Econometric Theory, **27**, 2011, 546–581.
doi:10.1017/S026646661000040X

A SPECTRAL METHOD FOR DECONVOLVING A DENSITY

Marine Carrasco
University of Montreal

Jean-Pierre Florens
Toulouse School of Economics

We propose a new estimator for the density of a random variable observed with an additive measurement error. This estimator is based on the spectral decomposition of the convolution operator, which is compact for an appropriate choice of reference spaces. The density is approximated by a sequence of orthonormal eigenfunctions of the convolution operator. The resulting estimator is shown to be consistent and asymptotically normal. While most estimation methods assume that the characteristic function (CF) of the error does not vanish, we relax this assumption and allow for isolated zeros. For instance, the CF of the uniform and symmetrically truncated normal distributions have isolated zeros. We show that, in the presence of zeros, the density is identified even though the convolution operator is not one-to-one. We propose two consistent estimators of the density. We apply our method to the estimation of the measurement error density of hourly income collected from survey data.

1. INTRODUCTION

Assume we observe n i.i.d. realizations, $y_1, \ldots, y_n$ of the random variable Y with unknown density h, and Y satisfies

$$Y = X + \varepsilon,$$

where X and ε are mutually independent continuously distributed random variables with probability density functions (p.d.f.) f and g, respectively so that h is the convolution between f and g. Moreover, X and ε are assumed to be unobserved scalar random variables. The aim of this paper is to give a new estimator of f, assuming g is known, that works even if the characteristic function of g has isolated zeros.

The authors thank James Stock for suggesting this line of research to them. They are grateful to Karim Abadir, Gordon Dahl, Joe Horowitz, Geert Ridder, Susanne Schennach, P.C.B Phillips (the editor), Oliver Linton (the co-editor), and three referees for their insightful comments. They thank Liang Hu for excellent research assistance. They also wish to thank the participants of the CEME conference (Rochester, 2001), the Econometric Society Winter Meeting (Atlanta, 2002), and the International Statistical Institute (Sydney, 2005). Carrasco gratefully acknowledges partial financial support from the National Science Foundation and SSHRC. Address correspondence to Marine Carrasco, Universite de Montreal, Departement de Sciences Economiques, CP 6128, succ Centre Ville, Montreal, QC H3C3J7, Canada; e-mail: marine.carrasco@umontreal.ca.

This problem consists in solving for f in the equation

$$h(y) = \int g(y-x) f(x)\,dx. \tag{1.1}$$

Equation (1.1) is an integral equation, and solving (1.1) is typically an ill-posed problem (Tikhonov and Arsenin, 1977). Indeed, the solution f is not continuous in h and hence a small perturbation in h may result in a big error in f. Consequently, some smoothing (or regularization) is needed, and the resulting estimator has a slow rate of convergence. The method we propose here consists in interpreting (1.1) as an integral equation

$$Tf = h, \tag{1.2}$$

where T is a compact operator with respect to well-chosen reference spaces and therefore admits a countably infinite number of singular values. We invert (1.2) using the singular value decomposition of T coupled with a Tikhonov regularization. Hence, our estimator denoted $\hat{f}$ does not rely on the choice of a kernel. Assuming that the characteristic function of g does not vanish, we show that our estimator is consistent and asymptotically normal. If we impose joint assumptions on f and g, more precisely if f is smoother than g, then our estimator achieves a much faster rate of convergence than that obtained without the joint assumptions. In particular, we show that if f and g are the pdf of two normal distributions and the variance of the error (g) is smaller than that of the signal (f), then the rate of the mean integrated square error (MISE) is $n^{-1/2}$, while the rate would be $(\ln(n))^{-2}$ if the only information available was that g is twice differentiable (Fan, 1993). It is interesting to note that slightly strengthening the assumptions may result in a big improvement in the convergence rates. We also propose a data-driven method for selecting the smoothing parameter. Moreover, we investigate the case where g depends on some unknown finite dimensional parameters that are estimated using an auxiliary sample. We show that if the size of the auxiliary model is large enough, the resulting estimator has the same properties as in the case where g is entirely known.

Another contribution of our paper is that we study the identification of f and show that f is identified when the characteristic function of g has isolated zeros, even though T is not injective in this case. Although the main identification result can be found in Devroye (1989), the analysis in terms of injectivity of T is new. Most papers require that the characteristic function (CF) of g be different from zero on the real line. This assumption, however, may be too restrictive. The class of densities for which the characteristic function has isolated real zeros is large and includes, among others, the uniform, the Epanechnikov, the triangular, the symmetrically truncated Laplace, and the symmetrically truncated normal distributions, as well as the convolution of any of these with another (arbitrary) density. Therefore having a method that applies to these cases is highly desirable. We propose two estimators that are robust to the presence of zeros. The first estimator consists in completing $\hat{f}$ by adding the projection of f on the space spanned by the eigenfunctions of T associated with zero. This new estimator is consistent but may take negative values. The second estimator of f is obtained by minimizing a penalized least-squares criterion under the constraint that f is a density. This second estimator is consistent and achieves a faster rate of convergence than the unconstrained estimator.

Now we briefly review the literature. The fact that the deconvolution is an ill-posed inverse problem has been known for a long time. For a survey on ill-posed problems in the statistical literature and examples on deconvolution, see Carroll, van Rooij, and Ruymgaart (1991) and van Rooij and Ruymgaart (1999). The most popular approach to deconvolution is the kernel estimator obtained by applying the Fourier inversion formula to the empirical characteristic function of X. This method was initiated by the seminal papers of Carroll and Hall (1988) and Stefanski and Carroll (1990), later followed by Fan (1991a, 1991b), among others. This technique cannot be applied when the CF of the error vanishes because of the resulting division by zero. Our method is related to those of Walter (1981), Van Rooij and Ruymgaart (1991), Efromovitch (1997), Pensky and Vidakovic (1999), and Carroll and Hall (2004), which also use series approximation.

As mentioned earlier, only a few papers deal with zeros in the CF of g. Devroye's (1989) estimator requires three smoothing parameters. The following papers focus on the deconvolution with a uniform error: Hall, Ruymgaart, van Gaans, and van Rooij (2001), Groeneboom and Jongbloed (2003), Johnstone and Raimondo (2004), and Johnstone, Kerkyacharian, Picard, and Raimondo (2004). In a panel data setting, Neumann (2007) proposes an estimator of the distribution function of X (but not its density), which is robust to the presence of zeros. Finally, Hu and Ridder (2007) show the identification of a model with mismeasured regressors when the CF of the measurement error has isolated zeros. One of the referees pointed out recent contributions by Hall and Meister (2007) and Meister (2007, 2008) that propose a similar solution to the problem of zeros. A detailed comparison between our estimator and theirs is given in Section 4.4 and shows that our paper still contributes significantly to the literature.

The deconvolution problem is encountered in many fields, including chemistry, physics, public health, signal restoration, and economics; see, e.g., Horowitz and Markatou (1996), Postel-Vinay and Robin (2002), and Hu and Ridder (2007). A similar problem is encountered in random coefficients binary choice models where the distribution of the coefficient is nonparametrically estimated; see Ichimura and Thompson (1998) and Gautier and Kitamura (2008). Gautier and Kitamura show that this problem can be recast as a deconvolution with a uniform error on $[-\pi/2, \pi/2]$. The application, which we investigate at the end of the paper, is relative to the measurement error in hourly earnings in the Consumer Population Survey. Although these data are widely used, they are known to be misreported. We estimate the density of the measurement error and find that indeed people tend to underreport their earnings.

The article is organized in the following way: In Section 2 we present the estimator. In Section 3 we establish its rate of convergence and asymptotic normality. Section 4 investigates the case where the characteristic function of the error has isolated zeros. A Monte Carlo study is presented in Section 5. Section 6 applies our method to the measurement error resulting from survey income data. Section 7 concludes. Appendix A explains how to compute the estimator in practice, and in particular how to estimate the eigenfunctions and eigenvalues via simulations. The proofs are in Appendix B.

2. METHOD

2.1. Intuition and Overview

We want to solve the integral equation (1.1) where g is known. Solving (1.1) is a linear inverse problem; see Carrasco, Florens, and Renault (2007) for a review on this topic. Here, T is regarded as an operator from a Hilbert space $\mathcal{H}$ into another Hilbert space $\mathcal{E}$. As we have some flexibility on the choice of $\mathcal{H}$ and $\mathcal{E}$, we select them so that T is compact and hence has a discrete singular value decomposition $\left(\varphi_j, \psi_j, \lambda_j\right)$, $j = 0, 1, 2, \ldots$. Solving (1.1) is an ill-posed problem, because the solution may not be unique and the solution is not stable. We address briefly these two issues. When zero is an eigenvalue of T, i.e., there exists f_0 such that $Tf_0 = 0$, T is not injective. Indeed, the solution to $Tf = h$ is not unique because, for any solution f_1, one can construct another solution, $f = f_1 + f_0$. Consider the least-squares solution[1] $f^\dagger$ of (1.1) of minimal norm. According to Nashed and Wahba (1974), this solution exists and is unique provided[2] $h \in \mathcal{R}(T) + \mathcal{R}(T)^\perp$. This pseudo-solution is given by

$$f^\dagger(x) = \sum_{\{j/|\lambda_j| \neq 0\}} \frac{1}{\lambda_j} \left\langle h, \psi_j \right\rangle \varphi_j(x) = \sum_{\{j/|\lambda_j| \neq 0\}} \left\langle f, \varphi_j \right\rangle \varphi_j(x). \quad \textbf{(2.1)}$$

We see that $f^\dagger$ coincides with f only if 0 is not an eigenvalue of T. A solution of the form $f^\dagger$ is not stable in the sense that a small perturbation in h may cause a large variation in $f^\dagger$. As a result, some stabilization or regularization of the solution is needed. We apply here the so-called Tikhonov regularization, which consists in adding a small penalization term to T^*T before inverting it (T^* denotes the adjoint of T). The regularized solution is given by

$$\hat{f} = \left(T^*T + \alpha I\right)^{-1} T^* h. \quad \textbf{(2.2)}$$

Using the spectral decomposition of T^*T, the solution (2.2) can be rewritten as

$$\hat{f}(x) = \sum_{j=0}^{\infty} \frac{1}{\lambda_j^2 + \alpha} \left\langle T^* h, \varphi_j \right\rangle \varphi_j(x) \quad \textbf{(2.3)}$$

$$= \sum_{j=0}^{\infty} \frac{\lambda_j}{\alpha + \lambda_j^2} \left\langle h, \psi_j \right\rangle \varphi_j(x), \quad \textbf{(2.4)}$$

where ψ_j is such that $T\varphi_j = \lambda_j \psi_j$. In practice, $\langle h, \psi_j \rangle$ is replaced by its sample counterpart, as explained later. The regularization parameter α is a smoothing parameter that needs to converge to zero at a certain rate, so that $\hat{f}$ converges to $f^\dagger$ as the sample size n goes to infinity. This method for estimating f has been mentioned in earlier work, see Walter (1981) and Donoho (1995), but has not been applied systematically because in general T is not a compact operator with respect to $L^2(\mathbb{R})$. Our first contribution consists in defining appropriate spaces of reference with respect to which T is compact and in showing that $\hat{f}$ is a consistent estimator of f provided T is injective. Our second contribution is to investigate the identification and estimation of f when the assumption T injective is not satisfied. Hers T noninjective corresponds to the case where the characteristic function of the error ε is equal to zero for some values. We show that f is identified, provided the zeros are isolated. The estimator (2.3) can still be computed but is no longer consistent. In Section 4 we propose two alternative estimators that are consistent.

2.2. Estimator

The method described above relies on a discrete spectrum of T. However, T considered as an operator from $L^2(\mathbb{R})$ into $L^2(\mathbb{R})$ provided with Lebesgue measure is in general not compact and hence has a continuous spectrum. We want to construct spaces of reference for which T is compact. Let π_X and π_Y be two nonnegative weighting functions; we impose further restrictions on those below. Denote $L^2_{\pi_Y}$ the space of square integrable real-valued functions with respect to π_Y :

$$L^2_{\pi_Y} = \left\{ \psi(y) \text{ such that } \int \psi(y)^2 \pi_Y(y)\,dy < \infty \right\}.$$

The inner product in $L^2_{\pi_Y}$ is defined as

$$\langle \psi_1, \psi_2 \rangle = \int \psi_1(y)\,\psi_2(y)\,\pi_Y(y)\,dy.$$

Similarly, we will define $L^2_{\pi_X}$ and $L^2_{\pi_X^m}$ associated with the functions π_X and π_X^m, respectively, to be introduced below. The inner product in $L^2_{\pi_X}$ is also denoted $\langle .,. \rangle$, and both norms in $L^2_{\pi_Y}$ and in $L^2_{\pi_X}$ are denoted $\|.\|$ without confusion. We define π_X^m as the solution of

$$\pi_X^m(x) = \int g(y-x)\,\pi_Y(y)\,dy. \tag{2.5}$$

Note that if π_Y is a density, then π_X^m can be interpreted as the marginal density of the joint distribution with density $g(y-x)\,\pi_Y(y)$. Now we impose the restrictions below.

Assumption 1.

(a) $L^2_{\pi_X} \subset L^2_{\pi_X^m}$.
(b) $\pi_X(x) = 0 \Rightarrow f(x) = 0$.
(c) $\int f^2(x)\pi_X(x)\,dx < \infty$.

Note that if X and ε are continuous random variables on $\mathbb{R}$ and f is square integrable with respect to Lebesgue measure (which is usually assumed in the deconvolution literature), then one can simply select $\pi_X = 1$ and π_Y an arbitrary pdf. The case $\pi_X = 1$ is important because the rate of convergence of our estimator is expressed in terms of a MISE defined with respect to π_X, and it is customary to define the MISE with respect to Lebesgue measure. On the other hand, in some applications (see Example 1 below), choosing π_X different from 1 simplifies the explicit derivation of the eigenvalues and eigenfunctions.

We formally define T as the operator from $L^2_{\pi_X}$ into $L^2_{\pi_Y}$ that associates to any function $\phi(x)$ of $L^2_{\pi_X}$ a function of $L^2_{\pi_Y}$ as

$$(T\phi)(y) = \int g(y-x)\phi(x)\,dx. \tag{2.6}$$

We define the adjoint, T^*, of T as the solution of $\langle T\varphi, \psi\rangle = \langle \varphi, T^*\psi\rangle$ for all $\varphi \in L^2_{\pi_X}$ and $\psi \in L^2_{\pi_Y}$. It associates to any function $\psi(y)$ of $L^2_{\pi_Y}$ a function of $L^2_{\pi_X}$:

$$\left(T^*\psi\right)(x) = \int \frac{g(y-x)\pi_Y(y)}{\pi_X(x)}\psi(y)\,dy.$$

For convenience, we denote its kernel $\pi_{Y|X}(y|x) = g(y-x)\pi_Y(y)/\pi_X(x)$. In the case where $\pi_X = \pi_X^m$ and π_Y is a density, T and T^* are conditional expectation operators. Indeed, $(T\phi)(y) = \mathrm{E}[\phi(X)|Y=y]$ and $(T^*\psi)(x) = \mathrm{E}[\psi(Y)|X=x]$, where X and Y are supposed to be drawn from π_X and π_Y, respectively. Note that Assumption 1(a) guarantees that $\phi \in L^2_{\pi_X} \Rightarrow \phi \in L^2_{\pi_X^m}$, which itself implies $T\phi \in L^2_{\pi_Y}$ by the law of iterated expectations.

In Assumption 2 below, we give a sufficient condition for T (and T^*) to be a Hilbert-Schmidt operator and therefore to be compact (see Dunford and Schwartz, 1963, p. 1130).

Assumption 2. We have

$$\iint (g(y-x))^2 \frac{\pi_Y(y)}{\pi_X(x)}\,dx\,dy < \infty.$$

Assumption 2 imposes some mild restrictions on $\pi_Y(y)$. Consider for illustration the case where ε follows a standard normal and $\pi_X(x) = 1$. We have $\int (g(y-x))^2\,dx = \frac{1}{2\sqrt{\pi}}$. Hence, Assumption 2 is not satisfied for $\pi_Y = 1$.

However, it is satisfied as soon as π_Y is an arbitrary density, including $\pi_Y = I_{[-1,1]}/2$.

To show consistency, we impose the standard identification condition (T injective), which will be relaxed in Section 3. Primitive sufficient conditions for injectivity are derived in Section 3.

Assumption 3. T is injective.

Assumption 4. There is a constant C such that var $\left[\pi_Y(Y_1)\psi_j(Y_1)\right] < C$ for all $j \geq 0$.

A sufficient condition for Assumption 4 is that the pdf h and π_Y belong to L_∞ that is $\sup|h| < \infty$ and $\sup|\pi_Y| < \infty$. Indeed, the variance equals

$$\operatorname{var}\left[\pi_Y(Y)\psi_j(Y)\right] = \int \pi_Y^2(y)\psi_j^2(y)h(y)\,dy - \left[\int \pi_Y(y)\psi_j(y)h(y)\,dy\right]^2.$$

It is enough to show that the first term is bounded. Indeed, $\int \pi_Y^2(y)\psi_j^2(y)h(y)\,dy \leq (\sup h)(\sup \pi_Y) < \infty$.

As a result of compactness, T has a discrete spectrum. Let $\mu_0 = 1 \geq \mu_1 \geq \mu_2 \ldots$ be the nonnegative eigenvalues of TT^* associated with the orthonormal eigenfunctions φ_j, $j = 0, 1, \ldots$. The $\{\mu_0, \mu_1, \ldots\}$ are also the eigenvalues of T^*T associated with the orthonormal eigenfunctions ψ_j, $j = 0, 1, 2, \ldots$. Let $\lambda_j = \sqrt{\mu_j}$, $j = 0, 1, 2, \ldots$. The λ_j are the singular values and φ_j, $j \geq 0$, ψ_j, $j \geq 0$, the singular functions of T and T^*, respectively. They satisfy

(i) $T\left[\varphi_j(x)\right] = \lambda_j \psi_j(y)$, $j \geq 0$;
(ii) $T^*\left[\psi_j(y)\right] = \lambda_j \varphi_j(x)$, $j \geq 0$;
(iii) $T^*T\left[\varphi_j(x)\right] = \lambda_j^2 \varphi_j(y)$, $j \geq 0$;
(iv) $TT^*\left[\psi_j(y)\right] = \lambda_j^2 \psi_j(y)$, $j \geq 0$.

Since g and π_Y are given, the eigenfunctions are either known explicitly (see Examples 1 and 2 below) or can be estimated via simulations as precisely as wanted (see Appendix A) so that we can consider them as known.

Equation (1.2) is approximated by a well-posed problem using the Tikhonov regularization method

$$\left(\alpha_n I + T^*T\right) f^{\alpha_n} = T^*h,$$

where the penalization term α_n plays the role of the smoothing parameter in the kernel estimation. Here f^{α_n} becomes

$$f^{\alpha_n}(x) = \sum_{j=0}^{\infty} \frac{1}{\alpha_n + \lambda_j^2} \left\langle T^*h, \varphi_j \right\rangle \varphi_j(x). \tag{2.7}$$

The only unknown is T^*h. Note that $(T^*h)(x) = \int h(y)\pi_{Y|X}(y|x)\,dy = \mathrm{E}\left[\pi_{Y|X}(Y|x)\right]$. A natural estimator of T^*h is given by

$$\left(\widehat{T^*h}\right)(x) = \frac{1}{n}\sum_{i=1}^{n}\pi_{Y|X}(y_i|x), \tag{2.8}$$

so that the estimator of f takes the form

$$\hat{f}(x) = \sum_{j=0}^{\infty}\frac{1}{\alpha_n+\lambda_j^2}\left\langle \frac{1}{n}\sum_{i=1}^{n}\pi_{Y|X}(y_i|.),\varphi_j(.)\right\rangle\varphi_j(x). \tag{2.9}$$

The function f^{α_n} can be rewritten in the alternative form: $f^{\alpha_n}(x) = \sum_{j=0}^{\infty}\frac{1}{\alpha_n+\lambda_j^2}\left\langle h, T\varphi_j\right\rangle\varphi_j(x) = \sum_{j=0}^{\infty}\frac{\lambda_j}{\alpha_n+\lambda_j^2}\left\langle h,\psi_j\right\rangle\varphi_j(x) = \sum_{j=0}^{\infty}\frac{\lambda_j}{\alpha_n+\lambda_j^2}\mathrm{E}\left[\overline{\psi_j(y_i)\pi_Y(y_i)}\right]\varphi_j(x)$. Hence another expression of $\hat{f}$ is given by

$$\hat{f}(x) = \sum_{j=0}^{\infty}\frac{\lambda_j}{\alpha_n+\lambda_j^2}\left[\frac{1}{n}\sum_{i=1}^{n}\overline{\psi_j(y_i)\pi_Y(y_i)}\right]\varphi_j(x). \tag{2.10}$$

This expression requires the estimation of ψ_j as well as that of φ_j, however the estimation of ψ_j can be obtained as a byproduct of that of φ_j without much extra calculation, as explained in Appendix A. Note that $\hat{f}$ is not always positive and does not integrate to 1. The popular deconvolution kernel estimator may also take negative values but integrates to 1. In Section 4.3 we propose an alternative estimator that is a density.

Example 1 (**Normal error**)
Assume $\varepsilon \sim \mathcal{N}(0,\sigma^2)$. We set $g(y-x) = \frac{1}{\sigma}\phi\left(\frac{y-x}{\sigma}\right)$, where ϕ denotes the pdf of a standard normal. A simple choice for π_Y is the density of a normal $\mathcal{N}(0,\sigma_Y^2)$. The fact that the true distribution may be totally different does not matter. We need to determine π_X, $\pi_{Y|X}$ and the singular value decomposition of T and T^*

(a) $\pi_X(x) = \pi_X^m(x) = \int g(y-x)\pi_Y(y)\,dy = \frac{1}{\sqrt{\sigma_Y^2+\sigma^2}}\phi\left(\frac{x}{\sqrt{\sigma_Y^2+\sigma^2}}\right)$, so that π_X is the density of a normal $\mathcal{N}(0,\sigma_Y^2+\sigma^2)$.

(b) The kernel of the operator T^* is given by $\pi_{Y|X}(y|x) = \frac{g(y-x)\pi_Y(y)}{\pi_X(x)} = \frac{1}{\sigma\sqrt{\rho}}\phi\left(\frac{y-\rho x}{\sigma\sqrt{\rho}}\right)$, where $\rho = \sigma_Y^2/(\sigma^2+\sigma_Y^2)$.

To calculate the eigenvalues and eigenfunctions, we need to compute T^*T. It is the integral operator from $L^2_{\pi_X}$ into $L^2_{\pi_X}$ defined by $(T^*T\varphi)(x) = \int k(x,s)\varphi(s)\,ds$ with kernel $k(x,s) = \int \pi_{Y|X}(y|s)g(y-x)\,dy = \frac{1}{\sigma\sqrt{1+\rho}}\phi\left(\frac{x-\rho s}{\sigma\sqrt{1+\rho}}\right)$. The eigenfunctions of T^*T, φ_j, are the (generalized) Hermite polynomials[3] orthonormal with respect to π_X and are associated with the eigenvalues $\lambda_j^2 = \rho^j$. Here $\varphi_j(x) = \frac{1}{\sqrt{j!}} \sum_{l=0}^{[j/2]} (-1)^l \frac{(2l)!}{2^l l!} \binom{j}{2l} \left(\frac{x}{\sigma_X}\right)^{j-2l}$, $j = 0, 1, 2, \ldots \varphi_j(x)$, $j=2,3,\ldots$ satisfy the following recursion:

$$\varphi_j(x) = \frac{1}{\sqrt{j}} \left\{ \left(\frac{x}{\sigma_X}\right) \varphi_{j-1}(x) - \sqrt{j-1}\varphi_{j-2}(x) \right\}, \quad \textbf{(2.11)}$$

with $\varphi_0(x) = 1$, $\varphi_1(x) = x/\sigma_X$.

(c) The operator TT^* is the integral operator from $L^2_{\pi_Y}$ to $L^2_{\pi_Y}$ defined by $(TT^*\psi)(y) = \int k(y,s)\psi(s)ds$ with kernel $k(y,s) = \int g(s-x)\pi_{Y|X}(y|x)dx = \frac{1}{\sigma\sqrt{\rho}\sqrt{1+\rho}}\phi\left(\frac{y-\rho s}{\sigma\sqrt{\rho}\sqrt{1+\rho}}\right)$. The eigenfunctions of TT^*, ψ_j, are the (generalized) Hermite polynomials orthonormal with respect to π_Y and are associated with the eigenvalues $\lambda_j^2 = \rho^j$. Here ψ_j are the same as φ_j, with σ_X replaced by σ_Y and x replaced by y.

Example 2 (Error with bounded support)
Here the support of the variable Y does not need to be known, however it is supposed to lie in a compact interval $\left[\underline{A}, \overline{A}\right]$ where $\underline{A}$ and $\overline{A}$ are assumed to be known but they could be estimated by the minimum and maximum observations of Y. Note that the supports of g and f are necessarily included in $\left[\underline{A}, \overline{A}\right]$. We also assume that g is symmetric around zero ($g(-x) = g(x)$). Let $\pi_X = \pi_Y$ be Lebesgue measure on $\left[\underline{A}, \overline{A}\right]$. Any function with bounded support $\left[\underline{A}, \overline{A}\right]$ can be extended to a periodic function of period $L = \overline{A} - \underline{A}$. Hence g admits a Fourier decomposition, $g(x) = \sum_{j\in\mathbf{Z}} \gamma_j \varphi_j(x)$, where $\gamma_j = \langle g, \varphi_j \rangle$ and $\varphi_j(x) = \frac{1}{\sqrt{L}} e^{ij2\pi x/L}$. Moreover, $\{\varphi_j(x)\}$ form an orthonormal basis of $L^2(\pi_X)$, where $L^2(\pi_X)$ denotes the space of square integrable complex-valued functions endowed with the inner product $\langle \varphi, \phi \rangle = \int_{\underline{A}}^{\overline{A}} \varphi(x)\overline{\phi(x)}dx$. By the symmetry of g, the operator T is self-adjoint and its eigenfunctions are $\{\varphi_j(x)\}$. Indeed, we have $(T\varphi_j)(y) = \int_{\underline{A}}^{\overline{A}} g(y-x)\varphi_j(x)dx = \int_{\underline{A}}^{\overline{A}} \sum_{z\in\mathbf{Z}} \gamma_z \varphi_z(y-x) \varphi_j(x)dx = \sum_{z\in\mathbf{Z}} \gamma_z e^{iz2\pi y/L} \int_{\underline{A}}^{\overline{A}} \frac{1}{\sqrt{L}} e^{-iz2\pi x/L} \varphi_j(x)dx = \sum_{z\in\mathbf{Z}} \gamma_z e^{iz2\pi y/L} \langle \varphi_j, \varphi_z \rangle = \sqrt{L}\gamma_j\varphi_j(y) \equiv \lambda_j\varphi_j(y)$. The λ_j can be calculated explicitly as $\lambda_j = \sqrt{L}\langle g, \varphi_j \rangle = \int_{\underline{A}}^{\overline{A}} g(x)e^{ij2\pi x/L}dx = \int_{-\infty}^{\infty} g(x)e^{ij2\pi x/L}dx = \Psi_\varepsilon\left(\frac{j2\pi}{L}\right)$, where Ψ_ε is the characteristic function of g. Note that the λ_j are real because g is even. Our approach differs slightly from that of Section 2.2 because we allow for complex-valued eigenfunctions. The advantage of the present approach is that the λ_j and associated eigenfunctions are known in closed form and do not need to be estimated. We have $\psi_j = \overline{\varphi_j}$. The operator T^*T has eigenfunctions $\{\varphi_j(x)\}$ associated with the eigenvalues λ_j^2, $j \in \mathbb{Z}$. Hence the form of the estimator is

$$\hat{f}(x) = \sum_{j\in\mathbf{Z}} \frac{\lambda_j}{\alpha_n + \lambda_j^2} \left[\frac{1}{n}\sum_{i=1}^{n} \varphi_j(y_i)\right] \overline{\varphi_j(x)}.$$

Here $\hat{f}$ can be seen as a regularized version of Fourier inversion formula. See Efromovich (1997) for a similar approach with a different regularization. Using the fact that $\lambda_{-j} = \lambda_j$ and $\varphi_{-j} = \overline{\varphi_j}$, $\hat{f}$ can be further simplified into

$$\hat{f}(x) = \frac{1}{L(\alpha_n + 1)} + 2\mathrm{Re}\left\{ \sum_{j=1,2...} \frac{\lambda_j}{\alpha_n + \lambda_j^2} \left[\frac{1}{n} \sum_{i=1}^{n} \varphi_j(y_i) \right] \overline{\varphi_j(x)} \right\}. \qquad \textbf{(2.12)}$$

In general, the eigenvalues and eigenfunctions cannot be derived in closed-form. In such circumstances, we rely on simulations to compute the spectral decomposition of the operator T. This is explained in Appendix A. In Section 3.4, we investigate the effect of these simulations on the rate of convergence of our estimator.

3. ASYMPTOTIC PROPERTIES AND SELECTION OF THE SMOOTHING PARAMETER

In this section we study the asymptotic properties of our estimator assuming that λ_j and ϕ_j are known to the researcher; that is, we do not take into account any simulation error.

3.1. Rate of the MISE

The criterion we use is the MISE with respect to π_X. That is,

$$MISE = \mathrm{E}\left[\left\| \hat{f} - f \right\|^2 \right] = \mathrm{E}\left[\int \left(\hat{f}(x) - f(x) \right)^2 \pi_X(x)\, dx \right].$$

The criterion usually employed in the kernel literature (e.g., Stefanski and Carroll, 1990) is the MISE with respect to Lebesgue measure on $\mathbb{R}$. Here $f(x)$ is not assumed to be square-integrable on $\mathbb{R}$**,** therefore we replace the integration with respect to Lebesgue by an integration with respect to $\pi_X(.)$. Remark that if $f(x)$ is square-integrable on $\mathbb{R}$, then we can take $\pi_X = 1$ and our MISE becomes the standard MISE.

The MISE can be rewritten as

$$\begin{aligned} MISE &= \int \mathrm{E}\left(\hat{f}(x) - f^{\alpha_n}(x) + f^{\alpha_n}(x) - f(x) \right)^2 \pi_X(x)\, dx \\ &= \int \mathrm{E}\left(\hat{f}(x) - f^{\alpha_n}(x) \right)^2 \pi_X(x)\, dy + \int \left(f^{\alpha_n}(x) - f(x) \right)^2 \pi_X(x)\, dx \\ &\equiv \mathrm{Var} + Bias^2 \end{aligned}$$

because $\mathrm{E}\left(\hat{f}\right) = f^{\alpha_n}$. As in the kernel estimation, the MISE displays a trade-off between the variance (decreasing in α_n) and the bias (increasing in α_n).

PROPOSITION 1. *Under Assumptions 1 to 4, we have*

$$MISE = \frac{1}{n}\sum_{j=0}^{\infty}\left(\frac{\lambda_j}{\alpha_n+\lambda_j^2}\right)^2 \operatorname{var}\left[\pi_Y(Y_i)\psi_j(Y_i)\right] + \alpha_n^2 \sum_{j=0}^{\infty}\frac{\left\langle f,\varphi_j\right\rangle^2}{\left(\alpha_n+\lambda_j^2\right)^2}. \tag{3.1}$$

The rate of convergence of the MISE depends on the rate at which the inner products $\left\langle f,\varphi_j\right\rangle$, the eigenvalues λ_j, and the terms $\operatorname{var}\left[\pi_Y(Y_i)\psi_j(Y_i)\right]$ converge to zero with j. Under Assumption 4, the term of variance is $O\left(1/\left(\alpha^2 n\right)\right)$. To obtain the rate of the bias, we need extra assumptions on the inner product $\left\langle f,\varphi_j\right\rangle$. Here we investigate the case where f satisfies

$$\sum_{j=1}^{\infty}\frac{\left\langle f,\varphi_j\right\rangle^2}{\lambda_j^{2\beta}} < \infty \tag{3.2}$$

for some $\beta > 0$. Condition (3.2) is equivalent to f belongs to the range of $(T^*T)^{\beta/2}$, in other words, there exists a function $\nu \in L^2_{\pi_X}$ such that $f = (T^*T)^{\beta/2}\nu$. Moreover, for $\beta = 1$, (3.2) is equivalent to f belongs to the range of T^*; see Carrasco, et al. (2007, Prop 3.6). This assumption is standard in the inverse problem literature and is starting to be used in econometrics (see Carrasco, et al.; Blundell, Chen, and Kristensen, 2007). Van Rooij and Ruymgaart (1999, Thm. 4.1) and Hall and Horowitz (2005, Assump. A3) use assumptions of the type $\lambda_j \approx j^{-a}$, $\left\langle f,\varphi_j\right\rangle \approx j^{-b}$ as $j \to \infty$, for some $b > 1/2$. We could use this assumption instead of (3.2), however it rules out the important case of exponentially declining eigenvalues that arise when the errors are normal. Under condition (3.2), the squared bias is $O\left(\alpha^{\beta\wedge 2}\right)$, where $\beta \wedge 2$ denotes the minimum between β and 2, hence the result below.

PROPOSITION 2. *Under Assumptions 1 to 4 and condition (3.2), by selecting a regularization parameter* $\alpha_n = dn^{-1/(\beta\wedge 2+2)}$ *for some* $d > 0$, *we have*

$$MISE = O\left(n^{-\beta\wedge 2/(\beta\wedge 2+2)}\right).$$

The convergence rate given in Proposition 2 is valid under very general hypotheses. It may be improved under a stronger assumption.

Assumption 4′. There exists γ such that

$$\gamma = \max\left\{\tilde{\gamma} \in [0,2] \text{ such that } \sum_{j=0}^{\infty}\lambda_j^{2(1-\tilde{\gamma})}\operatorname{Var}\left(\pi_Y(Y_i)\,\psi_j(Y_i)\right) < \infty\right\}.$$

This assumption is satisfied for $\gamma \geq 0$ under Assumption 4. Given g is known, γ itself can be considered to be known. Consider the case where γ may be positive. By an elementary extension of the proof of Proposition 2, we easily establish that the rate now becomes $MISE = O\left(n^{-\beta\wedge 2/(\beta\wedge 2+2-\gamma)}\right)$ with $\alpha = dn^{-1/(\beta\wedge 2+2-\gamma)}$ for some $d > 0$.

Note that when $\beta > 2$, the MISE in Proposition 2 is $O\left(n^{-1/2}\right)$. This rate could be improved to $O\left(n^{-\beta/(\beta+2)}\right)$ if an alternative regularization method were used, like the iterated Tikhonov, spectral cut-off, or Landweber-Fridman (see Kress, 1999). For a regularization by spectral cut-off, see Hall and Meister (2007). We do not investigate these alternative methods here. For normal errors, the rate in Proposition 2 is clearly much faster than the optimal rate derived by Fan (1993). The reason for this difference is that Fan assumes very little on the function f, while condition (3.2) restricts the class of admissible functions by imposing a relationship between the density of the signal X and that of the error ε. Further insights are provided by the next lemma.

LEMMA 1. *If g is even (or, equivalently, the error has a symmetric distribution around zero) and for $\pi_Y(y) = I[-1,1](y)/2$ and $\pi_X(x) = 1$ for all $x \in \mathbb{R}$, a sufficient condition for condition (3.2) to hold with $\beta = 1$ is*

$$\int \left| \frac{\phi_X(t)}{\Psi_\varepsilon(t)} \right| dt < \infty, \tag{3.3}$$

where ϕ_X and Ψ_ε are the characteristic functions of f and g, respectively.

Condition (3.3) requires that ϕ_X has thinner tails than Ψ_ε. Since the tail behavior of a CF is related to the smoothness of the pdf, this is equivalent to requiring that f be smoother than g (see Ushakov, 1999, Thm. 2.5.4). In the case of f Laplacian, this is a very weak requirement. In the case of f normal, it is less likely to be fulfilled. If both X and ε are normally distributed, (3.3) is satisfied if and only if the variance of the signal (X) is larger than the variance of the error (ε). Another interpretation of this condition is that f can be written as the convolution of g and another distribution. As Van Rooij and Ruymgaart (1991) point out, if g is smooth then h is also smooth, therefore if f is not a priori known to be smooth itself, the problem of recovering a potentially nonsmooth f from a sample of smooth h is particularly hard. To get further insights, we replace condition (3.2) by the following:

For all large j, we have

$$\left|\left\langle f, \varphi_j \right\rangle\right| = O\left(\left|\lambda_j\right|\right). \tag{3.4}$$

PROPOSITION 3. *Under Assumptions 1 to 4 and condition (3.4), by choosing a regularization parameter $\alpha_n = dn^{-1/2}$ for some $d > 0$, we have*

$$MISE = O\left(\frac{1}{n}\sum_{j=0}^{\infty}\left(\frac{\lambda_j}{\alpha_n + \lambda_j^2}\right)^2\right).$$

Proposition 3 permits us to give more precise rates of convergence in the case where the decay rate of λ_j is known.

Example 1 continued (Normal case)
Consider X normally distributed. The following is a corollary of Proposition 3.

COROLLARY 1. *Assume condition (3.4) holds. By choosing a regularization parameter $\alpha_n \le dn^{-1/2}$ for some $d > 0$, we have MISE $= O\left(n^{-1/2}\right)$.*

Example 2 continued (Bounded support case)
When X has bounded support and $\pi_X = \pi_Y = 1$, condition (3.4) is equivalent to requiring that $|\phi_X(t)| \le C\,|\Psi_\varepsilon(t)|$ for large t.

PROPOSITION 4. *Assume that condition (3.4) holds and that $\left|\lambda_j\right| = \left|\Psi_\varepsilon\left(\frac{j2\pi}{L}\right)\right| \sim j^{-a}$, $a > 1/2$. By choosing a regularization parameter $\alpha_n = dn^{-1}$ for some $d > 0$, we obtain MISE $= O\left(n^{-(2a-1)/2a}\right)$.*

For a uniform error[4], $a = 1$ and the rate is $n^{-1/2}$, whereas, for a triangular error, $a = 2$ and the rate is $n^{-3/4}$.

3.2. Asymptotic Normality

Carrasco et al. (2007, Sect. 4) proved the asymptotic normality of inner products $\left\langle \hat{f} - f, \varphi \right\rangle$ for some functions φ. To obtain this result, some restrictions on φ are needed. We could adopt this approach here, but we chose to study the pointwise asymptotic normality instead. The condition on φ will be replaced by a condition on x. Assumptions 5 to 7 below impose some restrictions on the eigenfunctions and the admissible range of values for x. As a result, our asymptotic normality will not hold for all x in general.

Because we have i.i.d. data, a sufficient condition for asymptotic normality

$$\frac{\hat{f}(x) - \mathrm{E}\hat{f}(x)}{\sqrt{\operatorname{var}\left(\hat{f}(x)\right)}} \xrightarrow{L} \mathcal{N}(0,1)$$

is that the Lyapounov's condition holds (Billingsley, 1995, Thm. 27.3); i.e., for some $\delta > 0$,

$$\frac{\mathrm{E}\left|Z_{n1}(x) - \mathrm{E}\left(Z_{n1}(x)\right)\right|^{2+\delta}}{n^{\delta/2}\left[\operatorname{var}\left(Z_{n1}(x)\right)\right]^{1+\delta/2}} \to 0, \tag{3.5}$$

where

$$Z_{ni}(x) = \sum_{j=0}^{\infty} \frac{1}{\alpha_n + \lambda_j^2} \left\langle \pi_{Y|X}(Y_i|.), \varphi_j(.) \right\rangle \varphi_j(x). \tag{3.6}$$

Note that from (2.9), $\hat{f} = \sum_{i=1}^{n} Z_{ni}/n$. The condition (3.5) is satisfied under the next assumptions.

Assumption 5. We have

$$\frac{1}{n^{1/2}} \sum_j \left(\frac{\lambda_j}{\alpha_n + \lambda_j^2} \right)^3 \mathrm{E}\left\{ \left[\pi_Y (Y_1) \psi_j (Y_1) - \mathrm{E}\left(\pi_Y \psi_j \right) \right]^3 \right\} \left| \varphi_j (x) \right|^3 \to 0.$$

This condition requires that α_n not go to zero too fast. It may not be satisfied for all x in the normal case because $\varphi_j (x)$ is not bounded. However, it will be satisfied for $|x| < 1$ when $\alpha_n = dn^{-1/2}$; see equation (B.6) in Appendix B.

PROPOSITION 5. *Under Assumptions 1–5, if $\alpha_n \to 0$ and $n \to \infty$, we have*

$$\frac{\hat{f}(x) - f^{\alpha}(x)}{\sqrt{\mathrm{var}\left(\hat{f}(x) \right)}} \xrightarrow{L} \mathcal{N}(0,1).$$

Note that $\mathrm{var}\left(\hat{f}(x) \right)$ is the first term in the right-hand side of equation (3.1). The following assumption insures that $\mathrm{var}\left(\hat{f}(x) \right)$ can be replaced by the sample variance.

Assumption 6.

$$\frac{1}{n} \sum_j \left(\frac{\lambda_j}{\alpha_n + \lambda_j^2} \right)^4 \mathrm{E}\left\{ \left[\pi_Y (Y_1) \psi_j (Y_1) - \mathrm{E}\left(\pi_Y (Y_1) \psi_j (Y_1) \right) \right]^4 \right\} \left| \varphi_j (x) \right|^4 \to 0.$$

LEMMA 2. *Under Assumptions 1–6, we have*

$$\frac{1}{n} \sum_{i=1}^{n} Z_{ni}(x) - \mathrm{E}\left(Z_{ni}(x) \right) \xrightarrow{P} 0,$$
$$\frac{1}{n} \sum_{i=1}^{n} Z_{ni}^2(x) - \mathrm{E}\left(Z_{ni}^2(x) \right) \xrightarrow{P} 0.$$

The next assumption guarantees that the bias goes to zero sufficiently fast so that f^{α} can be replaced by f.

Assumption 7.

$$\frac{\alpha_n^2 \sum_j \left(\frac{1}{\alpha_n + \lambda_j^2} \right)^2 \left\langle f, \varphi_j \right\rangle^2 \left| \varphi_j (x) \right|^2}{\frac{1}{n} \sum_j \left(\frac{\lambda_j}{\alpha_n + \lambda_j^2} \right)^2 \mathrm{E}\left\{ \left[\pi_Y \psi_j - \mathrm{E}\left(\pi_Y \psi_j \right) \right]^2 \right\} \left| \varphi_j (x) \right|^2} \to 0. \qquad \textbf{(3.7)}$$

If $\varphi_j(x)$ is uniformly bounded (as in the normal case around 0 or in the bounded support case), the numerator of (3.7) is also bounded because

$$\alpha_n^2 \sum_j \left(\frac{1}{\alpha_n + \lambda_j^2} \right)^2 \langle f, \varphi_j \rangle^2 \leq \sum_j \langle f, \varphi_j \rangle^2 = \|f\|^2 < \infty.$$

Hence Assumption 7 holds as soon as the denominator diverges, which is satisfied for $\alpha_n = o\left(n^{-2\beta/2\beta+1}\right)$ under the assumptions of Proposition 4 and for $\alpha_n = o\left(n^{-1}\right)$ for the normal case.

PROPOSITION 6. *Under Assumptions 1–7, if $\alpha_n \to 0$ and $n \to \infty$, we have*

$$\sqrt{n} \frac{\left(\hat{f}(x) - f(x)\right)}{s_n(x)} \xrightarrow{L} \mathcal{N}(0,1),$$

where $s_n^2(x) = \frac{1}{n}\sum_{i=1}^n \left(Z_{ni}(x) - \frac{1}{n}\sum_{i=1}^n Z_{ni}(x) \right)^2$ where $Z_{ni}(x)$ is given by (3.6).

Note that Proposition 6 does not claim that $\hat{f}(x)$ converges at a $\sqrt{n}$-rate of convergence because s_n typically diverges.

3.3. Automatic Selection of the Smoothing Parameter

From Proposition 2, we see that the rate of convergence of α_n depends on β, the regularity of the unknown function f. As β is unknown, the theoretical result of Proposition 2 is not very useful in practice. In this section we propose a data-driven method for selecting α_n that does not require the knowledge of β. Ideally, the penalization term α_n should be selected to minimize the MISE given in (3.1). As the MISE is unknown, it is replaced by an estimator. Denote $\hat{f}^1$ an estimator of f obtained using a nonoptimal α_n (quite small to avoid bias) denoted α_n^1. An estimator of $\langle f, \varphi_j \rangle$ is given by

$$\left\langle \hat{f}^1, \varphi_j \right\rangle = \frac{1}{n} \sum_{i=1}^n \frac{\lambda_j}{\alpha_n^1 + \lambda_j^2} \pi_Y(y_i) \psi_j(y_i).$$

Let $\hat{\lambda}_j$, $\hat{\varphi}_j$, and $\hat{\psi}_j$, $j = 0, 1, \ldots, B$, be the estimators of λ_j, φ_j, and ψ_j, respectively, obtained by the method described in Appendix A. Denote $\widehat{\text{var}}\left[\pi_Y(Y_i)\psi_j(Y_i)\right]$ the sample variance of $\pi_Y(y_i)\hat{\psi}_j(y_i)$. An estimator of the MISE is given by

$$M_n = \frac{1}{n} \sum_{j=0}^B \left(\frac{\hat{\lambda}_j}{\alpha_n + \hat{\lambda}_j^2} \right)^2 \widehat{\text{var}}\left[\pi_Y(Y_i)\psi_j(Y_i)\right]$$
$$+ \alpha_n^2 \sum_{j=0}^B \left(\frac{\hat{\lambda}_j}{\alpha_n^1 + \hat{\lambda}_j^2} \right)^2 \frac{\left\{ \frac{1}{n}\sum_{i=1}^n \pi_Y(y_i)\hat{\psi}_j(y_i) \right\}^2}{\left(\alpha_n + \hat{\lambda}_j^2\right)^2}.$$

This expression can be minimized numerically with respect to α_n to obtain the optimal smoothing parameter.

3.4. Estimation of *T*

Here, we investigate the effect of estimating the operator T by $\tilde{T}$ and T^* by $\tilde{T}^*$ such that

$$\left\|\tilde{T}-T\right\|^2 = O\left(\frac{1}{N}\right) \quad \text{and} \quad \left\|\tilde{T}^*-T^*\right\|^2 = O\left(\frac{1}{N}\right), \tag{3.8}$$

where N has a different interpretation depending on the type of approximation. There are two leading cases where T is approximated. The first one is the case where, although g is known, the spectral decomposition of T cannot be derived analytically, and one has to rely on simulations to compute the eigenvalues and eigenfunctions as described in Appendix A. Then using results on simulation-based estimators (see, e.g., Gourieroux and Monfort, 1996; Carrasco and Florens, 2002), the approximate operator $\tilde{T}$ satisfies (3.8), where $N = B$ is the number of simulations. The second case of interest is the case where the density of ε, g, is unknown and estimated. Although it is standard in the statistics literature to assume that g is known, it may not be very realistic in practice. In some circumstances, an auxiliary data set exists that can be used to estimate parametrically or nonparametrically the function g. In, for instance, Efromovich (1997), Johannes (2009), and Neumann (2007), g is estimated nonparametrically. In our application in Section 6, we postulate a parametric form for g and estimate the unknown parameters using an auxiliary sample. In the case where g is estimated parametrically, N in (3.8) is the size of the auxiliary sample. We compare the performance of our estimator $\hat{f}$ with the estimator $\tilde{f}$ obtained by using $\tilde{T}$ instead of T and $\tilde{T}^*$ instead of T^*:

$$\hat{f} = \left(\alpha_n I + T^*T\right)^{-1}\widehat{T^*h},$$
$$\tilde{f} = \left(\alpha_n I + \tilde{T}^*\tilde{T}\right)^{-1}\widehat{\tilde{T}^*h}.$$

We have the result below.

PROPOSITION 7. *Under Assumptions 1 to 3 and* $4'$, *we have*

$$\left\|\tilde{f}-\hat{f}\right\|^2 \sim \frac{1}{\alpha_n^2}\left\|\tilde{T}-T\right\|^2 = \frac{1}{\alpha_n^2 N}.$$

Let $N = n^{\nu}$ *for some* $\nu > 0$. *For* $\alpha_n = dn^{-1/(\beta\wedge 2+2-\gamma)}$, *we have that* $\left\|\tilde{f}-\hat{f}\right\|^2$ *converges faster than the MISE of* $\hat{f}$, *i.e.,* $n^{-\beta\wedge 2/(\beta\wedge 2+2-\gamma)}$, *provided* $\nu > (\beta\wedge 2+2)/\beta\wedge 2+2-\gamma$. *In this case,* $\tilde{f}$ *has the same asymptotic properties as* $\hat{f}$ *(the same rate of convergence and the same asymptotic distribution).*

It is interesting to note that γ large is beneficial for the rate of convergence of $\hat{f}$ but is somewhat detrimental when f is estimated by $\tilde{f}$. Note that if $N = n$, $\tilde{f}$ is still consistent for an appropriate choice of α_n, namely $\alpha_n = dn^{-1/(\beta \wedge 2+2)}$ for some $d > 0$, but the distribution of $\tilde{f}$ is different from that of $\hat{f}$.

4. CASE WITH ISOLATED ZEROS

4.1. Identification

In the deconvolution literature, it is usually assumed that the CF of ε, Ψ_ε, does not have real zeros. This rules out many well-known densities, as mentioned earlier. In this section we relax this assumption by supposing that Ψ_ε may have (possibly an infinity of) isolated real zeros: $t_1, t_2, \ldots$. For instance, the CF of a distribution with bounded support is analytic and therefore its zeros are necessarily isolated, although they need not be real (Lukacs 1970, Thm. 7.2.3). At the point t_1, we have

$$\Psi_Y(t_1) = \Psi_X(t_1)\Psi_\varepsilon(t_1) = 0. \tag{4.1}$$

Therefore the value of $\Psi_X(t_1)$ cannot be inferred from (4.1). But by the continuity of the CF (Lukacs 1970, Thm. 2.1.2.), $\Psi_X(t_1)$ can be recovered from the knowledge of $\Psi_X(t)$ in a neighborhood of t_1. Therefore there is no identification problem here. However, the presence of zeros has consequences on the way f can be estimated. The estimation of f will be discussed in the next subsections. Here we give results on identification.

Let T be, as before, the operator from $L^2(\pi_X)$ into $L^2(\pi_Y)$ defined by (2.6). We define the null space of T as $\mathcal{N}(T) = \left\{\varphi \in L^2(\pi_X) : T\varphi = 0\right\}$. Recall that T is injective if and only if $\mathcal{N}(T) = \{0\}$.

PROPOSITION 8. *Assume that Assumptions 1 and 2 hold. If* $\Psi_\varepsilon(t) \neq 0$ *for all* t*, then* T *is injective.*

Proposition 8 does not give an "if and only if" statement because, as illustrated below, $\Psi_\varepsilon(t)$ may be equal to 0 for some t while T is injective.

Example 2 continued

Consider $\varepsilon \sim U[-a,a]$. The CF of ε is $\sin(at)/at$ and is equal to zero for $t = j\pi/a$ with $j = \ldots, -2, -1, 1, 2, \ldots$. The eigenvalues of T are $\lambda_j = \sin(aj2\pi/L)/(aj2\pi/L)$ for all $j \in \mathbb{Z}$. Assume $L = 4$. If a is equal to 1, for instance, then $\lambda_j = 0$ for all even j. On the other hand, if a is irrational, then $\lambda_j \neq 0$ for all j in $\mathbb{Z}$ and hence T is injective. This result is exploited in Johnstone and Raimondo (2004).

Note that even if T is not injective on $L^2(\pi_X)$, it may be injective on a smaller space. Define $\mathcal{D}$ the space of the densities,

$$\mathcal{D} = \left\{\varphi \in L^2(\pi_X) : \varphi \geq 0 \text{ and } \int \varphi(x)\,dx = 1\right\}.$$

Now we consider $\tilde{T}$ the operator from $\mathcal{D}$ into $L^2(\pi_Y)$ defined by (2.6). It is the restriction of T on $\mathcal{D}$.

Assumption 3′. Assume Ψ_ε does not vanish on an interval but may have (possibly an infinity of) isolated zeros.

PROPOSITION 9. *Under Assumption 3′, $\tilde{T}$ is injective.*

COROLLARY 2. *Assume Assumptions 1, 2, and 3′ hold. Then, there is only one density of $L^2(\pi_X)$ that is the solution of $Tf = h$. In other words, f is identified.*

4.2. Estimation by Completion

Now we reexamine the estimation procedure of Section 2 to see what the limit of (2.9) is when Assumption 3 is replaced by 3′. When Assumption 3′ holds, the null space of T may not be empty, i.e., 0 may be an eigenvalue of T. If this happens, the solution to $Tf = h$ is not unique but, as mentioned in Section 2.1., the least-squares solution of minimal norm $f^\dagger$ exists and is unique. According to Nashed and Wahba (1974), $f^\dagger$ is the only solution of $\mathcal{N}(T)^\perp$ that satisfies $Tf = h$. Hence, the set of all least-squares solutions may be represented by $f^\dagger + \mathcal{N}(T)$. The estimators given by (2.9) or (2.10) are not consistent estimators of f but of $f^\dagger$. The results of Sections 2 and 3 (consistency, asymptotic normality) remain valid by replacing at the limit f by $f^\dagger$. Now $f^\dagger$ is not necessarily a density because its Fourier transform is not necessarily continuous (its Fourier transform is equal to zero at the points $t_1, t_2, \ldots$). However, the estimation of $f^\dagger$ may give valuable information on the shape of the density f. Moreover, f can be recovered from $f^\dagger$ by using the relationship

$$f = f^\dagger + \sum_{\{j/|\lambda_j|=0\}} \langle f, \varphi_j \rangle \varphi_j.$$

This suggests a way to construct an estimator of f by completing $f^\dagger$. This is illustrated in the example of random variables with bounded support.

Consider Example 2 of Section 2, where the support of Y is known to lie in an interval $[\underline{A}, \bar{A}]$. Assume Ψ_ε is real and is equal to zero at some isolated values t. The operator T has singular value zero associated with the singular functions e^{it_1x}, e^{-it_1x}, e^{it_2x}, e^{-it_2x}, …where the t_l are the zeros of Ψ_ε such that $j = Lt_l/(2\pi) \in \mathbb{Z}$. Indeed, by a change of variables, it is easy to verify that $\int g(y-x)e^{it_1x}dx = e^{it_1y}\int g(u)e^{-t_1u}du = e^{it_1y}\Psi_\varepsilon(t_1) = 0$. Hence the null space of T, $\mathcal{N}(T)$, is the closure of the space spanned by e^{it_1x}, e^{-it_1x}, e^{it_2x}, e^{-it_2x}, …. The eigenfunctions associated with zero are of the form $\varphi_j(x) = e^{it_lx}/\sqrt{L} = e^{ij2\pi x/L}/\sqrt{L}$ where $j = Lt_l/(2\pi) \in \mathbb{Z}$. The density f can be written as the sum of the pseudo-solution $f^\dagger$ and an element of $\mathcal{N}(T)$:

$$f = f^\dagger + \sum_{\{j/|\lambda_j|=0\}} \frac{1}{\sqrt{L}} \overline{\Psi_X\left(\frac{2\pi j}{L}\right)\varphi_j(x)}.$$

The unknown $\Psi_X(t_l)$ can be estimated using the continuity of the characteristic function by

$$\hat{\Psi}_X(t_l) = \frac{\hat{\Psi}_X(t_l - \tau) + \hat{\Psi}_X(t_l + \tau)}{2}, \tag{4.2}$$

where

$$\hat{\Psi}_X(t) = \frac{\hat{\Psi}_Y(t)}{\Psi_\varepsilon(t)} = \frac{\frac{1}{n}\sum_{i=1}^n e^{ity_i}}{\Psi_\varepsilon(t)}, \qquad t \notin \{t_1, t_2, \ldots\}. \tag{4.3}$$

is a $\sqrt{n}$-consistent estimator of $\Psi_X(t)$. Hence when τ goes to zero at an appropriate rate as n goes to infinity, $\widehat{\hat{f}}$ defined by

$$\widehat{\hat{f}}(x) = \hat{f}(x) + \sum_{\{j/|\lambda_j|=0\}} \frac{1}{\sqrt{L}} \overline{\hat{\Psi}_X\left(\frac{2\pi j}{L}\right)\varphi_j(x)}$$

should be a consistent estimator of f. We do not provide a proof of this result, but simulations show that this method works well in practice.

4.3. Estimation under Constraint

In theory, since g is known, the location of the zeros is given. In practice, there may be some densities for which locating the zeros may be problematic, but we do not address this issue here. In this section we propose an alternative method that does not require knowledge of the zeros. Since we know a priori that f is a density, we are going to exploit this information. We consider solving

$$Tf = h \tag{4.4}$$

for $f \in \mathcal{D}$ the subspace of $L^2(\pi_X)$ of density functions. Note that in spite of the linearity of T, problem (4.4) is now nonlinear because of the constraint. As $\mathcal{D}$ is a closed and convex set, the results of Engl, Hanke, and Neubauer (1996, Sect. 5.4) apply. In particular, the solution to (4.4) exists and is unique under Assumption 3′.

We briefly discuss the case where $T : L^2(\pi_X) \to L^2(\pi_Y)$ is injective. A fast way to estimate the constrained solution is to take a two-step approach. First, one determines the regularized solution $\hat{f}(x)$ of the unconstrained problem using (2.10). Second, one computes the metric projection of $\hat{f}(x)$ onto the set $\mathcal{D}$. Since $\mathcal{D}$ is closed and convex, the results on convergence and convergence rates of Section 2 remain valid for the constrained case.

Now, we turn to the important case where T is not injective. The two-step approach does not work any longer. Following Engl, et al. (1996), we propose to solve the following constrained optimization problem:

$$\min_{f\in\mathcal{D}}\left\{\left\|Tf-\hat{h}\right\|_{\pi_Y}^2+\alpha\|f\|_{\pi_X}^2\right\}, \tag{4.5}$$

where $\hat{h}$ is a nonparametric estimator of h, obtained for instance by kernel. Let us denote $\hat{f}_{\mathcal{D}}$ this solution.

PROPOSITION 10. *Assume that the estimator $\hat{h}$ satisfies $\left\|\hat{h}-h\right\|=O(\delta)$ and Assumptions 1,2, 3′, 4 and condition (3.2) hold. Let $\alpha=\delta^{2/(\beta\wedge 2+1)}$, then*

$$\left\|\hat{f}_{\mathcal{D}}-f\right\|_{\pi_X}^2=O\left(\delta^{2(\beta\wedge 2)/(\beta\wedge 2+1)}\right).$$

If $\hat{h}$ is the kernel estimator of a twice continuously differentiable density f, then $\delta=n^{-2/5}$ and $\left\|\hat{f}_{\mathcal{D}}-f\right\|_{\pi_X}^2=O\left(n^{-4/5(\beta\wedge 2)/(\beta\wedge 2+1)}\right)$. This can be compared with the MISE in the unconstrained case $MISE=O\left(n^{-(\beta\wedge 2)/(\beta\wedge 2+2)}\right)$. It turns out that the rate of convergence in the constrained case is faster than in the unconstrained case for all β. Note that the solution to problem (4.5) does not have a closed-form expression but can be computed numerically. In practice, the space $\mathcal{D}$ is replaced by a finite dimensional space $\mathcal{D}_n$, which can be a grid or a large dimensional sieve space. Some theoretical results on the effect of such an approximation can be found in Neubauer (1987). Some practical issues are discussed in Chernozhukov, Gagliardini, and Scaillet (2008, Sect. 6.1).

4.4. Comparison with Alternative Estimators

In this section we discuss some alternative estimators. First, we provide some intuition. Let us denote Ψ_Y, Ψ_X, and Ψ_ε the characteristic functions of Y, X, and ε, respectively. We have the relation $\Psi_Y=\Psi_X\Psi_\varepsilon$. Using the Fourier inversion formula, the density f of X satisfies

$$f(x)=\frac{1}{2\pi}\int e^{-itx}\frac{\Psi_Y(t)}{\Psi_\varepsilon(t)}dt. \tag{4.6}$$

This is the starting point of Hall and Meister (2007). First, they multiply the numerator and denominator in (4.6) by $\Psi_\varepsilon(-t)$ to obtain a real-valued, nonnegative function in the denominator. Then, they define the following estimator:

$$\hat{f}(x)=\mathrm{Re}\left\{\frac{1}{2\pi}\int e^{-itx}\frac{\Psi_\varepsilon(-t)\,|\Psi_\varepsilon(t)|^r\,\hat{\Psi}_Y(t)}{\{\max(|\Psi_\varepsilon(t)|,\alpha(t))\}^{r+2}}dt\right\}, \tag{4.7}$$

where $\alpha(t)$ is a smoothing parameter that depends on n and t. For the integral to be well defined, $|\Psi_\varepsilon(t)|^{r+1}$ needs to be integrable. If g is square integrable, it suffices to take $r \geq 1$. Now (4.7) involves a regularization of the convolution operator without making it compact first. The space of references is L^2 with respect to Lebesgue measure. In this case, the convolution operator has a continuous spectrum and as discussed in Carrasco et al. (2007, Sect. 5.4.2), $\Psi_\varepsilon(t)$ can be interpreted as the singular values of the convolution operator. Hall and Meister's regularization is similar to spectral cut-off. Their estimator has the same advantages as ours: It does not involve a kernel, and it applies even when $\Psi_\varepsilon(t)$ has isolated zeros. Moreover, it is shown to be consistent in the latter case, whereas ours needs to be modified. Why is their estimator consistent? It is because the spectrum of their convolution operator is continuous, and since $\Psi_\varepsilon(t)$ has only countably many zeros, the set of zeros has Lebesgue measure zero and their operator is still injective. In this framework, the optimal smoothing parameter $\alpha(t)$ is a function of n and t, with a rate of convergence that depends on the properties of $\Psi_\varepsilon(t)$ and the smoothness properties of f. Hall and Meister restrict their attention to a class of characteristic functions for which there exist $\mu \geq 1$, $\nu > 0$, $0 < C_1 < C_2 < \infty$, $\lambda > 0$, and $T > 0$ such that

$$C_1 |\sin(\lambda t)|^\mu |t|^{-\nu} \leq |\Psi_\varepsilon(t)| \leq C_2 |\sin(\lambda t)|^\mu |t|^{-\nu} \quad \text{for all } |t| > T, \tag{4.8}$$

and $\Psi_\varepsilon(t)$ does not vanish for $|t| \leq T$. This class includes all self-convolved uniform densities and their convolution with any ordinary-smooth density. They also consider a class corresponding to the convolution of uniform densities with any supersmooth density. Condition (4.8) and the other condition not reported here imply that the zeros occur at the points $t = j\pi/\lambda$, $j = 1, 2, \ldots$. It rules out all characteristic functions for which the zeros do not follow this pattern. For instance, the Epanechnikov distribution, which has density $g(\varepsilon) = \frac{3}{4}\left(1 - \varepsilon^2\right) I\{|\varepsilon| < 1\}$ and characteristic function $\Psi_\varepsilon(t) = \frac{3}{t^3}(\sin(t) - t\cos(t))$, is not part of this class. Note that we need not impose a restriction of the type (4.8) and therefore can cover a larger class of functions.

Let us mention two other recent papers. Meister (2007) is concerned with deconvolution when the density to be estimated has compact support and the characteristic function of the error may vanish. Finally, Meister (2008) focuses on the deconvolution with errors satisfying (4.8); he proposes to approximate the characteristic function $\Psi_X(t)$ by an expansion using Legendre polynomials for the t corresponding to zeros of Ψ_ε. This is an alternative approach to that proposed in Section 4.2.

5. SIMULATION STUDY

We conducted a Monte Carlo study to determine the performance of $\hat{f}$ in two cases corresponding to Examples 1 and 2:

Normal error. We consider the case where $\varepsilon \sim \mathcal{N}\left(0, \frac{1}{5}\right)$ $\left(\sigma^2 = 1/5\right)$ and X is a mixture of normal: $\mathcal{N}\left(\sqrt{\frac{2}{3}}, \frac{1}{3}\right)$ with probability 1/2 and $\mathcal{N}\left(-\sqrt{\frac{2}{3}}, \frac{1}{3}\right)$ with probability 1/2.

We choose σ_Y^2 so that $\rho = \sigma_Y^2 / \left(\sigma^2 + \sigma_Y^2\right)$ is large. We set $\sigma_Y^2 = 9$ so that $\sigma_X^2 = 46/5$ and $\rho = 45/46$. Using (2.11), we compute recursively the φ_j and ψ_j.

Uniform error. Now, we consider the case where ε is uniformly distributed on $[-1, 1]$ and X follows a triangular on [-1,1] so that $f(x) = (1 - |x|) I\{|x| < 1\}$ and $\phi_X(t) = 2(1 - \cos(t))/t^2$. Using the notation of Example 2, we set $\left[\underline{A}, \overline{A}\right] = [-2, 2]$ so that $L = 4$. The eigenfunctions and eigenvalues are $\varphi_j(x) = \frac{1}{2} e^{ij\pi x/2}$, $\lambda_j = \frac{\sin(\pi j/2)}{\pi j/2}$, $j \in \mathbf{Z}$. We see that λ_j equals zero for all even-valued j. The estimator $\hat{f}$ of $f^\dagger$ is given by (2.12). The estimator of f is

$$\widehat{\hat{f}}(x) = \hat{f}(x) + 2\,\mathrm{Re}\left\{\sum_{j=2,4,\ldots} \frac{1}{\sqrt{L}} \overline{\hat{\phi}_X\left(\frac{2\pi j}{L}\right)} \varphi_j(x)\right\},$$

where $\hat{\phi}_X$ is estimated using (4.2) and (4.3) and $\tau = 0.1$.

5.1. Simulation Design

The sample size is set at $n = 10{,}000$. To give an idea of the variance of our estimator, we run 100 replications and rank them in order of the size of the MISE. In Figure 1 we report the true density, the average of the 100 estimated densities, and the estimated densities corresponding to the 25th, 50th, and 75th percentiles of the MISE. The estimations are performed using the automatic bandwidth selection. We use the constrained optimization package in GAUSS “co” to get the optimal bandwidth.

1. Normal error

In equation (2.10) we truncate the sum in j to $J = 199$. We use as starting value for α, 0.001, and $\alpha_n^1 = 0.001$ to compute the MISE in order to select the optimal bandwidth. The automatic bandwidth selection gives a median α_n equal to 0.0054.

2. Uniform error

For calculating $\hat{f}(x)$, we take $j = 0, 1, 3, \ldots, 201$. For calculating the second part of $\widehat{\hat{f}}(x)$, we take $j = 2, 4, \ldots, 10$. The regularization parameter is selected in order to minimize the MISE of $\hat{f}$. Again, we take the starting value for the optimization and for the MISE, namely $\alpha = 0.02$. The median of the bandwidths selected using the automatic selection is $\alpha_n = 0.0297$.

5.2. Simulation Results

From Figure 1, we see that (i) the automatic bandwidth works well, (ii) $\widehat{\hat{f}}$ is (on average) very close to the true density even with an adhoc truncation of the series (here 10), and (iii) our estimator hasmore difficulty with the uniform error than with the normal error. This is not surprising given that (a) the error (a uniform on [-1,1]) is large relative to the signal (a triangular on [-1,1]), and (b) some of the eigenvalues are zeros and the completion of the series by projection makes the estimator more oscillatory.

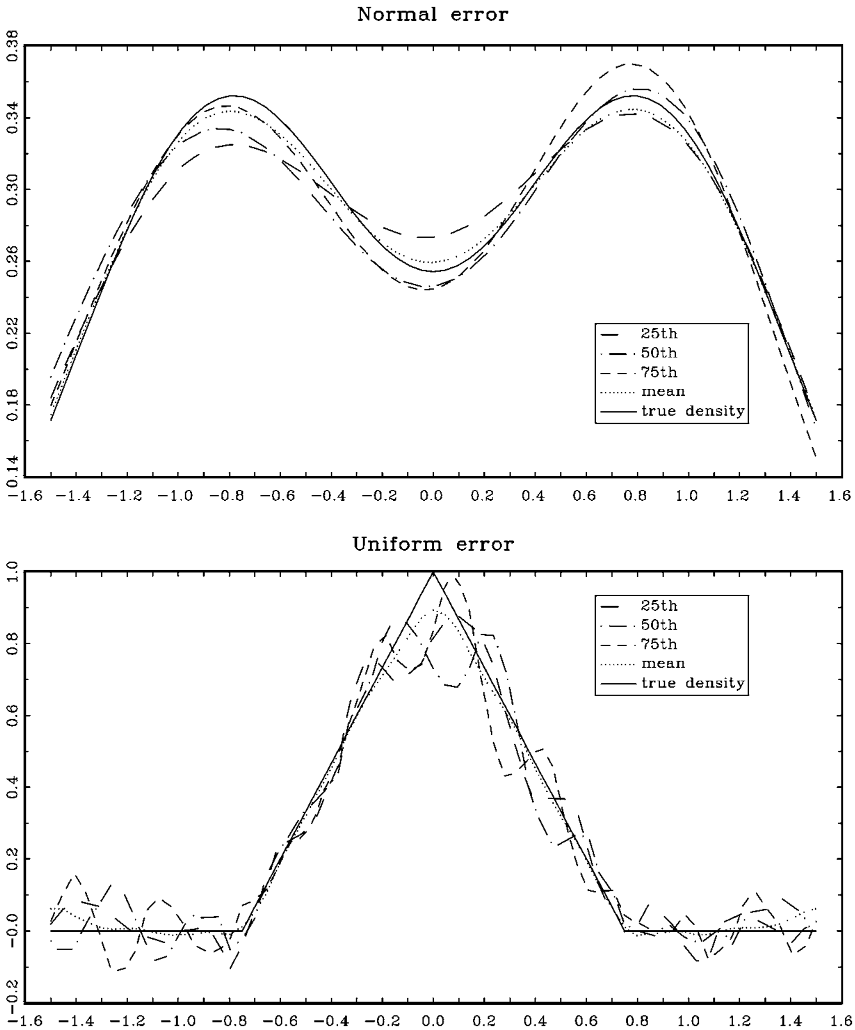

FIGURE 1. Simulations with 100 replications and $n = 10,000$. Automated bandwidth.

6. APPLICATION TO MEASUREMENT ERROR IN WAGE

The Current Population Survey (CPS) is a monthly survey of about 50,000 households by the U.S. Bureau of the Census. The CPS is publicly available[5] and provides detailed information on the labor force characteristics of the U.S. population. For these reasons, the CPS is widely used by economists. However, as the data are collected by interview from households, they are bound to be misreported. Our aim is to quantify the measurement error in the hourly earnings. Let Y be the reported hourly earnings; then Y is the sum of the true earnings; Y^*, and an error Z:

$$Y = Y^* + Z. \tag{6.1}$$

The density g_{Y^*} of Y^* is unknown, but we use data from the National Compensation Survey[6] (NCS) collected by the Bureau of Labor Statistics to select a parametric specification for g_{Y^*}. Earnings data of NCS are based on payroll data collected directly from the establishments and therefore can be considered as accurate data. The NCS does not provide individual data but reports the mean and the 10th, 25th, 50th, 75th, and 90th percentiles of the hourly earnings by occupations and regions. We focus on all occupations that enter in the category "Blue Collar," as we believe that it is a large but relatively homogeneous population. We use the information relative to the data collected between December 2001 and January 2003 on all of the United States. As we need a parametric specification of the true distribution, we assume that the earnings have a gamma $\mathcal{G}(\alpha, \beta, l)$ distribution. The density is given by

$$g_{Y^*}(x) = \frac{(x-l)^{\alpha-1} \exp(-(x-l)/\beta)}{\beta^\alpha \Gamma(\alpha)}, x > l.$$

The lower bound, $l = 5$, has been selected to be just below the federal minimum wage in 2002 ($5.15). We estimate (α, β) by the generalized method of moments, which consists in minimizing the Euclidean norm between the empirical quantiles of the distribution function and their theoretical values. We obtain $\alpha = 2.052$, $\beta = 4.699$. To verify that the gamma matches the true distribution, we report in Table 1 the percentiles found in the NCS publication and those of the gamma. The gamma is not a perfect match but is close enough for illustration purposes.

From the CPS, we extracted a sample of 9,335 individuals corresponding to the same occupations as for the NCS. The data are for January and September 2002 (this guarantees that the same household is not represented twice because the CPS uses an eight-month rotating survey). The hourly earnings range from $5 to $54. The percentiles of the CPS data are reported in the last row of Table 1. We see that people tend to underestimate their wages by 15% to 23%.

In (6.1) Y^* and Z are likely to be correlated, as people tend to underreport their income by more dollars when their hourly rate is higher in absolute term. This is clear from a comparison of the percentiles in Table 1. On the other hand, the

TABLE 1. Comparisons between hourly earnings distributions

		Percentiles				
	Mean	10	25	50	75	90
True earnings (source: NCS)	14.51	7.65	9.75	13.03	18	23.86
Gamma distribution	14.64	7.63	9.70	13.13	17.96	23.64
Reported earnings (source: CPS)	11.91	6.4	7.9	10.03	14.51	20.03

ratio Y/Y^* is likely to be independent of Y^*. We estimate the density of $V \equiv \ln(Y/Y^*)$, which is the solution of

$$\ln(Y) = \ln\left(Y^*\right) + V.$$

It is reasonable to assume that V is independent of $\ln(Y^*)$. As Y/Y^* is expected to be close to one, V is a good approximation for $Y/Y^* - 1 = (Y - Y^*)/Y^*$, which is the ratio of the measurement error over the true hourly rate. So here, $\ln(Y^*)$ plays the role of ε and V plays the role of X in the rest of the paper. We apply our method as if the distribution of $\ln(Y^*)$ were known, while in reality its parameters α and β have been estimated using an auxiliary sample. In Proposition 7 we established that provided that the size of the auxiliary sample is large enough, this approximation does not affect the asymptotic properties of our estimator.

The characteristic function of $\ln(Y^*)$ is not known in closed form. To determine whether it has zeros, we calculate it via simulations using a sample of 10,000 simulated data and find that it does not have zeros. To estimate the density, we apply the method described in Section 3 by setting π_Y and ω equal to the densities of a standard normal distribution and π_X equal to the marginal π_X^m defined in (2.5); the integral in (2.5) is computed by numerical integration. The eigenvalues and eigenfunctions are computed using 3,000 simulations, i.e., $B = B' = 3,000$. As the eigenvalues decline rapidly, we truncate the sum in j to $J = 23$. Figure 2 gives the plot of the estimated density of V for $\alpha_n = 0.05$. We see that the density is skewed to the left, suggesting that, as expected, people are likely to underreport their wages.

7. CONCLUSION

In this paper we approximate the function to be estimated by a sequence of orthonormal functions obtained from the singular value decomposition of the convolution operator. When the CF of the error does not vanish, we show that this estimator $\hat{f}$ is consistent and asymptotically normal. When the CF has isolated zeros, we find that $\hat{f}$ does not converge to the true density f, but to a pseudo-solution $f^\dagger$, which is the projection of f on the orthogonal complement to the null space of the convolution operator. It is, however, possible to recover the density by adding terms to $\hat{f}$. Finally, we also propose to estimate f by minimizing

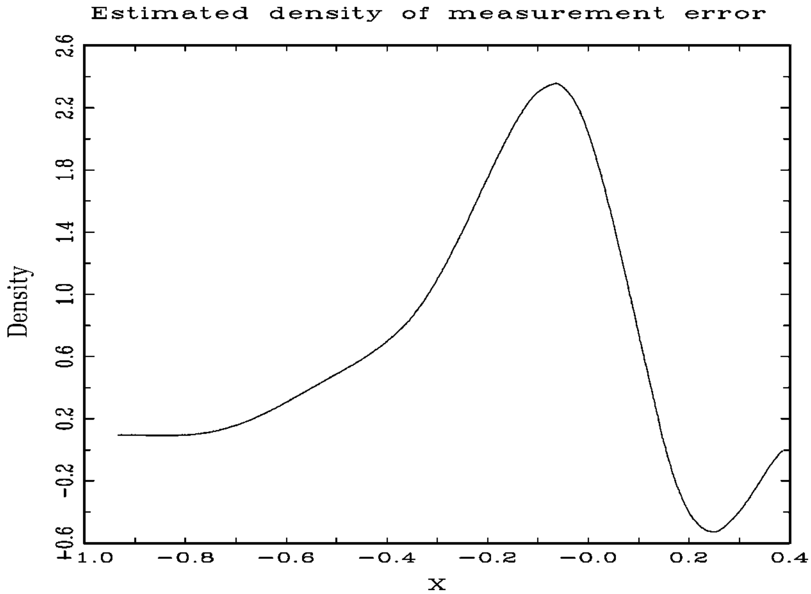

FIGURE 2. Density of measurement error in hourly rate

a penalized least-squares criterion under the constraint that f is a density. This second estimator is consistent and achieves a faster rate of convergence than the unconstrained estimator.

We restricted our analysis to the case where X and ε are univariate. However, some interesting applications involve multivariate variables; see, e.g., Gautier and Kitamura (2008). Our technique can be generalized to the multivariate setting in a straightforward manner. Indeed, the only difference is that the spaces $L^2_{\pi_X}$ and $L^2_{\pi_Y}$ will be defined for functions in $\mathbb{R}^p$ instead of $\mathbb{R}$. The computation of the eigenvalues and eigenfunctions described in Appendix A remains the same where x_c and y_b are drawn from a multivariate distribution. The dimension of the matrix M remains $B \times B$ regardless of the dimension of X and ε, so that the computational burden is not increased.

NOTES

1. An element $f \in \mathcal{H}$ is said to be a least-squares solution to (1.2) if $\inf\{\|T\varphi - h\| : \varphi \in \mathcal{H}\} = \|Tf - h\|$.

2. Here $\mathcal{R}(T)$ denotes the range of T. The assumption $h \in \mathcal{R}(T) + \mathcal{R}(T)^{\perp}$ is necessarily satisfied because the model is assumed to be correctly specified throughout the paper.

3. The φ_j are the standard Hermite polynomials (see the definition given in Wand and Jones, 1995, App. C), where x has been replaced by x/σ_X. To see this, we need to use the relation $\sigma_X^2 = \sigma^2/(1-\rho)$ and do a change of variable $z = x/\sigma_X$.

4. In the case of a uniform error, the operator T is in general not injective. Hence, as explained in Section 4, the estimator does not converge to the true density but to its projection on the orthogonal

of the null space of T. The rate of convergence given in Proposition 4 is then the rate toward this projection.

5. See http://www.bls.census.gov/cps/cpsmain.htm.
6. Data sets and descriptions are available at http://www.bls.gov/ncs/.

REFERENCES

Billingsley, P. (1995) *Probability and Measure*. Wiley.

Blundell, R., X. Chen, & D. Kristensen (2007) Semi-nonparametric IV estimation of shape-invariant Engel curves. *Econometrica* 75, 1613–1669.

Carrasco, M. & J.P. Florens (2002) Simulation based method of moments and efficiency. *Journal of Business & Economic Statistics* 20(4), 482–492.

Carrasco, M., J.P. Florens, & E. Renault (2007) Linear inverse problems in structural econometrics: Estimation based on spectral decomposition and regularization. In J.J. Heckman & E.E. Leamer (eds.), *The Handbook of Econometrics*, vol. 6, pp. 5633–5751. Elsevier.

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

Carroll, R. & P. Hall (2004) Low order approximations in deconvolution and regression with errors in variables. *Journal of the Royal Statistical Society Series B* 66, 31–46.

Carroll, R., A. Van Rooij, & F. Ruymgaart (1991) Theoretical aspects of iii-posed problems in statistics. *Acta Applicandae Mathematicae* 24, 113–140.

Chernozhukov V., P. Gagliardini, & O. Scaillet (2008) Nonparametric Instrumental Variable, Estimation of Quantile Structural Effects. Working paper, Swiss Finance Institute.

Devroye, L. (1986) *Non-Uniform Random Variate Generation*. Springer Verlag.

Devroye, L. (1989) Consistent deconvolution in density estimation. *Canadian Journal of Statistics* 17, 235–239.

Donoho, D. (1995) Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Applied and Computational Harmonic Analysis* 2, 101–126.

Dunford, N. & J. Schwartz (1963) *Linear Operators, Part II, Spectral Theory*. Wiley.

Efromovich (1997) Density estimation for the case of supersmooth measurement error. *Journal of the American Statistical Association* 92(438), 526–535.

Engl. H.W., M. Hanke, & A. Neubauer (1996) *Regularization of Inverse Problems*. Kluwer Academic.

Fan, J. (1991a) On the optimal rates of convergence for nonparametric deconvolution problems. *Annals of Statistics* 19, 1257–1272.

Fan, J. (1991b) Asymptotic normality for deconvolution kernel density estimators. *Sankhya* 53, 97–110.

Fan, J. (1993) Adaptively local one-dimentional subproblems with application to a deconvolution problem. *Annals of Statistics* 21, 600–610.

Gautier, E. & Y. Kitamura (2008) Nonparametric Estimation in Random Coefficients Binary Choice Models. Manuscript, Yale University.

Geweke, J. (1988) Antithetic acceleration of Monte Carlo integration in Bayesian inference. *Journal of Econometrics* 38, 73–90.

Gourieroux, C. & A. Monfort (1996) *Simulation-Based Econometric Methods*. Oxford University Press.

Groeneboom, P. & G. Jongbloed (2003) Density estimation in the uniform deconvolution model. *Statistica Neerlandica* 57, 136–157.

Hall, P. & J. Horowitz (2005) Nonparametric methods for inference in the presence of instrumental variables. *Annals of Statistics* 33, 2904–2929.

Hall, P. & A. Meister (2007) A ridge-parameter approach to deconvolution. *Annals of Statistics* 35, 1535–1558.

Hall, P., F. Ruymgaart, O. van Gaans, & A. van Rooij (2001) Inverting noisy integral equations using wavelet expansions: A class of irregular convolutions. In M.C.M. de Gunst, C.A.J. Klassen, & A.W. van der Vaart (eds.), *Festschrift in Honor of Willem van Zwet*. IMS Lectures Notes Monograph Series, vol. 36, pp. 533–546. Institute of Mathematical Statistics.

Horowitz, J. & M. Markatou (1996) Semiparametric estimation of regression models for panel data. *Review of Economic Studies* 63, 145–168.

Hu, Y. & G. Ridder (2007) Estimation of Nonlinear Models with Mismeasured Regressors Using Marginal Information. Mimeo, University of Southern California.

Ichimura, H. & T.S. Thompson (1998) Maximum likelihood estimation of a binary choice model with random coefficients of unknown distribution. *Journal of Econometrics* 86, 269–295.

Johannes, J. (2009) Deconvolution with unknown error distribution. *Annals of Statistics* 39, 2301–2323.

Johnstone, I., G. Kerkyacharian, D. Picard, & M. Raimondo (2004) Wavelet decomposition in a periodic setting. *Journal of the Royal Statistical Society Series B*, 66, 547–573.

Johnstone, I. & M. Raimondo (2004) Periodic boxcar deconvolution and diophantine approximation. *Annals of Statistics* 32, 1781–1804.

Kress, R. (1999) *Linear Integral Equations.* Springer.

Lukacs, E. (1970) *Characteristic Functions*. Statistical Monographs and Courses. Griffin.

Meister, A. (2007) Deconvolving compactly supported densities. *Mathematical Methods of Statistics* 16, 63–76.

Meister, A. (2008) Deconvolving from Fourier-oscillating error densities under decay and smoothness restrictions. *Inverse Problems* 24, 1–14.

Nashed, M.Z. & G. Wahba (1974) Generalized inverses in reproducing kernel spaces: An approach to regularization of linear operator equations. *SIAM Journal on Mathematical Analysis* 5, 974–987.

Neubauer, A. (1987) Finite-dimensional approaximation of constrained Tikhonov-regularized solutions of iii-posed linear operator equations. *Mathematics of Computation* 48, 565–583.

Neumann, M. (2007) Deconvolution from panel data with unknown error distribution. *Journal of Multivariate Analysis* 98, 1955–1968.

Pensky, M. & B. Vidakovic (1999) Adaptive wavelet estimator for nonparametric dendity deconvolution. *Annals of Statistics* 27, 2033–2053.

Postel-Vinay, F. & J.-M. Robin (2002) Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica* 70, 2295–2350.

Serfling, R. (1980) *Approximation Theorems of Mathematical Statistics*. Wiley.

Stefanski, L. & R. Carroll (1990) Deconvoluting kernel density estimators. *Statistics* 2, 169–184.

Tikhonov, A. & V. Arsenin (1977) *Solutions of Ill-posed Problems*. Winston & Sons.

Ushakov, N. (1999) *Selected Topics in Characteristic Functions*. VSP.

Van Rooij, A.C.M. & F.H. Ruymgaart (1991) Regularized deconvolution on the circle and the sphere. In G. Roussas (ed.), *Nonparametric Functional Estimation and Related Topics*, pp. 679–690. Kluwer Academic.

Van Rooij, A.C.M. & F.H. Ruymgaart (1999) On inverse estimation. In S. Ghosh (ed.), *Asymptotics, Nonparametrics and Time Series*, pp. 579–613. Dekker.

Walter, G. (1981) Orthogonal series estimators of the prior distribution. *Sankhia: The Indian Journal of Statistics Series A* 43, 228–245.

Wand, M.P. & M.C. Jones (1995) *Kernel Smoothing*, Chapman and Hall.

APPENDIX A: Implementation

In this Appendix we discuss the practical aspects of the estimation of f when no explicit expression of the eigenvalues and eigenfunctions is available. First, we explain how to estimate the eigenvalues and eigenfunctions. Second, we give the estimate of f.

Calculation of Eigenvalues and Eigenfunctions. We are looking for the solutions of

$$T^*T\varphi = \lambda^2\varphi. \tag{A.1}$$

If T and T^* are conditional expectation operators, they can be estimated, by kernel estimators, but there is a simpler way that applies in all cases.

(a) To estimate the operator T, we will use importance sampling (Geweke, 1988). Denote ω a pdf, such that it is easy to draw data from the distribution corresponding to ω either by inversion of the cdf or by a rejection method (see Devroye, 1986). The operator $(T\varphi)(y) = \int \varphi(x)g(y-x)\,dx = \int \frac{\varphi(x)g(y-x)}{\omega(x)}\omega(x)\,dx$ can be estimated by $\frac{1}{B'}\sum_{c=1}^{B'} \frac{\varphi(x_c)g(y-x_c)}{\omega(x_c)}$, where (x_c), $c = 1, \dots, B'$ is an i.i.d. sample drawn from ω.

(b) The operator $\left(T^*\psi\right)(x) = \int \psi(y)\pi_{Y|X}(y|x)\,dy = \frac{\int \psi(y)\pi_Y(y)g(y-x)dy}{\pi_X(x)}$ can be estimated by $\frac{1}{\pi_X(x)}\frac{1}{B}\sum_{b=1}^{B}\psi(y_b)\,g(y_b-x)$, where (y_b), $b = 1, \dots, B$ is an i.i.d. sample drawn from π_Y. This way we obtain estimators of T and T^* that are $\sqrt{B'}$ and $\sqrt{B}$ consistent and do not require a choice of a kernel and a bandwidth.

Therefore, $\left(T^*T\varphi\right)(x)$ can be approximated by

$$\frac{1}{\pi_X(x)}\frac{1}{B}\sum_{b=1}^{B}\left[\frac{1}{B'}\sum_{c=1}^{B'}\frac{\varphi(x_c)g(y_b-x_c)}{\omega(x_c)}\right]g(y_b-x).$$

This operator has a finite rank and has at most B eigenvalues. Note that the eigenfunctions are necessarily of the form

$$\varphi_j(x) = \sum_{b=1}^{B}\beta_b^j\frac{g(y_b-x)}{\pi_X(x)}. \tag{A.2}$$

Replacing φ_j by its expression, we see that solving (A.1) is equivalent to finding the eigenvalues and eigenvectors of the $B\times B-$matrix M with principal element

$$M_{b,l} = \frac{1}{BB'}\sum_{c=1}^{B'}\frac{g(y_l-x_c)g(y_b-x_c)}{\pi_X(x_c)\omega(x_c)}.$$

Let $\underline{\beta}^j = \left[\beta_1^j, \cdots, \beta_B^j\right]'$ be the jth eigenvector of M associated with λ_j^2; then the φ_j solution of (A.2) is the jth eigenfunction of T^*T associated with the same eigenvalue λ_j^2. The function φ_j can be evaluated at all points. Note that the φ_j associated with distinct eigenvalues are necessarily orthogonal, nevertheless, they need to be normalized. To normalize them, one can approximate the norm in the following way: $\|\varphi\|^2 = \int\varphi^2(x)\pi_X(x)\,dx = \int\varphi^2(x)\frac{\pi_X(x)}{\omega(x)}\omega(x)\,dx \simeq \frac{1}{B'}\sum_{c=1}^{B'}\varphi^2(x_c)\frac{\pi_X(x_c)}{\omega(x_c)}$. Denote $\hat{\varphi}_j$ and $\hat{\lambda}_j^2$ the estimators of the normalized φ_j and λ_j^2.

The operator $TT^*\psi(y)$ can be approximated by

$$\frac{1}{B^2}\sum_{b=1}^{B}\left[\sum_{c=1}^{B}\frac{\psi(x_c)g(y_c-x_b)g(y-x_b)}{\omega(x_b)\pi_X(x_b)}\right]$$

$$\equiv \frac{1}{B^2}\sum_{c=1}^{B}\varpi(y,y_c)\,\psi(x_c).$$

It is easy to verify that the eigenfunctions ψ_j are of the form $\sum_{c=1}^{B} \beta_c^j \varpi(y, y_c)$, where $\underline{\beta}^j = \left[\beta_1^j, \cdots, \beta_B^j\right]'$, $j = 1, \ldots, n$, are again the eigenvectors of M defined above. Hence the estimators of ψ_j are given by

$$\hat{\psi}_j(y) = \sum_{b=1}^{B} \beta_b^j \left[\sum_{l=1}^{B} \frac{g(y_b - x_l) g(y - x_l)}{\omega(x_l) \pi_X(x_l)}\right].$$

Calculation of $\hat{f}$. In formula (2.9) we need to compute the term $\left\langle \pi_{Y|X}(y_i|.), \varphi_j(.)\right\rangle = \int \pi_{Y|X}(y_i|x)\varphi_j(x)\pi_X(x)dx$. It can be approximated by $\widehat{\left\langle \pi_{Y|X}(y_i|.), \varphi_j(.)\right\rangle} = \frac{1}{B} \sum_{b=1}^{B} \frac{\pi_Y(y_i) g(y_i - x_b)}{\omega(x_b)} \hat{\varphi}_j(x_b)$, where (x_b), $b = 1, \ldots, B$ is an i.i.d. sample drawn from ω. Hence we obtain $\hat{f}$:

$$\hat{f}(x) = \sum_{j=1}^{B} \frac{1}{\alpha_n + \hat{\lambda}_j^2} \frac{1}{n} \sum_{i=1}^{n} \widehat{\left\langle \pi_{Y|X}(y_i|.), \varphi_j(.)\right\rangle} \hat{\varphi}_j(x).$$

APPENDIX B: Proofs

Proof of Proposition 1. We examine successively the terms of variance and bias.

Variance. Using the expression of $\hat{f}$ given in (2.9), we have

$$\mathrm{E}\left[\left(\hat{f}(x) - f^{\alpha_n}(x)\right)^2\right] = \frac{1}{n} \mathrm{var}\left[\sum_{j=0}^{\infty} \frac{1}{\alpha_n + \lambda_j^2} \left\langle \pi_{Y|X}(Y_i|.), \varphi_j(.)\right\rangle \varphi_j(x)\right].$$

Because the eigenfunctions φ_j are orthonormal with respect to π_X, we have

$$\int \mathrm{E}\left(\hat{f}(x) - f^{\alpha_n}(x)\right)^2 \pi_X(x)\, dx = \frac{1}{n} \sum_{j=0}^{\infty} \left(\frac{1}{\alpha_n + \lambda_j^2}\right)^2 \sigma_j^2,$$

with

$$\begin{aligned}
\sigma_j^2 &= \mathrm{var}\left[\left\langle \pi_{Y|X}(Y_i|.), \varphi_j(.)\right\rangle\right] \\
&= \mathrm{var}\left[\int \pi_{Y|X}(Y_i|x)\, \varphi_j(x) \pi_X(x)\, dx\right] \\
&= \mathrm{var}\left[\pi_Y(Y_i) \int \pi_{X|Y}(x|Y_i)\, \varphi_j(x)\, dx\right] \\
&= \mathrm{var}\left[\pi_Y(Y_i) \lambda_j \psi_j(Y_i)\right] \\
&= \lambda_j^2 \mathrm{var}\left[\pi_Y(Y_i) \psi_j(Y_i)\right]. \qquad \textbf{(B.1)}
\end{aligned}$$

so that the variance term is

$$\mathrm{Var} = \frac{1}{n} \sum_{j=0}^{\infty} \left(\frac{\lambda_j}{\alpha_n + \lambda_j^2}\right)^2 \mathrm{var}\left[\pi_Y(Y_i) \psi_j(Y_i)\right].$$

Bias. Using (2.7), f^{α_n} can be rewritten as

$$f^{\alpha_n} = \sum_{j=0}^{\infty} \frac{1}{\alpha_n + \lambda_j^2} \langle h, T\varphi_j \rangle \varphi_j = \sum_{j=0}^{\infty} \frac{\lambda_j}{\alpha_n + \lambda_j^2} \langle h, \psi_j \rangle \varphi_j = \sum_{j=0}^{\infty} \frac{\lambda_j^2}{\alpha_n + \lambda_j^2} \langle f, \varphi_j \rangle \varphi_j$$

because $h = Tf$. We have

$$\begin{aligned} f - f^{\alpha_n} &= \left(I - \left(\alpha_n I + T^* T \right)^{-1} T^* T \right) f \\ &= \alpha_n \left(\alpha_n I + T^* T \right)^{-1} f \\ &= \alpha_n \sum_{j=0}^{\infty} \frac{1}{\alpha_n + \lambda_j^2} \langle f, \varphi_j \rangle \varphi_j . \end{aligned}$$

It follows that

$$\left\| f - f^{\alpha_n} \right\|^2 = \alpha_n^2 \sum_{j=0}^{\infty} \frac{\langle f, \varphi_j \rangle^2}{\left(\alpha_n + \lambda_j^2 \right)^2} . \qquad \blacksquare$$

Proof of Proposition 2. Using $\alpha_n + \lambda_j^2 \geq \alpha_n$, the term of variance can be majored by

$$\begin{aligned} \text{Var} &= \frac{1}{n} \sum_{j=0}^{\infty} \left(\frac{\lambda_j}{\alpha_n + \lambda_j^2} \right)^2 \text{var}\left[\pi_Y(Y_i) \psi_j(Y_i) \right] \qquad \textbf{(B.2)} \\ &\leq \frac{1}{n\alpha_n^2} \sum_{j=0}^{\infty} \lambda_j^2 \text{var}\left[\pi_Y(Y_i) \psi_j(Y_i) \right] . \end{aligned}$$

Then by Assumption 4 and the fact that T is a Hilbert-Schmidt operator, we have

$$\text{Var} \leq \frac{C}{n\alpha_n^2} \sum_{j=0}^{\infty} \lambda_j^2 = O\left(\frac{1}{n\alpha_n^2} \right).$$

Assuming condition (3.2), it follows from Carrasco et al. (2007, Prop. 3.12) that $\left\| f - f^{\alpha_n} \right\|^2 = O\left(\alpha_n^{\beta \wedge 2} \right)$. Hence, we obtain a majoration of the MISE, $MISE \leq \frac{A}{n\alpha_n^2} + B\alpha_n^{\beta \wedge 2}$. For α_n of order $1/n^{(\beta \wedge 2+2)}$, we have $MISE \leq Cn^{-(\beta \wedge 2)/(\beta \wedge 2+2)}$. $\blacksquare$

Proof of Lemma 1. Condition (3.2) for $\beta = 1$ is satisfied if f belongs to the range of T^* :

$$\begin{aligned} T^* k = f &\Leftrightarrow, \\ \int g(y-x) \pi_Y(y) k(y) dy = \pi_X(x) f(x) &\Leftrightarrow, \\ \int g(y-x) k^*(y) dy = f^*(x), & \qquad \textbf{(B.3)} \end{aligned}$$

where $k^* \equiv \pi_Y k$, $f^* \equiv \pi_X f$. Denote $\mathcal{F}(g)$, $\mathcal{F}(k^*)$, $\mathcal{F}(f^*)$ the Fourier transforms of g, k^*, and f^*, respectively; that is, $\mathcal{F}(g)(t) = \int e^{-it\varepsilon} g(\varepsilon)\, d\varepsilon$. Now (B.3) is equivalent to

$$\begin{aligned}
&\mathcal{F}(g)\mathcal{F}(k^*) = \mathcal{F}(f^*) \Leftrightarrow, \\
&\qquad k^*(y) = \frac{1}{2\pi}\int e^{ity}\frac{\mathcal{F}(f^*)(t)}{\mathcal{F}(g)(t)}dt \Leftrightarrow, \\
&\qquad k(y) = \frac{1}{2\pi}\frac{1}{\pi_Y(y)}\int e^{ity}\frac{\mathcal{F}(f\pi_X)(t)}{\mathcal{F}(g)(t)}dt \quad \text{for any } y \text{ in the support of } \pi_Y,
\end{aligned}$$

by the inversion formula. The condition $\int |k(y)|^2 \pi_Y(y)\, dy < \infty$ is equivalent to

$$\int \frac{1}{\pi_Y(y)}\left|\int e^{ity}\frac{\mathcal{F}(f\pi_X)(t)}{\mathcal{F}(g)(t)}dt\right|^2 dy < \infty. \tag{B.4}$$

Take $\pi_Y = 0.5I[-1,1]$ and $\pi_X = 1$. Now (B.4) is satisfied as soon as

$$\int \left|\frac{\mathcal{F}(f)(t)}{\mathcal{F}(g)(t)}\right| dt < \infty.$$

Using a change of variables $t \to -t$, this is equivalent to

$$\int \left|\frac{\Psi_X(t)}{\Psi_\varepsilon(t)}\right| dt < \infty. \qquad \blacksquare$$

Proof of Proposition 3. Under condition (3.2) and Assumption 4, we have

$$\begin{aligned}
\text{Var} &\le C\frac{1}{n}\sum_{j=0}^{\infty}\left(\frac{\lambda_j}{\alpha_n + \lambda_j^2}\right)^2, \\
Bias^2 &= \alpha_n^2 \sum_{j=0}^{\infty}\frac{\langle f, \varphi_j\rangle^2}{\left(\alpha_n + \lambda_j^2\right)^2} \le \alpha_n^2 D \sum_{j=0}^{\infty}\left(\frac{\lambda_j}{\alpha_n + \lambda_j^2}\right)^2,
\end{aligned}$$

where C and D are some positive constant. Then for $\alpha_n \le dn^{-1/2}$, the rate of convergence of the MISE is given by

$$MISE \le \tilde{C}\frac{1}{n}\sum_{j=0}^{\infty}\left(\frac{\lambda_j}{\alpha_n + \lambda_j^2}\right)^2, \tag{B.5}$$

where $\tilde{C}$ is some positive constant. $\blacksquare$

Proof of Corollary 1. We look for an equivalent of the series in (B.5). For this, we use the following result: Let $f(j)$ be the element of a series and assume $f(j)$ is a positive and continuous decreasing function of j. Then it is easy to see that

$$\int_0^J f(s)\, ds + f(J) \le \sum_{j=0}^{J} f(j) \le \int_0^J f(s)\, ds + f(0) \quad \text{for all } J \ge 1.$$

When α_n goes to zero, an equivalent of the series is given by

$$\sum_{j=0}^{\infty} f(j) \sim \int_0^{\infty} f(s)\, ds.$$

In the normal case, the eigenvalues satisfy $\lambda_j = \rho^{j/2}$ with $|\rho| < 1$ so that as α_n goes to zero,

$$\sum_{j=0}^{\infty} \frac{\lambda_j^2}{\left(\alpha_n + \lambda_j^2\right)^2} \sim \int_0^{\infty} \frac{\rho^s}{(\alpha_n + \rho^s)^2} ds = -\frac{1}{\ln(\rho)} \left[\frac{1}{\alpha_n + \rho^s}\right]_0^{\infty} \sim -\frac{1}{\ln(\rho)\,\alpha_n}. \tag{B.6}$$

The rate for the MISE follows. ■

Proof of Proposition 4. Below, C and D denote arbitrary positive constants.

Variance.

$$\begin{aligned}
\operatorname{var}\left(\pi_Y(Y)\psi_j(Y)\right) &= \mathrm{E}\left[\psi_j(Y)^2\right] - \mathrm{E}\left[\psi_j(Y)\right]^2 \\
&= \frac{1}{L}\mathrm{E}\left(e^{ij4\pi Y/L}\right) - \frac{1}{L}\left[\mathrm{E}\left(e^{ij2\pi Y/L}\right)\right]^2 \\
&= \frac{1}{L}\Psi_Y\left(\frac{j4\pi}{L}\right) - \frac{1}{L}\Psi_Y\left(\frac{j2\pi}{L}\right)^2.
\end{aligned}$$

The second term on the right-hand side is negligible with respect to the first. Moreover, we have

$$\begin{aligned}
\Psi_Y\left(\frac{j4\pi}{L}\right) &= \Psi_\varepsilon\left(\frac{j4\pi}{L}\right)\Psi_X\left(\frac{j4\pi}{L}\right) \\
&= \Psi_\varepsilon\left(\frac{j4\pi}{L}\right)\langle f, \varphi_{2j}\rangle \\
&\leq \left|\lambda_j\right|^2
\end{aligned}$$

by condition (3.4). Hence, the variance is dominated by

$$\begin{aligned}
\mathrm{Var} &= \frac{1}{n}\sum_{j=0}^{\infty}\left(\frac{|\lambda_j|}{\alpha_n + |\lambda_j|^2}\right)^2 \operatorname{var}\left[\pi_Y(Y_i)\psi_j(Y_i)\right] \\
&\leq \frac{1}{n}\sum_{j=0}^{\infty}\left(\frac{|\lambda_j|^2}{\alpha_n + |\lambda_j|^2}\right)^2.
\end{aligned}$$

Using the same approach as in the proof of Corollary 1, we have

$$\begin{aligned}
\sum_{j=0}^{\infty}\left(\frac{|\lambda_j|^2}{\alpha_n + |\lambda_j|^2}\right)^2 &\sim \int_0^{\infty}\left(\frac{s^{-2a}}{\alpha_n + s^{-2a}}\right)^2 ds \\
&= \int_0^{\infty}\frac{1}{\left(\alpha_n s^{2a} + 1\right)^2} ds \\
&\leq C\alpha_n^{-1/(2a)},
\end{aligned}$$

where the last inequality follows from a change of variables $y = \alpha_n^{1/(2a)} s$. Hence $\mathrm{Var} \leq C n^{-1} \alpha_n^{-1/(2a)}$.

Bias.

$$Bias^2 \leq \alpha_n^2 \sum_{j=0}^{\infty} \left(\frac{|\lambda_j|}{\alpha_n + |\lambda_j|^2} \right)^2.$$

We have

$$\sum_{j=0}^{\infty} \left(\frac{|\lambda_j|}{\alpha_n + |\lambda_j|^2} \right)^2 \sim \int_0^{\infty} \frac{s^{-2a}}{\left(\alpha_n + s^{-2a}\right)^2} ds = \int_0^{\infty} \frac{s^{2a}}{\left(\alpha_n s^{2a} + 1\right)^2} ds.$$

Using an integration by parts with $u = s$, $v' = s^{2a-1} / \left(\alpha_n s^{2a} + 1\right)^2$, we obtain

$$\int_0^{\infty} \frac{s^{2a}}{\left(\alpha_n s^{2a} + 1\right)^2} ds \sim \frac{1}{\alpha_n} o(1) + \frac{1}{\alpha_n 2a} \int_0^{\infty} \frac{1}{\left(\alpha_n s^{2a} + 1\right)} ds \leq D \frac{1}{\alpha_n} \frac{1}{\alpha_n^{1/(2a)}}.$$

Hence $Bias^2 \leq D\alpha_n^{1-1/(2a)}$ and $MISE \leq C n^{-1} \alpha_n^{-1/(2a)} + D\alpha_n^{1-1/(2a)}$. Finally, for a choice $\alpha_n = dn^{-1}$, we get the result. ∎

Proof of Proposition 5. The argument (x) in $Z_{n1}(x)$ is omitted in the proof to simplify notations. First we check that $\mathrm{var}(Z_{n1})$ is bounded from below.

$$\mathrm{var}(Z_{n1}) = \sum_j \frac{1}{\left(\alpha_n + \lambda_j^2\right)^2} \sigma_j^2 \left|\varphi_j(x)\right|^2 + 2 \sum_{j<k} \frac{1}{\left(\alpha_n + \lambda_j^2\right)\left(\alpha_n + \lambda_k^2\right)} \sigma_{ij} \varphi_j(x) \overline{\varphi_k(x)},$$

where

$$\sigma_{ij} = cov\left(\left\langle \pi_{Y|X}(Y_i|.), \varphi_j(.)\right\rangle, \left\langle \pi_{Y|X}(Y_i|.), \varphi_k(.)\right\rangle\right) = \lambda_j \lambda_k cov\left(\pi_Y \psi_j, \pi_Y \psi_k\right),$$

using the same rewriting as in (B.1). As $\mathrm{var}(Z_{n1})$ is a sum of positive terms, it is bounded from below. To establish (3.5) for $\delta = 1$, we need to show that

$$\frac{\mathrm{E}|Z_{n1} - \mathrm{E}(Z_{n1})|^3}{n^{1/2}} \to 0.$$

Using $\left\langle \pi_{Y|X}(y_1|.) - T^* h, \varphi_j(.)\right\rangle = \lambda_j \left[\pi_Y(y_1)\psi_j(y_1) - \mathrm{E}\left(\pi_Y \psi_j\right)\right]$, x_{n1} can be rewritten as

$$Z_{n1} - \mathrm{E}(Z_{n1}) = \sum_{j=0}^{\infty} \frac{\lambda_j}{\alpha_n + \lambda_j^2} \left[\pi_Y \psi_j - \mathrm{E}\left(\pi_Y \psi_j\right)\right] \varphi_j(x).$$

We have

$$\mathrm{E}\left|Z_{n1}-\mathrm{E}\left(Z_{n1}\right)\right|^3 \leq \sum_{j=0}^{\infty}\left(\frac{\lambda_j}{\alpha_n+\lambda_j^2}\right)^3$$

$$\times\,\mathrm{E}\left\{\left[\pi_Y\psi_j-\mathrm{E}\left(\pi_Y\psi_j\right)\right]^3\right\}\left|\varphi_j(x)\right|^3+\text{cross}=\text{products}.$$

The cross-products are dominated by the first term. The result follows from Assumption 5. ■

Proof of Lemma 2.

$$\operatorname{var}\left(Z_{n1}\right)=O\left(\sum_j\left(\frac{\lambda_j}{\alpha_n+\lambda_j^2}\right)^2\mathrm{E}\left\{\left[\pi_Y\psi_j-\mathrm{E}\left(\pi_Y\psi_j\right)\right]^2\right\}\left|\varphi_j(x)\right|^2\right).$$

Under Assumption 6, $\frac{1}{n}\operatorname{var}(Z_{n1})\to 0$, which implies the weak law of large numbers (WLLN) by Theorem C of Serfling (1980, p. 27). For the WLLN of Z_{ni}^2, we use

$$\operatorname{var}\left(Z_{n1}^2\right)=O\left(\sum_j\left(\frac{\lambda_j}{\alpha_n+\lambda_j^2}\right)^4\mathrm{E}\left\{\left[\pi_Y\psi_j-\mathrm{E}\left(\pi_Y\psi_j\right)\right]^4\right\}\left|\varphi_j(x)\right|^4\right).\quad ■$$

Proof of Proposition 6. Under Assumption 7, we have

$$\frac{\left|f-f^{\alpha}\right|^2}{\operatorname{var}\left(\hat{f}\right)}=\frac{\alpha_n^2\left|\sum_{j=0}^{\infty}\frac{1}{\alpha_n+\lambda_j^2}\left\langle f,\varphi_j\right\rangle\varphi_j\right|^2}{\frac{1}{n}\operatorname{var}\left(Z_{n1}\right)}$$

converges to zero. By Assumption 6 and Lemma 2, $\operatorname{var}(Z_{n1})$ can be replaced by the sample variance. ■

Proof of Proposition 8. Assume T is not injective, i.e., there exists a nonzero function φ in $L^2(\pi_X)$ such that $T\varphi=0$. We want to show that necessarily $\Psi_\varepsilon(t)=0$ for some t. Let $\mathcal{F}_\varphi$ denote the Fourier transform of φ; i.e., $\mathcal{F}_\varphi=\int e^{itx}\varphi(x)\,dx$ for an arbitrary function φ. By the convolution theorem, it follows that

$$\begin{aligned}(T\varphi)(y)&=\int g(y-x)\varphi(x)\,dx=0\quad\text{for all } y\\&\Leftrightarrow\mathcal{F}_g(t)\mathcal{F}_\varphi(t)=0\quad\text{for all } t\\&\Leftrightarrow\Psi_\varepsilon(t)\mathcal{F}_\varphi(t)=0\quad\text{for all } t.\end{aligned}$$

Since $\mathcal{F}_\varphi(t)$ cannot be equal to zero for all t, Ψ_ε necessarily has some zeros. ■

Proof of Proposition 9. Let f_1 and f_2 be two densities so that $\widetilde{T}f_1=h$ and $\widetilde{T}f_2=h$. It follows that $\widetilde{T}(f_1-f_2)=0$ and, by the convolution theorem; $\Psi_\varepsilon(t)\left(\mathcal{F}_{f_1}(t)-\mathcal{F}_{f_2}(t)\right)=0$ for all t. Hence the CF of f_1 and f_2 may differ only on the isolated points $t_1,t_2,\ldots$ where $\Psi_\varepsilon(t)=0$. By the continuity of the CF, $\mathcal{F}_{f_1}(t)$ and $\mathcal{F}_{f_2}(t)$ have to agree, and therefore $f_1=f_2$. ■

Proof of Proposition 10. We have

$$\left\|\hat{f}_{\mathcal{D}}-f\right\|^2\leq\left\|\hat{f}_{\mathcal{D}}-f_{\mathcal{D}}^{\alpha}\right\|^2+\left\|f_{\mathcal{D}}^{\alpha}-f\right\|^2,$$

where $f_{\mathcal{D}}^{\alpha}$ is the solution to (4.5) where $\hat{h}$ has been replaced by h. The rate of the first term on the right-hand side is given by

$$\left\| \hat{f}_{\mathcal{D}} - f_{\mathcal{D}}^{\alpha} \right\| = O\left(\frac{\delta}{\sqrt{\alpha}} \right).$$

Indeed, according to Engl et al. (1996, Thm. 5.16), we have

$$\left\| \hat{f}_{\mathcal{D}} - f_{\mathcal{D}}^{\alpha} \right\| \leq \frac{\left\| Q\left(\hat{h} - h\right) \right\|}{\sqrt{\alpha}} \leq \frac{\|Q\| \left\| \hat{h} - h \right\|}{\sqrt{\alpha}} = O\left(\frac{\delta}{\sqrt{\alpha}} \right),$$

where Q is the orthogonal projector of $L^2(\pi_Y)$ onto $\overline{\mathcal{R}(T)}$. The rate of the regularization bias in the constrained case cannot be slower than in the unconstrained case because the true f is known to be a density. Hence, we have

$$\left\| f_{\mathcal{D}}^{\alpha} - f \right\|^2 = O\left(\alpha^{\beta \wedge 2} \right).$$

It follows that

$$\left\| \hat{f}_{\mathcal{D}} - f \right\|^2 = O\left(\frac{\delta^2}{\alpha} + \alpha^{\beta \wedge 2} \right). \tag{B.7}$$

Setting the two terms in the right-hand side of (B.7) equal to each other yields the result. ■