Robust Inference in Deconvolution[*]

Kengo Kato[†] Yuya Sasaki[‡] Takuya Ura[§]

Abstract

Kotlarski's identity has been widely used in applied economic research based on repeated-measurement or panel models with latent variables. However, how to conduct inference for these models has been an open question for two decades. This paper addresses this open problem by constructing a novel confidence band for the density function of a latent variable in repeated measurement error model. The confidence band builds on our finding that we can rewrite Kotlarski's identity as a system of linear moment restrictions. Our approach is robust in that we do not require the completeness. The confidence band controls the asymptotic size uniformly over a class of data generating processes, and it is consistent against all fixed alternatives. Simulation studies support our theoretical results.

Keywords: deconvolution, measurement error, robust inference, uniform confidence band.

1 Introduction

Empirical researchers are often interested in recovering features of unobserved variables in economic models. With an availability of repeated measurements or panel data, Kotlarski's identity (Kotlarski, 1967) – see also Rao (1992) – is one of the most popular tools used to identify probability density functions of unobserved latent variables in additive-error models. Examples of research topics that use Kotlarski's identity include, but are not limited to, empirical auctions (e.g., Li, Perrigne, and Vuong, 2000; Krasnokutskaya, 2011), education and labor economics (e.g., Carneiro, Hansen, and Heckman, 2003; Cunha, Heckman, and Navarro, 2005; Cunha, Heckman, and Schennach, 2010;

[*]The first arXiv date: August 28, 2018. This paper supersedes the previous version which was entitled "Inference Based on Kotlsarki's Identity." K. Kato is partially supported by NSF grants DMS-1952306 and DMS-2014636. We benefited from very useful comments by Chris Taber, Andres Santos, and two anonymous referees. We also would like to thank numerous scholars, seminar participants at Aarhus University, Duke University, KU Leuven, Tinbergen Institute, Universidad Carlos III de Madrid, University of California Berkeley, University of California Davis, University of California Los Angeles, University of California San Diego, University of Illinois at Urbana-Champaign, University of Oxford, University of Texas Austin, and University of Wisconsin Madison, and conference participants at AMES 2019, NASM 2019, New York Camp Econometrics XVI, and 2019 Triangle Econometrics Conference for very helpful comments. The usual disclaimer applies.

[†]K. Kato: Department of Statistics and Data Science, Cornell University, 1194 Comstock Hall, Ithaca, NY 14853. Email: kk976@cornell.edu

[‡]Y. Sasaki: Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235-1819. Email: yuya.sasaki@vanderbilt.edu

[§]T. Ura: Department of Economics, University of California, Davis, 1151 Social Sciences and Humanities, Davis, CA 95616. Email: takura@ucdavis.edu

Arcidiacono, Aucejo, Fang, and Spenner, 2011; Bonhomme and Sauder, 2011; Kennan and Walker, 2011; Taber and Vejlin, 2020), and earnings dynamics (e.g., Bonhomme and Robin, 2010; Botosaru and Sasaki, 2018; Hu, Moffitt, and Sasaki, 2019). In these applications, researchers are interested in identifying the probability density function f_X of a latent variable X among others. The variable X of interest is not observed in data, but two measurements (Y_1, Y_2) are available in data with classical errors, $U_1 = Y_1 - X$ and $U_2 = Y_2 - X$. Kotlarski's identity is a nonparametric closed-form identifying restriction for the probability density function f_X of X implied by this setup.

The existing econometric literature on Kotlarski's identity focuses on identification and consistent estimation of f_X and related objects (e.g., Li and Vuong, 1998; Li, 2002; Schennach, 2004a,b, 2008; Bonhomme and Robin, 2010; Evdokimov, 2010; Zinde-Walsh, 2014; Song, Schennach, and White, 2015; Firpo, Galvao, and Song, 2017) – also see surveys on this literature by Chen, Hong, and Nekipelov (2011) and Schennach (2016). On the other hand, satisfactory inference methods for f_X are missing in this literature – in fact, even the *sharp* rate of convergence is unknown for the estimators based on Kotlarski's identity under *unrestrictive* assumptions,[1] and hence a limit distribution result is unavailable under such assumptions. Indeed some empirical papers implement nonparametric bootstrap without a theoretical guarantee. Other empirical papers, including many of those listed above, often consider parametric estimation and parametric inference given the ill-posedness of the deconvolution problem as well as the lack of available inference methods. In light of the current unavailability of theoretically supported methods of inference, we propose a method of inference for f_X. Furthermore, we propose a method of inference that is robust against possible identification failure.

This paper develops a confidence band for f_X. Our construction of confidence bands works as follows. First, we derive linear complex-valued moment restrictions by modifying the proof of Kotlarski's identity (Kotlarski, 1967) – see also Rao (1992). Second, we let the Hermite orthogonal sieve (cf. Chen, 2007) approximate unknown probability density functions. Third, for a given sieve dimension and for a given class of probability density functions, we compute a bias bound for the linear complex-valued moment restrictions, and slack the linear complex-valued moment restrictions by this bias bound. Fourth, applying Chernozhukov, Chetverikov, and Kato (2018), we compute the supremum of the self-normalized process of the slacked linear complex-valued moment restrictions as the test statistic for each point in a set of sieve coefficients. Fifth, inverting this test statistic in the spirit of Anderson and Rubin (1949) yields a confidence set of sieve approximations to possible probability density functions. Sixth, for a given sieve dimension and for a given class for probability density functions, we compute a bias bound for sieve approximations of probability density functions, and the desired confidence band is obtained by uniformly enlarging the set of sieve approximations by this bias bound.

[1]By the *unrestrictive* assumptions, we specifically mean assumptions that do not impose either known error distribution or symmetric error distribution. Under such settings, the existing convergence rates are not shown to be sharp to the best of our knowledge. A recent paper by Kurisu and Otsu (2019) obtains improved convergence rates compared with those of Li and Vuong (1998) and Bonhomme and Robin (2010), although they are not shown to be sharp either.

The process of identifying f_X in additive measurement error models is called *deconvolution* – for solving convolution integral equations. There are a number of existing papers on nonparametric inference in deconvolution. Bissantz, Dümbgen, Holzmann, and Munk (2007), Bissantz and Holzmann (2008), van Es and Gugushvili (2008), Lounici and Nickl (2011), and Schmidt-Hieber, Munk, and Dümbgen (2013) develop uniform confidence bands for f_X under the assumption of known error distributions.[2] In most economic applications, however, it is not plausible to assume that the error distributions are known. More recently, Kato and Sasaki (2018) and Adusumilli, Kurisu, Otsu, and Whang (2019) develop uniform confidence bands for f_X and the distribution function, respectively, without assuming that the error distributions are known, but they both assume that at least one error distribution is symmetric.[3] Kotlarski's identity is a powerful device for new identification results which require neither the known error distribution assumption nor the symmetric error distribution assumption. This useful feature attracts many economic applications including those listed above, but no econometrician has developed a method of inference in this framework for twenty years ever since its first introduction by Li and Vuong (1998) until our present paper.

It is not surprising that such an inference method has been missing for long in the literature, given the technical difficulties of the problem. Deconvolution is an ill-posed inverse problem, and inference under this problem is known to be challenging – see Bissantz et al. (2007); Bissantz and Holzmann (2008); Lounici and Nickl (2011); Horowitz and Lee (2012); Hall and Horowitz (2013); Schmidt-Hieber et al. (2013); Babii (2018); Chen and Christensen (2018); Kato and Sasaki (2018, 2019); Adusumilli et al. (2019) for existing papers developing confidence bands in ill-posed inverse problems for example. We take a robust inference approach à la Anderson and Rubin (1949), and directly work with the moment restrictions based on Kotlarski's identity. A positive side product of taking this approach is that we do not need to assume the non-vanishing characteristic functions (i.e., we do not need the completeness), which is commonly assumed for nonparametric identification or inversion.

It is also worth mentioning that we chose to use the Hermite orthogonal sieve among other sieves in this paper. The Hermite orthogonal sieve has been in fact already known in the literature to be useful to approximate "smooth density with unbounded support" (Chen, 2007) – also see her discussion of Gallant and Nychka (1987) therein. In addition to this known advantage, we also find this sieve particularly useful for the deconvolution problem. Note that the deconvolution problem involves applications of the Fourier transform operation and the inverse Fourier transform operation. To our convenience, the Hermite functions are eigen-functions of the Fourier transform operator. While we deal with simultaneous restrictions in terms of density and characteristic functions, we can use the Hermite orthogonal sieve to approximate both the density and characteristic functions

[2]These paper are based on the literature on deconvolution under known error distribution (e.g., Carroll and Hall, 1988; Stefanski and Carroll, 1990; Fan, 1991b; Carrasco and Florens, 2011). Fan (1991a) develops a point-wise asymptotic inference result in this framework.

[3]These paper are based on the literature on deconvolution under unknown error distribution with auxiliary data or symmetric error distributions (e.g., Diggle and Hall, 1993; Horowitz and Markatou, 1996; Neumann and Hössjer, 1997; Efromovich, 1997; Delaigle, Hall, and Meister, 2008; Johannes, 2009; Comte and Lacour, 2011; Delaigle and Hall, 2015).

without having to apply the Fourier transform or the Fourier inverse because of the eigen-function property. This convenient property saves computational time and resources as costly numerical integration within each iteration of a numerical optimization routine would be necessary if any other sieve were used.

The rest of the paper is organized as follows. Section 2 derives linear complex-valued moment restrictions based on Kotlarski's identity. Section 3 presents how to construct the confidence band for f_X based on the Hermite orthogonal basis. Section 4 presents asymptotic properties of the confidence band. Section 5 applies our proposed method to wildcat auctions, in which we investigate the distribution of the ex post values, $\exp(X)$, of the mineral rights. Section 6 illustrates simulation studies. The paper concludes in Section 7. All mathematical derivations and details are collected in the appendix.

2 Linear Complex-Valued Moment Restrictions

Consider the repeated measurement model

$$\begin{cases} Y_1 = X + U_1, \\ Y_2 = X + U_2, \end{cases} \tag{1}$$

where Y_1 and Y_2 are observed, but none of X, U_1, or U_2 is observed. We are interested in making inference on the probability density function f_X of X. We equip this model with the following assumption.

Assumption 1.

(i) X, U_1, and U_2 are continuous random variables with finite first moments.

(ii) U_1 has mean zero, and X, U_1, and U_2 are mutually independent.

This assumption is standard in the literature on identification and estimation based on Kotlarski's identity (e.g., Li and Vuong, 1998). In fact, the existing literature imposes an additional assumption, namely the identification condition (non-vanishing characteristic function or the completeness) – see Lemma 1 ahead for a specific condition. We do not invoke such an identification assumption for the purpose of identification-robust inference – see Remark 1 ahead for further details. Note also that this assumption (or any additional assumption that we make ahead) does not require a large support for either X, U_1 or U_2. This point reassures the identification-robustness.

We now fix basic notations. In what follows, $\mathbb{E}_P$ and $\mathbb{V}_P$ denote the expectation and variance operators, respectively, with respect to a joint distribution P of (Y_1, Y_2). Analogously, $\mathbb{E}_n$ and $\mathbb{V}_n$ denote the expectation and variance operators, respectively, with respect the empirical distribution of n independent copies of (Y_1, Y_2). We let $i = \sqrt{-1}$ denote the imaginary unit. For the set of absolutely integrable functions, $\mathcal{L}^1$, we define the Fourier transform $\mathcal{F}$ on $\mathcal{L}^1$ by $[\mathcal{F}f](t) = \int_{-\infty}^{\infty} \exp(itx)f(x)dx$, and its inverse transform is $[\mathcal{F}^{-1}\phi](x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-itx)\phi(t)dt$ – see Folland

(2007). In light of Assumption 1 (i), we let f_X, f_{U_1}, and f_{U_2} denote the density functions of X, U_1, and U_2, respectively. Further, we denote the characteristic functions of them by $\phi_X = \mathcal{F} f_X$, $\phi_{U_1} = \mathcal{F} f_{U_1}$, and $\phi_{U_2} = \mathcal{F} f_{U_2}$. We first review the existing result of the identification.

Lemma 1 (Kotlarski's Identity)**.** *For every joint distribution P of (Y_1, Y_2) satisfying Assumption 1 for (1) and $\mathbb{E}_P[\exp(itY_2)] \neq 0$ for all $t \in \mathbb{R}$, we have*

$$\phi_X(t) = \exp\left(\int_0^t \frac{i\mathbb{E}_P[Y_1 \exp(i\tau Y_2)]}{\mathbb{E}_P[\exp(i\tau Y_2)]} d\tau \right). \tag{2}$$

This lemma presents Kotlarski's identity due to Kotlarski (1967) – see also Rao (1992). Since it is stated as a lemma, Kotlarski's identity is also known as Kotlarski's lemma or the lemma of Kotlarski in the econometrics literature. Li and Vuong (1998) first introduced it into econometrics and statistics, followed by a series of extensions (Li, 2002; Schennach, 2004a; Bonhomme and Robin, 2010; Evdokimov, 2010). Some of these extensions relax the assumptions for identification and estimation in various ways. We do not need to rely on the prototypical assumptions for our purpose of inference, even though they are stated in Lemma 1 for convenience of a concise review. Lemma 1 shows that the characteristic function ϕ_X of X is explicitly identified by the joint distribution of (Y_1, Y_2). Under the additional assumption of absolutely integrable characteristic function ϕ_X, the formula $f_X = \mathcal{F}^{-1}\phi_X$ in turn yields the identification of the probability density function f_X of X.

Uniform convergence rates for the estimator of f_X based on Kotlarski's identity are discovered in the existing literature (Li and Vuong, 1998; Li, 2002; Schennach, 2004a; Bonhomme and Robin, 2010; Evdokimov, 2010), but the *sharp* rates under *unrestrictive* assumptions are still unknown – see Footnote 1. In particular, limit distribution results under such assumptions are still unknown in the existing literature. This paper does not aim to derive a non-degenerate limit distribution for any estimator, but it aims to conduct an inference on f_X. With this said, our proposed inference approach does not rely on an explicit identifying formula. We argue that rewriting Kotlarski's identity in terms of moment restrictions suffices and serves even more conveniently for the sake of conducting inference. Here are the moment restrictions obtained from rewriting Kotlarski's identity.

Theorem 1 (Linear Complex-Valued Moment Restrictions)**.** *For every joint distribution P of (Y_1, Y_2) satisfying Assumption 1 for (1),*

$$\mathbb{E}_P\left[\left(iY_1 \phi_X(t) - \phi_X^{(1)}(t) \right) \exp(itY_2) \right] = 0 \tag{3}$$

holds for every real t.

A proof is provided in Appendix A.1.

Remark 1. *Taking a few more steps beyond the claim in Theorem 1 will lead us to the identification result of Lemma 1 under the additional assumption of the invertibility or non-vanishing character-*

istic functions (also known as the completeness) – see D'Haultfoeuille (2011).[4] *Specifically, under the completeness, (3) has the unique solution equal to (2). For the purpose of inference, however, it is not essential to solve the inverse problem, and thus we stop short of obtaining the explicit formula (2), and only use the moment condition (3). This idea is analogous to that of Santos (2011, 2012), where robust inference for functional parameters is conducted without assuming the completeness. Since the assumption of completeness is not testable in general (Canay, Santos, and Shaikh, 2013), we propose the robust inference approach based on (3) instead of (2).*

Remark 2. *In addition to ϕ_X, Kotlarski's identity also identities ϕ_{U_1} and ϕ_{U_2} by ϕ_{Y_1}/ϕ_X and ϕ_{Y_2}/ϕ_X, respectively. Analogously, we may augment the moment restriction (3) in Theorem 1 with $\mathbb{E}_P\left[\exp(itY_1) - \phi_X(t)\phi_{U_1}(t)\right] = 0$ to partially identify (ϕ_X, ϕ_{U_1}) jointly, and similarly, we may augment the moment restriction (3) in Theorem 1 with $\mathbb{E}_P\left[\exp(itY_2) - \phi_X(t)\phi_{U_2}(t)\right] = 0$ to partially identify (ϕ_X, ϕ_{U_2}) jointly. These augmented moment restriction can be obtained without strengthening Assumption 1.*

Remark 3. *For the rest of the paper, we focus on the moment restriction in* (3). *This moment restriction is an implication from Assumption 1 for (1) and it holds for the true characteristic function ϕ_X of X. It is worth noting that this characterization may not be sharp, i.e., there could be other moment restrictions implied by Assumption 1.*

3 Confidence Band with the Hermite Orthogonal Basis

In this section, we introduce a confidence band for f_X using the Hermite function basis.[5] We consider a pre-specified significance level $\alpha \in (0, 1/2)$ throughout.

3.1 The Hermite Orthonormal Basis

We recommend the Hermite orthonormal basis in particular for its convenient properties and its nice compatibility with the deconvolution framework – a Hermite function is an eigenfunction of the Fourier transform and the Fourier inverse.[6] The Hermite functions take the form

$$\psi_j(x) = \frac{1}{\sqrt{2^j j! \sqrt{\pi}}} \cdot \exp(-x^2/2) \cdot H_j(x) \tag{4}$$

for $j = 0, 1, \ldots$, where H_j is the Hermite polynomial defined by

$$H_j(x) = (-1)^j \cdot \exp(x^2) \cdot \frac{d^j}{dx^j} \exp(-x^2).$$

[4]Evdokimov and White (2012) provides a relaxed assumption for the identification.

[5]We remark that our confidence band can be constructed using a different basis as well. For example, if we know that f_X has a bounded support, it is preferred to use a basis with a bounded support. For general discussions on sieve basis, see Chen (2007).

[6]The Hermite orthonormal sieve is not location invariant, and hence we recommend to location- and scale normalize the observed data using the empirical moments.

As emphasized earlier, the Hermite functions are the eigenfunctions of the Fourier transform operator, and specifically, $\phi_j = \mathcal{F}\psi_j = i^j\sqrt{2\pi}\psi_j$ holds.

3.2 Basis Expansion and Approximation for the Density Function

Provided that f_X belongs to the $\mathcal{L}^2$ space, we have the basis expansion

$$f_X = \sum_{j=0}^{\infty} \langle f_X, \psi_j \rangle \cdot \psi_j$$

with the Hermite functions $\{\psi_j\}$, where $\langle \cdot, \cdot \rangle$ denotes the inner product defined by $\langle f_1, f_2 \rangle = \int_{\mathbb{R}} f_1(x) f_2(x) dx$ (cf. Blanchard and Bruening, 2002, Theorem 16.3.1). Instead of using all the terms, we focus on the first $(q+1)$ terms. In other words, we use

$$\sum_{j=0}^{q} \langle f_X, \psi_j \rangle \cdot \psi_j,$$

to approximate f_X. We hereafter use $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_q)^T$ for a generic value for the unknown parameter vector $(\langle f_X, \psi_j \rangle : j = 0, 1, \ldots, q)^T$. For any $q \in \mathbb{N}$, we consider the set of sieve coefficients given by

$$\Theta^{q+1} = \{\boldsymbol{\theta} \in \mathbb{R}^{q+1} : |\theta_j| \le \sup_{x \in \mathbb{R}} |\psi_j(x)|\}. \tag{5}$$

In order to approximate f_X well by finite terms from the Hermite orthonormal basis, we impose the following restrictions on f_X.

Assumption 2. *f_X belongs to a known set, $\mathcal{L}$, of probability density function f's with $f \in \mathcal{L}^2$ and $\mathcal{F}f \in \mathcal{L}^1$ satisfying $|\langle f, \psi_j \rangle| \le j^{-3}$ for every $j \ge q+1$.*

We impose $f \in \mathcal{L}^2$ to approximate $\mathcal{L}$ by the Hermite orthonormal basis, whereas we impose $\mathcal{F}f \in \mathcal{L}^1$ for applying the inverse Fourier transform. For the last part of the assumption, we restrict the coefficient behavior for the Hermite basis expansion of the density function f_X. With this condition, we can bound the error from approximating the functions, ϕ_X and $\phi_X^{(1)}$, which appear in the moment condition (3). This condition implies the differentiability of the density f_X, and therefore it excludes non-differentiable density functions such as the uniform distribution and the triangular distribution. When we can take a sufficiently large value of q, a sufficient condition for $|\langle f_X, \psi_j \rangle| \le j^{-3}, \forall j \ge q+1$ is that the function $x^\kappa f_X(x)$ and the first κ-th derivatives of $f_X(t)$ are bounded and integrable for some integer $\kappa \ge 7$.[7]

3.3 Inequality Constraints for $(\langle f_X, \psi_j \rangle : j = 0, 1, \ldots, q)$

The moment restrictions of the form (3) impose equality restrictions. With this said, the finite-dimensional sieve approximation by Θ^{q+1} entails a truncation bias. This bias causes the equality

[7] Boyd (1984, p.385) shows $j^{\kappa/2}|\langle f_X, \psi_j \rangle| = O(1)$ as $j \to \infty$. Since we can bound $O(1)$ by $j^{(\kappa-6)/2}$ for sufficiently large j, we have $|\langle f, \psi_j \rangle| \le j^{-3}$ for every sufficiently large j.

restrictions into inequality restrictions with slackness by a uniform bias bound that can be determined by the restriction in Assumption 2. In this section, we present the resultant inequality restrictions and propose concrete choices of the slackness parameters.

To construct a confidence set for $(\langle f_X, \psi_j \rangle : j = 0, 1, \ldots, q)$, we are going to derive the two types of inequality constraints about $\boldsymbol{\theta}$, which are true when $\boldsymbol{\theta} = (\langle f_X, \psi_j \rangle : j = 0, 1, \ldots, q)^T$. The first type derives from the moment restrictions (3) in Theorem 1, and the second derives from natural restrictions on any density function.

First, using the moment restrictions of the form (3) in Theorem 1, we obtain the moment inequality constraints of the form

$$-\delta_q \leq \mathrm{Re}\left(\mathbb{E}_P\left[\left(iY_1 \sum_{j=0}^{q} \langle f_X, \psi_j \rangle \cdot \phi_j(t) - \sum_{j=0}^{q} \langle f_X, \psi_j \rangle \cdot \phi_j^{(1)}(t)\right) \exp(itY_2)\right]\right) \leq \delta_q \tag{6}$$

$$-\delta_q \leq \mathrm{Im}\left(\mathbb{E}_P\left[\left(iY_1 \sum_{j=0}^{q} \langle f_X, \psi_j \rangle \cdot \phi_j(t) - \sum_{j=0}^{q} \langle f_X, \psi_j \rangle \cdot \phi_j^{(1)}(t)\right) \exp(itY_2)\right]\right) \leq \delta_q, \tag{7}$$

where $\mathrm{Re}(\cdot)$ (respectively, $\mathrm{Im}(\cdot)$) denotes the real (respectively, imaginary) part of a complex number. Here, δ_q is the uniform approximation bound for the truncation bias, and is defined in (14) ahead. These inequalities are formally derived in Appendix A.2 as Lemma 2. Note that there are $4L$ inequalities in total if we use L grid points $t_1, \cdots, t_L$ of frequencies to evaluate the moment restrictions - $2L$ inequalities for the real part and $2L$ inequalities for the imaginary part. We can choose $t_1, \ldots, t_L$ such that they are equally distributed in the interval $[-h^{-1}, h^{-1}]$ selected according to Appendix B.2.

To economize our writings for (6) and (7), we introduce additional notations. For every function $\psi \in \mathcal{L}$ and for every frequency $t \in \mathbb{R}$, define

$$R_{\psi,t}(y_1, y_2) = -\cos(ty_2)(y_1 \mathrm{Im}(\phi(t)) + \mathrm{Re}(\phi^{(1)}(t))) - \sin(ty_2)(y_1 \mathrm{Re}(\phi(t)) - \mathrm{Im}(\phi^{(1)}(t))) \quad \text{and} \tag{8}$$

$$I_{\psi,t}(y_1, y_2) = \cos(ty_2)(y_1 \mathrm{Re}(\phi(t)) + \mathrm{Im}(\phi^{(1)}(t))) - \sin(ty_2)(y_1 \mathrm{Im}(\phi(t)) - \mathrm{Re}(\phi^{(1)}(t))), \tag{9}$$

where $\phi = \mathcal{F}\psi$. Further, stack these functions across $\psi \in \{\psi_0, \ldots, \psi_q\}$ to in turn define the random vector

$$\mathbf{R}_t = (R_{\psi_0,t}(Y_1, Y_2), \ldots, R_{\psi_q,t}(Y_1, Y_2))^T \qquad \text{and} \tag{10}$$

$$\mathbf{I}_t = (I_{\psi_0,t}(Y_1, Y_2), \ldots, I_{\psi_q,t}(Y_1, Y_2))^T. \tag{11}$$

With these notations, we now represent (6) and (7), respectively, by

$$|\mathbb{E}_P[\mathbf{R}_{t_l}]^T (\langle f_X, \psi_j \rangle : j = 0, 1, \ldots, q)^T| \leq \delta_q \qquad \text{and} \qquad |\mathbb{E}_P[\mathbf{I}_{t_l}]^T (\langle f_X, \psi_j \rangle : j = 0, 1, \ldots, q)^T| \leq \delta_q.$$

Second, since $\sum_{j=0}^{\infty} \langle f_X, \psi_j \rangle \cdot \int \psi_j(x)dx = \int f_X(x)dx = 1$ and $\sum_{j=0}^{\infty} \langle f_X, \psi_j \rangle \cdot \psi_j(x) = f_X(x) \geq 0$ hold as natural restrictions on any density function, we obtain the additional inequality constraints of the form

$$\left|\sum_{j=0}^{q}\langle f_X,\psi_j\rangle\cdot\int\psi_j(x)dx-1\right|\leq\eta_q \text{ and } \sum_{j=0}^{q}\langle f_X,\psi_j\rangle\cdot\psi_j(x)\geq-\eta_q. \tag{12}$$

Here, η_q denotes the uniform approximation bound for the truncation bias, and is defined in (13) ahead. The inequalities are formally derived in Appendix A.2 as Lemma 2.

With the coefficient conditions in Assumption 2, we can derive known approximation-error bounds, η_q and δ_q, used in (12)–(7), according to the following rule:

$$\eta_q = \max\left\{\sum_{j=q+1}^{\infty} j^{-3}\sup_{x\in\mathbb{R}}|\psi_j(x)|\,,\ \sum_{j=q+1}^{\infty} j^{-3}\int|\psi_j(x)|dx\right\} \tag{13}$$

$$\delta_q = \sum_{j=q+1}^{\infty} j^{-3}\left(\mathbb{E}_P\left[|Y_1|\right]\sup_{t\in\mathbb{R}}|\phi_j(t)|+\sup_{t\in\mathbb{R}}|\phi_j^{(1)}(t)|\right), \tag{14}$$

where $\phi_j=\mathcal{F}\psi_j$.[8] The idea to obtain the approximation-error bounds from restrictions on the function space (Assumption 2) is similar to those of in the preceding papers by Schennach (2015) and Armstrong and Kolesár (2018).

The following lemma summarizes and formalizes the statements we made in this subsection.

Lemma 2. *Under Assumptions 1 and 2, we have*

$$\left|\left(\int\boldsymbol{\psi}(x)dx\right)^T\boldsymbol{\theta}-1\right|\leq\eta_q, \tag{15}$$

$$\inf_{x\in\mathbb{R}}\boldsymbol{\psi}(x)^T\boldsymbol{\theta}\geq-\eta_q, \tag{16}$$

$$|\mathbb{E}_P[\mathbf{R}_{t_l}]^T\boldsymbol{\theta}|\leq\delta_q \tag{17}$$

$$|\mathbb{E}_P[\mathbf{I}_{t_l}]^T\boldsymbol{\theta}|\leq\delta_q \tag{18}$$

when $\boldsymbol{\theta}=(\langle f_X,\psi_j\rangle : j=0,1,\ldots,q)^T$, where $\boldsymbol{\psi}(x)=(\psi_0(x),\cdots,\psi_q(x))^T$, and η_q and δ_q are defined in Equations (13)-(14).

A proof is provided in Appendix A.2.

We can also show that, under some conditions on q and $t_1,\ldots,t_L$, the constraints in Lemma 2 approximately characterize the moment condition (3). The following lemma provides a mathematical statement of this approximation in a large sample.

Lemma 3. *Suppose that t_l's are equally distributed over $[-h^{-1},h^{-1}]$ for some positive integer h. Consider a probability density function $f\in\mathcal{L}$ which does not satisfy the moment condition* (3) *for*

[8] Note that $\mathbb{E}_P\left[|Y_1|\right]$ needs to be estimated. For simplicity, we assume that $\mathbb{E}_P\left[|Y_1|\right]$ is a known constant when we define δ_q.

some $\bar{t} \in [-h^{-1}, h^{-1}]$*. When* q *and* Lh *are sufficiently large,*

$$\max_{l=1,\ldots,L} \max\left\{|\mathbb{E}_P[\mathbf{R}_{t_l}]^T\boldsymbol{\theta}|, |\mathbb{E}_P[\mathbf{I}_{t_l}]^T\boldsymbol{\theta}|\right\} > \delta_q,$$

where $\boldsymbol{\theta} = (\langle f, \psi_j\rangle : j = 0, 1, \ldots, q)^T$ *and* δ_q *is defined in Eq. (14).*

A proof is provided in Appendix A.3.

3.4 Confidence Region for $(\langle f_X, \psi_j\rangle : j = 0, 1, \ldots, q)$

In light of the inequality conditions (17) and (18) involving the population moments, we construct a test statistic and a critical value based on their sample counterparts.[9] We define our test statistic by

$$T(\boldsymbol{\theta}) = \sqrt{n} \max_{1\le l\le L} \max\left\{\frac{|\mathbb{E}_n[\mathbf{R}_{t_l}]^T\boldsymbol{\theta}| - \delta_q}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_n(\mathbf{R}_{t_l})\boldsymbol{\theta}}}, \frac{|\mathbb{E}_n[\mathbf{I}_{t_l}]^T\boldsymbol{\theta}| - \delta_q}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_n(\mathbf{I}_{t_l})\boldsymbol{\theta}}}\right\}$$

for each $\boldsymbol{\theta} \in \Theta^{q+1}$. We define the critical value $c(\alpha, \boldsymbol{\theta})$ of this statistic $T(\boldsymbol{\theta})$ by the conditional $(1-\alpha)$-th quantile of the multiplier bootstrap statistic

$$\sqrt{n} \max_{1\le l\le L} \max\left\{\frac{\left|\mathbb{E}_n[\boldsymbol{\epsilon}(\mathbf{R}_{t_l} - \mathbb{E}_n[\mathbf{R}_{t_l}])]^T\boldsymbol{\theta}\right|}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_n(\mathbf{R}_{t_l})\boldsymbol{\theta}}}, \frac{\left|\mathbb{E}_n[\boldsymbol{\epsilon}(\mathbf{I}_{t_l} - \mathbb{E}_n[\mathbf{I}_{t_l}])]^T\boldsymbol{\theta}\right|}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_n(\mathbf{I}_{t_l})\boldsymbol{\theta}}}\right\}$$

given the data, where $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n$ are independent standard normal random variables independent of the data. As a more conservative yet simpler alternative following Chernozhukov et al. (2018, eq. (19)), we may also employ the critical value defined by

$$c(\alpha) = \frac{\Phi^{-1}(1-\alpha/(4L))}{1 - \Phi^{-1}(1-\alpha/(4L))^2/n},$$

where Φ is the cumulative distribution function of the standard normal distribution.

We can construct the confidence region for $(\langle f_X, \psi_j\rangle : j = 0, 1, \ldots, q)$ as the set of $\boldsymbol{\theta}$'s satisfying

[9] The underlying idea of our test statistic and critical value is to discretize the continuum of moment conditions (3) into a set of finite but many moment inequalities on the sieve coefficients $\boldsymbol{\theta}$, and to calibrate critical values by the multiplier bootstrap for the maximum statistic as in Chernozhukov et al. (2018). A natural question would be whether we could directly use the continuum of the moment conditions (3) without discretization, similarly to, e.g., Andrews and Shi (2013, 2017). One difficulty, however, is that the moment functions corresponding to the moment condition (3) at given f_X, i.e., $\{(y_1, y_2) \mapsto R_{f_X,t}(y_1, y_2), I_{f_X,t}(y_1, y_2) : t \in \mathbb{R}\}$, is not likely to be Donsker, in view of the fact that, e.g., the function class $\{y_2 \mapsto \cos(ty_2) : t \in \mathbb{R}\}$ is non-Donsker as soon as Y_2 has a density (which is the case under our assumption), so that the "manageability" condition in Andrews and Shi (2013, 2017) would not be satisfied in our case. Indeed, the preceding function class is not Glivenko-Cantelli from the Riemann-Lebesgue lemma and discreteness of the empirical distribution; see Feuerverger and Mureika (1977) for details. Another potential approach would be to apply the method of continuum of moment conditions developed in Carrasco and Florens (2000). Their analysis relies on point-identification of the parameter of interest and, more importantly, focuses on a finite dimensional parameter of interest (so the convergence rate of their estimator is the parametric rate), so that their approach is not directly applicable to our problem.

the three conditions

$$T(\boldsymbol{\theta}) \leq c(\alpha, \boldsymbol{\theta}), \quad \left| \sum_{j=0}^{q} \theta_j \cdot \int \psi_j(x)dx - 1 \right| \leq \eta_q, \text{ and } \inf_{x\in\mathbb{R}} \sum_{j=0}^{q} \theta_j \cdot \psi_j(x) \geq -\eta_q.$$

We denote this set by $\mathcal{C}_{(\langle f_X, \psi_j \rangle : j=0,1,\ldots,q)}(\alpha)$.

3.5 Confidence Band for f_X

Given the confidence region for $(\langle f_X, \psi_j \rangle : j = 0, 1, \ldots, q)$ provided by $\mathcal{C}_{(\langle f_X, \psi_j \rangle : j=0,1,\ldots,q)}(\alpha)$, we can in turn construct a confidence band for f_X by

$$\mathcal{C}_n(\alpha) = \left\{ f \in \mathcal{L} : \sup_{x\in\mathbb{R}} \left| f(x) - \sum_{j=0}^{q} \theta_j \cdot \psi_j(x) \right| \leq \eta_q \text{ for some } \boldsymbol{\theta} \in \mathcal{C}_{(\langle f_X, \psi_j \rangle : j=0,1,\ldots,q)}(\alpha) \right\}. \tag{19}$$

The following algorithm describes a concrete practical procedure for constructing this confidence band.[10]

Algorithm 1.

1. *For each $x \in \mathbb{R}$, compute*

$$f^L(x) = \min_{\boldsymbol{\theta}\in\Theta^{q+1}} \boldsymbol{\psi}(x)^T \boldsymbol{\theta} \quad \textit{subject to} \quad \begin{aligned} & T(\boldsymbol{\theta}) \leq c(\alpha, \boldsymbol{\theta}) \\ & \boldsymbol{\psi}(x)^T \boldsymbol{\theta} \geq -\eta_q \textit{ for all } x \in \mathbb{R} \\ & \left| \sqrt{2\pi} \boldsymbol{\psi}(0)^T \textit{diag}\,(1, 0, -1, 0 \ldots)\, \boldsymbol{\theta} - 1 \right| \leq \eta_q \end{aligned}$$

2. *For each $x \in \mathbb{R}$, compute*

$$f^U(x) = \max_{\boldsymbol{\theta}\in\Theta^{q+1}} \boldsymbol{\psi}(x)^T \boldsymbol{\theta} \quad \textit{subject to} \quad \begin{aligned} & T(\boldsymbol{\theta}) \leq c(\alpha, \boldsymbol{\theta}) \\ & \boldsymbol{\psi}(x)^T \boldsymbol{\theta} \geq -\eta_q \textit{ for all } x \in \mathbb{R} \\ & \left| \sqrt{2\pi} \boldsymbol{\psi}(0)^T \textit{diag}\,(1, 0, -1, 0 \ldots)\, \boldsymbol{\theta} - 1 \right| \leq \eta_q \end{aligned}$$

[10] We denote by $\mathrm{diag}\,(1, 0, -1, 0 \ldots)$ the diagonal matrix whose $(j+1)$th diagonal element is $i^j \cdot 1\{j \text{ is even}\}$ for every $j = 0, 1, \ldots, q$. Namely, the main diagonal of $\mathrm{diag}\,(1, 0, -1, 0 \ldots)$ repeats $1, 0, -1, 0$. The constraint $\left| \sqrt{2\pi} \boldsymbol{\psi}(0)^T \mathrm{diag}\,(1, 0, -1, 0 \ldots)\, \boldsymbol{\theta} - 1 \right| \leq \eta_q$ follows from combining

$$\left| \sum_{j=0}^{q} \theta_j \cdot \int \psi_j(x)dx - 1 \right| \leq \eta_q$$

and

$$\int \psi_j(x)dx = \begin{cases} \sqrt{2\pi}\psi_j(0) & \text{if the remainder from dividing } j \text{ by 4 is 0} \\ 0 & \text{if the remainder from dividing } j \text{ by 4 is 1} \\ -\sqrt{2\pi}\psi_j(0) & \text{if the remainder from dividing } j \text{ by 4 is 2} \\ 0 & \text{if the remainder from dividing } j \text{ by 4 is 3.} \end{cases}$$

The calculations of $\int \psi_j(x)dx$ use the fact that ψ_j's are the Hermite functions.

3. The confidence band is set to $\left[f^L(x)-\eta_q,\ f^U(x)+\eta_q\right]$, $x\in I$.

3.6 Discussion on Sieve Dimension q

The sieve dimension q can be chosen by adapting a bandwidth selection method suggested in Bissantz et al. (2007) in density deconvolution with known error distribution; similar bandwidth selection rules are also used in Kato and Sasaki (2018) and Adusumilli et al. (2019) in the deconvolution literature. Given q, choose tolerance levels η_q and δ_q depending on q, and then construct the upper and lower functions $f^U(x)=f_q^U(x)$ and $f^L(x)=f_q^L(x)$ according to Algorithm 1. Then use the midpoint $\hat f_q(x)=\{f_q^U(x)+f_q^L(x)\}/2$ as a surrogate of a point estimate of $f(x)$. Realizing that a sieve dimension corresponds to the reciprocal of a bandwidth, we suggest the following rule to choose q. Construct a candidate set for q as $\{q_{\min},\dots,q_{\max}\}$, and compute the L^∞-distance between the density estimates with adjacent sieve dimensions, $d^{(\infty)}_{q,q+1}=\sup_{x\in I}|\hat f_{q+1}(x)-\hat f_q(x)|$. Then we choose the smallest q such that $d^{(\infty)}_{q,q+1}$ is larger than $\rho\cdot d^{(\infty)}_{q_{\min},q_{\min}+1}$ for some $\rho>1$ (or alternatively we can choose the largest q such that $d^{(\infty)}_{q,q+1}$ is smaller than $\rho\cdot d^{(\infty)}_{q_{\min},q_{\min}+1}$). In practice, it is recommended to make use of visual information on how $d^{(\infty)}_{q,q+1}$ behaves as q decreases when determining the sieve dimension.

4 Properties of the Confidence Band

In this section, we present theoretical properties of the confidence band (19) with n i.i.d. observations of (Y_1,Y_2). Let $\mathcal{P}$ denote a given space to which the joint distribution of (Y_1,Y_2) belongs. For every $P\in\mathcal{P}$, define the identified set

$$\mathcal{L}_0(P)=\left\{f\in\mathcal{L}\ :\ \phi=\mathcal{F}f \text{ and } \mathbb{E}_P\left[\left(iY_1\phi(t)-\phi^{(1)}(t)\right)\exp(itY_2)\right]=0 \text{ for every } t\in\mathbb{R}\right\}\tag{20}$$

as the set of density functions for which the linear complex-valued moment restriction (3) is satisfied.

4.1 Size Control

We make the following assumption for a uniform size control.

Assumption 3. *(i)* $\mathbb{E}_P\left[Y_1^2\right]<\infty$ *for all* $P\in\mathcal{P}$. *(ii) There are constants* $0<c_1<1/2$ *and* $C_1>0$ *such that*

$$\left(M^3_{L,q,3}(\boldsymbol{\theta},P)\vee M^2_{L,q,4}(\boldsymbol{\theta},P)\vee B_{L,q}(\boldsymbol{\theta},P)\right)^2\log^{7/2}(4Ln)\le C_1 n^{1/2-c_1}$$

for all $P\in\mathcal{P}$ *and* $\theta\in\Theta^{q+1}$, *where*

$M_{L,q,k}(\boldsymbol{\theta},P)=\max_{1\le l\le L}\max\left\{\mathbb{E}_P\left[\left|\frac{(\mathbf{R}_{t_l}-\mathbb{E}_P[\mathbf{R}_{t_l}])^T\boldsymbol{\theta}}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_P(\mathbf{R}_{t_l})\boldsymbol{\theta}}}\right|^k\right]^{1/k},\mathbb{E}_P\left[\left|\frac{(\mathbf{I}_{t_l}-\mathbb{E}_P[\mathbf{I}_{t_l}])^T\boldsymbol{\theta}}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_P(\mathbf{I}_{t_l})\boldsymbol{\theta}}}\right|^k\right]^{1/k}\right\}$

and $B_{L,q}(\boldsymbol{\theta},P)=\mathbb{E}_P\left[\max_{1\le l\le L}\max\left\{\left|\frac{(\mathbf{R}_{t_l}-\mathbb{E}_P[\mathbf{R}_{t_l}])^T\boldsymbol{\theta}}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_P(\mathbf{R}_{t_l})\boldsymbol{\theta}}}\right|^4,\left|\frac{(\mathbf{I}_{t_l}-\mathbb{E}_P[\mathbf{I}_{t_l}])^T\boldsymbol{\theta}}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_P(\mathbf{I}_{t_l})\boldsymbol{\theta}}}\right|^4\right\}\right]^{1/4}.$

Remark 4. *When $\mathbb{E}_P\left[Y_1^4\right] < \infty$, Assumption 3 (ii) holds if*

$$(q+1)/\min_{t\in[-T,T]}\min\left\{\mathrm{eig}_{\min}(\mathbb{V}_P(\mathbf{R}_t)),\mathrm{eig}_{\min}(\mathbb{V}_P(\mathbf{I}_t))\right\} = O(n^{1/6-c_1/3}\log^{-7/6}(4Ln)). \tag{21}$$

A derivation of this condition is in Appendix A.7. Eq. (21) restricts how fast q and T can increase to infinity.[11] *On the other hand, Eq. (21) does not restrict the choice of L in the sense that, as long as L grows at a polynomial rate of n, the term $\log^{-7/6}(4Ln)$ is negligible on the right hand side.*

The following theorem states the uniform size control by our proposed inference method.

Theorem 2 (Size Control)**.** *Suppose that Assumption 2 and 3 holds. Then, there exist positive constants c and C depending only on c_1 and C_1 such that*

$$\inf_{P\in\mathcal{P}}\inf_{f\in\mathcal{L}_0(P)}\mathbb{P}_P(f\in\mathcal{C}_n(\alpha))\geq 1-\alpha-Cn^{-c}.$$

4.2 Power

We introduce the following short-hand notation for the random variable defined as the maximum deviation of the sample variance from the population variance.

$$B_V = \sup_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)}\max_{l=1,\dots,L}\max\left\{\boldsymbol{\theta}^T(\mathbb{V}_n(\mathbf{R}_{t_l})-\mathbb{V}_P(\mathbf{R}_{t_l}))\boldsymbol{\theta},\boldsymbol{\theta}^T(\mathbb{V}_n(\mathbf{I}_{t_l})-\mathbb{V}_P(\mathbf{I}_{t_l}))\boldsymbol{\theta}\right\}$$

The following theorem shows a power property of our proposed inference method.

Theorem 3 (Power)**.** *Define*

$$\mathbf{B}_{q+1,\eta_q}(f) = \left\{\boldsymbol{\theta}\in\Theta^{q+1}:\sup_{x\in I}|f(x)-\boldsymbol{\psi}(x)^T\boldsymbol{\theta}|\leq\eta_q\right\}.$$

For every $P\in\mathcal{P}$, every $f\in\mathcal{L}$, every $\nu>0$, and every $b\in(0,1)$, if there is $t_*\in\{t_1,\dots,t_L\}$ such that at least one of the following statements holds:

$$\begin{aligned}\sqrt{n}\inf_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)}\frac{\mathbb{E}_P[\mathbf{R}_{t_*}]^T\boldsymbol{\theta}-\delta_q}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_P(\mathbf{R}_{t_*})\boldsymbol{\theta}+\nu}} \geq \frac{1}{b}\cdot\mathbb{E}_P\left[\sup_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)}\frac{|\mathbb{G}_n[\mathbf{R}_{t_*}]^T\boldsymbol{\theta}|}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_n(\mathbf{R}_{t_*})\boldsymbol{\theta}}}\right]\\ +\sqrt{2\log(4L)}+\sqrt{2\log(1/\alpha)}\end{aligned} \tag{22}$$

$$\begin{aligned}\sqrt{n}\inf_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)}\frac{-\mathbb{E}_P[\mathbf{R}_{t_*}]^T\boldsymbol{\theta}-\delta_q}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_P(\mathbf{R}_{t_*})\boldsymbol{\theta}+\nu}} \geq \frac{1}{b}\cdot\mathbb{E}_P\left[\sup_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)}\frac{|\mathbb{G}_n[\mathbf{R}_{t_*}]^T\boldsymbol{\theta}|}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_n(\mathbf{R}_{t_*})\boldsymbol{\theta}}}\right]\\ +\sqrt{2\log(4L)}+\sqrt{2\log(1/\alpha)}\end{aligned} \tag{23}$$

[11] Loosely speaking, the term $(q+1)/\min_{t\in[-T,T]}\min\left\{\mathrm{eig}_{\min}(\mathbb{V}_P(\mathbf{R}_t)),\mathrm{eig}_{\min}(\mathbb{V}_P(\mathbf{I}_t))\right\}$ is increasing in q and T. When q increases, the numerator increases and the minimum eigenvalues of the $(q+1)$ square matrices, $\mathbb{V}_P(\mathbf{R}_t)$ and $\mathbb{V}_P(\mathbf{I}_t)$, can be smaller. Moreover, when T increases, the denominator decreases.

$$\sqrt{n}\inf_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)}\frac{\mathbb{E}_P[\mathbf{I}_{t_*}]^T\boldsymbol{\theta}-\delta_q}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_P(\mathbf{I}_{t_*})\boldsymbol{\theta}+\nu}}\geq\frac{1}{b}\cdot\mathbb{E}_P\left[\sup_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)}\frac{|\mathbb{G}_n[\mathbf{I}_{t_*}]^T\boldsymbol{\theta}|}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_n(\mathbf{I}_{t_*})\boldsymbol{\theta}}}\right]+\sqrt{2\log(4L)}+\sqrt{2\log(1/\alpha)} \tag{24}$$

$$\sqrt{n}\inf_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)}\frac{-\mathbb{E}_P[\mathbf{I}_{t_*}]^T\boldsymbol{\theta}-\delta_q}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_P(\mathbf{I}_{t_*})\boldsymbol{\theta}+\nu}}\geq\frac{1}{b}\cdot\mathbb{E}_P\left[\sup_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)}\frac{|\mathbb{G}_n[\mathbf{I}_{t_*}]^T\boldsymbol{\theta}|}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_n(\mathbf{I}_{t_*})\boldsymbol{\theta}}}\right]+\sqrt{2\log(4L)}+\sqrt{2\log(1/\alpha)}, \tag{25}$$

where $\mathbb{G}_nf=\sqrt{n}\,(\mathbb{E}_n-\mathbb{E}_P)\,f$ *denotes the empirical process evaluated at* f*, then*

$$\mathbb{P}_P(f\notin\mathcal{C}_n(\alpha))\geq\mathbb{P}_P\left(B_V\leq\nu\right)-b.$$

A proof is provided in Appendix A.5. According to this theorem, for any density function $f\in\mathcal{L}$ such that at least one of the moment inequalities violated at some frequency point t_* in the grid $\{t_1,\ldots,t_L\}$, then the probability that this density function does not belong to the confidence band is bounded below by $\mathbb{P}_P\left(B_V\leq\nu\right)-b$. Choosing sequences of ν and b so that $\mathbb{P}_P\left(B_V\leq\nu\right)-b\to 1$ as $n\to\infty$, therefore, this theorem implies the consistency against all fixed alternatives as formally stated in the corollary below. For the following corollary, let q and L depend on n such that $q\to\infty$ and $L\to\infty$ as $n\to\infty$. Let

$$\begin{aligned}\lambda^P_{\max}(t)&=\sup_{\boldsymbol{\theta}\in\Theta_{q+1}}\max\{\boldsymbol{\theta}^T\mathbb{V}_P(\mathbf{R}_t)\boldsymbol{\theta},\boldsymbol{\theta}^T\mathbb{V}_P(\mathbf{I}_t)\boldsymbol{\theta}\}\\ a^P_n(t)&=\max\left\{\mathbb{E}_P\left[\sup_{\theta\in\Theta_{q+1}}\frac{|\mathbb{G}_n[\mathbf{R}_t]^T\boldsymbol{\theta}|}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_n(\mathbf{R}_t)\boldsymbol{\theta}}}\right],\mathbb{E}_P\left[\sup_{\theta\in\Theta_{q+1}}\frac{|\mathbb{G}_n[\mathbf{I}_t]^T\boldsymbol{\theta}|}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_n(\mathbf{I}_t)\boldsymbol{\theta}}}\right]\right\}.\end{aligned}$$

For two sequences of positive numbers a_n and b_n, we write $a_n\gg b_n$ if $a_n/b_n\to\infty$. The next corollary states the consistency against all fixed alternatives.

Corollary 1 (Consistency)**.** *Pick any* $P\in\mathcal{P}$ *and* $f\in\mathcal{L}$ *such that for some sequence* $t_*=t_*(n)\in\{t_1,\ldots,t_L\}$,

$$\sqrt{n}\inf_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)}\max\{|\mathbb{E}_P[\mathbf{R}_{t_*}]^T\boldsymbol{\theta}-\delta_q|,|\mathbb{E}_P[\mathbf{I}_{t_*}]^T\boldsymbol{\theta}-\delta_q|\}\gg\max\left\{a^P_n(t_*)\sqrt{\lambda^P_{\max}(t_*)},\sqrt{\log(1+L)}\right\}.$$

Then $\mathbb{P}_P(f\notin\mathcal{C}_n(\alpha))\to 1$ *provided that*

$$\sup_{\boldsymbol{\theta}\in\Theta_{q+1}}\max_{l=1,\ldots,L}\max\left\{\boldsymbol{\theta}^T(\mathbb{V}_n(\mathbf{R}_{t_l})-\mathbb{V}_P(\mathbf{R}_{t_l}))\boldsymbol{\theta},\boldsymbol{\theta}^T(\mathbb{V}_n(\mathbf{I}_{t_l})-\mathbb{V}_P(\mathbf{I}_{t_l}))\boldsymbol{\theta}\right\}=o_P(\lambda^P_{\max}(t_*)).$$

A proof is provided in Appendix A.6.

5 Application to Wildcat Auctions

In this section, we present an empirical application of our proposed method to wildcat auctions. The data set that we use is the Outer Continental Shelf (OCS) Auction Data. This data set records bids for mineral rights on oil and gas on offshore lands off the coasts of Texas and Louisiana in the gulf of Mexico. We refer interested readers to Hendricks, Porter, and Boudreau (1987), for example, for details of this data set.

Among other types of sales, we focus on wildcat sales, i.e., sales of rights for oil and gas tracts whose geological or seismic characteristics are unknown to bidding firms. The sales follows the first-price sealed-bid auction mechanism. In this mechanism, bidding firms simultaneously submit sealed bids. The bidder with the highest bid pays the price which they submitted to receive the right for the tract.

Prior to an auction, firms which are planning to participating in the auction can carry out a seismic investigation to estimate the value of the rights. Results of these seismic investigation provide firm 1 (respectively, firm 2) with the *ex ante* value Y_1 (respectively, Y_2) of the mineral right in the logarithm of US dollars per acre. These *ex ante* values, Y_1 and Y_2, are two measurements of the *ex post* value X, also known as the common component, in the logarithm of US dollars per acre, with *ex ante* assessment errors, U_1 and U_2, respectively, also known as the private components, in the logarithm of US dollars per acre.

In this setting, we obtain the system of repeated measurement error equations

$$\begin{cases} Y_1 = X + U_1, \\ Y_2 = X + U_2, \end{cases}$$

as is the case with the main equation (1) of our model framework. We assume that the *ex post* value X and the two firms' *ex ante* assessment errors, U_1 and U_2, are continuously distributed and are mutually independent. Furthermore, we also assume the rational expectations, that is, $E[U_1] = 0$. These conditions satisfy our Assumption 1.

Li et al. (2000) apply the method of Li and Vuong (1998) to this setup and this data set in order to estimate f_X, but they do not conduct a statistical inference. Krasnokutskaya (2011) similarly estimates the density function f_X in a different framework, and draws its confidence intervals via nonparametric bootstrap. They first recover firms' *ex ante* values (Y_1, Y_2) from bid data through the method of Guerre, Perrigne, and Vuong (2000) based on an equilibrium restriction (Bayesian Nash equilibrium) for the first-price sealed-bid auction mechanism. Following their approach, we also directly take these *ex ante* values (Y_1, Y_2) as the data to be used as an input for our analysis, in light of the faster convergence rate of the preliminary estimation than the convergence rate of the deconvolution estimation. The sample consists of 169 tracts with 2 firms in each tract.

Applying our proposed method, we draw the 95% confidence band $\mathcal{C}_n(0.05)$ for the density function f_X of the *ex post* values X on $I = [4.16,\ 6.16]$ as follows. Following the settings for our simulations to be presented in Section 6, we set $L = 50$ for the grid size of frequencies, and

$q = 2$ such that the dimensionality of $\boldsymbol{\theta}$ is 3.[12] The frequency domain of integration for the Fourier transform is set to be the interval $[-h^{-1}, h^{-1}]$, where $h = 0.289$ is chosen based on the procedure outlined in Appendix B.2. In other words, we construct $\mathbf{R}_{t_l}$ as in (10) and $\mathbf{I}_{t_l}$ as in (11) at each of the 50 equally spaced grid points $t_1, ..., t_{50}$ from $-1/0.289$ to $1/0.289$. With these $\mathbf{R}_{t_l}$ and $\mathbf{I}_{t_l}$ for $t_1, ..., t_{50}$, we construct the test statistic

$$T(\boldsymbol{\theta}) = \sqrt{n} \max_{1 \leq l \leq L} \max \left\{ \frac{|\mathbb{E}_n[\mathbf{R}_{t_l}]^T \boldsymbol{\theta}| - \delta_q}{\sqrt{\boldsymbol{\theta}^T \mathbb{V}_n(\mathbf{R}_{t_l}) \boldsymbol{\theta}}}, \frac{|\mathbb{E}_n[\mathbf{I}_{t_l}]^T \boldsymbol{\theta}| - \delta_q}{\sqrt{\boldsymbol{\theta}^T \mathbb{V}_n(\mathbf{I}_{t_l}) \boldsymbol{\theta}}} \right\},$$

where $L = 50$ and (14) yields the approximation bound $\delta_q = 0.139$ for the current application. Let $c(0.05)$ denote the critical value obtained by setting $\alpha = 0.05$, $L = 50$, and $n = 169$ in

$$c(\alpha) = \frac{\Phi^{-1}(1 - \alpha/(4L))}{1 - \Phi^{-1}(1 - \alpha/(4L))^2/n}.$$

Finally, using (13), we obtain the approximation error bound of $\eta_q = 0.048$ for the density function. With these settings, the lower bound of the confidence band $\mathcal{C}_{n(0.05)}$ is $f^L(x) - \eta_q$ where

$$\begin{aligned} f^L(x) = \min_{\boldsymbol{\theta} \in \Theta^{q+1}} \boldsymbol{\psi}(x)^T \boldsymbol{\theta} \quad \text{subject to} \quad & T(\boldsymbol{\theta}) \leq c(0.05) \\ & \boldsymbol{\psi}(x)^T \boldsymbol{\theta} \geq -\eta_q \text{ for all } x \in \mathbb{R} \\ & \left| \sqrt{2\pi} \boldsymbol{\psi}(0)^T \mathrm{diag}\,(1, 0, -1, 0 \dots)\, \boldsymbol{\theta} - 1 \right| \leq \eta_q \end{aligned}$$

for all $x \in I$, and the upper bound of it is $f^U(x) + \eta_q$ where

$$\begin{aligned} f^U(x) = \max_{\boldsymbol{\theta} \in \Theta^{q+1}} \boldsymbol{\psi}(x)^T \boldsymbol{\theta} \quad \text{subject to} \quad & T(\boldsymbol{\theta}) \leq c(\alpha) \\ & \boldsymbol{\psi}(x)^T \boldsymbol{\theta} \geq -\eta_q \text{ for all } x \in \mathbb{R} \\ & \left| \sqrt{2\pi} \boldsymbol{\psi}(0)^T \mathrm{diag}\,(1, 0, -1, 0 \dots)\, \boldsymbol{\theta} - 1 \right| \leq \eta_q \end{aligned}$$

for all $x \in I$.[13] We numerically solve these constrained optimization problems via Newton's method with the penalty factor of 1,000 and the penalty exponent of 2 for the constraints. Note that the test statistic $T(\boldsymbol{\theta})$ in the constraint derives from objects in the frequency domain, and evaluation of it in general would require a numerical integration for Fourier transform *within each iteration* of the numerical optimization routine. We recommend to use the Hermite orthogonal sieve to substantially reduce this computational burden. Since the Hermite functions are the eigenfunctions of the Fourier transform and inverse operators, we only need to multiply by the eigenvalues and thus do not need to conduct a numerical integration within each iteration thanks to the Hermite orthogonal sieve.

[12] Given that our sample size in this section is small, we use a small value of the sieve dimension and control the magnitude of the variance at the expense of the bias. In fact, the area of the light gray shade representing the bias is already much smaller than that of the dark gray area representing the stochastic part even with this small sieve dimension in Figure 1 to be discussed ahead.

[13] See Footnote 10 for the definition of $\mathrm{diag}\,(1, 0, -1, 0 \dots)$.

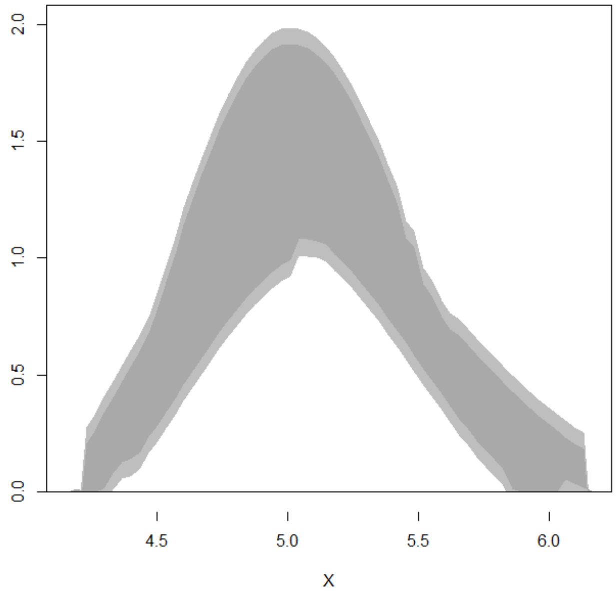


Figure 1: 95% confidence band $\mathcal{C}_n(0.05)$ of the density function f_X of the *ex post* values X of the mineral right in the logarithm of US dollars per acre. The dark gray shade represent the band based on the stochastic part, and the light gray shade represent the uniform bias bound η_q.

Figure 1 depicts the resultant confidence band $\mathcal{C}_n(0.05)$.

Density functions *per se* are often of research interest in the auction literature (e.g., Li et al., 2000; Krasnokutskaya, 2011). With this said, many parameters of economic interest can be also obtained as functionals of the density function f_X. We next use $\mathcal{C}_n(0.05)$ to obtain confidence intervals for parameters of economic interest as functionals of the density function f_X. Specifically, we are interested in the average *ex post* values of the mineral rights in US dollars per acre. Note that X is the *logarithm* of the pecuniary unit, and hence, parameters of more economic interest would be statistics about $\exp(X)$. (If the average of X were the parameter of interest, one could simply estimate its mean by $E[X] = E[Y_1]$ and obtain its confidence interval without relying on the deconvolution approach.) We obtain the 95% confidence interval for the average *ex post* values, $E[\exp(X)]$, of mineral rights in US dollars per acre by

$$\left\{ \int \exp(x) f(x) dx \;\middle|\; f \in \mathcal{C}_n(0.05) \right\}.$$

Similarly, we can obtain the 95% confidence interval for the τ-th quantile of the *ex post* values $\exp(X)$ of mineral rights in US dollars per acre by

$$\left\{ \inf \left\{ \exp(x) \;\middle|\; \int_{-\infty}^{x} f(x')dx' \geq \tau \right\} \;\middle|\; f \in \mathcal{C}_n(0.05) \right\}.$$

	Mean	Quantiles				
	$E[\exp(X)]$	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
Upper Bound	219	114	150	190	267	362
Lower Bound	146	81	100	131	166	218

Table 1: 95% confidence intervals of the mean and quantiles of the *ex post* values of mineral rights on oil and gas on offshore lands off the coasts of Texas and Louisiana in the gulf of Mexico in US dollars per square acre. The first column shows the 95% confidence interval for the mean. The remaining five columns show the 95% confidence intervals for the 0.1-th, 0.3-th, 0.5-th, 0.7-th and 0.9-th quantiles.

Table 1 summarizes the 95% confidence interval for the mean $E[\exp(X)]$ and the 95% confidence interval for the quantile $Q_{\exp(X)}(\tau)$ with $\tau = 0.1, 0.3, 0.5, 0.7, 0.9$. The 95% confidence interval for the average *ex post* value $E[\exp(X)]$ in US dollars per acre is $[146, 219]$. This is close to, but a little above the 95% confidence interval for the median, which is $[131, 190]$. The 95% confidence intervals for the five quantile points highlight a large variation of the *ex post* values of the mineral rights across tracts.

6 Simulation Studies

In this section, we present and discuss finite-sample performance of the proposed method by simulation studies. Simulation outcomes that we present include the size under the null of the true distribution, the power under alternative distributions, and the lengths of confidence bands. The lengths will be further decomposed into the bias bound η_q and the remaining lengths due to the stochastic part.

6.1 Simulation Setting

We employ three distribution families to generate the latent variable X – the normal distribution, the skew normal distribution, and the t distribution. We employ the skew normal distribution and the t distribution to see whether our method is effective for asymmetric distributions and super-Gaussian tails, respectively. Specifically, we generate a random sample of (X, U_1, U_2) mutually independently according to the marginal laws:

$$\begin{aligned}
&\text{Model 1:} \quad X \sim N(\xi_1, \xi_2^2), && U_1 \sim N(0, \sigma_{U_1}^2), \quad U_2 \sim N(0, \sigma_{U_2}^2) \\
&\text{Model 2:} \quad X \sim SN(\xi_1, \xi_2, \xi_3), && U_1 \sim N(0, \sigma_{U_1}^2), \quad U_2 \sim N(0, \sigma_{U_2}^2) \\
&\text{Model 3:} \quad X \sim t_{\xi_4}, && U_1 \sim N(0, \sigma_{U_1}^2), \quad U_2 \sim N(0, \sigma_{U_2}^2)
\end{aligned}$$

Here, $N(\xi_1, \xi_2^2)$ denotes the normal distribution with mean ξ_1 and variance ξ_2^2, $SN(\xi_1, \xi_2, \xi_3)$ denotes the skew normal distribution with location ξ_1, scale ξ_2, and shape ξ_3, and t_{ξ_4} denotes the t distribution with ξ_4 degrees of freedom. The distribution parameters for the latent variable X are set to $(\xi_1, \xi_2) = (0, 1)$ for Model 1, $(\xi_1, \xi_2, \xi_3) = (0, 1, 1)$ for Model 2, and $\xi_4 = 10$ for Model 3. The choice of the normal error distribution, which is an instance of super-smooth distributions, imposes a difficult case in deconvolution – see Li and Vuong (1998). The error variance parameters are set to $\sigma_{U_1} = \sigma_{U_2} = 0.5$ in each of the three models. We conduct experiments with three sample sizes $n = 250$, 500, and 1,000, and run 2,500 Monte Carlo iterations for each set of simulations.

In the simulation studies, we use $q = 2$. We have experimented with alternative sieve dimensions $q \in \{4, 6\}$, but the results are qualitatively similar. The number of frequency grid points is set to $L = 50$. The interval on which the confidence band is formed is set to $I = \left[\mathbb{E}[X] - 2\sqrt{\mathrm{Var}(X)}, \mathbb{E}[X] + 2\sqrt{\mathrm{Var}(X)}\right]$, where $\mathbb{E}[X]$ and $\mathrm{Var}(X)$ are the theoretical mean and the theoretical variance, respectively, of X under the relevant model. Throughout, we use the critical value $c(\alpha, \boldsymbol{\theta})$ based on 1,000 multiplier bootstrap iterations. The level is set to $\alpha = 0.05$ throughout.

6.2 Simulation Results

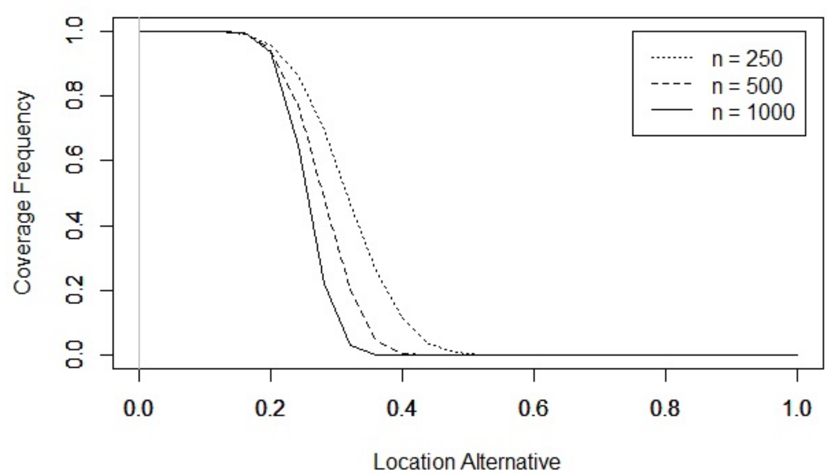
Figure 2 (A) shows the simulated frequencies that the confidence band formed under Model 1 covers alternative probability density functions for $N(\xi_1, \xi_2^2)$ indexed by location parameter values $\xi_1 \in [0.0, 1.0]$ while the scale parameter is fixed at the true value $\xi_2 = 1.0$. The coverage frequency under $\xi_1 = 0.0$ indicate (the complement of) the size, whereas the coverage frequencies under $\xi_1 \in (0.0, 1.0]$ indicate (the complement of) the power. Similarly, Figure 2 (B) shows the simulated frequencies that the confidence band formed under Model 1 covers alternative probability density functions for $N(\xi_1, \xi_2^2)$ indexed by scale parameter values $\xi_2 \in [1.0, 2.0]$ while the location parameter is fixed at the true value $\xi_1 = 0.0$. These results show the correct size and increasing power. The size entails over-coverage, which is still consistent with our theory on size control. Although we present these coverage and power results only for Model 1, we observe similar qualitative patterns for Models 2 and 3.

We display instances of confidence bands in Figure 3. The gray shades indicate the confidence bands including the bias bound and the stochastic parts together. The internal dark gray shades include only the stochastic parts. We also plot the true density functions and Li-Vuong estimates (with the choice of tuning parameter according to Appendix B.2) as solid and dashed curves, respectively. While such instances of confidence bands will not tell us any evidence on the statistical properties, they at least inform how a confidence band may look in applications.

7 Conclusion

Since its introduction to econometrics by Li and Vuong (1998), Kotlarski's identity (Kotlarski, 1967) – see also Rao (1992) – has been widely used in empirical economics. Examples include applications to empirical auctions (e.g., Li et al., 2000; Krasnokutskaya, 2011), education and labor economics (e.g., Carneiro et al., 2003; Cunha et al., 2005, 2010; Arcidiacono et al., 2011; Bonhomme and Sauder, 2011; Kennan and Walker, 2011; Taber and Vejlin, 2020), and earnings dynamics (e.g., Bonhomme and Robin, 2010; Botosaru and Sasaki, 2018; Hu et al., 2019). Despite

(A) Coverage frequencies under location alternatives in Model 1



(B) Coverage frequencies under scale alternatives in Model 1

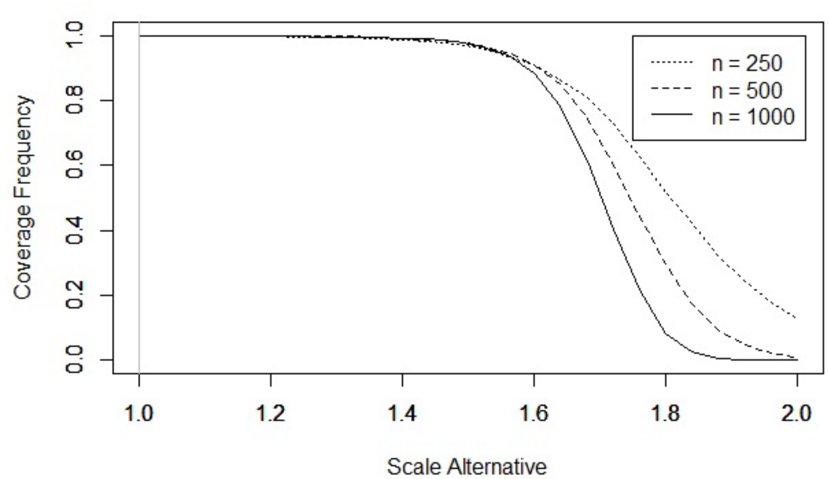


Figure 2: The simulated frequencies that the confidence band formed under Model 1 covers alternative probability density functions for $N(\xi_1, \xi_2^2)$. Panel (A) runs across alternative location parameter values $\xi_1 \in [0.0, 1.0]$ while the scale parameter is fixed at the true value $\xi_2 = 1.0$. Panel (B) runs across alternative scale parameter values $\xi_2 \in [1.0, 2.0]$ while the location parameter is fixed at the true value $\xi_1 = 0.0$.

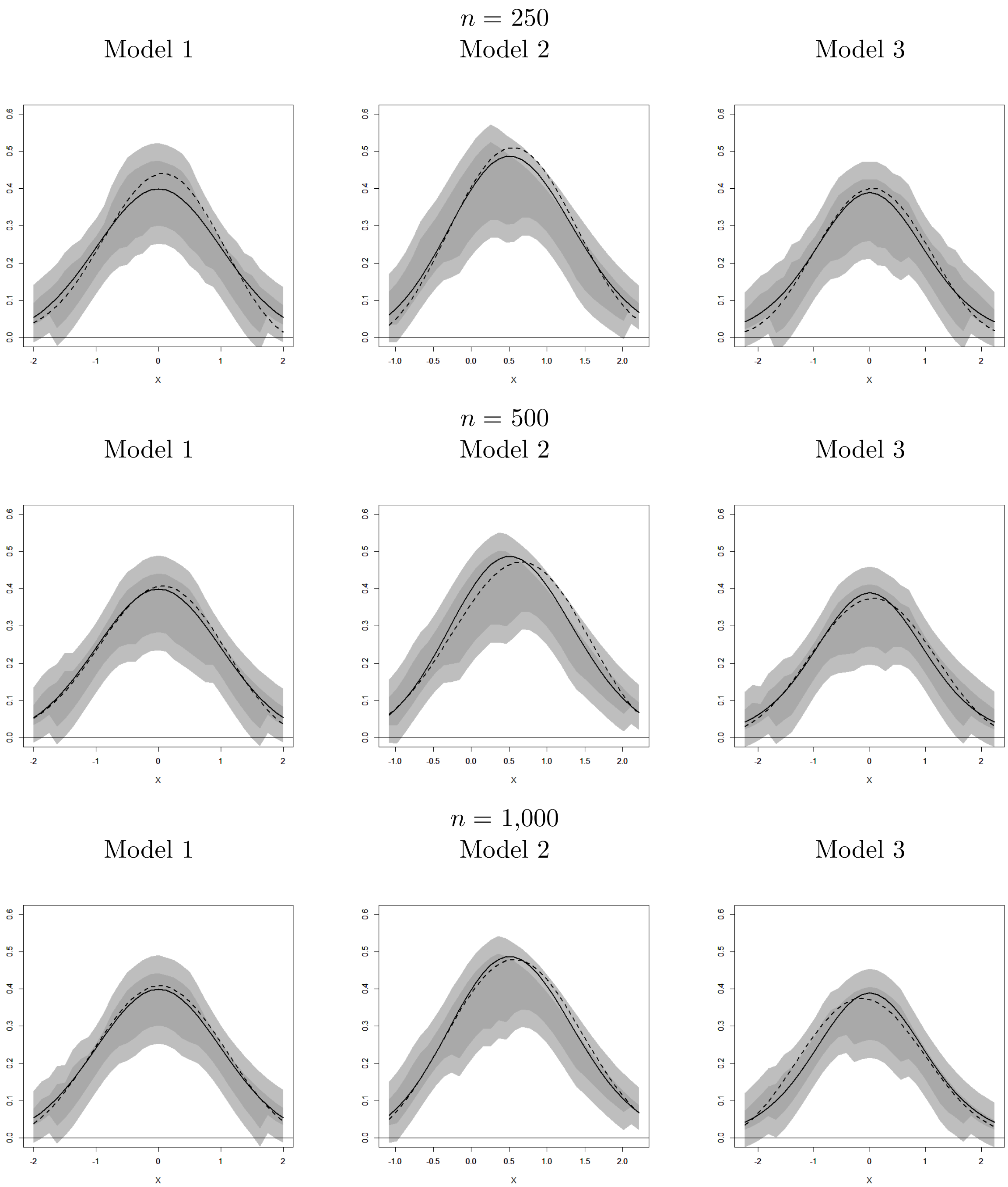


Figure 3: Instances of confidence bands. The gray shades indicate the confidence bands. The internal dark gray shades indicate the stochastic parts of the confidence bands. The solid and dashed curves indicate the true density functions and Li-Vuong estimates, respectively.

its popular use in applications, a method of inference based on Kotlarski's identity has long been missing in the literature. After twenty years since Li and Vuong (1998), we now propose a method of inference based on Kotlarski's identity. Specifically, we develop confidence bands for the probability density function f_X of X in the repeated measurement model where two measurements (Y_1, Y_2) of unobserved variable X are available in data with additive independent errors, $U_1 = Y_1 - X$ and $U_2 = Y_2 - X$.

Our construction of confidence bands can be summarized as follows. First, we derive linear complex-valued moment restrictions based on Kotlarski's identity. Second, we let the Hermite orthogonal sieve approximate unknown probability density functions. Third, for a given sieve dimension and for a given class for probability density functions, we compute a bias bound for the linear complex-valued moment restrictions, and slack the linear complex-valued moment restrictions by this bias bound. Fourth, we compute the uniform norm of the self-normalized process of the slacked linear complex-valued moment restrictions as the test statistics for each point in a set of sieve coefficients. Fifth, inverting this test statistic yields a confidence set of sieve approximations to possible probability density functions. Sixth, for a given sieve dimension and for a given class for probability density functions, we compute a bias bound for sieve approximations of probability density functions, and the desired confidence band is obtained by uniformly enlarging the set of sieve approximations by this bias bound.

We not only provide a method that is guaranteed to work theoretically, but also care for its practicality. The Fourier transform and the inverse Fourier transform operations are known to be computationally costly in the deconvolution literature. By exploiting the property of the Hermite functions as eigen-functions of the Fourier transform operator, we propose to let the Hermite orthogonal sieve approximate both the density and characteristic functions without having to implement numerical integrations within each iteration of a numerical optimization routine. This convenient feature of the proposed method saves computational resources. Simulation studies indeed conclude reasonably fast with informative inference results. The results evidence the efficacy of the proposed method. Since latent-variable models with repeated measurements and panel data are of use in a number of applied fields, including empirical auctions, income dynamics, and labor economics, we hope that our method will contribute to the practice of economic analyses in these and other topics.

We conjecture that the proposed methodology can be extended to related models. For example, Cunha et al. (2010) consider a nonlinear factor model for the evolution of unobserved multidimensional skills. Our current methodology is not readily applicable to their model because we focus on the univariate analysis for X. We speculate that our proposed strategy will work for the inference in a more complicated model, such as that of Cunha et al. (2010). Namely, as in this paper, it could be possible to construct a confidence set for the sieve coefficient of the unknown functions and use a bias bound (η_q in this paper) to enlarge the confidence set. Since it is beyond the scope of this paper, we leave such extensions for future work.

References

Adusumilli, K., D. Kurisu, T. Otsu, and Y.-J. Whang (2019): "Inference on distribution functions under measurement error," *Journal of Econometrics*.

Anderson, T. W. and H. Rubin (1949): "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *The Annals of Mathematical Statistics*, 20, 46–63.

Andrews, D. W. K. and X. Shi (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81, 609–666.

——— (2017): "Inference based on many conditional moment inequalities," *Journal of Econometrics*, 196, 275 – 287.

Arcidiacono, P., E. M. Aucejo, H. Fang, and K. I. Spenner (2011): "Does affirmative action lead to mismatch? A new test and evidence," *Quantitative Economics*, 2, 303–333.

Armstrong, T. B. and M. Kolesár (2018): "Simple and Honest Confidence Intervals in Nonparametric Regression," Working paper.

Babii, A. (2018): "Honest Confidence Sets in Nonparametric IV Regression and Other Ill-posed Models," Working paper.

Bissantz, N., L. Dümbgen, H. Holzmann, and A. Munk (2007): "Nonparametric Confidence Bands in Deconvolution Density Estimation," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69, 483–506.

Bissantz, N. and H. Holzmann (2008): "Statistical Inference for Inverse Problems," *Inverse Problems*, 24, 034009.

Blanchard, P. and E. Bruening (2002): *Mathematical Methods in Physics: Distributions, Hilbert Space Operators, and Variational Methods*, Birkhäuser, 1 ed.

Bonhomme, S. and J.-M. Robin (2010): "Generalized Non-Parametric Deconvolution with an Application to Earnings Dynamics," *The Review of Economic Studies*, 77, 491–533.

Bonhomme, S. and U. Sauder (2011): "Recovering Distributions in Difference-in-Differences Models: A Comparison of Selective and Comprehensive Schooling," *Review of Economics and Statistics*, 93, 479–494.

Botosaru, I. and Y. Sasaki (2018): "Nonparametric Heteroskedasticity in Persistent Panel Processes: An Application to Earnings Dynamics," *Journal of Econometrics*, 203, 283–296.

Boyd, J. P. (1984): "Asymptotic coefficients of Hermite function series," *Journal of Computational Physics*, 54, 382–410.

Canay, I. A., A. Santos, and A. M. Shaikh (2013): "On the testability of identification in some nonparametric models with endogeneity," *Econometrica*, 81, 2535–2559.

Carneiro, P., K. T. Hansen, and J. J. Heckman (2003): "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College," Tech. rep., National Bureau of Economic Research.

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

Carrasco, M. and J.-P. Florens (2011): "A spectral method for deconvolving a density," *Econometric Theory*, 27, 546–581.

Carroll, R. J. and P. Hall (1988): "Optimal Rates of Convergence for Deconvolving a Density," *Journal of the American Statistical Association*, 83, 1184–1186.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. E. Leamer, Elsevier, chap. 76, 5549–5632.

Chen, X. and T. M. Christensen (2018): "Optimal Sup-norm Rates and Uniform Inference on Nonlinear Functionals of Nonparametric IV Regression," *Quantitative Economics*, 9, 39–84.

Chen, X., H. Hong, and D. Nekipelov (2011): "Nonlinear Models of Measurement Errors," *Journal of Economic Literature*, 49, 901–937.

Chernozhukov, V., D. Chetverikov, and K. Kato (2018): "Inference on causal and structural parameters using many moment inequalities," *Review of Economic Studies*, forthcoming.

Comte, F. and C. Lacour (2011): "Data-Driven Density Estimation in the Presence of Additive Noise with Unknown Distribution," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73, 601–627.

Cunha, F., J. Heckman, and S. Navarro (2005): "Separating Uncertainty from Heterogeneity in Life Cycle Earnings," *Oxford Economic Papers*, 57, 191–261.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78, 883–931.

Delaigle, A. and I. Gijbels (2004): "Practical Bandwidth Selection in Deconvolution Kernel Density Estimation," *Computational Statistics & Data Analysis*, 45, 249–267.

Delaigle, A. and P. Hall (2015): "Methodology for Non-Parametric Deconvolution When the Error Distribution Is Unknown," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78, 231–252.

Delaigle, A., P. Hall, and A. Meister (2008): "On Deconvolution with Repeated Measurements," *Annals of Statistics*, 36, 665–685.

D'Haultfoeuille, X. (2011): "On the completeness condition in nonparametric instrumental problems," *Econometric Theory*, 27, 460–471.

Diggle, P. J. and P. Hall (1993): "A Fourier Approach to Nonparametric Deconvolution of a Density Estimate," *Journal of the Royal Statistical Society. Series B (Methodological)*, 55, 523–531.

Efromovich, S. (1997): "Density Estimation for the Case of Supersmooth Measurement Error," *Journal of the American Statistical Association*, 92, 526–535.

Evdokimov, K. (2010): "Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity," Working Paper.

Evdokimov, K. and H. White (2012): "Some extensions of a lemma of Kotlarski," *Econometric Theory*, 28, 925–932.

Fan, J. (1991a): "Asymptotic Normality for Deconvolution Kernel Density Estimators," *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 53, 97–110.

——— (1991b): "On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems," *The Annals of Statistics*, 19, 1257–1272.

Feuerverger, A. and R. A. Mureika (1977): "Empirical characteristic function and its applications," *Annals of Statistics*, 5, 88–97.

Firpo, S., A. F. Galvao, and S. Song (2017): "Measurement errors in quantile regression models," *Journal of econometrics*, 198, 146–164.

Folland, G. B. (2007): *Real Analysis: Modern Techniques and Their Applications*, Wiley, 2 ed.

Gallant, A. and D. W. Nychka (1987): "Semi-nonparametric Maximum Likelihood Estimation," *Econometrica*, 55, 363–90.

Guerre, E., I. Perrigne, and Q. Vuong (2000): "Optimal nonparametric estimation of first-price auctions," *Econometrica*, 68, 525–574.

Hall, P. and J. Horowitz (2013): "A Simple Bootstrap Method for Constructing Nonparametric Confidence Bands for Functions," *Annals of Statistics*, 41, 1892–1921.

Hendricks, K., R. H. Porter, and B. Boudreau (1987): "Information, returns, and bidding behavior in OCS auctions: 1954-1969," *The Journal of Industrial Economics*, 517–542.

Horowitz, J. L. and S. Lee (2012): "Uniform Confidence Bands for Functions Estimated Nonparametrically with Instrumental Variables," *Journal of Econometrics*, 168, 175–188.

Horowitz, J. L. and M. Markatou (1996): "Semiparametric Estimation of Regression Models for Panel Data," *The Review of Economic Studies*, 63, 145–168.

Hu, Y., R. Moffitt, and Y. Sasaki (2019): "Semiparametric Estimation of the Canonical Permanent-Transitory Model of Earnings Dynamics," *Quantitative Economics*, 10, 1495–1536.

Johannes, J. (2009): "Deconvolution with Unknown Error Distribution," *Annals of Statistics*, 37, 2301–2323.

Kato, K. and Y. Sasaki (2018): "Uniform confidence bands in deconvolution with unknown error distribution," *Journal of Econometrics*, 207, 129–161.

——— (2019): "Uniform confidence bands for nonparametric errors-in-variables regression," *Journal of Econometrics*.

Kennan, J. and J. R. Walker (2011): "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 79, 211–251.

Kotlarski, I. (1967): "On Characterizing the Gamma and the Normal Distribution." *Pacific Journal of Mathematics*, 20, 69–76.

Krasnokutskaya, E. (2011): "Identification and Estimation of Auction Models with Unobserved Heterogeneity," *The Review of Economic Studies*, 78, 293–327.

Kurisu, D. and T. Otsu (2019): "On the uniform convergence of deconvolution estimators from repeated measurements," Tech. Rep. 604, Suntory and Toyota International Centres for Economics and Related Disciplines.

Li, T. (2002): "Robust and Consistent Estimation of Nonlinear Errors-in-Variables Models," *Journal of Econometrics*, 110, 1 – 26.

Li, T., I. Perrigne, and Q. Vuong (2000): "Conditionally Independent Private Information in OCS Wildcat Auctions," *Journal of Econometrics*, 98, 129 – 161.

Li, T. and Q. Vuong (1998): "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators," *Journal of Multivariate Analysis*, 65, 139–165.

Lounici, K. and R. Nickl (2011): "Global uniform risk bounds for wavelet deconvolution estimators," *Annals of Statistics*, 39, 201–231.

Neumann, M. H. and O. Hössjer (1997): "On the Effect of Estimating the Error Density in Nonparametric Deconvolution," *Journal of Nonparametric Statistics*, 7, 307–330.

Rao, B. (1992): *Identifiability in Stochastic Models: Characterization of Probability Distributions*, Probability and mathematical statistics, Academic Press.

Santos, A. (2011): "Instrumental variable methods for recovering continuous linear functionals," *Journal of Econometrics*, 161, 129–146.

——— (2012): "Inference in nonparametric instrumental variables with partial identification," *Econometrica*, 80, 213–275.

Schennach, S. M. (2004a): "Estimation of Nonlinear Models with Measurement Error," *Econometrica*, 72, 33–75.

——— (2004b): "Nonparametric Regression in the Presence of Measurement Error," *Econometric Theory*, 20, 1046–1093.

——— (2008): "Quantile regression with mismeasured covariates," *Econometric Theory*, 24, 1010–1043.

——— (2015): "A Bias Bound Approach to Nonparametric Inference," CeMMAP working papers CWP71/15, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

——— (2016): "Recent Advances in the Measurement Error Literature," *Annual Review of Economics*, 8, 341–377.

Schmidt-Hieber, J., A. Munk, and L. Dümbgen (2013): "Multiscale methods for shape constraints in deconvolution: Confidence statements for qualitative features," *Ann. Statist.*, 41, 1299–1328.

Song, S., S. M. Schennach, and H. White (2015): "Estimating nonseparable models with mismeasured endogenous variables," *Quantitative Economics*, 6, 749–794.

Stefanski, L. A. and R. J. Carroll (1990): "Deconvolving Kernel Density Estimators," *Statistics*, 21, 169–184.

Taber, C. and R. Vejlin (2020): "Estimation of a Roy/search/compensating differential model of the labor market," *Econometrica*, 88, 1031–1069.

van Es, B. and S. Gugushvili (2008): "Weak Convergence of the Supremum Distance for Supersmooth Kernel Deconvolution," *Statistics & Probability Letters*, 78, 2932–2938.

Zinde-Walsh, V. (2014): "Identification and Well-Posedness in Nonparametric Models with Independence Conditions," in *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, ed. by J. S. Racine, L. Su, and A. Ullah, Oxford: Oxford University Press.

Appendix

A Proofs for the Results in the Main Text

A.1 Proof of Theorem 1 (Linear Complex-Valued Moment Restrictions)

Proof. By Assumption 1 (ii), we have the following three equations:

$$\begin{aligned}
\mathbb{E}_P\left[iX\exp(itX)\exp(itU_2)\right] &= \mathbb{E}_P\left[iX\exp(itX))\right]\mathbb{E}_P\left[\exp(itU_2)\right] \\
\mathbb{E}_P\left[iU_1\exp(itX)\exp(itU_2)\right] &= \mathbb{E}_P\left[i\mathbb{E}_P\left[U_1 \mid X, U_2\right]\exp(itX)\exp(itU_2)\right] = 0 \\
\mathbb{E}_P\left[\exp(itX)\exp(itU_2)\right] &= \mathbb{E}_P\left[\exp(itX)\right]\mathbb{E}_P\left[\exp(itU_2)\right]
\end{aligned}$$

for every real t, where we use $E[U_1 \mid X, U_2] = 0$ and the independence between X and U_2. By (1),

$$\begin{aligned}
\mathbb{E}_P\left[\left(iY_1\phi_X(t) - \phi_X^{(1)}(t)\right)\exp(itY_2)\right] &= \mathbb{E}_P\left[iX\exp(itX)\exp(itU_2)\right]\phi_X(t) \\
&\quad + \mathbb{E}_P\left[iU_1\exp(itX)\exp(itU_2)\right]\phi_X(t) \\
&\quad - \mathbb{E}_P\left[\exp(itX)\exp(itU_2)\right]\phi_X^{(1)}(t) \\
&= \mathbb{E}_P\left[iX\exp(itX))\right]\mathbb{E}_P\left[\exp(itU_2)\right]\phi_X(t) \\
&\quad - \mathbb{E}_P\left[\exp(itX)\right]\mathbb{E}_P\left[\exp(itU_2)\right]\phi_X^{(1)}(t) \\
&= 0
\end{aligned}$$

for every real t, where the last equality follows from $\phi_X^{(1)}(t) = \mathbb{E}_P\left[iX\exp(itX))\right]$. □

A.2 Proof of Lemma 2

Proof. Regarding the first statement, the Hermite basis expansion for the pdf $f_X = \sum_{j=0}^{\infty}\langle f_X, \psi_j\rangle \cdot \psi_j(x)$ implies

$$\left|\left(\int \boldsymbol{\psi}(x)dx\right)^T \boldsymbol{\theta} - 1\right| = \left|\left(\int \boldsymbol{\psi}(x)dx\right)^T \boldsymbol{\theta} - \int f_X(x)dx\right| = \left|\int \sum_{j=q+1}^{\infty}\langle f_X, \psi_j\rangle \cdot \psi_j(x)dx\right|.$$

By the triangle inequality, Assumption 2, and the definition of η_q,

$$\left|\left(\int \boldsymbol{\psi}(x)dx\right)^T \boldsymbol{\theta} - 1\right| \leq \sum_{j=q+1}^{\infty} |\langle f_X, \psi_j\rangle| \int |\psi_j(x)|\, dx \leq \sum_{j=q+1}^{\infty} j^{-3} \int |\psi_j(x)|\, dx \leq \eta_q.$$

Regarding the second statement, the Hermite basis expansion implies

$$\inf_{x\in\mathbb{R}} \boldsymbol{\psi}(x)^T\boldsymbol{\theta} \geq \inf_{x\in\mathbb{R}} \left(\boldsymbol{\psi}(x)^T\boldsymbol{\theta} - f_X(x)\right) = -\sup_{x\in\mathbb{R}} \sum_{j=q+1}^{\infty} \langle f_X, \psi_j\rangle \cdot \psi_j(x).$$

By the triangle inequality, Assumption 2, and the definition of η_q,

$$\inf_{x\in\mathbb{R}} \boldsymbol{\psi}(x)^T\boldsymbol{\theta} \geq -\sum_{j=q+1}^{\infty} |\langle f_X, \psi_j\rangle| \cdot \sup_{x\in\mathbb{R}} |\psi_j(x)| \geq -\sum_{j=q+1}^{\infty} j^{-3} \sup_{x\in\mathbb{R}} |\psi_j(x)| \geq -\eta_q.$$

Regarding the last two statements, $\mathbb{E}_P\left[\left(iY_1\phi_X(t) - \phi_X^{(1)}(t)\right)\exp(itY_2)\right] = 0$ implies

$$\begin{aligned}
&\sqrt{|\mathbb{E}_P[\mathbf{R}_{t_l}]^T\boldsymbol{\theta}|^2 + |\mathbb{E}_P[\mathbf{I}_{t_l}]^T\boldsymbol{\theta}|^2} \\
&\quad = \left|\mathbb{E}_P\left[\left(iY_1 \sum_{j=0}^{q}\langle f_X, \psi_j\rangle \cdot \phi_j(t) - \sum_{j=0}^{q}\langle f_X, \psi_j\rangle \cdot \phi_j^{(1)}(t)\right)\exp(itY_2)\right]\right| \\
&\quad = \left|\mathbb{E}_P\left[\left(iY_1 \sum_{j=q+1}^{\infty}\langle f_X, \psi_j\rangle \cdot \phi_j(t) - \sum_{j=q+1}^{\infty}\langle f_X, \psi_j\rangle \cdot \phi_j^{(1)}(t)\right)\exp(itY_2)\right]\right|.
\end{aligned}$$

By the triangle inequality, Assumption 2, and the definition of δ_q,

$$\begin{aligned}
&\sqrt{|\mathbb{E}_P[\mathbf{R}_{t_l}]^T\boldsymbol{\theta}|^2 + |\mathbb{E}_P[\mathbf{I}_{t_l}]^T\boldsymbol{\theta}|^2} \\
&\quad \leq \mathbb{E}_P\left[\left(|Y_1| \sum_{j=q+1}^{\infty} |\langle f_X, \psi_j\rangle| \cdot |\phi_j(t)| + \sum_{j=q+1}^{\infty} |\langle f_X, \psi_j\rangle| \cdot |\phi_j^{(1)}(t)|\right)|\exp(itY_2)|\right] \\
&\quad = \mathbb{E}_P\left[|Y_1|\right] \sum_{j=q+1}^{\infty} |\langle f_X, \psi_j\rangle| \cdot |\phi_j(t)| + \sum_{j=q+1}^{\infty} |\langle f_X, \psi_j\rangle| \cdot |\phi_j^{(1)}(t)| \\
&\quad \leq \sum_{j=q+1}^{\infty} j^{-3}\left(\mathbb{E}_P\left[|Y_1|\right]|\phi_j(t)| + |\phi_j^{(1)}(t)|\right) \\
&\quad \leq \delta_q.
\end{aligned}$$

□

A.3 Proof of Lemma 3

Proof. There is some constant $\varepsilon > 0$ such that

$$\left|\mathbb{E}_P\left[\left(iY_1\phi(\bar{t}) - \phi^{(1)}(\bar{t})\right)\exp(i\bar{t}Y_2)\right]\right| > 2\varepsilon,$$

where $\phi = \mathcal{F}f$. Suppose q and Lh are sufficiently large so that

$$\varepsilon \geq \sum_{j=q+1}^{\infty} j^{-3}\left(\mathbb{E}_P\left[|Y_1|\right]\sup_{t\in\mathbb{R}}|\phi_j(t)| + \sup_{t\in\mathbb{R}}|\phi_j^{(1)}(t)|\right) = \delta_q$$

and that

$$\varepsilon \geq \left(\sup_{t\in\mathbb{R}}|\phi^{(2)}(t)| + \mathbb{E}_P\left[|Y_1| + |Y_2| + 1\right]\sup_{t\in\mathbb{R}}|\phi^{(1)}(t)| + \mathbb{E}_P\left[|Y_1| + |Y_1Y_2|\right]\right)\frac{2}{Lh}.$$

Note that $\sup_{t\in\mathbb{R}}|\phi^{(1)}(t)|$ and $\sup_{t\in\mathbb{R}}|\phi^{(2)}(t)|$ are finite by Assumption 2. By the second inequality, we have

$$\begin{aligned}
\left|\frac{\partial}{\partial t}\mathbb{E}_P\left[\left(iY_1\phi(t) - \phi^{(1)}(t)\right)\exp(itY_2)\right]\right| \leq & (\mathbb{E}_P\left[|Y_1|\right] + 1)|\phi^{(1)}(t)| \\
& + \mathbb{E}_P\left[|Y_1|\right] + |\phi^{(2)}(t)| \\
& + \mathbb{E}_P\left[|Y_1Y_2|\right] + \mathbb{E}_P\left[|Y_2|\right]|\phi^{(1)}(t)| \\
= & |\phi^{(2)}(t)| + \mathbb{E}_P\left[|Y_1| + |Y_2| + 1\right]|\phi^{(1)}(t)| + \mathbb{E}_P\left[|Y_1| + |Y_1Y_2|\right] \\
\leq & \frac{Lh}{2}\varepsilon.
\end{aligned}$$

Since $|t_l - \bar{t}| \leq 2/(Lh)$ for some $l = 1, \ldots, L$, we have

$$\begin{aligned}
|\mathbb{E}_P\left[\left(iY_1\phi(t_l) - \phi^{(1)}(t_l)\right)\exp(it_lY_2)\right]| \geq & |\mathbb{E}_P\left[\left(iY_1\phi(\bar{t}) - \phi^{(1)}(\bar{t})\right)\exp(i\bar{t}Y_2)\right]| \\
& - \sup_{t\in\mathbb{R}}\left|\frac{\partial}{\partial t}\mathbb{E}_P\left[\left(iY_1\phi(t) - \phi^{(1)}(t)\right)\exp(itY_2)\right]\right||t_l - \bar{t}| \\
> & 2\varepsilon - \frac{Lh}{2}\varepsilon\frac{2}{Lh} \\
\geq & \delta_q.
\end{aligned}$$

□

A.4 Proof of Theorem 2 (Size Control)

Proof. Let $P \in \mathcal{P}$ and $f \in \mathcal{L}_0(P)$. Let $\boldsymbol{\theta} = (\langle f, \psi_j\rangle : j = 0, 1, \ldots, q)$. By Lemma 2 and Assumption 3, Chernozhukov, Chetverikov, and Kato (2018, Theorem A.1) implies $\mathbb{P}_P(T(\boldsymbol{\theta}) \leq c(\alpha, \boldsymbol{\theta})) \geq 1 - \alpha - Cn^{-c}$ for some positive constants c and C depending on c_1 and C_1. It suffices to show $T(\boldsymbol{\theta}) \leq c(\alpha, \boldsymbol{\theta}) \implies f \in \mathcal{C}_n(\alpha)$. The condition $T(\boldsymbol{\theta}) \leq c(\alpha, \boldsymbol{\theta})$ implies $\boldsymbol{\theta} \in \mathcal{C}_{(\langle f_X, \psi_j\rangle : j=0,1,\ldots,q)}(\alpha)$.

Since

$$\sup_{x\in\mathbb{R}}\left|f(x)-\sum_{j=0}^{q}\theta_j\cdot\psi_j(x)\right|=\sup_{x\in\mathbb{R}}\left|\sum_{j=0}^{\infty}\langle f,\psi_j\rangle\cdot\psi_j(x)-\sum_{j=0}^{q}\theta_j\cdot\psi_j(x)\right|=\sup_{x\in\mathbb{R}}\left|\sum_{j=q+1}^{\infty}\langle f,\psi_j\rangle\cdot\psi_j(x)\right|,$$

by the triangle inequality, Assumption 2, and the definition of η_q, we have

$$\sup_{x\in\mathbb{R}}\left|f(x)-\sum_{j=0}^{q}\theta_j\cdot\psi_j(x)\right|\leq\sum_{j=q+1}^{\infty}|\langle f,\psi_j\rangle|\sup_{x\in\mathbb{R}}|\psi_j(x)|\leq\sum_{j=q+1}^{\infty}j^{-3}\sup_{x\in\mathbb{R}}|\psi_j(x)|\leq\eta_q.$$

By the definition of $\mathcal{C}_n(\alpha)$, we have $f\in\mathcal{C}_n(\alpha)$. □

A.5 Proof of Theorem 3 (Power)

Proof. This proof focuses on the case in (22). The proofs for the cases of (23)–(25) are similar. By the definition of $\mathcal{C}_n(\alpha)$, we can write

$$\begin{aligned}
\mathbb{P}_P(f\notin\mathcal{C}_n(\alpha)) &= \mathbb{P}_P\left(T(\boldsymbol{\theta})>c(\alpha,\boldsymbol{\theta})\text{ for every }\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)\right)\\
&\geq \mathbb{P}_P\left(T(\boldsymbol{\theta})>\sqrt{2\log(4L)}+\sqrt{2\log(1/\alpha)}\text{ for every }\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)\right)\\
&= \mathbb{P}_P\left(\inf_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)}T(\boldsymbol{\theta})>\sqrt{2\log(4L)}+\sqrt{2\log(1/\alpha)}\right),
\end{aligned}$$

where the inequality follows from

$$c(\alpha,\boldsymbol{\theta})\leq\sqrt{2\log(4L)}+\sqrt{2\log(1/\alpha)}$$

– see Chernozhukov, Chetverikov, and Kato (2018, Lemma D.4). If $B_V\leq\nu$, then

$$\begin{aligned}
\inf_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)}T(\boldsymbol{\theta}) &\geq \inf_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)}\sqrt{n}\frac{\mathbb{E}_n[\mathbf{R}_{t_*}]^T\boldsymbol{\theta}-\delta_q}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_n(\mathbf{R}_{t_*})\boldsymbol{\theta}}}\\
&\geq \inf_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)}\frac{\mathbb{G}_n[\mathbf{R}_{t_*}]^T\boldsymbol{\theta}}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_n(\mathbf{R}_{t_*})\boldsymbol{\theta}}}+\inf_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)}\sqrt{n}\frac{\mathbb{E}_P[\mathbf{R}_{t_*}]^T\boldsymbol{\theta}-\delta_q}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_n(\mathbf{R}_{t_*})\boldsymbol{\theta}}}\\
&\geq -\sup_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)}\frac{|\mathbb{G}_n[\mathbf{R}_{t_*}]^T\boldsymbol{\theta}|}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_n(\mathbf{R}_{t_*})\boldsymbol{\theta}}}+\inf_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)}\sqrt{n}\frac{\mathbb{E}_P[\mathbf{R}_{t_*}]^T\boldsymbol{\theta}-\delta_q}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_P(\mathbf{R}_{t_*})\boldsymbol{\theta}+\nu}}\\
&\geq -\sup_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)}\frac{|\mathbb{G}_n[\mathbf{R}_{t_*}]^T\boldsymbol{\theta}|}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_n(\mathbf{R}_{t_*})\boldsymbol{\theta}}}+\sqrt{2\log(4L)}+\sqrt{2\log(1/\alpha)}\\
&\quad+\frac{1}{b}\cdot\mathbb{E}_P\left[\sup_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)}\frac{|\mathbb{G}_n[\mathbf{R}_{t_*}]^T\boldsymbol{\theta}|}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_n(\mathbf{R}_{t_*})\boldsymbol{\theta}}}\right]
\end{aligned}$$

by (22). Thus, we obtain

$$
\begin{aligned}
\mathbb{P}_P(f \notin \mathcal{C}_n(\alpha)) &\geq \mathbb{P}_P\left(\left\{\frac{\sup_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)} \frac{|\mathbb{G}_n[\mathbf{R}_{t*}]^T\boldsymbol{\theta}|}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_n(\mathbf{R}_{t*})\boldsymbol{\theta}}}}{\mathbb{E}_P\left[\sup_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)} \frac{|\mathbb{G}_n[\mathbf{R}_{t*}]^T\boldsymbol{\theta}|}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_n(\mathbf{R}_{t*})\boldsymbol{\theta}}}\right]} < \frac{1}{b}\right\} \cap \{B_V \leq \nu\}\right) \\
&\geq \mathbb{P}_P\left(\frac{\sup_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)} \frac{|\mathbb{G}_n[\mathbf{R}_{t*}]^T\boldsymbol{\theta}|}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_n(\mathbf{R}_{t*})\boldsymbol{\theta}}}}{\mathbb{E}_P\left[\sup_{\boldsymbol{\theta}\in\mathbf{B}_{q+1,\eta_q}(f)} \frac{|\mathbb{G}_n[\mathbf{R}_{t*}]^T\boldsymbol{\theta}|}{\sqrt{\boldsymbol{\theta}^T\mathbb{V}_n(\mathbf{R}_{t*})\boldsymbol{\theta}}}\right]} < \frac{1}{b}\right) - (1 - \mathbb{P}_P(B_V \leq \nu)) \\
&\geq 1 - b - (1 - \mathbb{P}_P(B_V \leq \nu)),
\end{aligned}
$$

where the last inequality is due to Markov's inequality. Therefore, the statement of the theorem follows. □

A.6 Proof of Corollary 1 (Consistency against all fixed alternatives)

Proof. Pick $\nu = \lambda^P_{\max}(t_*)$. With this choice of ν, under our assumption, one of (22)-(25) holds for large n for some $b = b_n \to 0$ sufficiently slowly. The conclusion then follows since $\mathbb{P}_P(B_V \leq \nu) \to 1$. □

A.7 Derivation of Eq. (21) (A Sufficient Condition for Assumption 3 (ii))

Proof. Note that

$$
\begin{aligned}
\left|(\mathbf{R}_t - \mathbb{E}_P[\mathbf{R}_t])^T\boldsymbol{\theta}\right| &\leq \sup_{\psi=\psi_0,\ldots,\psi_q} |R_{\psi,t}(Y_1,Y_2) - \mathbb{E}_P[R_{\psi,t}(Y_1,Y_2)]| \cdot \|\boldsymbol{\theta}\| \\
&\leq \sup_{\psi=\psi_0,\ldots,\psi_q} \left(2|Y_1| + 2\mathbb{E}_P[|Y_1|] + 4|\phi^{(1)}(t)|\right) \cdot \|\boldsymbol{\theta}\| \\
&= \left(2|Y_1| + 2\mathbb{E}_P[|Y_1|] + 4\sup_{\psi=\psi_0,\ldots,\psi_q} |\phi^{(1)}(t)|\right) \cdot \|\boldsymbol{\theta}\|
\end{aligned}
$$

and

$$
\left|(\mathbf{I}_t - \mathbb{E}_P[\mathbf{I}_t])^T\boldsymbol{\theta}\right| \leq \left(2|Y_1| + 2\mathbb{E}_P[|Y_1|] + 4\sup_{\psi=\psi_0,\ldots,\psi_q} |\phi^{(1)}(t)|\right) \cdot \|\boldsymbol{\theta}\|,
$$

because

$$
\begin{aligned}
|R_{\psi,t}(Y_1,Y_2)| &= \left|-\cos(tY_2)(Y_1\mathrm{Im}(\phi(t)) + \mathrm{Re}(\phi^{(1)}(t))) - \sin(tY_2)(Y_1\mathrm{Re}(\phi(t)) - \mathrm{Im}(\phi^{(1)}(t)))\right| \\
&\leq |Y_1| + |\mathrm{Re}(\phi^{(1)}(t))| + |Y_1| + |\mathrm{Im}(\phi^{(1)}(t))| \\
&\leq 2|Y_1| + 2|\phi^{(1)}(t)|.
\end{aligned}
$$

Also note that, for the Hermite functions, we have

$$\begin{aligned}
|\psi_j^{(1)}| &= |i^j\sqrt{2\pi}\psi_j^{(1)}| \\
&= \sqrt{2\pi}|\psi_j^{(1)}| \\
&= \sqrt{2\pi}\left|\sqrt{j/2}\psi_{j-1}^{(1)} - \sqrt{(j+1)/2}\psi_{j+1}^{(1)}\right| \\
&\leq \sqrt{2\pi}(\sqrt{j/2}|\psi_{j-1}^{(1)}| + \sqrt{(j+1)/2}|\psi_{j+1}^{(1)}|) \\
&\leq \sqrt{2\pi}(\sqrt{j/2} + \sqrt{(j+1)/2}) \times 1.086435\pi^{-1/4} \\
&\leq 4\sqrt{j+1}.
\end{aligned}$$

Since $\boldsymbol{\theta}^T\mathbb{V}_P(\mathbf{R}_{t_l})\boldsymbol{\theta} \geq \|\boldsymbol{\theta}\|^2\mathrm{eig}_{\min}(\mathbb{V}_P(\mathbf{R}_t))$ and $\boldsymbol{\theta}^T\mathbb{V}_P(\mathbf{I}_{t_l})\boldsymbol{\theta} \geq \|\boldsymbol{\theta}\|^2\mathrm{eig}_{\min}(\mathbb{V}_P(\mathbf{I}_t))$, we have

$$\begin{aligned}
M_{L,q,k}(\boldsymbol{\theta}, P) &\leq \frac{\max_{1\leq l\leq L}\max\left\{\mathbb{E}_P\left[\left|(\mathbf{R}_{t_l} - \mathbb{E}_P[\mathbf{R}_{t_l}])^T\boldsymbol{\theta}\right|^k\right]^{1/k}, \mathbb{E}_P\left[\left|(\mathbf{I}_{t_l} - \mathbb{E}_P[\mathbf{I}_{t_l}])^T\boldsymbol{\theta}\right|^k\right]^{1/k}\right\}}{\sqrt{\min_{1\leq l\leq L}\min\left\{\boldsymbol{\theta}^T\mathbb{V}_P(\mathbf{R}_{t_l})\boldsymbol{\theta}, \boldsymbol{\theta}^T\mathbb{V}_P(\mathbf{I}_{t_l})\boldsymbol{\theta}\right\}}} \\
&\leq \frac{4\mathbb{E}_P\left[|Y_1|^k\right]^{1/k} + 4\mathbb{E}_P[|Y_1|] + 8\sup_{\psi=\psi_0,\ldots,\psi_q}|\phi^{(1)}(t)|}{\sqrt{\min_{t\in[-T,T]}\min\left\{\mathrm{eig}_{\min}(\mathbb{V}_P(\mathbf{R}_t)), \mathrm{eig}_{\min}(\mathbb{V}_P(\mathbf{I}_t))\right\}}} \\
&\leq \frac{4\mathbb{E}_P\left[|Y_1|^k\right]^{1/k} + 4\mathbb{E}_P[|Y_1|] + 32\sqrt{q+1}}{\sqrt{\min_{t\in[-T,T]}\min\left\{\mathrm{eig}_{\min}(\mathbb{V}_P(\mathbf{R}_t)), \mathrm{eig}_{\min}(\mathbb{V}_P(\mathbf{I}_t))\right\}}}
\end{aligned}$$

and

$$\begin{aligned}
B_{L,q}(\boldsymbol{\theta}, P) &\leq \frac{\mathbb{E}_P\left[\max_{1\leq l\leq L}\max\left\{\left|(\mathbf{R}_{t_l} - \mathbb{E}_P[\mathbf{R}_{t_l}])^T\boldsymbol{\theta}\right|^4, \left|(\mathbf{I}_{t_l} - \mathbb{E}_P[\mathbf{I}_{t_l}])^T\boldsymbol{\theta}\right|^4\right\}\right]^{1/4}}{\sqrt{\min_{1\leq l\leq L}\min\left\{\boldsymbol{\theta}^T\mathbb{V}_P(\mathbf{R}_{t_l})\boldsymbol{\theta}, \boldsymbol{\theta}^T\mathbb{V}_P(\mathbf{I}_{t_l})\boldsymbol{\theta}\right\}}} \\
&\leq \frac{4\mathbb{E}_P\left[|Y_1|^4\right]^{1/4} + 4\mathbb{E}_P[|Y_1|] + 8\sup_{\psi=\psi_0,\ldots,\psi_q}|\phi^{(1)}(t)|}{\sqrt{\min_{t\in[-T,T]}\min\left\{\mathrm{eig}_{\min}(\mathbb{V}_P(\mathbf{R}_t)), \mathrm{eig}_{\min}(\mathbb{V}_P(\mathbf{I}_t))\right\}}} \\
&\leq \frac{4\mathbb{E}_P\left[|Y_1|^4\right]^{1/4} + 4\mathbb{E}_P[|Y_1|] + 32\sqrt{q+1}}{\sqrt{\min_{t\in[-T,T]}\min\left\{\mathrm{eig}_{\min}(\mathbb{V}_P(\mathbf{R}_t)), \mathrm{eig}_{\min}(\mathbb{V}_P(\mathbf{I}_t))\right\}}}
\end{aligned}$$

Since $\mathbb{E}_P\left[Y_1^4\right] < \infty$, we have

$$\begin{aligned}
&\left(M_{L,q,3}^3(\boldsymbol{\theta}, P) \vee M_{L,q,4}^2(\boldsymbol{\theta}, P) \vee B_{L,q}(\boldsymbol{\theta}, P)\right)^2 \\
&= O\left(\left(\frac{\sqrt{q+1}}{\sqrt{\min_{t\in[-T,T]}\min\left\{\mathrm{eig}_{\min}(\mathbb{V}_P(\mathbf{R}_t)), \mathrm{eig}_{\min}(\mathbb{V}_P(\mathbf{I}_t))\right\}}}\right)^6\right) \\
&= O(n^{1/2-c_1}\log^{-7/2}(4Ln)).
\end{aligned}$$

□

B Identification and Estimation from the Previous Literature

This appendix section presents the identification and estimation for the characteristic function φ_X and the density function f_X of X based on Li and Vuong (1998). Moreover, a choice of the tuning parameter based on (Delaigle and Gijbels, 2004) is also reviewed. Although the main text of this paper is focused on inference, one would also want to present estimates along with confidence bands as we presented in Figure 3. This appendix section provides a method of obtaining estimates for convenience of readers.

B.1 Identification and Estimation of the Characteristic Functions

For a joint distribution P of (Y_1, Y_2), Li and Vuong (1998) show that the characteristic functions of X and U_1 are identified by

$$\varphi_X(t) = \exp\left(\int_0^t \frac{i\mathbb{E}_P\left[Y_1 e^{i\tau Y_2}\right]}{\mathbb{E}_P\left[e^{i\tau Y_2}\right]} d\tau\right) \qquad \text{and} \tag{26}$$

$$\varphi_{U_1}(t) = \frac{\mathbb{E}_P\left[e^{itY_1}\right]}{\exp\left(\int_0^t \frac{i\mathbb{E}_P\left[Y_1 e^{i\tau Y_2}\right]}{\mathbb{E}_P\left[e^{i\tau Y_2}\right]} d\tau\right)}$$

respectively, under the assumption of nonvanishing characteristic function of Y_2 in addition to Assumption 1. The sample-counterpart estimator of (26) reads

$$\widehat{\varphi}_X(t) = \exp\left(\int_0^t \frac{i\mathbb{E}_n\left[Y_1 e^{i\tau Y_2}\right]}{\mathbb{E}_n\left[e^{i\tau Y_2}\right]} d\tau\right). \tag{27}$$

Similarly,

$$\widehat{\varphi}_{U_1}(t) = \frac{\mathbb{E}_n\left[e^{itY_1}\right]}{\exp\left(\int_0^t \frac{i\mathbb{E}_n\left[Y_1 e^{i\tau Y_2}\right]}{\mathbb{E}_n\left[e^{i\tau Y_2}\right]} d\tau\right)}.$$

B.2 Tuning Parameter

To estimate the probability density function f_X of X using the characteristic function estimator (26), we need to impose a regularization by limiting the integration for the Fourier transform to a compact interval $[-h^{-1}, h^{-1}]$ for some "bandwidth" h. Finite-sample choice methods of choosing the limit frequency h are proposed in the literature of deconvolution kernel density estimation. One of the most widely used approaches is to minimize the MISE (Stefanski and Carroll, 1990) or its asymptotically dominating part (Delaigle and Gijbels, 2004):

$$AMISE(h) = \frac{1}{2\pi nh} \int \left| \frac{\phi_K(t)}{\varphi_{U_1}(t/h)} \right|^2 dt + \frac{h^4}{4} \int u^2 K(u) du \cdot \int f_X^{(2)}(x)^2 dx.$$

where φ_K, supported on $[-1, 1]$, is $\mathcal{F}K$ for some kernel function K.

There are alternative ways to compute $\int f_X^{(2)}(x)^2 dx$. Based on Parseval's identity, Delaigle and Gijbels (2004) suggest

$$\int f_X^{(2)}(x)^2 dx = \frac{1}{2\pi h^5} \int t^4 \frac{|\varphi_X(t/h)|^2 |\varphi_K(t)|^2}{|\varphi_{U_1}(t/h)|^2} dt.$$

Combining the above two equations together yields

$$AMISE(h) = \frac{1}{2\pi nh} \int \left| \frac{\phi_K(t)}{\varphi_{U_1}(t/h)} \right|^2 dt + \frac{1}{8\pi h} \int u^2 K(u) du \cdot \int t^4 \frac{|\varphi_X(t/h)|^2 |\varphi_K(t)|^2}{|\varphi_{U_1}(t/h)|^2} dt.$$

With this formula, one may choose h to minimize the plug-in counterpart of $AMISE(h)$, replacing the unknown characteristic functions φ_X and φ_{U_1} by the sample counterparts $\widehat{\varphi}_X$ and $\widehat{\varphi}_{U_1}$, respectively, in Appendix B.1.

Since the set $[-h^{-1}, h^{-1}]$ of frequencies is used for estimation, it is also a natural idea to use this set $[-h^{-1}, h^{-1}]$ of frequencies for inference as well, although our theory for inference does not require such a finite limit unlike the estimation which requires regularization.

B.3 Estimation of the Density Function

With the estimated characteristic function (27) and the bandwidth parameter h chosen in Section B.2, the density function may be estimated by

$$\widehat{f}_X(x) = \frac{1}{2\pi} \int e^{-itx} \varphi_K(th) \widehat{\varphi}_X(t) dt.$$

The "Li-Vuong estimates" shown in Section 6 are based on the above formula together with the tuning parameter chosen according to the procedure outlined in Appendix B.2.