Density Deconvolution with Laplace Errors and Unknown Variance

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

Paper No. 225
March 2020
CENTER FOR POLICY RESEARCH – Spring 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
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
Michael Wasylenko, ECON
Peter Wilcoxon, PAIA
Maria Zhu, ECON

GRADUATE ASSOCIATES

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

STAFF

Joanna Bailey, Research Associate
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
Abstract

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

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
1 Introduction

Deconvolution is a versatile technique for recovering the density of a random variable of interest \( u \) in the presence of an independent and additive measurement error or noise \( v \). The vast majority of deconvolution methods are developed for the setting where a random sample of observations from a contaminated random variable is observable \( \varepsilon = u + v \) and the error distribution \( f_v \) is assumed to be completely known (e.g., Neumann, 1997; Johannes, 2009; Wang and Ye, 2012). Stefanski and Carroll (1990) give an early treatment in which they consider kernel estimation of a continuous and bounded target density convolved with errors from a “fully known” normal density and derive the convergence rate for a family of deconvolution kernel density estimators with different assumptions on the density of the errors.\(^1\)

For normal errors, they show that the target density estimator has a uniform convergence rate of \( \ln n \), with \( n \) being the sample size. Fan (1991) shows that optimal convergence rates for nonparametric deconvolution density estimators depend on the smoothness of the error distribution. For the super-smooth family, such as the normal distribution, the fastest convergence rate is \( O((\ln n)^c) \), while for the ordinary-smooth family, for example the Laplace distribution, the fastest attainable rate is polynomial, \( O(n^c) \), with \( c \) a constant depending on the smoothness of the assumed target density. While normality is a common and convenient assumption for the error distribution in a deconvolution problem, it comes at the cost of logarithmic convergence rates for the target density.

Recent contributions to the literature relax the “fully-known” assumption on the error distribution, see Butucea and Matias (2004), Meister (2006), and Butucea, Matias, and Pouet (2008) and the references therein. Meister (2006) dispenses with the assumption of known variance for zero-mean normal errors and develops “semi-uniformly consistent” estimators.

\(^1\)Typically the normal errors are assumed zero-mean with known variance, which is often unrealistic in empirical exercises.
for both the target density and the unknown error variance.\textsuperscript{2}

Meister (2006) assumes that the target density is from the ordinary-smooth family of
distributions which places a lower bound on the rate of decay of its characteristic function’s
setting with a composed error (notably the stochastic frontier model).\textsuperscript{3} They develop a
semi-uniformly consistent estimator of the target density, if the error distribution is known
to be normal up to its variance and the target density is ordinary-smooth.

Although infrequently studied, an alternative approach is to assume that the error distri-
bution is from the ordinary-smooth family (Laplace, gamma, etc.). Meister (2004b) shows
that if the error distribution is misspecified, it is always better to assume Laplace errors
rather than normal, because normal errors produce infinite risk while Laplace errors have
finite risk in the deconvolution setting. A similar result arises in the simulations of Hor-
race and Parmeter (2018) who find that the Mean Squared Error (MSE) of the parametric
stochastic frontier model is smaller with Laplace errors than with normal errors under mis-
specification.\textsuperscript{4} Additionally, they show that there exist many interesting and unique features
of the stochastic frontier model when the error term is Laplace instead of normal.\textsuperscript{5} Moreover,
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, maxi-
mum likelihood estimation with Laplace errors produces the least absolute deviations (LAD)

\textsuperscript{2}Semi-uniform consistency means that the estimator is uniformly consistent for the target density but
individually for the error distribution with bounded variance, i.e. uniformity does not hold over all error
distributions.

\textsuperscript{3}Antecedent to Horrace and Parmeter (2011), Horowitz and Markatou (1996) consider the linear regres-
sion case where panel data are available and neither component of the composed error’s density is known.
Essentially, the information in the time-dimension of the panel replaces the distributional assumptions on
the errors.

\textsuperscript{4}This is in the non-deconvolution setting, where the MSE (the risk) is for maximum likelihood estimation
of the variance parameters.

\textsuperscript{5}For example, they find that the distribution of inefficiency conditional on the composed error is constant
(in the composed error) for positive values of the composed error, but varies for negative values. Hence,
there are ties for the most efficient firm in the sample in the Laplace stochastic frontier model.
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 noise distribution to be fully known, using results of Meister (2006). We modify the “variance truncation device” developed in Meister (2006) to incorporate the informational content of the variance of the noisy random variable under examination. Target density estimation is drastically improved in the Laplace measurement error scenario and, following Meister (2004b), is robust to misspecification of the error distribution. Moreover, we offer practical guidance and an adaptive procedure for selecting the smoothness parameters which are key to implementation of the proposed techniques. This adaptive procedure is new in the literature and offers sound footing for use of these methods in practice. Lastly, two practical applications of this method are provided: a regression stochastic frontier model (SFM), widely used in the efficiency and productivity literature, and a pure deconvolution problem for measurement errors without regression.

The paper is organized as follows. In Section 2 we discuss the basic issues surrounding deconvolution inherent to the error component regression model and introduce the modified variance truncation device under Laplace errors. Section 3 derives large sample properties of the estimator under certain regularity conditions. Two extensions are discussed 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 Laplace error. 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:

\[ y_j = x_j' \beta + u_j + v_j = x_j' \beta + \varepsilon_j, \quad j = 1, 2, \ldots, n. \tag{1} \]

Here \( j \) indexes individuals or firms, \( \beta \) is a parameter vector of dimension \( q \) to be estimated and exogenous covariates \( x \in \mathbb{R}^q \). \( \varepsilon \) is the composed error term and \( v \) is the random noise. Depending on the assumptions of the unknown error component \( u \), it can be a cross sectional stochastic frontier model (e.g., \( u \sim |N(0, \sigma_u^2)| \)) or 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 statistical measurement error model (e.g., \( \beta = 0, y_j = u_j + v_j \)). Instead of focusing on estimation of the unknown parameter \( \beta \), our goal is to estimate the density of the error component \( u \) without parametric assumptions on its distributional family. Absent \( x_j' \beta \) (i.e., \( \beta = 0 \)), equation (1) represents the classic or “pure” deconvolution problem. A large statistical literature investigates this restricted model with known or partially-known error distribution \( f_v \) (see Meister, 2009).\(^6\)

In this setting, deconvolution is complicated by the fact that only cross sectional data are available and the random noise \( v \) is not fully known.

Following the literature (Fan, 1991a; 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

\(^6\)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, \( y_{jt} \).
between $v_j$ and $u_j$ in Assumption 1,

$$h_\varepsilon(\tau) = h_v(\tau)h_u(\tau)$$  \hspace{1cm} (2)$$

We restrict attention to densities that satisfy the following two assumptions.

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

By assumption 2, the characteristic function of $v$ is $h_v(\tau) = \frac{1}{1 + b^2 \tau^2}$. Then we have

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

**Assumption 3.** *Assume $u$ is ordinary-smooth defined by Fan (1991a), namely, $u$ belongs to the family $\mathcal{F}_u = \{h_u : C_1|\tau|^{-\delta} \leq |h_u(\tau)| \leq C_2|\tau|^{-\delta}, \text{ for } |\tau| \geq T > 0\}$ where $0 < C_1 < C_2$ and $\delta > 1$. Assume $C_1$ and $\delta$ are known.*

Assumption 2 is standard and restricts $v$ to the class of Laplacian distributed random variables with mean zero and unknown variance $2b^2$. Clearly, we require $h_v(\tau) \neq 0$ for any $\tau$. Assumption 3 dictates tail behavior of the characteristic function of $u$. The upper and lower bounds ensure the rate of decay of the tails of the characteristic function does not approach zero too rapidly or too slowly and is needed for identification. $C_1$ and $C_2$ are constants that become irrelevant when $T$ gets large, while $\delta$ ensures polynomial tail behavior and includes a wide array of densities (Horrace and Parmeter, 2011). The assumption that $C_1$ and $\delta$ are known is not necessary. We adopt this convention here to more clearly illustrate the key intuition of our results. The more general case of unknown $C_1$ and $\delta$ will be discussed in Section 4. Examples of distributions satisfying Assumption 3 are the Laplace
The latter is commonly used in the parametric stochastic frontier literature.

Under Assumption 2 and 3, the Fourier inversion formula identifies the first derivative of the distribution of \( u \), which equals the density of \( u \),

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

where \( i = \sqrt{-1} \). If \( v \sim \mathcal{G} = \{ N(0, \sigma^2) : \sigma^2 > 0 \ \text{unknown} \} \), under Assumptions 1-3, Meister (2006) demonstrates that there is no uniformly consistent estimator of \( f_u(z) \). He shows that one can estimate \( f_u(z) \) semi-uniformly consistently 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 distribution case with unknown variance. As we will demonstrate, one still pays a price for not knowing variance, but the cost is not as high as in the case with normally distributed errors.

Since \( h_\varepsilon \) is unknown, we may rely on the empirical characteristic function to recover the density of \( u \) based on equation (4),

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

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

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

(6)

Replacing \( h_\varepsilon \) with \( \hat{h}_\varepsilon \) or \( \hat{h}_\varepsilon \) in equation (6) 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,

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

(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 illustration of the case where \( C_1 \) and \( \delta \) are unknown. Under Assumption 3 in which \( C_1 \) and \( \delta \) are known, \( w_n = k_n \).

Due to the upper and lower bound conditions on the target density function in Assumption 3, we are able to estimate the unknown error variance consistently. Therefore, setting \( \tilde{b}_n^2 = k_n^{-2} \left( \frac{C_1 k_n^{-\delta}}{h_\varepsilon(k_n)} - 1 \right) \) with constants \( \delta > 1 \) and \( C_1 > 0 \), we propose an explicit truncation

\[ \tilde{b}_n^2 = k_n^{-2} \left( \frac{C_1 k_n^{-\delta}}{h_\varepsilon(k_n)} - 1 \right) \]
device for the unknown variance:

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

where \( \delta > 1 \) and \( C_1 > 0 \) are the parameters of the target density in Assumption 3, and \( b_n^2 = \frac{1}{2} Var(\tilde{\epsilon}) \), which is half the total variance of the estimated sum of the error components.

The intuition is that we choose an increasing sequence to cover the unknown variance, \( \tilde{b}_n^2 \), and bound it by the half variance.\(^8\) 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 \( v \). This is trivially satisfied under Assumption 1. Compared with 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 \( b_n^2 = \frac{1}{4} \ln \ln n \) 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 bound, \( \hat{b}_n^2 \), 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 estimator in equation 7.

\(^8\)Recall that for a Laplace distribution as defined in Assumption 2, the variance is \( var(v) = 2b^2 \). Moreover, \( Var(v) < Var(\tilde{\epsilon}) \) under Assumption 1. Hence, a natural upper bound for \( b^2 \) is one-half the variance.
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. All proofs appear in an appendix.

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

Assumption 5. The estimator $\beta_n$ 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 $\beta = 0$, and Assumption 5 is trivially satisfied. Moreover, the function $x_j'\beta$ can be estimated with a nonparametric $n^a$ convergence rate and $a = \frac{2}{4+q}$.

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

Lemma 1. For Assumptions 1,3-5 and $L_n = \{\text{Laplace}(0, b) : b^2 \in (0, b_n^2]\}$, 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 $B \leq \text{const}_1 * w_n^{1-2\delta}$,

$$V \leq \text{const}_2 * n^{-1} w_n (1 + b_n^2 w_n^2)^2 + \text{const}_3 * n^{-1} w_n^3 (1 + b_n^2 w_n^2)^2,$$

$$E \leq \text{const}_4 * \sup_{g \in \mathcal{L}_n} \sup_{f \in \mathcal{F}_u} \left( w_n \int_{-1}^1 |h_u(sw_n)|^2 \left( \frac{d_n}{w_n} \right)^2 ds + w_n \int_{-1}^1 |h_u(w_n s)|^2 \frac{b_4}{b^4} \ast P_{f,g}(\hat{b}_n^2 - b^2 > d_n) ds \right), \text{ with } d_n := \frac{1}{w_n}; f \text{ and } g \text{ are the probability density function in distribution family } \mathcal{F}_u \text{ and } \mathcal{L}_n \text{ respectively. } \text{const}_j \text{ are positive constants for } j = 1, 2, 3, 4.$$. 
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. The $B$ term is a bias component of the estimator 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 a bias-variance hybrid and the second integral looks like a variance. This entire bound exhibits the usual bias-variance trade-off in nonparametric density estimation. One point worth mentioning is that the second addend of $V$, arises from the regression function which does not appear in the pure deconvolution setting of Meister (2006).

The convergence rate of $E$ is not straight-forward. We need the following Lemma to assist in determining this rate.

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

Proof is in the appendix. Compared to the normal deconvolution of Horrace and Parmeter (2011), estimation of $\varepsilon$ 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 $\beta_n$ and $\hat{f}_u$ converge at polynomial rates, so there is an additional effect for convergence of the target density. Given that we replace $\varepsilon$ 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

---

9In Meister (2006), the bounding of the normal variance is what leads to semi-uniform 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.

10The 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.
in $E$, ensures convexity of the entire bound with respect to the bandwidth $k_n$. Therefore, the optimal bandwidth $w_n$ and the entire convergence rate of the density estimator can be determined.

Notice that neither of the proofs of 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 in the Mean Integrated Squared Error (MISE) of Lemma 1, this assumption plays a role. That is, if we fully know the smoothness parameters (i.e., $C_1$ and $\delta$) 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 $\delta$ are known.

**Theorem 1.** Under Assumption 1, 3-5, consider (7) and take $\{b_n^2\}_{n \in \mathbb{N}} = \frac{1}{2} \text{Var}(\hat{\varepsilon})$ and $w_n = k_n$ with $\{k_n\}_{n \in \mathbb{N}} = \{\left(\frac{n}{b_n^2}\right)^{\frac{1}{1+2\delta}}\}_{n \in \mathbb{N}}$ if $1 < \delta \leq 3/2$; $\{k_n\}_{n \in \mathbb{N}} = \{\left(\frac{n}{b_n^2}\right)^{\frac{1}{3+4\delta}}\}_{n \in \mathbb{N}}$ if $\delta > 3/2$.

For any $g \in \mathcal{L}_n$, the proposed deconvolution kernel density estimator in equation (7) is bounded from above as follows:

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

and

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

where $\delta$ is defined by Assumption 3.

Proof is in the appendix. The proposed density estimator is semi-uniformly consistent, namely, $\hat{f}_u$ is uniformly consistent over a given class of Laplace distribution $\mathcal{L}_n$. The optimal convergence rate for an ordinary-smooth target density is achieved in a minimax sense.
It is similar to the conclusions in Fan (1991a), even though in this exercise the variance of the error 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 (1991a). At first glance the proposed theorem is similar to Theorem 3 in Meister (2006), but there are three major differences: (i) the upper bound of the random 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 \( \frac{1}{2} \text{Var}(\varepsilon) \)); (ii) the chosen sequences are functions of the target density parameter \( \delta \) which is due to the characteristic function of the Laplace distribution; (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 errors. 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 errors. It is more nuanced in the context of Laplace errors, given the polynomial rate of convergence. This has important implications if one were to estimate the \( \beta \) 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 \( \delta \) are unknown in Assumption 3 and (ii) deploying nonparametric regression to estimate \( \beta \) in the calculation of \( \hat{\varepsilon} \). It is rare in applications that researchers have information on the target density. This leads to uncertainty in \( C_1 \) and \( \delta \), two
parameters which are important in implementation of our estimator and the Meister (2006) estimator. 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 errors due to their logarithmic convergence rates.

4.1 Selection of Unknown $C_1$ and $\delta$

In the usual case that $\delta$ and $C_1$ are unknown and, therefore, might be misspecified,\footnote{Actually, if one wants to assume the random noise is super-smooth with similarity index $s$, the smoothness parameter $\delta$ of target density can be estimated as well as the $s$ by an adaptive procedure proposed by Butucea, Matias and Pouet (2008).} we could apply the following selection rule due to Meister (2006):

**Selection rule 1.** If $C_1$ and $\delta$ are unknown, we specify one set of \{$C_1, \delta$\} and choose

$$w_n = k_n / \ln k_n.$$ 

An alternative rule may be based on our procedure when $\delta$ 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 data set to achieve a reliable estimate of the target density. This can be seen from following theorem.

**Assumption 6.** Assume $u$ belongs to the family $\mathcal{F}_u = \{h_u$ characteristic function : $C_1|\tau|^{-\delta} \leq |h_u(\tau)| \leq C_2|\tau|^{-\delta}$, for $|\tau| \geq T > 0\}$ where $0 < C_1 < C_2$ and $\delta > 1$. Assume $C_1$ and $\delta$ are unknown.
Theorem 2. Under Assumption 1, 4, 5 and 6 consider equation (7) with \( \{b_n^2\}_{n \in \mathbb{N}} = \frac{1}{2} \text{Var}(\hat{\varepsilon}) \)
and \( w_n = k_n / \ln k_n \)
with \( \{k_n\}_{n \in \mathbb{N}} = \left\{ \left( \frac{n}{b_n^2} \right)^{\frac{1}{\delta + 2}} \right\}_{n \in \mathbb{N}} \)
if \( 1 < \delta \leq 3/2 \); \( \{k_n\}_{n \in \mathbb{N}} = \left\{ \left( \frac{n}{b_n^2} \right)^{\frac{1}{\delta + 2}} \right\}_{n \in \mathbb{N}} \)
if \( \delta > 3/2 \). For any \( g \in \mathcal{L}_n \), the proposed deconvolution kernel density estimator in equation (7) is bounded from above as following:

\[
\sup_{f_u \in \mathcal{F}_u} E_{f,g} ||\hat{f}_u - f_u||^2_{L^2} \leq \left( \frac{n}{\ln n} \right)^{-\frac{2\delta - 1}{\delta + 2}} \quad \text{if} \quad 1 < \delta \leq 3/2
\]

and

\[
\sup_{f_u \in \mathcal{F}_u} E_{f,g} ||\hat{f}_u - f_u||^2_{L^2} \leq \left( \frac{n}{\ln n} \right)^{-\frac{2\delta - 1}{\delta + 2}} \quad \text{if} \quad \delta > 3/2
\]

where \( \delta \) is defined by Assumption 6.

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. We list smoothness parameters (\( \delta \) and \( C_1 \)) of typical continuous distributions in Table 1. For example, the symmetric uniform and the uniform distributions have \( \delta = 1 \), and we cannot identify the target density in these cases, because the uniform densities are uninformative. For the twice-convolved Laplace \( \delta = 4 \). Following Meister (2006) and Horrace and Parmeter (2011), this is the target density that we use in our simulation exercise.

Based on the conclusion of Theorem 2 and Table 1, we propose a rule-of-thumb adaptive procedure as follows:

**Step 1** Set initial estimates for \( C_1 \) and \( \delta \). A useful rule-of-thumb is \( C_1 \) is commonly between 0 and 1; \( \delta \) is between 1 and 10.
Step 2 Treating this $C_1$ and $\delta$ as “known,” select $k_n = w_n$ and apply the proposed deconvolution techniques to construct the estimated target density, $\hat{f}_{\text{known}}(u)$, say.

Step 3 Now, with the same $C_1$ and $\delta$ 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 $\hat{f}_{\text{unknown}}(u)$, say.

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

One caveat with this iterative approach is that $\Delta$ 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 errors, and not a density.

4.2 Nonparametric estimation of $\beta$

If one is unsure of the linear relation between the dependent variable and the covariates, equation (1) can be generalized to the nonparametric case as follows:

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

(9)
where $g(.)$ is unknown and $x \in \mathcal{R}^q$. Under certain regularity conditions\textsuperscript{12}, a straightforward nonparametric kernel estimator for the unknown function $g(x)$ is:

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

where $K(.)$ is the standard product kernel with bandwidth $h$. 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:

$$
\sup_{x \in S} |\hat{g}(x) - g(x)| = O \left( \frac{(\ln n)^{0.5}}{(nh_1 \cdots h_q)^{0.5}} + \sum_{s=1}^{q} h_s^2 \right), \text{a.s.}
$$

Assuming each bandwidth $(h_s)$ has the same order of magnitude, the optimal choice of $h_s$ that minimizes $MSE[\hat{g}(x)]$ is $h_s \sim n^{-\frac{1}{4+q}}$, 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, $n^n(\hat{\varepsilon} - \varepsilon) = O_p(1)$ 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) consider equation (7) and take \( \{b_n^2\}_{n \in \mathbb{N}} = \frac{1}{2} \text{Var}(\hat{\varepsilon}) \) and \( w_n = k_n \) with \( \{k_n\}_{n \in \mathbb{N}} = \{(\frac{n}{b_n})^\frac{2a}{1+2a}\}_{n \in \mathbb{N}} \) if \( 1 < \delta \leq 3/2; \{k_n\}_{n \in \mathbb{N}} = \{(\frac{n}{b_n})^\frac{2a}{3+2a}\}_{n \in \mathbb{N}} \) if \( \delta > 3/2 \). For any $g \in \mathcal{L}_n$, the proposed

\textsuperscript{12}Details see Condition 2.1 in Li and Racine (2007).
deconvolution kernel density estimator in equation (7) is bounded from above as follows:

\[
\sup_{f_u \in F_u} E_{f,g} \left| \hat{f}_u - f_u \right|^2_{L_2} \leq n^{- \frac{2a(2\delta - 1)}{6 + 2\delta}} \text{ if } 1 < \delta \leq 3/2
\]

and

\[
\sup_{f_u \in F_u} E_{f,g} \left| \hat{f}_u - f_u \right|^2_{L_2} \leq n^{- \frac{2a(2\delta - 1)}{3 + 4\delta}} \text{ if } \delta > 3/2
\]

where \( a = \frac{2}{4+q} \) and \( \delta \) 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 design from Meister (2006) and Horrace and Parmeter (2011) except that we consider performance of the Laplace deconvolution with both Laplace errors (correctly specified) and normal errors (misspecified). We focus on sample sizes of \( n = 500, 1000, \) and 3000 with the linear model:

\[
y_j = 4 + 3x_j + v_j + u_j, \quad j = 1, \ldots, n.
\] (10)

The \( x_j \)'s 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 which the variance is varied to produce several signal-to-noise settings. The \( u_j \)'s are generated from the twice convolved, zero-mean Laplace density for which the probability
density function is $L(x) = (4)^{-1}e^{-|x|(|x| + 1)}$.\(^\text{13}\) We fix the variance of $u$ to 2. In this setting it is known that $C_1 = 1/4$, $\delta = 4$ and $T = 1$.\(^\text{14}\)

Following Theorem 1, we choose $b_n^2 = \frac{1}{2}Var(\hat{\varepsilon})$ where $\hat{\varepsilon}$ is the residual from the first-step ordinary least squares (OLS) estimation, and $k_n = n^{1/7}*(b_n^2)^{4/7} = n^{1/15}*(b_n^2)^{-4/15}$ correspondingly as $\delta = 4 > 3/2$. 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 $Var(u)$ and $Var(v)$: $\gamma := \frac{\sigma_u^2}{\sigma_v^2} \in \{1/2, 1, 2\}$. 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):

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

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 $\{500, 1000, 3000\}$ and compare the proposed estimator (CHP) to the true unknown density (True). The graphical fit of the proposed

---

\(^\text{13}\)This follows from the setting in Meister (2006).

\(^\text{14}\)We are not concerned with $C_2$, since it has no bearing on any calculations for the estimator.
estimator is quite good with only 500 observations (Figure 1). Most of the bias comes from estimation around the mode.\textsuperscript{15} As the sample size increases, the RMISE of the proposed estimator (\textit{CHP}) decreases from 0.0148 (Figure 1) to 0.0142 (Figure 2) and to 0.0125 (Figure 3).

Figure 4-6 show the results for a single draw ($R = 1$) and fixed sample size $n = 1000$ but varying the signal-to-noise ratio $\sigma_u^2/\sigma_v^2=2/1, 2/2, 2/4$. The proposed estimator (\textit{CHP}) works very well when $\sigma_u^2/\sigma_v^2=2/1$ with 1000 observations. As the signal-to-noise ratio decreases, the RIMSE of proposed estimator (\textit{CHP}) increases from 0.0136 (Figure 4) to 0.0142 (Figure 5) and to 0.0180 (Figure 6). Even for the most noisy 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 2 contains detailed results from $R = 500$ simulations with varying sample sizes \{500, 1000, 3000\} 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 3000 (down the rows), demonstrating the consistency of the proposed estimator (\textit{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 $\hat{b}_n^2$ 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.

\textsuperscript{15}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 Errors

To understand the impact of misspecification of the error distribution, we consider the performance of the proposed estimator when the true error 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 ($n = 3000$). 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 $n = 1000$ 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) and 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 3 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.\(^{16}\) 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 $\delta$ are known and the estimated density assuming these parameters are unknown. We restrict the range of $u$ to be $(-5, 5)$ and 128 points are evaluated 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 $n = 1000$ 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

\(^{16}\)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.\footnote{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 128 points in their support. Therefore, the relative fit of the densities with known and unknown parameters will be variable 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}

A more comprehensive analysis is conducted in Figures 14-16. Figure 14 shows the Euclidean distance of the estimated densities: $\Delta = \|\hat{f}_{\text{known}} - \hat{f}_{\text{unknown}}\|^2$, as a function of the smoothness parameters for a single draw ($R = 1$). Figure 15 and Figure 16 show the Euclidean distance between the true density and the estimated density taking the chosen $C_1$ and $\delta$ as known, $\|\hat{f}_{\text{known}} - f_{\text{true}}\|^2$, and unknown, $\|\hat{f}_{\text{unknown}} - f_{\text{true}}\|^2$, 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.

Though it is a useful tool, two caveats are worth mentioning. First, our Laplace de-convolution estimator assumes that the error 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.\footnote{We reduce the replication size from 500 to save computation time.} The results are presented in Table 4.\footnote{We report the RMISE of $\hat{f}_{\text{known}}$ here.} Similar to Table 2, the convergence pattern still holds
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, \( \delta \) and \( C_1 \), based on minimizing the Euclidean distance in Table 4. 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), \( \beta \) 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 \( \varepsilon \) (e.g., Jondrow et al. 1982). Much of the received 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 ar-
ticulated 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).20

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 \( E(u|\varepsilon) \).21

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 and commercial loans and real estate loans; and securities, \( Y_3 \), including all non-loan financial

---

20 The data are publicly available on the Journal of Applied Econometrics website.
21 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 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

$$c_j = \alpha + x_j'\beta + u_j + v_j \quad j = 1, \ldots, n, \quad (12)$$

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, $\hat{e}_j$. The asymmetry of the distribution (skew equals 1.550) suggests non-zero cost inefficiency. Selecting $\delta = 3$ and $C_1 = 1$ and using Theorem 1, the deconvolution estimator yields an estimate of $\sigma_v^2$ equal to 0.0403. A plot of the density estimate, $\hat{f}_u(u)$, is shown as the dashed line (CHP) in Figure 18. Third, using the procedure of Hall and Simar (2002) with a bandwidth of 0.3052,

\footnote{It is interesting to note that with a skew of 1.55, this provides evidence against use of the half-normal distribution.}
\footnote{For the Laplace distribution, $\delta = 2$; for convolved Laplace, $\delta = 4$. The choice $\delta = 3$ is between Laplace and convolved Laplace.}
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.\(^{24}\)

Figure 18 shows all three density estimators for US bank inefficiency in 2005. Notice that even without boundary correction, the deconvolution estimator (CHP) has a thinner right tail than the estimated half normal density (SFA). With boundary correction, the deconvolution estimator (CHP \( E(u)_{bc} \)) performs much better with a smaller \( E(u) \) which 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; 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

\(^{24}\)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 (1995) and Delaigle and Gijbels (2002).

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 (1995), Delaigle and Gijbels (2002) 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 (2002): $\log(fat + 5)$.

To these data we apply 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 observed data assuming there is no measurement error.\footnote{We apply the default package “ksdensity” in matlab for the $ErrorFree$ case.}

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, $\Delta$, over a grid of $\delta \in \{1.25, 1.5, 1.75, 2, 2.25, 2.5, 2.75, 3\}$ and $C_1 \in \{0.1, 0.25, 0.40, 0.55, 0.70, 0.85, 1\}$.\footnote{We also tried larger range of $\delta$ and narrow down to this specific range by searching the minimum of $\Delta$.} Figure 19 shows the surface of the Euclidean distance as a function of the smoothness parameters over the grid. The $\Delta$ increases as $C_1$ rise from 0 to 1 except when $\delta$ is around 0.25.
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 $\delta = 3/2$ 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.\textsuperscript{27} 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.\textsuperscript{28}

The $CHP$ density estimator based on the NHANES III data is quite similar to that of
Delaigle and Gijbels (2002), 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 $\delta$ and $C_1$ are considered
for the $CHP$ estimator: $C_1 = 1$ and $\delta = 3/2$; $C_1 = 1$ and $\delta = 2$; $C_1 = 0.6$ and $\delta = 3/2$;\textsuperscript{27} It seems to violate the independence assumption between the target variable and the measurement error.
\textsuperscript{28} 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.

28
$C_1 = 0.6$ and $\delta = 2$ in Figure 21. The baseline ($C_1 = 1, \delta = 3/2$) 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 $C_1 = 0.6, \delta = 3/2$ (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.\footnote{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 $\delta$. Recall that the $\delta_n^2$ is bound between 0 and $0.5\text{var}(\bar{e})$. When it happens, the deconvolution estimators will be very similar to the $ErrorFree$ estimator.} 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 $\delta$ 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 error term with unknown variance.

The density deconvolution estimator with Laplace error has at least two attractive characteristics for applied researchers: 1) it possesses a faster convergence rate than that of normal
distributed errors (i.e., $O(n^c)$ versus $O((\ln n)^c)$) and 2) it is robust to misspecification of the true underlying error distribution. A third (potential) feature that practitioners may find appealing is the Laplace errors generate different insights than normal errors: 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 interest component’s and noise’s 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 the adaptive function estimation in nonparametric regression with one-sided errors. Another interesting strand in this area is to investigate the distribution of the 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 (2017) applies it with a real data set.

References


A General Appendix

Definition: \( \varepsilon \) is ordinary-smooth of order \( \delta \) (Fan 1991a): 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 \( f \in L^2(\mathbb{R}) \),
\[
\int_{-\infty}^{\infty} |\hat{f}(z)|^2 dz = \int_{-\infty}^{\infty} |f(\tau)|^2 d\tau
\]
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:

\[
\sup_{g \in \mathcal{L}_n} \sup_{f \in \mathcal{F}_u} E_{f,g} \left| \hat{f}_u - f_u \right|_{L^2}^2 = (2\pi)^{-1} \sup_{g \in \mathcal{L}_n} \sup_{f \in \mathcal{F}_u} \left( \int_{-w_n}^{w_n} E_{f,g} \left| e^{-i\tau z} \hat{h}_{\varepsilon}(1 + \hat{b}_n^2 \tau^2) - h_u(\tau) \right|^2 d\tau + \int_{|\tau| > w_n} |e^{-i\tau z} h_u(\tau)|^2 d\tau \right)
\]
\[
\text{Parseval } = (2\pi)^{-1} \sup_{g \in \mathcal{L}} \sup_{f \in \mathcal{F}} \left( \int_{-w}^{w} E_{f,g} |\hat{h}_\varepsilon(1 + \hat{b}_n^2 \tau^2) - h_u(\tau)|^2 d\tau + \int_{|\tau| > w} |h_u(\tau)|^2 d\tau \right)
\]

\[
\leq (2\pi)^{-1} \left( \sup_{g \in \mathcal{L}} \sup_{f \in \mathcal{F}} 2 \int_{-w}^{w} |h_u(\tau)|^2 d\tau + \sup_{g \in \mathcal{L}} \sup_{f \in \mathcal{F}} 2 \int_{-w}^{w} E_{f,g} |(1 + \hat{b}_n^2 \tau^2)(\hat{h}_\varepsilon(\tau) - h_c(\tau))|^2 d\tau + \right)
\]

\[
\sup_{g \in \mathcal{L}} \sup_{f \in \mathcal{F}} 2 \int_{-w}^{w} E_{f,g} |h_c(\tau)|/ \frac{1}{(1 + \hat{b}_n^2 \tau^2)} - h_u(\tau)|^2 d\tau)
\]

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 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 the additional component of 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 contamination is constructed with \(\hat{\varepsilon}\) instead of \(\varepsilon\). 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 the Laplace contamination and nonparametric first-step regression rather than just normal noise for the linear (stochastic frontier) model.

(1) \(B \leq \text{const} \times w_n^{1-2\delta}\) 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,

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

where $a = \frac{2}{4+q}$ for the nonparametric first-step regression and $a = 1/2$ 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. $^{30}$ So

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

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

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

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

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

$^{30}$Basically, there are $2a$ instead of 1 in the power of $n$. 

37
where $V_1 \leq const \cdot (n^{-1}w_n^3)(1 + b_n^2w_n^2)^2$ and $V_2 \leq const \cdot (n^{-1}w_n)(1 + b_n^2w_n^2)^2$.

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

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

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

where

$$E_{f,g} \left\{ \frac{s^2 w_n^2 (\hat{b}_n^2 - b^2)}{1 + b^2 s^2 w_n^2} \right\}^2 = E_{f,g} \left\{ \frac{s^2 w_n^2 (\hat{b}_n^2 - b^2)}{1 + b^2 s^2 w_n^2} \right\}^2 |\hat{b}_n^2 - b^2| \leq d_n \right\} + E_{f,g} \left\{ \frac{s^2 w_n^2 (\hat{b}_n^2 - b^2)}{1 + b^2 s^2 w_n^2} \right\}^2 |\hat{b}_n^2 - b^2| > d_n \right\}$$

$$\leq \left| \frac{s^2 w_n^2 d_n}{1 + s^2 w_n^2 b_n^2} \right|^2 + \left| \frac{s^2 w_n^2 b_n^2}{1 + b^2 w_n^2 s^2} \right|^2 Pr(|\hat{b}_n^2 - b^2| > d_n) \right\} \left( \hat{b}_n^2 \leq b_n^2, b^2 \leq b_n^2 \right)$$

$$\leq \left( \frac{s^2 w_n^2 d_n}{s^2 w_n^2 b_n^2} \right)^2 + \left( \frac{s^2 w_n^2 b_n^2}{s^2 b^2 w_n^2} \right)^2 Pr(|\hat{b}_n^2 - b^2| > d_n) \right\} \left( w_n \to \infty \text{ as } n \to \infty \right)$$

$$\leq \left( \frac{d_n}{b_n^2} \right)^2 + \left( \frac{b_n^2}{b^2} \right)^2 Pr(|\hat{b}_n^2 - b^2| > d_n)$$

$$= const \cdot w_n^{-2} + const \cdot (b_n^2)^2 Pr(|\hat{b}_n^2 - b^2| > d_n)$$

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

38
C Proof of Lemma 2

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

First,

\[
\sup_{g \in L_n} \sup_{f \in F_u} P_{f,g}(|\hat{b}_n^2 - b^2| > d_n) = \sup_{g \in L_n} \sup_{f \in F_u} P_{f,g}(k_n^{-\delta} \frac{C_1}{\hat{h}_\varepsilon(k_n)} - 1) > d_n + b^2
\]

\[
= \sup_{g \in L_n} \sup_{f \in F_u} P_{f,g}(\frac{C_1 k_n^{-\delta}}{\hat{h}_\varepsilon(k_n)} - 1) > d_n k_n^2 + b^2 k_n^2 = P_{f,g}(|\hat{h}_\varepsilon(k_n)| < \frac{C_1 k_n^{-\delta}}{1 + d_n k_n^2 + b^2 k_n^2})
\]

\[
\leq \sup_{g \in L_n} \sup_{f \in F_u} P_{f,g}(|\hat{h}_\varepsilon(k_n)| < \alpha_n \frac{C_1 k_n^{-\delta}}{1 + b^2 k_n^2})
\]

\[
= \sup_{g \in L_n} \sup_{f \in F_u} P_{f,g}(\hat{h}_\varepsilon(k_n) < \alpha_n |h_\varepsilon(k_n)|)
\]

where \( \alpha_n = \frac{1+b^2 k_n^2}{1+d_n k_n^2 + b^2 k_n^2} \), 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 \( \delta \) and \( C_1 \) case and \( d_n = w_n^{-1} = O(ln k_n/k_n) \), \( d_n * k_n^2 = O(ln(k_n)k_n) \) for other cases.\(^{31}\)

So there exists a constant \( c \in (0,1) \) that guaranteed so that above formula is bounded above by \( \sup_{g \in L_n} \sup_{f \in F_u} P_{f,g}(\hat{h}_\varepsilon(k_n) < \alpha_n |h_\varepsilon(k_n)|) \leq \sup_{g \in L_n} \sup_{f \in F_u} P_{f,g}(\hat{h}_\varepsilon(k_n) < c |h_\varepsilon(k_n)|) \) which by Chebyshev’s inequality yields

\[
\leq (1 - c)^{-2} \sup_{g \in L_n} \sup_{f \in F_u} |h_\varepsilon(k_n)|^{-2} E_\varepsilon |\hat{h}_\varepsilon(k_n) - h_\varepsilon(k_n)|^2
\]

\[
\leq 2(1 - c)^{-2} \sup_{g \in L_n} \sup_{f \in F_u} |h_\varepsilon(k_n)|^{-2} [E_\varepsilon |\hat{h}_\varepsilon(k_n) - h_\varepsilon(k_n)|^2 + E_\varepsilon |\hat{h}_\varepsilon(k_n) - h_\varepsilon(k_n)|^2]
\]

\[
\leq 2(1-c)^{-2} \sup_{g \in L_n} \sup_{f \in F_u} |h_\varepsilon(k_n)|^{-2} [E_\varepsilon |O_p(k_n^{-1})\frac{1}{n} \sum_j exp(ik_n \varepsilon_j)|^2 + E_\varepsilon |\frac{1}{n} \sum_j exp(ik_n \varepsilon_j - h_\varepsilon(k_n)|^2]
\]

\(^{31}\)This is discussed in the extension section.
\[ = const \ast (E_1 + E_2) \]

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

Then the second addend can be bounded in a similar way.

\[
\sup_{g \in L_n} \sup_{f \in F_u} P_{f,g}(\hat{b}_n^2 - b^2 < -d_n) = \sup_{g \in L_n} \sup_{f \in F_u} P_{f,g}(k_n^{-2}(\frac{C_1 k_n^{-\delta}}{\hat{h}_\varepsilon(k_n)} - 1) < b^2 - d_n) 
\]

\[
\leq \sup_{g \in L_n} \sup_{f \in F_u} P_{f,g}(\hat{h}_\varepsilon(k_n) > \gamma_n |h_\varepsilon(k_n)|) 
\]

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

So there exists a constant \(C \in (0,1)\) that guaranteed so that above formula is bounded above by \(\sup_{g \in L_n} \sup_{f \in F_u} P_{f,g}(\hat{h}_\varepsilon(k_n) > \gamma_n |h_\varepsilon(k_n)|) \leq \sup_{g \in L_n} \sup_{f \in F_u} P_{f,g}(\hat{h}_\varepsilon(k_n) > C |h_\varepsilon(k_n)|)\) which by Chebyshev’s inequality yields

\[
\leq (C - 1)^{-2} \sup_{g \in L_n} \sup_{f \in F_u} |h_\varepsilon(k_n)|^{-2} E_\varepsilon |\hat{h}_\varepsilon(k_n) - h_\varepsilon(k_n)|^2 
\]

\[
\leq 2(C - 1)^{-2} \sup_{g \in L_n} \sup_{f \in F_u} |h_\varepsilon(k_n)|^{-2} [E_\varepsilon |\hat{h}_\varepsilon(k_n) - \hat{h}_\varepsilon(k_n)|^2 + E_\varepsilon |\hat{h}_\varepsilon(k_n) - h_\varepsilon(k_n)|^2] 
\]

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 can obtain the upper bound of MISE of the density \( \hat{f}_n \) is

\[
\max(B, V, E) = \max \left\{ \text{const}_1 w_n^{1-2\delta}, \text{const}_2 n^{-1} w_n (1 + b_n^2 w_n^2)^2 + \text{const}_3 n^{-1} (1 + b_n^2 w_n^2)^2 w_n^3, \right.
\]
\[
\left. \text{const}_4 w_n^{1-2\delta} w_n^{-2} + \text{const}_5 k_n^{2\delta} b_n^4 (1 + b_n^2 k_n^2) (1 + k_n^2) n^{-1} \right\}
\]

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

(i) If \( 1 < \delta \leq 3/2 \), \( k_n = \frac{1}{25+6} \cdot (b_n^2)^{\frac{1}{25+6}} \). Consequently, a \( n^{\frac{(25-1)}{25+6} \cdot (b_n^2)^{\frac{25-1}{25+6}}} \to D_1 n^{\frac{(25-1)}{25+6}} \) convergence rate is determined by the equality of the first term and the first addend of the third term where \( D_1 = \text{var}(\varepsilon)^{\frac{(25-1)}{25+6}} \).

(ii) If \( \delta > 3/2 \), \( k_n = \frac{1}{12+3} \cdot (b_n^2)^{\frac{1}{12+3}} \). Consequently, a \( n^{\frac{12-1}{12+3} \cdot (b_n^2)^{\frac{12-1}{12+3}}} \to D_2 n^{\frac{12-1}{12+3}} \) convergence rate is determined by the equality of the first term and the second addend of the third term where \( D_2 = \text{var}(\varepsilon)^{\frac{(25-1)}{25+3}} \).

E Proof of Theorem 2

For the case where \( C_1 \) and \( \delta \) are unknown, similar argument applies, \( k_n \) stays the same with guess \( C_1 \) and \( \delta \) since \( w_n = k_n / \ln k_n = O(k_n) \) and the convergence rates are \( n^{\frac{(25-1)}{25+6} \cdot (\ln n)^{\frac{25-1}{25+6}}} \), if \( 1 < \delta \leq 3/2 \) and \( n^{\frac{25-1}{48+3} \cdot (\ln n)^{\frac{25-1}{48+3}}} \), if \( \delta > 3/2 \).\(^{32}\)

\(^{32}\)See the rule-of-thumb adaptive procedure in section 4.1.
F Proof of Theorem 3

When a nonparametric kernel estimation is implemented for the first-step regression, we can easily derive similar Lemmas as Lemma 1 and Lemma 2 for the parametric case as follows:

**Lemma 1′.** For Assumptions 3-5, Condition 2.1 in Li and Racine (2007) and \( L_n = \{\text{Laplace}(0, b) : b^2 \in (0, b_n^2)\} \), the Mean Integrated Squared Error (MISE) of (7) is

\[
\sup_{g \in L_n} \sup_{f \in F_u} E_{f,g} \left\| \hat{f}_u - f_u \right\|_{L_2}^2 \leq B + V + E,
\]

where \( B \leq \text{const}_1 \ast w_n^{1-2\delta} \),

\[
V \leq \text{const}_2 \ast n^{-2a} w_n (1 + b_n^2 w_n^2)^2 + \text{const}_3 \ast n^{-2a} w_n^3 (1 + b_n^2 w_n^2)^2,
\]

\[
E \leq \text{const}_4 \ast \sup_{g \in L_n} \sup_{f \in F_u} \left( w_n \int_1^{\infty} |h_u(s w_n)|^2 \left( \frac{d_n}{s^2} \right)^2 ds + w_n \int_1^{1} |h_u(w_n s)|^2 \frac{h_n^4}{h_n^4} \ast P_{f,g} (|\hat{b}_n^2 - b^2| > d_n) ds \right),
\]

with \( d_n := \frac{1}{w_n} \); \( f \) and \( g \) are the probability density function in distribution family \( F_u \) and \( L_n \) respectively. \( \text{const}_j \) 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 \( \sup_{g \in L_n} \sup_{f \in F_u} P_{f,g} (|\hat{b}_n^2 - b^2| > d_n) \leq \text{const} \ast k_n^{2\delta} (1 + b_n^2 k_n^2)(1 + k_n^2) \ast n^{-2a} \).

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

\[
\max(B, V, E) = \max \left[ \text{const}_1 \ast w_n^{1-2\delta}, \text{const}_2 \ast n^{-2a} w_n (1 + b_n^2 w_n^2)^2 + \text{const}_3 \ast n^{-2a} (1 + b_n^2 w_n^2)^2 w_n^3, \right.
\]

\[
\text{const}_4 \ast w_n^{1-2\delta} w_n^{-2} + \text{const}_5 \ast k_n^{2\delta} b_n^4 (1 + b_n^2 k_n^2)(1 + k_n^2) n^{-2a} \left] \right.
\]

Under Assumption 3, if \( C_1 \) and \( \delta \) are known, then \( w_n = k_n, b_n^2 = 0.5 \text{var}(\hat{\varepsilon}) \), and mini-
mizing the above maximum leads to

(i) If $1 < \delta \leq 3/2$, $k_n = n^{\frac{2a}{23+6}} \cdot (b_n^2)^{\frac{1}{23+6}}$. Consequently, a $n^{-\frac{2a(2\delta-1)}{23+6}} \cdot (b_n^2)^{\frac{2a(2\delta-1)}{23+6}} \rightarrow D_1 \cdot n^{-\frac{2a(2\delta-1)}{23+6}}$ convergence rate is determined by the equality of the first term and the first addend of the third term where $D_1 = var(\varepsilon)^{\frac{2a(2\delta-1)}{23+6}}$.

(ii) If $\delta > 3/2$, $k_n = n^{\frac{2a}{4\delta+3}} \cdot (b_n^2)^{-\frac{1}{4\delta+3}}$. Consequently, a $n^{-\frac{2a(2\delta-1)}{4\delta+3}} \cdot (b_n^2)^{\frac{2a(2\delta-1)}{4\delta+3}} \rightarrow D_2 \cdot n^{-\frac{2a(2\delta-1)}{4\delta+3}}$ convergence rate is determined by the equality of the first term and the second addend of the third term where $D_2 = var(\varepsilon)^{\frac{2a(2\delta-1)}{4\delta+3}}$.
### Table 1: Smoothing Parameters of Some Popular Continuous Distributions

<table>
<thead>
<tr>
<th>Name</th>
<th>Parameter</th>
<th>Density</th>
<th>Chara. Function</th>
<th>Smoothness Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symm. Uniform</td>
<td>$a &gt; 0$</td>
<td>$\frac{1}{2a}1_{[-a,a]}(x)$</td>
<td>$\frac{\sin(at)}{at}$</td>
<td>$0^*$ 1 1</td>
</tr>
<tr>
<td>Laplace</td>
<td>$b &gt; 0$</td>
<td>$\frac{1}{\pi b^2} e^{-\frac{x^2}{b^2}}$</td>
<td>$\frac{</td>
<td>\cos(\delta) - \cos(\alpha)</td>
</tr>
<tr>
<td>Uniform</td>
<td>$a,b (b &gt; a)$</td>
<td>$\frac{1}{\pi (b-a)} 1_{[a,b]}(x)$</td>
<td>$\frac{\sinh(a)}{a b}$</td>
<td>$\frac{1}{b^2} 1$</td>
</tr>
<tr>
<td>$\chi^2_k$</td>
<td>$k &gt; 0$</td>
<td>$\frac{1}{(1/b^2)^{k/2}} 1^{k-1/2} e^{-\frac{x^2}{b^2}}$</td>
<td>$\frac{</td>
<td>\Gamma(k - 1/2) - \Gamma(k + 1/2)</td>
</tr>
<tr>
<td>Gamma</td>
<td>$k &gt; 0, \theta &gt; 0$</td>
<td>$\frac{1}{(1/\theta)^{k}} e^{-\frac{x^2}{\theta}}$</td>
<td>$\frac{1}{</td>
<td>\theta</td>
</tr>
<tr>
<td>Twice-convolved Laplace</td>
<td>$b &gt; 0$</td>
<td>$\frac{1}{\pi b^2} e^{-\frac{</td>
<td>x</td>
<td>}{b} + b}$</td>
</tr>
<tr>
<td>Cauchy</td>
<td>$\mu = 0, \theta &gt; 0$</td>
<td>$\frac{1}{\pi \theta^2} e^{-\frac{</td>
<td>x</td>
<td>^2}{\theta^2}}$</td>
</tr>
<tr>
<td>Normal</td>
<td>$\mu = 0, \sigma^2 &gt; 0$</td>
<td>$\frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$</td>
<td>$e^{-\frac{x^2}{2\sigma^2}}$</td>
<td>NA NA $\infty$</td>
</tr>
</tbody>
</table>

Notes: The ordinary-smoothness parameter are defined by the Fan(1991a): $C_1 |\tau| - \delta \leq |h_x(\tau)| \leq C_2 |\tau| - \delta$ 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: RMISE for Laplacian Noise Deconvolution

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\sigma_u^2/\sigma_v^2 = 2/1$</th>
<th>$\sigma_v^2/\sigma_u^2 = 2/2$</th>
<th>$\sigma_u^2/\sigma_v^2 = 2/4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>0.0162</td>
<td>0.0155</td>
<td>0.0204</td>
</tr>
<tr>
<td>1000</td>
<td>0.0150</td>
<td>0.0143</td>
<td>0.0197</td>
</tr>
<tr>
<td>3000</td>
<td>0.0138</td>
<td>0.0126</td>
<td>0.0190</td>
</tr>
</tbody>
</table>

Notes: Replication 500 times. $\frac{\sigma_u^2}{\sigma_v^2}$ stands for the signal-to-noise ratio
Table 3: RMISE under Misspecification: Normal Noise Deconvolution

<table>
<thead>
<tr>
<th>n</th>
<th>$\sigma_u^2/\sigma_v^2 = 2/1$</th>
<th>$\sigma_u^2/\sigma_v^2 = 2/2$</th>
<th>$\sigma_u^2/\sigma_v^2 = 2/4$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CHP</td>
<td>Meister06</td>
<td>CHP</td>
</tr>
<tr>
<td>500</td>
<td>0.0155</td>
<td>0.0128</td>
<td>0.0186</td>
</tr>
<tr>
<td>1000</td>
<td>0.0143</td>
<td>0.0116</td>
<td>0.0168</td>
</tr>
<tr>
<td>3000</td>
<td>0.0129</td>
<td>0.0108</td>
<td>0.0152</td>
</tr>
</tbody>
</table>

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

Table 4: Simulation by Rule-of-Thumb Adaptive Procedure with Laplace Noise

<table>
<thead>
<tr>
<th>N</th>
<th>$\sigma_u^2/\sigma_v^2 = 2/1$</th>
<th>$\sigma_u^2/\sigma_v^2 = 2/2$</th>
<th>$\sigma_u^2/\sigma_v^2 = 2/4$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMISE</td>
<td>Ave. $\delta$</td>
<td>Ave. $C_1$</td>
</tr>
<tr>
<td>500</td>
<td>0.0139</td>
<td>2.02</td>
<td>0.10</td>
</tr>
<tr>
<td>1000</td>
<td>0.0125</td>
<td>2.00</td>
<td>0.10</td>
</tr>
<tr>
<td>3000</td>
<td>0.0110</td>
<td>2.00</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Notes: Replication 100 times. $\sigma_u^2/\sigma_v^2$ stands for the signal-to-noise ratio.
Figure 1: Laplace Deconvolution (CHP): $n = 500, \sigma_u^2 / \sigma_v^2 = 1$ 

Figure 2: Laplace Deconvolution (CHP): $n = 1000, \sigma_u^2 / \sigma_v^2 = 2/2$ 

Figure 3: Laplace Deconvolution (CHP): $n = 3000, \sigma_u^2 / \sigma_v^2 = 2/2$
Figure 4: Laplace Deconvolution (CHP): $n = 1000, \sigma_u^2/\sigma_v^2 = 2/1$

Figure 5: Laplace Deconvolution (CHP): $n = 1000, \sigma_u^2/\sigma_v^2 = 2/2$

Figure 6: Laplace Deconvolution (CHP): $n = 1000, \sigma_u^2/\sigma_v^2 = 2/4$
Figure 7: Misspecified Laplace (CHP) Deconvolution: \( n = 500, \sigma_u^2/\sigma_v^2 = 2/2 \)

Figure 8: Misspecified Laplace (CHP) Deconvolution: \( n = 1000, \sigma_u^2/\sigma_v^2 = 2/2 \)

Figure 9: Misspecified Laplace (CHP) Deconvolution: \( n = 3000, \sigma_u^2/\sigma_v^2 = 2/2 \)
Figure 10: Misspecified Laplace (CHP) Deconvolution: $n = 1000, \sigma_u^2 / \sigma_v^2 = 2/1$

Figure 11: Misspecified Laplace (CHP) Deconvolution: $n = 1000, \sigma_u^2 / \sigma_v^2 = 2/2$

Figure 12: Misspecified Laplace (CHP) Deconvolution: $n = 1000, \sigma_u^2 / \sigma_v^2 = 2/4$
Figure 13: Deconvolution with Unknown Smooth Parameters, \( n = 1000, \frac{\sigma_u^2 / \sigma_v^2}{2/2} \)
Figure 14: Euclidean Distance Between $\hat{f}_{\text{unknown}}$ and $\hat{f}_{\text{known}}$

Figure 15: Euclidean Distance Between $\hat{f}_{\text{known}}$ and True Density
Figure 16: Euclidean Distance Between $\hat{f}_{\text{unknown}}$ 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 $\delta$