CENTER FOR POLICY RESEARCH

THE MAXWELL SCHOOL

WORKING PAPER SERIES

Density Deconvolution with Laplace Errors and Unknown Variance

Jun Cai, William C. Horrace, and Christopher F. Parmeter

Paper No. 225
June 2020 (Revised from March 2020)

ISSN: 1252-3066

SYRACUSE UNIVERSITY
Maxwell School
Center for Policy Research

426 Eggers Hall
Syracuse University
Syracuse, NY 13244-1020
(315) 443-3114/email: ctrpol@syr.edu
http://www.maxwell.syr.edu/CPR_Working_Papers.aspx

Abstract

We consider density deconvolution with zero-mean Laplace noise in the context of an error component regression model. We adapt the minimax deconvolution methods of Meister (2006) to allow estimation of the unknown noise variance. We propose a semi-uniformly consistent estimator for an ordinary-smooth target density and a modified "variance truncation device" for the unknown noise variance. We provide a simulation study and practical guidance for the choice of smoothness parameters of the ordinary-smooth target density. We apply restricted versions of our estimator to a stochastic frontier model of US banks and to a measurement error model of daily saturated fat intake.

JEL No.: C12, C14, C44, D24

Keywords: Efficiency Estimation, Laplace Distribution, Stochastic Frontier

Authors: Jun Cai, Department of Economics, Center for Policy Research, Syracuse University, jcai106@syr.edu; William C. Horrace, Department of Economics, Center for Policy Research, Syracuse University, whorrace@syr.edu; Christopher F. Parmeter, Department of Economics, University of Miami, cparmeter@bus.miami.edu

CENTER FOR POLICY RESEARCH – Summer 2020

Leonard M. Lopoo, Director
Professor of Public Administration and International Affairs (PAIA)

Associate Directors

Margaret Austin
Associate Director, Budget and Administration

John Yinger
Trustee Professor of Economics (ECON) and Public Administration and International Affairs (PAIA)
Associate Director, Center for Policy Research

SENIOR RESEARCH ASSOCIATES

Badi Baltagi, ECON
Robert Bifulco, PAIA
Leonard Burman, PAIA
Carmen Carrión-Flores, ECON
Alfonso Flores-Lagunes, ECON
Sarah Hamersma, PAIA
Madonna Harrington Meyer, SOC
Colleen Heflin, PAIA
William Horrace, ECON
Yilin Hou, PAIA
Hugo Jales, ECON
Jeffrey Kubik, ECON
Yoonseok Lee, ECON
Amy Lutz, SOC
Yingyi Ma, SOC
Katherine Michelmore, PAIA
Jerry Miner, ECON
Shannon Monnat, SOC
Jan Ondrich, ECON
David Popp, PAIA
Stuart Rosenthal, ECON
Michah Rothbart, PAIA
Alexander Rothenberg, ECON
Rebecca Schewe, SOC
Amy Ellen Schwartz, PAIA/ECON
Ying Shi, PAIA
Saba Siddiki, PAIA
Perry Singleton, ECON
Yulong Wang, ECON
Peter Wilcoxen, PAIA
Maria Zhu, ECON

GRADUATE ASSOCIATES

Rhea Acuña, PAIA
Mariah Brennan, SOC. SCI.
Ziqiao Chen, PAIA
Yoon Jung Choi, PAIA
Dahae Choo, ECON
Stephanie Coffey, PAIA
Giuseppe Germinario, ECON
Myriam Gregoire-Zawilski, PAIA
Jeehee Han, PAIA
Mary Helander, SOC. SCI.
Hyoung Kwon, PAIA
Mattie Mackenzie-Liu, PAIA
Maeve Maloney, ECON
Austin McNeill Brown, SOC. SCI.
Qasim Mehdi, PAIA
Claire Pendergrast, SOC
Krushna Ranaware, SOC
Christopher Rick, PAIA
Huong Tran, ECON
Joaquin Urrego, ECON
Yao Wang, ECON
Yi Yang, ECON
Xiaoyan Zhang, Human Dev.
Bo Zheng, PAIA
Dongmei Zuo, SOC. SCI.

STAFF

Joseph Boskovski, Manager, Maxwell X Lab
Katrina Fiacchi, Administrative Specialist
Michelle Kincaid, Senior Associate, Maxwell X Lab
Emily Minnoe, Administrative Assistant
Candi Patterson, Computer Consultant
Samantha Trajkovski, Postdoctoral Scholar
Laura Walsh, Administrative Assistant

1 Introduction

Deconvolution uses kernel techniques to estimate the density (the *target density*) of a random variable (u) in the presence of an independent and additive noise term (v). Most deconvolution estimators are for a random cross-section of observations from a noisy random variable (i.e., $\varepsilon = u + v$), where the noise distribution (f_v) is known. If we know f_v and (hence) its characteristic function, then under regularity conditions we can calculate the empirical characteristic function of ε and use the Fourier inversion formula to consistently point estimate f_u. Fan (1991) shows that convergence rates for kernel deconvolution estimators depend on the smoothness of the noise distribution, where smoothness is characterized by the tail behavior of the associated characteristic function. Specifically, if v is from the super-smooth family (e.g., normal or Cauchy), the fastest convergence rate is logarithmic in the sample size (n), and if noise is from the ordinary-smooth family (e.g. Laplace or gamma), the fastest rate is polynomial in n.[1]

However, in applications (like the stochastic frontier model) it may be more practical to assume that the noise distribution is known *up to its variance*. Hence, Meister (2006) develops a semi-uniformly consistent estimator of the target density and the unknown noise variance, when the noise density is super-smooth (e.g., normal) and the target density is ordinary-smooth (e.g., gamma), which bounds the decay of the tails of its characteristic function.[2] Horrace and Parmeter (2011) adapt the estimator of Meister (2006) to the stochastic frontier model (Aigner et al., 1977), where the noisy random variable (ε) is appended to a linear regression model, v is normally distributed, and u is ordinary-smooth and non-negative.[3]

[1] We give a precise definition of smoothness in the sequel. Deconvolution applications for v normal (super-smooth) abound. See Stefanski and Carroll, 1990; Neumann, 1997; Johannes, 2009; Wang and Ye, 2012.

[2] Others are Butucea and Matias (2004) and Butucea, Matias, and Pouet (2008). The Meister (2006) estimator is uniformly consistent relative to the target distributional family but individually relative the noise distributional family. That is, consistency of the estimator does not hold uniformly over all noise distributions.

[3] Horowitz and Markatou (1996) consider deconvolution in the linear regression model for panel data.

That is, for a linear production function with normally distributed (super-smooth) noise (v), we may estimate the density of technical inefficiency (u), if it belongs to the ordinary-smooth family (e.g., exponential or gamma). Unfortunately, the convergence results of Fan (1991) still apply: both the Meister (2006) and Horrace and Parmeter (2011) estimators converge at logarithmic rates. Therefore, it is natural to consider a version of Horrace and Parmeter (2011) where noise is Laplace (ordinary-smooth), so as to achieve polynomial convergence rates for estimators of the density of technical inefficiency. This is the goal of this paper.

Laplace noise is not unprecedented in the literature. Horrace and Parmeter (2018) develop a parametric stochastic frontier model with Laplace noise which possess useful features for ranking and selecting efficient firms.[4] Meister (2004) shows that in a deconvolution problem if the noise distribution is misspecified, it is always better to assume Laplace noise rather than normal, because normal noise produces infinite risk while Laplace noise produces finite risk. A similar result arises in the simulations of Horrace and Parmeter (2018) who find that the mean squared error (MSE) of the parametric stochastic frontier model is smaller with Laplace noise than with normal noise under misspecification of the noise distribution. Errors-in-variable models have recently considered Laplace errors. See Carroll et al. (2006), Koul and Song (2014), Song et al. (2016), Cao (2016) and references therein. Finally, maximum likelihood estimation with Laplace errors produces the least absolute deviations (LAD) estimator, and applications of this method are plentiful in statistics, finance, engineering, and other applied sciences (see Dodge, 1987, 1992, 1997 and Dodge and Falconer, 2002).

Our aim here is to provide a complete account of Laplace kernel deconvolution and to develop a regression-based deconvolution estimator that does not require the variance of the Laplace distribution to be known. We modify the "variance truncation device" of Meister (2006) to bound of the variance of the noise (v) with the variance of the noisy random

[4] Horrace and Parmeter do maximum likelihood estimation of the stochastic frontier model, not deconvolution.

variable (ε). Target density estimation is drastically improved (in terms of convergence) with Laplace noise and is robust to misspecification of the noise distribution (per Meister, 2004). Moreover, we offer practical guidance and an adaptive procedure for selecting the smoothness parameters which are key to implementation of the proposed techniques (and which will be discussed later). This adaptive procedure is new in the literature and offers sound footing for practical use of these methods. Lastly, we apply the Laplace deconvolution estimator to two restricted versions of the model: a stochastic (cost) frontier model (SFM), where u is restricted non-positive, and a pure deconvolution problem, where the linear regression parameters are restricted to equal zero.

The paper is organized as follows. In Section 2 we discuss the basic issues surrounding deconvolution in the regression model and introduce the modified variance truncation device under Laplace errors (noise). Section 3 derives large sample properties of the estimator under certain regularity conditions. Two extensions are considered in Section 4. Section 5 contains a variety of Monte Carlo results demonstrating the finite sample performance of the proposed estimator as well as issues pertaining to robustness of the choice of the Laplace noise. In Section 6 we provide two practical applications to illustrate the utility of the proposed methodology. Conclusions are in Section 7.

2 The Laplace Convolution Problem

Consider the error component model (ECM) in the cross sectional setting:

$$y_j = x_j'\beta + u_j + v_j = x_j'\beta + \varepsilon_j, \quad j = 1, \ldots, n. \tag{1}$$

Here j indexes individuals or firms, β is a parameter vector of dimension q to be estimated and exogenous covariates are $x \in \mathcal{R}^q$. The ε is a composed error term, u is the target error component, and v is statistical noise. Depending on assumptions on u, the model in (1) can be a cross sectional stochastic frontier model (e.g., $u \sim Exp(\sigma_u^2)$), a linear regression with measurement error (e.g., $y_j = x_j^* \beta + v_j$, where $x_j^* = x_j + e_j$, $u_j = \beta * e_j$), or a pure measurement error model (e.g., $\beta = 0$). A large statistical literature investigates the $\beta = 0$ model with known or partially-known error distribution of v (see Meister, 2009).[5] In this setting, deconvolution is complicated by the fact that only cross sectional data are available. Following the literature (i.e., Fan, 1991; Meister, 2006; Horrace and Parmeter, 2011), we make the following assumptions on the random components of the model and the covariates when present.

Assumption 1. *The x_j, v_j and u_j are pairwise independent for all $j = 1, \ldots, n$.*

Let the probability densities of the error components be $f_v(z)$, $f_u(z)$ and $f_\varepsilon(z)$ with corresponding characteristic functions $h_v(\tau)$, $h_u(\tau)$ and $h_\varepsilon(\tau)$. Based on the independence between v_j and u_j in Assumption 1,

$$h_\varepsilon(\tau) = h_v(\tau) h_u(\tau). \tag{2}$$

We restrict v to the family of Laplace densities with the following assumption.

Assumption 2. *The distribution of v is a member of the Laplace family with zero mean and unknown variance, i.e. $\mathcal{L} = \{Laplace(0, b) : b^2 > 0\}$.*

Hence, the density of v is known up to its variance ($2b^2$), and the characteristic function of v is $h_v(\tau) = (1 + b^2\tau^2)^{-1}$, so that,

$$h_u(\tau) = \frac{h_\varepsilon(\tau)}{h_v(\tau)} = (1 + b^2\tau^2) h_\varepsilon(\tau). \tag{3}$$

[5]Neumann (1997), Johannes (2009), and Wang and Ye (2012) study deconvolution with *fully unknown* error distribution but require either an additional sample of the error or repeated observations, y_{jt}.

We restrict u to be ordinary-smooth (Fan, 1991) with the following assumption.

Assumption 3. *Assume u is ordinary-smooth. Namely, u belongs to the family* $\mathcal{F}_u = \big\{h_u : C_1|\tau|^{-\delta} \leq |h_u(\tau)| \leq C_2|\tau|^{-\delta}, for \quad |\tau| \geq T > 0\big\}$ *where* $0 < C_1 < C_2$ *and* $\delta > 1$*,* $\delta \neq 2$.

Assumption 3 dictates tail behavior of the characteristic function of u (smoothness of the density of u), and positive constants C_1, C_2 and δ are *smoothness parameters*. The lower bound, C_1, and upper bound, C_2, ensure the rate of decay of the tails of the characteristic function does not approach zero too rapidly or too slowly and are needed for identification. Constants C_1 and C_2 become irrelevant when T gets large. Practically speaking, we only use the lower bound to define our variance truncation device, so only C_1 is relevant to our estimator. We assume C_1 and δ to be known for now but will relax this in the sequel.[6]

Constant δ is the *smoothness order*, ensuring polynomial tail behavior of the characteristic function, and includes a wide array of nonparametric and analytical families (Horrace and Parmeter, 2011). Common families and their polynomial smoothness orders are tabulated in Table 1. For example, the Symmetric Uniform family of distributions has polynomial order $\delta = 1$, and the Laplace family has $\delta = 2$. We restrict $\delta \neq 2$ in Assumption 3 so that the target density cannot be Laplace, allowing our estimator to appropriately assign the target and noise distributions. That is, if u and v are both Laplace, we cannot determine which distribution is the target and which is the noise.[7] In the parlance of frontier estimation, when $\delta = 2$ we cannot distinguish the signal from the noise. Letting $\delta = 2$ does not preclude deconvolution *per se*. For example, the deconvolution convolution estimator of Dattner et al. (2011) relies on very general classes of distributions for the target and noise densities that includes the Laplace-Laplace convolution as a special case, and consistent target density estimation is achieved as long as the error variance is known. The restriction in Assumption 3

[6] Knowing C_1 and δ does not imply knowing $V(u)$ nor does it uniquely determine the analytic family.

[7] We are grateful to an anonymous reviewer who alerted us to this identification issue. It should be noted that the restriction eliminates a broad class of ordinary-smooth distributions, not just the Laplace.

that $\delta > 1$ does not preclude a nonparamteric family of densities in Table 1 that is arbitrarily close to a family with $\delta = 1$, like the Uniform or the Exponential (i.e, a Gamma with $k = 1$ in the table), which have both been employed in Stochastic Frontier Analysis.

Note that Meister (2006) assumes different distributional families for u and v (i.e., ordinary-smooth and super-smooth, respectively) and that simplifies derivation of the convex upper bound of the criterion function in that paper. The intuition is that as n goes to infinity the tail of h_v (normal noise) decays faster than that of h_u. Turning to Table 1, we see that the normal distribution has polynomial order $\delta \to \infty$, so the intuition is justified.[8] In the current paper similar intuition applies, but the key here is that the tails of characteristic function of u and v decay at *different* rates with the polynomial order of the noise decay fixed at 2 by design.

Under Assumptions 2 and 3, the Fourier inversion formula returns the density of u,

$$f_u(z) = \frac{1}{2\pi} \int e^{-i\tau z}(1 + b^2\tau^2)h_\varepsilon(\tau)d\tau, \tag{4}$$

where $i = \sqrt{-1}$. If noise $v \sim \mathcal{G} = \{N(0, \sigma^2) : \sigma^2 > 0\}$, Meister (2006) shows that there is no uniformly consistent estimator of $f_u(z)$ when σ^2 is unknown. His deconvolution estimator of $f_u(z)$ is semi-uniformly consistent in the sense that for a given density in $\mathcal{G}$ whose variance is bounded, a deconvolution estimator is uniformly consistent but not uniformly consistent over all densities within $\mathcal{G}$. This is the price one pays for not knowing the variance. Here we focus on the Laplace noise case with unknown variance. As we shall demonstrate, with Laplace noise one still pays a price for not knowing the variance, but the cost is not as high as in the case with normally distributed noise.

Since h_ε is unknown, we may rely on the empirical characteristic function to recover the

[8] Indeed neither the Normal nor Cauchy families of distribution are ordinary-smooth; they are *super-smooth*. See Fan (1991).

density of u based on equation (4),

$$\hat{h}_{\varepsilon}(\tau) = \left| \frac{1}{n} \sum_{j=1}^{n} e^{i\tau\varepsilon_j} \right|. \tag{5}$$

As mentioned previously, ε_j is unobserved when $\beta \neq 0$. Therefore, we must estimate it by consistently estimating the unknown parameter β first. That is, for a consistent estimator β_n, define $\hat{\varepsilon}_j = y_j - x_j'\beta_n$. Again, we take advantage of the empirical characteristic function of the residuals, which is defined as

$$\hat{h}_{\hat{\varepsilon}}(\tau) = \left| \frac{1}{n} \sum_{j=1}^{n} e^{i\tau\hat{\varepsilon}_j} \right|. \tag{6}$$

Replacing h_{ε} with $\hat{h}_{\varepsilon}$ or $\hat{h}_{\hat{\varepsilon}}$ in equation (4) does not ensure that the integration exists, so we convolve the integrand with a smoothing kernel (Stefanski and Carroll, 1990). Define a random variable z with the usual Parzen (1962) kernel density $K(z)$ and corresponding (invertible) characteristic function $h_K(\tau)$. Finite support of the characteristic function $h_K(\tau)$ is required to ensure the integrand exists and the resulting estimate is a valid density function.

Using $K(z) = (\pi z)^{-1} sin(z), (h_K(\tau) = 1\{|\tau| \leq 1\})$, our estimator of the density of u is,

$$\hat{f}_u(z) = \frac{1}{2\pi} \int_{-w_n}^{w_n} e^{-i\tau z}(1 + \hat{b}_n^2\tau^2) \left| \frac{1}{n} \sum_{j=1}^{n} e^{i\tau\hat{\varepsilon}_j} \right| d\tau, \tag{7}$$

where the limits of the integration are a function of an increasing sequence of positive constants w_n, which represent the degree of smoothing. In the sequel, $\{w_n\}_{n\in\mathbb{N}}$, $\{k_n\}_{n\in\mathbb{N}}$ and $\{b_n^2\}_{n\in\mathbb{N}}$ denote sequences of positive numbers which will be determined later. k_n is an intermediate sequence that will be useful for the case where C_1 and δ are unknown. When C_1

and δ are known, set $w_n = k_n$.[9]

Due to the upper and lower bound conditions on the target density function in Assumption 3, we propose an estimator of unknown error variance parameter, b^2. Therefore, setting $\tilde{b}_n^2 = k_n^{-2}\left(\frac{C_1 k_n^{-\delta}}{\hat{h}_{\hat{\varepsilon}}(k_n)} - 1\right)$ with constants $\delta > 1$ and $C_1 > 0$, we propose an explicit truncation device for the unknown variance parameter:

$$\hat{b}_n^2 = \begin{cases} 0 & \text{if } \tilde{b}_n^2 < 0 \\ \tilde{b}_n^2 & \text{if } \tilde{b}_n^2 \in [0, b_n^2] \\ b_n^2 & \text{if } \tilde{b}_n^2 > b_n^2, \end{cases} \tag{8}$$

where the variance parameter bound is $b_n^2 = \frac{1}{2}V(\hat{\varepsilon})$, half the variance of the estimated sum of the error components. The intuition is that we choose an increasing sequence to cover the unknown variance parameter, $\tilde{b}_n^2$, but bound it by half the total variance.[10] This is a modified version of the variance truncation device of Meister (2006).

What distinguishes our truncation device from that in Meister (2006) is that the variance of the estimated compound error is incorporated as a natural upper bound of the unknown variance of random noise v. Compared to the variance truncation device of Meister (2006), ours is more informative and converges faster, while still covering the unknown error variance associated with Laplace errors. Meister (2006) uses the bound $b_n^2 = \frac{1}{4}\ln\ln n$ for deconvolution with normal errors, and his bound arises directly from the characteristic function of the normal distribution and implicitly requires a very large sample size n. The modified truncation device, $\hat{b}_n^2$, is an important contribution of this paper which can also be applied in the setting of Meister (2006). Its attractiveness and usefulness will be demonstrated in the simulation section. We now discuss semi-uniform consistency of the Laplace deconvolution

[9] In Section 4, we propose setting $w_n = k_n/\ln k_n$ in the case C_1 and δ are not fully known.

[10] Recall that for a Laplace distribution as defined in Assumption 2, the variance is $V(v) = 2b^2$. Moreover, $V(v) < V(\varepsilon)$ under Assumption 1. Hence, a natural upper bound for b^2 is one-half the variance.

estimator in equation 7.

3 Asymptotic Theory

To demonstrate that the unknown variance deconvolution estimator retains its asymptotic properties when the composed error is estimated, we introduce two additional conditions that will be useful in the Lemmas and Theorem to follow.

Assumption 4. *The distribution of x has bounded support.*

Assumption 5. *The estimator β_n converges at a rate of square root n. That is, $\sqrt{n}(\beta_n - \beta) = O_p(1)$ as $n \to \infty$.*

Assumption 4 follows Horowitz and Markatou (1996) while Assumption 5 guarantees that the difference between the composed errors and estimated errors is asymptotically negligible. In the pure deconvolution problem, $\beta = 0$, Assumption 5 is trivially satisfied. Moreover, the conditional mean function $x_j'\beta$ may suffer from misspecification but can be estimated with a nonparametric n^a convergence rate and $a = \frac{2}{4+q}$. We will discuss this case in the extensions in Section 4.

To establish semi-uniform consistency of $\hat{f}_u$, we introduce the following lemmas.

Lemma 1. *For Assumptions 1, and 3-5 and $\mathcal{L}_n = \{Laplace(0, b) : b^2 \in (0, b_n^2]\}$, the mean integrated squared error (MISE) of (7) is*

$$\sup_{g \in \mathcal{L}_n} \sup_{f \in \mathcal{F}_u} E_{f,g} ||\hat{f}_u - f_u||_{L_2}^2 \leq B + V + E,$$

where $B \leq const_1 \times w_n^{1-2\delta}$,

$V \leq const_2 \times n^{-1} w_n (1 + b_n^2 w_n^2)^2 + const_3 \times n^{-1} w_n^3 (1 + b_n^2 w_n^2)^2$,

$E \leq const_4 \times \sup_{g \in \mathcal{L}_n} \sup_{f \in \mathcal{F}_u} \Big(w_n \int_{-1}^{1} |h_u(s w_n)|^2 \left(\frac{d_n}{b^2}\right)^2 ds + w_n \int_{-1}^{1} |h_u(w_n s)|^2 \frac{b_n^4}{b^4} \times P_{f,g}(|\hat{b}_n^2 - b^2| > d_n) ds \Big)$, with $d_n := \frac{1}{w_n}$; f and g are the probability density function in distribution family $\mathcal{F}_u$ and $\mathcal{L}_n$, respectively, and $const_j$ are positive constants for $j = 1, 2, 3, 4$.

The proof is in the appendix. Notice the distinction between $\mathcal{L}_n$ above and $\mathcal{L}$ in Assumption 2. The former is the family of Laplace distributions with an upper bound on the variance and is a subset of the latter.[11] Following Horrace and Parmeter (2011), the B term is a bias component which is bounded by the ordinary-smoothness of f_u under Assumption 3. The V terms are variance components. The E term is a hybrid bias-variance component in which the first integral behaves like squared bias and the second integral looks like a variance. This entire bound exhibits the usual bias-variance trade-off in nonparametric density estimation. Note that the second addend of V arises from the regression function, which does not appear in the pure deconvolution setting of Meister (2006).

Establishing the convergence rate of E is not straight-forward. We need the following Lemma to assist in determining it.

Lemma 2. *Let d_n, f and g be the same as in Lemma 1. Then $\sup_{g \in \mathcal{L}_n} \sup_{f \in \mathcal{F}_u} P_{f,g}\big(|\hat{b}_n^2 - b^2| > d_n\big) \leq const \times n^{-1} k_n^{2\delta} (1 + b_n^2 k_n^2)(1 + k_n^2)$.*

The proof is in the appendix. Compared to deconvolution with normal noise in Horrace and Parmeter (2011), estimation of ε matters here. That is, the conditional mean function in Horrace and Parmeter (2011) is linear, so their estimated error converges at a rate $n^{1/2}$, which is much faster than the logarithmic rate of their target density estimator. Therefore, estimation of the error can effectively be ignored. Here, both β_n and $\hat{f}_u$ converge at poly-

[11] In Meister (2006), the bounding of the normal variance is what leads to semi-uniformly consistency (as opposed to uniform consistency). Here, for Laplace errors, we still impose this "strong" condition for ease of proof. However, it may not be a necessary condition.

nomial rates, so there is an additional effect on the convergence rate of the estimator of the target density.[12] Given that we replace ε with a consistent estimator, we have an additional term k_n^2 in Lemma 2, as well as the characteristic function of the Laplace distribution, embodied in the term $(1+b_n^2k_n^2)$. The second addend of E in Lemma 1, together with the upper bound of B and the first term in E, ensures convexity of the entire bound with respect to the bandwidth parameter k_n. Therefore, the optimal bandwidth w_n, which is a function of k_n, and the entire convergence rate of the density estimator can be determined.

Notice that neither of the proofs of the above two lemmas leverage anything on the assumption that the smoothness parameters of the target density are known (or not). However, for joint minimization of the upper bounds of MISE of Lemma 1, this assumption plays a role. That is, if the smoothness parameters are fully known (i.e., C_1 and δ) tight bounds can be achieved by setting $w_n = k_n$; otherwise, the best general upper bound can be reached by setting $w_n = k_n / \ln k_n$. The latter case is considered in the next section. First, we introduce the following theorem when C_1 and δ are known.

Theorem 1. *Assume δ and C_1 are known. Under Assumption 1, 3-5, set $\{b_n^2\}_{n\in\mathbb{N}} = \frac{1}{2}V(\hat{\varepsilon})$ and $w_n = k_n$ with $\{k_n\}_{n\in\mathbb{N}} = \{(\frac{n}{b_n^2})^{\frac{1}{6+2\delta}}\}_{n\in\mathbb{N}}$, if $1 < \delta \leq 1.5$ or $\{k_n\}_{n\in\mathbb{N}} = \{(\frac{n}{b_n^8})^{\frac{1}{3+4\delta}}\}_{n\in\mathbb{N}}$, if $\delta > 1.5$. For any $g \in \mathcal{L}_n$, the proposed deconvolution kernel density estimator in equation (7) is bounded from above as follows:*

$$\sup_{f_u \in \mathcal{F}_u} E_{f,g}||\hat{f}_u - f_u||_{L_2}^2 \leq n^{-\frac{2\delta-1}{6+2\delta}} \quad if \quad 1 \leq \delta \leq 1.5,$$

[12]The compound effect of estimating the regression function will slow the target density rate compared to pure (non-regression) deconvolution, but the final rate is not a simple algebraic sum of the rates.

and

$$\sup_{f_u \in \mathcal{F}_u} E_{f,g}||\hat{f}_u - f_u||^2_{L_2} \leq n^{-\frac{2\delta-1}{3+4\delta}} \quad if \quad \delta > 1.5$$

where δ is defined in Assumption 3.

The proof is in the appendix. The proposed density estimator is semi-uniformly consistent. That is, $\hat{f}_u$ is uniformly consistent over a given class of Laplace distributions $\mathcal{L}_n$. The optimal convergence rate for an ordinary-smooth target density is achieved in a minimax sense. It is similar to the conclusions in Fan (1991), even though in this exercise the variance of the noise distribution is unknown and the composed error needs to be estimated. The polynomial convergence rate plays a role in the following sense. After imposing the modified variance truncation device, which is the proposed best choice one can use for unknown variance, and after deriving the optimal sequences for convergence (i.e., the order of the positive sequence $\{k_n\}_{n\in\mathbb{N}}$), we still achieve a polynomial convergence rate which is consistent with the lower bound derived by Fan (1991).

At first glance the Theorem 1 is similar to Theorem 2 in Meister (2006), but there are three major differences: (i) the upper bound of the noise v is not a known constant but a consistently estimated (at $\sqrt{n}$ rate) quantity (i.e., $\frac{1}{4}\ln\ln n$ versus $\frac{1}{2}V(\hat{\varepsilon})$); (ii) the chosen sequences are functions of the target density smoothness order, δ, which is due to the characteristic function of the Laplace noise, leading to different convergence rates (or effective sample size as shown in Table 2); and (iii) we consider estimation in the regression setting, which is more general than the pure deconvolution setting ($\beta = 0$), and yields different convergence rates with Laplace noise. In Horrace and Parmeter (2011) this last difference was easily handled, given the slow convergence of the density estimator due to the assumption of super-smooth noise. It is more nuanced in the context of Laplace noise, given the polynomial rate of convergence. This has important implications if one were to estimate the unknown conditional mean using nonparametric methods. We discuss this and other extensions of the Laplace deconvolution estimator in the next section.

4 Some Useful Extensions

We discuss two useful extensions to the Laplace deconvolution estimator which are likely to arise in applications: (i) C_1 and δ are unknown in Assumption 3 and (ii) deploying nonparametric regression to estimate the unknown conditional mean needed to subsequently recover $\hat{\varepsilon}$. It is rare in applications that researchers have information on the target density. This leads to uncertainty in C_1 and δ, two parameters which are important in the implementation of our estimator.[13] Also, if we wish to follow the work of Fan, Li and Weersink (1996) and estimate the unknown regression function nonparametrically, then we must think carefully about the relative polynomial convergence rates of the deconvolution estimator and the nonparametric regression estimator. This is not a consideration with normal noise due to the logarithmic convergence rates it produces.

4.1 Selection of Unknown C_1 and δ

In the usual case that δ and C_1 are unknown and, therefore, might be misspecified,[14] we could apply the following selection rule due to Meister (2006):

Selection rule 1. *If C_1 and δ are unknown, we specify one set of $\{C_1, \delta\}$ and choose $w_n = k_n / \ln k_n$.*

[13]... and the estimator of Meister (2006) as well.

[14]Actually, if one wants to assume the random noise is super-smooth with similarity index s, the smoothness parameter δ of target density can be estimated as well as the s by an adaptive procedure proposed by Butucea, Matias and Pouet (2008).

An alternative rule may be based on our procedure when δ and C_1 are known. First, we specify one set of parameters $\{C_1, \delta\}$ to pin down the variance truncation device defined in Section 2, and then by Lemmas 1 and 2 we determine the optimal choice for the sequence $\{k_n\}_{n\in\mathbb{N}}$. The trade-off is a slower convergence rate of the estimated target density compared with that in the fully-known case due to lack of information about the target density. This implicitly requires a larger n to achieve a reliable estimate of the target density. This can be seen from following theorem.

Theorem 2. *Assume δ and C_1 are unknown. Under Assumption 1, 3, 4, and 5 set $\{b_n^2\}_{n\in\mathbb{N}} = \frac{1}{2}V(\hat{\varepsilon})$ and $w_n = k_n/\ln k_n$ with $\{k_n\}_{n\in\mathbb{N}} = \{(\frac{n}{b_n^2})^{\frac{1}{6+2\delta}}\}_{n\in\mathbb{N}}$, if $1 < \delta \leq 1.5$, or $\{k_n\}_{n\in\mathbb{N}} = \{(\frac{n}{b_n^8})^{\frac{1}{3+4\delta}}\}_{n\in\mathbb{N}}$, if $\delta > 1.5$. For any $g \in \mathcal{L}_n$, the proposed deconvolution kernel density estimator in equation (7) is bounded from above as following:*

$$\sup_{f_u\in\mathcal{F}_u} E_{f,g}||\hat{f}_u - f_u||_{L_2}^2 \leq (n/\ln n)^{-\frac{2\delta-1}{6+2\delta}} \quad if \quad 1 < \delta \leq 1.5$$

and

$$\sup_{f_u\in\mathcal{F}_u} E_{f,g}||\hat{f}_u - f_u||_{L_2}^2 \leq (n/\ln n)^{-\frac{2\delta-1}{3+4\delta}} \quad if \quad \delta > 1.5$$

where δ is defined by Assumption 3.

The proof is similar to that of Theorem 1 in the appendix and is contained therein. The only difference between the bounds in Theorem 1 and in Theorem 2 is that the bounds are negative exponents of n in the former and of $n/\ln n$ in the latter, and this is the price one pays for not knowing the smoothness parameters of the target density. Based on the Theorem 2 and Table 1, we propose a rule-of-thumb adaptive procedure as follows:

Step 1: Set initial estimates for C_1 and δ. A useful rule-of-thumb is C_1 is commonly between

0 and 1; δ is between 1 and 10.

Step 2: Treating this C_1 and δ as "known," select $k_n = w_n$ and apply the proposed deconvolution techniques to construct the estimated target density, $\hat{f}_{known}(u)$, say.

Step 3: Now, with the same C_1 and δ assume they are unknown and select $w_n = k_n / \ln k_n$. Again, apply the proposed deconvolution estimator to construct the estimated target density as $\hat{f}_{unknown}(u)$, say.

Step 4: Compare the *vector* of values $\hat{f}_{known}(u)$ and $\hat{f}_{unknown}(u)$ over a discretized support with a Euclidean distance measure (e.g., $\Delta = ||\hat{f}_{known}(u) - \hat{f}_{unknown}(u)||^2$). Iterate Steps 1 to 3 until Δ is smaller than a pre-specified threshold, say 0.0001.

One caveat with this iterative approach is that Δ may be quite large initially. The essential point is that more information about the underlying distribution is revealed after several trials with combinations of the smoothness parameters. This is similar in spirit to the adaptive procedure proposed by Butucea, Matias and Pouet (2008), but their targets are a "self-similarity index" and a smoothness parameter with super-smooth noise, and not a target density.

4.2 Nonparametric Estimation of the Conditional Mean

If one is unsure of the linear specification of the conditional mean, equation (1) can be generalized to the nonparametric case as follows:

$$y_j = g(x_j) + u_j + v_j \quad j = 1, 2, \ldots, n \tag{9}$$

where $g(.)$ is unknown and $x \in \mathcal{R}^q$. Under certain regularity conditions,[15] a straightforward nonparametric kernel estimator for the unknown function $g(x)$ is:

$$\hat{g}(x) = \frac{\sum_{j=1}^{n} Y_j K(\frac{X_j - x}{\lambda})}{\sum_{j=1}^{n} K(\frac{X_j - x}{\lambda})}$$

where $K(\cdot)$ is the standard Gaussian kernel with bandwidth λ. Note that since the convergence rate of the nonparametric estimator is a polynomial function of the number of covariates, this may impact application of the Laplace deconvolution estimator.

By Theorem 2.6 (with Condition 2.1) of Li and Racine (2007), the convergence rate of the estimated function is:

$$\sup_{x \in S} |\hat{g}(x) - g(x)| = O\left(\frac{(\ln n)^{0.5}}{(n\lambda_1 \cdots \lambda_q)^{0.5}} + \sum_{s=1}^{q} \lambda_s^2 \right) \quad a.s.$$

Assuming each bandwidth (λ_s) has the same order of magnitude, the optimal choice of λ_s that minimizes $MSE[\hat{g}(x)]$ is $\lambda_s \sim n^{-\frac{1}{4+q}}$, and the resulting MSE is therefore of order $O(n^{-\frac{4}{4+q}})$. Consequently, the estimated error, $\hat{\varepsilon}$, is n^a consistent where $a = \frac{2}{4+q}$. That is, $n^a(\hat{\varepsilon} - \varepsilon) = O_p(1)$ as $n \to \infty$.

Similarly, we can establish the convergence rate as follows:

Theorem 3. *Under Assumptions 3-5, and Condition 2.1 in Li and Racine (2007) set* $\{b_n^2\}_{n\in\mathbb{N}} = \frac{1}{2}V(\hat{\varepsilon})$ *and* $w_n = k_n$ *with* $\{k_n\}_{n\in\mathbb{N}} = \{(\frac{n}{b_n^2})^{\frac{2a}{6+2\delta}}\}_{n\in\mathbb{N}}$*, if* $1 < \delta \leq 1.5$*, or* $\{k_n\}_{n\in\mathbb{N}} = \{(\frac{n}{b_n^8})^{\frac{2a}{3+4\delta}}\}_{n\in\mathbb{N}}$*, if* $\delta > 1.5$*. For any* $g \in \mathcal{L}_n$*, the proposed deconvolution kernel density estima-*

[15] Details see Condition 2.1 in Li and Racine (2007).

tor in equation (7) *is bounded from above as follows:*

$$\sup_{f_u \in \mathcal{F}_u} E_{f,g}||\hat{f}_u - f_u||^2_{L_2} \leq n^{-\frac{2a(2\delta-1)}{6+2\delta}} \quad if \quad 1 < \delta \leq 1.5$$

and

$$\sup_{f_u \in \mathcal{F}_u} E_{f,g}||\hat{f}_u - f_u||^2_{L_2} \leq n^{-\frac{2a(2\delta-1)}{3+4\delta}} \quad if \quad \delta > 1.5$$

where $a = \frac{2}{4+q}$ *and* δ *is defined by Assumption 3.*

The proof is very similar to the proof of Theorem 1 in the appendix, and a sketch of the proof is contained therein.

5 Monte Carlo Simulations

We present a Monte Carlo study of the finite sample properties of the Laplace deconvolution estimator. For ease of comparison, we follow the sample designs of Meister (2006) and Horrace and Parmeter (2011) except that we consider performance of the Laplace deconvolution with both Laplace noise (correctly specified) and normal noise (misspecified). We focus on sample sizes of $n = 500$, 1,000, and 3,000 with the linear model:

$$y_j = 4 + 3x_j + v_j + u_j, \quad j = 1, \ldots, n. \tag{10}$$

The x_js are generated from a standard normal distribution. Random noise v_j is generated from either a standard Laplace (correctly specified) or normal (misspecified) distribution for a range of values of the variance to produce several signal-to-noise settings. The u_js are generated from the twice convolved, zero-mean Laplace density for which the probability density function is $\tilde{L}(x) = \frac{1}{4}e^{-|x|}(|x|+1)$.[16] We fix the variance of u to 2. In this setting it is known that $C_1 = 1/4$, $\delta = 4$ and $T = 1$.[17]

Following Theorem 1, we choose $b_n^2 = \frac{1}{2}V(\hat{\varepsilon})$ where $\hat{\varepsilon}$ is the residual from the first-step ordinary least squares (OLS) estimation, and $k_n = n^{\frac{1}{4\delta+3}}(b_n^2)^{\frac{-4}{4\delta+3}} = n^{\frac{1}{19}}(b_n^2)^{-\frac{4}{19}}$, correspondingly as $\delta = 4 > 1.5$. To explore the impact of the relative ratio of the component variances, we consider different scenarios of the signal-to-noise ratio which is defined as the ratio of $V(u)$ and $V(v)$: $\gamma := \sigma_u^2/\sigma_v^2 \in \{1/2, 1, 2\}$. We also apply our Laplace deconvolution estimator in the misspecified case where the errors are normally distributed. We compare the performance of our estimator under misspecification to the normal deconvolution estimator of Meister (2006) which is correctly specified. Even in this case, our estimator performs fairly well. We also explore the finite sample performance of our proposed rule-of-thumb adaptive procedure when the smoothness parameters of the target density are unknown.

The performance of our estimator is assessed through the root mean integrated square error (RMISE):

$$RMISE(\hat{f}_u) = \sqrt{\frac{1}{R}\sum_{l=1}^{R}\frac{1}{M}\sum_{i=1}^{M}(\hat{f}_l(u_i) - f(u_i))^2} \tag{11}$$

where R is the number of replications and $M = 256$ is the number of evaluation points over $u \in (-5, 5)$, which is fixed across the R replications.

5.1 Laplace Deconvolution with Laplace Errors

First, we consider the case that the random noise v_j is correctly specified (i.e., v_j is drawn from a Laplace distribution with variance 1). Figures 1-3 show the results for a single random draw ($R = 1$) across various sample sizes {500, 1,000, 3,000} and compare the proposed estimator (CHP) to the true unknown density ($True$). The graphical fit of the

[16] This follows from the setting in Meister (2006).

[17] We are not concerned with C_2, since it has no bearing on any calculations for the estimator.

proposed estimator is quite good with only 500 observations (Figure 1). Most of the bias comes from estimation around the mode.[18] As the sample size increases, the RMISE of the proposed estimator (CHP) decreases from 0.0148 (Figure 1) to 0.0142 (Figure 2) and to 0.0125 (Figure 3).

Figures 4-6 show the results for a single draw ($R = 1$) and fixed sample size n =1,000 but varying the signal-to-noise ratio $\sigma_u^2/\sigma_v^2 = 2/1, 2/2, 2/4$. The proposed estimator (CHP) works very well when $\sigma_u^2/\sigma_v^2 = 2/1$ with 1,000 observations. As the signal-to-noise ratio decreases, the RIMSE of proposed estimator (CHP) increases from 0.0136 (Figure 4) to 0.0142 (Figure 5) and to 0.0180 (Figure 6). Even for the noisiest case (Figure 6) with $\sigma_u^2/\sigma_v^2 = 2/4$, the fit is very good except in an interval around the mode.

Table 3 contains detailed results from $R = 500$ simulations with varying sample sizes {500, 1,000, 3,000} and signal-to-noise ratios {1/2, 1, 2}. For each signal-to-noise setting (each column), the RMISE decreases monotonically as the sample size increases from 500 to 3,000 (down the rows), demonstrating the consistency of the proposed estimator (CHP). Unexpectedly, the RMISE is not increasing as the signal-to-noise ratio increases across the columns. This is an atypical finding that is due to the variance truncation device: when the variance of the random noise is relatively small, the estimated variance parameter $\hat{b}_n^2$ is more likely to be closer to zero which dilutes the ability of the deconvolution estimator to recover the target density. Alternatively, when the variance of the random noise is relatively large, the estimated variance is no longer near zero, but the performance of the deconvolution estimator deteriorates as there is little information in the target density taken from the compound errors. This is a limitation of the variance truncation device.

[18]Estimation of a density around the mode is difficult due to the derivative at the mode being zero (Henderson and Parmeter, 2015).

5.2 Laplace Deconvolution with Misspecified Noise

To understand the impact of misspecification of the noise distribution, we consider the performance of the proposed estimator when the true noise is distributed normal. We compare the performance of our proposed estimator (CHP) with that of Meister (2006).

As a first pass on the empirical performance, Figures 7-9 show the results for the case with fixed $\sigma_u^2/\sigma_v^2 = 2/2$ for a single draw ($R = 1$) across various sample sizes. The proposed estimator (CHP) shows decent performance even with sample size of $n = 500$ (Figure 7). The figure contains plots of the proposed estimator (CHP), the estimator of Meister (2006) ($Meister06$), and the true normal density ($True$). As the sample size increases, the RMISE of the proposed estimator (CHP) changes from 0.0151 (Figure 7) to 0.0156 (Figure 8) to 0.0137 (Figure 9). Our estimator (CHP) performs as well as Meister's when the sample size is large ($n = 3,000$). An intuitive explanation is that the proposed estimator converges faster than Meister's estimator (even under misspecification).

Figures 10-12 show the results for $R = 1$ and fixed sample size $n = 1,000$ across the various signal-to-noise ratios. The proposed estimator (CHP) performs quite well in the least noisy case even though the error distribution is misspecified. As the signal-to-noise ratio decreases, the RIMSE of the proposed estimator increases from 0.0155 (Figure 10) to 0.0156 (Figure 11) and to 0.0191 (Figure 12) whereas the RMISE of Meister's estimator increases from 0.0120 (Figure 10) to 0.0172 (Figure 11) to 0.0260 (Figure 12). When the signal-to-noise ratio decreases from 1 to 0.5 (Figures 11 and 12, respectively) the misspecified estimator even outperforms Meister's estimator.

Table 4 presents the results of $R = 500$ replications across various sample sizes and signal-to-noise ratios under misspecification. Though misspecified, the RMISE of the proposed estimator decreases monotonically as the sample size increases (down each column) for each signal-to-noise ratio setting, and it is comparable to that of Meister's correctly specified estimator. In the most noisy setting, $\sigma_u^2/\sigma_v^2 = 2/4$, the proposed estimator outperforms Meister's estimator across all sample sizes. This may be due to the faster convergence rate of the proposed estimator coupled with the fact that the characteristic functions of the normal and the Laplace are quite similar.[19] Fixing the sample size (within each row), both RMISEs increase when the signal-to-noise ratio decreases as the information that can be recovered is reduced. Overall, the proposed estimator is robust to misspecification of the error distribution and its convergence rate is faster than that of Meister's estimator.

5.3 Deconvolution With Unknown Smoothness Parameters

To verify the feasibility and performance of the proposed rule-of-thumb adaptive procedure for unknown smoothness parameters of section 4.1, a set of simulations are performed. We employ the same simulation design. Specifically, the true target density is still a twice-convolved Laplace with true smoothness parameters of $C_1 = 1/4$ and $\delta = 4$. We search on a two-dimension grid of $C_1 \in \{0.1, 0.25, 0.40, 0.55, 0.70, 0.85\}$ and $\delta \in \{2, 4, 6, 8\}$ to minimize the Euclidean distance of the two estimated densities: the estimated density assuming the chosen C_1 and δ are known and the estimated density assuming these parameters are unknown. We restrict the range of u to be $(-5, 5)$ and evaluate over 128 evenly spaced points within this range.

Figure 13 shows the estimated densities (labeled CHP for the estimate with known smoothness parameters and CHP_{UN} for the estimate with unknown parameter) and the true density (labeled $True$) for one simulation ($R = 1$) with sample size $n = 1,000$ and signal-to-noise ratio equal to 1. The chosen smoothness parameters are: $C_1 = 0.1$ and $\delta = 2$. Even though the chosen smoothness parameters are misspecified (not exactly equal to the their true values $C_1 = 1/4$ and $\delta = 4$), the overall fit of the density with estimated parameters

[19] Actually the characteristic function of the Laplace distribution is the second order Taylor expansion of that of a normal random variable with same variance (Hesse, 1999).

is quite good (CHP_{UN}) and appears to be better than the fit assuming the true values of the parameters, particularly around the mode.[20]

A more comprehensive analysis is conducted in Figures 14-16. Figure 14 shows the Euclidean distance of the estimated densities: $\Delta = ||\hat{f}_{known} - \hat{f}_{unknown}||^2$, as a function of the smoothness parameters for a single draw ($R = 1$). Figures 15 and 16 show the Euclidean distance between the true density and the estimated density taking the chosen C_1 and δ as known, $||\hat{f}_{known} - f_{true}||^2$, and unknown, $||\hat{f}_{unknown} - f_{true}||^2$, respectively. A straight comparison of the three figures indicates that the convergence pattern is almost identical which means that minimizing the Euclidean distance of the estimated densities (Figure 14) is almost equivalent to minimizing the Euclidean distance of the estimated density and the true underlying density (Figures 15 and 16). Obviously, the Euclidean distance is smaller for values around the true smoothness parameters ($C_1 = 1/4$ and $\delta = 4$) in this context.

Although it is a useful tool, our adaptive procedure comes with two caveats. First, our Laplace deconvolution estimator assumes that the noise distribution is Laplace. If this assumption is violated, the adaptive procedure may not perform as well as we see here. Second, the Euclidean distance between the true density and the estimated density achieves small values in a range of smooth parameters rather than at one specific point in Figure 14. It indicates that the proposed rule-of-thumb adaptive procedure is informative for providing a small range of the smoothness parameters rather than one optimal point.

To calculate the RMISE when the smoothness parameters are unknown, we replicate the above simulations for $R = 100$ with various sample sizes and signal-to-noise ratios.[21] The results are presented in Table 5.[22] Similar to Table 3, the convergence pattern still holds

[20] The reader is reminded that the fit of the estimated densities, whether with or without known smoothness parameters, is a function of the Euclidean distance evaluated over the 128 points in their support. Therefore, the relative fit of the densities with known and unknown parameters will vary over this support. That is, we should not expect the density with known parameters to always have better fit than the estimated density with unknown parameters. This is reflected in Figure 13

[21] We reduce the replication size from 500 to save computation time.

[22] We report the RMISE of $\hat{f}_{known}$ here.

when the sample size increases with fixed signal-to-noise ratios. That is, reading down the columns, RMISE is decreasing in the sample size. As we read across RMISE columns within a row, the RMISE is decreasing slightly and then increasing. We also report the chosen smoothness parameters, δ and C_1, based on minimizing the Euclidean distance in Table 5. They vary slightly around 2 and 0.1, respectively. They are not always accurate (compared to the true values) but still render reasonably good estimates of the target density.

6 Applications

In this section two applications demonstrate the utility of the proposed method. We consider the parametric Laplace stochastic frontier model (Horrace and Parmeter, 2018), a regression-based application of the method, and a second application where the outcome of interest, daily saturated fat intake, is contaminated with measurement error (which we assume to be Laplace) and $\beta = 0$ in equation (1). In the first application we assume the smoothness parameters are known; in the second we use our adaptive rule-of-thumb to select them.

6.1 Stochastic Frontier Analysis

A typical parametric stochastic frontier model is equation (1), but restricting $u < 0$ (for a production frontier) or $u > 0$ (for a cost frontier). Given distributional assumptions on inefficiency, u (e.g., exponential or half-normal) and noise, v (e.g., normal or Laplace), β may be consistently estimated and used to calculate the conditional distribution of firm-level inefficiency, which is typically characterized by the empirical distribution of u conditional on ε (e.g., Jondrow et al. 1982). Much of the existing literature assumes normality of v (i.e., super-smooth v) and then applies maximum likelihood estimation (MLE). Relaxing parametric assumptions on the inefficiency distribution in these models is important, as articulated by Kneip, Simar, and Van Keilegom (2015, p.380) who note that "...there does usually not exist any information justifying particular distributional assumptions on (inefficiency)." Additionally, Tsionas (2017, p.1169) suggests that a model constructed to provide microfoundations for the presence of inefficiency "...does not make a prediction about the distribution." These statements underlie the importance of seeking alternative estimation approaches to recover important features of the stochastic frontier model; those approaches which eschew restrictive parametric assumptions are likely to curry favor among practitioners and regulators alike.

There is also no reason to favor normally distributed errors in the stochastic frontier model (Horrace and Parmeter, 2018). As such we apply our Laplace deconvolution estimator to estimate the distribution of inefficiency from a cost frontier for US banks. The data come from Feng and Serletis (2009) and are obtained from the Reports of Income and Condition (Call Reports).[23]

The data are a sample of US banks covering the period from 1998 to 2005 (inclusive). After deleting banks with negative or zero input prices, we are left with a balanced panel of 6,010 banks observed annually over the 8-year period. A more detailed description of the data may be found in Feng and Serletis (2009). For our purposes we ignore the panel structure of the data and choose the most recent year data, 2005, for our example. The goal of this exercise is to estimate the marginal distribution of u and compare it with the typical half-normal distribution which informs practical choice of parametric assumption on u, which , in turn, informs estimation of $E(u|\varepsilon)$.[24]

The data contain information on three output quantities and three input prices. The three outputs are consumer loans, Y_1; non-consumer loans, Y_2, which consists of industrial

[23]The data are publicly available on the Journal of Applied Econometrics data archive website http://qed.econ.queensu.ca/jae/2009-v24.1/feng-serletis/.

[24]Once $\hat{f}_u$ is obtained, one can estimate the efficiency score using numerical integration on a grid of $\hat{\varepsilon}$. To avoid an overloading of present paper, we stick to the estimation of marginal density of u.

and commercial loans and real estate loans; and securities, Y_3, including all non-loan financial and physical assets minus the sum of consumer loans, non-consumer loans, securities and equity. All outputs are deflated by the Consumer Price Index (CPI) to the base year of 1988. The three input prices are: the wage rate for the labor, P_1; the interest rate for borrowed funds, P_2 and the prices of physical capital, P_3. The total cost, C, is the sum of three corresponding input costs: total salaries and benefits, expenses on premises and equipment, and total interest expenses. Our specification of output and input prices is the same as (or very similar to) what is typical in the literature (see, for example, Feng and Serletis, 2009; Kumbhakar and Tsionas, 2005.) The cost frontier model is

$$c_j = \alpha + x_j'\beta + u_j + v_j \quad j = 1, \ldots, n, \tag{12}$$

where $c_j = \ln C_j$; $x_j = \ln X_j$ with X_j including the three output quantities and three input prices: Y_1, Y_2, Y_3, P_1, P_2, P_3; and $u_j > 0$ is firm-specific inefficiency.

We estimate the distribution of cost inefficiency in three ways. First, we estimate a fully parametric model, assuming v is distributed $N(0, \sigma_v^2)$ and u is distributed $|N(0, \sigma_u^2)|$. Our maximum likelihood estimates of the distributional parameters are $\hat{\sigma}_u = 1.294$ and $\hat{\sigma}_v = 0.989$, implying $E(u) = \hat{\sigma}_u\sqrt{2/\pi} = 1.033$. Then, our estimate of the density of u is $|N(0, 1.294^2)|$, which is shown as the dotted line (SFA) in Figure 18. Second, we estimate equation (12) by OLS. Figure 17 shows a histogram of the OLS residuals, $\hat{\varepsilon}_j$. The asymmetry of the distribution (skew equals 1.550) suggests non-zero cost inefficiency.[25] Selecting $\delta = 3$ and $C_1 = 1$ and using Theorem 1, the deconvolution estimator yields an estimate of σ_v^2 equal to 0.0403.[26] A plot of the density estimate, $\hat{f}_u(u)$, is shown as the dashed line (CHP) in

[25] It is interesting to note that with a skew of 1.55, this provides evidence against use of the half-normal distribution.

[26] For the Laplace distribution, $\delta = 2$; for convolved Laplace, $\delta = 4$. The choice $\delta = 3$ is between Laplace and convolved Laplace.

Figure 18. Third, using the procedure of Hall and Simar (2002) with a bandwidth of 0.3052, we detect a jump discontinuity point in $\hat{f}_u(u)$ at $u = -0.355$ which implies an estimate of $\hat{E}(u) = 0.355$. Then using the boundary kernel proposed by Zhang and Karunamuni (2000), with an estimated error variance of 0.0403 (as before), the boundary bias corrected density estimate is shown as the solid line (CHP_$E(u)$_bc) in Figure 18.[27]

Figure 18 shows all three density estimators for US bank inefficiency in 2005. Notice that even without a boundary correction, the deconvolution estimator (CHP) has a thinner right tail than the estimated half normal density (SFA). With boundary correction in place, the deconvolution estimator (CHP_$E(u)$_bc) implies that US banks in 2005 have a much smaller average inefficiency than parametric SFA would have predicted. This corresponds to the fact that in 1998 there are 10,139 banks in the US and this number declined to 8,390 in 2005 due to industry consolidation (Feng and Serletis, 2009).

Finally, there are at least two reasons to employ the proposed estimator: 1) the proposed method provides a robustness check for the distributional assumptions made in a parametric stochastic frontier model and 2) the skewness of the OLS residuals is greater than one, which invalidates the choice of the half-normal assumption for the distribution of u (which has maximal skewness of 1 by definition).

6.2 Daily Saturated Fat Intake With Measurement Errors

The data come from Wave III (1988-1994) of the National Health and Nutrition Examination Survey, abbreviated NHANES III. Our interest is the survey response to daily saturated fat intake of 3,551 women between the ages of 25 and 50. This data set is ideally suited to our Laplace deconvolution estimator as it is well established that saturated fat consumption is recorded with measurement errors. In fact, previous analysis of the NHANES Wave I

[27] For Laplace deconvolution, we can apply directly Example 1 in Zhang and Karunamuni (2000).

(1971-75) and Wave II (1976-1980) data suggest that more than 50% of the variability in the observed data may be due to measurement errors. See Stefanski and Carroll (1990), Carroll, Ruppert and Stefanski (2006) and Delaigle and Gijbels (2004).

The data were originally recorded to explore the relationship between breast cancer and dietary fat intake, see Jones et al. (1987). Stefanski and Carroll (1990) were the first to consider nonparametric deconvolution techniques to estimate the underlying true density of saturated fat intake, using NHANES I. Subsequently, Carroll, Ruppert and Stefanski (2006), Delaigle and Gijbels (2004) and others applied deconvolution estimators to NHANES II. In each of these applications a normal error distribution was assumed. To the best of our knowledge we are the first to apply deconvolution techniques to NHANES III (and certainly the first to apply Laplace deconvolution to any of these data). Here, saturated fat (fat) is measured in milligrams per day, and we apply the same data transformation as Delaigle and Gijbels (2004): $log(fat + 5)$.

To these data we implement a) the proposed estimator with Laplace errors (CHP), b) the estimator with normal errors due to Meister (2006) ($Meister$), and c) an error free estimator ($ErrorFree$), based on pure kernel density estimation of the observed data assuming there is no measurement error.[28]

First, we apply the proposed rule-of-thumb adaptive procedure to get a preliminary estimate of the smoothness parameters since they are unknown. Specifically, we search for the minimum of the Euclidean distance between the density estimator with unknown smoothness parameters and density estimator with known smoothness parameter, Δ, over a grid of $\delta \in \{1.25, 1.5, 1.75, 2, 2.25, 2.5, 2.75, 3\}$ and $C_1 \in \{0.1, 0.25, 0.40, 0.55, 0.70, 0.85, 1\}$.[29] Figure 19 shows the surface of the Euclidean distance as a function of the smoothness parameters over the grid. The Δ increases as C_1 rises from 0 to 1 except when δ is around

[28] We use the package "ksdensity" in Matlab for the $ErrorFree$ case.

[29] We also tried larger range of δ and narrow down to this specific range by searching the minimum of Δ.

2. It seems that $\delta = 1.5$ and $\delta = 3$ yield the minimum distance. It turns out that when $\delta = 3$, the estimated density decreases very quickly and goes below zero and becomes volatile when $log(fat + 5) < 2$ or $log(fat + 5) > 4.5$. Therefore, we consider the $\delta = 1.5$ case to be optimal. Specifically, we choose $C_1 = 1$ and $\delta = 1.5$ as our baseline model. We then consider alternative specifications of the smoothness parameters as a robustness check.

Figure 20 presents the final results of the analysis. The estimated error variance is 0.065 based on the CHP estimator and 0.525 based on the $Meister$ estimator in the baseline model. The $Meister$ error variance estimate is exceedingly large compared to the variance of the observed (convoluted) data, 0.236.[30] The CHP error variance estimate is more reasonable in the sense of being less than the total observed variance, and its corresponding signal-to-noise ratio is 0.275. This is consistent with the finding in the existing literature that about 30-50% of the variability of observed data is due to measurement error. The tail behaviors in Figure 20 shows that the $Meister$ estimator assigns more variance to the error variance than expected and it decreases to zero very quickly. The CHP estimator extracts the target density information based on the smoothness assumptions, which gives a reasonable variance estimate and tends to have longer tails.[31]

The CHP density estimator based on the NHANES III data is quite similar to that of Delaigle and Gijbels (2004), despite the fact that they used the NHANES II data, assumed the error to be normal, along with differing identification assumptions. They experiment with different "known" values of the signal-to-noise ratio, while we have to select the smoothness parameters. The minor difference is that our estimated tails are slightly thicker than theirs, however the means of the estimated densities are nearly identical.

As a robustness check, different combinations for the values of δ and C_1 are considered for the CHP estimator: $C_1 = 1$ and $\delta = 1.5$; $C_1 = 1$ and $\delta = 2$; $C_1 = 0.6$ and $\delta = 1.5$;

[30]It seems to violate the independence assumption between the target variable and the measurement error.

[31]Under Assumption 1, i.e., u and v are independent, the variance of Y should be the sum of the variances of u and v. Empirically, this may not be the case for real data.

$C_1 = 0.6$ and $\delta = 2$ in Figure 21. The baseline ($C_1 = 1$, $\delta = 1.5$) is in the upper-left panel of the figure. As we move to different panels in the figures we change the values of the smoothness parameters, so the CHP estimator is changing across panels, while the $ErrorFree$ estimator is fixed. For $C_1 = 0.6$, $\delta = 1.5$ (lower-left panel), the estimated error variance of CHP is 0.019 which is less than the baseline model, and it has less fat tails. For $C_1 = 1$, $\delta = 2$ (upper-right panel), the estimated error variance of CHP is 0 which makes it nearly coincide with the $ErrorFree$ case.[32] This means that it is more difficult for information on the measurement error to be be disentangled under these smoothness assumptions. We can also vary C_1 to recover certain information concerning the noise or the error term. For instance, $C_1 = 0.6$, $\delta = 2$ (lower-right panel), the estimated error variance of CHP is still 0 which renders an identical deconvolution density estimate. It seems that the variability of δ dominates that of C_1. This is intuitive as $\tau \to \infty$, the effect of C_1 is ignorable in Assumption 3.

7 Conclusion

This paper proposes a semiparametric estimator for a cross-sectional error component model. Instead of focusing on the estimation of the model parameters with the typical assumption of normality, we are interested in the density of the target error component. To estimate the target density without fully known random noise, we modify the variance truncation device proposed by Meister (2006) and extend the methodology to the framework of an error component model with a Laplace noise term with unknown variance.

The density deconvolution estimator with Laplace noise has at least two attractive characteristics for applied researchers: 1) it possesses a faster convergence rate than that of

[32]One point worth mentioning is that these minimax deconvolution techniques can produce error variance estimates equal to zero as we vary the choice of C_1 and δ. Recall that $\hat{b}_n^2$ is bound between 0 and $0.5V(\hat{\varepsilon})$. When it happens, the deconvolution estimators will be very similar to the $ErrorFree$ estimator.

normal distributed noise (i.e., $O(n^c)$ versus $O((\ln n)^c)$) and 2) it is robust to misspecification of the true underlying noise distribution. A third (potential) feature that practitioners may find appealing is the Laplace noise generates different insights than normal noise: for example, the LAD estimator rather than OLS, the Laplace stochastic frontier model (Horrace and Parmeter, 2018) and the L-SIMEX estimator (Koul and Song, 2014).

For future research, it may be useful to extend the model to panel data and use it to estimate both the target and noise distributions nonparametrically. For example, with a nonparametric production or cost function this would imply a fully nonparametric stochastic frontier model. Jirak, Meister and Reiss (2014) studied adaptive function estimation in nonparametric regression with one-sided errors. Another interesting strand in this area is to investigate the distribution of unobserved heterogeneity with proposed deconvolution techniques. Recently, Evdokimov (2010) takes an initial step to explore that in a panel data model and Ju, Gan and Li (2019) apply it to a labor data set.

References

[1] Butucea, C. and C. Matias, Minimax Estimation of the Noise Level and of the Deconvolution Density in a Semiparametric Convolution Model. Bernoulli. 2005, pp. 309-340.

[2] Butucea, C., C. Matias, and C. Pouet, Adaptivity in Convolution Models with Partially Known Noise Distribution. Electronic Journal of Statistics. 2008, pp. 897-915.

[3] Cao, C. D. Linear Regression with Laplace Measurement Errors. Master Thesis. Department of Statistic, Kansas State University.

[4] Carroll, R.J., D. Ruppert, L. A. Stefanski, and C. M., Crainiceanu. Measurement Error in Nonlinear Models: A Modern Perspective, volume 2nd Ed. CRC Press, 2006.

[5] Delaigle, A. and I. Gijbels. Practical bandwidth selection in deconvolution kernel density estimation. Computational Statistics & Data Analysis. 2004. 45 (2) pp. 249-267.

[6] Evdokimov, K., Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity. 2010, Working paper version.

[7] Fan, J., On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems. Annals of Statistics. 1991, pp. 1257-1272.

[8] Feng, G. H., and A. Serletis, Efficiency and Productivity of the US Banking Industry,1998-2005: Evidence from the Fourier Cost Function Satisfying Global Regularity Conditions. Journal of Applied Econometrics. 2009, 24, pp.105-138.

[9] Greene, W. H., A gamma-distributed stochastic frontier model, Journal of Econometrics. 1990, 46(1-2), 141-164.

[10] Greene, W. H., Simulated likelihood estimation of the normal-gamma stochastic frontier function, Journal of Productivity Analysis. 2003, 19(2), 179-190.

[11] Henderson, D. J. and C. F. Parmeter. Applied Nonparametric Econometrics. 2015, Cambridge University Press.

[12] Hesse, C. H., Data-driven Deconvolution. Nonparametric Statistics. 1999, 10, pp. 343-373.

[13] Horowitz, J. L., and M. Markatou, Semiparametric Estimation of Regression Models for Panel Data. Review of Economics Studies. 1996, 63, pp. 145-168.

[14] Horrace, W. C., Moments of the Truncated Normal Distribution. Journal of Productivity Analysis. 2015, 43, pp. 133-138.

[15] Horrace, W. C., and C.F. Parmeter, Semiparametric Deconvolution with Unknown Error Variance. Journal of Productivity Analysis. 2011, 35, pp. 129-141.

[16] Horrace, W. C., and C.F. Parmeter, A Laplace Stochastic Frontier Model. Econometric Reviews. 2018, 37, pp. 260-280.

[17] Jirak, M., A. Meister, and M. Reiss. Adaptive function estimation in nonparametric regression with one-sided errors. Annals of Statistics. 2014, 42(5), pp. 1970-2002.

[18] Jondrow, J., C. K. Lovell, I. S. Materov, and P. Schmidt, On the Estimation of Technical Inefficiency in the Stochastic Frontier Model. Journal of Econometrics. 1982, 19, pp. 233-238.

[19] Jones, D.Y., et al. Dietary fat and breast cancer in the National Health and Nutrition Examination Survey I Epidemiologic Follow-up Study. Journal of the National Cancer Institute. 1987. 79(3), pp. 465-71.

[20] Johannes, J., Deconvolution with Unknown Error Distribution. Annals of Statistics. 2009, 37, pp. 2301-2323.

[21] Ju, G., L. Gan, and Q. Li. Nonparametric panel estimation of labor supply. Journal of Business & Economic Statistics. 2019, 37(2), pp. 260-274.

[22] Koul, H.L., and W.X. Song. Simulation Extrapolation Estimation in Parametric Models with Laplace Measurement Error. Electronic Journal of Statistics. 2014, 8, pp. 1973-1995.

[23] Kneip, A., L. Simar, and I. Van Keilegom. Frontier estimation in the presence of measurement error with unknown variance. Journal of Econometrics. 2015, 184(2), pp. 379-393.

[24] Kumbhakar, S. C. and E. G. Tsionas, Measuring Technical and Allocative Inefficiency in the Translog Cost System: A Bayesian Approach. Journal of Econometrics. 2005, 126, pp. 355-384.

[25] Kumbhakar S. C., B.U. Park, L. Simar, E.G. Tsionas. Nonparametric stochastic frontiers: a local likelihood approach. Journal of Econometrics. 2007, 137(1), pp. 1-27.

[26] Kumbhakar, S.C., C.F. Parmeter, and V. Zelenyuk. Stochastic Frontier Analysis: Foundations and Advances. CEPA Working Papers Series WP022018, School of Economics, University of Queensland, Australia.

[27] Li, Q. and J.S. Racine. Nonparametric Econometrics: Theory and Practice. 2007, Princeton University Press.

[28] Meeusen, W. and J. van den Broeck. Efficiency estimation from Cobb-Douglas production functions with composed error, International Economic Review. 1977, 18(2), pp. 435-444.

[29] Meister, A., On the Effect of Misspecifying the Error Density in a Nonparametric Deconvolution Problem. Canadian Journal of Statistics. 2004, 32, pp. 439-449.

[30] Meister, A., Density Estimation with Normal Measurement Error with Unknown Variance. Statistica Sinica. 2006, 16, pp. 195-211.

[31] Meister, A., Deconvolution Problems in Nonparametric Statistics. 2009. Lecture Notes in Statistics 193. Springer, Berlin

[32] Neumann, M. H., On the effect of Estimating the Error Density in Nonparametric Deconvolution. Nonparametric Statistics. 1997, 7, pp. 307-330.

[33] Nguyen, N. B., Estimation of technical efficiency in stochastic frontier analysis. 2010, PhD thesis, Bowling Green State University.

[34] Parmeter, C. F., H. J. Wang, and S. C. Kumbhakar. Nonparametric estimation of the determinants of inefficiency. Journal of Productivity Analysis. 2017, 47, pp. 205-221.

[35] Parzen, E., On Estimation of A Probability Density Function and Mode. Annals of Math Statistics. 1962, 35, pp. 1065-1076.

[36] Ritter, C. and L. Simar. Pitfalls of normal-gamma stochastic frontier models. Journal of Productivity Analysis. 1997, 8(2), pp. 167-182.

[37] Simar, L., I. van Keilegom, and V. Zelenyuk. Nonparametric least squares methods for stochastic frontier models. Journal of Productivity Analysis. 2017, 47, pp. 189-204.

[38] Song, W.X., J.H. Shi, and C.X. Zhang. Linear errors-in-variables regression and Tweedie's formula. 2016. Working paper

[39] Stevenson, R. Likelihood functions for generalized stochastic frontier estimation. Journal of Econometrics. 1980, 13(1), pp. 58-66.

[40] Tran, K. C. and E. G. Tsionas. Local GMM estimation of semiparametric panel data with smooth coefficient models. Econometric Reviews. 2010, 29, pp. 39-61.

[41] Wang, W. S., C. Amsler, and P. Schmidt, Goodness of fit tests in stochastic frontier models. Journal of Productivity Analysis. 2011, 35, pp. 95-118.

[42] Wang, W. S. and P. Schmidt, On the Distribution of Estimated Technical Efficiency in Stochastic Frontier Models. Journal of Econometrics. 2009, 148, pp. 36-45.

[43] Wang, X. F. and D. Ye, The Effects of Error Magnitude and Bandwidth Selection for Deconvolution with Unknown Error Distribution. Journal of Nonparametric Statistics. 2012, 24, pp. 153-167.

[44] Zhang, S. and R. J. Karunramuni, Boundary Bias correction for Nonparametric Deconvolution. Annals of Institute Statistic Mathematics. 2000, 52, pp. 612-629.

A General Appendix

Definition: ε is ordinary-smooth of order δ (Fan 1991): characteristic function $\phi_\varepsilon(t)$ satisfies $d_0|t|^{-\delta} \leq |\phi_\varepsilon(t)| \leq d_1|t|^{-\delta}$ as $t \to \infty$.This is literally the same with the Assumption 3, just replacing $\phi_\varepsilon(t)$ with $h_\varepsilon(\tau)$.

A generalized result of Parseval's identity (or the Plancherel theorem) asserts that the integral of the square of the Fourier transform of a function is equal to the integral of the square of the function itself.

In one-dimension, for $f \in L_2(R)$,

$$\int_{-\infty}^{\infty} |\hat{f}(z)|^2 dz = \int_{-\infty}^{\infty} |f(\tau)|^2 d\tau$$

where $\hat{f}(z) = \int_{-\infty}^{\infty} e^{-i\tau z} f(\tau) d\tau$ is the Fourier transform of the function $f(\tau)$.

B Proof of Lemma 1

There is a N so that $w_n > T$ holds for all $n \geq N$. Hence the upper and lower bound of the Fourier Transform can be used. Similar to Lemma 1 in Meister(2006), using Parseval's identity and Fubini's theorem, we have:

$$\sup_{g \in \mathcal{L}_n} \sup_{f \in \mathcal{F}_u} E_{f,g} ||\hat{f}_u - f_u||_{L_2}^2 = \big(2\pi\big)^{-1} \sup_{g \in \mathcal{L}_n} \sup_{f \in \mathcal{F}_u} \Big(\int_{-w_n}^{w_n} E_{f,g} |e^{-i\tau z} \big(\hat{h}_{\hat{\varepsilon}}(1 + \hat{b}_n^2 \tau^2) - h_u(\tau)\big)|^2 d\tau +$$

$$\int_{|\tau| > w_n} |e^{-i\tau z} h_u(\tau)|^2 d\tau \Big)$$

$$\overset{Parseval}{=} (2\pi)^{-1} \sup_{g\in\mathcal{L}_n} \sup_{f\in\mathcal{F}_u} \Big(\int_{-w_n}^{w_n} E_{f,g}|\hat{h}_{\hat{\varepsilon}}(1+\hat{b}_n^2\tau^2) - h_u(\tau)|^2 d\tau + \int_{|\tau|>w_n} |h_u(\tau)|^2 d\tau \Big)$$

$$\leq (2\pi)^{-1} \Big(\sup_{g\in\mathcal{L}_n} \sup_{f\in\mathcal{F}_u} 2 \int_{w_n}^{\infty} |h_u(\tau)|^2 d\tau + \sup_{g\in\mathcal{L}_n} \sup_{f\in\mathcal{F}_u} 2 \int_{-w_n}^{w_n} E_{f,g}|(1+\hat{b}_n^2\tau^2)(\hat{h}_{\hat{\varepsilon}}(\tau) - h_\varepsilon(\tau))|^2 d\tau +$$

$$\sup_{g\in\mathcal{L}_n} \sup_{f\in\mathcal{F}_u} 2 \int_{-w_n}^{w_n} E_{f,g}|h_\varepsilon(\tau)/\frac{1}{(1+\hat{b}_n^2\tau^2)} - h_u(\tau)|^2 d\tau \Big)$$

The first term, which we call B, represents the bias which does not depend on the fact that the convoluted errors are estimated and can be bounded as in Lemma 1 of Meister (2006). The second term can be split into two pieces, V_1 and V_2, where V_1 is similar to V in Lemma 1 of Meister (2006) while V_2 is an additional component of the variance due to estimating the composed errors. Our third term, which we call E, can be found almost as that in Lemma 1 of Meister (2006) but the form of the bound is more complicated due to the fact that the empirical characteristic function used to construct the variance of the Laplace noise is constructed with $\hat{\varepsilon}$ instead of ε. The nonparametric regression in the first step impacts the convergence rate through the estimation of $\hat{\varepsilon}$.

The following proof is similar to Meister (2006) and Horrace and Parmeter (2011) except now we deal with Laplace noise and a nonparametric first-step regression estimator rather than just normal noise for the linear (stochastic frontier) model. There are three steps to the proof.

(1) $B \leq const \times w_n^{1-2\delta}$ by Assumption(3) $C_1|\tau|^{-\delta} \leq |h_u(\tau)| \leq C_2|\tau|^{-\delta}$ where $0 < C_1 < C_2$ and $\delta > 1$.

(2) By assumption 5,

$$\hat{h}_{\hat{\varepsilon}}(\tau) = \Big|\frac{1}{n}\sum_{j=1}^{n} e^{i\tau\hat{\varepsilon}_j}\Big| = \Big|\frac{1}{n}\sum_{j=1}^{n} e^{i\tau\varepsilon_j}(1+O_p(\tau n^{-a}))\Big| = (1+O_p(\tau n^{-a}))|\hat{h}_{\varepsilon}(\tau)|$$

where $a = \frac{2}{4+q}$ for the nonparametric first-step regression and $a = 0.5$ for parametric first-step regression, e.g, translog in the stochastic frontier model. We focus on the parametric setting hereafter for the main formulas and lay out the details of the differences when first-step nonparametric regression is implemented.[33] So

$$\hat{h}_{\hat{\varepsilon}}(\tau) = |\frac{1}{n}\sum_{j=1}^{n} e^{i\tau\hat{\varepsilon}_j}| = |\frac{1}{n}\sum_{j=1}^{n} e^{i\tau\varepsilon_j}(1+O_p(\tau n^{-1/2}))| = (1+O_p(\tau n^{-1/2}))|\hat{h}_{\varepsilon}(\tau)|$$

Let $A(\hat{h}_\varepsilon) = \int_{-w_n}^{w_n} E_{f,g}|\hat{h}_\varepsilon(\tau) - h_\varepsilon(\tau)|^2 d\tau = \int_{-w_n}^{w_n} E_{f,g}|\frac{1}{n}\sum_{j=1}^{n} e^{i\tau\varepsilon_j} - E(e^{i\tau\varepsilon})|^2 d\tau = O_p(n^{-1}w_n)$,

$$\sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u} 2\int_{-w_n}^{w_n} E_{f,g}(1+\hat{b}_n^2\tau^2)^2|\hat{h}_{\hat{\varepsilon}}(\tau) - h_\varepsilon(\tau)|^2 d\tau \le 4(1+\hat{b}_n^2 w_n^2)^2 \sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u}\int_{-w_n}^{w_n}[E_{f,g}|\hat{h}_{\hat{\varepsilon}}(\tau) - \hat{h}_\varepsilon(\tau)|^2 +$$

$$E_{f,g}|\hat{h}_\varepsilon(\tau) - h_\varepsilon(\tau)|^2]d\tau$$

$$= 4(1+\hat{b}_n^2 w_n^2)^2 \sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u}\int_{-w_n}^{w_n} E_{f,g}|\hat{h}_{\hat{\varepsilon}}(\tau) - \hat{h}_\varepsilon(\tau)|^2 d\tau + 4(1+\hat{b}_n^2 w_n^2)^2 A(\hat{h}_\varepsilon)$$

$$\le 4(1+b_n^2 w_n^2)^2 \sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u}\int_{-w_n}^{w_n} \tau^2 E_{f,g}(n^{-1}\sum_{j=1}^{n}|\hat{\varepsilon}_j - \varepsilon_j|)^2 d\tau + 4(1+b_n^2 w_n^2)^2 A(\hat{h}_\varepsilon) = V_1 + V_2$$

[33] Basically, there is $2a$ instead of 1 in the power of n.

where $V_1 \leq const \times (n^{-1}w_n^3)(1 + b_n^2 w_n^2)^2$ and $V_2 \leq const \times (n^{-1}w_n)(1 + b_n^2 w_n^2)^2$.

(3) Similar to Lemma 1 in Meister (2006), for the last term we can derive:

$$E = \sup_{g \in \mathcal{L}_n} \sup_{f \in \mathcal{F}_u} 2 \int_{-w_n}^{w_n} E_{f,g} \left| \frac{h_u(\tau)(1 + \hat{b}_n^2 \tau^2)}{1 + b^2 \tau^2} - h_u(\tau) \right|^2 d\tau$$

$$\stackrel{\tau = s w_n}{=} \sup_{g \in \mathcal{L}_n} \sup_{f \in \mathcal{F}_u} 2 \int_{-1}^{1} E_{f,g} \left| \frac{s^2 w_n^2 (\hat{b}_n^2 - b^2)}{1 + b^2 s^2 w_n^2} \right|^2 |h_u(s w_n)|^2 w_n ds,$$

where

$$\begin{aligned}
E_{f,g} \left| \frac{s^2 w_n^2 (\hat{b}_n^2 - b^2)}{1 + b^2 s^2 w_n^2} \right|^2 &= E_{f,g} \left| \frac{s^2 w_n^2 (\hat{b}_n^2 - b^2)}{1 + b^2 s^2 w_n^2} \right|^2 \chi(|\hat{b}_n^2 - b^2| \leq d_n) + E_{f,g} \left| \frac{s^2 w_n^2 (\hat{b}_n^2 - b^2)}{1 + b^2 s^2 w_n^2} \right|^2 \chi(|\hat{b}_n^2 - b^2| > d_n) \\
&\leq \left| \frac{s^2 w_n^2 d_n}{1 + s^2 w_n^2 b^2} \right|^2 + \left| \frac{s^2 w_n^2 b_n^2}{1 + b^2 w_n^2 s^2} \right|^2 Pr(|\hat{b}_n^2 - b^2| > d_n) \\
&\leq \left(\frac{s^2 w_n^2 d_n}{s^2 w_n^2 b^2} \right)^2 + \left(\frac{s^2 w_n^2 b_n^2}{s^2 b^2 w_n^2} \right)^2 Pr(|\hat{b}_n^2 - b^2| > d_n) \\
&\leq \left(\frac{d_n}{b^2} \right)^2 + \left(\frac{b_n^2}{b^2} \right)^2 Pr(|\hat{b}_n^2 - b^2| > d_n) \\
&= const \times w_n^{-2} + const \times (b_n^2)^2 Pr(|\hat{b}_n^2 - b^2| > d_n)
\end{aligned}$$

Where the last inequality for the first term comes from the fact that $d_n = O(w_n^{-1})$.

C Proof of Lemma 2

Let d_n and f, g be the same as in Lemma 1, the term $\sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u} P_{f,g}(|\hat{b}_n^2 - b^2| > d_n)$ is bounded by two addends. We derive an upper bound for each of them. First,

$$\begin{aligned}
\sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u} P_{f,g}(\hat{b}_n^2 - b^2 > d_n) &= \sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u} P_{f,g}(k_n^{-2}(\frac{C_1 k_n^{-\delta}}{\hat{h}_{\hat{\varepsilon}}(k_n)} - 1) > d_n + b^2) \\
&= \sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u} P_{f,g}((\frac{C_1 k_n^{-\delta}}{\hat{h}_{\hat{\varepsilon}}(k_n)} - 1) > d_n k_n^2 + b^2 k_n^2) \\
&= P_{f,g}(|\hat{h}_{\hat{\varepsilon}}(k_n)| < \frac{C_1 k_n^{-\delta}}{1 + d_n k_n^2 + b^2 k_n^2}) \\
&\leq \sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u} P_{f,g}(|\hat{h}_{\hat{\varepsilon}}(k_n)| < \alpha_n \frac{C_1 k_n^{-\delta}}{1 + b^2 k_n^2}) \\
&= \sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u} P_{f,g}(\hat{h}_{\hat{\varepsilon}}(k_n) < \alpha_n |h_\varepsilon(k_n)|)
\end{aligned}$$

where $\alpha_n = \frac{1+b^2k_n^2}{1+d_nk_n^2+b^2k_n^2}$, hence, $\alpha_n \to 0$ as $d_n = w_n^{-1} = O(k_n^{-1})$, $d_n k_n^2 = O(k_n)$ for known δ and C_1 case and $d_n = w_n^{-1} = O(lnk_n/k_n), d_n k_n^2 = O(ln(k_n)k_n)$ for other cases.[34]

There exists a constant $c \in (0, 1)$ that guarantees that the above formula is bounded above by $\sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u} P_{f,g}(\hat{h}_{\hat{\varepsilon}}(k_n) < \alpha_n |h_\varepsilon(k_n)|) \leq \sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u} P_{f,g}(\hat{h}_{\hat{\varepsilon}}(k_n) < c|h_\varepsilon(k_n)|)$ which by Chebyshev's inequality yields

$$\begin{aligned}
&\leq (1-c)^{-2} \sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u} |h_\varepsilon(k_n)|^{-2} E_\varepsilon |\hat{h}_{\hat{\varepsilon}}(k_n) - h_\varepsilon(k_n)|^2 \\
&\leq 2(1-c)^{-2} \sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u} |h_\varepsilon(k_n)|^{-2} [E_\varepsilon |\hat{h}_{\hat{\varepsilon}}(k_n) - \hat{h}_\varepsilon(k_n)|^2 + E_\varepsilon |\hat{h}_\varepsilon(k_n) - h_\varepsilon(k_n)|^2] \\
&\leq 2(1-c)^{-2} \sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u} |h_\varepsilon(k_n)|^{-2} [E_\varepsilon |O_p(k_n n^{-1}) \frac{1}{n} \sum_j exp(ik_n\varepsilon_j)|^2 + E_\varepsilon |\frac{1}{n} \sum_j exp(ik_n\varepsilon_j) - h_\varepsilon(k_n)|^2] \\
&= const \times (E_1 + E_2),
\end{aligned}$$

[34] This is discussed in Section 4.

where the first term is bounded by $|h_\varepsilon(k_n)|^{-2} \leq k_n^{2\delta+2}(1+b_n^2k_n^2)$ as that for V_1; $E_1 \leq const \times k_n^{2\delta+2}(1+b_n^2k_n^2)n^{-1}$ and $E_2 \leq const \times k_n^{2\delta}(1+b_n^2k_n^2)n^{-1}$ are similar to that in Lemma 2 of Meister (2006).

The second addend can be bounded in a similar way:

$$\sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u} P_{f,g}(\hat{b}_n^2 - b^2 < -d_n) = \sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u} P_{f,g}(k_n^{-2}(\frac{C_1k_n^{-\delta}}{\hat{h}_{\hat{\varepsilon}}(k_n)} - 1) < b^2 - d_n)$$
$$\leq \sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u} P_{f,g}(\hat{h}_{\hat{\varepsilon}}(k_n) > \gamma_n|h_\varepsilon(k_n)|)$$

where $\gamma_n = \frac{1+b^2k_n^2}{1+b^2k_n^2-d_nk_n^2}$, hence, $\gamma_n \to 1^+$ as $d_n = w_n^{-1} = O(k_n^{-1})$, $d_nk_n^2 = O(k_n)$ for known δ and C_1 case and $d_n = w_n^{-1} = O(lnk_n/k_n), d_nk_n^2 = O(ln(k_n)k_n)$ for other cases.

Again there exists a constant $C \in (0,1)$ that guarantees the above formula is bounded above by $\sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u} P_{f,g}(\hat{h}_{\hat{\varepsilon}}(k_n) > \gamma_n|h_\varepsilon(k_n)|) \leq \sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u} P_{f,g}(\hat{h}_{\hat{\varepsilon}}(k_n) > C|h_\varepsilon(k_n)|)$ which by Chebyshev's inequality yields

$$\leq (C-1)^{-2}\sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u}|h_\varepsilon(k_n)|^{-2}E_\varepsilon|\hat{h}_{\hat{\varepsilon}}(k_n) - h_\varepsilon(k_n)|^2$$
$$\leq 2(C-1)^{-2}\sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u}|h_\varepsilon(k_n)|^{-2}[E_\varepsilon|\hat{h}_{\hat{\varepsilon}}(k_n) - \hat{h}_\varepsilon(k_n)|^2 + E_\varepsilon|\hat{h}_\varepsilon(k_n) - h_\varepsilon(k_n)|^2],$$

which leads to the same upper bound as derived for the first addend.

D Proof of Theorem 1

Combining the results from Lemma 1 and 2, we obtain the upper bound of MISE of the density $\hat{f}_u$ as

$$\max(B, V, E) = \max\Big[const_1 \times w_n^{1-2\delta}, const_2 \times n^{-1} w_n (1 + b^2 w_n^2)^2 + const_3 \times n^{-1}(1 + b_n^2 w_n^2)^2 w_n^3,$$
$$const_4 \times w_n^{1-2\delta} w_n^{-2} + const_5 \times k_n^{2\delta} b_n^4 (1 + b_n^2 k_n^2)(1 + k_n^2) n^{-1}\Big]$$

Under Assumption 3, if C_1 and δ are known, then $w_n = k_n$, $b_n^2 = 0.5V(\hat{\varepsilon}\,)$, and minimizing the above maximum leads to

(i) If $1 < \delta \leq 1.5$, $k_n = n^{\frac{1}{2\delta+6}}(b_n^2)^{\frac{-1}{2\delta+6}}$. Consequently, a $n^{-\frac{(2\delta-1)}{2\delta+6}}(b_n^2)^{\frac{(2\delta-1)}{2\delta+6}} \to D_1 n^{-\frac{(2\delta-1)}{2\delta+6}}$ convergence rate is determined by the equality of the first term and the first addend of the third term where $D_1 = V(\varepsilon)^{\frac{(2\delta-1)}{2\delta+6}}$.

(ii) If $\delta > 1.5$, $k_n = n^{\frac{1}{4\delta+3}}(b_n^2)^{\frac{-4}{4\delta+3}}$. Consequently, a $n^{-\frac{2\delta-1}{4\delta+3}}(b_n^2)^{\frac{(2\delta-1)}{4\delta+3}} \to D_2 n^{-\frac{2\delta-1}{4\delta+3}}$ convergence rate is determined by the equality of the first term and the second addend of the third term where $D_2 = V(\varepsilon)^{\frac{(2\delta-1)}{4\delta+3}}$. We exclude the case with $\delta = 2$ (Laplace-Laplace convolution) here as $\hat{b}_n^2 < b_n^2 = 0.5V(\hat{\varepsilon})$ converges to $\min\{V(u), V(v)\}$ which cannot distinguish the target from the noise[35].

E Proof of Theorem 2

For the case where C_1 and δ are unknown, similar argument applies, k_n stays the same with guess C_1 and δ since $w_n = k_n / \ln k_n = O(k_n)$ and the convergence rates are $n^{-\frac{(2\delta-1)}{2\delta+6}}(\ln n)^{\frac{(2\delta-1)}{2\delta+6}}$ if $1 < \delta \leq 1.5$ and $n^{-\frac{2\delta-1}{4\delta+3}}(\ln n)^{\frac{2\delta-1}{4\delta+3}}$ if $\delta > 1.5$.[36]

[35]This is a rare case related to identification given that $\delta = 2$ is negligible in the range of $\delta > 1$ but it does not impact the estimation

[36]See the rule-of-thumb adaptive procedure in section 4.1.

F Proof of Theorem 3

When nonparametric kernel estimation is implemented for the first-step regression, we can easily derive similar Lemmas (as those for the parametric case) as follows:

Lemma $1'$**.** *For Assumptions 3-5, Condition 2.1 in Li and Racine (2007) and* $\mathcal{L}_n = \{Laplace(0,b) : b^2 \in (0, b_n^2]\}$*, the MISE of (7) is*

$$\sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u} E_{f,g}||\hat{f}_u - f_u||_{L_2}^2 \leq B + V + E,$$

where $B \leq const_1 \times w_n^{1-2\delta}$*,*

$V \leq const_2 \times n^{-2a} w_n (1 + b_n^2 w_n^2)^2 + const_3 \times n^{-2a} w_n^3 (1 + b_n^2 w_n^2)^2$ *,*

$E \leq const_4 \times \sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u}\Big(w_n \int_{-1}^{1} |h_u(sw_n)|^2 \left(\frac{d_n}{b^2}\right)^2 ds + w_n \int_{-1}^{1} |h_u(w_n s)|^2 \frac{b_n^4}{b^4} \times P_{f,g}(|\hat{b}_n^2 - b^2| > d_n) ds\Big)$*, with* $d_n := \frac{1}{w_n}$*;* f *and* g *are the probability density function in distribution family* $\mathcal{F}_u$ *and* $\mathcal{L}_n$ *respectively.* $const_j$ *are positive constants for* $j = 1, 2, 3, 4$*.*

Lemma $2'$**.** *Let* d_n *and* f, g *be the same as in Lemma 3, then* $\sup_{g\in\mathcal{L}_n}\sup_{f\in\mathcal{F}_u} P_{f,g}(|\hat{b}_n^2 - b^2| > d_n) \leq const \times k_n^{2\delta}(1 + b_n^2 k_n^2)(1 + k_n^2) n^{-2a}$*.*

Then by a parallel argument, combining the results from Lemma 1 and 2, we can obtain the upper bound of MISE of the density $\hat{f}_u$ as

$$\max(B, V, E) = \max\Big[const_1 \times w_n^{1-2\delta}, const_2 \times n^{-2a} w_n (1 + b^2 w_n^2)^2 + const_3 \times n^{-2a}(1 + b_n^2 w_n^2)^2 w_n^3,$$
$$const_4 \times w_n^{1-2\delta} w_n^{-2} + const_5 \times k_n^{2\delta} b_n^4 (1 + b_n^2 k_n^2)(1 + k_n^2) n^{-2a}\Big]$$

Under Assumption 3, if C_1 and δ are known, then $w_n = k_n$, $b_n^2 = 0.5V(\hat{\varepsilon})$, and minimizing the above maximum leads to

(i) If $1 < \delta \leq 1.5$, $k_n = n^{\frac{2a}{2\delta+6}} \times (b_n^2)^{\frac{-1}{2\delta+6}}$. Consequently, an $n^{-\frac{2a(2\delta-1)}{2\delta+6}} \times (b_n^2)^{\frac{2a(2\delta-1)}{2\delta+6}} \to D_1 \times n^{-\frac{2a(2\delta-1)}{2\delta+6}}$ convergence rate is determined by the equality of the first term and the first addend of the third term where $D_1 = V(\varepsilon)^{\frac{2a(2\delta-1)}{2\delta+6}}$.

(ii) If $\delta > 1.5$, $k_n = n^{\frac{2a}{4\delta+3}} \times (b_n^2)^{\frac{-4}{4\delta+3}}$. Consequently, an $n^{-\frac{2a(2\delta-1)}{4\delta+3}} \times (b_n^2)^{\frac{2a(2\delta-1)}{4\delta+3}} \to D_2 \times n^{-\frac{2a(2\delta-1)}{4\delta+3}}$ convergence rate is determined by the equality of the first term and the second addend of the third term where $D_2 = V(\varepsilon)^{\frac{2a(2\delta-1)}{4\delta+3}}$. We exclude the case with $\delta = 2$ (Laplace-Laplace convolution) here as similar reasoning in the proof of Theorem 1 applies.

Table 1: Smoothness Parameters of Some Popular Continuous Distributions

Name	Parameter	Density	Chara. Function	Smoothness Parameters C_1	C_2	δ	T
Symm. Uniform	$a > 0$	$\frac{1}{2a}1_{[-a,a]}(x)$	$\frac{\sin(at)}{at}$	0^+	$\frac{1}{a}$	1	0^+
Laplace	$b > 0$	$\frac{1}{2b}e^{-\frac{\lvert x\rvert}{b}}$	$\frac{1}{1+b^2t^2}$	$\frac{1}{1+b^2}$	$\frac{1}{b^2}$	2	1
Uniform	$a, b(b > a)$	$\frac{1}{2(b-a)}1_{[a,b]}(x)$	$\frac{e^{itb}-e^{ita}}{it(b-a)}$	$\frac{\lvert\cos(b)-\cos(a)\rvert}{b-a}$	$\frac{2}{b-a}$	1	0^+
χ^2_k	$k > 0$	$\frac{1}{2^{k/2}\Gamma(k/2)}x^{k/2-1}e^{-\frac{x}{2}}$	$\frac{1}{(1-2it)^{k/2}}$	$\frac{1}{(2^{k/2})^+}$	1	$k/2$	1
Gamma	$k > 0, \theta > 0$	$\frac{1}{\Gamma(k)\theta^k}x^{k-1}e^{-\frac{x}{\theta}}$	$\frac{1}{(1-i\theta t)^k}$	$\frac{1}{(\theta^k)^+}$	$1(\theta > 1)$	k	$\frac{1}{\theta}$
Twice-convolved Laplace	$b > 0$	$\frac{1}{4b}e^{-\frac{\lvert x\rvert}{b}}(\lvert x\rvert + b)$	$\frac{1}{(1+b^2t^2)^2}$	$\frac{1}{4}$	1	4	$\frac{1}{b^2}$
Cauchy	$\mu = 0, \theta > 0$	$\frac{\theta}{\pi(\theta^2+x^2)}$	$e^{-\theta\lvert t\rvert}$	NA	NA	∞	0^+
Normal	$\mu = 0, \sigma^2 > 0$	$\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$	$e^{-\frac{1}{2}\sigma^2t^2}$	NA	NA	∞	0^+

Notes: The ordinary-smoothness parameter are defined by the Fan(1991): $C_1|\tau|^{-\delta} \leq |h_x(\tau)| \leq C_2|\tau|^{-\delta}$ for $|\tau| \geq T > 0$ where $0 < C_1 < C_2$, $\delta > 1$ and $h_x(\tau)$ is the characteristic function of the corresponding distribution. $\Gamma(s) = \int_0^\infty t^{s-1}e^{-t}dt$. The last two rows are from the super-smooth family.

Table 2: Effective Sample Size Compared to Ordinary Least Squares (OLS)

Estimator	Convergence Rate	Sample Size n: 100	1,000	3,000	5,000	10,000
OLS	$n^{-1/2}$	100	1,000	3,000	5,000	10,000
Laplace v:	$n^{-2/9}$ ($\delta = 1.5$)	8	22	35	44	60
	$n^{-1/3}$ ($\delta = 3$)	22	100	208	292	464
	$n^{-7/19}$ ($\delta = 4$)	30	162	365	531	886
Normal v:	$\frac{\ln(\ln n)}{\ln n}$ ($\delta = 1.5$)	9	13	15	16	17
	$\left(\frac{\ln(\ln n)}{\ln n}\right)^2$ ($\delta = 3, 4$)	83	163	219	250	296

Notes: Assume δ is known for both deconvolution cases (Laplace v and Normal v). For an $n^{-\alpha}$ convergence rate, the effect sample size is calculated by $n^{2\alpha}$. Similarly, for a $(\ln(\ln n)/\ln n)^2$ convergence rate, it could be calculated as $(\ln n/\ln(\ln n))^{2\times 2}$.

Table 3: RMISE for Laplacian Noise Deconvolution

n	$\sigma_u^2/\sigma_v^2 = 2/1$	$\sigma_u^2/\sigma_v^2 = 2/2$	$\sigma_u^2/\sigma_v^2 = 2/4$
500	0.0162	0.0155	0.0204
1,000	0.0150	0.0143	0.0197
3,000	0.0138	0.0126	0.0190

Notes: Replication 500 times. $\frac{\sigma_u^2}{\sigma_v^2}$ stands for the signal-to-noise ratio

Table 4: RMISE under Misspecification: Normal Noise Deconvolution

n	$\sigma_u^2/\sigma_v^2 = 2/1$		$\sigma_u^2/\sigma_v^2 = 2/2$		$\sigma_u^2/\sigma_v^2 = 2/4$	
	CHP	Meister06	CHP	Meister06	CHP	Meister06
500	0.0155	0.0128	0.0186	0.0170	0.0242	0.0340
1,000	0.0143	0.0116	0.0168	0.0156	0.0234	0.0337
3,000	0.0129	0.0108	0.0152	0.0146	0.0230	0.0330

Notes: Replication 500 times. $\frac{\sigma_u^2}{\sigma_v^2}$ stands for the signal-to-noise ratio.

Table 5: Simulation by Rule-of-Thumb Adaptive Procedure with Laplace Noise

N	$\sigma_u^2/\sigma_v^2 = 2/1$			$\sigma_u^2/\sigma_v^2 = 2/2$			$\sigma_u^2/\sigma_v^2 = 2/4$		
	RMISE	Ave. δ	Ave. C_1	RMISE	Ave. δ	Ave. C_1	RMISE	Ave. δ	Ave. C_1
500	0.0139	2.02	0.10	0.0133	2.02	0.10	0.0244	2.04	0.10
1,000	0.0125	2.00	0.10	0.0112	2.00	0.10	0.0231	2.08	0.10
3,000	0.0110	2.00	0.10	0.0094	2.00	0.10	0.0221	2.00	0.10

Notes: Replication 100 times. $\frac{\sigma_u^2}{\sigma_v^2}$ stands for the signal-to-noise ratio.

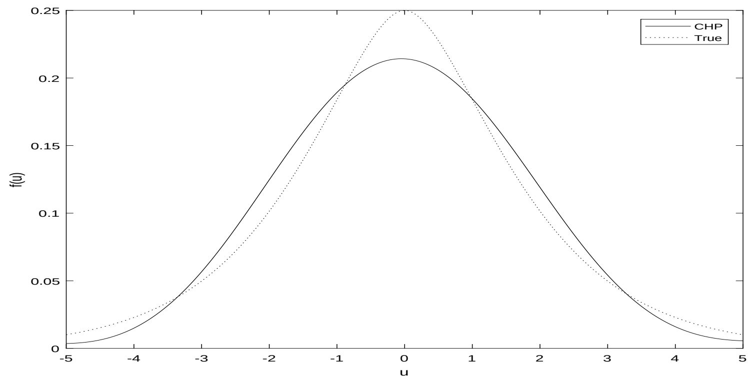

Figure 1: Laplace Deconvolution (CHP): $n = 500, \sigma_u^2/\sigma_v^2 = 2/2$

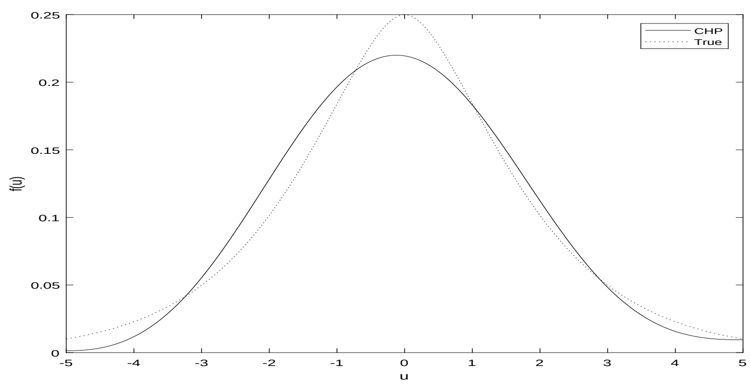

Figure 2: Laplace Deconvolution (CHP): $n = 1,000, \sigma_u^2/\sigma_v^2 = 2/2$

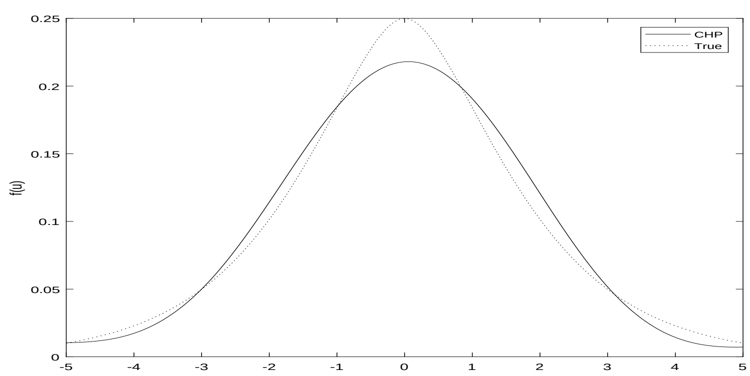

Figure 3: Laplace Deconvolution (CHP): $n = 3,000, \sigma_u^2/\sigma_v^2 = 2/2$

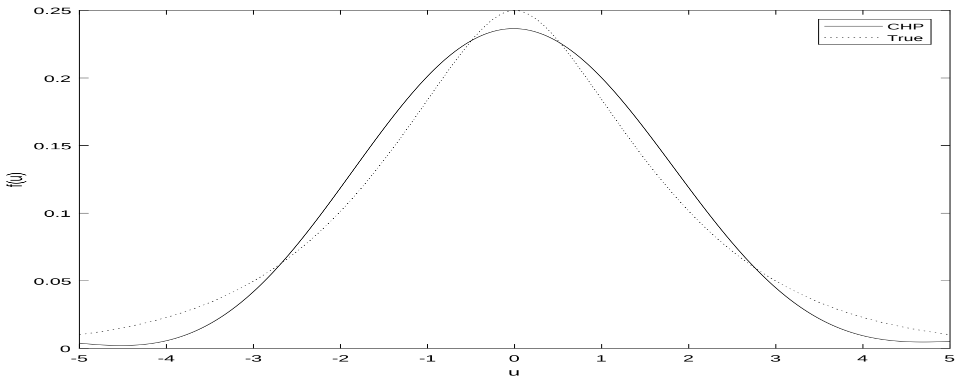

Figure 4: Laplace Deconvolution (CHP): $n = 1,000, \sigma_u^2/\sigma_v^2 = 2/1$

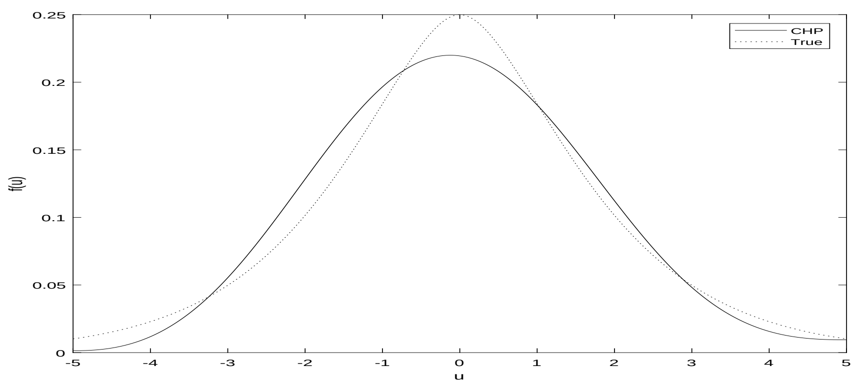

Figure 5: Laplace Deconvolution (CHP): $n = 1,000, \sigma_u^2/\sigma_v^2 = 2/2$

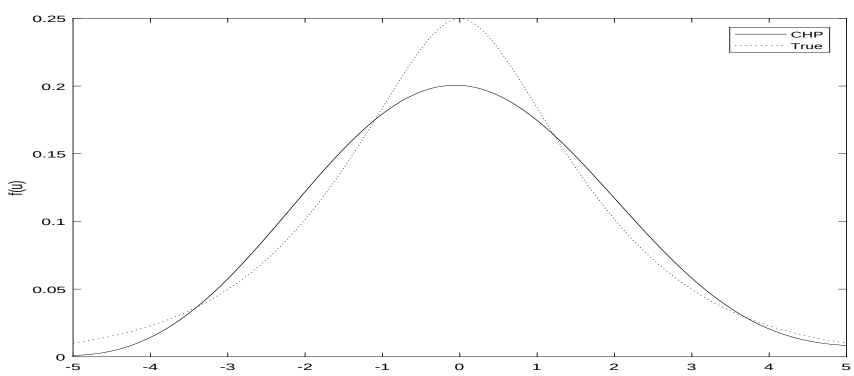

Figure 6: Laplace Deconvolution (CHP): $n = 1,000, \sigma_u^2/\sigma_v^2 = 2/4$

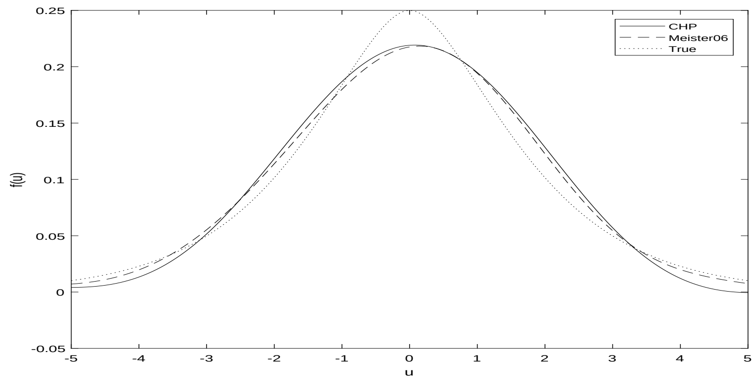

Figure 7: Misspecified Laplace (CHP) Deconvolution: $n = 500, \sigma_u^2/\sigma_v^2 = 2/2$

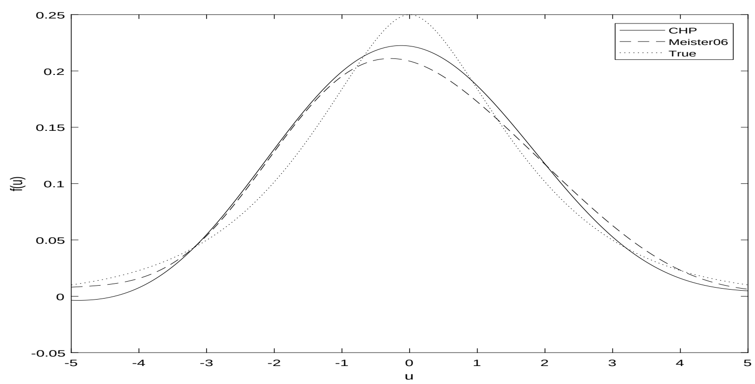

Figure 8: Misspecified Laplace (CHP) Deconvolution: $n = 1,000, \sigma_u^2/\sigma_v^2 = 2/2$

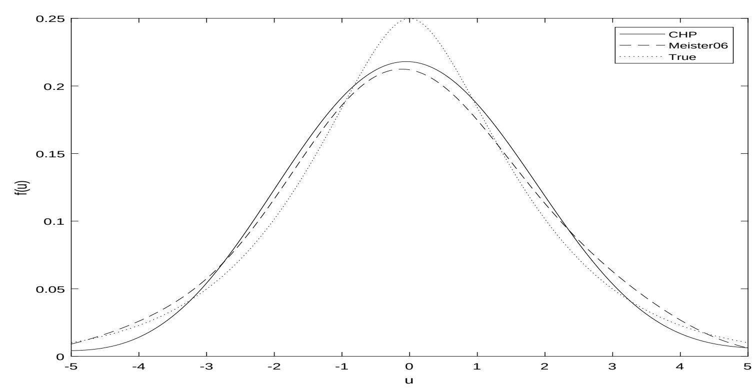

Figure 9: Misspecified Laplace (CHP) Deconvolution: $n = 3,000, \sigma_u^2/\sigma_v^2 = 2/2$

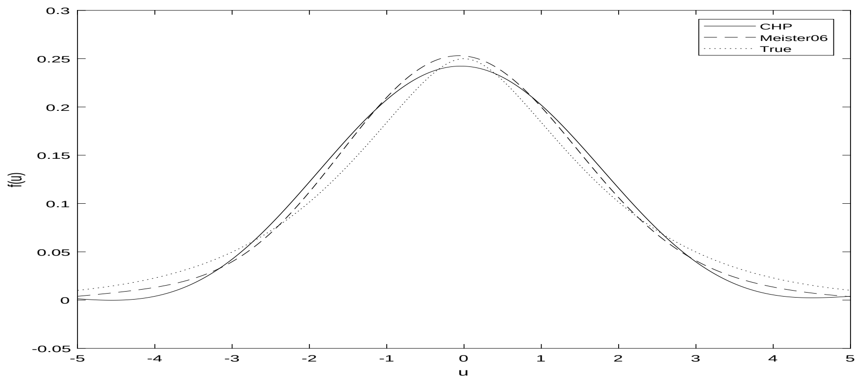

Figure 10: Misspecified Laplace (CHP) Deconvolution: $n = 1,000, \sigma_u^2/\sigma_v^2 = 2/1$

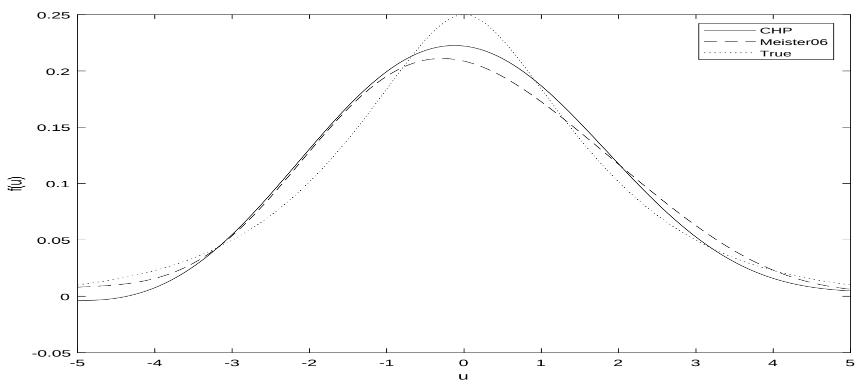

Figure 11: Misspecified Laplace (CHP) Deconvolution: $n = 1,000, \sigma_u^2/\sigma_v^2 = 2/2$

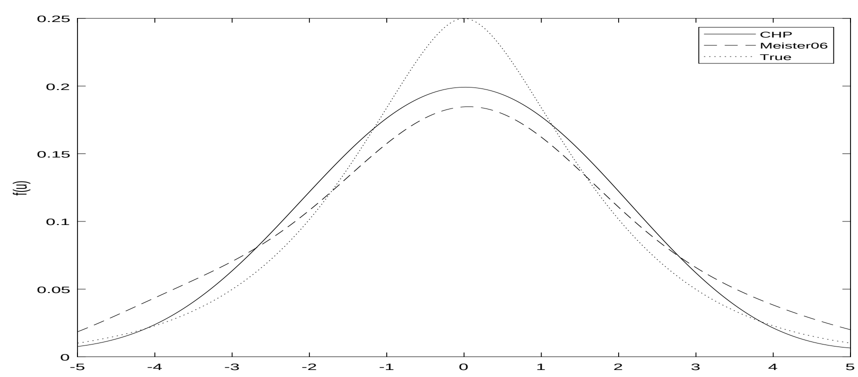

Figure 12: Misspecified Laplace (CHP) Deconvolution: $n = 1,000, \sigma_u^2/\sigma_v^2 = 2/4$

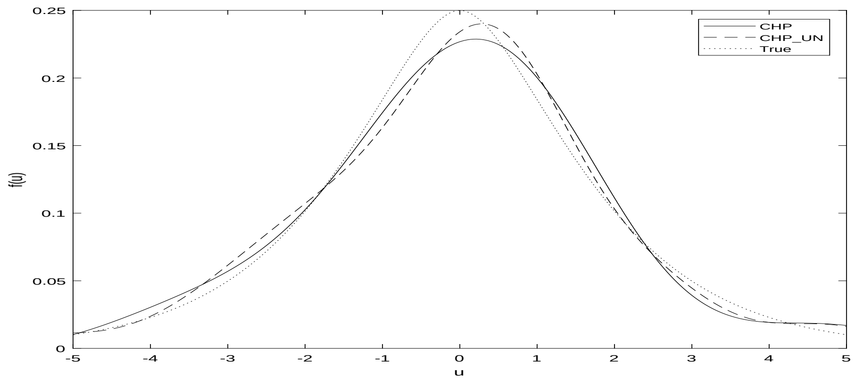

Figure 13: Deconvolution with Unknown Smooth Parameters, $n = 1,000$, $\sigma_u^2/\sigma_v^2 = 2/2$

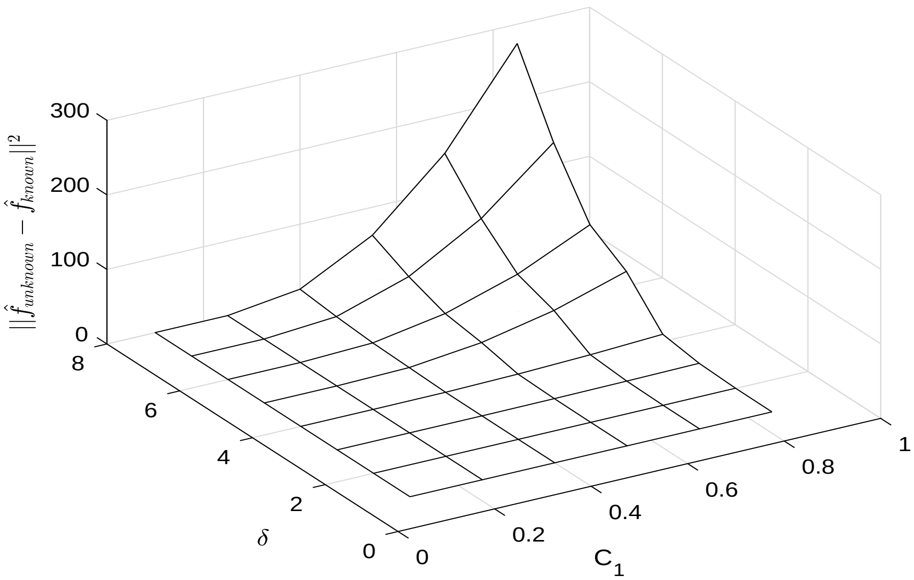

Figure 14: Euclidean Distance Between $\hat{f}_{unknown}$ and $\hat{f}_{known}$

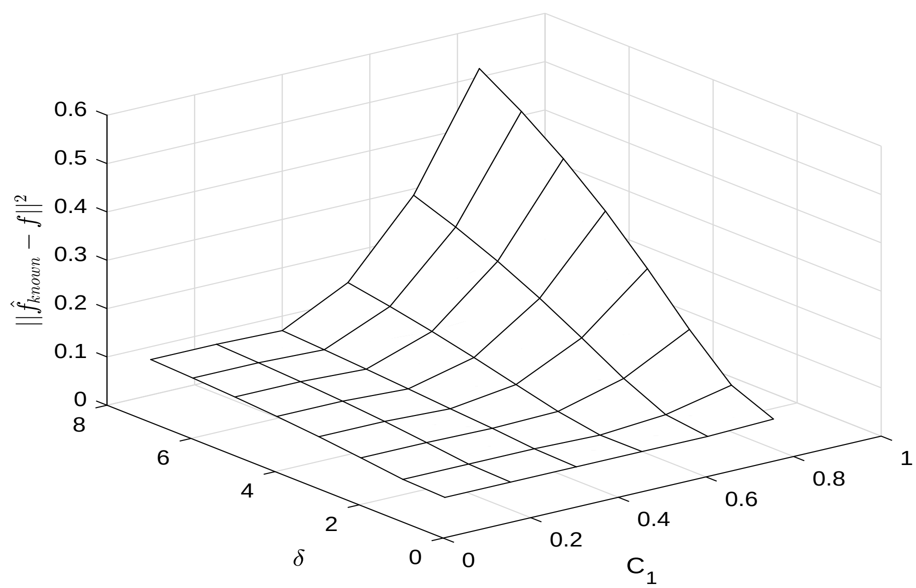

Figure 15: Euclidean Distance Between $\hat{f}_{known}$ and True Density

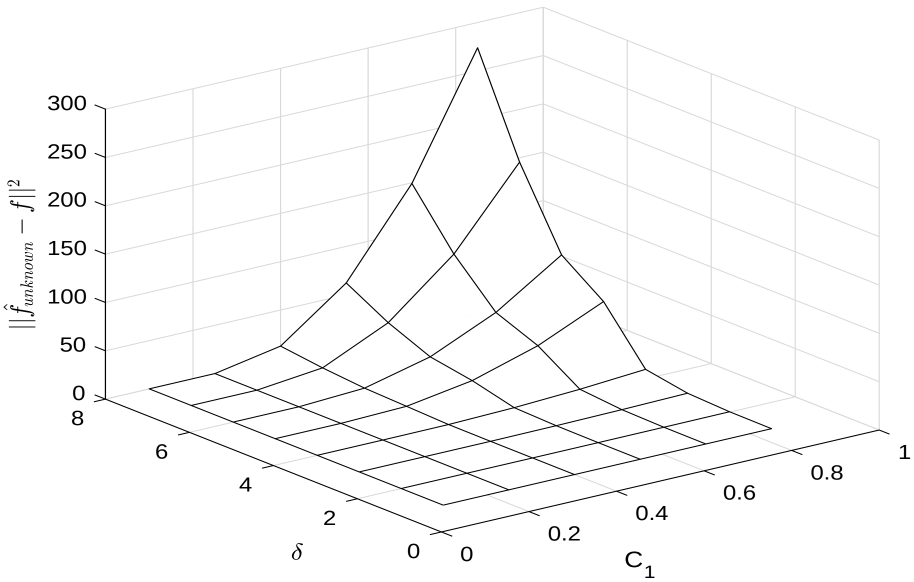

Figure 16: Euclidean Distance Between $\hat{f}_{unknown}$ and True Density

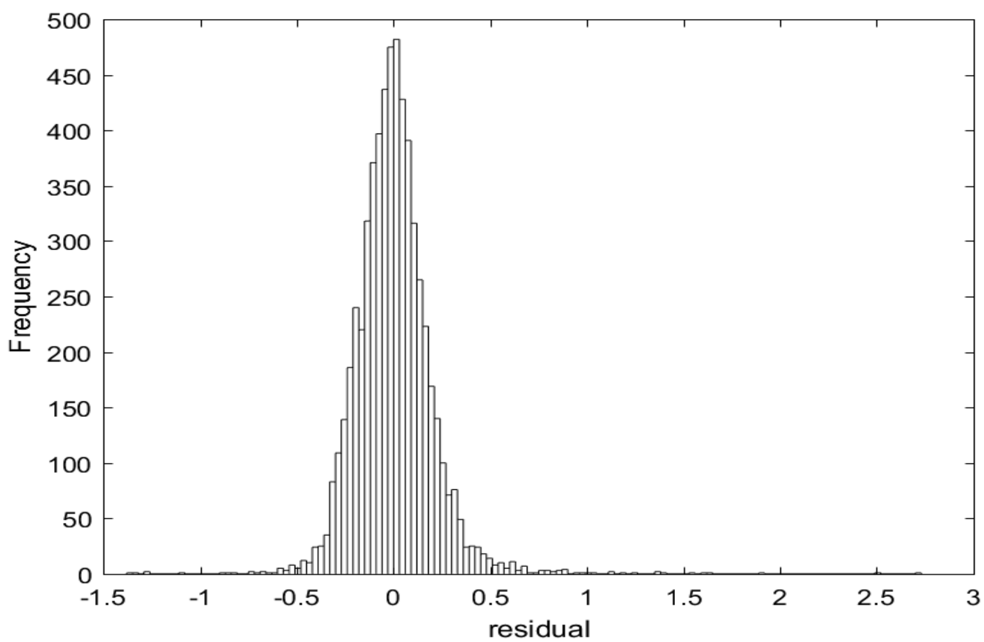

Figure 17: Histogram of the Residuals

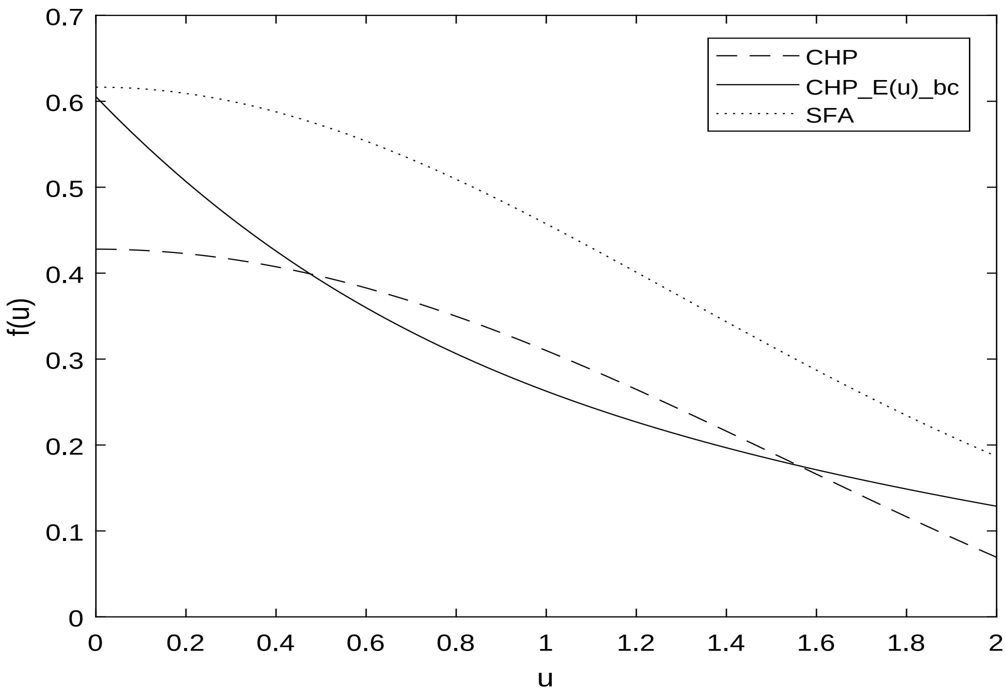

Figure 18: Estimated density of inefficiency

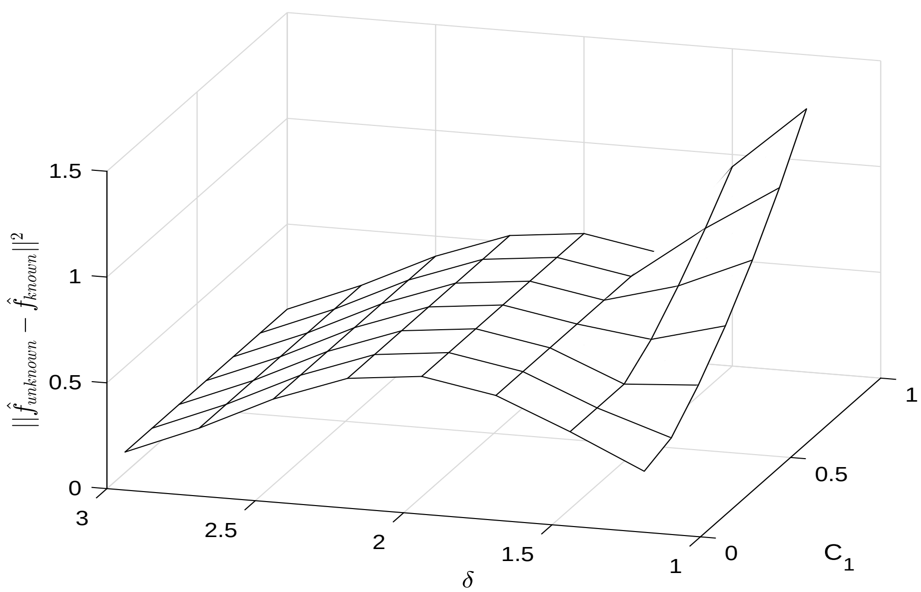

Figure 19: Euclidean Distance Between $\hat{f}_{unknown}$ and $\hat{f}_{known}$

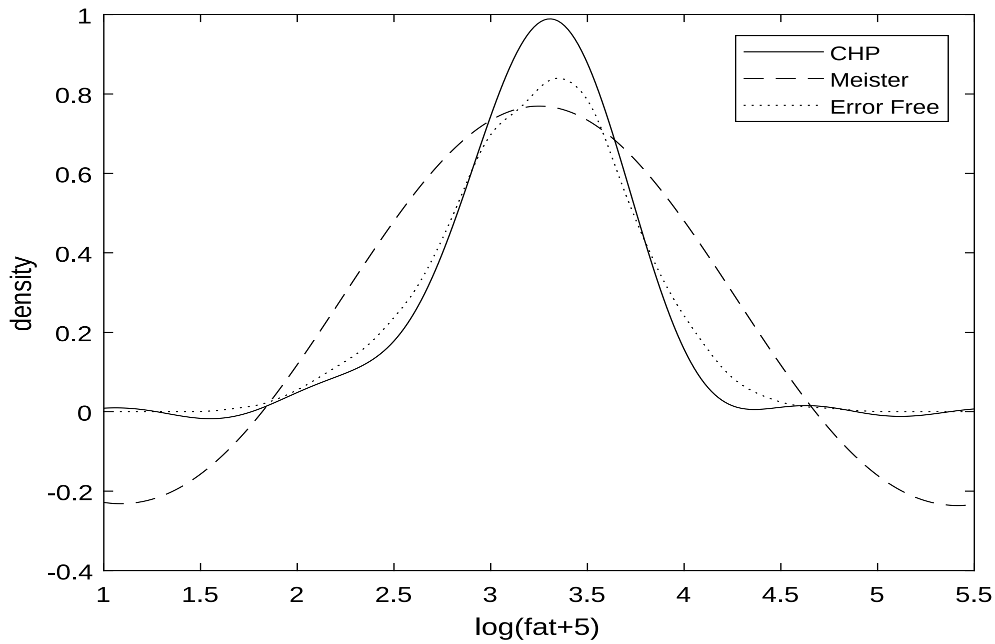

Figure 20: Density of the Logarithm of Daily Saturated Fat Intake, $C_1 = 1$, $\delta = 1.5$

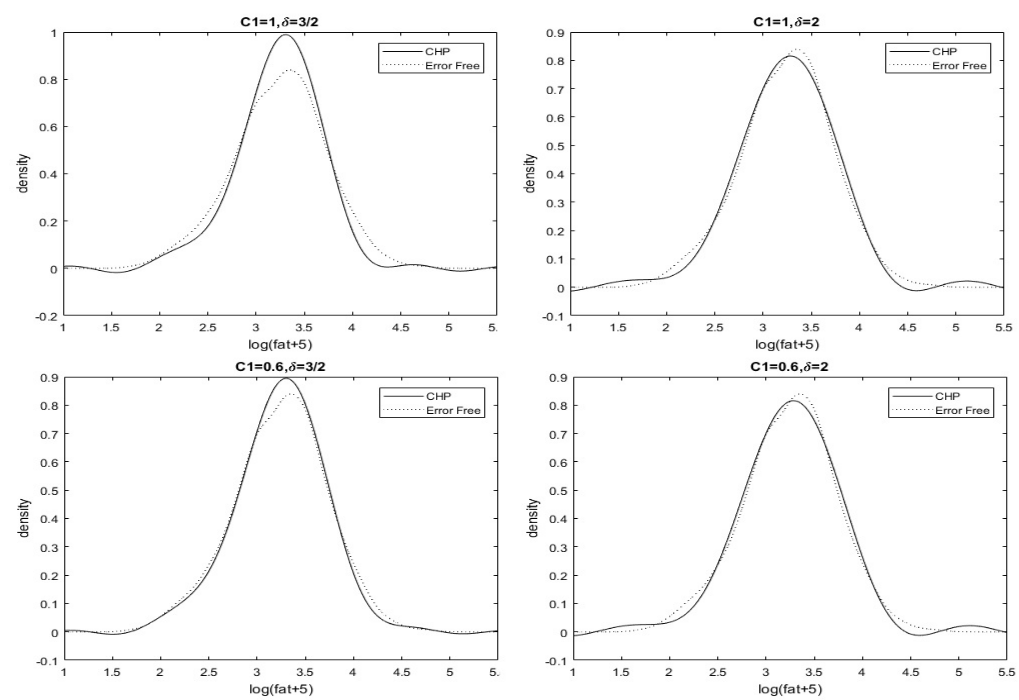

Figure 21: Saturated Fat Intake with Various Values of C_1 and δ