Deconvolution in Nonparametric Statistics

Kris De Brabanter[1,2] and Bart De Moor[1,2] *

1- Katholieke Universiteit Leuven - Department of Electrical Engineering (ESAT-SCD)
Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

2- Katholieke Universiteit Leuven - IBBT-KU Leuven Future Health Department
Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

Abstract. In this tutorial paper we give an overview of deconvolution problems in nonparametric statistics. First, we consider the problem of density estimation given a contaminated sample. We illustrate that the classical Rosenblatt-Parzen kernel density estimator is unable to capture the full shape of the density while the presented method experiences almost no problems. Second, we use the previous estimator in a nonparametric regression framework with errors-in-variables.

1 Introduction

Deconvolution problems occur in many fields of nonparametric statistics, for example, density estimation based on contaminated data [1], nonparametric regression with errors-in-variables [2], image and signal deblurring [3]. During the last decades, these topics have received considerable attention. As applications of deconvolution procedures concern many real-life problems in econometrics, biometrics, medical statistics and image reconstruction. On the other hand, some rigorous results from Fourier analysis, functional analysis and probability theory are required to understand the construction of deconvolution techniques and their properties.

The general problem of deconvolution in statistics can be described as follows: Our goal is to estimate a function f while an empirical access is restricted to some quantity

$$z = f * G = \int f(x - y)\, dG(y), \tag{1}$$

that is, the convolution of f and some probability distribution G. Hence, the function f can be estimated from some observations only indirectly. The strategy is estimating z first; this means producing an empirical version $\hat{z}$ of z followed by a deconvolution procedure to $\hat{z}$ to estimate f. Therefore, we have to invert the convolution operator with G where some regularization is required to guarantee that $\hat{z}$ is contained in the invertibility domain of the deconvolution operator.

*BDM is full professor at the Katholieke Universiteit Leuven, Belgium. Research supported by Onderzoeksfonds KU Leuven/Research Council KUL: GOA/11/05 Ambiorics, GOA/10/09 MaNet , CoE EF/05/006 Optimization in Engineering (OPTEC) en PFV/10/002 (OPTEC), IOF-SCORES4CHEM, several PhD/postdoc & fellow grants; Flemish Government:FWO: PhD/postdoc grants, projects: G0226.06 (cooperative systems and optimization), G0321.06 (Tensors), G.0302.07 (SVM/Kernel), G.0320.08 (convex MPC), G.0558.08 (Robust MHE), G.0557.08 (Glycemia2), G.0588.09 (Brain-machine) research communities (WOG: ICCoS, ANMMM, MLDM); G.0377.09 (Mechatronics MPC) IWT: PhD Grants, Eureka-Flite+, SBO LeCoPro, SBO Climaqs, SBO POM, O&O-Dsquare Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, Dynamical systems, control and optimization, 2007-2011);IBBT EU: ERNSI; FP7-HD-MPC (INFSO-ICT-223854), COST intelliCIS, FP7-EMBOCON (ICT-248940), FP7-SADCO (MC ITN-264735), ERC HIGHWIND (259 166) Contract Research: AMINAL Other: Helmholtz: viCERP ACCM. The scientific responsibility is assumed by its authors.

The estimator $\hat{z}$ has to be chosen with respect to the statistical experiment. Obviously, to ensure that the specific convolution operator is known, we have to assume that the distribution G is known. Although not realistic in many practical problems, full knowledge of G is assumed in the classical deconvolution approaches [4]. Then, G may be used in the construction of deconvolution estimators. Recent approaches relaxing the exact knowledge of G can be found in [5, 6]. Nevertheless, as one faces troubles of identifiability in problems with unknown G, either more restrictive conditions on f [6], additional data [7] or repeated measurements are required [6]. While there are discrete deconvolution problems [8] where all probability mass of G is concentrated on a finite set, all problems discussed in this tutorial paper deal with continuous convolution models, where G has a density function g in the Lebesgue sense. Then, g is called the error density or blurring density, according to the corresponding model

$$z = f * g = \int f(x-y)g(y)\,dy.$$

Then, the integral is to be understood in the Lebesgue sense; and f and g are real-valued functions mapping into $\mathbb{R}$.

Let us roughly explain why Fourier methods are very popular in deconvolution problems. The Fourier transform of a distribution G, defined by

$$\mathcal{G}(t) = \int \exp(itx)\,dG(x), \quad t \in \mathbb{R}$$

and also the Fourier transform of a function f (not necessarily a density function), defined in the same way when $dG(x)$ is replaced by $f(x)dx$, are utilized. Throughout this tutorial paper, the Fourier transform is denoted by $\mathcal{G}$ and $\mathcal{F}$, respectively. Using the Fourier transform is motivated by the fact that it changes convolution into simple multiplication. More concretely, (1) is equivalent with

$$\mathcal{Z} = \mathcal{F} \cdot \mathcal{G}.$$

It is now clear that the construction of f from z just becomes dividing $\mathcal{Z}$ (empirically accessible) by $\mathcal{G}$ in the Fourier domain. As rough guidelines we give the following scheme for the construction of deconvolution estimators:

1. Estimate $\mathcal{Z}$ based on empirical information, denoted by $\hat{\mathcal{Z}}$.
2. Calculate $\hat{\mathcal{Z}}$ and divide it by $\mathcal{G}$, leading to $\hat{\mathcal{F}}$.
3. Regularize $\hat{\mathcal{F}}$ so that the inverse Fourier transform $\hat{f}$ exists.

However, this scheme seems straightforward but the mathematical effort for the regularization must not be underestimated.

This paper is organized as follows. Section 2 discusses the estimation of a density via kernel methods when only a contaminated sample from the density is available. Section 3 describes the errors-in-variables problem. Section 4 presents a short overview regarding the current research in this area. Section 5 states the conclusions.

2 Density deconvolution

2.1 Assumptions and general estimation procedure

In many real-life situations, direct data are not available since measurement error occurs. Then, we observe the contaminated data $Y_1, \ldots, Y_n$ instead of the true data $X_1, \ldots, X_n$. The elementary model of noisy data is the additive measurement error, that is, any empirical access is restricted to the data $Y_1, \ldots, Y_n$ with

$$Y_j = X_j + \varepsilon_j, \quad j \in \{1, \ldots, n\}$$

instead of the incorrupted independent and identically distributed (i.i.d.) random variables $X_1, \ldots, X_n$. Nevertheless, our goal is still to estimate the density f of the incorrupted, but unobserved random variable X. The i.i.d. random variables $\varepsilon_1, \ldots, \varepsilon_n$ represent the error or the contamination of the data; the density of the random variables ε, consequently called error density, is denoted by g. Further, we assume that X_j and ε_j are real valued, independent and that $\mathbf{E}(\varepsilon_j|X_j) = 0$ and $\mathbf{Var}(\varepsilon_j|X_j) < \infty$.

Following the classical approach to this model, g is assumed to be exactly known; although in many real-life situations this condition cannot be justified. However, in most practical applications, we are able to estimate the error density g from replicated measurements. Discussion on the case of unknown g is a current topic of research in the area of nonparametric statistics and therefore we will give a brief summary in Section 4 so, in what follows, we assume that g is perfectly known so that we can fully concentrate on the deconvolution step itself. It is an elementary result of probability theory that the density of the sum of two independent random variables is equal to the convolution of the densities of both addends. Hence,

$$z = f * g = \int f(x-y)g(y)\,dy,$$

where z denotes the density of the observation Y. Since any direct empirical access is restricted to z, the problem is included in the general deconvolution framework.

Following the general deconvolution scheme in the introduction, the first step is estimating the density z of the observations Y_j. But we see that the main intention at this stage is making the Fourier transform $\mathcal{Z}$ empirically accessible. As the characteristic function Ψ_Y of a random variable Y is just the Fourier transform of the density of Y, we have

$$\mathcal{Z}(t) = \int \exp(itx)z(x)\,dx = \mathbf{E}\exp(itY) = \Psi_Y(t).$$

By replacing the expectation by averaging with respect to the i.i.d. data yields

$$\hat{\Psi}_Y(t) = \frac{1}{n}\sum_{j=1}^{n} \exp(itY_j).$$

There is a simple multiplicative link between $\mathcal{Z}$ and $\mathcal{F}$ because

$$\begin{aligned}\mathcal{Z}(t) &= \Psi_Y(t) = \mathbf{E}\exp(it(X+\varepsilon)) = \mathbf{E}[\exp(itX)\exp(it\varepsilon)] \\ &= \mathbf{E}\exp(itX)\cdot\mathbf{E}\exp(it\varepsilon) = \Psi_X(t)\cdot\Psi_\varepsilon(t) = \mathcal{F}(t)\cdot\mathcal{G}(t),\end{aligned}$$

where the independence of X and ε implies the independence of $\exp(itX)$ and $\exp(it\varepsilon)$. It would be reasonable to consider

$$\hat{\Psi}_X(t) = \frac{1}{n}\frac{\sum_{j=1}^n \exp(itY_j)}{\mathcal{G}(t)}$$

as an estimator of $\mathcal{F}(t)$ assuming $\mathcal{G}$ is bounded away from zero. It is easy to show that this estimator is unbiased and consistent (by the strong law of large numbers). Hence, by taking the inverse Fourier transform, a naive estimator of f is

$$\hat{f}_{\text{naive}}(x) = \frac{1}{2\pi}\int \exp(-itx)\hat{\Psi}_X(t)\,dt,$$

where the integral is taken over the whole real line. However, the estimator $\hat{f}_{\text{naive}}$ is not well-defined as $\hat{\Psi}_X$ is neither integrable nor square integrable over $\mathbb{R}$. Unlike its true counterpart $\mathcal{F}$ to be estimated, which is square-integrable whenever f is square-integrable, due to Parseval's identity. Apparently, for large $|t|$, $\hat{\Psi}_X(t)$ is no good estimator for $\mathcal{F}(t)$ as the tail behavior is significantly different. Therefore, there is some necessity to regularize $\hat{\Psi}_X$ before the Fourier inversion is employed.

2.2 Rozenblatt-Parzen kernel density deconvolution estimator

One of the most well-known methods for density estimation based on direct data (i.e., the error-free case) is the Rosenblatt-Parzen kernel density estimator [9], defined by

$$\hat{z}(x) = \frac{1}{nh}\sum_{j=1}^n K\left(\frac{x-Y_j}{h}\right), \tag{2}$$

with kernel function $K : \mathbb{R} \to \mathbb{R}^+$ and a bandwidth parameter $h > 0$. If $K \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$, that is, the intersection of the sets of all absolutely or square integrable functions over the whole real line, respectively, in Lebesgue sense, the estimator $\hat{z}$ also lies in $L_1(\mathbb{R})\cap L_2(\mathbb{R})$ almost surely so that its Fourier transform exist. It is given by

$$\begin{aligned}\mathcal{Z}(t) &= \frac{1}{nh}\int\sum_{j=1}^n \exp(itx)K\left(\frac{x-Y_j}{h}\right)dx = \frac{1}{nh}\sum_{j=1}^n\int \exp(itx)K\left(\frac{x-Y_j}{h}\right)dx \\ &= \frac{1}{n}\sum_{j=1}^n\int\exp(it(uh+Y_j))K(u)du = \frac{1}{n}\sum_{j=1}^n\exp(itY_j)\int\exp(ithu)K(u)du \\ &= \hat{\Psi}_Y(t)\cdot\mathcal{K}(th),\end{aligned}$$

by using $u = (x - Y_j)/h$ and let $\mathcal{K}$ denote the Fourier transform of the kernel K. Then,

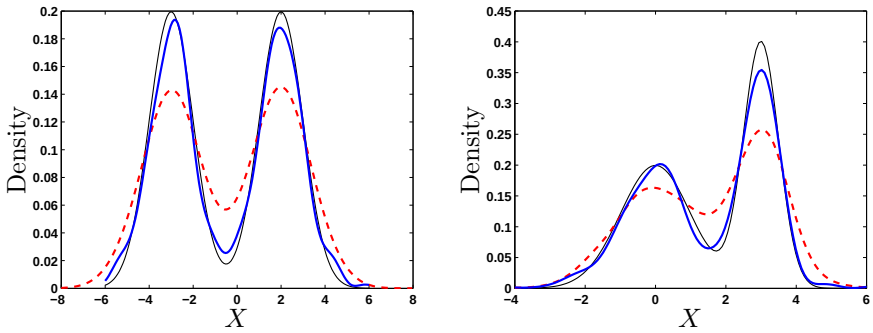
$$\hat{\Psi}_X(t) = \frac{\hat{\Psi}_Y(t)\mathcal{K}(th)}{\mathcal{G}(t)} \tag{3}$$

is a second empirical version for $\mathcal{F}$. There are kernel functions whose Fourier transforms are bounded and compactly supported, for example, the de la Vallée Poussin kernel $K(x) = (1/2\pi)[\sin(\frac{1}{2}x)/\frac{1}{2}x]^2$ having Fourier transform $\mathcal{K}(t) = \max(1 - |t|, 0)$. It can be shown that both integrability and square-integrability of (3) hold. Hence, the inverse Fourier transform can be applied to (3) leading to the estimator

$$\hat{f}(x) = \frac{1}{2\pi} \int \exp(-itx)\mathcal{K}(th) \frac{\frac{1}{n}\sum_{j=1}^{n} \exp(itY_j)}{\mathcal{G}(t)}\, dt, \tag{4}$$

which is well-defined for any nonvanishing $\mathcal{G}$ whenever $\mathcal{K}$ is compactly supported. This in contrast to the naive estimator $\hat{f}_{\text{naive}}$. The estimator (4) has become known as the standard deconvolution kernel density estimator (see [10, 4, 11] for a rigorous study of this estimator).

Example 1 *To illustrate the effect of a contaminated sample on density estimation we consider the following example for 1000 data points. We applied the classical Rosenblatt-Parzen kernel density estimator and the estimator* (4) *to simulated examples from two densities f_X: (1) $X \sim 0.5N(-3, 1^2) + 0.5N(2, 1^2)$ and (2) $0.5N(0, 1^2) + 0.5N(3, (1/2)^2)$. The error density g is Laplace distributed, $\mathcal{L}(\mu, b)$, with location parameter $\mu = 0$ and scale parameter $b = 0.5$. The Fourier transform of the error density is $\mathcal{G}(t) = 4/(4 + t^2)$. For the two examples we have chosen de la Vallée Poussin kernel.*



(a) (b)

Fig. 1: (a), (b) Effect of a contaminated sample on density estimation for two normal mixtures. The thine line is the true density, bold line is the estimated density based on (4) and bold dashed line represents the estimate based on the Rosenblatt-Parzen kernel density estimator.

3 Nonparametric regression with errors-in-variables

3.1 Errors-in-variables problem formulation

As a broad field in statistics in general, the investigation of the link or the dependence between some quantity, which is affected by random noise, and some circumstances, under which the quantity is observed, is referred to as regression. We assume that those circumstances may be represented by a real number X, which is called the covariate or the independent variable. In the standard nonparametric measurement error model, we assume that the covariates X can only be observed with some additive independent noise. Therefore, we change the observation scheme into the i.i.d. dataset $(W_1, Y_1), \ldots, (W_n, Y_n)$, where

$$W_j = X_j + \delta_j \qquad \text{and} \qquad Y_j = m(X_j) + \varepsilon_j \quad \text{for } j = 1, \ldots, n \tag{5}$$

and m is the regression function. The covariate errors δ_j are i.i.d. unobservable random variables having error density g. Note that they are different from the regression errors ε_j. The δ_j are stochastically independent of the X_j and the Y_j. As in the previous section on density deconvolution, we assume that g is known in the standard setting, while the distribution of the ε_j need not be known.

3.2 Kernel regression with errors-in-variables

In case the covariates are not affected by contamination, the Nadaraya-Watson estimator [12, 13] is a well-known kernel regression estimator. It is defined as follows

$$\hat{m}(x) = \sum_{j=1}^{n} \frac{K\left(\frac{x - X_j}{h}\right) Y_j}{\sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right)}, \tag{6}$$

with $K : \mathbb{R} \to \mathbb{R}$ and bandwidth $h > 0$. As we may equivalently multiply the numerator and the denominator by $1/(nh)$, we realize the close relation to the kernel density estimator (2). Indeed, the kernel density estimator of f_X based on the i.i.d. data $X_1, \ldots, X_n$ occurs as the denominator of $\hat{m}$.

We can now focus on extending the Nadaraya-Watson estimator to our contaminated data $(W_1, Y_1), \ldots, (W_n, Y_n)$. The denominator of the Nadaraya-Watson estimator may be replaced by the deconvolution kernel density estimator (4) using the data $W_1, \ldots, W_n$, which are additively corrupted by unobservable random variables with density g. Then the denominator is an empirical version of the density f_X as in the error-free setting. We must also alter the numerator of the Nadaraya-Watson estimator so it does not require knowledge of the unobservable data $X_1, \ldots, X_n$ but only uses the data $W_1, \ldots, W_n$. This can be done as follows. The kernel deconvolution estimator (4), based on the data $W_1, \ldots, W_n$, can be written as

$$\frac{\sum_{j=1}^{n} \int \exp(-itx) \mathcal{K}(th) \frac{\exp(itW_j)}{\mathcal{G}(t)} dt}{2n\pi} = \frac{\sum_{j=1}^{n} \int \exp\left[-i\left(\frac{x - W_j}{h}\right) u\right] \frac{\mathcal{K}(u)}{\mathcal{G}(\frac{u}{h})} du}{2\pi n h},$$

by using the substitution $u = th$. Then, the latter can be written as

$$\hat{f}(x) = \frac{1}{nh} \sum_{j=1}^{n} H\left(\frac{x - W_j}{h}\right),$$

where

$$H(x) = \frac{1}{2\pi} \int \exp(-ixu) \frac{\mathcal{K}(u)}{\mathcal{G}(\frac{u}{h})}\, du. \tag{7}$$

By appealing to (4), (6) and (7), the following kernel regression estimator involving errors-in-variables is given by [14]

$$\hat{m}(x) = \frac{\frac{1}{nh} \sum_{j=1}^{n} H\left(\frac{x - W_j}{h}\right) Y_j}{\frac{1}{2n\pi} \sum_{j=1}^{n} \int \exp(-itx) \mathcal{K}(th) \frac{\exp(itW_j)}{\mathcal{G}(t)}\, dt}. \tag{8}$$

The steps in deriving the Nadaraya-Watson kernel regression estimator involving errors-in-variables boil down to replacing the unobserved $K\left(\frac{x-X_j}{h}\right)$ by an observable quantity $H\left(\frac{x-W_j}{h}\right)$ satisfying (see also [15])

$$\mathbf{E}\left[H\left(\frac{x - W_j}{h}\right) | X_j\right] = K\left(\frac{x - X_j}{h}\right).$$

In the usual nomenclature of measurement error models, this simply means that $H\left((x - W_j)/h\right)$ is an unbiased score for the kernel function $K\left((x - X_j)/h\right)$.

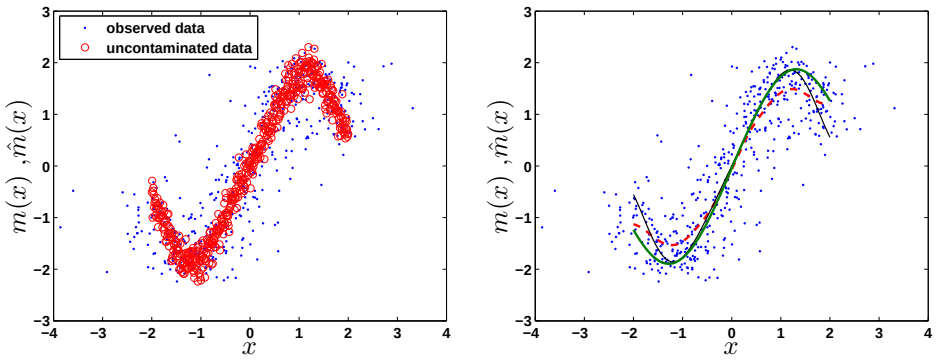
Example 2 *To illustrate the effect of a contaminated sample on regression estimation via model* (5)*, we consider the following example for 350 equispaced data points. We consider the following regression function*

$$m(x) = 2x \exp(-10x^4/81) \quad \textit{with} \quad x \in [-2, 2]$$

with $\varepsilon \sim N(0, 0.2^2)$*,* $\delta \sim \mathcal{L}(0, 0.2)$ *and the* δ_j *are independent of the* (X_j, Y_j)*. Figure 2 shows the result for the artificial data set and illustrate the difference between the observed and uncontaminated data.*

Remark 1 (Berkson regression) *There is another errors-in-variables problem in nonparametric regression, which is closely related to the model* (5)*, but not identical. In literature, it is usually referred to as the Berkson regression model. It was first mentioned in the paper of [16] which had been published before conventional nonparametric techniques such as kernel smoothing were introduced. The main difference between the Berkson model and* (5) *concerns the fact that, in the Berkson context, the covariate is affected by additive noise after it was measured. In the Berkson model, we observe the i.i.d. data* $(X_1, Y_1), \ldots, (X_n, Y_n)$ *where*

$$Y_j = m(X_j + \delta_j) + \varepsilon_j, \quad j = 1, \ldots, n.$$



(a) (b)

Fig. 2: (a) Difference between the observed and uncontaminated data;(b) Effect of a contaminated sample on regression estimation. The thine line is the true regression function, the bold line is the regression estimate based on (8) and the bold dashed line represents the regression estimation using the standard Nadaraya-Watson estimator assuming the error-free case.

4 Current state-of-the-art

The current state-of-the-art regarding deconvolution methods is to relax the assumption that the error density g has to be known. Several approaches exist to estimate this density. A first approach is based on additional data. This means that the error density g is unknown but can be estimated directly from i.i.d. data $\varepsilon'_1, \ldots, \varepsilon'_n$, which are collected in a separate independent experiment. This model was studied in [7, 17] and in the book [18]. Of course, its applicability is restricted to cases where the system of measurement can be calibrated somehow. In particular, the model should be considered when, in some cases, the same individual or quantity can be observed both in an error-free way (call this measurement $X'_{j,1}$) and by a measurement procedure, which is affected by nonnegligible noise (denote this observation by $X'_{j,2}$). Then put $\varepsilon'_j = X'_{j,2} - X'_{j,1}$ where ε'_j is indeed a direct observation from the error distribution. The previously estimated error density may be employed in the deconvolution step for that latter dataset.

A second approach is based on replicated measurements. Here, the same incorrupted but unobserved random variable X_j is independently measured for several times, but each measurement is affected by error. Suppose we observe data $Y_{j,k}$, $j = 1, \ldots, n$ and $k = 1, \ldots, m_j$ with $m_j \geq 2$ defined by

$$Y_{j,k} = X_j + \varepsilon_{j,k}.$$

Then each $\varepsilon_{j,k}$ has error density g. Consider the accessible differences (set $m_j = 2$)

$$\Delta Y_j = Y_{j,1} - Y_{j,2} = \varepsilon_{j,1} - \varepsilon_{j,2}, \quad j \in \{1, \ldots, n\}.$$

It can be shown that characteristic function of ΔY_j yields

$$\psi_{\Delta Y_j}(t) = \mathbf{E}\exp(it\Delta Y_j) = |\mathcal{G}(t)|^2.$$

Hence, under some conditions on $\mathcal{G}$, a reasonable estimate would be

$$\hat{\mathcal{G}}(t) = \left|\frac{1}{n}\sum_{j=1}^{n}\exp(it\Delta Y_j)\right|^{1/2}$$

as an estimator for $\mathcal{G}(t)$. Such approaches have been studied in [19] and [15].

Other topics within this area such as model selection methods [1], optimal kernels for deconvolution [20] and establishing all its statistical properties is an active research area. But perhaps one of the most challenging part is to implement all these methods in a numerically stable way, see e.g. [8] and [6].

5 Conclusion

In this tutorial paper we have given an overview of deconvolution problems in statistics. We have illustrated that classical kernel density and regression estimation suffer when the independent variables are contaminated with measurement error. This is due to the fact the classical estimators inherently assume an error-free independent variable. In order to allow that the estimator can deal with these measurement errors or error-in-variables, the classical estimators need to be modified. The key method in the deconvolution approach is the Fourier transform.

References

[1] A. Delaigle and I. Gijbels. Data-driven boundary estimation in deconvolution problems. *Computational Statistics & Data Analysis*, 50(8):1965–1994, 2006.

[2] A. Delaigle and A. Meister. Rate-optimal nonparametric estimation in classical and Berkson errors-in-variables. *Journal of Statistical Planning and Inference*, 141(1):102–114, 2011.

[3] P. Qiu. *Image Processing and Jump Regression Analysis*. John Wiley & Sons, New York, 2005.

[4] L. Devroye. Consistent deconvolution in density estimation. *The Canadian Journal of Statistics*, 17(2):235–239, 1989.

[5] A. Delaigle. An alternative view of the deconvolution problem. *Statistica Sinica*, 18(3):1025–1045, 2008.

[6] A. Meister. *Deconvolution Problems in Nonparametric Statistics*. Springer-Verlag, Berlin Heidelberg, 2009.

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

[8] P. Hall and P. Qiu. Discrete-transform approach to deconvolution problems. *Biometrika*, 92(1):135–148, 2005.

[9] E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.

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

[11] L.A. Stefanski and R.J. Carroll. Deconvoluting kernel density estimators. *Statistics*, 21:169–184, 1990.

[12] E.A. Nadaraya. On estimating regression. *Theory of Probability and its Applications*, 9(1):141–142, 1964.

[13] G.S. Watson. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 26(4):359–372, 1964.

[14] J. Fan and Y.K. Truong. Nonparametric regression with errors in variables. *The Annals of Statistics*, 21(4):1900–1925, 1993.

[15] A. Delaigle, J. Fan, and R.J. Carroll. A design-adaptive local polynomial estimator for the errors-in-variables problem. *Journal of the American Statistical Association*, 104(485):348–359, 2009.

[16] J. Berkson. Are there two regression problems? *Journal of the American Statistical Association*, 45:164–180, 1950.

[17] M.H. Neumann. On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics*, 7:307–330, 1997.

[18] S. Efromovich. *Nonparametric Curve Estimation: Methods, Theory and Applications.* Springer-Verlag, 1999.

[19] J.L. Horowitz and M. Markatou. Semiparametric estimation of regression models for panel data. *The Review of Economic Studies*, 63:145–168, 1996.

[20] A. Delaigle and P. Hall. On optimal kernel choice for deconvolution. *Statistics & Probability Letters*, 76:1594–1602, 2006.