Econometric Theory, **38**, 2022, 454–496.
doi:10.1017/S0266466621000220

NONPARAMETRIC SIGNIFICANCE TESTING IN MEASUREMENT ERROR MODELS

HAO DONG
Southern Methodist University

LUKE TAYLOR
Aarhus University

We develop the first nonparametric significance test for regression models with classical measurement error in the regressors. In particular, a Cramér-von Mises test and a Kolmogorov–Smirnov test for the null hypothesis $E\left[Y|X^*,Z^*\right]=E\left[Y|X^*\right]$ are proposed when only noisy measurements of X^* and Z^* are available. The asymptotic null distributions of the test statistics are derived, and a bootstrap method is implemented to obtain the critical values. Despite the test statistics being constructed using deconvolution estimators, we show that the test can detect a sequence of local alternatives converging to the null at the $\sqrt{n}$-rate. We also highlight the finite sample performance of the test through a Monte Carlo study.

1. INTRODUCTION

R. A. Fisher stated, "in the investigation of living beings by biological methods, statistical tests of significance are essential" (Fisher, 1929, p. 191). In this paper, we propose the first nonparametric significance test for data contaminated with classical measurement error. The test is flexible in that the measurement error may enter through the variables being tested, the controlling variables, or a combination of both.

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 that use such estimators have incorrect size. 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.

The authors would like to thank the Editor, Associate Editor, three anonymous referees, Karun Adusumilli, Taisuke Otsu, Ricardo Reis, and Peter Robinson for helpful comments. Financial support from the SMU Dedman College Research Fund (12-412268; Dong), and the Aarhus University Research Fund (AUFF-26852; Taylor) is gratefully acknowledged. Address correspondence to Luke Taylor, Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus, Denmark; e-mail: lntaylor@econ.au.dk.

© The Author(s), 2021. Published by Cambridge University Press.

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

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, "nonsmoothing" tests were developed, which detect local alternatives at the $\sqrt{n}$-rate (see, for example, Bierens, 1982, 1990; Stute, 1997). This is in contrast to "smoothing" tests, which typically attain slower than $\sqrt{n}$-convergence (see, for example, Härdle and Mammen, 1993). However, smoothing tests have been shown to have advantages over nonsmoothing tests in detecting high-frequency alternatives (Fan and Li, 2000; Su and White, 2008) and are asymptotically pivotal. Nonetheless, the level accuracy of smoothing tests that use an approximation of the asymptotic distribution can be quite poor (Härdle and Mammen, 1993), and their performance can depend heavily on the bandwidth when using bootstrap critical values (Lavergne and Vuong, 2000).[1]

There is a plethora of research on nonparametric significance testing when data are uncontaminated (see, for example, Robinson, 1989; Chen and Fan, 1999; Delgado and González-Manteiga, 2001; DM henceforth). But as yet there appear to be no results on such tests in the presence of measurement error, although there has been work carried out on other testing problems in this setting. In particular, model specification testing has received recent attention, with Hall and Ma (2007b) proposing a nonsmoothing test and Otsu and Taylor (2020) providing results for a smoothing test. Although specification tests share some similar characteristics with significance tests, these tests are fundamentally different in the nature of the null hypothesis. The null model in the case of a specification test is typically parametric (for example, a linear specification), whereas a significance test has a nonparametric null. Indeed, in the majority of nonsmoothing specification tests, there is no need for any smoothing. In contrast, nonsmoothing significance tests still require nonparametric estimation (despite the somewhat misleading name).

There is also a growing literature on the related problem of variable selection in nonparametric and semiparametric models with measurement error (see, for example, Ma and Li, 2010; Zhang et al. 2017). Although the problem of variable selection is similar to significance testing, the two areas are quite distinct. Variable selection is concerned with finding the most parsimonious model while retaining as much predictive ability as possible; this typically involves the use of information criteria such as AIC or BIC. On the other hand, significance tests take no stance on the performance of a given model; they aim only to test the plausibility of a null hypothesis.

The test proposed in this paper uses kernel deconvolution to deal with the measurement error problem. Deconvolution has received considerable attention in recent decades, and it is impractical to mention more than just a fraction of the varied work in this literature. Within statistics and econometrics, deconvolution

[1] González-Manteiga and Crujeiras (2013) provide a useful review of this vast literature.

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

has been used primarily for density estimation (Carroll and Hall, 1988; Stefanski and Carroll, 1990) and estimation of a regression function (Fan and Truong, 1993; Butucea and Taupin, 2008); with this, inference procedures have also been developed (Bissantz et al., 2007; Kato and Sasaki, 2018, 2019). In addition, it is increasingly popular to construct deconvolution methods that do not require *a priori* knowledge of the measurement error distribution (Delaigle, Hall, and Meister, 2008; Hu and Ridder, 2012; Delaigle and Hall, 2016). See Schennach (2016) for a recent review of this growing literature.

This paper is organized as follows. Section 2 outlines the hypothesis of interest and the test statistics, as well as discussing possible alternatives to our test. Section 3 presents the asymptotic properties of our test statistics under the null hypothesis and the alternative hypothesis, respectively. Section 4 provides a consistent bootstrap procedure to compute critical values. Section 5 uses repeated measurements to relax the known error distribution assumption. Section 6 considers the small sample performance of our test through a Monte Carlo study. Finally, Section 7 concludes. We relegate mathematical proofs of theorems to Appendix A.

2. TESTING FRAMEWORK

Let $Y \in \mathbb{R}$ be a response variable and $(X^*, Z^*) \in \mathbb{R}^2$ be a pair of regressors. We assume that (X^*, Z^*) are unobservable due to measurement error, but that we can observe noisy measurements, X and Z, generated by

$$X = X^* + \epsilon, \quad E[\epsilon] = 0, \tag{2.1}$$

$$Z = Z^* + \nu, \quad E[\nu] = 0, \tag{2.2}$$

where (ϵ, ν) is a pair of classical measurement errors, i.e., ϵ and ν are independent of (X^*, Y, Z^*). Although independence between (ϵ, ν) and (X^*, Y, Z^*) is assumed, this is not in fact necessary; we only require $f_X^{\mathrm{ft}}(t) = f^{\mathrm{ft}}(t) f_\epsilon^{\mathrm{ft}}(t)$ and $f_Z^{\mathrm{ft}}(t) = f_{Z^*}^{\mathrm{ft}}(t) f_\nu^{\mathrm{ft}}(t)$, for all $t \in \mathbb{R}$, where f_A denotes the density of a random variable A, f denotes the density of X^*, and $m^{\mathrm{ft}}(t) \equiv \frac{1}{2\pi} \int e^{\mathrm{i}tu} m(u) du$ denotes the Fourier transform of a function m with $\mathrm{i} = \sqrt{-1}$.

In this paper, we are interested in testing the significance of Z^* when both X^* and Z^* are mismeasured, i.e., we wish to test the hypothesis

$$\mathrm{H}_0 : E[Y|X^*, Z^*] = E[Y|X^*] \quad \text{a.s.},$$

based on an independent and identically distributed (i.i.d.) sample of (X, Y, Z). Notice that the null hypothesis is equivalent to the conditional moment restriction

$$E\big[(Y - E[Y|X^*])|X^*, Z^*\big] = 0 \quad \text{a.s.} \tag{2.3}$$

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

In the spirit of Bierens (1982, 1990), (2.3) can be written as an unconditional moment restriction of the form

$$T(\xi) \equiv E\big[(Y - r(X^*))f(X^*)\mathcal{W}(X^*, Z^*; \xi)\big] = 0 \quad \text{for all } \xi \in \Xi,$$

where $r(x) = E[Y|X^* = x]$, $f(x) > 0$ a.s.,[2] and $\mathcal{W}(x, z; \xi)$ is a "generically totally revealing" function (Stinchcombe and White, 1998) indexed by $\xi \in \Xi$ for some compact set $\Xi \subseteq \mathbb{R}^2$ with nonempty interior. Common choices for $\mathcal{W}(x, z; \xi)$ include $e^{\mathrm{i}(x,z)\xi}$ and $e^{(x,z)\xi}$ as used in Bierens (1982, 1990), respectively, and $\mathbb{I}\{(x, z) \leq \xi\}$ proposed by Stute (1997), where $\mathbb{I}\{\cdot\}$ denotes the indicator function. In this paper, we take $\mathcal{W}(x, z; \xi) = e^{\mathrm{i}(x,z)\xi}$, as this simplifies many of the technical proofs. However, we conjecture that very similar results can be obtained using any generically totally revealing function.

In the error-free case, both X^* and Z^* are observable, and $T(\xi)$ can be estimated by

$$\check{T}_n(\xi) = \frac{1}{n}\sum_{j=1}^{n}\big(Y_j - \check{r}\big(X_j^*\big)\big)\check{f}\big(X_j^*\big)\, e^{\mathrm{i}(X_j^*, Z_j^*)\xi},$$

where $\check{r}(x)$ is a nonparametric kernel regression estimator of $r(x)$ and $\check{f}(x)$ is a kernel density estimator of $f(x)$, as in DM. In our case, $\check{T}_n(\xi)$ is infeasible as X_j^* and Z_j^* are unobservable due to measurement error. However, $T(\xi)$ can still be estimated by evaluating $r(x)$, $f(x)$, and the joint density $h(x, y, z)$ of (X^*, Y, Z^*) in the expression

$$T(\xi) = \iiint (y - r(x))f(x)h(x, y, z)e^{\mathrm{i}(x,z)\xi}\,dxdydz, \tag{2.4}$$

using deconvolution methods.

In particular, $r(x)$, $f(x)$, and $h(x, y, z)$ can be estimated by

$$\hat{r}(x) = \frac{\sum_{j=1}^{n} Y_j \mathbb{K}_\epsilon\left(\frac{x - X_j}{b_n}\right)}{\sum_{j=1}^{n} \mathbb{K}_\epsilon\left(\frac{x - X_j}{b_n}\right)}, \quad \hat{f}(x) = \frac{1}{nb_n}\sum_{j=1}^{n}\mathbb{K}_\epsilon\left(\frac{x - X_j}{b_n}\right),$$

$$\hat{h}(x, y, z) = \frac{1}{nb_n^3}\sum_{j=1}^{n} K\left(\frac{y - Y_j}{b_n}\right)\mathbb{K}_\epsilon\left(\frac{x - X_j}{b_n}\right)\mathbb{K}_\nu\left(\frac{z - Z_j}{b_n}\right),$$

where the deconvolution kernel associated with a random error η, denoted as $\mathbb{K}_\eta$, is defined as

$$\mathbb{K}_\eta(u) = \frac{1}{2\pi}\int e^{-\mathrm{i}tu}\frac{K^{\mathrm{ft}}(t)}{f_\eta^{\mathrm{ft}}(t/b_n)}dt,$$

with a standard kernel function K and a bandwidth b_n.

[2]The multiplication by $f(X^*)$ in the definition of $T(\xi)$ is used only to remove the random denominator in $E[Y|X^*]$ and hence simplify the analysis.

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

Substituting these estimators into (2.4) leads to the following estimator for $T(\xi)$:

$$\begin{aligned}
T_n(\xi) &= \iiint (y-\hat{r}(x))\hat{f}(x)\hat{h}(x,y,z)e^{\mathrm{i}(x,z)\xi}\,dxdydz \\
&= \frac{1}{n^2 b_n^4}\sum_{j=1}^{n}\sum_{k=1}^{n}\iint \left\{ \begin{array}{c} \mathbb{K}_\epsilon\left(\frac{x-X_j}{b_n}\right)\mathbb{K}_\epsilon\left(\frac{x-X_k}{b_n}\right)\mathbb{K}_\nu\left(\frac{z-Z_k}{b_n}\right)e^{\mathrm{i}(x,z)\xi} \\ \times \int \left(y-Y_j\right)K\left(\frac{y-Y_k}{b_n}\right)dy \end{array} \right\} dxdz \\
&= \frac{1}{n^2 b_n^3}\sum_{j=1}^{n}\sum_{k=1}^{n}\iint \left\{ \begin{array}{c} \mathbb{K}_\epsilon\left(\frac{x-X_j}{b_n}\right)\mathbb{K}_\epsilon\left(\frac{x-X_k}{b_n}\right)\mathbb{K}_\nu\left(\frac{z-Z_k}{b_n}\right)e^{\mathrm{i}(x,z)\xi} \\ \times \int \left(Y_k-Y_j+b_n y\right)K\left(y\right)dy \end{array} \right\} dxdz \\
&= \frac{1}{n^2 b_n^3}\sum_{j\neq k}^{n}(Y_k-Y_j)\iint \mathbb{K}_\epsilon\left(\frac{x-X_j}{b_n}\right)\mathbb{K}_\epsilon\left(\frac{x-X_k}{b_n}\right)\mathbb{K}_\nu \\
&\quad \times\left(\frac{z-Z_k}{b_n}\right)e^{\mathrm{i}(x,z)\xi}dxdz,
\end{aligned}$$

where the third equality follows by a change of variables and the last equality holds when $\int K(u)du = 1$ and $\int uK(u)du = 0$.

Note that smoothing along Y is not needed in the construction of $T_n(\xi)$, even though it is used in the construction of $\hat{h}$. Also note that the same bandwidth is used within each kernel function; however, this is only to keep the notation simple and is not necessary. In Section 6, we discuss a bandwidth selection scheme that can be applied individually to each bandwidth and which we use in the analysis of the finite sample performance of our test. Finally, f_ϵ and f_ν are assumed to be known in the construction of $T_n(\xi)$, which rarely holds in practice. In Section 5, we show how an additional noisy measurement of (X^*, Z^*) can be used to relax this known error distribution assumption, and prove the properties of the resulting test.

The test statistic is constructed as a continuous functional of $T_n(\xi)$. In this paper, we consider two popular examples of such functionals: the Cramér-von Mises-type and Kolmogorov–Smirnov-type functionals. Corresponding test statistics are given by

$$\begin{aligned}
CM_n &= \int_{\Xi}|T_n(\xi)|^2 d\xi, \\
KS_n &= \sup_{\xi\in\Xi}|T_n(\xi)|.
\end{aligned}$$

Throughout this paper, to keep the notation simple, we focus on the case of only two regressors which are both affected by measurement error. The proposed test, however, can easily be adapted to the general multivariate case with a combination of error-free and contaminated regressors. For example, when $X^* = (X_1^*, \ldots, X_{d_x}^*)$ is correctly measured and $Z^* = (Z_1^*, \ldots, Z_{d_z}^*)$ is unobservable due to measurement error, let $Z = (Z_1, \ldots, Z_{d_z})$ be a noisy measurement of Z^* generated by $Z_d = Z_d^* + \nu_d$ for $d = 1, \ldots, d_z$. Following a similar route to the bivariate fully contaminated

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

case, the test statistic can be constructed as $CM'_n = \int_\Xi |T'_n(\xi)|^2 d\xi$ or $KS'_n = \sup_{\xi\in\Xi} |T'_n(\xi)|$, where

$$T'_n(\xi) = \frac{1}{n^2 b_n^3} \sum_{j\neq k} (Y_k - Y_j) \iint \left\{ \begin{array}{c} \prod_{d=1}^{d_x} K\left(\frac{x_d - X^*_{d,j}}{b_n}\right) \prod_{d=1}^{d_x} K\left(\frac{x_d - X^*_{d,k}}{b_n}\right) \\ \times \prod_{d=1}^{d_z} \mathbb{K}_{\nu_d}\left(\frac{z_d - Z_{d,k}}{b_n}\right) e^{\mathrm{i}(x',z')\xi} \end{array} \right\} dxdz,$$

with $x = (x_1, \ldots, x_{d_x})'$, $z = (z_1, \ldots, z_{d_z})'$, and $\xi = (\xi_1, \ldots, \xi_{d_x+d_z})'$. We conjecture that analogous results to our main theorems can be established for such cases.

It is important to emphasize that conducting a conventional nonparametric significance test using the mismeasured regressors would, in general, lead to misleading results. To see this, note that by ignoring the measurement error, a conventional test would work with a null hypothesis of the form

$$\mathrm{H}'_0 : E[Y|X,Z] = E[Y|X] \quad \text{a.s.}$$

However, H'_0 fails under H_0 unless $\int E[Y|X^* = x][f_{X^*|X,Z}(x) - f_{X^*|X}(x)]dx = 0$, which would generally fail if $f_{X^*|X,Z} \neq f_{X^*|X}$, where $f_{X^*|X,Z}$ and $f_{X^*|X}$ separately denote the conditional density of X^* given (X,Z) and the conditional density of X^* given X. Hence, a conventional test which ignores measurement error has, in general, incorrect size.

There are, however, special cases where the measurement error can indeed be ignored and a conventional test using the contaminated sample can be conducted. One such example is when Z^* is independent of X^*. In such a situation, $f_{X^*|X,Z} = f_{X^*|X}$ given (ϵ, ν) are classical, and H'_0 holds if and only if H_0 is true. A conventional test can then be conducted without accounting for the measurement error issue.

A second example occurs when we wish to test the joint significance of all regressors. In such a situation, we test $E[Y|Z^*] = E[Y]$, which is equivalent to testing $E[Y|Z] = E[Y]$. Hence, the test constructed in this paper is applicable when there is interest in testing the significance of a nontrivial subset of regressors, which are not independent of other regressors, and when the regressors are contaminated with classical measurement error.

A potential alternative to our test could be constructed from linearizing the model by taking, for example, a finite polynomial in the regressors, estimating the regression coefficients using an IV approach, and conducting a Wald test. Note that an instrument is required in this case: typically a repeated measurement of the mismeasured regressor. Furthermore, for each transformation of each variable, the instrument is the analogous transformation of the repeated measurement. Of course, this is a simplification of the model and, although being valid, would not lead to a consistent test if the model is misspecified. However, for practical purposes, we may hope that this approach fares well. Unfortunately, as our simulation results in Section 6 indicate, this is not the case, in general. The reason is twofold. Nonlinear transformations of mismeasured regressors generally exacerbate the measurement error problem. In addition, the strength of an instrument typically deteriorates with nonlinear transformations.

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

Finally, it is also worth noting that while our test is able to accommodate discrete regressors if they are correctly measured, discrete mismeasured variables cannot be accounted for as they would violate the classical measurement error assumption.

3. ASYMPTOTIC THEORY

In this section, the asymptotic distributions of the test statistics CM_n and KS_n are established under the null hypothesis H_0 and under a sequence of local alternatives H_{1n} that converge to H_0 at the parametric rate $n^{-1/2}$. Let $g(x) = r(x)f(x)$ and $g_{X^*,Z^*}(x,z) = E[Y|X^* = x, Z^* = z]f_{X^*,Z^*}(x,z)$, where f_{X^*,Z^*} denotes the joint density of (X^*, Z^*). In addition, let $\binom{a}{b} = \frac{a!}{(a-b)!b!}$ denote the binomial coefficient for nonnegative integers $a \geq b$. For a function $m : \mathbb{R}^d \to \mathbb{R}$, denote $m^{(\delta_1,\ldots,\delta_d)}(u) = \partial^{\delta_1+\cdots+\delta_d} m(u)/\partial u_1^{\delta_1} \cdots \partial u_d^{\delta_d}$, where d is any positive integer, $u = (u_1, \ldots, u_d)$ is a d-dimensional vector, and $(\delta_1, \ldots, \delta_d)$ is a d-dimensional vector of nonnegative integers.

3.1. Asymptotic Null Distribution

To derive the asymptotic null distributions of the test statistics CM_n and KS_n, we first impose the following regularity conditions.

Assumption D.

(i) $\{X_j, Y_j, Z_j\}_{j=1}^n$ is an i.i.d. sample of (X, Y, Z), where (X, Z) satisfies (2.1) and (2.2) and $E|Y|^2 < \infty$.

(ii) (ϵ, ν) are mutually independent and are independent of (X^*, Y, Z^*) with characteristic functions $f_\epsilon^{\text{ft}}(t) \neq 0$ and $f_\nu^{\text{ft}}(t) \neq 0$, for all $t \in \mathbb{R}$. In addition, $\inf_{\xi \in \Xi} |f_\nu^{\text{ft}}(\xi_2)| > \underline{c}$ for a positive constant $\underline{c}$.

(iii) $b_n \to 0$ as $n \to \infty$.

Assumption D(i) requires random sampling and the existence of the second moment of Y. Assumption D(ii) is standard in the classical measurement error literature.[3] The assumptions $f_\epsilon^{\text{ft}}(t) \neq 0$ and $f_\nu^{\text{ft}}(t) \neq 0$, for all $t \in \mathbb{R}$, are needed

[3]The independence between (ϵ, ν) and Y required by Assumption D(ii) is imposed for identification. This full independence between the measurement error and the dependent variable is commonly used in the literature on nonparametric regression with errors-in-variables (Fan and Truong, 1993; Delaigle and Meister, 2007; Meister, 2009). However, it has recently been shown that this can be relaxed to a mean independence condition (Kato and Sasaki, 2019). In our setting, the identification of $T(\xi)$ hinges on the identification of $E[Y|X^*]$ and $E[Y|X^*, Z^*]$, which, as in the regression case, rely on

$$\{E[Y|X = \cdot]f_X(\cdot)\}(x) = \int \{E[Y|X^* = \cdot]f_{X^*}(\cdot)\}(x-u)f_\epsilon(u)du.,$$

and an analogous result for $\{E[Y|X = \cdot, Z = \cdot]f_{X,Z}(\cdot,\cdot)\}(x,z)$. Thus, for the identification of $T(\xi)$, instead of full independence between (ϵ, ν) and Y, it is sufficient to require $E[Y|X^*, Z^*, \epsilon, \nu] = E[Y|X^*, Z^*]$. We are grateful to the Co-Editor for pointing this out.

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

for kernel-based deconvolution methods and are satisfied by many conventional distributions; however, a notable exception is the uniform distribution, whose characteristic function is oscillatory and has infinitely many zeros. If this assumption is unlikely to hold, a possible solution involves using a ridge parameter approach (see, for example, Hall and Meister, 2007; Meister, 2009). Given f_ν^{ft} is nonvanishing everywhere and Ξ is compact, the assumption $\inf_{\xi\in\Xi}|f_\nu^{\text{ft}}(\xi_2)| > \underline{c}$ is innocuous, and is imposed here for regularity purposes. Assumption D(iii) simply states that the bandwidth must decay to zero as the sample size grows.

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 errors. For the ordinary smooth case, where the error characteristic function decays at a polynomial rate, the following assumptions are imposed.

Assumption O.

(i) f, g, f_{X^*,Z^*}, and g_{X^*,Z^*} are p-times continuously differentiable with bounded and integrable derivatives, where p is a positive integer satisfying $p > 1+2\alpha$, and satisfy $E[Y^2|f^{(l)}(X)|^2] < \infty$, $E|g^{(l)}(X)|^2 < \infty$, $E[Y^2|f^{(l,0)}_{X^*,Z^*}(X)|^2] < \infty$, and $E|g^{(l,0)}_{X^*,Z^*}(X)|^2 < \infty$ for $0 \le l \le p$, and

$$\begin{aligned}
\left|f^{(p)}(x+c) - f^{(p)}(x)\right| &< m_{1,p}(x)|c|,\\
\left|g^{(p)}(x+c) - g^{(p)}(x)\right| &< m_{2,p}(x)|c|,\\
\max_{0\le l\le p}\left|f^{(l,p-l)}_{X^*,Z^*}(x+c_x,z+c_z) - f^{(l,p-l)}_{X^*,Z^*}(x,z)\right| &< m_{3,p}(x,z)\sqrt{c_x^2+c_z^2},\\
\max_{0\le l\le p}\left|g^{(l,p-l)}_{X^*,Z^*}(x+c_x,z+c_z) - g^{(l,p-l)}_{X^*,Z^*}(x,z)\right| &< m_{4,p}(x,z)\sqrt{c_x^2+c_z^2},
\end{aligned}$$

for some bounded and integrable functions $m_{1,p}(x)$, $m_{2,p}(x)$, $m_{3,p}(x,z)$, and $m_{4,p}(x,z)$.

(ii) The characteristic function of ϵ is of the form

$$f_\epsilon^{\text{ft}}(t) = \frac{1}{c_0^{\text{os}} + c_1^{\text{os}}t + \cdots + c_\alpha^{\text{os}}t^\alpha},$$

for all $t \in \mathbb{R}$ and for some finite constants $c_0^{\text{os}}, c_1^{\text{os}}, \ldots, c_\alpha^{\text{os}}$ with $c_0^{\text{os}} = 1$ and $\alpha > 0$.

(iii) K is differentiable to order p and satisfies

$$\int K(u)du = 1, \quad \int u^{p+1}K(u)du \neq 0, \quad \int u^l K(u)du = 0, \text{ for } l = 1,2,\ldots,p.$$

In addition, K^{ft} is compactly supported on $[-1,1]$, symmetric around zero, and bounded.

(iv) $nb_n^{2p} \to 0$ and $nb_n^{2+4\alpha} \to \infty$ as $n \to \infty$.

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

(v) For $c_l^{\mathrm{os}}(\xi_1) = (-\mathrm{i})^l \sum_{h=l}^{\alpha} c_h^{\mathrm{os}} \binom{h}{l} \xi_1^{h-l}$, we have

$$E\left[\begin{array}{c} \sup_{\xi\in\Xi}\left|\sum_{l=0}^{\alpha} c_l^{\mathrm{os}}(\xi_1)\int\{g^{(l,0)}_{X^*,Z^*} - Yf^{(l,0)}_{X^*,Z^*}\}(X,z)e^{\mathrm{i}(X,z)\xi}dz\right|^2 \\ +\sup_{\xi\in\Xi}\left|\sum_{l=0}^{\alpha} c_l^{\mathrm{os}}(\xi_1)\{g^{(l)} - Yf^{(l)}\}(X)\right|^2 \end{array}\right] < \infty.$$

Assumption O(i) imposes further conditions on the existence of moments, and gives smoothness restrictions on the structural functions, which are equivalent to Assumptions 3 and 5 in Powell, Stock, and Stoker (1989).

Assumption O(ii) is an ordinary smooth condition. Specifically, it requires that $f_\epsilon^{\mathrm{ft}}(t)$ decays to zero at a polynomial rate as $|t| \to \infty$. At the cost of additional notation, this assumption can be relaxed to $f_\epsilon^{\mathrm{ft}}(t) = \exp(\mathrm{i}t\varsigma) / \left(c_0^{\mathrm{os}} + c_1^{\mathrm{os}} t + \cdots + c_\alpha^{\mathrm{os}} t^\alpha\right)$ for some real number ς, as in Fan (1995). In that paper, the author shows that this form includes typical examples of ordinary smooth densities such as Laplace and gamma distributions. Compared to the conventional ordinary smooth assumption, which only imposes bounds on the tail behavior of the error characteristic function, Assumption O(ii) is stronger in the sense that it characterizes the exact limiting behavior of $f_\epsilon^{\mathrm{ft}}(t)$ as $|t| \to \infty$, which is necessary to characterize the asymptotic distribution of our test statistics. It is also worth noting that Assumption O(ii) only concerns f_ϵ, while f_ν is left unrestricted except for the nonvanishing nature of f_ν^{ft} anywhere and $\inf_{\xi\in\Xi}|f_\nu^{\mathrm{ft}}(\xi_2)| > \underline{c}$ as required by Assumption D. This is because our test statistics are constructed asymmetrically with respect to X and Z. In particular, as Z only appears under the alternative, due to the presence of the generically totally revealing function $e^{\mathrm{i}(x,z)\xi}$ and integration with respect to z, to characterize the asymptotic distributions of the test statistics, we only need to characterize the behavior of f_ν^{ft} over a compact set. In contrast, X appears in both the null and alternative models. Thus, even after considering the presence of the generically totally revealing function $e^{\mathrm{i}(x,z)\xi}$ and integration with respect to x, the exact limiting behavior of $f_\epsilon^{\mathrm{ft}}(t)$ as $|t| \to \infty$ is still needed to characterize the asymptotic distribution of the test statistic. An implication of this insight is that if only regressors which are being tested are contaminated with error, our test would be adaptive in the sense that the asymptotic distribution of our test statistics can be characterized in a uniform way irrespective of the smoothness of the error distribution.

Assumption O(iii) requires a higher-order kernel to remove the asymptotic bias of the nonparametric estimators and is used in the proof of Lemma 3. Besides removing the estimation bias, the higher-order kernel is also used to characterize the integration involving the deconvolution kernel in Lemma 5. Higher-order kernels are common in deconvolution problems; indeed, Meister (2009) discusses how kernels of any order can be constructed quite simply.

The first condition of Assumption O(iv) is required to ensure that the bias from the nonparametric estimators is asymptotically negligible and the second condition is needed for the error of the Hájek projection to be asymptotically negligible. Note that Assumption O(iv) implicitly requires $p > 1 + 2\alpha$, i.e., the structural functions

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

have to be smooth enough to use the Hájek projection while undersmoothing, which is guaranteed by Assumption O(i).

Finally, Assumption O(v) is a high-level assumption on the boundedness of the asymptotic variance of $T_n(\xi)$; an analogous assumption is made in Fan (1995).

To derive the asymptotic distributions of CM_n and KS_n under the null hypothesis H_0 for the ordinary smooth case, we must characterize the limiting behavior of the empirical process $T_n(\xi)$.

THEOREM 1. *Suppose Assumptions D and O hold. Then, under* H_0*, we have*

$$\sqrt{n}T_n(\xi) \Rightarrow G^{os}(\xi),$$

where $\Rightarrow$ *denotes weak convergence and* $G^{os}(\xi)$ *is a Gaussian process with mean zero and covariance structure* $Var[r_\infty^{os}(X,Y,Z;\xi)]$ *for*

$$r_\infty^{os}(X,Y,Z;\xi) = \sum_{l=0}^{\alpha} c_l^{os}(\xi_1) \left[\begin{array}{l} \{Yf^{(l)} - g^{(l)}\}(X)e^{i(X,Z)\xi} f_v^{ft-1}(\xi_2) + \\ \int \{g^{(l,0)}_{X^*,Z^*} - Yf^{(l,0)}_{X^*,Z^*}\}(X,z)e^{i(X,z)\xi} dz \end{array} \right].$$

Based on Theorem 1, using the continuous mapping theorem (see, for example, van der Vaart and Wellner, 1996, Thm. 1.3.6), the asymptotic null distributions of CM_n and KS_n for the ordinary smooth case are established as follows.

COROLLARY 1. *Suppose Assumptions D and O hold. Then, under* H_0*, we have*

$$nCM_n \xrightarrow{d} \int_\Xi \left|G^{os}(\xi)\right|^2 d\xi,$$
$$\sqrt{n}KS_n \xrightarrow{d} \sup_{\xi\in\Xi} \left|G^{os}(\xi)\right|.$$

Corollary 1 shows that in the ordinary smooth case, the Cramér-von Mises statistic and Kolmogorov–Smirnov statistic converge to the squared L_2-norm and the Skorohod norm of a Gaussian process, respectively, which are characterized by the limit of the empirical process $T_n(\xi)$. For the Cramér-von Mises statistic CM_n, following Bierens and Ploberger (1997), the limiting distribution can be further characterized as an infinite weighted sum of independent chi-square random variables.

It is worth noting that despite using nonparametric deconvolution estimation, our tests achieve a $\sqrt{n}$-rate. This is, however, not the first time that a parametric rate has been obtained for functionals of nonparametric estimators in a measurement error context. Among others, Zhu and Cui (2003) and Hall and Ma (2007a) propose semiparametric estimators for regression functions which also achieve a $\sqrt{n}$-rate of convergence, and Schennach (2004) develops a $\sqrt{n}$-consistent nonlinear regression estimator which uses a nonparametric estimate of the measurement error density in its construction. These results mirror analogous findings when averaging nonparametric kernel estimators using uncontaminated data (see, for example, Powell et al., 1989; DM, among many others).

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

For the second case, known as supersmooth measurement error, where the error characteristic function decays at an exponential rate, we focus on the Gaussian distribution and impose the following assumptions.

Assumption S.

(i) f, g, f_{X^*,Z^*}, and g_{X^*,Z^*} are infinitely differentiable.
(ii) ϵ follows a Gaussian distribution with characteristic function of the form

$$f_\epsilon^{\mathrm{ft}}(t) = e^{-\mu t^2},$$

for all $t \in \mathbb{R}$ and some positive constant μ.
(iii) K is infinitely differentiable and satisfies

$$\int K(u)du = 1, \quad \int u^l K(u)du = 0, \text{ for all } l \in \mathbb{N}.$$

In addition, K^{ft} is compactly supported on $[-1, 1]$, symmetric around zero, and bounded.
(iv) $nb_n^2 e^{-16\mu b_n^{-2}} \to \infty$ as $n \to \infty$.
(v) For $c_l^{\mathrm{ss}}(\xi_1) = (-\mathrm{i})^l \sum_{h \geq l/2}^{\infty} \frac{\mu^h}{h!} \binom{2h}{l} \xi_1^{2h-l}$, we have

$$E\left[\begin{array}{c} \sup_{\xi \in \Xi} \left| \sum_{l=0}^{\infty} c_l^{\mathrm{ss}}(\xi_1) \int \{ g_{X^*,Z^*}^{(l,0)} - Y f_{X^*,Z^*}^{(l,0)} \}(X,z) e^{\mathrm{i}(X,z)\xi} dz \right|^2 \\ + \sup_{\xi \in \Xi} \left| \sum_{l=0}^{\infty} c_l^{\mathrm{ss}}(\xi_1) \{ g^{(l)} - Y f^{(l)} \}(X) \right|^2 \end{array} \right] < \infty.$$

Assumption S(i) is restrictive but appears necessary when dealing with supersmooth measurement error. In particular, it is required in Lemma 4, which shows that supersmooth error can be treated as an infinitely smooth version of ordinary smooth error. The conditions on the density functions simply require the distributions of the latent regressors to be supersmooth themselves. Indeed, Meister (2009) discusses how the class of infinitely differentiable functions still contain a "comprehensive nonparametric class of densities" (p. 44). For the conditional mean functions, there are still a broad class of functions which satisfy such a smoothness restriction, including—but not limited to—all polynomials, circular functions, exponentials, and products or sums of such functions.

Assumption S(ii) says ϵ follows a Gaussian distribution, which represents the most common example of a supersmooth distribution. This assumption is primarily required to use Lemma 6. While we acknowledge that this imposes a strong distributional assumption on ϵ, we continue to allow the distribution of v and the regression function to remain unspecified. Furthermore, at the cost of additional notation, we can relax this condition to allow ϵ to come from a convolution of a Gaussian density with an arbitrary ordinary smooth density satisfying Assumption O(ii). Note that this is not the first time the supersmooth case has been restricted in order to obtain results (for example, Van Es and Uh, 2005; Kato and Sasaki, 2018), and in many cases, the supersmooth case is omitted entirely (for example, Delaigle, Hall, and Jamshidi, 2015; Kato and Sasaki, 2019).

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

Assumption S(iii) requires K to be an infinite-order kernel, so that it can adapt to the smoothness of f, g, f_{X^*,Z^*}, and g_{X^*,Z^*}, as used in the proof of Lemma 6. A common choice of K satisfying this restriction is the flat-top kernel (McMurry and Politis, 2004).

As a result of the infinite-order kernel removing the bias in the nonparametric estimator, in contrast to Assumption O(iv) for the ordinary smooth case, Assumption S(iv) does not impose an upper bound on the bandwidth, which is usually required to control the magnitude of the estimation bias. Similar to Assumption O(iv) for the ordinary smooth case, a lower bound on the bandwidth is imposed in Assumption S(iv). In particular, Assumption S(iv) is satisfied by a bandwidth of order $\ln(n)^{-1/2}$.

Finally, Assumption S(v) is a high-level assumption on the boundedness of the asymptotic variance of $T_n(\xi)$, which is analogous to Assumption O(v) for the ordinary smooth case.

The limiting behavior of the empirical process $T_n(\xi)$ under the null hypothesis H_0 for the supersmooth case is given in the following theorem.

THEOREM 2. *Suppose Assumptions D and S hold. Then, under* H_0*, we have*

$$\sqrt{n}T_n(\xi) \Rightarrow G^{ss}(\xi),$$

where $G^{ss}(\xi)$ *is a Gaussian process with mean zero and covariance structure* $Var[r^{ss}_{\infty}(X,Y,Z;\xi)]$ *for*

$$r^{ss}_{\infty}(X,Y,Z;\xi) = \sum_{l=0}^{\infty} c_l^{ss}(\xi_1) \begin{bmatrix} \{Yf^{(l)} - g^{(l)}\}(X) e^{i(X,Z)\xi} f_v^{ft-1}(\xi_2) + \\ \int \{g^{(l,0)}_{X^*,Z^*} - Yf^{(l,0)}_{X^*,Z^*}\}(X,z) e^{i(X,z)\xi} dz \end{bmatrix}.$$

Based on Theorem 2, using the continuous mapping theorem, the asymptotic null distributions of CM_n and KS_n for the supersmooth case are established as follows.

COROLLARY 2. *Suppose Assumptions D and S hold. Then, under* H_0*, we have*

$$nCM_n \xrightarrow{d} \int_{\Xi} \left|G^{ss}(\xi)\right|^2 d\xi,$$

$$\sqrt{n}KS_n \xrightarrow{d} \sup_{\xi \in \Xi} \left|G^{ss}(\xi)\right|.$$

Corollary 2 includes parallel results for the supersmooth case to the results in Corollary 1. Fan (1995) derived an average derivative estimator using deconvolution kernel techniques and went on to prove the $\sqrt{n}$-convergence of the estimator. Our derivation of the limits of the empirical process $T_n(\xi)$ in both Theorems 1 and 2 follow a similar route as in Fan (1995) due to the link between average derivative estimators and nonsmoothing tests. However, Fan (1995) was restricted to the ordinary smooth case. The key to our result for the supersmooth case lies in

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

Lemma 4, where we show that supersmooth error can be thought of as an ordinary smooth problem with an infinite smoothness parameter.

3.2. Asymptotic Power

In this section, we analyze the power properties of the test against a sequence of local alternatives H_{1n} that converges to H_0 at the parametric rate $n^{-1/2}$. In particular, we assume a local linear alternative of the form

$$H_{1n} : E[Y|X^*, Z^*] = E[Y|X^*] + \frac{\Delta(X^*, Z^*)}{\sqrt{n}} \quad \text{a.s.},$$

where $\Delta : \mathbb{R}^2 \to \mathbb{R}$ is a bounded nonzero function. The local power properties of CM_n and KS_n under H_{1n} can be derived based on the following results.

THEOREM 3. *Under H_{1n}, if Assumptions D and O hold, we have*

$$\sqrt{n}T_n(\xi) \Rightarrow G^{\text{os}}(\xi) + \mu_\Delta(\xi),$$

and if Assumptions D and S hold, we have

$$\sqrt{n}T_n(\xi) \Rightarrow G^{\text{ss}}(\xi) + \mu_\Delta(\xi),$$

where $G^{\text{os}}(\xi)$ and $G^{ss}(\xi)$ are the same Gaussian processes as defined in Theorems 1 and 2, respectively, and $\mu_\Delta(\xi)$ is a deterministic shift function given by

$$\mu_\Delta(\xi) = E[\Delta(X^*, Z^*) f(X^*) e^{i(X^*, Z^*)\xi}].$$

Theorem 3 shows that under both ordinary smooth and supersmooth measurement error, the asymptotic behavior of $T_n(\xi)$ under H_{1n} can be characterized by the same Gaussian process used to characterize the asymptotic behavior of $T_n(\xi)$ under H_0 plus a deterministic shift function $\mu_\Delta(\xi)$, which essentially characterizes the extra bias brought by H_{1n}. Using the continuous mapping theorem, tests based on a continuous functional of $T_n(\xi)$ will have nontrivial power against local alternatives converging to the null at the rate $n^{-1/2}$.

4. COMPUTATION OF CRITICAL VALUES

The asymptotic distributions derived in Theorems 1 and 2 can be used to obtain critical values. However, as explained in Bierens and Ploberger (1997), the limiting distributions 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, we implement a bootstrap procedure.

Measurement error models provide 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 is possible to

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

follow an approach similar to Hall and Ma (2007b): estimating the density of the true regressor and measurement error using deconvolution techniques, applying a wild bootstrap approach for the regression error, and sampling from these respective densities. However, the estimated densities suffer from the slow rates of convergence associated with deconvolution estimation, and the approach is computationally expensive. In addition, the choice of several tuning parameters is needed. Instead, we consider a simple empirical bootstrap approach.

We focus on the bootstrap for the KS_n test statistic, but very similar results can be obtained for the CM_n statistic. A bootstrap sample $\{X_j^\#, Y_j^\#, Z_j^\#\}_{j=1}^n$ is constructed by independently drawing with replacement from the set of observed $\{X_j, Y_j, Z_j\}_{j=1}^n$, and the bootstrap statistic is calculated as $\sqrt{n}KS_n^\# = \sup_{\xi\in\Xi}\left|\sqrt{n}\left\{T_n^\#(\xi) - E\left[T_n^\#(\xi)\right]\right\}\right|$, where

$$T_n^\#(\xi) = \frac{1}{n^2 b_n^3}\sum_{j\neq k}^{n}(Y_k^\# - Y_j^\)\iint \mathbb{K}_\epsilon\left(\frac{x - X_j^\#}{b_n}\right)\mathbb{K}_\epsilon\left(\frac{x - X_k^\#}{b_n}\right)\mathbb{K}_\nu \times\left(\frac{z - Z_k^\#}{b_n}\right)e^{\mathrm{i}(x,z)\xi}dxdz.$$

To analyze the properties of this procedure, we strengthen Assumptions O(v) and S(v) as follows.

Assumption B. $E[\int_\Xi r_\infty^{os}(X,Y,Z;\xi)^4 d\xi] < \infty$ holds under Assumption O or $E[\int_\Xi r_\infty^{ss}(X,Y,Z;\xi)^4 d\xi] < \infty$ holds under Assumption S.

Assumption B places bounds on the fourth moment of the dominant term from the Hájek projection of our empirical process. This assumption is needed to apply Chernozhukov et al. (2016, Thms. 2.1 and 2.3).

Define $\hat{c}_\alpha$ as the $(1-\alpha)$th quantile of $KS_n^\#$. The validity of the proposed empirical bootstrap procedure is then given in the following theorem.

THEOREM 4. *Suppose Assumptions D, O/S, and B hold, then under* H_0,

$$P\{KS_n \leq \hat{c}_\alpha\} \geq 1 - \alpha - o(1).$$

Theorem 4 provides the asymptotic validity of the empirical bootstrap for our Kolmogorov–Smirnov statistic and shows that quantiles of the bootstrap statistic distribution conditional on the data can be used as critical values for our test. Furthermore, under H_{1n}, since $|\mu_\Delta(\xi)| > 0$, for all ξ, and $\hat{c}_\alpha$ does not depend on the validity of H_0, the power of the test using $\hat{c}_\alpha$ as critical values is guaranteed. Similar results for the empirical bootstrap in a measurement error setting were found by Kato and Sasaki (2018, 2019) in the construction of confidence bands for density estimation and nonparametric regression, respectively, and by Adusumilli et al. (2020) for confidence bands for the cumulative distribution function.

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

5. CASE OF UNKNOWN MEASUREMENT ERROR DISTRIBUTION

In this section, we discuss how to relax the known error distribution assumption. Suppose that, in addition to X and Z, we also observe X^r and Z^r, which are generated by

$$X^r = X^* + \epsilon^r, \tag{5.1}$$

$$Z^r = Z^* + \nu^r, \tag{5.2}$$

where (ϵ^r, ν^r) are i.i.d. as (ϵ, ν). Furthermore, we assume the following.

Assumption R.

(i) $\{X_j^r, Z_j^r\}_{j=1}^n$ is an i.i.d. sample of (X^r, Z^r) satisfying (5.1) and (5.2).
(ii) (ϵ^r, ν^r) are i.i.d. as (ϵ, ν), $E|\epsilon|^{2+\varsigma_1} < \infty$ and $E|\nu|^{2+\varsigma_2} < \infty$ for some positive constants ς_1 and ς_2, and f_ϵ^{ft} and f_ν^{ft} are real-valued, i.e., f_ϵ and f_ν are symmetric around zero.

Assumption R(ii) requires that (ϵ^r, ν^r) is an independent copy of the symmetrically distributed (ϵ, ν), which is conventional in the deconvolution literature when the error density is unknown (see, for example, Delaigle et al., 2008). It should be noted that alternative estimation schemes for f_ϵ^{ft} and f_ν^{ft} exist that rely on weaker assumptions than have been made here. In particular, Li and Vuong (1998) use a lemma by Kotlarski (1967) to construct an estimator of an error characteristic function which relaxes the symmetry assumption but still requires independent repeated measurements. While Delaigle and Hall (2016) develop a clever method which does not require repeated measurements; their work requires symmetry in the measurement error and asymmetry in the true regressor. A testing procedure based on such estimators can be constructed in an analogous fashion as outlined above; however, the analysis of the properties of such tests is beyond the scope of this paper.

Under Assumption R, f_ϵ^{ft} and f_ν^{ft} can be estimated by (Delaigle et al., 2008)

$$\hat{f}_\epsilon^{\mathrm{ft}}(t) = \left| \frac{1}{n} \sum_{j=1}^n \cos\{t(X_j - X_j^r)\} \right|^{1/2}, \quad \hat{f}_\nu^{\mathrm{ft}}(t) = \left| \frac{1}{n} \sum_{j=1}^n \cos\{t(Z_j - Z_j^r)\} \right|^{1/2}.$$

Let $\hat{\mathbb{K}}_\epsilon$ and $\hat{\mathbb{K}}_\nu$ separately denote the deconvolution kernels based on the estimated error characteristic functions $\hat{f}_\epsilon^{\mathrm{ft}}$ and $\hat{f}_\nu^{\mathrm{ft}}$, and are defined as

$$\hat{\mathbb{K}}_\epsilon(u) = \frac{1}{2\pi} \int e^{-\mathrm{i}tu} \frac{K^{\mathrm{ft}}(t)}{\hat{f}_\epsilon^{\mathrm{ft}}(t/b_n)} dt, \quad \hat{\mathbb{K}}_\nu(u) = \frac{1}{2\pi} \int e^{-\mathrm{i}tu} \frac{K^{\mathrm{ft}}(t)}{\hat{f}_\nu^{\mathrm{ft}}(t/b_n)} dt.$$

By replacing $\mathbb{K}_\epsilon$ and $\mathbb{K}_\nu$ by $\hat{\mathbb{K}}_\epsilon$ and $\hat{\mathbb{K}}_\nu$, respectively, in the case when the error distribution is unknown, $T(\xi)$ can be estimated by

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

$$\hat{T}_n(\xi) = \frac{1}{n^2 b_n^3} \sum_{j \neq k}^{n} (Y_k - Y_j) \iint \hat{\mathbb{K}}_\epsilon \left(\frac{x - X_j}{b_n} \right) \hat{\mathbb{K}}_\epsilon \left(\frac{x - X_k}{b_n} \right) \hat{\mathbb{K}}_\nu \times \left(\frac{z - Z_k}{b_n} \right) e^{\mathrm{i}(x,z)\xi} dx dz,$$

and the test statistics are constructed as

$$\widehat{CM}_n = \int_\Xi |\hat{T}_n(\xi)|^2 d\xi, \qquad \widehat{KS}_n = \sup_{\xi \in \Xi} |\hat{T}_n(\xi)|.$$

To derive the asymptotic distributions of $\widehat{CM}_n$ and $\widehat{KS}_n$, as seen in Theorems 1–3, it is suffice to characterize the asymptotic behavior of $\hat{T}_n(\xi)$. Let

$$\hat{r}_\infty^{\epsilon,\nu}(X, X^r, Y, Z, Z^r; \xi) = \frac{1}{4\pi} \int \left\{ \begin{array}{l} \left\{ 1 - \frac{\cos(t(X - X^r))}{|f_\epsilon^{\mathrm{ft}}(t)|^2} \right\} \left\{ \begin{array}{l} f^{\mathrm{ft}}(t) g_{X^*,Z^*}^{\mathrm{ft}}(\xi_1 - t, \xi_2) - g^{\mathrm{ft}}(t) f_{X^*,Z^*}^{\mathrm{ft}}(\xi_1 - t, \xi_2) \\ + f^{\mathrm{ft}}(\xi_1 - t) g_{X^*,Z^*}^{\mathrm{ft}}(t, \xi_2) - g^{\mathrm{ft}}(\xi_1 - t) f_{X^*,Z^*}^{\mathrm{ft}}(t, \xi_2) \end{array} \right\} \\ + \left\{ 1 - \frac{\cos(t(Z - Z^r))}{|f_\nu^{\mathrm{ft}}(\xi_2)|^2} \right\} \left\{ f^{\mathrm{ft}}(t) g_{X^*,Z^*}^{\mathrm{ft}}(\xi_1 - t, \xi_2) - g^{\mathrm{ft}}(t) f_{X^*,Z^*}^{\mathrm{ft}}(\xi_1 - t, \xi_2) \right\} \end{array} \right\} dt.$$

In the ordinary smooth case, to characterize the asymptotic behavior of $\hat{T}_n(\xi)$, we further impose the following assumptions in addition to Assumption O.

Assumption O'.

(i) $n b_n^{2+12\alpha} \log(1/b_n)^{-4} \to \infty$ as $n \to \infty$.
(ii) $E[\int_\Xi \hat{r}_\infty^{\mathrm{os}}(X, X^r, Y, Z, Z^r; \xi)^2 d\xi] < \infty$, where

$$\hat{r}_\infty^{\mathrm{os}}(X, X^r, Y, Z, Z^r; \xi) = r_\infty^{\mathrm{os}}(X, Y, Z; \xi) + \hat{r}_\infty^{\epsilon,\nu}(X, X^r, Y, Z, Z^r; \xi).$$

Assumption O'(i) is a stronger bandwidth condition than Assumption O(iv), and is required to ensure that the higher-order terms in the error brought by the estimation of the error characteristic functions are asymptotically negligible. Assumption O'(ii) is a high-level assumption on the boundedness of the asymptotic variance of $\hat{T}_n(\xi)$, as Assumption O(v) for $T_n(\xi)$. In particular, the first term $r_\infty^{\mathrm{os}}(X, Y, Z; \xi)$ characterizes the estimation error when the error distribution is known, and the additional term $\hat{r}_\infty^{\epsilon,\nu}(X, X^r, Y, Z, Z^r; \xi)$ characterizes the main terms in the error brought by the estimation of the error characteristic functions. The following theorem summarizes the asymptotic behavior of $\hat{T}_n(\xi)$ for the ordinary smooth case under the null hypothesis H_0 and under the alternative hypothesis H_{1n}.

THEOREM 5. *Suppose Assumptions D, O, R, and O' hold.*

(i) *Under* H_0*, we have*

$$\sqrt{n} \hat{T}_n(\xi) \Rightarrow \hat{G}^{\mathrm{os}}(\xi),$$

where $\hat{G}^{\mathrm{os}}(\xi)$ *denotes a Gaussian process with mean zero and covariance structure* $Var[\hat{r}_\infty^{\mathrm{os}}(X, X^r, Y, Z, Z^r; \xi)]$.

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

(ii) *Under* $\mathrm{H}_{1,n}$*, we have*

$$\sqrt{n}\hat{T}_n(\xi) \Rightarrow \hat{G}^{\mathrm{os}}(\xi) + \mu_\Delta(\xi),$$

where $\mu_\Delta(\xi)$ *is the same deterministic shift function as defined in Theorem 3.*

Theorem 5 shows that in the ordinary smooth case when the error distribution is unknown, $\hat{T}_n(\xi)$ converges at a $\sqrt{n}$-rate to a Gaussian process $\hat{G}^{\mathrm{os}}(\xi)$, which, by the continuous mapping theorem, implies that the Cramér-von Mises statistic and Kolmogorov–Smirnov statistic converge to the squared L_2-norm and sup-norm of the Gaussian process $\hat{G}^{\mathrm{os}}(\xi)$, respectively. The Gaussian process $\hat{G}^{\mathrm{os}}(\xi)$ here, however, is different from $G^{\mathrm{os}}(\xi)$, the limit of the empirical process $T_n(\xi)$ in the case when the error distribution is known, as the estimation error of the error characteristic functions is non-negligible. In contrast, the bias term does not depend on the error distributions, so the deterministic shift function $\mu_\Delta(\xi)$ is the same as that obtained in the case when the error distribution is known.

For the supersmooth case, the following assumptions are needed in addition to Assumption S to characterize the asymptotic behavior of $\hat{T}_n(\xi)$.

Assumption S'.

(i) $nb_n^2 e^{-48\mu b_n^{-2}} \log(1/b_n)^{-4} \to \infty$ as $n \to \infty$.
(ii) $E[\int_\Xi \hat{r}_\infty^{\mathrm{os}}(X, X^r, Y, Z, Z^r; \xi)^2 d\xi] < \infty$, where

$$\hat{r}_\infty^{\mathrm{ss}}(X, X^r, Y, Z, Z^r; \xi) = r_\infty^{\mathrm{ss}}(X, Y, Z; \xi) + \hat{r}_\infty^{\epsilon,\nu}(X, X^r, Y, Z, Z^r; \xi).$$

Similar to the ordinary smooth case, Assumption S'(i) requires a stronger bandwidth condition than Assumption S(iv) to ensure that the higher-order terms in the error brought by the estimation of the error characteristic functions are asymptotically negligible, and Assumption S'(ii) requires the boundedness of the asymptotic variance of $\hat{T}_n(\xi)$. The asymptotic behavior of $\hat{T}_n(\xi)$ for the supersmooth case under the null hypothesis H_0 and under the alternative hypothesis H_{1n} are summarized in the following theorem.

THEOREM 6. *Suppose Assumptions D, S, R, and S' hold.*

(i) *Under* H_0*, we have*

$$\sqrt{n}\hat{T}_n(\xi) \Rightarrow \hat{G}^{\mathrm{ss}}(\xi),$$

where $\hat{G}^{\mathrm{os}}(\xi)$ *denotes a Gaussian process with mean zero and covariance structure* $Var[\hat{r}_\infty^{ss}(X, X^r, Y, Z, Z^r; \xi)]$.
(ii) *Under* $\mathrm{H}_{1,n}$*, we have*

$$\sqrt{n}\hat{T}_n(\xi) \Rightarrow \hat{G}^{\mathrm{ss}}(\xi) + \mu_\Delta(\xi),$$

where $\mu_\Delta(\xi)$ *is the same deterministic shift function as defined in Theorem 3.*

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

Theorem 6 includes parallel results for the supersmooth case to the results in Theorem 5 for the ordinary smooth case. Dong, Otsu, and Taylor (2020) showed that if the error density is symmetric around zero, the asymptotic distribution of the average derivative estimator is the same irrespective of whether the error density is estimated or not. Thus, given the link between average derivative estimators and nonsmoothing tests, the conclusions of Theorems 5 and 6 seem somewhat surprising. The reason for the discrepancy between the two results resides in the asymmetric nature of $T(\xi)$, which is not present in the average derivative. In particular, the asymmetry is twofold. First, $T(\xi)$ is based on ff_{X^*,Z^*}, while the average derivative is based on ff'. Second, $T(\xi)$ involves a revealing function $e^{\mathrm{i}(x,z)\xi}$ for an arbitrary $\xi \in \Xi$, while the average derivative restricts $\xi = (0,0)$. Due to the asymmetry of $T(\xi)$, even when the error density is symmetric around zero, the estimation error of the error characteristic functions is still non-negligible in our test.

Given the non-negligibility of the estimation error of the error characteristic functions, the empirical bootstrap proposed in Section 4 will be inconsistent; this is in contrast to the results of Kato and Sasaki (2018, 2019) and Adusumilli et al. (2020). Since the measurement errors are unobservable, it is not possible to account for their effect using an empirical bootstrap. However, because the estimator for the error characteristic function is constructed using a sample mean, it is possible to construct a multiplier bootstrap version of each estimator, i.e.,

$$\hat{f}_{\epsilon}^{\mathrm{ft}\#}(t) = \left|\frac{1}{n}\sum_{j=1}^{n}\eta_{1,j}\cos\{t(X_j - X_j^r)\}\right|^{1/2}, \quad \hat{f}_{\nu}^{\mathrm{ft}\#}(t) = \left|\frac{1}{n}\sum_{j=1}^{n}\eta_{2,j}\cos\{t(Z_j - Z_j^r)\}\right|^{1/2},$$

where η_1 and η_2 are i.i.d. random variables with unit mean and unit variance. These bootstrap estimates can be plugged into $T_n^{\#}(\xi)$ to give a bootstrap statistic for $\hat{T}_n(\xi)$, denoted as $\hat{T}_n^{\#}(\xi)$. However, it is beyond the scope of this paper to prove the consistency of this bootstrap procedure.

6. SIMULATION

Since this is the first nonparametric significance test able 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 DM 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, and its square, are used as the instruments.

The true, unobservable regressors (X^*, Z^*) are each distributed as $N(0,1)$ with a correlation of 0.4. The contaminated regressors are given by $X = X^* + \epsilon$ and $Z = Z^* + \nu$. In order to fairly compare our test with the available alternatives, we do not assume we know $f_{\epsilon}^{\mathrm{ft}}$ or f_{ν}^{ft}.[4] Instead, we assume a repeated measurement of each contaminated regressor is available, given by $X^r = X^* + \epsilon^r$ and $Z^r = Z^* + \nu^r$, where ϵ^r is distributed independently and identically to ϵ (as are ν^r and ν).

[4]The results are slightly better when $f_{\epsilon}^{\mathrm{ft}}$ and f_{ν}^{ft} are assumed known; however, the difference is small.

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

For the ordinary smooth case, both ϵ and ν are drawn independently from the Laplace distribution with a standard deviation of $\frac{1}{3}$, giving a signal-to-noise ratio of 3. For the supersmooth case, we use a zero-mean Gaussian error with standard deviation also equal to $\frac{1}{3}$. Since both distributions are symmetric around zero, we can use the repeated data to estimate the error characteristic functions as outlined in Section 5.

The infinite-order flat-top kernel of McMurry and Politis (2004) is used throughout all simulations. This kernel is defined by its Fourier transform

$$K^{\text{ft}}(t) = \begin{cases} 1 & \text{if } |t| \leq 0.05, \\ \exp\left\{\frac{-\exp(-(|t|-0.05)^2)}{(|t|-1)^2}\right\} & \text{if } 0.05 < |t| < 1, \\ 0 & \text{if } |t| \geq 1, \end{cases}$$

and satisfies Assumptions O(iii) and S(iii). Results are reported for two sample sizes $n = \{250, 500\}$.

Unfortunately, there are currently no theoretical results on the optimal choice of bandwidth for testing in the presence of measurement error. It is beyond the scope of this paper to investigate such issues. However, we suggest using the procedure put forth by Bissantz et al. (2007) to select the optimal bandwidths for each individual deconvolution kernel estimator, i.e., for $\tilde{r}, \tilde{f}$, and $\tilde{h}$. The bandwidth selection procedure involves two steps. First, an oversmoothing pilot bandwidth is selected, denoted b_n^0. It is crucial that the initial bandwidth is larger than the optimal choice; however, beyond this, the procedure is relatively insensitive to this initial selection. The plug-in bandwidth of Delaigle and Gijbels (2004) is used in these simulations, but, as suggested in Bissantz et al. (2007), this choice is multiplied by two to ensure it is oversmoothing. Second, a set of potential bandwidths is created, $b_{n,j} = b_n^0(j/J)$, for $j = 1, \ldots, J$. Denote by $\tilde{f}_j$ the estimator $\tilde{f}$ which uses the jth bandwidth in this set (an analogous procedure is used for $\tilde{r}$ and $\tilde{h}$). The selected bandwidth is the largest bandwidth $b_{n,j}$ such that $d_{j-1,j}^{(\infty)} \equiv ||\tilde{f}_{j-1} - \tilde{f}_j||_\infty$ is larger than $\rho d_{J-1,J}^{(\infty)}$, for some $\rho > 1$. To select J and ρ, we follow the suggestion of Kato and Sasaki (2018) and choose $J = 4\log(n)$ and $\rho = 0.4\log(n)$.

For the test of DM, a rule-of-thumb bandwidth, $b_{\text{dm}} = n^{-1/3}$, is used; this is taken from the simulations carried out in DM. Results are also reported for half of this value and double this value.

The critical values for our test are constructed using the bootstrap procedure outlined at the end of Section 5 with 499 replications. For DM, the bootstrap procedure denoted C_n^{**} in their paper is used. The perturbation random variable for their bootstrap is the Mammen two-point distribution. All results are based on 1,000 Monte Carlo replications.

Table 1 shows results for the level accuracy of the three tests. The columns labeled "CM" and "KS" report the proportion of rejections for the Cramér-von Mises test and the Kolmogorov–Smirnov test proposed in this paper, respectively. The column labeled "DM" refers to the test of Delgado and González-Manteiga (2001), and "IV" displays results for the Wald test based on the IV quadratic

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

TABLE 1. $Y = 1 + X + X^2 + U$

n	Level	Ordinary smooth						Supersmooth					
		CM	KS	IV	DM			CM	KS	IV	DM		
					$0.5b_{\text{dm}}$	b_{dm}	$2b_{\text{dm}}$				$0.5b_{\text{dm}}$	b_{dm}	$2b_{\text{dm}}$
250	**5%**	6.0	6.7	8.1	15.1	14.5	25.1	6.7	6.7	10.1	15.3	18.8	24.9
	10%	13.2	13.5	12.1	25.4	26.0	42.3	13.1	13.4	15.1	25.2	30.1	39.7
500	**5%**	6.3	6.8	11.2	22.1	26.5	30.1	6.0	6.4	8.8	28.5	29.0	36.7
	10%	12.3	13.5	18.0	35.2	40.7	47.8	12.2	13.1	15.5	41.8	43.2	51.6

TABLE 2. $Y = 1 + X + X^2 + \frac{2}{\sqrt{n}}Z^2 + U$

n	Level	Ordinary smooth						Supersmooth					
		CM	KS	IV	DM			CM	KS	IV	DM		
					$0.5b_{\text{dm}}$	b_{dm}	$2b_{\text{dm}}$				$0.5b_{\text{dm}}$	b_{dm}	$2b_{\text{dm}}$
250	**5%**	34.2	29.5	**53.9**	14.4	13.0	12.1	28.3	26.8	**52.8**	16.1	16.6	22.2
	10%	46.6	41.7	**66.4**	30.6	30.7	29.0	45.8	41.5	**67.5**	28.4	29.7	30.9
500	**5%**	30.3	23.9	**44.3**	14.2	16.9	18.4	32.4	28.5	**50.5**	18.7	20.1	23.9
	10%	47.6	35.9	**58.7**	28.6	29.0	30.2	52.2	41.3	**66.2**	28.8	29.9	34.3

TABLE 3. $Y = 1 + X + X^2 + \frac{2}{\sqrt{n}}\sin(Z) + U$

n	Level	Ordinary smooth						Supersmooth					
		CM	KS	IV	DM			CM	KS	IV	DM		
					$0.5b_{\text{dm}}$	b_{dm}	$2b_{\text{dm}}$				$0.5b_{\text{dm}}$	b_{dm}	$2b_{\text{dm}}$
250	**5%**	**41.0**	39.9	7.5	24.2	25.7	25.5	**36.9**	36.1	5.5	25.0	29.2	32.4
	10%	51.8	**52.2**	13.3	38.2	40.1	42.2	**51.5**	48.6	10.4	37.1	38.3	38.6
500	**5%**	32.8	**37.5**	6.7	23.2	26.7	27.7	39.1	**41.8**	6.2	30.0	32.5	33.4
	10%	46.2	**49.4**	11.2	39.8	40.2	40.4	**52.7**	52.4	10.5	40.0	41.6	43.8

regression. Tables 2–4 display the power results for three different alternatives, with the most powerful method in each case highlighted in bold. Note that the power results are empirically adjusted to give a fair comparison of the three methods.[5]

The results are encouraging and appear to reflect the theoretical findings. The bootstrap procedure keeps the size of the Cramér-von Mises test and the

[5]Empirically adjusted power results are obtained in the following way. Using the null model from Table 1, for each parameter setting and for each nominal level, we first obtain the quantile at which the empirical rejection rate is equal to the nominal level, say the $\hat{Q}_{\alpha}^{th}$ quantile for an α nominal level. In all subsequent models, the critical value is taken as the $\hat{Q}_{\alpha}^{th}$ quantile from the bootstrap distribution, rather than the typical $(1-\alpha)$ quantile.

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

TABLE 4. $Y = 1 + X + X^2 + \frac{5}{\sqrt{n}} X \sin(Z) + U$

n	Level	Ordinary smooth						Supersmooth					
		CM	KS	IV	DM			CM	KS	IV	DM		
					$0.5b_{\text{dm}}$	b_{dm}	$2b_{\text{dm}}$				$0.5b_{\text{dm}}$	b_{dm}	$2b_{\text{dm}}$
250	**5%**	34.6	**35.6**	6.3	20.8	22.5	20.5	30.1	**30.9**	5.0	20.4	25.5	26.8
	10%	43.8	**46.2**	13.0	35.6	38.0	38.7	**44.5**	43.3	11.8	33.6	34.3	32.4
500	**5%**	27.8	**40.6**	5.0	21.8	22.2	24.4	32.7	**39.0**	5.1	25.4	26.6	30.2
	10%	33.4	**46.5**	11.3	38.2	35.1	36.3	46.5	**50.5**	10.0	34.4	35.2	40.8

Kolmogorov–Smirnov test reasonably close to the nominal value for both types of measurement error but with a general over-rejection. The Wald test based on the IV also shows an over-rejection in all cases which is slightly more severe than our tests. In contrast, the DM test displays severe size distortions which are exacerbated when using a larger bandwidth; the test appears to be almost unusable in some cases. However, we should keep in mind that this test is not designed to deal with situations of measurement error.

Table 2 displays power results for a quadratic alternative. In this case, the Wald test is conducted on a correctly specified model and acts as a benchmark with which to compare the other tests. As expected, the Wald test dominates the tests of this paper and that of DM. However, it is promising to see that, at least at the 10% level, the power of our tests do not fall far behind that of the Wald test, and in all cases dominate the test of DM. Given that the alternative is drifting at the rate $n^{-1/2}$, it is also encouraging to see that the power of our tests show no substantial decline—and even increase in some situations. Interestingly, there appears to be little difference in power in the ordinary smooth and supersmooth settings for our tests.

In Tables 3 and 4, a sinusoidal and an interactive sinusoidal alternative are considered, respectively. The results are broadly similar in each case. It seems that neither the Cramér-von Mises test nor the Kolmogorov–Smirnov test dominates the other; the Kolmogorov–Smirnov test is preferable with an interactive sinusoidal alternative, while the Cramér-von Mises test has better power against a quadratic alternative. The Wald test struggles to detect the higher nonlinearity in Tables 3 and 4, reflecting the arguments given in Section 2, while our tests and the DM test perform well in these settings. However, in all cases, our tests show the highest power.

7. CONCLUSION

This paper develops, to the best of our knowledge, the first nonparametric significance test for regression models with classical measurement error in the regressors. The test is flexible in that the measurement error can enter the model through

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

the regressors of interest, through the controlling variables, or through both. Furthermore, we allow for the characteristic function of the error distribution to be estimated. We derive the asymptotic properties of the test and show it can detect local alternatives at the $\sqrt{n}$-rate for both ordinary smooth and Gaussian measurement errors. As a cost of this $\sqrt{n}$-rate, the asymptotic distribution is case-dependent and difficult to estimate in practice. However, a consistent bootstrap procedure is provided to obtain critical values that is shown to work well in finite samples. Although a practical approach to the choice of bandwidth is given, the problem of choosing an optimal bandwidth for testing in the presence of measurement error is still an open issue that needs to be resolved.

APPENDICES

A. Proofs of Theorems

A.1. *Proof of Theorem 1.* Define $d_j = (X_j, Y_j, Z_j)$ and $r_n(d_j;\xi) = E[p_n(d_j,d_k;\xi)|d_j]$ with

$$q_n(d_j,d_k;\xi) = \frac{1}{2b_n^3}\iint (Y_k - Y_j)\mathbb{K}_\epsilon\left(\frac{x-X_j}{b_n}\right)\mathbb{K}_\epsilon\left(\frac{x-X_k}{b_n}\right)\mathbb{K}_\nu\left(\frac{z-Z_k}{b_n}\right)e^{\mathrm{i}(x,z)\xi}\,dxdz.$$

Then, we can write $T_n(\xi) = \frac{(n-1)}{n}U_n(\xi)$, where $U_n(\xi) = \binom{n}{2}^{-1}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n} p_n(d_j,d_k;\xi)$, for $\xi = (\xi_1,\xi_2)$, is a second-order U-statistic with symmetric kernel $p_n(d_j,d_k;\xi) = q_n(d_j,d_k;\xi)+q_n(d_k,d_j;\xi)$ and Hájek projection $U_n^*(\xi) = E[r_n(d_1;\xi)]+\frac{2}{n}\sum_{j=1}^{n}\{r_n(d_j;\xi)-E[r_n(d_j;\xi)]\}$. By Lemma A.3 of Ahn and Powell (1993), we have

$$\sup_{\xi\in\Xi}|U_n(\xi)-U_n^*(\xi)| = o_p\left(n^{-1/2}\right), \tag{A.1}$$

if

$$E\left[\sup_{\xi\in\Xi} p_n(d_j,d_k;\xi)^2\right] = o(n). \tag{A.2}$$

To show (A.2), note that

$$\begin{aligned} &E\left[\sup_{\xi\in\Xi} p_n(d_j,d_k;\xi)^2\right] \\ &\le \frac{2}{b_n^6}E\left[\begin{array}{c}\sup_{\xi\in\Xi}\left|\iint \mathbb{K}_\epsilon\left(\frac{x-X_j}{b_n}\right)\mathbb{K}_\epsilon\left(\frac{x-X_k}{b_n}\right)\mathbb{K}_\nu\left(\frac{z-Z_k}{b_n}\right)e^{\mathrm{i}(x,z)\xi}\,dxdz\right|^2 \\ \times\left\{E[Y_j^2|X_j^*]+E[Y_k^2|X_k^*,Z_k^*]\right\}\end{array}\right] \end{aligned}$$

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

$$
\begin{aligned}
&= \frac{2}{b_n^2} E\left[\begin{array}{c} \sup_{\xi\in\Xi} \left| \iint \mathbb{K}_\epsilon\left(\tilde{x} + \frac{X_k - X_j}{b_n}\right) \mathbb{K}_\nu(\tilde{z})\, e^{\mathrm{i}(X_k + b_n\tilde{x}, Z_k + b_n\tilde{z})\xi} d\tilde{x} d\tilde{z} \right|^2 \\ \times \left\{ E[Y_j^2 | X_j^*] + E[Y_k^2 | X_k^*, Z_k^*] \right\} \end{array} \right] \\
&\leq \frac{1}{2\pi^2 b_n^2} E\left[\begin{array}{c} \left\{ \int \frac{|K^{\mathrm{ft}}(t)|}{|f_\epsilon^{\mathrm{ft}}(t/b_n)|} \left(\sup_{\xi\in\Xi} \frac{|K^{\mathrm{ft}}(-t + b_n\xi_1)||K^{\mathrm{ft}}(b_n\xi_2)|}{|f_\epsilon^{\mathrm{ft}}(-t/b_n + \xi_1)||f_\nu^{\mathrm{ft}}(\xi_2)|} \right) dt \right\}^2 \\ \times \left\{ E[Y_j^2 | X_j^*] + E[Y_k^2 | X_k^*, Z_k^*] \right\} \end{array} \right] \\
&= O\left(b_n^{-2} \left(\inf_{|t| \leq 2b_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)| \right)^{-4} \right), \qquad \textbf{(A.3)}
\end{aligned}
$$

where the first step follows by Jensen's inequality, the law of iterated expectations, and the independence between (ϵ, ν) and Y as imposed in Assumption D(ii), the second step follows by the change of variables $\tilde{x} = \frac{x - X_k}{b_n}$ and $\tilde{z} = \frac{z - Z_k}{b_n}$, the third step uses Lemma 1, and the last step follows from $E[Y^2] < \infty$ as assumed in Assumption D(i), $\inf_{\xi\in\Xi} |f_\nu^{\mathrm{ft}}(\xi_2)| > \underline{c}$ as assumed in Assumption D(ii), and the fact that K^{ft} is supported on $[-1, 1]$ and is bounded as assumed in Assumption O(iii). Then, (A.2) follows by (A.3) and Assumption O(ii) and (iv).

For the bias term of $U_n^*(\xi)$, under H_0, we claim that it is asymptotically negligible, i.e.,

$$
\sup_{\xi\in\Xi} |E[r_n(d_1; \xi)]| = o\left(n^{-1/2}\right). \qquad \textbf{(A.4)}
$$

To show (A.4), using Lemma 2, we write $E[r_n(d_1; \xi)]$ as

$$
E[r_n(d_1; \xi)] = \frac{1}{b_n^3} \iint \left\{ \begin{array}{c} E\left[K\left(\frac{x - X^*}{b_n}\right)\right] E\left[YK\left(\frac{x - X^*}{b_n}\right) K\left(\frac{z - Z^*}{b_n}\right)\right] \\ -E\left[YK\left(\frac{x - X^*}{b_n}\right)\right] E\left[K\left(\frac{x - X^*}{b_n}\right) K\left(\frac{z - Z^*}{b_n}\right)\right] \end{array} \right\} e^{\mathrm{i}(x,z)\xi} dx dz.
$$

Then, (A.4) follows by Lemma 3. In particular, $\sup_{\xi\in\Xi} |E[r_n(d_1; \xi)]| = O(b_n^p) = o\left(n^{-1/2}\right)$ under Assumption O.

For the main term of $U_n^*(\xi)$, we decompose $2r_n(d_j; \xi) = r_{1,n}(d_j; \xi) + r_{2,n}(d_j; \xi)$, where

$$
\begin{aligned}
r_{1,n}(d_j; \xi) &= \iiint \left\{ \begin{array}{c} \mathbb{K}_\epsilon(\tilde{x}) \mathbb{K}_\nu(\tilde{z}) K(x^*) e^{\mathrm{i}(X_j + b_n\tilde{x}, Z_j + b_n\tilde{z})\xi} \\ \times \{Y_j f - g\}(X_j + b_n(\tilde{x} - x^*)) \end{array} \right\} dx^* d\tilde{x} d\tilde{z}, \\
r_{2,n}(d_j; \xi) &= \iiint \left\{ \begin{array}{c} \mathbb{K}_\epsilon(\tilde{x}) K(x^*) K(z^*) e^{\mathrm{i}(X_j + b_n\tilde{x}, z)\xi} \\ \times \{g_{X^*,Z^*} - Y_j f_{X^*,Z^*}\}(X_j + b_n(\tilde{x} - x^*), z - b_n z^*) \end{array} \right\} dx^* dz^* d\tilde{x} dz,
\end{aligned}
$$

which are obtained using Lemma 2 and the change of variables $\tilde{x} = \frac{x - X_j}{b_n}$ and $\tilde{z} = \frac{z - Z_j}{b_n}$.

Since f, g, f_{X^*,Z^*}, and g_{X^*,Z^*} are p-times differentiable with $p > 1 + 2\alpha$ under Assumption O, by expanding $\{Y_j f - g\}(X_j + b_n(\tilde{x} - x^*))$ around X_j to the pth order, we have $r_{1,n}(d_j; \xi) = r_{1,n}^*(d_j; \xi) + t_{1,n}(d_j; \xi)$, where

$$
r_{1,n}^*(d_j; \xi) = \sum_{l=0}^{p} \frac{b_n^l K^{\mathrm{ft}}(b_n\xi_2)}{l! f_\nu^{\mathrm{ft}}(\xi_2)} \int \mathbb{K}_\epsilon(\tilde{x}) \tilde{x}^l e^{\mathrm{i}b_n\tilde{x}\xi_1} d\tilde{x} \{Y_j f^{(l)} - g^{(l)}\}(X_j) e^{\mathrm{i}(X_j, Z_j)\xi},
$$

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

$$t_{1,n}(d_j;\xi)=\sum_{l=0}^{p}\frac{b_n^p(-1)^{p-l}K^{\mathrm{ft}}(b_n\xi_2)}{l!\,(p-l)!f_v^{\mathrm{ft}}(\xi_2)}\iint\left\{\begin{array}{c}\mathbb{K}_\epsilon(\tilde{x})\tilde{x}^l e^{\mathrm{i}b_n\tilde{x}\xi_1}K(x^*)x^{*p-l}\\ \times\begin{bmatrix}\{Y_jf^{(p)}-g^{(p)}\}(\bar{X}_j)\\ -\{Y_jf^{(p)}-g^{(p)}\}(X_j)\end{bmatrix}e^{\mathrm{i}(X_j,Z_j)\xi}\end{array}\right\}dx^*d\tilde{x},$$

for some $\bar{X}_j$ lying between X_j and $X_j+b_n(\tilde{x}-x^*)$, which are obtained using the binomial theorem, Lemma 1, and the properties of the kernel function K as assumed in Assumption O(iii).

For the main term $r^*_{1,n}(d_j;\xi)$, we have

$$\sup_{\xi\in\Xi}\left|\frac{2}{n}\sum_{j=1}^{n}\{r^*_{1,n}(d_j;\xi)-r^{\mathrm{os}}_{1,\infty}(d_j;\xi)-E[r^*_{1,n}(d_j;\xi)-r^{\mathrm{os}}_{1,\infty}(d_j;\xi)]\}\right|=o_p(n^{-1/2}),\quad\textbf{(A.5)}$$

for $r^{\mathrm{os}}_{1,\infty}(d_j;\xi)=\sum_{l=0}^{\alpha}c_l^{\mathrm{os}}(\xi_1)\{Y_jf^{(l)}-g^{(l)}\}(X_j)\,e^{\mathrm{i}(X_j,Z_j)\xi}f_v^{\mathrm{ft}-1}(\xi_2)$, which follows by

$$Var\left(\sup_{\xi\in\Xi}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\{r^*_{1,n}(d_j;\xi)-r^{\mathrm{os}}_{1,\infty}(d_j;\xi)\}\right|\right)\le E\left[\sup_{\xi\in\Xi}\{r^*_{1,n}(d_j;\xi)-r^{\mathrm{os}}_{1,\infty}(d_j;\xi)\}^2\right]$$

$$=O\left(\sup_{\xi\in\Xi}|K^{\mathrm{ft}}(b_n\xi_2)-1|^2E\left[\sup_{\xi\in\Xi}\left|\sum_{l=1}^{\alpha}c_l^{\mathrm{os}}(\xi_1)\{Yf^{(l)}-g^{(l)}\}(X)\right|^2\right]+b_n^2\right)=o(1),$$

where the first equality follows by Lemma 5 and $\inf_{\xi\in\Xi}|f_v^{\mathrm{ft}}(\xi_2)|>\underline{c}$ as assumed in Assumption D(ii), and the second equality follows by Assumption O(v), the compactness of Ξ, and $K^{\mathrm{ft}}(b_n\xi_2)\to 1$ as $n\to\infty$.

For the residual term $t_{1,n}(d_j;\xi)$, we claim that it is negligible, i.e.,

$$\sup_{\xi\in\Xi}\left|\frac{2}{n}\sum_{j=1}^{n}\{t_{1,n}(d_j;\xi)-E[t_{1,n}(d_j;\xi)]\}\right|=o_p(n^{-1/2}),\quad\textbf{(A.6)}$$

which follows by

$$Var\left(\sup_{\xi\in\Xi}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}t_{1,n}(d_j;\xi)\right|\right)\le E\left[\sup_{\xi\in\Xi}t_{1,n}(d_j;\xi)^2\right]$$

$$=O\left(b_n^{2p}\max_{0\le l\le p}E\left[\iint\left\{\begin{array}{c}\left|\mathbb{K}_\epsilon(\tilde{x})\tilde{x}^lK(x^*)x^{*\alpha+1-l}\right|\\ \times\left|\{Yf^{(p)}-g^{(p)}\}(\bar{X})-\{Yf^{(p)}-g^{(p)}\}(X)\right|\end{array}\right\}dx^*d\tilde{x}\right]^2\right)$$

$$=O\left(b_n^{2p+2}\max_{0\le l\le p}E\left[\iint\left\{\begin{array}{c}\left|\mathbb{K}_\epsilon(\tilde{x})\tilde{x}^lK(x^*)x^{*p-l}(\tilde{x}-x^*)\right|\\ \times\{|Y||m_{1,p}|+|m_{2,p}|\}(X)\end{array}\right\}dx^*d\tilde{x}\right]^2\right)$$

$$=O\left(b_n^{2p+2}\max_{0\le l\le p+1}\left(\int\left|\mathbb{K}_\epsilon(\tilde{x})\tilde{x}^l\right|d\tilde{x}\right)^2E[\{Y^2m_{1,p}^2+m_{2,p}^2\}(X)]\right)$$

$$=O(b_n^{2(p-\alpha)+2})=o(1),$$

where the first equality follows by $\inf_{\xi\in\Xi}|f_v^{\mathrm{ft}}(\xi_2)|>\underline{c}$ as assumed in Assumption D(ii) and the boundedness of K^{ft} as assumed in Assumption O(iii), the second equality follows by

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

Assumption O(i), and the last equality follows by $\sup_{0\le l\le p+1}\int\left|\mathbb{K}_\epsilon(u)u^l\right|du = O(b_n^{-\alpha})$, which is obtained using Lemma 4, the boundedness of $m_{1,p}$ and $m_{2,p}$ as assumed in Assumption O(i), and the existence of the second moment of Y as assumed in Assumption D(i).

Then, (A.5) and (A.6) together imply

$$\sup_{\xi\in\Xi}\left|\frac{2}{n}\sum_{j=1}^{n}\{r_{1,n}(d_j;\xi)-r^{\mathrm{os}}_{1,\infty}(d_j;\xi)-E[r_{1,n}(d_j;\xi)-r^{\mathrm{os}}_{1,\infty}(d_j;\xi)]\}\right| = o_p(n^{-1/2}). \quad \textbf{(A.7)}$$

By a similar argument, for $r^{\mathrm{os}}_{2,\infty}(d_j;\xi)=\sum_{l=0}^{\alpha}c_l^{\mathrm{os}}(\xi_1)\int\{g^{(l,0)}_{X^*,Z^*}-Y_jf^{(l,0)}_{X^*,Z^*}\}(X_j,z)$ $e^{\mathrm{i}(X_j,z)\xi}dz$, we can show $\sup_{\xi\in\Xi}\left|\frac{2}{n}\sum_{j=1}^{n}\{r_{2,n}(d_j;\xi)-r^{\mathrm{os}}_{2,\infty}(d_j;\xi)-E[r_{2,n}(d_j;\xi)-r^{\mathrm{os}}_{2,\infty}(d_j;\xi)]\}\right|=o_p(n^{-1/2})$, which together with (A.7) imply

$$\sup_{\xi\in\Xi}\left|\frac{2}{n}\sum_{j=1}^{n}\{r_n(d_j;\xi)-r^{\mathrm{os}}_{\infty}(d_j;\xi)-E[r_n(d_j;\xi)-r^{\mathrm{os}}_{\infty}(d_j;\xi)]\}\right| = o_p(n^{-1/2}). \quad \textbf{(A.8)}$$

Note that Assumption O(v) implies $E\left[\int_\Xi r^{\mathrm{os}}_\infty(d_j;\xi)^2d\xi\right]<\infty$. Then, by Theorem 3.9 of Chen and White (1998), we have

$$\frac{2}{\sqrt{n}}\sum_{j=1}^{n}\{r^{\mathrm{os}}_{\infty}(d_j;\xi)-E[r^{\mathrm{os}}_{\infty}(d_j;\xi)]\}\Rightarrow G^{\mathrm{os}}(\xi). \quad \textbf{(A.9)}$$

The conclusion follows by combining (A.1), (A.4), (A.8), and (A.9).

A.2. *Proof of Theorem 2.* The proof follows along the same lines as Theorem 1, except that (A.2) follows by (A.3) (obtained using the boundedness of K^{ft} as assumed in Assumption S(iii) instead of Assumption O(iii)) and Assumption S(ii) and (iv). In addition, $E[r_n(d_1;\xi)]=0$, for any $\xi\in\Xi$, under Assumption S, i.e., the bias is completely removed by using the infinite-order kernel. To achieve this, we must expand $\{Y_jf-g\}(X_j+b_n(\tilde{x}-x^*))$ around X_j to an infinite order instead of the pth order to adapt to the smoothness of the measurement error under Assumption S. In particular, by expanding $\{Y_jf-g\}(X_j+b_n(\tilde{x}-x^*))$ around X_j to an infinite order and using the binomial theorem, Lemma 1, and the properties of the kernel function K as assumed in Assumption S(iii), we obtain

$$r_{1,n}(d_j;\xi)=\sum_{l=0}^{\infty}\frac{b_n^lK^{\mathrm{ft}}(b_n\xi_2)}{l!f_v^{\mathrm{ft}}(\xi_2)}\int\mathbb{K}_\epsilon(\tilde{x})\tilde{x}^le^{\mathrm{i}b_n\tilde{x}\xi_1}d\tilde{x}\{Y_jf^{(l)}-g^{(l)}\}(X_j)\,e^{\mathrm{i}(X_j,Z_j)\xi}.$$

It is worth noting that we do not have the residual term $t_{1,n}$ as in the ordinary smooth case, because the expansion here is to the infinite order. For this $r_{1,n}(d_j;\xi)$, we have

$$\sup_{\xi\in\Xi}\left|\frac{2}{n}\sum_{j=1}^{n}\{r_{1,n}(d_j;\xi)-r^{\mathrm{ss}}_{1,\infty}(d_j;\xi)-E[r_{1,n}(d_j;\xi)-r^{\mathrm{ss}}_{1,\infty}(d_j;\xi)]\}\right| = o_p(n^{-1/2}), \quad \textbf{(A.10)}$$

for $r^{\mathrm{ss}}_{1,\infty}(d_j;\xi)=\sum_{l=0}^{\infty}c_l^{\mathrm{ss}}(\xi_1)\{Y_jf^{(l)}-g^{(l)}\}(X_j)\,e^{\mathrm{i}(X_j,Z_j)\xi}f_v^{\mathrm{ft}-1}(\xi_2)$, which follows by

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

$$Var\left(\sup_{\xi\in\Xi}\left|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\{r_{1,n}(d_j;\xi)-r^{ss}_{1,\infty}(d_j;\xi)\}\right|\right)\leq E\left[\sup_{\xi\in\Xi}\{r_{1,n}(d_j;\xi)-r^{ss}_{1,\infty}(d_j;\xi)\}^2\right]$$

$$=O\left(\sup_{\xi\in\Xi}|K^{ft}(b_n\xi_2)-1|^2E\left[\sup_{\xi\in\Xi}\left|\sum_{l=1}^{\infty}c^{ss}_l(\xi_1)\{Y_jf^{(l)}-g^{(l)}\}(X_j)\right|^2\right]\right)=o(1),$$

where the first equality follows by Lemma 6 and $\inf_{\xi\in\Xi}|f^{ft}_\nu(\xi_2)|>\underline{c}$ as assumed in Assumption D(ii), and the last equality follows by Assumption S(v), the compactness of Ξ, and $K^{ft}(b_n\xi_2)\to 1$ as $n\to\infty$.

By a similar argument, for $r^{ss}_{2,\infty}(d_j;\xi)]=\sum_{l=0}^{\infty}c^{ss}_l(\xi_1)\int\{g^{(l,0)}_{X^*,Z^*}-Y_jf^{(l,0)}_{X^*,Z^*}\}(X_j,z)e^{i(X_j,z)\xi}dz$, we can show $\sup_{\xi\in\Xi}\left|\frac{2}{n}\sum_{j=1}^{n}\{r_{2,n}(d_j;\xi)-r^{ss}_{2,\infty}(d_j;\xi)-E[r_{2,n}(d_j;\xi)-r^{ss}_{2,\infty}(d_j;\xi)]\}\right|=o_p(n^{-1/2})$, which together with (A.10) imply

$$\sup_{\xi\in\Xi}\left|\frac{2}{n}\sum_{j=1}^{n}\{r_n(d_j;\xi)-r^{ss}_{\infty}(d_j;\xi)-E[r_n(d_j;\xi)-r^{ss}_{\infty}(d_j;\xi)]\}\right|=o_p(n^{-1/2}). \tag{A.11}$$

Note that Assumption S(v) implies $E\left[\int_\Xi r^{ss}_\infty(d_j;\xi)^2d\xi\right]<\infty$. Then, by Theorem 3.9 of Chen and White (1998), we have

$$\frac{2}{\sqrt{n}}\sum_{j=1}^{n}\{r^{ss}_{\infty}(d_j;\xi)-E[r^{ss}_{\infty}(d_j;\xi)]\}\Rightarrow G^{ss}(\xi). \tag{A.12}$$

The conclusion then follows by combining (A.1), (A.4), (A.11), and (A.12).

A.3. *Proof of Theorem 3.* The proof follows along the same lines as Theorems 1 and 2, except that under H_{1n}, instead of (A.4), we have

$$\sup_{\xi\in\Xi}|E[r_n(d_1;\xi)]-n^{-1/2}\mu_\Delta(\xi)|=o(n^{-1/2}), \tag{A.13}$$

which immediately follows by Lemma 3 and H_{1n}.

A.4. *Proof of Theorem 4.* We first introduce some notation needed in this proof. For some probability measure Q on a measurable space $(S,\mathcal{S})$ and a class of measurable functions $\mathcal{F}$ on S such that $\mathcal{F}\subset L^2(Q)$, let $N(\mathcal{F},|\cdot|_{Q,2},\delta)$ denote the δ-covering number for $\mathcal{F}$ with respect to the $L^2(Q)$-seminorm $|\cdot|_{Q,2}$. An envelope function $F:S\to[0,\infty]$ for $\mathcal{F}$ is characterized by the following property $F(u)\geq\sup_{f\in\mathcal{F}}|f(u)|$, for all $u\in S$ (Van Der Vaart and Wellner, 1996).

Since the proofs for the ordinary smooth and supersmooth cases are very similar (given the results from Theorems 1 and 2), we provide the proof only for the ordinary smooth case. Define

$$\bar{r}_n(\xi)=\frac{2}{\sqrt{n}}\sum_{j=1}^{n}\{r_n(d_j;\xi)-E[r_n(d_j;\xi)]\},$$

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

for $r_n\left(d_j;\xi\right)$ defined as in the proof of Theorem 1. Note that

$$\sup_{\xi\in\Xi}\left|\sqrt{n}\{T_n(\xi)-E[T_n(\xi)]\}-\bar{r}_n(\xi)\right|=o_p(n^{-1/2}),$$

from the proof of Theorem 1. Define the class of functions $\mathcal{F}_n^{(r)}=\{(x,y,z)\mapsto \bar{r}_n(x,y,z;\xi):\xi\in\Xi\}$. For $f\in\mathcal{F}_n^{(r)}$, consider the empirical process indexed by $\mathcal{F}_n^{(r)}$: $v_n(f)=\frac{1}{\sqrt{n}}\sum_{j=1}^n f(X_j,Y_j,Z_j)$. We first use Chernozhukov et al. (2016, Thm. 2.1) to show $\sup_{f\in\mathcal{F}_n^{(r)}}|v_n(f)|$ can be approximated by the supremum of a tight Gaussian process, $G_n(f)$, with zero mean and the same covariance function as $v_n(f)$. Thus, we begin by checking the required conditions for this approximation result.

As shown in the proof of Theorem 1, $\sup_{f\in\mathcal{F}_n^{(r)}}|f(x,y,z)|=O_p(1)$. Define $\sup_{f\in\mathcal{F}_n^{(r)}}|f(x,y,z)|=C_1h(x,y,z)$, where $C_1<\infty$ and $h(x,y,z)=O_p(1)$, and an envelope function for $\mathcal{F}_n^{(r)}$ as $F_n^{(r)}(x,y,z)=C_2h(x,y,z)$, where $C_1<C_2<\infty$. With this envelope function, Lemma 7 gives

$$\sup_{Q}N(\mathcal{F}_n^{(r)},|\cdot|_{Q,2},\delta)=O_p(1).$$

Using this result and the bound on the envelope function $F_n^{(r)}$, there exists a tight Gaussian process, $G_n(f)$, on $l^\infty(\mathcal{F}_n^{(r)})$ with mean zero and the same covariance function as $\{v_n(f):f\in\mathcal{F}_n^{(r)}\}$. Following Kato and Sasaki (2019, Thm. 3.1), extend v_n linearly to $\mathcal{F}_n^{(r)}\cup(-\mathcal{F}_n^{(r)})=\{f,-f:f\in\mathcal{F}_n^{(r)}\}$, so that $|v_n|_{\mathcal{F}_n^{(r)}}=\sup_{f\in\mathcal{F}_n^{(r)}\cup(-\mathcal{F}_n^{(r)})}v_n(f)$. Using Giné and Nickl (2016, Thm. 3.7.28), it can be shown that G_n has uniformly continuous paths on the symmetric convex hull of $\mathcal{F}_n^{(r)}$ for the L^2 pseudometric and $\{G_n(f):f\in\mathcal{F}_n^{(r)}\cup(-\mathcal{F}_n^{(r)})\}$ is a tight Gaussian process in $l^\infty(\mathcal{F}_n^{(r)}\cup(-\mathcal{F}_n^{(r)}))$ with zero mean and the same covariance function as $\{v_n(f):f\in\mathcal{F}_n^{(r)}\cup(-\mathcal{F}_n^{(r)})\}$.

The proof of Theorem 1 shows that $\sup_{\xi\in\Xi}|\bar{r}_n(d_j;\xi)-\frac{2}{\sqrt{n}}\sum_{j=1}^n\{r_\infty^{os}(d_j;\xi)-E[r_\infty^{os}(d_j;\xi)]\}|=o_p(1)$ and, by Assumption 4, $\sup_{\xi\in\Xi}E|r_\infty^{os}(X,Y,Z;\xi)|^{2+l}=O_p(1)$; thus, $\sup_{f\in\mathcal{F}_n^{(r)}}E|f(X,Y,Z;\xi)|^{2+l}=O_p(1)$, for $l=0,1,2$. Hence, Chernozhukov et al. (2016, Thm. 2.1) can be applied to $\mathcal{F}_n^{(r)}\cup(-\mathcal{F}_n^{(r)})$ with, in the notation of that paper, $B(f)=0$, $q=4$, $A=O(1)$, $V=O(1)$, $b=O(1)$, $\sigma=O(1)$, and $\gamma=O(1/\log(n))$ to show that a random variable V_n with the same distribution as $\sup_{f\in\mathcal{F}_n^{(r)}}|G_n(f)|$ exists and

$$\left|\sup_{f\in\mathcal{F}_n^{(r)}}|v_n(f)|-V_n\right|=O_p\left(\frac{\log(n)^{5/4}}{n^{1/4}}+\frac{\log(n)}{n^{1/6}}\right)=o_p(1).$$

Define $\bar{r}_n^G(\xi)=G_n(\bar{r}_n(\xi))$, for $\xi\in\Xi$, so that $\bar{r}_n^G(\xi)$ is a tight Gaussian process in $l^\infty(\Xi)$ with zero mean and the same covariance function as $\bar{r}_n(\xi)$, giving $\sup_{\xi\in\Xi}|\bar{r}_n^G(\xi)|$ the same distribution as V_n. Now,

$$\left|\sup_{\xi\in\Xi}|\sqrt{n}\{T_n(\xi)-E[T_n(\xi)]\}|-V_n\right|$$

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

$$\leq \left| \sup_{\xi \in \Xi} |\sqrt{n}\{T_n(\xi) - E[T_n(\xi)]\}| - \sup_{\xi \in \Xi} |\bar{r}_n(\xi)| \right| + \left| \sup_{\xi \in \Xi} |\bar{r}_n(\xi)| - V_n \right| = o_p(1),$$

which implies $P\left\{ \left| \sup_{\xi \in \Xi} \left| \sqrt{n}\{T_n(\xi) - E[T_n(\xi)]\} \right| - V_n \right| > \epsilon_n \right\} \leq \epsilon_n$, for some $\epsilon_n = O(n^{-c})$ with $c > 0$ (Dudley, 2002, Thm. 9.2.2). Combining these results, we have

$$\begin{aligned} & P\left\{ \sup_{\xi \in \Xi} |\sqrt{n}\{T_n(\xi) - E[T_n(\xi)]\}| \leq w \right\} \\ & \quad \geq P\left\{ \sup_{\xi \in \Xi} |\bar{r}_n^G(\xi)| \leq w - \epsilon_n \right\} - \epsilon_n \\ & \quad \geq P\left\{ \sup_{\xi \in \Xi} |\bar{r}_n^G(\xi)| \leq w \right\} - 4\epsilon_n \left(1 + E[\sup_{\xi \in \Xi} |\bar{r}_n^G(\xi)|] \right), \end{aligned}$$

where the second inequality follows from Chernozhukov et al. (2015, Thm. 3). Thus, it remains to bound $E[\sup_{\xi \in \Xi} |\bar{r}_n^G(\xi)|]$. In our case, Dudley's entropy integral bound gives

$$E\left[\sup_{\xi \in \Xi} |\bar{r}_n^G(\xi)| \right] = O\left(\int_0^1 \sqrt{\sup_Q N(\mathcal{F}_n^{(r)}, |\cdot|_{Q,2}, \delta)} d\delta \right) = O(1),$$

where the second equality uses Lemma 7. Therefore,

$$P\left\{ \sup_{\xi \in \Xi} |\sqrt{n}\{T_n(\xi) - E[T_n(\xi)]\}| \leq w \right\} \geq P\left\{ \sup_{\xi \in \Xi} |\bar{r}_n^G(\xi)| \leq w \right\} - o(1).$$

The final two steps of the proof are to show

$$P\left\{ \sup_{\xi \in \Xi} |\bar{r}_n^{\#}(\xi)| \leq w \middle| d_n \right\} \geq P\left\{ \sup_{\xi \in \Xi} |\sqrt{n}\{T_n^{\#}(\xi) - E[T_n^{\#}(\xi)]\}| \leq w \middle| d_n \right\} - o(1), \tag{A.14}$$

$$P\left\{ \sup_{\xi \in \Xi} |\bar{r}_n^G(\xi)| \leq w \right\} \geq P\left\{ \sup_{\xi \in \Xi} |\bar{r}_n^{\#}(\xi)| \leq w \middle| d_n \right\} - o(1), \tag{A.15}$$

where $\bar{r}_n^{\#}(\xi) = \bar{r}_n(X^{\#}, Y^{\#}, Z^{\#}; \xi)$ and $d_n = \{X_j, Y_j, Z_j\}_{j=1}^n$.

Inequality (A.14) can be shown by a straightforward application of the arguments of Theorem 1. For (A.15), we appeal to Chernozhukov et al. (2016, Thm. 2.3). To this end, define $v_n^{\#}(f) = \frac{1}{\sqrt{n}} \sum_{k=1}^n f(X_k^{\#}, Y_k^{\#}, Z_k^{\#})$, where $f \in \mathcal{F}_n^{(r)}$. Taking $B(f) = 0$, $q = 4$, $A = O(1)$, $V = O(1)$, $b = O(1)$, $\sigma = O(1)$, and $\gamma = O(1/\log(n))$, Chernozhukov et al. (2016, Thm. 2.3) gives that there exists a random variable $V_n^{\#}$ with the same distribution (conditional on d_j) as $\sup_{f \in \mathcal{F}_n^{(r)}} |G_n(f)|$ and

$$\left| \sup_{f \in \mathcal{F}_n^{(r)}} |v_n^{\#}(f)| - V_n^{\#} \right| = O_p\left(\frac{\log(n)^{9/4}}{n^{1/4}} + \frac{\log(n)^2}{n^{1/4}} + \frac{\log(n)}{n^{1/6}} \right) = o_p(1).$$

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

Therefore, for some $\epsilon_n = O(n^{-c})$, for $c > 0$,

$$P\left\{\sup_{\xi\in\Xi}|\bar{r}_n^{\#}(\xi)| \le w \middle| d_n\right\} \ge P\left\{\sup_{\xi\in\Xi}|\bar{r}_n^{G}(\xi)| \le w-\epsilon_n\right\} - \epsilon_n$$
$$\ge P\left\{\sup_{\xi\in\Xi}|\bar{r}_n^{G}(\xi)| \le w\right\} - o(1),$$

where the second inequality follows again from Chernozhukov et al. (2015, Thm. 3) and Dudley's entropy integral bound. Combining these results, we have

$$P\left\{\sup_{\xi\in\Xi}|\sqrt{n}\{T_n(\xi) - E[T_n(\xi)]\}| \le w\right\} \ge P\left\{\sup_{\xi\in\Xi}|\bar{r}_n^{G}(\xi)| \le w\right\} - o(1)$$
$$\ge P\left\{\sup_{\xi\in\Xi}|\bar{r}_n^{\#}(\xi)| \le w \middle| d_n\right\} - o(1) \ge P\left\{\sup_{\xi\in\Xi}|\sqrt{n}\{T_n^{\#}(\xi) - E[T_n^{\#}(\xi)]\}| \le w \middle| d_n\right\} - o(1),$$

and the conclusion follows by taking w to be $\hat{c}_\alpha^{KS}$, i.e., the $(1-\alpha)$th quantile of $KS_n^{\#}$.

$$P\left\{\sup_{\xi\in\Xi}|\sqrt{n}\{T_n(\xi) - E[T_n(\xi)]\}| \le w\right\} \ge 1-\alpha+o(1).$$

A.5. *Proof of Theorem 5.* Let $\hat{\zeta}_\epsilon(t) = \frac{1}{n}\sum_{l=1}^n \zeta_{\epsilon,l}(t)$ and $\hat{\zeta}_\nu(t) = \frac{1}{n}\sum_{l=1}^n \zeta_{\nu,l}(t)$, for $\zeta_{\epsilon,l}(t) = \cos(t(X_l - X_l^r))$ and $\zeta_{\nu,l}(t) = \cos(t(Z_l - Z_l^r))$, and $\zeta_\epsilon(t) = E[\zeta_{\epsilon,1}(t)]$ and $\zeta_\nu(t) = E[\zeta_{\nu,1}(t)]$. Note that $\hat{f}_\epsilon^{\mathrm{ft}}(t) = \left|\hat{\zeta}_\epsilon(t)\right|^{1/2}, \hat{f}_\nu^{\mathrm{ft}}(t) = \left|\hat{\zeta}_\nu(t)\right|^{1/2}, f_\epsilon^{\mathrm{ft}}(t) = |\zeta_\epsilon(t)|^{1/2}$, and $f_\nu^{\mathrm{ft}}(t) = |\zeta_\nu(t)|^{1/2}$.

For a random error $\eta \in \{\epsilon, \nu\}$, define

$$A_\eta(u) = \frac{1}{2\pi}\int e^{-\mathrm{i}tu}\frac{K^{\mathrm{ft}}(t)}{f_\eta^{\mathrm{ft}}(t/b_n)}\hat{\Pi}_\eta(t/b_n)dt,$$
$$R_\eta(u) = \frac{1}{2\pi}\int e^{-\mathrm{i}tu}\frac{K^{\mathrm{ft}}(t)}{f_\eta^{\mathrm{ft}}(t/b_n)}\hat{\Pi}_\eta^{\mathrm{res}}(t/b_n)dt,$$

where $\hat{\Pi}_\eta(t) = \frac{1}{n}\sum_{l=1}^n \Pi_{\eta,l}(t)$, for $\Pi_{\eta,l}(t) = \frac{\zeta_\eta(t)-\zeta_{\eta,l}(t)}{2\zeta_\eta(t)}$ and

$$\hat{\Pi}_\eta^{\mathrm{res}}(t) = \frac{\left(2|\hat{\zeta}_\eta(t)|^{1/2} + |\zeta_\eta(t)|^{1/2}\right)\left(|\zeta_\eta(t)|^{1/2} - |\hat{\zeta}_\eta(t)|^{1/2}\right)^2}{|\zeta_\eta(t)|^{3/2}|\hat{\zeta}_\eta(t)|^{1/2}}.$$

By expanding $\hat{\zeta}_\eta(t/b_n)$ around $\zeta_\eta(t/b_n)$, we obtain

$$\hat{\mathbb{K}}_\eta(u) = \mathbb{K}_\eta(u) + A_\eta(u) + R_\eta(u).$$

Here, $A_\eta(u)$ is the Fréchet derivative of $\hat{\mathbb{K}}_\eta(u)$ as a functional of $\hat{\zeta}_\eta(t/b_n)$ at $\zeta_\eta(t/b_n)$. Also note that $\hat{\Pi}_\eta(t)$ is of order $(|\zeta_\eta(t)|^{1/2} - |\hat{\zeta}_\eta(t)|^{1/2})$ and $\hat{\Pi}_\eta^{\mathrm{res}}(t)$ is of order $(|\zeta_\eta(t)|^{1/2} -$

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

$|\hat{\zeta}_\eta(t)|^{1/2})^2$. So, $R_\eta(u)$ is a higher-order term than $A_\eta(u)$, and should be dominated by $A_\eta(u)$ asymptotically.

Decompose

$$\hat{T}_n(\xi) = \hat{S}_n(\xi) + T_{1,n}(\xi) + \cdots + T_{4,n}(\xi),$$

where

$$\hat{S}_n(\xi) = \frac{1}{n^2 b_n^3} \sum_{j \neq k}^{n} (Y_k - Y_j) \iint \left\{ \begin{array}{l} \mathbb{K}_\epsilon\left(\frac{x-X_j}{b_n}\right)\mathbb{K}_\epsilon\left(\frac{x-X_k}{b_n}\right)\mathbb{K}_\nu\left(\frac{z-Z_k}{b_n}\right) \\ +A_\epsilon\left(\frac{x-X_j}{b_n}\right)\mathbb{K}_\epsilon\left(\frac{x-X_k}{b_n}\right)\mathbb{K}_\nu\left(\frac{z-Z_k}{b_n}\right) \\ +\mathbb{K}_\epsilon\left(\frac{x-X_j}{b_n}\right)A_\epsilon\left(\frac{x-X_k}{b_n}\right)\mathbb{K}_\nu\left(\frac{z-Z_k}{b_n}\right) \\ +\mathbb{K}_\epsilon\left(\frac{x-X_j}{b_n}\right)\mathbb{K}_\epsilon\left(\frac{x-X_k}{b_n}\right)A_\nu\left(\frac{z-Z_k}{b_n}\right) \end{array} \right\} e^{\mathrm{i}(x,z)\xi}\, dxdz,$$

$$T_{1,n}(\xi) = \frac{1}{n^2 b_n^3} \sum_{j \neq k}^{n} (Y_k - Y_j) \iint \mathbb{K}_\epsilon\left(\frac{x-X_j}{b_n}\right)\mathbb{K}_\epsilon\left(\frac{x-X_k}{b_n}\right)R_\nu\left(\frac{z-Z_k}{b_n}\right) e^{\mathrm{i}(x,z)\xi}\, dxdz,$$

$$T_{2,n}(\xi) = \frac{1}{n^2 b_n^3} \sum_{j \neq k}^{n} (Y_k - Y_j) \iint \left\{ \begin{array}{l} A_\epsilon\left(\frac{x-X_j}{b_n}\right)\mathbb{K}_\epsilon\left(\frac{x-X_k}{b_n}\right) \\ +\mathbb{K}_\epsilon\left(\frac{x-X_j}{b_n}\right)A_\epsilon\left(\frac{x-X_k}{b_n}\right) \end{array} \right\} \times \left\{ \begin{array}{l} A_\nu\left(\frac{z-Z_k}{b_n}\right) \\ +R_\nu\left(\frac{z-Z_k}{b_n}\right) \end{array} \right\} e^{\mathrm{i}(x,z)\xi}\, dxdz,$$

$$T_{3,n}(\xi) = \frac{1}{n^2 b_n^3} \sum_{j \neq k}^{n} (Y_k - Y_j) \iint \left\{ \begin{array}{l} R_\epsilon\left(\frac{x-X_j}{b_n}\right)\mathbb{K}_\epsilon\left(\frac{x-X_k}{b_n}\right) \\ +\mathbb{K}_\epsilon\left(\frac{x-X_j}{b_n}\right)R_\epsilon\left(\frac{x-X_k}{b_n}\right) \end{array} \right\} \hat{\mathbb{K}}_\nu \times \left(\frac{z-Z_k}{b_n}\right) e^{\mathrm{i}(x,z)\xi}\, dxdz,$$

$$T_{4,n}(\xi) = \frac{1}{n^2 b_n^3} \sum_{j \neq k}^{n} (Y_k - Y_j) \iint \left\{ \begin{array}{l} A_\epsilon\left(\frac{x-X_j}{b_n}\right) \\ +R_\epsilon\left(\frac{x-X_j}{b_n}\right) \end{array} \right\} \left\{ \begin{array}{l} A_\epsilon\left(\frac{x-X_k}{b_n}\right) \\ +R_\epsilon\left(\frac{x-X_k}{b_n}\right) \end{array} \right\} \hat{\mathbb{K}}_\nu \times \left(\frac{z-Z_k}{b_n}\right) e^{\mathrm{i}(x,z)\xi}\, dxdz.$$

First, we show $T_{1,n}(\xi), \ldots, T_{4,n}(\xi)$ are asymptotic negligible, that is,

$$\sup_{\xi \in \Xi} \left|T_{1,n}(\xi)\right|, \ldots, \sup_{\xi \in \Xi} \left|T_{4,n}(\xi)\right| = o_p(n^{-1/2}). \tag{A.16}$$

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

For $T_{1,n}(\xi)$, we have

$$T_{1,n}(\xi) = \frac{1}{2\pi n^2 b_n} \sum_{j\neq k}^{n} (Y_k - Y_j) \left\{ \begin{array}{l} \iint \left\{ \begin{array}{c} \frac{1}{2\pi} \int e^{-\mathrm{i}(t_1+t_2-b_n\xi_1)\tilde{x}} d\tilde{x} \\ \times \frac{K^{\mathrm{ft}}(t_1)K^{\mathrm{ft}}(t_2) e^{\mathrm{i}\left(\frac{t_1 X_j + t_2 X_k}{b_n}\right)}}{f_\epsilon^{\mathrm{ft}}(t_1/b_n) f_\epsilon^{\mathrm{ft}}(t_2/b_n)} \end{array} \right\} dt_1 dt_2 \\ \times \int \left\{ \begin{array}{c} \frac{1}{2\pi} \int e^{-\mathrm{i}(t_3-b_n\xi_2)\tilde{z}} d\tilde{z} \\ \times \frac{K^{\mathrm{ft}}(t_3) e^{\mathrm{i}t_3 Z_k/b_n}}{f_\nu^{\mathrm{ft}}(t_3/b_n)} \hat{\Pi}_\nu^{\mathrm{res}}(t_3/b_n) \end{array} \right\} dt_3 \end{array} \right\}$$

$$= \frac{1}{2\pi n^2 b_n} \sum_{j\neq k}^{n} (Y_k - Y_j) \left\{ \begin{array}{c} \int \frac{K^{\mathrm{ft}}(t) K^{\mathrm{ft}}(-t+b_n\xi_1) e^{\mathrm{i}\left(\frac{t(X_j - X_k)}{b_n}\right)}}{f_\epsilon^{\mathrm{ft}}(t/b_n) f_\epsilon^{\mathrm{ft}}(-t/b_n+\xi_1)} dt \\ \times \frac{K^{\mathrm{ft}}(b_n\xi_2) e^{\mathrm{i}(X_k, Z_k)\xi}}{f_\nu^{\mathrm{ft}}(\xi_2)} \hat{\Pi}_\nu^{\mathrm{res}}(\xi_2) \end{array} \right\}$$

$$= O_p\left(\frac{1}{n b_n (\inf_{|t|\leq 2b_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)|)^2} \right), \tag{A.17}$$

where the second equality follows by $\int \delta(u-b)m(u)du = m(b)$ with the Dirac delta function $\delta(u) = \frac{1}{2\pi}\int e^{-\mathrm{i}ut}dt$, and the last equality follows from the fact that K^{ft} is compactly supported and $\sup_{\xi\in\Xi} |\hat{\Pi}_\nu^{\mathrm{res}}(\xi_2)| = O_p(n^{-1})$, which follows from Lemma 9 and the compactness of Ξ.

By similar arguments, we can show

$$T_{2,n}(\xi) = O_p\left(\frac{\sup_{|t|\leq 2b_n^{-1}} |\hat{\Pi}_\epsilon(t)|}{n^{1/2} b_n (\inf_{|t|\leq 2b_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)|)^2} \right),$$

$$T_{3,n}(\xi) = O_p\left(\frac{\sup_{|t|\leq 2b_n^{-1}} |\hat{\Pi}_\epsilon^{\mathrm{res}}(t)|}{b_n (\inf_{|t|\leq 2b_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)|)^2} \right),$$

$$T_{4,n}(\xi) = O_p\left(\frac{\sup_{|t|\leq 2b_n^{-1}} |\hat{\Pi}_\epsilon(t)|^2}{b_n (\inf_{|t|\leq 2b_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)|)^2} \right), \tag{A.18}$$

and (A.16) follows by (B.1), (B.2), Assumptions O and O'.

Let $a_j = (Y_j, X_j, Z_j, X_j^r, Z_j^r)$ and

$$\hat{p}_n(a_j, a_k, a_l) = \left\{ \begin{array}{c} \hat{q}_n(a_j, a_k, a_l) + \hat{q}_n(a_j, a_l, a_k) + \hat{q}_n(a_k, a_j, a_l) \\ + \hat{q}_n(a_l, a_j, a_k) + \hat{q}_n(a_k, a_l, a_j) + \hat{q}_n(a_l, a_k, a_j) \end{array} \right\},$$

where

$$\hat{q}_n(a_j, a_k, a_l) =$$

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

$$\frac{1}{6b_n^3}\iint\left\{\begin{array}{l}\mathbb{K}_\epsilon\left(\frac{x-X_j}{b_n}\right)Y_k\mathbb{K}_\epsilon\left(\frac{x-X_k}{b_n}\right)\mathbb{K}_\nu\left(\frac{z-Z_k}{b_n}\right)\\ -Y_j\mathbb{K}_\epsilon\left(\frac{x-X_j}{b_n}\right)\mathbb{K}_\epsilon\left(\frac{x-X_k}{b_n}\right)\mathbb{K}_\nu\left(\frac{z-Z_k}{b_n}\right)\\ +\frac{1}{2\pi}\int\left\{\begin{array}{l}e^{-\mathrm{i}t\left(\frac{x-X_j}{b_n}\right)}Y_k\mathbb{K}_\epsilon\left(\frac{x-X_k}{b_n}\right)\mathbb{K}_\nu\left(\frac{z-Z_k}{b_n}\right)\\ -Y_je^{-\mathrm{i}t\left(\frac{x-X_j}{b_n}\right)}\mathbb{K}_\epsilon\left(\frac{x-X_k}{b_n}\right)\mathbb{K}_\nu\left(\frac{z-Z_k}{b_n}\right)\\ +\mathbb{K}_\epsilon\left(\frac{x-X_j}{b_n}\right)Y_ke^{-\mathrm{i}t\left(\frac{x-X_k}{b_n}\right)}\mathbb{K}_\nu\left(\frac{z-Z_k}{b_n}\right)\\ -Y_j\mathbb{K}_\epsilon\left(\frac{x-X_j}{b_n}\right)e^{-\mathrm{i}t\left(\frac{x-X_k}{b_n}\right)}\mathbb{K}_\nu\left(\frac{z-Z_k}{b_n}\right)\end{array}\right\}\frac{\Pi_{\epsilon,l}(t/b_n)K^{\mathrm{ft}}(t)}{f_\epsilon^{\mathrm{ft}}(t/b_n)}dt\\ +\frac{1}{2\pi}\int\left\{\begin{array}{l}\mathbb{K}_\epsilon\left(\frac{x-X_j}{b_n}\right)Y_k\mathbb{K}_\epsilon\left(\frac{x-X_k}{b_n}\right)e^{-\mathrm{i}t\left(\frac{z-Z_k}{b_n}\right)}\\ -Y_j\mathbb{K}_\epsilon\left(\frac{x-X_j}{b_n}\right)\mathbb{K}_\epsilon\left(\frac{x-X_k}{b_n}\right)e^{-\mathrm{i}t\left(\frac{z-Z_k}{b_n}\right)}\end{array}\right\}\frac{\Pi_{\nu,l}(t/b_n)K^{\mathrm{ft}}(t)}{f_\nu^{\mathrm{ft}}(t/b_n)}dt\end{array}\right\}$$
$$\times e^{\mathrm{i}(x,z)\xi}\,dxdz.$$

Furthermore, decompose

$$\hat{S}_n(\xi)=n^{-2}(n-1)(n-2)\hat{U}_n(\xi)+S_{1,n}(\xi)+\cdots+S_{4,n}(\xi),$$

where

$$\hat{U}_n(\xi)=\binom{n}{3}^{-1}\sum_{j=1}^{n-2}\sum_{k=j+1}^{n-1}\sum_{l=k+1}^{n}p_n(a_j,a_k,a_l),$$

$$S_{1,n}(\xi)=\frac{6}{n^3}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}[p_n(a_j,a_k,a_j)+p_n(a_k,a_j,a_k)],$$

$$S_{2,n}(\xi)=\frac{6}{n^3}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}[p_n(a_k,a_j,a_j)+p_n(a_j,a_k,a_k)],$$

$$S_{3,n}(\xi)=\frac{6}{n^3}\sum_{j=1}^{n-1}\sum_{k=j+1}^{n}[p_n(a_j,a_j,a_k)+p_n(a_k,a_k,a_j)],$$

$$S_{4,n}(\xi)=\frac{6}{n^3}\sum_{j=1}^{n}p_n(a_j,a_j,a_j).$$

We claim that $S_{1,n}(\xi),\ldots,S_{4,n}(\xi)$ are asymptotically negligible, i.e.,

$$\sup_{\xi\in\Xi}\left|S_{1,n}(\xi)\right|,\ldots,\sup_{\xi\in\Xi}\left|S_{4,n}(\xi)\right|=o_p(n^{-1/2}). \tag{A.19}$$

As $S_{3,n}(\xi)=S_{4,n}(\xi)=0$, we focus on $S_{1,n}(\xi)$ and $S_{2,n}(\xi)$. Note that

$$\hat{q}_n(a_j,a_k,a_l)$$

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

$$= \frac{Y_k - Y_j}{12\pi b_n} \iiint \left\{ \begin{array}{c} \left\{ \frac{1}{2\pi} \int e^{-\mathrm{i}(t_1+t_2-b_n\xi_1)\tilde{x}} d\tilde{x} \right\} \left\{ \frac{1}{2\pi} \int e^{-\mathrm{i}(t_3-b_n\xi_2)\tilde{z}} d\tilde{z} \right\} \\ \times \frac{K^{\mathrm{ft}}(t_1)K^{\mathrm{ft}}(t_2)K^{\mathrm{ft}}(t_3) e^{\mathrm{i}\left(\frac{t_1 X_j + t_2 X_k + t_3 Z_k}{b_n}\right)}}{f_\epsilon^{\mathrm{ft}}(t_1/b_n) f_\epsilon^{\mathrm{ft}}(t_2/b_n) f_v^{\mathrm{ft}}(t_3/b_n)} \\ \times \left\{ 1 + \Pi_{\epsilon,l}(t_1/b_n) + \Pi_{\epsilon,l}(t_2/b_n) + \Pi_{v,l}(t_3/b_n) \right\} \end{array} \right\} dt_1 dt_2 dt_3$$

$$= \frac{Y_k - Y_j}{12\pi b_n} \left\{ \begin{array}{c} \int \frac{K^{\mathrm{ft}}(t) K^{\mathrm{ft}}(-t+b_n\xi_1) e^{\mathrm{i}t\left(\frac{X_j - X_k}{b_n}\right)}}{f_\epsilon^{\mathrm{ft}}(t/b_n) f_\epsilon^{\mathrm{ft}}(-t/b_n+\xi_1)} dt \quad \frac{K^{\mathrm{ft}}(b_n\xi_2)}{f_v^{\mathrm{ft}}(\xi_2)} e^{\mathrm{i}(X_k, Z_k)\xi} \\ \times \left\{ 1 + \Pi_{\epsilon,l}(t/b_n) + \Pi_{\epsilon,l}(-t/b_n+\xi_1) + \Pi_{v,l}(\xi_2) \right\} \end{array} \right\},$$

where the first equality follows from the change of variables $\tilde{x} = \frac{x}{b_n}$ and $\tilde{z} = \frac{z}{b_n}$ and the second equality follows by the property of the Dirac delta function, which implies

$$S_{1,n}(\xi), S_{2,n}(\xi) = O_p\left(\frac{\sup_{|t| \le 2b_n^{-1}} |\Pi_{\epsilon,l}(t)|}{n b_n (\inf_{|t| \le 2b_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)|)^2} \right), \tag{A.20}$$

and (A.19) follows by (A.20), Assumptions O and O', and Lemma 10.

Thus, our focus turns to $\hat{U}_n(\xi)$. Let $\hat{U}_n^*(\xi) = E[\hat{r}_n(a_1;\xi)] + \frac{3}{n}\sum_{j=1}^n \{\hat{r}_n(a_j;\xi) - E[\hat{r}_n(a_j;\xi)]\}$, for $\hat{r}_n(a_j) = E[\hat{p}_n(a_j, a_k, a_l)|a_j]$. By the Hájek projection (see, for example, Ahn and Powell, 1993, Lem. A.3), if

$$E\left[\sup_{\xi \in \Xi} \hat{p}_n(a_j, a_k, a_l)^2 \right] = o(n), \tag{A.21}$$

then

$$\sup_{\xi \in \Xi} \left| \hat{U}_n(\xi) - \hat{U}_n^*(\xi) \right| = o_p(n^{-1/2}). \tag{A.22}$$

To show (A.21), note that

$$\begin{aligned} & E\left[\sup_{\xi \in \Xi} \hat{p}_n(a_j, a_k, a_l)^2 \right] \\ & \le 6E\left[\sup_{\xi \in \Xi} \hat{q}_n(a_j, a_k, a_l)^2 \right] \\ & \le \frac{\sup_{\xi \in \Xi} \left| \frac{K^{\mathrm{ft}}(b_n\xi_2)}{f_v^{\mathrm{ft}}(\xi_2)} \right|}{12\pi^2 b_n^2} \\ & \quad \times \iint \left\{ \begin{array}{c} \frac{K^{\mathrm{ft}}(t_1)K^{\mathrm{ft}}(-t_1+b_n\xi_1)K^{\mathrm{ft}}(t_2)K^{\mathrm{ft}}(-t_2+b_n\xi_1)}{f_\epsilon^{\mathrm{ft}}(t_1/b_n) f_\epsilon^{\mathrm{ft}}(-t_1/b_n+\xi_1) f_\epsilon^{\mathrm{ft}}(t_2/b_n) f_\epsilon^{\mathrm{ft}}(-t_2/b_n+\xi_1)} \\ \times E\left[(Y_k^2 + Y_j^2) e^{\mathrm{i}(t_1+t_2)X_j/b_n} e^{-\mathrm{i}(t_1+t_2)X_k/b_n} \right] \\ \times E\left[\begin{array}{c} \left\{ 1 + \Pi_{\epsilon,l}(t_1/b_n) + \Pi_{\epsilon,l}(-t_1/b_n+\xi_1) + \Pi_{v,l}(\xi_2) \right\} \\ \times \left\{ 1 + \Pi_{\epsilon,l}(t_2/b_n) + \Pi_{\epsilon,l}(-t_2/b_n+\xi_1) + \Pi_{v,l}(\xi_2) \right\} \end{array} \right] \end{array} \right\} dt_1 dt_2 \\ & = O\left(\frac{(\sup_{|t| \le 2b_n^{-1}} |\Pi_{\epsilon,l}(t)|)^2}{b_n^2 (\inf_{|t| \le 2b_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)|)^4} \right), \end{aligned} \tag{A.23}$$

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

where the first step follows from Jensen's inequality and the last step follows by the boundedness of $E[Y^2|X^*]$ guaranteed by Assumption D(i). Then, (A.21) follows by (A.23), Assumptions O and O', and Lemma 10.

Decompose $\hat{r}_n(a_j;\xi) = \hat{r}_n^*(a_j;\xi) + \hat{r}_n^\epsilon(a_j;\xi) + \hat{r}_n^\nu(a_j;\xi)$, where

$$\hat{r}_n^*(a_j;\xi) = \iint \frac{e^{\mathrm{i}(x,z)\xi}}{3b_n^3} \left\{ \begin{array}{c} \mathbb{K}_\epsilon\left(\frac{x-X_j}{b_n}\right) E\left[YK\left(\frac{x-X^*}{b_n}\right)K\left(\frac{z-Z^*}{b_n}\right)\right] \\ -Y_j\mathbb{K}_\epsilon\left(\frac{x-X_j}{b_n}\right) E\left[K\left(\frac{x-X^*}{b_n}\right)K\left(\frac{z-Z^*}{b_n}\right)\right] \\ +E\left[K\left(\frac{x-X^*}{b_n}\right)\right] Y_j\mathbb{K}_\epsilon\left(\frac{x-X_j}{b_n}\right)\mathbb{K}_\nu\left(\frac{z-Z_j}{b_n}\right) \\ -E\left[YK\left(\frac{x-X^*}{b_n}\right)\right]\mathbb{K}_\epsilon\left(\frac{x-X_j}{b_n}\right)\mathbb{K}_\nu\left(\frac{z-Z_j}{b_n}\right) \\ +E\left[K\left(\frac{x-X^*}{b_n}\right)\right] E\left[YK\left(\frac{x-X^*}{b_n}\right)K\left(\frac{z-Z^*}{b_n}\right)\right] \\ -E\left[YK\left(\frac{x-X^*}{b_n}\right)\right] E\left[K\left(\frac{x-X^*}{b_n}\right)K\left(\frac{z-Z^*}{b_n}\right)\right] \end{array} \right\} dxdz,$$

$$\hat{r}_n^\epsilon(a_j;\xi) = \iiint \frac{\Pi_{\epsilon,j}(t/b_n)K^{\mathrm{ft}}(t)e^{\mathrm{i}(x,z)\xi}}{6\pi b_n^3} \times \left\{ \begin{array}{c} E\left[e^{-\mathrm{i}t\left(\frac{x-X^*}{b_n}\right)}\right] E\left[YK\left(\frac{x-X^*}{b_n}\right)K\left(\frac{z-Z^*}{b_n}\right)\right] \\ -E\left[Ye^{-\mathrm{i}t\left(\frac{x-X^*}{b_n}\right)}\right] E\left[K\left(\frac{x-X^*}{b_n}\right)K\left(\frac{z-Z^*}{b_n}\right)\right] \\ +E\left[K\left(\frac{x-X^*}{b_n}\right)\right] E\left[Ye^{-\mathrm{i}t\left(\frac{x-X^*}{b_n}\right)}K\left(\frac{z-Z^*}{b_n}\right)\right] \\ -E\left[YK\left(\frac{x-X^*}{b_n}\right)\right] E\left[e^{-\mathrm{i}t\left(\frac{x-X^*}{b_n}\right)}K\left(\frac{z-Z^*}{b_n}\right)\right] \end{array} \right\} dtdxdz,$$

$$\hat{r}_n^\nu(a_j;\xi) = \iiint \frac{\Pi_{\nu,j}(t/b_n)K^{\mathrm{ft}}(t)e^{\mathrm{i}(x,z)\xi}}{6\pi b_n^3} \times \left\{ \begin{array}{c} E\left[K\left(\frac{x-X^*}{b_n}\right)\right] E\left[YK\left(\frac{x-X^*}{b_n}\right)e^{-\mathrm{i}t\left(\frac{z-Z^*}{b_n}\right)}\right] \\ -E\left[YK\left(\frac{x-X^*}{b_n}\right)\right] E\left[K\left(\frac{x-X^*}{b_n}\right)e^{-\mathrm{i}t\left(\frac{z-Z^*}{b_n}\right)}\right] \end{array} \right\} dtdxdz,$$

which are obtained using Lemma 2.

For $\hat{U}_n^*(\xi)$, we first claim that it has the same bias term as $U_n^*(\xi)$, i.e.,

$$E[\hat{r}_n(a_j;\xi)] = E[r_n(d_j;\xi)], \tag{A.24}$$

which follows from $E[\hat{r}_n^\epsilon(a_j;\xi)] = E[\hat{r}_n^\nu(a_j;\xi)] = 0$ and $E[\hat{r}_n^*(a_j;\xi)] = E[r_n(d_j;\xi)]$. Here, $E[\hat{r}_n^\epsilon(a_j;\xi)] = E[\hat{r}_n^\nu(a_j;\xi)] = 0$ is an immediate result of $E[\Pi_{\epsilon,j}(t/b_n)] = E[\Pi_{\nu,j}(t/b_n)] = 0$, and $E[\hat{r}_n^*(a_j;\xi)] = E[r_n(d_j;\xi)]$ is shown by using Lemma 2.

For the main term in $\hat{U}_n^*(\xi)$, we first note that

$$3\{\hat{r}_n^*(a_j;\xi) - E[\hat{r}_n^*(a_j;\xi)]\} = 2\{r_n(d_j;\xi) - E[r_n(d_j;\xi)]\}.$$

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

Also note that

$$\hat{r}_n^{\epsilon}(a_j;\xi) = \int\cdots\int \frac{\Pi_{\epsilon,j}(t/b_n)K^{\mathrm{ft}}(t)}{6\pi b_n^3} \times \left\{ \begin{array}{c} \left\{ \begin{array}{c} f^{\mathrm{ft}}(t/b_n)g_{X^*,Z^*}(x^*,z^*) \\ -g^{\mathrm{ft}}(t/b_n)f_{X^*,Z^*}(x^*,z^*) \\ +f(x^*)\left\{g_{X^*,Z^*}(\cdot,z^*)\right\}^{\mathrm{ft}}(t/b_n) \\ -g(x^*)\left\{f_{X^*,Z^*}(\cdot,z^*)\right\}^{\mathrm{ft}}(t/b_n) \end{array} \right\} \\ \times e^{\mathrm{i}((\xi_1-t/b_n)x+\xi_2 z)}K\left(\frac{x-x^*}{b_n}\right)K\left(\frac{z-z^*}{b_n}\right) \end{array} \right\} dxdzdx^*dz^*dt$$

$$= \int \frac{\Pi_{\epsilon,j}(t/b_n)K^{\mathrm{ft}}(t)}{6\pi b_n} \times \left\{ \begin{array}{c} \left\{ \begin{array}{c} f^{\mathrm{ft}}(t/b_n)\iint e^{\mathrm{i}\{(\xi_1-t/b_n)x^*+\xi_2 z^*\}}g_{X^*,Z^*}(x^*,z^*)dx^*dz^* \\ -g^{\mathrm{ft}}(t/b_n)\iint e^{\mathrm{i}\{(\xi_1-t/b_n)x^*+\xi_2 z^*\}}f_{X^*,Z^*}(x^*,z^*)dx^*dz^* \\ +\iint e^{\mathrm{i}((\xi_1-t/b_n)x^*+\xi_2 z^*)}f(x^*)\left\{g_{X^*,Z^*}(\cdot,z^*)\right\}^{\mathrm{ft}}(t/b_n)dx^*dz^* \\ -\iint e^{\mathrm{i}((\xi_1-t/b_n)x^*+\xi_2 z^*)}g(x^*)\left\{f_{X^*,Z^*}(\cdot,z^*)\right\}^{\mathrm{ft}}(t/b_n)dx^*dz^* \end{array} \right\} \\ \times\iint e^{\mathrm{i}\{(\xi_1 b_n-t)\tilde{x}+\xi_2 b_n\tilde{z}\}}K(\tilde{x})K(\tilde{z})\,d\tilde{x}d\tilde{z} \end{array} \right\} dt$$

$$= \int \frac{\Pi_{\epsilon,j}(t)K^{\mathrm{ft}}(tb_n)K^{\mathrm{ft}}((\xi_1-t)b_n)K^{\mathrm{ft}}(\xi_2 b_n)}{6\pi} \times \left\{ \begin{array}{c} f^{\mathrm{ft}}(t)g^{\mathrm{ft}}_{X^*,Z^*}(\xi_1-t,\xi_2)-g^{\mathrm{ft}}(t)f^{\mathrm{ft}}_{X^*,Z^*}(\xi_1-t,\xi_2) \\ +f^{\mathrm{ft}}(\xi_1-t)g^{\mathrm{ft}}_{X^*,Z^*}(t,\xi_2)-g^{\mathrm{ft}}(\xi_1-t)f^{\mathrm{ft}}_{X^*,Z^*}(t,\xi_2) \end{array} \right\} dt,$$

where the first equality follows by Lemma 11 and the second equality follows from the change of variables $\tilde{x} = (x-x^*)/b_n$ and $\tilde{z} = (z-z^*)/b_n$, which, by using the dominant convergence theorem, implies

$$\sup_{\xi\in\Xi}\left|\hat{r}_n^{\epsilon}(a_j;\xi) - \int \frac{\Pi_{\epsilon,j}(t)}{6\pi}\left\{ \begin{array}{c} f^{\mathrm{ft}}(t)g^{\mathrm{ft}}_{X^*,Z^*}(\xi_1-t,\xi_2)-g^{\mathrm{ft}}(t)f^{\mathrm{ft}}_{X^*,Z^*}(\xi_1-t,\xi_2) \\ +f^{\mathrm{ft}}(\xi_1-t)g^{\mathrm{ft}}_{X^*,Z^*}(t,\xi_2)-g^{\mathrm{ft}}(\xi_1-t)f^{\mathrm{ft}}_{X^*,Z^*}(t,\xi_2) \end{array} \right\} dt\right| = o_p(1).$$

By similar arguments, we have

$$\sup_{\xi\in\Xi}\left|\hat{r}_n^{\nu}(a_j;\xi) - \frac{\Pi_{\nu,j}(\xi_2)}{6\pi}\int \left\{ \begin{array}{c} f^{\mathrm{ft}}(t)g^{\mathrm{ft}}_{X^*,Z^*}(\xi_1-t,\xi_2) \\ -g^{\mathrm{ft}}(t)f^{\mathrm{ft}}_{X^*,Z^*}(\xi_1-t,\xi_2) \end{array} \right\} dt\right| = o_p(1).$$

Then, by Theorem 3.9 of Chen and White (1998) and Assumption O', we have

$$\frac{3}{\sqrt{n}}\sum_{j=1}^{n}\{\hat{r}_n(a_j;\xi) - E[\hat{r}_n(a_j;\xi)]\} \Rightarrow \hat{G}^{\mathrm{os}}(\xi). \tag{A.25}$$

Combining (A.16), (A.19), (A.22), (A.24), and (A.25), if H_0 holds, (A.4) implies

$$\sqrt{n}\hat{T}_n(\xi) \Rightarrow \hat{G}^{\mathrm{os}}(\xi),$$

and if H_{1n} holds, (A.13) implies

$$\sqrt{n}\hat{T}_n(\xi) \Rightarrow \hat{G}^{\mathrm{os}}(\xi) + \mu_{\Delta}(\xi).$$

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

A.6. *Proof of Theorem 6.* This proof follows along the same lines as Theorem 5, except that (A.16) follows by (A.17), (A.18), (B.1), (B.2), and Assumptions S and S'. In addition, (A.19) follows by (A.20), Assumptions S and S', and Lemma 10. Finally, (A.21) follows by (A.23), Assumptions S and S', and Lemma 10.

Then, by Theorem 3.9 of Chen and White (1998) and Assumption S', we have

$$\frac{3}{\sqrt{n}}\sum_{j=1}^{n}\{\hat{r}_n(a_j;\xi)-E[\hat{r}_n(a_j;\xi)]\}\Rightarrow \hat{G}^{\mathrm{ss}}(\xi). \tag{A.26}$$

Combining (A.16), (A.19), (A.22), (A.24), and (A.26), if H_0 holds, (A.4) implies

$$\sqrt{n}\hat{T}_n(\xi)\Rightarrow \hat{G}^{\mathrm{ss}}(\xi),$$

and if H_{1n} holds, (A.13) implies

$$\sqrt{n}\hat{T}_n(\xi)\Rightarrow \hat{G}^{\mathrm{ss}}(\xi)+\mu_{\Delta}(\xi).$$

B. Lemmas

LEMMA 1. *For a random error η with $f_{\eta}^{ft}(t)\neq 0$, for all $t\in\mathbb{R}$, we have*

$$\int \mathbb{K}_{\eta}(u)\,e^{ib_nu\xi}\,du=\frac{K^{ft}(b_n\xi)}{f_{\eta}^{ft}(\xi)},$$

$$\int \mathbb{K}_{\eta}(u+c)\,\mathbb{K}_{\eta}(u)\,e^{ib_nu\xi}\,du=\frac{1}{2\pi}\int\frac{e^{-\mathrm{i}tc}K^{ft}(t)K^{ft}(-t+b_n\xi)}{f_{\eta}^{ft}(t/b_n)f_{\eta}^{ft}(-t/b_n+\xi)}dt.$$

Proof. The conclusion follows by

$$\int \mathbb{K}_{\eta}(u)\,e^{\mathrm{i}b_nu\xi}\,du=\int\left(\frac{1}{2\pi}\int e^{-\mathrm{i}(t-b_n\xi)u}du\right)\frac{K^{\mathrm{ft}}(t)}{f_{\eta}^{\mathrm{ft}}(t/b_n)}dt,$$

$$\int \mathbb{K}_{\eta}(u+c)\,\mathbb{K}_{\eta}(u)\,e^{\mathrm{i}b_nu\xi}\,du=\frac{1}{2\pi}\iint\left\{\begin{array}{c}\frac{1}{2\pi}\int e^{-\mathrm{i}(t_1+t_2-b_n\xi)u}du\\ \times\frac{e^{-\mathrm{i}t_1c}K^{\mathrm{ft}}(t_1)K^{\mathrm{ft}}(t_2)}{f_{\eta}^{\mathrm{ft}}(t_1/b_n)f_{\eta}^{\mathrm{ft}}(t_2/b_n)}\end{array}\right\}dt_1dt_2,$$

and $\int\delta(u-b)m(u)du=m(b)$ with the Dirac delta function $\delta(u)\equiv\frac{1}{2\pi}\int e^{-\mathrm{i}ut}dt$. ■

LEMMA 2. *Under Assumption D, for $\{k_1,k_2\}\in\{0,1\}$, we have*

$$E\left[Y^{k_1}\mathbb{K}_{\epsilon}\left(\frac{x-X}{b_n}\right)\mathbb{K}_{\nu}^{k_2}\left(\frac{z-Z}{b_n}\right)\right]=E\left[Y^{k_1}K\left(\frac{x-X^*}{b_n}\right)K^{k_2}\left(\frac{z-Z^*}{b_n}\right)\right].$$

Proof. As the proofs are similar, we focus on the case $k_1=k_2=1$. The result follows by

$$E\left[Y\mathbb{K}_{\epsilon}\left(\frac{x-X}{b_n}\right)\mathbb{K}_{\nu}\left(\frac{z-Z}{b_n}\right)\right]$$

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

$$= \frac{1}{4\pi^2}\iint e^{-\mathrm{i}(t_1 x + t_2 z)/b_n}\left\{E[Y|X,Z]f_{X,Z}\right\}^{\mathrm{ft}}(t_1/b_n, t_2/b_n)\frac{K^{\mathrm{ft}}(t_1)K^{\mathrm{ft}}(t_2)}{f_\epsilon^{\mathrm{ft}}(t_1/b_n)f_v^{\mathrm{ft}}(t_2/b_n)}dt_1 dt_2$$
$$= \frac{1}{4\pi^2}\iint e^{-\mathrm{i}(t_1 x + t_2 z)/b_n} g_{X^*,Z^*}^{\mathrm{ft}}(t_1/b_n, t_2/b_n)K^{\mathrm{ft}}(t_1)K^{\mathrm{ft}}(t_2)dt_1 dt_2$$
$$= E\left[YK\left(\frac{x-X^*}{b_n}\right)K\left(\frac{z-Z^*}{b_n}\right)\right],$$

where the second step uses Assumption D(ii) and the last step follows by the convolution theorem. ■

LEMMA 3. *Under Assumptions D and O/S, we have*

$$\iint\left\{\ b_n^{-3}E\left[YK\left(\tfrac{x-X^*}{b_n}\right)\right]E\left[K\left(\tfrac{x-X^*}{b_n}\right)K\left(\tfrac{z-Z^*}{b_n}\right)\right] - g(x)f_{X^*,Z^*}(x,z)\ \right\}e^{i(x,z)\xi}dxdz$$
$$= \tau_{1,n},$$
$$\iint\left\{\ b_n^{-3}E\left[K\left(\tfrac{x-X^*}{b_n}\right)\right]E\left[YK\left(\tfrac{x-X^*}{b_n}\right)K\left(\tfrac{z-Z^*}{b_n}\right)\right] - f(x)g_{X^*,Z^*}(x,z)\ \right\}e^{i(x,z)\xi}dxdz$$
$$= \tau_{2,n},$$

where $\tau_{k,n} = O(b_n^p)$ *under Assumption O and* $\tau_{k,n} = 0$ *under Assumption S, for* $k = 1, 2$.

Proof. As the proofs are similar, we focus on the statement concerning $\tau_{2,n}$. Define

$$J_{1,n}(x) = b_n^{-1}E\left[K\left(\frac{x-X^*}{b_n}\right)\right], \quad J_{2,n}(x,z) = b_n^{-2}E\left[YK\left(\frac{x-X^*}{b_n}\right)K\left(\frac{z-Z^*}{b_n}\right)\right].$$

Under Assumption S, f and g_{X^*,Z^*} are both infinitely differentiable and K is an infinite-order kernel, which implies $J_{1,n}(x) = f(x)$ and $J_{2,n}(x,z) = g_{X^*,Z^*}(x,z)$.

Under Assumption O, f and g_{X^*,Z^*} are both p-times continuously differentiable and K is a pth order kernel, which implies $J_{1,n}(x) = f(x) + b_n^p\Delta_{1,n}(x)$ and $J_{2,n}(x,z) = g_{X^*,Z^*}(x,z) + b_n^p\Delta_{2,n}(x,z)$, where $\Delta_{1,n}(x) = \frac{1}{p!}\int f^{(p)}(\bar{x})K(x^*)x^{*p}dx^*$ for some $\bar{x}$ lying between x and $x - b_n\tilde{x}^*$ and $\Delta_{2,n}(x,z) = \sum_{l=1}^{p}\frac{1}{l!(p-l)!}\iint g_{X^*,Z^*}^{(l,p-l)}(\bar{\bar{x}},\bar{\bar{z}})K(x^*)K(z^*)x^{*l}z^{*p-l}dx^*dz^*$ for some $(\bar{\bar{x}},\bar{\bar{z}})$ lying between (x,z) and $(x - b_n\tilde{x}^*, z - b_n\tilde{z}^*)$. The Lipschitz condition on $f^{(p)}$ and $g_{X^*,Z^*}^{(l,p-l)}$ as assumed in Assumption O(i) implies

$$|\Delta_{1,n}(x)| \le \frac{1}{p!}\{\mu_K^p|f^{(p)}(x)| + b_n\mu_K^{p+1}|m_{1,p}(x)|\},$$
$$|\Delta_{2,n}(x,z)| \le \sum_{l=1}^{p}\frac{1}{l!\,(p-l)!}\{\mu_K^l\mu_K^{p-l}|g_{X^*,Z^*}^{(l,p-l)}(x,z)|$$
$$+ b_n(\mu_K^{l+1}\mu_K^{p-l} + \mu_K^l\mu_K^{p+1-l})|m_{4,p}(x,z)|\},$$

where $\mu_K^d = \int K(u)|u|^d du$. Then, the conclusion follows by

$$\iint\left\{J_{1,n}(x)J_{2,n}(x,z) - f(x)g_{X^*,Z^*}(x,z)\right\}e^{\mathrm{i}(x,z)\xi}dxdz$$

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

$$= \iint \left\{ \begin{array}{c} \{J_{1,n} - f\}(x) g_{X^*,Z^*}(x,z) \\ + f(x)\{J_{2,n} - g_{X^*.Z^*}\}(x,z) \\ + \{J_{1,n} - f\}(x)\{J_{2,n} - g_{X^*.Z^*}\}(x,z) \end{array} \right\} e^{\mathrm{i}(x,z)\xi} dxdz,$$

and the boundedness and the integrability of f, $f^{(p)}$, g_{X^*,Z^*}, $g_{X^*,Z^*}^{(l,p-l)}$, $m_{1,p}$, and $m_{4,p}$ as assumed in Assumption O(i). ■

LEMMA 4.

$$\mathbb{K}_\epsilon(x) = \begin{cases} \sum_{h=0}^{\alpha} c_h^{\mathrm{os}} (i/b_n)^h K^{(h)}(x) & \textit{under Assumption O} \\ \sum_{h=0}^{\infty} \frac{\mu^h}{h!} (i/b_n)^{2h} K^{(2h)}(x) & \textit{under Assumption S} \end{cases}.$$

Proof. Under Assumption O, the conclusion follows by Lemma 3.1 of Fan (1995). Under Assumption S, the conclusion follows by Lemma 5 of Dong et al. (2020). ■

LEMMA 5. *Under Assumption O, we have*

$$\int \mathbb{K}_\epsilon(x) x^l e^{\mathrm{i}b_n x\xi_1} dx = \mathbb{I}\{l \le \alpha\} c_l^{\mathrm{os}}(\xi_1) l! b_n^{-l} + \mathrm{i}b_n \sum_{h=0}^{\alpha} \frac{c_h^{\mathrm{os}} (-1)^h \xi_1^{h+1}}{(h+1)!} \int K^{(h)}(x) x^l \bar{x}_h^{h+1} dx,$$

for $l = 0, 1, \ldots, p$, where $\bar{x}_h$ lying between 0 and x, for $h = 1, \ldots, \alpha$.

Proof. Note that

$$\int \mathbb{K}_\epsilon(x) x^l e^{\mathrm{i}b_n x\xi_1} dx = \sum_{h=0}^{\alpha} c_h^{\mathrm{os}} (\mathrm{i}/b_n)^h \int K^{(h)}(x) x^l e^{\mathrm{i}b_n x\xi_1} dx$$

$$= \sum_{h=0}^{\alpha} c_h^{\mathrm{os}} (\mathrm{i}/b_n)^h \left\{ \sum_{k=0}^{h} \frac{(\mathrm{i}b_n\xi_1)^k}{k!} \int K^{(h)}(x) x^{l+k} dx + \frac{(\mathrm{i}b_n\xi_1)^{h+1}}{(h+1)!} \int K^{(h)}(x) x^l \bar{x}_h^{h+1} dx \right\},$$

where the first equality follows by Lemma 4 under Assumption O and the second equality follows by expanding $e^{\mathrm{i}b_n x\xi_1}$ around $x = 0$ to the hth order and $\bar{x}_h$ lies between 0 and x, for $h = 1, \ldots, \alpha$. Then, the conclusion follows by using $\int K^{(h)}(u) u^{l+k} du = \mathbb{I}\{h = l+k\} h! (-1)^h$, for $h \ge l$ and $k = 0, 1, \ldots, \alpha$, which follows by integration by parts and the properties of the kernel function K as assumed in Assumption O(iii). ■

LEMMA 6. *Under Assumption S, we have*

$$\int \mathbb{K}_\epsilon(x) x^l e^{\mathrm{i}b_n \xi_1 x} dx = c_l^{\mathrm{ss}}(\xi_1) l! b_n^{-l}.$$

Proof. Note that

$$\int \mathbb{K}_\epsilon(x) x^l e^{\mathrm{i}b_n \xi_1 x} dx = \sum_{h=0}^{\infty} \frac{\mu^h}{h!} (\mathrm{i}/b_n)^{2h} \int K^{(2h)}(x) x^l e^{\mathrm{i}b_n \xi_1 x} dx$$

$$= \sum_{h=0}^{\infty} \frac{\mu^h}{h!} (\mathrm{i}/b_n)^{2h} \sum_{k=0}^{\infty} \frac{(\mathrm{i}b_n\xi_1)^k}{k!} \int K^{(2h)}(x) x^{l+k} dx,$$

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

where the first equality follows by Lemma 4 under Assumption S and the second equality follows by expanding $e^{ib_n x\xi_1}$ around $x=0$ to an infinite order. Then, the conclusion follows by $\int K^{(2h)}(u)u^{l+k}du = \mathbb{I}\{2h=l+k\}(l+k)!\,(-1)^{l+k}$, which follows by integration by parts and the properties of the kernel function K as assumed in Assumption S(iii). ■

LEMMA 7. *There exist constants $A, v \geq 0$ independent of n such that*

$$\sup_{Q} N(\mathcal{F}_n^{(r)}, |\cdot|_{Q,2}, \delta) \leq \left(\frac{A|F_n^{(r)}|_{Q,2}}{\delta}\right)^v = O_p(1),$$

for all $0 < \delta \leq 1$, and where $\sup_Q$ is taken over all finitely discrete distributions on $\mathbb{R}^3$.

Proof. See Lemma 1 of Kato and Sasaki (2018). See also Corollary A.1 of Chernozhukov, Chetverikov, and Kato (2014). ■

LEMMA 8. *Under Assumptions D and R, for $\eta \in \{\epsilon, v\}$, we have*

$$\sup_{|t|\leq \varrho_n} |\hat{\zeta}_\eta(t) - \zeta_\eta(t)| = O_p\left(n^{-1/2}\log(\varrho_n)\right).$$

Proof. Note that $|\hat{\zeta}_\eta(t) - \zeta_\eta(t)| \leq |\frac{1}{n}\sum_{j=1}^n e^{\mathrm{i}t(\eta_j - \eta_j^r)} - \zeta_\eta(t)|$. The conclusion then follows by $E|\eta|^{2+\varsigma} < \infty$, for some $\varsigma > 0$ (Assumption R(ii)), and Kurisu and Otsu (2021, Lem. 1). ■

LEMMA 9. *Under Assumptions D and R, for $\eta \in \{\epsilon, v\}$, we have*

$$\sup_{|t|\leq \varrho_n} |\hat{\Pi}_\eta(t)| = O_p\left(\frac{\log(\varrho_n)}{n^{1/2}\{\inf_{|t|\leq\varrho_n}|f_\eta^{ft}(t)|\}^2}\right).$$

Moreover, if $n^{-1/2}\log(\varrho_n)\left(\inf_{|t|\leq\varrho_n}|f_\eta^{ft}(t)|\right)^{-2} \to 0$, we also have

$$\sup_{|t|\leq \varrho_n} |\hat{\Pi}_\eta^{\mathrm{res}}(t)| = O_p\left(\frac{\log(\varrho_n)^2}{n\{\inf_{|t|\leq\varrho_n}|f_\eta^{ft}(t)|\}^4}\right).$$

Proof. The first statement follows by Lemma 8 and $\sup_{|t|\leq\varrho_n}|\hat{\Pi}_\eta(t)| = \frac{\sup_{|t|\leq\varrho_n}|\hat{\zeta}_\eta(t)-\zeta_\eta(t)|}{2\{\inf_{|t|\leq\varrho_n}|f_\eta^{\mathrm{ft}}(t)|\}^2}$. For the second statement, note that

$$\begin{aligned}
\sup_{|t|\leq\varrho_n}|\hat{\Pi}_\eta^{\mathrm{res}}(t)| &\leq \sup_{|t|\leq\varrho_n}\left|\frac{2(\zeta_\eta(t)-\hat{\mu}_s(t))(|\zeta_\eta(t)|^{1/2}-|\hat{\mu}_s(t)|^{1/2})}{\zeta_\eta(t)|\hat{\zeta}_\eta(t)|^{1/2}}\right| \\
&= \sup_{|t|\leq\varrho_n}\left|\frac{2(\hat{\zeta}_\eta(t)-\zeta_\eta(t))^2}{\zeta_\eta(t)|\hat{\zeta}_\eta(t)|^{1/2}(|\zeta_\eta(t)|^{1/2}+|\hat{\zeta}_\eta(t)|^{1/2})}\right| \\
&\leq \sup_{|t|\leq\varrho_n}\left|\frac{2(\hat{\zeta}_\eta(t)-\zeta_\eta(t))^2}{\zeta_\eta(t)\{(\hat{\zeta}_\eta(t)-\zeta_\eta(t))+\zeta_\eta(t)\}}\right|
\end{aligned}$$

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

$$\leq \frac{2(\sup_{|t|\leq \varrho_n} |\hat{\zeta}_\eta(t) - \zeta_\eta(t)|)^2}{\inf_{|t|\leq \varrho_n} \zeta_\eta(t) \inf_{|t|\leq \varrho_n} \{(\hat{\zeta}_\eta(t) - \zeta_\eta(t)) + \zeta_\eta(t)\}}.$$

The conclusion then follows by Lemma 8 and $n^{-1/2}\log(\varrho_n)(\inf_{|t|\leq \varrho_n} |f_\eta^{\mathrm{ft}}(t)|)^{-2} \to 0$. ■

Immediate results following Lemma 9 are

$$\sup_{|t|\leq 2b_n^{-1}} |\hat{\Pi}_\epsilon(t)| = \begin{cases} O_p\left(n^{-1/2}\log(1/b_n) b_n^{-2\alpha}\right) & \text{under Assumption O} \\ O_p\left(n^{-1/2}\log(1/b_n) e^{8\mu b_n^{-2}}\right) & \text{under Assumption S} \end{cases}, \tag{B.1}$$

$$\sup_{|t|\leq 2b_n^{-1}} |\hat{\Pi}_\epsilon^{\mathrm{res}}(t)| = \begin{cases} O_p\left(n^{-1}\log(1/b_n)^2 b_n^{-4\alpha}\right) & \text{under Assumption O} \\ O_p\left(n^{-1}\log(1/b_n)^2 e^{16\mu b_n^{-2}}\right) & \text{under Assumption S} \end{cases}. \tag{B.2}$$

LEMMA 10. *Under Assumption D, we have*

$$\sup_{|t|\leq 2b_n^{-1}} |\Pi_{\epsilon,l}(t)| = \begin{cases} O_p\left(b_n^{-2\alpha}\right) & \textit{under Assumption O} \\ O_p\left(e^{8\mu b_n^{-2}}\right) & \textit{under Assumption S} \end{cases}.$$

Proof. The conclusion follows by

$$\sup_{|t|\leq 2b_n^{-1}} |\Pi_{\epsilon,l}(t)| \leq \frac{(\sup_{|t|\leq 2b_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)|)^2 + 1}{2(\inf_{|t|\leq 2b_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)|)^2} = O\left(\frac{1}{(\inf_{|t|\leq 2b_n^{-1}} |f_\epsilon^{\mathrm{ft}}(t)|)^2}\right).$$

■

LEMMA 11. *Under Assumption D, we have*

$$E\left[e^{-it\left(\frac{x-X^*}{b_n}\right)}\right] = e^{-itx/b_n} f^{ft}(t/b_n),\ E\left[Ye^{-it\left(\frac{x-X^*}{b_n}\right)}\right] = e^{-itx/b_n} g^{ft}(t/b_n),$$

$$E\left[e^{-it\left(\frac{x-X^*}{b_n}\right)} K\left(\frac{z-Z^*}{b_n}\right)\right] = e^{-itx/b_n} \int \left\{g_{X^*,Z^*}(\cdot,z^*)\right\}^{ft}(t/b_n) K\left(\frac{z-z^*}{b_n}\right) dz^*,$$

$$E\left[Ye^{-it\left(\frac{x-X^*}{b_n}\right)} K\left(\frac{z-Z^*}{b_n}\right)\right] = e^{-itx/b_n} \int \left\{f_{X^*,Z^*}(\cdot,z^*)\right\}^{ft}(t/b_n) K\left(\frac{z-z^*}{b_n}\right) dz^*.$$

Proof. As the proofs are similar, we focus on the last statement, which follows by

$$\begin{aligned} E\left[e^{-\mathrm{i}t\left(\frac{x-X^*}{b_n}\right)} K\left(\frac{z-Z^*}{b_n}\right)\right] &= \iint e^{-\mathrm{i}t\left(\frac{x-x^*}{b_n}\right)} f_{X^*,Z^*}(x^*,z^*) K\left(\frac{z-z^*}{b_n}\right) dx^* dz^* \\ &= e^{-\mathrm{i}tx/b_n} \int \left\{ \int e^{\mathrm{i}tx^*/b_n} f_{X^*,Z^*}(x^*,z^*) dx^* \right\} K\left(\frac{z-z^*}{b_n}\right) dz^* \\ &= e^{-\mathrm{i}tx/b_n} \int \left\{f_{X^*,Z^*}(\cdot,z^*)\right\}^{\mathrm{ft}}(t/b_n) K\left(\frac{z-z^*}{b_n}\right) dz^*. \end{aligned}$$

■

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

REFERENCES

Adusumilli, K., D. Kurisu, T. Otsu & Y. J. Whang (2020) Inference on distribution functions under measurement error. *Journal of Econometrics* 215(1), 131–164.

Ahn, H. & J. L. Powell (1993) Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics* 58(1–2), 3–29.

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

Bierens, H. J. (1990) A consistent conditional moment test of functional form. *Econometrica* 58(6), 1443–1458.

Bierens, H. J. & W. Ploberger (1997) Asymptotic theory of integrated conditional moment tests. *Econometrica* 65(5), 1129–1151.

Bissantz, N., L. Dümbgen, H. Holzmann & A. Munk (2007) Non-parametric confidence bands in deconvolution density estimation. *Journal of the Royal Statistical Society: Series B* 69(3), 483–506.

Butucea, C. & M. L. Taupin (2008) New M-estimators in semi-parametric regression with errors in variables. *Annales de l'IHP Probabilités et Statistiques* 44(3), 393–421.

Carroll, R. J. & P. Hall (1988) Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association* 83(404), 1184–1186.

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

Chen, X. & H. White (1998) Central limit and functional central limit theorems for Hilbert-valued dependent heterogeneous arrays with applications. *Econometric Theory* 14(2), 260–284.

Chernozhukov, V., D. Chetverikov & K. Kato (2014) Gaussian approximation of suprema of empirical processes. *Annals of Statistics* 42(4), 1564–1597.

Chernozhukov, V., D. Chetverikov & K. Kato (2015) Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probability Theory and Related Fields* 162(1–2), 47–70.

Chernozhukov, V., D. Chetverikov & K. Kato (2016) Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings. *Stochastic Processes and Their Applications* 126(12), 3632–3651.

Delaigle, A. & P. Hall (2016) Methodology for non-parametric deconvolution when the error distribution is unknown. *Journal of the Royal Statistical Society: Series B* 78(1), 231–252.

Delaigle, A., P. Hall & F. Jamshidi (2015) Confidence bands in non-parametric errors-in-variables regression. *Journal of the Royal Statistical Society: Series B* 77(1), 149–169.

Delaigle, A., P. Hall & A. Meister (2008) On deconvolution with repeated measurements. *Annals of Statistics* 36(2), 665–685.

Delaigle, A., & Gijbels, I. (2004) Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Annals of the Institute of Statistical Mathematics*, 56(1), 19–47.

Delaigle, A., & Meister, A. (2007) Nonparametric regression estimation in the heteroscedastic errors-in-variables problem. *Journal of the American Statistical Association*, 102(480), 1416–1426.

Delgado, M. A. & W. González-Manteiga (2001) Significance testing in nonparametric regression based on the bootstrap. *Annals of Statistics* 29, 1469–1507.

Dong, H., T. Otsu, & L. Taylor (2020) Average derivative estimation under measurement error. *Econometric Theory*, first published online 13 November 2020. https://doi.org/10.1017/S0266466620000432.

Dudley, R. M. (2002) *Real Analysis and Probability.* Cambridge University Press.

Fan, J. & Y. K. Truong (1993) Nonparametric regression with errors in variables. *Annals of Statistics* 21(4), 1900–1925.

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

Fan, Y., & Li, Q. (2000) Consistent model specification tests: Kernel-based tests versus Bierens' ICM tests. *Econometric Theory*, 1016–1041.

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

Fisher, R. A. (1929) The statistical method in psychical research. *Proceedings of the Society for Psychical Research* 39, 189–192.
Giné, E. & R. Nickl (2016) *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge University Press.
González-Manteiga, W. & R. M. Crujeiras (2013) An updated review of goodness-of-fit tests for regression models. *Test* 22(3), 361–411.
Hall, P. & Y. Ma (2007a) Semiparametric estimators of functional measurement error models with unknown error. *Journal of the Royal Statistical Society: Series B* 69(3), 429–446.
Hall, P. & Y. Ma (2007b) Testing the suitability of polynomial models in errors-in-variables problems. *Annals of Statistics* 35(6), 2620–2638.
Hall, P. & A. Meister (2007) A ridge-parameter approach to deconvolution. *Annals of Statistics* 35(4), 1535–1558.
Härdle, W. & E. Mammen (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21(4), 1926–1947.
Hu, Y. & G. Ridder (2012) Estimation of nonlinear models with mismeasured regressors using marginal information. *Journal of Applied Econometrics* 27(3), 347–385.
Kato, K. & Y. Sasaki (2018) Uniform confidence bands in deconvolution with unknown error distribution. *Journal of Econometrics* 207(1), 129–161.
Kato, K. & Y. Sasaki (2019) Uniform confidence bands for nonparametric errors-in-variables regression. *Journal of Econometrics* 213(2), 516–555.
Kotlarski, I. (1967) On characterizing the gamma and the normal distribution. *Pacific Journal of Mathematics* 20(1), 69–76.
Kurisu, D. & T. Otsu (2021) On uniform convergence of deconvolution estimator from repeated measurements. *Econometric Theory*, first published online 25 January 2021. https://doi.org/10.1017/S0266466620000572.
Lavergne, P. & Q. H. Vuong (2000) Nonparametric significance testing. *Econometric Theory* 16(4), 576–601.
Li, T. & Q. H. Vuong (1998) Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis* 65(2), 139–165.
Ma, Y. & R. Li (2010) Variable selection in measurement error models. *Bernoulli* 16(1), 274–300.
McMurry, T. L. & D. N. Politis (2004) Nonparametric regression with infinite order flat-top kernels. *Journal of Nonparametric Statistics* 16, 549–562.
Meister, A. (2009) *Deconvolution Problems in Nonparametric Statistics*. Springer.
Otsu, T. & L. Taylor (2020) Specification testing for errors-in-variables models. *Econometric Theory*, first published online 19 June 2020. https://doi.org/10.1017/S0266466620000262.
Powell, J. L., J. H. Stock & T. M. Stoker (1989) Semiparametric estimation of index coefficients. *Econometrica* 57(6), 1403–1430.
Robinson, P. M. (1989) Hypothesis testing in semiparametric and nonparametric models for econometric time series. *The Review of Economic Studies* 56(4), 511–534.
Schennach, S. M. (2004) Estimation of nonlinear models with measurement error. *Econometrica* 72(1), 33–75.
Schennach, S. M. (2016) Recent advances in the measurement error literature. *Annual Review of Economics* 8, 341–377.
Stefanski, L. A. & R. J. Carroll (1990) Deconvolving kernel density estimators. *Statistics* 21(2), 169–184.
Stinchcombe, M. B. & H. White (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14(3), 295–325.
Stute, W. (1997) Nonparametric model checks for regression. *Annals of Statistics* 25(2), 613–641.
Su, L. & H. White (2008) A nonparametric Hellinger metric test for conditional independence. *Econometric Theory* 24(4), 829–864.
Van Der Vaart, A. W. & J. A. Wellner (1996) *Weak Convergence and Empirical Processes*. Springer.
Van Es, A. J. & H. W. Uh (2005) Asymptotic normality for kernel type deconvolution estimators. *Scandinavian Journal of Statistics* 32, 467–483.

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press

Zhang, X., H. Wang, Y. Ma & R. J. Carroll (2017) Linear model selection when covariates contain errors. *Journal of the American Statistical Association* 112(520), 1553–1561.
Zhu, L. & H. Cui (2003) A semi-parametric regression model with errors in variables. *Scandinavian Journal of Statistics* 30(2), 429–442.

https://doi.org/10.1017/S0266466621000220 Published online by Cambridge University Press