
\end{aligned}
\tag{S.A.3}
$$

for some positive constant $C''$ because $\int_{-p+1}^{p} |\Psi(u)| du$ is bounded and $2^{-j}$ shrinks by the definition of wavelet. A positive constant $\bar{f}'$ is defined in Assumptions 2.

Since $2^{-lj} \psi_{j,\kappa}(Z_i)$ is orthonormal and $\mathcal{I}_j$ includes $2^{jl}$ location shifts, we obtain

$$E[\delta_{n,\theta_0}(Z_i)^2] = \lambda^2 \rho_n^2 2^{-jl} \sum_{\kappa \in \mathcal{K}_j} E\left[ \psi_{j,\kappa}(Z_i)^2 \right] = \lambda^2 \rho_n^2 2^{jl}. \tag{S.A.4}$$

Equations (S.A.1), (S.A.3), and (S.A.4) yields $\{E[\delta_{\check{\theta}_{m_n}}(Z_i)^2]\}^{1/2} \geq \lambda \rho_n(2^{jl/2} - C'C'')$. Thus, $E[\delta_{n,\theta_0}(Z_i)^2] \geq \rho_n^2$ when $n$ is large enough.

$\square$

## S.A-2    Proof of Lemma 2

*Proof.* From the mean value theorem, we obtain

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [g(X_{i^*}, \theta_0) - g(X_{i^*}, \hat{\theta})] u_i$$

$$= \sqrt{n}(\hat{\theta} - \theta_0)' \frac{1}{n} \sum_{i=1}^{n} \underline{\mu}_i + \sqrt{n}(\hat{\theta} - \theta_0)' \frac{1}{n} \sum_{i=1}^{n} \overline{\mu}_i + \sqrt{n}(\hat{\theta} - \theta_0)' \mu_n \sqrt{n}(\hat{\theta} - \theta_0),$$

$$(\text{S.A.5})$$

where $\sqrt{n}(\hat{\theta} - \theta_0) = O_p(1)$, $\underline{\mu}_i \equiv \sum_{j<i} K_{i,j} \frac{\partial}{\partial \theta} g(X_j, \theta_0) u_i$, $\overline{\mu}_i \equiv \sum_{j>i} K_{i,j} \frac{\partial}{\partial \theta} g(X_j, \theta_0) u_i$, and $\mu_n \equiv \frac{1}{n\sqrt{n}} \sum_{i=1}^{n} \sum_{j\neq i} K_{i,j} \frac{\partial}{\partial \theta \partial \theta'} g(X_j, \theta)|_{\theta=\tilde{\theta}} u_i$ for an interior point $\tilde{\theta}$ between $\hat{\theta}$ and $\theta_0$. Note that $\underline{\mu}_i$ and $\overline{\mu}_i$ are martingale difference sequences with respect to $\sigma$-fields generated by $\{X_1, X_2, \ldots, X_{i-1}, Z_1, Z_2, \ldots, Z_n\}$, and $\{X_{i+1}, X_{i+2}, \ldots, X_n, Z_1, Z_2, \ldots, Z_n\}$, respectively. The variances of $\frac{1}{n} \sum_{i=1}^{n} \underline{\mu}_i$ and $\frac{1}{n} \sum_{j=1}^{n} \overline{\mu}_i$ can be straightforwardly shown to be $O(1/n)$. Thus, $\frac{1}{n} \sum_{j=1}^{n} \underline{\mu}_i \overset{p}{\to} 0$ and $\frac{1}{n} \sum_{j=1}^{n} \overline{\mu}_i \overset{p}{\to} 0$ from the Chebyshev's inequality. We can also show $\mu_n = o_p(1)$ by using the bounded second moments for $u_j$ and $\frac{\partial}{\partial \theta \partial \theta'} g(X_j, \theta_0)$. $\qquad \square$

## S.A-3 Proof of Lemma 3

*Proof.* We define $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_i u_{i*} = \frac{1}{\sqrt{n}} \sum_{i=2}^{n} \sum_{j=1}^{i-1} W_{i,j} u_i u_j \equiv \sum_{i=2}^{n} \epsilon_{n,i}$, where $W_{i,j} \equiv K_{i,j} + K_{j,i}$. Let $\mathcal{F}_{n,i}$ be a $\sigma$-field generated by $\{Y_1, Y_2, \ldots, Y_i, X_1, \ldots, X_i, Z_1, \ldots, Z_n\}$. It is obvious that $\mathcal{F}_{n,i}$ form a filtration, that is, $\mathcal{F}_{n,k} \subset \mathcal{F}_{n,k+1}$ holds, and $\epsilon_{n,i}$ is a martingale difference with respect to $\mathcal{F}_{n,i}$. Note that $\mu^2 \equiv \lim_{n\to\infty} \frac{1}{n} \sum_{i=2}^{n} \sum_{j=1}^{i-1} W_{i,j}^2 u_j^2 \sigma^2(Z_i) < \infty$, because $E(u_i^2 | \mathcal{F}_{n,i-1}) = \sigma^2(Z_i) < \infty$ by Assumption 1 and $\sum_{i=1}^{n} W_{i,j}^2 \leq \infty$. Furthermore, $\sum_{i=1}^{n} E[\epsilon_{n,i}^2 \mathbb{1}\{|\epsilon_{n,i}| \geq \epsilon\}] \leq \frac{n}{\epsilon} E[|\epsilon_{n,i}|^3] \leq O(n^{-1/2})$, by the boundedness of $E(u^4 | Z)$. Thus, applying Theorem 35.12 of Billingsley (2012) yields $\sum_{i=2}^{n} \epsilon_{n,i} \overset{d}{\to} N(0, \mu^2)$. $\qquad \square$

## S.A-4 Proof of Lemma 4

*Proof.* We show that $\hat{\mu}^2$ converges to $\mu^2$ defined above almost surely and $\mu^2$ is equivalent to $\lim_{n\to\infty} \frac{1}{n} \sum_{i=2}^{n} \sum_{j=1}^{i-1} W_{i,j}^2 \sigma^2(Z_j)\sigma^2(Z_i)$. $\hat{\mu}^2$ is represented as follows:

$$\hat{\mu}^2 = \frac{1}{n} \sum_{i=2}^{n} \sum_{j<i}^{n} W_{i,j}^2 u_i^2 u_j^2 + B, \tag{S.A.6}$$

where $B$ includes terms that converges to zero in probability. The convergence can be shown straightforwardly by using the $\sqrt{n}$-consistency of parameter estimates in Assumption 10, uniform convergence of the first and the second derivative of $g(x,\theta)$ with respect to $\theta \in \Theta$ under Assumptions 1, 3, 4, 5, and 10, and the boundedness given in Assumptions 1 and 6.

We apply Theorem 2.17 of Hall and Heyde (1980) to show the probability limit of the first term of equation (S.A.6). We define $\frac{1}{n} \sum_{i=2}^{n} \sum_{j<i}^{n} W_{i,j}^2 u_i^2 u_j^2 = \sum_{i=2}^{n} \nu_{n,i}$, where $\nu_{n,i} \equiv \frac{1}{n} \sum_{j=1}^{i-1} W_{i,j}^2 u_i^2 u_j^2$ is a martingale with respect to $\mathcal{F}_{n,i}$. According to Theorem 2.17 of Hall and Heyde (1980), $\sum_{i=2}^{n} \nu_{n,i}$ converges to $\lim_{n\to\infty} \sum_{i=2}^{n} E(\nu_{n,i}|\mathcal{F}_{n,i-1})$ almost surely because $\lim_{n\to\infty} \sum_{i=2}^{n} E(|\nu_{n,i}||\mathcal{F}_{n,i-1}) < \infty$.

Let $\sum_{i=2}^{n} E(\nu_{n,i}|\mathcal{F}_{n,i-1}) = \sum_{j=1}^{n-1} v_{n,j}$, where $v_{n,j} \equiv \frac{1}{n} \sum_{i=j+1}^{n} W_{i,j}^2 \sigma^2(Z_i) u_j^2$. Let $\overline{\mathcal{F}}_{n,j}$ be a $\sigma$-field generated by $\{Y_j, Y_{j+1}, \ldots, Y_n, X_j, X_{j+1}, \ldots, X_n, Z_1, Z_2, \ldots, Z_n\}$. Then, $v_{n,j}$ is a reversed martingale with respect to $\overline{\mathcal{F}}_{n,j}$. By applying Theorem 2.17 of Hall and Heyde (1980), $\sum_{j=1}^{n-1} v_{n,j}$ converges almost surely to $\lim_{n\to\infty} \sum_{i=2}^{n} E(v_{n,j}|\overline{\mathcal{F}}_{n,j+1}) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} W_{i,j}^2 \sigma^2(Z_i)\sigma^2(Z_j)$ because straightforward calculation leads to $\sum_{i=2}^{n} E(|v_{n,j}||\overline{\mathcal{F}}_{n,j+1}) \leq \frac{1}{n} \sum_{i=2}^{n} \sum_{i=j+1}^{n} W_{i,j}^2 \sigma^2(Z_i)\sigma^2(Z_j) < \infty$. $\square$

## S.A-5 Proof of Lemma 5

*Proof.* Under the alternative, we can show

$$E(u_i^{*2}|Z_i) < \infty, \tag{S.A.7}$$

5

uniformly in $m(\cdot) \in \mathcal{M}_\epsilon$. Indeed, $E(u_i^{*2}|Z_i)$ can be decomposed into $m(Z_i)^p$, $E(\omega_i^p|Z_i)$, $E[g(X,\theta)^p|z]$ for $p = 2$, and cross products of them with $p = 1$. Since $m(\cdot)$ belongs to the class of bounded functions $\mathcal{M}_\epsilon$ where $m(Z_i)^p$ is bounded uniformly $m(\cdot) \in \mathcal{M}_\epsilon$ for any $n$ under the alternative. $E(\omega_i^p|Z_i)$ and $E[g(X,\theta)^p|z]$ are bounded by Assumption 1 and 7, respectively. Since $\sum_{j=1}^{i-1} W_{i,j}^2$ is bounded by a constant, $\bar{\mu}$ is bounded from above uniformly in $m(\cdot) \in \mathcal{M}_\epsilon$.

Note that we obtain $\hat{\mu}^2 = \frac{1}{n} \sum_{i=2}^n \sum_{j<i}^n W_{i,j}^2 u_i^{*2} u_j^{*2} + D$ where $D$ includes terms that converges to zero in probability. This equation is a version of equation (S.A.6) in Lemma 4, and differs from (S.A.6) in points that it has $\theta^*$ instead of $\theta_0$ and error term under the pseudo true value $u_i^*$ instead of $u_i$. Thus, the limiting behavior of $\hat{\mu}^2$ can be shown by going along with the proof of Lemma 4 except points on which asymptotic behavior of parameter estimates under the alternative affects. Especially, the convergence can be shown straightforwardly by using the $\sqrt{n}$-consistency of parameter estimates in Assumption 11, uniform convergence of the first and the second derivative of $g(x,\theta)$ with respect to $\theta \in \Theta$ under Assumptions 1, 3, 4, 5, and 11, and the boundedness in Assumptions 4 and 6. The boundedness for the conditional expectation of error terms is now guaranteed by equation (S.A.7) under Assumptions 1 and 7. $\qquad \square$

## S.A-6   Proof of Lemma 6

*Proof.* Since $\sqrt{n}(\hat{\theta} - \theta^*) = O_p(1)$ uniformly in $m(\cdot) \in \mathcal{M}_\epsilon$ from Assumption 11, we have $C_1 = O_p(1)(C_1' + C_1'')$, where

$$C_1' \equiv \frac{1}{n} \sum_{i=1}^n \delta_{\theta^*}(Z_i) \frac{\partial}{\partial \theta} g(X_{i^*}, \theta^*),$$

$$C_1'' \equiv \frac{1}{n\sqrt{n}} \sum_{i=1}^n \delta_{\theta^*}(Z_i) \frac{\partial}{\partial \theta \partial \theta} g(X_{i^*}, \theta)|_{\theta = \tilde{\theta}^*}.$$

First, applying the Schwarz inequality yields $C_1'' = o_p(1)$, since $E\left[|\delta_{\theta^*}(Z_i)|^2\right]$ is

bounded by Assumption 8 under $H_{n,1}$ and the second derivative of $g$ converges to its expectation uniformly in $\theta \in \Theta$ under Assumptions 1, 3, 5, and 11.

Second, by Assumption 6, there is a constant $c > 0$ such that

$$E(\|C_1'\|) \leq \sum_{j \neq i} E \left\{ K_{i,j} \, |\delta_{\theta^*}(Z_i)| \, E \left[ \left\| \frac{\partial}{\partial \theta} g(X_j, \theta^*) \right\| \Big| Z_j \right] \right\} \leq cE \left( |\delta_{\theta^*}(Z_i)| \right) < \infty.$$

(S.A.8)

From the Markov's inequality, $P(\sup_m \|C_1'\| > c) < E(\sup_m |\delta_\theta(Z_i)|) < \infty$, which indicates $C_1'$ is stochastically bounded. Therefore, we yield $\sup_{m(\cdot) \in \mathcal{M}(\kappa n^{-1/4})} C_1' = O_p(1)$. □

# Supplemental Material B: Simulation

Tables S.B.1 and S.B.2 show Monte Carlo results for the power of the test. The results correspond to power functions illustrated in Figures 2 and 3, respectively.

Table S.B.1: Monte Carlo results: power of the proposed test.

| | 0.035 | 0.06 | 0.085 | 0.11 | 0.135 | 0.16 | 0.185 | 0.21 | 0.235 | 0.26 | 0.285 | 0.31 | 0.335 | 0.36 | 0.385 | 0.41 | 0.435 | 0.46 | 0.485 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Null (7), Alternative (10) DGP1 | | | | | | | | | | | | | | | | | | | |
| $n = 200$ | 0.96 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 | 0.98 | 0.95 | 0.88 | 0.73 | 0.50 |
| $n = 500$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 | 0.89 |
| $n = 1000$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 |
| Null (7), Alternative (10) DGP2 | | | | | | | | | | | | | | | | | | | |
| $n = 200$ | 0.96 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.97 | 0.94 | 0.87 | 0.73 | 0.56 | 0.36 |
| $n = 500$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.97 | 0.91 | 0.72 |
| $n = 1000$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 | 0.93 |
| Null (7), Alternative (10) DGP3 | | | | | | | | | | | | | | | | | | | |
| $n = 200$ | 0.96 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 | 0.98 | 0.96 | 0.90 | 0.80 | 0.65 | 0.44 |
| $n = 500$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 | 0.94 | 0.81 |
| $n = 1000$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.97 |
| Null (8), Alternative (10) DGP1 | | | | | | | | | | | | | | | | | | | |
| $n = 200$ | 0.06 | 0.10 | 0.28 | 0.58 | 0.84 | 0.94 | 0.98 | 0.98 | 0.99 | 0.99 | 0.99 | 0.98 | 0.96 | 0.93 | 0.86 | 0.74 | 0.56 | 0.37 | 0.23 |
| $n = 500$ | 0.08 | 0.41 | 0.92 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.97 | 0.86 | 0.63 |
| $n = 1000$ | 0.19 | 0.93 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 | 0.91 |
| Null (8), Alternative (10) DGP2 | | | | | | | | | | | | | | | | | | | |
| $n = 200$ | 0.05 | 0.09 | 0.28 | 0.57 | 0.82 | 0.92 | 0.97 | 0.98 | 0.98 | 0.99 | 0.98 | 0.95 | 0.91 | 0.83 | 0.72 | 0.58 | 0.42 | 0.26 | 0.16 |
| $n = 500$ | 0.08 | 0.42 | 0.92 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 | 0.97 | 0.91 | 0.72 | 0.49 |
| $n = 1000$ | 0.20 | 0.93 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.96 | 0.78 |
| Null (8), Alternative (10) DGP3 | | | | | | | | | | | | | | | | | | | |
| $n = 200$ | 0.06 | 0.09 | 0.28 | 0.58 | 0.83 | 0.93 | 0.97 | 0.98 | 0.98 | 0.99 | 0.98 | 0.97 | 0.94 | 0.88 | 0.78 | 0.65 | 0.47 | 0.32 | 0.19 |
| $n = 500$ | 0.08 | 0.41 | 0.92 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 | 0.94 | 0.80 | 0.56 |
| $n = 1000$ | 0.19 | 0.93 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 | 0.84 |
| Null (9), Alternative (10) DGP1 | | | | | | | | | | | | | | | | | | | |
| $n = 200$ | 0.02 | 0.03 | 0.03 | 0.04 | 0.05 | 0.05 | 0.07 | 0.10 | 0.12 | 0.13 | 0.14 | 0.15 | 0.14 | 0.11 | 0.08 | 0.06 | 0.04 | 0.02 | 0.02 |
| $n = 500$ | 0.04 | 0.05 | 0.05 | 0.07 | 0.10 | 0.18 | 0.28 | 0.39 | 0.47 | 0.57 | 0.62 | 0.62 | 0.57 | 0.48 | 0.36 | 0.26 | 0.16 | 0.10 | 0.07 |
| $n = 1000$ | 0.04 | 0.05 | 0.08 | 0.10 | 0.26 | 0.45 | 0.67 | 0.84 | 0.93 | 0.97 | 0.98 | 0.97 | 0.97 | 0.92 | 0.81 | 0.65 | 0.42 | 0.24 | 0.14 |
| Null (9), Alternative (10) DGP2 | | | | | | | | | | | | | | | | | | | |
| $n = 200$ | 0.02 | 0.03 | 0.03 | 0.04 | 0.05 | 0.05 | 0.07 | 0.09 | 0.10 | 0.12 | 0.11 | 0.11 | 0.11 | 0.09 | 0.06 | 0.04 | 0.03 | 0.02 | 0.02 |
| $n = 500$ | 0.04 | 0.05 | 0.05 | 0.07 | 0.10 | 0.17 | 0.25 | 0.36 | 0.43 | 0.52 | 0.54 | 0.53 | 0.46 | 0.38 | 0.27 | 0.19 | 0.13 | 0.08 | 0.07 |
| $n = 1000$ | 0.04 | 0.06 | 0.08 | 0.10 | 0.24 | 0.41 | 0.64 | 0.81 | 0.90 | 0.94 | 0.95 | 0.94 | 0.92 | 0.83 | 0.68 | 0.49 | 0.29 | 0.17 | 0.11 |
| Null (9), Alternative (10) DGP3 | | | | | | | | | | | | | | | | | | | |
| $n = 200$ | 0.02 | 0.03 | 0.03 | 0.04 | 0.05 | 0.05 | 0.07 | 0.09 | 0.11 | 0.12 | 0.13 | 0.13 | 0.12 | 0.10 | 0.08 | 0.05 | 0.03 | 0.02 | 0.02 |
| $n = 500$ | 0.04 | 0.05 | 0.05 | 0.07 | 0.11 | 0.18 | 0.27 | 0.37 | 0.45 | 0.55 | 0.58 | 0.56 | 0.51 | 0.42 | 0.30 | 0.21 | 0.14 | 0.10 | 0.07 |
| $n = 1000$ | 0.04 | 0.05 | 0.09 | 0.10 | 0.25 | 0.43 | 0.66 | 0.83 | 0.92 | 0.95 | 0.96 | 0.95 | 0.94 | 0.86 | 0.75 | 0.57 | 0.36 | 0.20 | 0.12 |

*Note:* This table corresponds to Figurer 2. DGP 1: $\{\rho, \eta\} = \{0.8, 0.1\}$; DGP 2: $\{\rho, \eta\} = \{0.8, 0.5\}$; DGP 3: $\{\rho, \eta\} = \{0.7, 0.1\}$.

Table S.B.2: Monte Carlo results: power of the kernel smoothing test.

| | 0.035 | 0.06 | 0.085 | 0.11 | 0.135 | 0.16 | 0.185 | 0.21 | 0.235 | 0.26 | 0.285 | 0.31 | 0.335 | 0.36 | 0.385 | 0.41 | 0.435 | 0.46 | 0.485 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Null (7), Alternative (10) DGP1 | | | | | | | | | | | | | | | | | | | |
| $n = 200$ | 0.82 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 | 0.97 |
| $n = 500$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| $n = 1000$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Null (7), Alternative (10) DGP2 | | | | | | | | | | | | | | | | | | | |
| $n = 200$ | 0.81 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 | 0.91 |
| $n = 500$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| $n = 1000$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Null (7), Alternative (10) DGP3 | | | | | | | | | | | | | | | | | | | |
| $n = 200$ | 0.77 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 | 0.93 |
| $n = 500$ | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| $n = 1000$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Null (8), Alternative (10) DGP1 | | | | | | | | | | | | | | | | | | | |
| $n = 200$ | 0.07 | 0.16 | 0.39 | 0.63 | 0.80 | 0.89 | 0.95 | 0.97 | 0.97 | 0.97 | 0.96 | 0.97 | 0.95 | 0.89 | 0.82 | 0.70 | 0.55 | 0.38 | 0.23 |
| $n = 500$ | 0.34 | 0.80 | 0.97 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 | 0.91 | 0.71 |
| $n = 1000$ | 0.79 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 |
| Null (8), Alternative (10) DGP2 | | | | | | | | | | | | | | | | | | | |
| $n = 200$ | 0.06 | 0.15 | 0.38 | 0.64 | 0.77 | 0.88 | 0.94 | 0.96 | 0.96 | 0.96 | 0.94 | 0.93 | 0.89 | 0.81 | 0.71 | 0.58 | 0.44 | 0.28 | 0.18 |
| $n = 500$ | 0.32 | 0.75 | 0.95 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 | 0.95 | 0.80 | 0.57 |
| $n = 1000$ | 0.75 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.92 |
| Null (8), Alternative (10) DGP3 | | | | | | | | | | | | | | | | | | | |
| $n = 200$ | 0.06 | 0.12 | 0.33 | 0.56 | 0.73 | 0.85 | 0.92 | 0.95 | 0.96 | 0.96 | 0.95 | 0.95 | 0.93 | 0.85 | 0.76 | 0.64 | 0.48 | 0.32 | 0.20 |
| $n = 500$ | 0.26 | 0.63 | 0.91 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.97 | 0.86 | 0.66 |
| $n = 1000$ | 0.72 | 0.97 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.96 |
| Null (9), Alternative (10) DGP1 | | | | | | | | | | | | | | | | | | | |
| $n = 200$ | 0.01 | 0.01 | 0.01 | 0.00 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.02 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| $n = 500$ | 0.02 | 0.02 | 0.01 | 0.02 | 0.03 | 0.03 | 0.04 | 0.04 | 0.05 | 0.05 | 0.05 | 0.06 | 0.06 | 0.04 | 0.04 | 0.03 | 0.03 | 0.03 | 0.03 |
| $n = 1000$ | 0.04 | 0.03 | 0.02 | 0.03 | 0.04 | 0.06 | 0.09 | 0.10 | 0.13 | 0.14 | 0.14 | 0.13 | 0.13 | 0.10 | 0.10 | 0.08 | 0.07 | 0.06 | 0.05 |
| Null (9), Alternative (10) DGP2 | | | | | | | | | | | | | | | | | | | |
| $n = 200$ | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| $n = 500$ | 0.02 | 0.02 | 0.01 | 0.02 | 0.02 | 0.03 | 0.03 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.05 | 0.04 | 0.04 | 0.03 | 0.03 |
| $n = 1000$ | 0.04 | 0.03 | 0.02 | 0.04 | 0.04 | 0.06 | 0.08 | 0.09 | 0.12 | 0.12 | 0.12 | 0.11 | 0.10 | 0.08 | 0.08 | 0.07 | 0.05 | 0.04 | 0.05 |
| Null (9), Alternative (10) DGP3 | | | | | | | | | | | | | | | | | | | |
| $n = 200$ | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.00 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 | 0.01 |
| $n = 500$ | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.05 | 0.04 | 0.04 | 0.04 | 0.04 | 0.03 | 0.04 |
| $n = 1000$ | 0.04 | 0.03 | 0.02 | 0.03 | 0.04 | 0.05 | 0.08 | 0.09 | 0.11 | 0.12 | 0.11 | 0.11 | 0.10 | 0.08 | 0.08 | 0.07 | 0.05 | 0.05 | 0.05 |

*Note:* This table corresponds to Figurer 3. DGP 1: $\{\rho, \eta\} = \{0.8, 0.1\}$; DGP 2: $\{\rho, \eta\} = \{0.8, 0.5\}$; DGP 3: $\{\rho, \eta\} = \{0.7, 0.1\}$.

## Supplemental Material C: Empirical Application

In the estimation of Engel curves, two possible sources of endogeneity exist. First, total expenditure seems to be simultaneously determined with expenditure for each good. Second, expenditure data may be measured with errors. The conventional IV approach applied to (11) fails to obtain consistent parameter estimates because measurement errors are non-linear in (11).

IV estimator is consistent under the model specification (assumption) proposed by Lewbel (1996) and Battistin and De Nadai (2015), In this Appendix, their approaches are explained and their specifications are tested by the proposed test.

### S.C-1   Model

Three estimation approaches are considered: IV, Lewbel's (1996), and Battistin and De Nadai's (2015) approaches. Although all of these approaches aim to estimate the same parameter $\beta_j$, they differ in assumptions on the source of endogeneity, which lead to different model specifications.

Let $Z_i$ be a vector of instruments that includes exogenous variables $W_{i,j}$. In the IV approach, it is assumed that

$$E\left(\left.\frac{y_{i,j}}{X_i}\right| Z_i\right) = \alpha_{0,j} + \alpha'_{1,j}W_{i,j} + \beta_j E(\log X_i | Z_i).$$

The parameters are identified through a 2SLS regression of $y/X$ on constant $W$ and $\log X$, where $\log X$ is instrumented by $Z$.

We assume additive measurement errors such that $y_{i,j} = \tilde{y}_{i,j} + \tilde{X}_i v_j$, where $\tilde{y}_{i,j}$ and $\tilde{X}_i \equiv \sum_j \tilde{y}_{i,j}$ are the real expenditures, and $v_i$ is a mean zero random variable independent of $\tilde{X}_i$, $W_i$, $\epsilon_{i,j}$, and $Z_i$. Note that equation (11) holds under the real

expenditures, which implies

$$\frac{y_{i,j}}{X_i} = \frac{\tilde{y}_{i,j}/\tilde{X}_i + v_j}{V} = \frac{\alpha_{0,j} + \alpha'_{1,j}W_{i,j} + \beta_j \log \tilde{X}_i + \epsilon_{i,j} + v_j}{V}, \qquad \text{(S.C.1)}$$

where $V \equiv 1 + \sum_j v_j$, and thus $X_i = \tilde{X}_i V$.

The independence and zero mean of $v_i$ implies that $E(X_i|Z_i) = E(\tilde{X}_i|Z_i)E(V) = E(\tilde{X}_i|Z_i)$, $E(W_{i,j}X_i|Z_i) = E(W_{i,j}\tilde{X}_i|Z_i)$, and $E(X_i \log X_i|Z_i) = E[\tilde{X}_i V(\log \tilde{X}_i V)|Z_i] = E[\tilde{X}_i \log \tilde{X}_i|Z_i] + E(\tilde{X}_i|Z_i)E(V \log V|Z_i)$. Thus, multiplying either side of equation (S.C.1) by $X_i$ and taking conditional expectations with respect to $Z_i$ yields

$$E\left(y_{i,j}|Z_i\right) = \tilde{\alpha}_{0,j}E(X_i|Z_i) + \alpha'_{1,j}E(W_{i,j}X_i|Z_i) + \beta_j E(X_i \log X_i|Z_i) + E(\tilde{X}_i \epsilon_{i,j}|Z_i),$$

where $\tilde{\alpha}_{0,j} \equiv \alpha_{0,j} - \beta_j E(V \log V|Z_i)$. In Lewbel's (1996) approach, it is assumed that no endogeneity caused by simultaneity exists so that $E(\epsilon_{i,j}|\tilde{X}_i) = 0$. Then, the parameters are identified through a 2SLS regression of $y$ on $X$, $WX$, and $X \log X$ without a constant and $Z$ as instruments.

To address the violation of $E(\tilde{X}_i \epsilon_{i,j}|Z_i) = 0$ assumed in Lewbel (1996), Battistin and De Nadai (2015) use a control function approach. Let $\eta_i$ be the residual term from the regression of $\log X_i$ on the set of instruments $Z_i$ and $\tilde{\eta}_i$ be the residual using $\log \tilde{X}_i$ instead of $\log X_i$. The authors set a parametric assumption that $E(\epsilon_{i,j}|Z_i, \tilde{\eta}_i) = \rho_i \tilde{\eta}_i$, which yields $E(\tilde{X}_i \epsilon_{i,j}|Z_i) = E[\tilde{X}_i E(\epsilon_{i,j}|Z_i, \tilde{\eta}_i)|Z_i] = \rho_i E(\tilde{X}_i \tilde{\eta}_i|Z_i)$.

Since $E(X_i \eta_i|Z_i) = E(\tilde{X}_i \tilde{\eta}_i|Z_i) + \text{cov}(V, \log V)$ by using $\eta_i = \tilde{\eta}_i + \log V - E(\log V)$, we obtain

$$E\left(y_{i,j}|Z_i\right) = \alpha_{0,j}E(X_i|Z_i) + \alpha'_{1,j}E(W_{i,j}X_i|Z_i) + \beta_j E(X_i \log X_i|Z_i) + \rho_i E(X_i \eta_{i,j}|Z_i),$$

where $\alpha_{0,j} \equiv \alpha_{0,j} - \beta_j E(V \log V|Z_i) - \rho_i \text{cov}(V, \log V)$. By replacing $\eta$ with its fitted values $\hat{\eta}$, parameters, including $\rho_j$, are identified through a 2SLS regression of $y$ on $X$,

$WX$, $X \log X$, and $X\hat{\eta}$ without a constant and $Z$ as instruments.

In summary, each of IV, Lewbel's (1996), and Battistin and De Nadai's (2015) approaches has its own econometric specification for Engel curves. They are represented in the following moment restrictions:

$$E\left(\epsilon_{i,j}^{\mathrm{IV}}\middle|\, Z_i\right) = 0, \quad E\left(\epsilon_{i,j}^{\mathrm{L}}\middle|\, Z_i\right) = 0, \ \text{ and } \ E\left(\epsilon_{i,j}^{\mathrm{BN}}\middle|\, Z_i\right) = 0, \qquad \text{(S.C.2)}$$

where

$$\epsilon_{i,j}^{\mathrm{IV}} \equiv y_{i,j}/X_i - \alpha_{0,j} - \alpha'_{1,j}W_{i,j} - \beta_j \log X_i$$

$$\epsilon_{i,j}^{\mathrm{L}} \equiv y_{i,j} - \tilde{\alpha}_{0,j}X_i - \alpha'_{1,j}W_{i,j}X_i - \beta_j X_i \log X_i$$

$$\epsilon_{i,j}^{\mathrm{BN}} \equiv y_{i,j} - \alpha_{0,j}X_i - \alpha'_{1,j}W_{i,j}X_i - \beta_j X_i \log X_i - \rho_i X_i \hat{\eta}_{i,j}.$$

When $\epsilon_{i,j}$ is exogenous to both $X_i$ and $Z_i$, all moment restrictions in (S.C.2) hold. Consider that the source of endogeneity is only the simultaneous determination (or omitted variables). Then, moment restriction of the IV approach holds, while that of Battistin and De Nadai's (2015) approach holds only if the parametric specification for $E(\epsilon_{i,j}|Z_i, \tilde{\eta}_i)$ is correct. The moment restriction of Lewbel's (1996) approach may not hold, since $E(\epsilon_{i,j}|\tilde{X}_i) = E(\epsilon_{i,j}|X_i) \neq 0$. By contrast, when endogeneity arises only from the measurement error of the form discussed above, moment restrictions of Lewbel's (1996) and Battistin and De Nadai's (2015) approaches hold, while those of the IV approach fail. When both simultaneity and measurement errors are present, only the moment restrictions of Battistin and De Nadai's (2015) approach hold under the correct parametric assumption for $E(\epsilon_{i,j}|Z_i, \tilde{\eta}_i)$.

Table S.C.1: Test for Engel curve specification using SHIW 2010 data.

|  | IV | | Lewbel (1996) | | BN (2015) | |
|---|---|---|---|---|---|---|
| No children | 0.414 | (0.339) | 5.161 | (0.000) | 2.230 | (0.013) |
| One child | 2.765 | (0.003) | 3.801 | (0.000) | 4.929 | (0.000) |
| More than one child | 3.193 | (0.001) | 5.612 | (0.000) | 5.595 | (0.000) |

*Note:* Presented are the test statistics $T_n$ in equation (3). Sample size for groups "No children," "One child," and "More than one child" are 345, 709, and 1257, respectively. P-values are given in parentheses. BN (2015) denotes Battistin and De Nadai (2015).

## S.C-2 Test Results

Table S.C.1 presents $T_n$ values. The null hypothesis is that the specification of the model is true. For households without children (the first row), the test rejects Lewbel's (1996) specification at the 1% significance level. This result coincides with the suggestion of Battistin and De Nadai (2015) that total expenditure endogeneity caused by simultaneity might be a more serious problem than measurement error, at least in these data. For households with one child and more than one child (the second and third rows, respectively), all model specifications are rejected even at the 1% significance level.