Journal of Econometrics 168 (2012) 259–269

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Well-posedness of measurement error models for self-reported data☆

Yonghong An[a,*], Yingyao Hu[b]

[a] University of Connecticut, United States
[b] Johns Hopkins University, United States

ARTICLE INFO

Article history:
Received 18 December 2009
Received in revised form
26 July 2011
Accepted 18 January 2012
Available online 11 February 2012

JEL classification:
C14

Keywords:
Well-posed
Ill-posed
Inverse problem
Fredholm integral equation
Deconvolution
Rate of convergence
Measurement error model
Self-reported data
Survey data

ABSTRACT

This paper considers the widely admitted ill-posed inverse problem for measurement error models: estimating the distribution of a latent variable X^* from an observed sample of X, a contaminated measurement of X^*. We show that the inverse problem is well-posed for self-reporting data under the assumption that the probability of truthful reporting is nonzero, which is supported by empirical evidences. Comparing with ill-posedness, well-posedness generally can be translated into faster rates of convergence for the nonparametric estimators of the latent distribution. Therefore, our optimistic result on well-posedness is of importance in economic applications, and it suggests that researchers should not ignore the point mass at zero in the measurement error distribution when they model measurement errors with self-reported data. We also analyze the implications of our results on the estimation of classical measurement error models. Then by both a Monte Carlo study and an empirical application, we show that failing to account for the nonzero probability of truthful reporting can lead to significant bias on estimation of the latent distribution.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Empirical studies in microeconomics usually involve survey samples, where personal information is reported by the interviewees themselves, and therefore, the corresponding variables in the samples are subject to measurement errors. The measurement error problem can be summarized as estimating the distribution of a latent variable X^*, $f_{X^*}(\cdot)$, from an observed sample of X, a contaminated measurement of X^*, as follows[1]:

$$f_X(x) = \int f_{X|X^*}(x|x^*)f_{X^*}(x^*)dx^*, \tag{1}$$

where both X^* and X have continuous support.

The conditional density $f_{X|X^*}$ describes the behavior of the measurement errors defined as $X - X^*$. We focus on the estimation of the true model f_{X^*} given the measurement error structure $f_{X|X^*}$ and a sample of X. A straightforward estimator is to solve for f_{X^*} from Eq. (1) with f_X replaced by its sample counterpart. In fact, Eq. (1) is a Fredholm integral equation of the first kind, which is notoriously ill-posed.[2]

The ill-posed inverse problems have been widely studied in statistics literature, and the main efforts in solving the problems were put into various regularization methods pioneered by Tikhonov (1963). In econometrics literature, economists also focus on constructing estimators and deriving optimal convergence rates of the estimators based on various regularization methods in a general setting, such as Eq. (1). (e.g., see Blundell et al. (2007), Chen and Reiss (2011), and Hall and Horowitz (2005))

In this paper, however, we show that the widely admitted ill-posed problem above is actually well-posed for self-reporting data, under the condition that interviewees report truthfully with

☆ We are grateful to an associate editor, two anonymous referees, Chris Bollinger, Arthur Lewbel, Tong Li, Susanne Schennach, Stephen Shore, Richard Spady, Tiemen Woutersen, and seminar participants at the ESWC 2010 for helpful comments or discussions. We also thank Han Hong for sharing the dataset and Wendy Chi for proofreading the draft. All errors remain our own.

* Corresponding author.
E-mail addresses: yonghong.an@uconn.edu (Y. An), yhu@jhu.edu (Y. Hu).

[1] The measurement error problem may also involve in estimating interested parameters that appear in an equation with X^*, and the estimation may (e.g., Li (2002)) or may not involve estimating f_{X^*}. In this paper, we focus on the nonparametric estimation of f_{X^*}. As we argue in the paper, estimating f_{X^*} generalizes many other interesting problems in economic applications.

[2] According to Hadamard (1923), a well-posed problem has the following three properties. (1) A solution exists. (2) The solution is unique. (3) The solution depends continuously on the data. If any of the three conditions is violated, then the problem is ill-posed.

0304-4076/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2012.01.036

a nonzero probability. The property of truthful-reporting can be observed from validation studies by Bollinger (1998) and Chen et al. (2008). Based on this property, we prove that the inverse problem Eq. (1) is in fact a *Fredholm integral equation of the second kind*, which is generally well-posed. We further employ the existing results in the literature to show that comparing with the case of ill-posedness, well-posedness can generally be translated into faster rates of convergence for the estimators of $f_{X^*}(\cdot)$. Hence the property of positive truth-reporting probability may help us gain great advantage in estimating the unknown distribution $f_{X^*}(\cdot)$. Therefore, we advocate that it is best for economists to exploit the property of self-reporting data while solving the inverse problems in measurement error models with a generally ill-posed setup, such as Eq. (1).

To further implore the implications of our results on well-posedness, we analyze the well-known classical measurement error case, where the error structure $f_{X|X^*}(x|x^*)$ is reduced to $f_\epsilon(x - x^*)$. In this case, estimating the unknown density $f_{X^*}(\cdot)$ is a deconvolution problem. We provide sufficient conditions under which a general deconvolution problem is well-posed, and we also present the convergence rate of the deconvolution estimator $\widehat{f}_{X^*}(\cdot)$. In general, this rate is faster than the existing ones in the literature (e.g., see Fan (1991)).

This paper points out that if self-reported errors satisfy that there is a nonzero probability of being zero, then the inverse problems in measurement error models are well-posed. In both general and classical measurement error cases, we show that for well-posed inverse problems, the achievable rates of convergence for estimating f_{X^*} may be much faster than that available in the literature. These results imply that the estimation of the latent model f_{X^*} from the observed sample of X may not be as technically challenging as previously thought. In this sense, our findings in this paper are important in economic applications. The importance of our findings is also due to the fact that the theoretical framework Eq. (1) generalizes many other interesting problems in economics. For instance, estimating the nonparametric structural function from an instrumental variable model in Newey and Powell (2003) is equivalent to estimating f_{X^*}. The estimation of consumption based asset pricing Euler equations in Lewbel and Linton (2010) can also be described in the same framework as ours.[3]

We organize the rest of the paper as follows. In Section 2, we present a general setup of the inverse problem in measurement error models. In Section 3, we show the well-posedness of measurement error models for self-reporting data, and discuss the rates of convergence for $\widehat{f}_{X^*}$ when the problem is well-posed. In Section 4, we analyze the well-posedness in the case of classical measurement errors and present the convergence rate for the deconvolution estimator. In Section 5, we provide Monte Carlo evidence on the improvement that the property can make in estimating f_{X^*}. In Section 6, we present an empirical illustration, using the data-set that matches self-reported earning from the CPS to employer-reported social security earnings (SSR) from 1978. Section 7 concludes. Proofs are in the Appendix.

2. A general setup

We are interested in the nonparametric estimation of the distribution of a latent variable X^*, $f_{X^*}(\cdot)$, given the known measurement error structure $f_{X|X^*}$ and a sample of X. The random sample $\{X_i\}_{i=1,\ldots,n}$ contains the contaminated measurements of the true values X_i^* in each observation i. The estimation of $f_{X^*}(\cdot)$ is based on solving Eq. (1). Without loss of generality, we assume that the supports of X and X^* are the real line $\mathbb{R}$ and the inverse problem is defined on the L^p $(1 \le p \le \infty)$ space over the real line, i.e., $L^p(\mathbb{R})$, with $f_X, f_{X^*} \in L^p$ unless we specify the space otherwise.

For simplicity, we alternatively express the inverse problem as an operator equation:

$$f_X = L_{X|X^*} f_{X^*}, \tag{2}$$

where the operator $L_{X|X^*} : L^p(\mathbb{R}) \to L^p(\mathbb{R})$ is defined as

$$(L_{X|X^*}h)(x) = \int f_{X|X^*}(x|x^*)h(x^*)dx^*, \quad \forall h \in L^p(\mathbb{R}).$$

The well-posedness of the inverse problem (2) is then defined as follows.

Definition 1 (*Carrasco et al., 2007, p. 5670*). The equation $L_{X|X^*}f_{X^*} = f_X$ $(f_{X^*}, f_X \in L^p)$ is well-posed if $L_{X|X^*}$ is bijective and the inverse operator $L^{-1}_{X|X^*} : L^p(\mathbb{R}) \to L^p(\mathbb{R})$ is continuous. Otherwise, the equation is ill-posed.

In this paper, we intend to focus on the estimation, instead of identification, of the latent model $f_{X^*}(\cdot)$. Hence we make the following assumption.

Condition 1. *The density of measurement error $f_{X|X^*}$ is known and the operator $L_{X|X^*} : L^p(\mathbb{R}) \to L^p(\mathbb{R})$ is injective.*[4]

This assumption guarantees that the left inverse of $L_{X|X^*}$ exists and f_{X^*} is uniquely identified from Eq. (2).[5] Therefore, we can identify and estimate f_{X^*} as follows:

$$f_{X^*} = L^{-1}_{X|X^*} f_X.$$

As in many empirical applications, however, we only observe a random sample of X instead of the density f_X itself. We have to replace f_X by its estimator based on the random sample $\{X_i\}$. Let $\widehat{f}_X$ denote an estimator of f_X, then the latent model f_{X^*} can be estimated as

$$\begin{aligned}\widehat{f}_{X^*} &= L^{-1}_{X|X^*}\widehat{f}_X \\ &= f_{X^*} + L^{-1}_{X|X^*}(\widehat{f}_X - f_X).\end{aligned}$$

Since the injectivity of $L_{X|X^*}$ is assumed above, we still need its surjectivity and the continuity of $L^{-1}_{X|X^*}$ to assure the well-posedness of the problem (2).

In economic applications, the main concern for well-posedness of this inverse problem is the continuous dependence of the estimator $\widehat{f}_{X^*}$ on the data of X, i.e., whether the bias in $\widehat{f}_{X^*}$, $L^{-1}_{X|X^*}(\widehat{f}_X - f_X)$, is dependent on the estimation error in $\widehat{f}_X$ continuously. Notice that whether the problem is well-posed or not is completely determined by the operator $L_{X|X^*}$: if the inverse $L^{-1}_{X|X^*}$ is not continuous, then the problem becomes ill-posed and a small estimation error in $\widehat{f}_X$ might cause a huge bias in $\widehat{f}_{X^*}$. Actually, when the problem is ill-posed on the space L^p, it may still be well-posed on some subsets of L^p if some prior information of f_{X^*} is available. In this case, the problem is *conditionally* well-posed.[6] It is beyond the scope of this paper to discuss conditional well-posedness, and we only focus on ill/well-posedness of the problem (2).

[3] Also see Carrasco et al. (2007) for more interesting examples.

[4] We assume that $f_{X|X^*}$ is known here, and more properties of $f_{X|X^*}$ will be specified when they are needed.

[5] Given an operator $F : \Upsilon \to \Psi$, if there exists an operator $G : \Psi \to \Upsilon$ such that GF is the identity operator I on Υ, then G is said to be a left inverse of F. G exists if and only if F is injective. See Naylor and Sell (2000), pp. 32–33 for details.

[6] A rigorous definition of conditionally well-posed is as follows (Petrov and Sizikov, 2005, p. 157): An operator equation $L_{X|X^*}f_{X^*} = f_X$ with $f_{X^*}, f_X \in L^p(\mathbb{R})$ is conditionally well-posed if

(i) It is known *a priori* that a solution of the problem above exists and belongs to a specific set $\Upsilon \subset L^p(\mathbb{R})$.
(ii) The operator $L_{X|X^*}$ is a one-to-one mapping of Υ onto $L_{X|X^*}\Upsilon \equiv \Psi$.
(iii) The operator $L^{-1}_{X|X^*}$ is continuous on $\Psi \subset L^p(\mathbb{R})$.

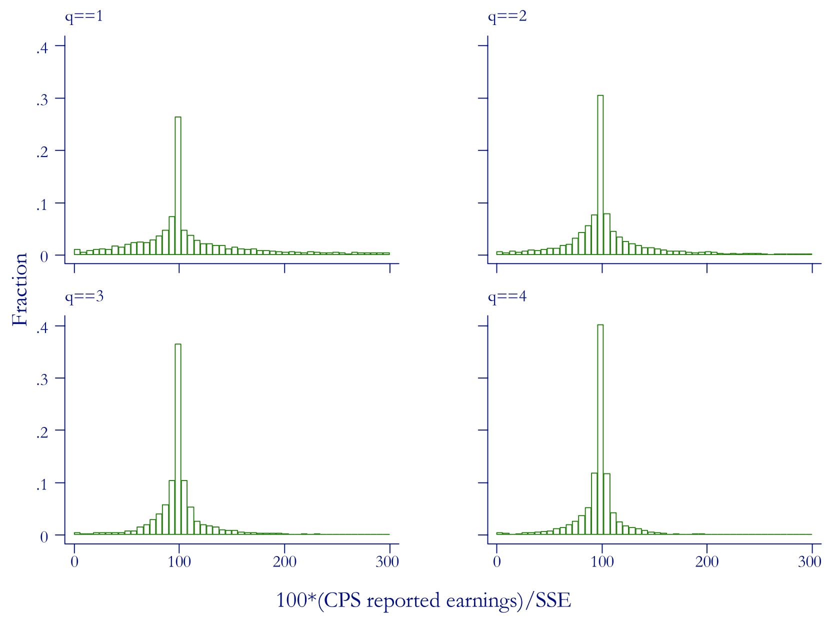

Fig. 1. Histograms of measurement error in earnings, by quartile of true (Social Security) earnings. The figure was excerpted from Chen et al. (2008), p. 50. The link of the paper is: http://cowles.econ.yale.edu/P/cd/d16a/d1644.pdf.

3. Measurement error models for self-reporting data

In this section, we show the well-posedness of measurement error models for self-reporting data and discuss the convergence rate for the nonparametric estimator of the latent distribution, $\widehat{f_{X^*}}$. We first present a property observed in validation studies that individuals report the true values with a nonzero probability. As a consequence, the problem (2) becomes a Fredholm equation of the second kind and is well-posed. Next, we discuss the rates of convergence for $\widehat{f_{X^*}}$ in both well-posed and ill-posed problems. By comparing the rates in two scenarios, we try to emphasize the importance of well-posedness in economic applications.

3.1. A property of self-reporting errors

This subsection discusses the properties of the operator $L_{X|X^*}$ in measurement error models for self-reporting data. We show why and how self-reporting errors are essentially distinct from the traditional measurement errors.

The traditional measurement error models describe the errors generated from measuring a true value, such as, height or temperature, using certain measurement equipment, e.g., a ruler or a thermometer. Such errors are generally assumed to be independent of the true values, which makes perfect sense because the errors are mainly caused by the equipment or measuring methods. However, most measurement errors in economic variables are not caused by measurement but by misreporting. This is due to the fact that most of economic studies are based on self-reported survey data, such as Current Population Survey (CPS) and Panel Study of Income Dynamics (PSID). Therefore, it is essential for economists to take into account the properties of the self-reporting errors before using the traditional measurement error models.

Self-reporting errors have been studied thoroughly in the literature. In a validation study by Chen et al. (2008), they provide an important empirical evidence on the exact distribution of self-reporting errors for earnings. The authors use the data set that matches self-reported earning from the CPS to employer-reported social security earnings (SSR) from 1978 (the CPS/SSR Exact Match File). By quartile of Social Security Earnings, the four sub-figures in Fig. 1 show histograms of percentage of the ratio between self-reported and social security earnings. An observation from the figure is that there are mass points where self-reported earnings equal social security earnings, i.e., the probability of reporting truthfully is strictly positive.[7]

In fact, Bollinger (1998) provides estimates of the probability of reporting truthfully in CPS. The paper utilizes the same CPS/SSR exact match file above to show that 11.7% of the men and 12.7% of the women report their earnings correctly. In addition, he finds that the probability of reporting truthfully does not vary much with the true income.

Similar observations also apply to the discrete variables. Bound et al. (2001) provides the discrete version of $f_{X|X^*}$ in different economic data, where the misclassification probability matrices corresponding to $f_{X|X^*}$ are all strictly diagonally dominant, i.e., the probability of telling the truth is much larger than that of reporting any other values. For instance, when employees are asked to report their occupation classification, self-reported data are agree with company reported ones for 70% of the current occupations, and for 60% of the occupations more than four years ago.

Employing the CPS/SSR data set we discussed above, we plot histogram of social security earnings X^* for those X^* that are exactly equal to X, the self-reported earnings in Fig. 2. The histogram shows that people report truthfully almost at every earning level, which implies that they report truthfully not just because their earning levels are easy to remember.

These validation studies suggest that there is a nonzero probability that people report the truth even for a continuous variable, i.e., the distribution of self-reporting errors has a mass point at zero. This observation may be explained by the following reporting process shown in Fig. 3: if an interviewee remembers the true value, she first decides whether to intentionally misreport the truth or not. Empirical evidences above suggest that she would report the truth with a nonzero probability; if she does not remember the true value, she provides an estimate of the true

[7] We would like to emphasize that by truthful reporting, we mean X^* is equal to X exactly. Due to the discrete nature of the histograms, this point may not be illustrated clearly by the figure. However, this property can be directly observed from the CPS/SSR data.

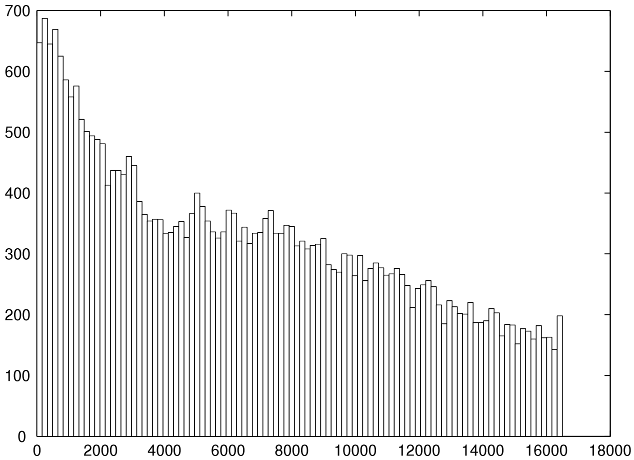

Fig. 2. Histogram of X^* given $X^* = X$.

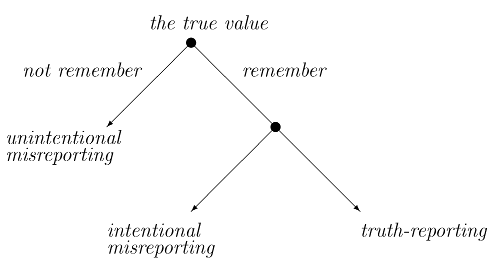

Fig. 3. Illustration of self-reporting.

value, which may be considered as unintentional misreporting.[8] Admittedly, we cannot distinguish intentional misreporting from unintentional misreporting without further information. The example on self-reported occupations we discussed above can be rationalized by our conjecture: as time goes, the probability of remembering the occupations decreases, which leads to the decreasing agreement rate between self-reported occupations and company reported ones.[9]

Based on these observations from the validation studies, it is natural to make the following assumption in measurement error models for self-reporting data.

Condition 2. *The probability of telling the truth conditional on the true values is bounded away from zero, i.e.*

$$\lambda(x^*) \equiv \Pr(X = X^* | X^* = x^*) \geq c > 0 \quad \text{for any } x^*.$$

Consequently, the self-reporting error distribution may be written as:

$$f_{X|X^*}(x|x^*) = \lambda(x^*) \times \delta(x - x^*) + [1 - \lambda(x^*)] \times g(x|x^*), \tag{3}$$

where $\delta(\cdot)$ is a Dirac delta function and $g(x|x^*)$ is the conditional density corresponding to misreporting errors.[10]

3.2. Well-posedness with self-reporting errors

Given the property of the self-reporting error in economic data, the corresponding models of measurement error in Eq. (3) becomes

$$\begin{aligned} f_X(x) &= \int f_{X|X^*}(x|x^*) f_{X^*}(x^*) dx^* \\ &= \lambda(x) f_{X^*}(x) + \int g(x|x^*)[1 - \lambda(x^*)] f_{X^*}(x^*) dx^*, \end{aligned}$$

which is a Fredholm equation of the second kind. We may also describe it as an operator equation,

$$\begin{aligned} f_X &= L_{X|X^*} f_{X^*} \\ &= [D_\lambda + L_g(I - D_\lambda)] f_{X^*}, \end{aligned} \tag{4}$$

where I is an identity operator defined on L^p, $D_\lambda : L^p(\mathbb{R}) \to L^p(\mathbb{R})$ is the multiplication operator defined as

$$(D_\lambda h)(z) = \lambda(z) h(z), \quad 0 < \lambda(z) \leq 1, \tag{5}$$

and the operator $L_g : L^p(\mathbb{R}) \to L^p(\mathbb{R})$ is defined as

$$(L_g h)(x) = \int g(x|x^*) h(x^*) dx^*. \tag{6}$$

Since $\lambda(z) \geq c > 0$, the operator Eq. (4) can be written as

$$D_\lambda^{-1} f_X = [I + D_\lambda^{-1} L_g(I - D_\lambda)] f_{X^*}, \tag{7}$$

where the only unknown is still f_{X^*}. Moreover, Eq. (7) belongs to Fredholm equations of the second kind. Since it is known that Fredholm equations of the second kind are well-posed under certain conditions, our goal here is to apply the existing results to show the well-posedness of problem (2) under Condition 2. For this purpose, we need to assume the compactness of the operator L_g.

Condition 3. *Operator $L_g : L^p(\mathbb{R}) \to L^p(\mathbb{R})$ defined in Eq. (6) is compact.*

The compactness of operator L_g may correspond to different properties of the density $g(\cdot|\cdot)$ in L^p space for different p. For example, in L^2 space, an integral operator is a Hilbert–Schmidt operator and consequently is compact if the kernel of the operator is square integrable (see e.g. Pedersen (1999), pp. 92–94.).[11] Therefore if we assume $\|g(\cdot|\cdot)\|_2 < \infty$, then the operator L_g is compact, i.e., in $L^2(\mathbb{R})$ space a sufficient condition for compactness of the operator L_g is that the error density $g(\cdot|\cdot)$ is square integrable.

Now we are ready to present the main result on the well-posedness of problem (2).

Theorem 1. *Under Conditions 1–3 the problem (2) is well-posed.*

Proof. See Appendix. □

This theorem suggests that the observed property of misreporting errors has a strong implication for modeling measurement error problems with survey data. Without Condition 2, the problem (2) is ill-posed, which implies that the estimation of the latent model f_{X^*} is quite technically challenging. However, Condition 2, which is directly supported by empirical evidences, dramatically reverses the pessimistic perspective on this inverse problem. Theorem 1implies that the estimator of f_{X^*} based on Eq. (2) with self-

---

8 This conjecture may help us justify whether the self-reported data on a variable has the property. For instance, validation data are also available in biostatistics according to Carroll et al. (2006). However, the existence of validation data in biostatistics generally does not imply that the error distribution has a point mass. This is because the variables biostatistician mainly focus on are different from wage we analyzed in the sense that in biostatistics the interviewees hardly know the exact value they need to report. For example, a person hardly knows his weight as accurate as measured by weight scales (true values). Consequently, there is no way for interviewees report the true value with positive probability.

9 However, we do have the risk that the nonzero point mass is an untestable assumption in many existing surveys for which validation samples are not available, especially for a continuous X^*. In fact, validation samples are rare in the literature, for example, the 1978 SSR validation sample we discussed in the paper has been used for about thirty years (recently the data set was used by Chen et al. (2005)).

10 The misreporting error density $g(x|x^*)$ is corresponding to both unintentional and intentional misreporting in Fig. 3, and the two sources are indistinguishable without further information.

11 Let k be a function of two variables $(s, t) \in I \times I = I^2$, where I is a finite or infinite real interval. Then a linear integral operator K on $L^2(I)$ is called a Hilbert–Schmidt operator if the kernel k is in $L^2(I \times I)$, i.e., $\|k\|_2 = \int_I \int_I |k(s,t)|^2 ds dt < \infty$.

reported data should perform well in general if the misreporting errors have a nonzero probability of being equal to zero. The virtue of honesty literally makes the inverse problem (2) well-posed.

Furthermore, the optimistic result in Theorem 1 may also have implications on certain instrumental variable models (e.g., see Newey and Powell (2003)). We may consider the latent variable X^* as the endogenous variable and X as its exogenous instruments. Then the results in Theorem 1 imply that an instrumental variable model may also be well-posed when $\Pr(X^* = X|X^*) > 0$, i.e. the variable X^* is exogenous with a nonzero probability.[12]

3.3. Rates of convergence

In economic applications, the main difficulty caused by ill-posedness is the slow rate of convergence for the nonparametric estimator $\widehat{f}_{X^*}$. Hence, we only focus on the connection between ill/well-posedness and the rates of convergence.[13] Specifically, we discuss the rates of convergence for $\widehat{f}_{X^*}$ in both well-posed and ill-posed problems. By comparing the rates in two scenarios, we try to emphasize the importance of well-posedness in economic applications whenever we need to nonparametrically estimate the unknown density f_{X^*}.

We first analyze the rate of convergence for a well-posed problem as in Eq. (7). For the convenience of our analysis, we rewrite the problem as

$$(I - K)f_{X^*} = \omega, \tag{8}$$

where $K \equiv D_\lambda^{-1} L_g (D_\lambda - I)$, $\omega \equiv D_\lambda^{-1} f_X$. Let $\widehat{K}_n$ and $\widehat{\omega}_n$ respectively denote the estimates of K and ω, where n is the sample size. For such a problem, Carrasco et al. (2007) give the rate of convergence in estimating f_{X^*}, we restate the result in the following proposition.

Proposition 1 (*Carrasco et al., 2007, p. 5729*). *For a well-posed problem Eq. (8), if we have $\|\widehat{K}_n - K\| = o(1)$ and $\|(\widehat{\omega}_n + \widehat{K}_n f_{X^*}) - (\omega + K f_{X^*})\| = O(\frac{1}{a_n})$, then $\|\widehat{f}_{X^*} - f_{X^*}\| = O(\frac{1}{a_n})$, where $a_n \to \infty$ as $n \to \infty$.*

The results above are under L^2-norm. For a linear operator K, the norm $\|K\|$ is defined as $\sup_{\|\phi\|=1} \|K\phi\|$, where ϕ is any vector in a normed vector space.

Thus far, we assumed that both the probability $\lambda(x^*)$, and the error density $f_{X|X^*}$ are known. If this is the case, the proposition above shows that the convergence rate for the estimator $\widehat{f}_{X^*}$ is the same as that for the estimator $\widehat{f}_X(x)$, which has a rate of kernel density estimation with uncontaminated observations. More generally, if we estimate linear functionals of f_{X^*}, e.g., moments of X^*, a parametric rate may be obtained under suitable regularity conditions, as in Shen (1997).

In many economic applications, the error density $f_{X|X^*}$ and the probability $\lambda(x^*)$ may be unknown and need to be estimated. The impacts of estimating $f_{X|X^*}$ and $\lambda(x^*)$ on the statistical properties of $\widehat{f}_{X^*}$ can be analyzed using Proposition 1: when the rates of $\widehat{f}_{X|X^*}$ and $\widehat{\lambda}(x^*)$ are not slower than that of $\widehat{f}_X$, then estimating $f_{X|X^*}$ and $\lambda(x^*)$ has no impact on the rate of $\widehat{f}_{X^*}$; otherwise, the rate of $\widehat{f}_{X^*}$ is determined by the slower rate of $\widehat{f}_{X|X^*}$ and $\widehat{\lambda}(x^*)$. The impacts of estimating error density $f_{X|X^*}$ on the rate of $\widehat{f}_{X^*}$ are also addressed by a few existing papers in different settings of the inverse problem. For classical measurement error, Li and Vuong (1998) consider how estimating unknown error density $f_{X|X^*}$ affects the nonparametric estimation of $\widehat{f}_{X^*}$ in the case where repeated measurements of X are observed. Allowing arbitrary correlation between the measurement error and the true data, Chen et al. (2005) analyze parametric estimation of $\widehat{f}_{X^*}$ using auxiliary data of X^* and X when the error density is unknown. However, it is beyond the scope of this paper to analyze how estimation of the error density $f_{X|X^*}$ affects asymptotic properties of the nonparametric estimator $\widehat{f}_{X^*}$ in general when measurement error has a mass point at zero.

Without Condition 2, the ill-posedness of the problem Eq. (2) leads to a notoriously slow rate of convergence for the estimator $\widehat{f}_{X^*}$.[14] However, there does not exist a general rate of $\widehat{f}_{X^*}$ for an ill-posed problem since ill-posed problems can be further classified as mildly ill-posed and severely ill-posed ones according to the properties of the operator $L_{X|X^*}$, and the rates in two cases can be very different (e.g., see Chen and Reiss (2011) for details). Nevertheless, the slow rate for ill-posed problems can be well illustrated by the case of deconvolution: for example, if the error distribution $f_{X|X^*}$ is such that its Fourier transform decays exponentially, then the rate for the estimator $\widehat{f}_{X^*}$ is of logarithmic order.[15]

4. A further discussion on the classical error case

In this section, we further explore the implications of Theorem 1 in a special case: the measurement error is classical, i.e., the measurement error ϵ is independent of the true value X^*.

For classical measurement errors, the error density $f_{X|X^*}(x|x^*)$ is reduced to $f_\epsilon(x - x^*)$. Furthermore, it is known that the independence of X^* and ϵ implies that the characteristic functions of f_X, f_{X^*}, and f_ϵ (denoted by $\phi_X(\cdot)$, $\phi_{X^*}(\cdot)$, and $\phi_\epsilon(\cdot)$, respectively) have the following relationship:

$$\phi_X(t) = \phi_{X^*}(t)\phi_\epsilon(t).$$

Condition 1 guarantees that $\phi_\epsilon(t) \neq 0$ for any real t. Therefore, the density f_{X^*} can be recovered from deconvolution, i.e.,

$$f_{X^*} = \frac{1}{2\pi} \int e^{-itx} \phi_{X^*}(t) dt = \frac{1}{2\pi} \int e^{-itx} \frac{\phi_X(t)}{\phi_\epsilon(t)} dt.$$

In empirical applications, the density $f_X(x)$ needs to be estimated from the observed data $\{X_i\}_{i=1,\dots,n}$. A popular estimator for f_X is as follows:

$$\hat{f}_X = \frac{1}{2\pi} \int e^{-itx} \hat{\phi}_X(t) dt \tag{9}$$

$$\hat{\phi}_X(t) = \hat{\phi}_n(t)\phi_K\left(\frac{t}{T_n}\right),$$

where $\hat{\phi}_n(t)$ is an empirical characteristic function defined by $\hat{\phi}_n(t) = \sum_{i=1}^n e^{itX_i}/n$, and $\phi_K(\frac{t}{T_n})$ is the Fourier transform of the kernel function $K(\cdot)$ with bandwidth $\frac{1}{T_n}$. The smoothing parameter T_n depends on the sample size n. In other words, a different T_n implies a different estimator $\widehat{f}_X$ for f_X. We may pick a kernel $K(\cdot)$ such that $\phi_K(t) = 0$ for $|t| > 1$. To assure $\hat{\phi}_X(t)$ uniformly converge to $\phi_X(t)$ over $[-T_n, T_n]$ at a geometric rate with respect to the sample size n, Hu and Ridder (2010) suggest that we need

$$T_n = O\left(\frac{n}{\log n}\right)^\gamma \quad \text{for } \gamma \in \left(0, \frac{1}{2}\right). \tag{10}$$

[12] We thank Richard Spady for pointing this out.

[13] Given that Condition 2 is satisfied, the asymptotic properties of $\widehat{f}_{X^*}$ does not change with the probability $\lambda(x^*)$. For finite samples, as we show in Monte Carlo evidence, a larger or smaller probability $\lambda(x^*)$ does affect the property of the estimator $\widehat{f}_{X^*}$.

[14] If suitable regularization schemes are employed to approximate the solution of an ill-posed problem, the rate of convergence may be fast under some additional assumptions, please see Carrasco et al. (2007) for detailed discussions. We only focus on the comparison of ill-posed problems and well-posed ones, hence we do not discuss the regularization schemes and the assumptions that lead to fast rates in this paper.

[15] We will further discuss the rates of convergence in both ill-posed and well-posed problems for deconvolution estimators in the next section.

Consequently the deconvolution estimator of f_{X^*}, $\hat{f}_{X^*}(x^*)$ is

$$\widehat{f}_{X^*}(x^*) = \frac{1}{2\pi}\int e^{-itx^*}\frac{\hat{\phi}_X(t)}{\phi_\epsilon(t)}dt = \frac{1}{2\pi}\int_{-T_n}^{T_n} e^{-itx^*}\frac{\widehat{\phi}_n(t)\phi_K(t/T_n)}{\phi_\epsilon(t)}dt.$$

It is known in the literature that the general deconvolution problem described above is ill-posed, and the rate of convergence for the estimator $\widehat{f}_{X^*}$ is very slow (e.g., see Fan (1991) for a general analysis of the deconvolution problem). In the remaining parts of this section, we first show that the general deconvolution problem above is well-posed under some mild conditions corresponding to Condition 2. Next, we present the rate of convergence for the deconvolution estimator $\widehat{f}_{X^*}$ when the problem is well-posed. Our analysis in this section will be conducted on both L^2 and L^∞ spaces.

4.1. Well-posedness for classical measurement errors

In this subsection, we assume that $\lambda(x^*) = \lambda$ is a constant for simplicity.[16] Our discussion can be extended to the general case straightforwardly. Accordingly, the error distribution is

$$f_{X|X^*}(x|x^*) = f_\epsilon(x - x^*) = \lambda \times \delta(x - x^*) + (1-\lambda) \times g_{\tilde{\epsilon}}(x - x^*).$$

Let $\phi_\epsilon(t)$ and $\phi_{\tilde{\epsilon}}(t)$ denote the characteristic functions of f_ϵ and $g_{\tilde{\epsilon}}$, respectively. The equation above then implies that

$$\phi_\epsilon(t) = \lambda + (1-\lambda)\phi_{\tilde{\epsilon}}(t).$$

The ch.f. $\phi_\epsilon(t)$ is in fact bounded away from zero by a constant. Define the space of all the bounded functions with a bounded Fourier transform as

$$L^\infty_{bc} = \left\{ f \in L^\infty(\mathbb{R}) : \sup_{t\in\mathbb{R}} |\phi_f(t)| < \infty \right\}.$$

We have the following results.

Proposition 2. (i) *Suppose Conditions 1 and 2 hold and the error distribution $g_{\tilde{\epsilon}}$ satisfies*

$$\int |\phi_{\tilde{\epsilon}}(t)|dt < \infty.$$

Then problem (2) is well-posed with $L_{X|X^*} : L^\infty_{bc} \to L^\infty_{bc}$.

(ii) *Suppose Conditions 1 and 2 hold and the error distribution $g_{\tilde{\epsilon}}$ satisfies*

$$\int_{\mathbb{R}}\int_{\mathbb{R}} |g_{\tilde{\epsilon}}(x - x^*)|^2 dx dx^* < \infty. \tag{11}$$

Then problem (2) is well-posed with $L_{X|X^*} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$.

Proof. See Appendix. □

This proposition is just a specific case of Theorem 1. Instead of imposing the compactness of the operator $L_{X|X^*}$ as in Theorem 1, here we just make some less abstract assumptions on the error distribution $g_{\tilde{\epsilon}}$ (or its Fourier transform) to assure the well-posedness. Even though the results in this proposition are not as general as Theorem 1, they might be very useful in applications since we will show that the results permit us to obtain a consistent estimator of f_{X^*} with a desirable convergence rate from the sample $\{X_i\}$ for a very general error distribution.

4.2. Rates of convergence for deconvolution estimators

The deconvolution estimator has been studied thoroughly in the literature and the convergence rates of $\widehat{f}_{X^*}$ are established under various circumstances. In this section, we try to associate the existing results with the deconvolution problem. We illustrate how ill-posedness and well-posedness can be translated to different rates of convergence for $\widehat{f}_{X^*}$ when the measurement error is classical.

Hesse (1995) demonstrates that the best rates of convergence for $\widehat{f}_{X^*}$ are $(\log n/n)^{2/5}$ and $n^{-4/5}$ under L^∞-norm and L^2-norm, respectively, if the observations are "partially contaminated".[17] For self-reported data, the existence of partially contaminated observations is equivalent to the fact that the truth-telling probability $\lambda(x^*) > 0$. Hence the rates for a well-posed problem with classical measurement errors are the same as that in Hesse (1995). To restate the result, we first specify the conditions.[18]

A1. f_{X^*}, f'_{X^*}, and f''_{X^*} are uniformly absolutely bounded.
A2. For any $x > 0$, there exists some $\rho > 0$ such that $P(|X_i| > x) \le x^{-\rho}$.
A3. The kernel function $K(\cdot)$ satisfies $\int K(u)du = 1$, $\int uK(u)du = 0$, and $\int u^2K(u)du < \infty$.
A4. The characteristic functions $\phi_K(t)$ and $\phi_{X^*}(t)$ are twice continuously differentiable; $\phi_K(t) = 0$ for $|t| > 1$, and $\inf_t |\phi_\epsilon(t)| \ge \lambda$.
A5. f_{X^*} has square integrable continuous second derivative.
A6. $|\phi_{\tilde{\epsilon}}(t)| \to 0$ as $|t| \to +\infty$.

It is easy to check that the assumptions above all hold for our setting of the inverse problem. Specifically, the relationship $\phi_\epsilon(t) = \lambda + (1-\lambda)\phi_{\tilde{\epsilon}}(t)$ guarantees that $\inf_t |\phi_\epsilon(t)| \ge \lambda$ holds in A4. With these assumptions, we state the convergence rate of $\widehat{f}_{X^*}$ in the following proposition.

Proposition 3 (*Hesse 1995*).

(i) *Under conditions* A1–A4 *and the choice of optimal bandwidth* $\frac{1}{T_n} = c(\frac{\log n}{n})^{1/5}$, *we have*

$$\lim_{n\to\infty} \sup \left(\frac{n}{\log n}\right)^{2/5} \sup_{x^*\in\mathbb{R}} |\widehat{f}_{X^*} - f_{X^*}| < \infty, \quad a.s.$$

where c is a positive constant, and $\widehat{f}_{X^*}$ is the deconvolution estimator of f_{X^*} with sample size n.

(ii) *Let conditions* A1, A3, A4, A5, *and* A6 *hold, then we have*

$$MISE(h_{opt}) = O(n^{-4/5}),$$

where

$$MISE(h) \equiv E\left[\int_{\mathbb{R}} (\widehat{f}_{X^*}(u) - f_{X^*}(u))^2 du\right],$$

and h_{opt} is the optimal choice for the bandwidth h,

$$h_{opt} = n^{-1/5}\left(\int_{\mathbb{R}} u^2K(u)du\right)^{-2/5}\left(\int_{\mathbb{R}} (K(u)/\lambda)^2du\right)^{1/5} \times \left(\int_{\mathbb{R}} (f''_{X^*}(u))^2du\right)^{-1/5}.$$

According to Proposition 3, both the rate $(\log n/n)^{2/5}$ and $n^{-4/5}$ are general and achievable for any distribution of measurement error. Hence the result permits us to demonstrate how well-posedness

[16] This assumption can be rationalized by the results in Bollinger (1998) that the probability of reporting truthfully does not vary much with the true income in CPS/SSR data. However, this assumption may not be directly testable in some survey samples where validation samples are not available.

[17] (Hesse, 1995) defines partially contaminated data as $\Pr(X = X^*) = p$, $\Pr(X = X^* + \epsilon) = 1 - p$ and $0 < p < 1$.

[18] Please see Hesse (1995) for explanations of these conditions.

can be translated into faster rates of convergence than that of ill-posed problems in general cases. We take the rate in L^∞-norm as an illustrating example. It is known that the general deconvolution problem is ill-posed when measurement error is super-smooth,[19] and the rates of convergence are of logarithmic order for both pointwise convergence (Carroll and Hall 1988) and in mean square error (Fan 1991). Explicitly, when the distribution of measurement error is standard normal, Carroll and Hall (1988) show that the rate is $(\log n)^{-k/2}$, where the unknown function f_{X^*} has up to k-th bounded derivatives. Apparently, the rate for a well-posed problem is much faster than that for an ill-posed one. When the distribution of measurement error is ordinary smooth, the rate $(\log n/n)^{2/5}$ may be slower than the rate for a classical deconvolution estimator (i.e., $\lambda = 0$). For instance, according to Carroll and Hall (1988), when the distribution of measurement error is gamma with shape parameter α, and k is defined as above, then the fastest achievable rate for a deconvolution estimator is $n^{-k/(2k+2\alpha+1)}$. When $\alpha > (3k-1)/2$, this fastest rate is slower than $(\log n/n)^{2/5}$; when $k > 2$ and $\alpha < (k-2)/4$, the fastest rate is faster than the rate $(\log n/n)^{2/5}$.[20]

The analysis above illustrates that the positive truth-telling probability in self-reported data plays a crucial role in deconvolving a density. The reason is that the positive probability leads to the well-posedness of the deconvolution problem and consequently results in faster rates of convergence for all super-smooth and some ordinary smooth error distributions.

5. Simulation studies: deconvolution with normal error

In this section, we conduct a simulation study to investigate the performance of various deconvolution estimators when the distribution of errors has a mass point at zero.

We consider

$$X = X^* + \epsilon,$$

where X^* is distributed according to a truncated standard normal on the interval $[-1, 1]$. In this study, we estimate the density of X^* from a sample of X, and the known density of errors $f_\epsilon(\cdot)$. Following our discussions in previous sections, the density $f_\epsilon(x - x^*)$ is assumed to be

$$\lambda\delta(x - x^*) + (1 - \lambda)g(x - x^*),$$

where $\lambda \neq 0$, and $g(x - x^*)$ is distributed according to a standard normal. We focus on the deconvolution density estimator

$$\hat{f}_{X^*}(x^*) = \frac{1}{2\pi}\int e^{-itx^*}\frac{\hat{\phi}_X(t)}{\phi_\epsilon(t)}dt,$$

where $\hat{\phi}_X(t) = \hat{\phi}_n(t)\phi_K(\frac{t}{T_n})$ and $\hat{\phi}_n(t) = \frac{1}{n}\sum_{i=1}^n e^{itX_i}$. The kernel K is taken as the normalized sinc function:

$$\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x},$$

and its ch.f. $\phi_K(t)$ is the rectangular function

$$\phi_K(t) = \begin{cases} 0 & \text{if } |t| > \frac{1}{2} \\ \frac{1}{2} & \text{if } |t| = \frac{1}{2} \\ 1 & \text{if } |t| < \frac{1}{2}. \end{cases} \tag{12}$$

We present simulation results for sample size $n = 1000$ in Figs. 4–6 where $T_n = 2.0, T_n = 2.2$ and $T_n = 2.3$, respectively. In each figure, we pick three different values of λ: 2%, 5% and 10%. In all graphs, "estimated density" is the deconvolution estimator $\hat{f}_{X^*}$ given we model the error distribution correctly, while "naïve estimate" is the counterpart given we model the error distribution mistakenly, i.e, $\lambda = 0$. We also include in each plot the 5% and 95% pointwise confidence intervals calculated using bootstrap resampling (200 times) for both "estimated density" and "naïve estimate".

The graphs show that the "estimated density" tracks the true density f_{X^*} much closer than the "naive estimate" does for all the values of λ. We also observe from the graphs that for a fixed T_n, the performance of naïve estimator is getting worse when λ increases, which is natural since the larger λ is, the less accurate of the approximation by $\lambda = 0$ to the true value of λ. For a given λ, the naive estimator is more sensitive to T_n than our consistent estimator because deconvolution with a normal is an ill-posed problem.

6. Empirical illustration

In this section, we illustrate our method empirically by using the data-set we analyzed in Section 3. Besides in Chen et al. (2008) and Bollinger (1998), the data-set has also been used in Bound and Krueger (1991) to study the extent of measurement error in earnings, and in Chen et al. (2005) to study the problem of parameter inference in econometric models when the data are measured with error. A full description of the data-set can be found in Bound and Krueger (1991).

For this data-set, Chen et al. (2008) argued that the error densities are different for different income levels and low income individuals tend to overreport their earnings. In order to reduce bias of estimation, we divide the data into four sub-samples based on SSR: sub-sample 1, 2, 3, 4 contain observations with SSR below the first quartile, between the first and the second quartile, between the second and the third quartile, and above the third quartile, respectively. We also drop those observations with SSR being the topcoded values \$16,500 to reduce bias may caused by the topcoding.[21] Following the literature we introduced above, we assume that the error ϵ, which is defined as $\epsilon = \log X - \log X^*$ is distributed according to the density[22]

$$f_\epsilon(\epsilon) = \lambda\delta(\epsilon) + (1 - \lambda)\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\epsilon-\mu)^2}{2\sigma^2}}. \tag{13}$$

To conduct our analysis, we employ a two-step estimation procedure. First, we estimate parameters λ, μ, and σ for each sub-sample: λ is estimated as the relative frequency of $\epsilon = 0$; while μ and σ are estimated by maximum likelihood estimation with those observation $\epsilon = 0$ dropped from the sample. The estimated results are presented in Table 1.[23]

19 According to Fan (1991), the distribution of the error ϵ is supersmooth of order β if $\phi_\epsilon(t)$ satisfies

$$c_0|t|^{-d}\exp(-|t|^\beta/\rho) \le |\phi_\epsilon(t)| \le c_1|t|^{-d_1}\exp(-|t|^\beta/\rho), \quad \text{as } |t| \to \infty,$$

for some positive constants c_0, c_1, β, ρ and some constants d, d_1. The distributions of normal and Cauchy are examples of this category of distributions. Similarly, the distribution of ϵ is ordinary smooth if $\phi_\epsilon(t)$ satisfies

$$c_0|t|^{-d} \le |\phi_\epsilon(t)| \le c_1|t|^{-d}, \quad \text{as } |t| \to \infty,$$

for some positive constants c_0, c_1, d. The ordinary smooth distributions include gamma, double exponential and symmetric gamma, etc.

20 To derive the condition $\alpha > (3k-1)/2$, we consider that $\log n < \sqrt{n}$ as $n \to \infty$. Hence if $\alpha > (3k-1)/2$, then $(n/\log n)^{2/5} > (\sqrt{n})^{2/5} > n^{k/(2k+2\alpha+1)}$ as $n \to \infty$. Analogously, $(n/\log n) < n$ as $n \to \infty$. Consequently, if $\alpha < (k-2)/4$ and $k > 2$, then $(n/\log n)^{2/5} < n^{2/5} < n^{k/(2k+2\alpha+1)}$ as $n \to \infty$.

21 See Chen et al. (2008) for detailed description of the topcoding.

22 Variable X denotes self-reported earnings, and X^* denotes SSR earnings, which we treat as "true" earnings. We drop those 85 observations with $X = 0$ (3 of them with $X^* = 16{,}500$, too).

23 Standard errors of estimated parameters are computed by bootstrap resampling (200 times).

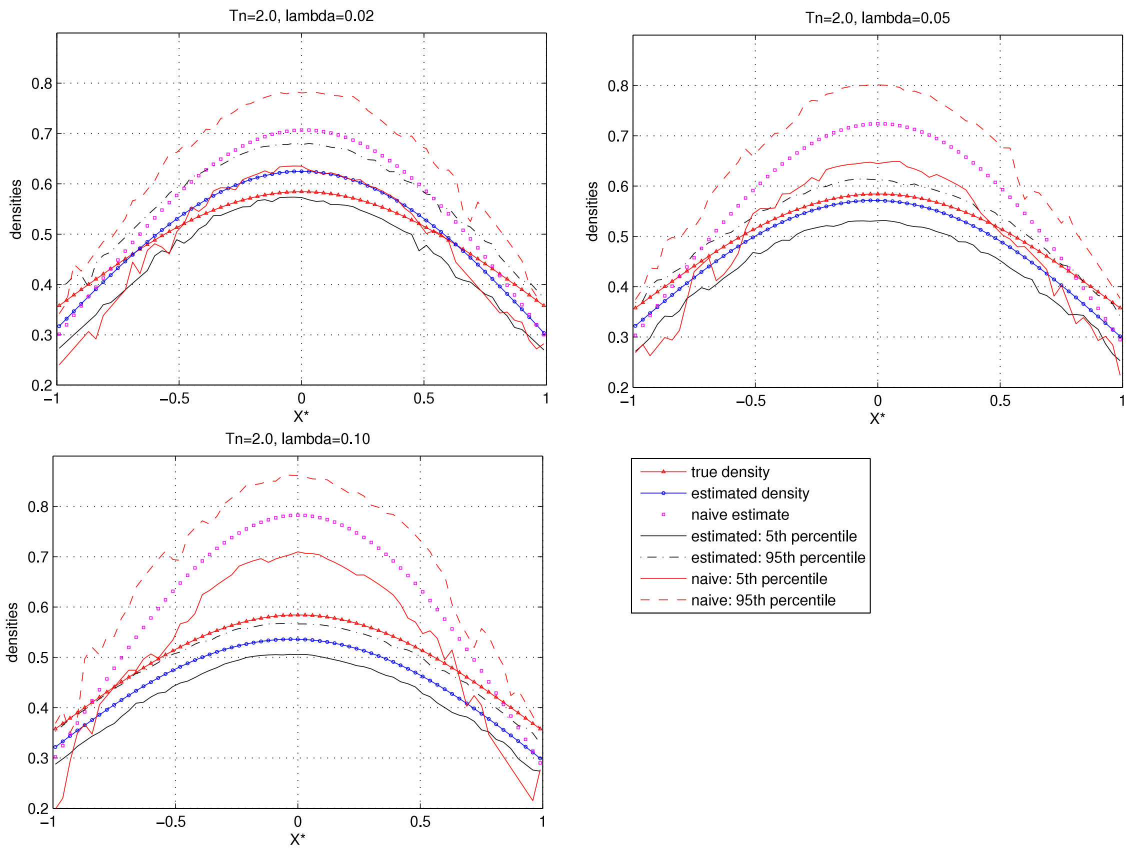

Fig. 4. Simulation results: $T_n = 2.0$.

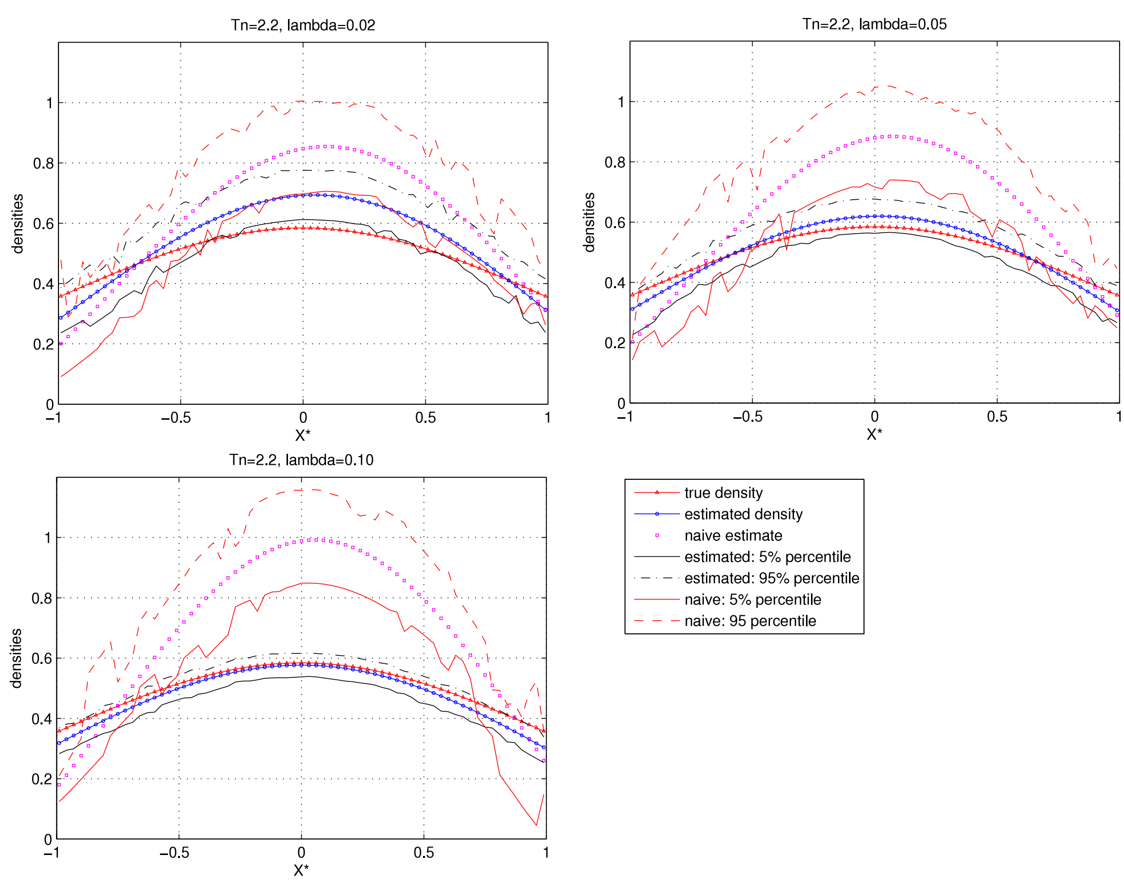

Fig. 5. Simulation results: $T_n = 2.2$.

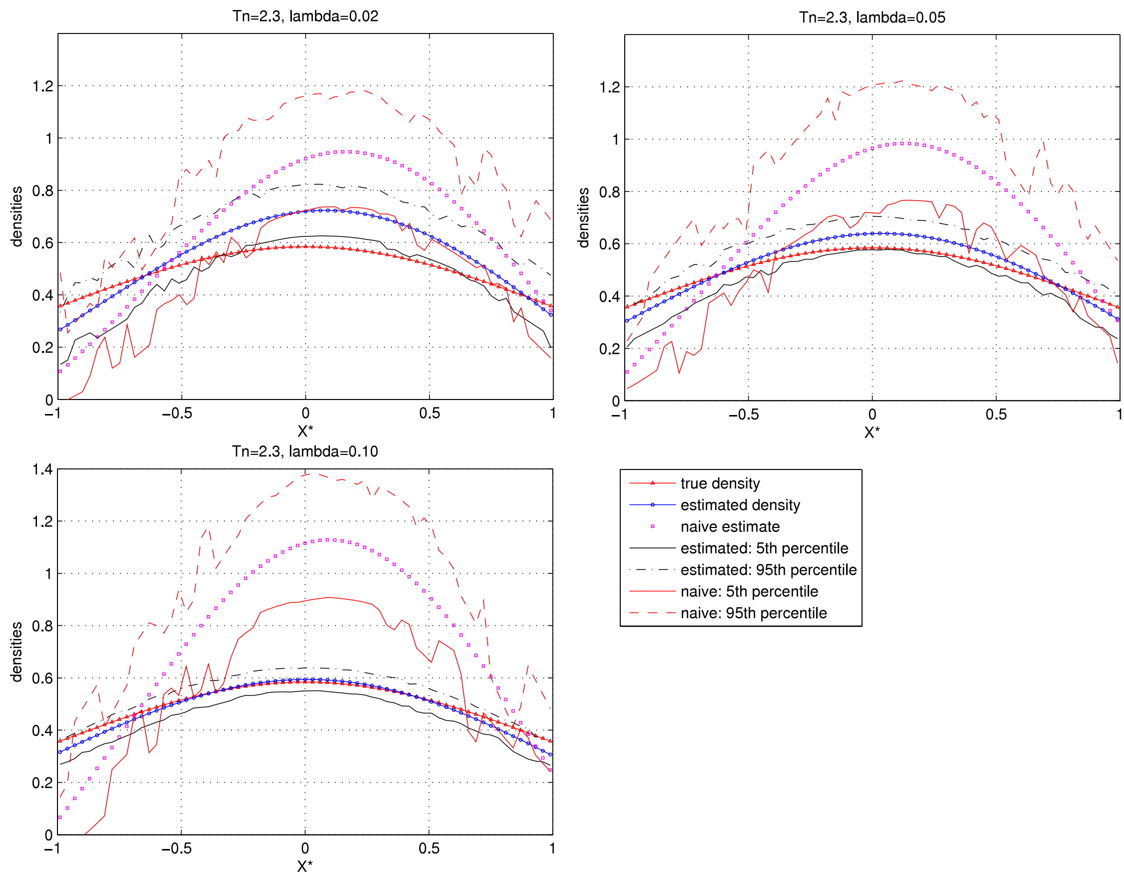

Fig. 6. Simulation results: $Tn = 2.3$.

With the estimated parameters, we employ the method of deconvolution to estimate the density of SSR, f_{X^*} in the second step. Our estimated results are presented in Fig. 7. In each of the four subplots, we present the "true" density of SSR (kernel estimate of the density), "naïve density", the "estimated density", and the 5%–95% pointwise confidence intervals of the last two, where our estimated density uses the estimates of the parameters in the error distribution presented in the third column of Table 1, while the naive density estimator uses the estimates in the fourth column of the table. The kernel function we used in the estimation is the same as the one in the section of simulation. Because of the sample differences, we utilize different T_n for four sub-samples: $T_n = 1.9, 3.4, 5.1$ and 6.6 for sub-sample 1, 2, 3, and 4, respectively. In accordance with the distinct values of T_n, the bandwidths were taken to be 0.4, 0.36, 0.48, and 0.18 for the estimation in four sub-samples (in the order of 1, 2, 3, 4).

The results show that our estimates track the true kernel densities very close and outperform the naïve estimates for all four sub-samples. Although neither the 5%–95% confidence intervals of our estimated densities nor that of the naive densities are able to contain the entire true densities, our estimates have much smaller bias than the naïve ones. The estimated results imply that failing to account for the property we discussed in Section 3.1 can lead to significant bias of $\hat{f}_{X^*}$.

Table 1
Estimation results of parameters.

Data	Parameters	Estimates with $\lambda \neq 0$ for our density estimator	Estimates with $\lambda = 0$ for naïve estimator
Sub-sample 1	μ	0.4733 (0.0148)	0.4315(0.0131)
	σ	1.2467 (0.0186)	1.1979 (0.0160)
	λ	0.0883 (0.0033)	–
Sub-sample 2	μ	0.0229 (0.0069)	0.0248(0.0061)
	σ	0.5734 (0.0145)	0.5326 (0.0100)
	λ	0.0965 (0.0033)	–
Sub-sample 3	μ	−0.0136 (0.0041)	−0.0113(0.0035)
	σ	0.3334 (0.0091)	0.3124(0.0074)
	λ	0.0958 (0.0031)	–
Sub-sample 4	μ	−0.0361 (0.0036)	−0.0313(0.0028)
	σ	0.2758 (0.0069)	0.2582 (0.0068)
	λ	0.0940 (0.0033)	–

7. Conclusions

In this paper, we consider the widely admitted ill-posed inverse problem for measurement error models. We show that measurement error models for self-reporting data are well-posed under the assumption that the probability of reporting truthfully is nonzero, which is supported by empirical evidences. This optimistic result suggests that researchers should not ignore the point mass at zero in the measurement error distribution when they model measurement errors in self-reported data. In fact, this discontinuity in the error distribution implies that in general we may achieve much faster rate of convergence for an estimator of the latent distribution than people thought before in the literature. To illustrate the implications of our main results, we analyze the well-known classical measurement errors case and provide the conditions under which the deconvolution problem is well-posed. When the deconvolution problem is well-posed, we also present the convergence rate of the deconvolution estimator for the latent distribution. The well-posedness of our measurement error models also implies that of certain instrumental variable models. We will explore this possibility in our future research.

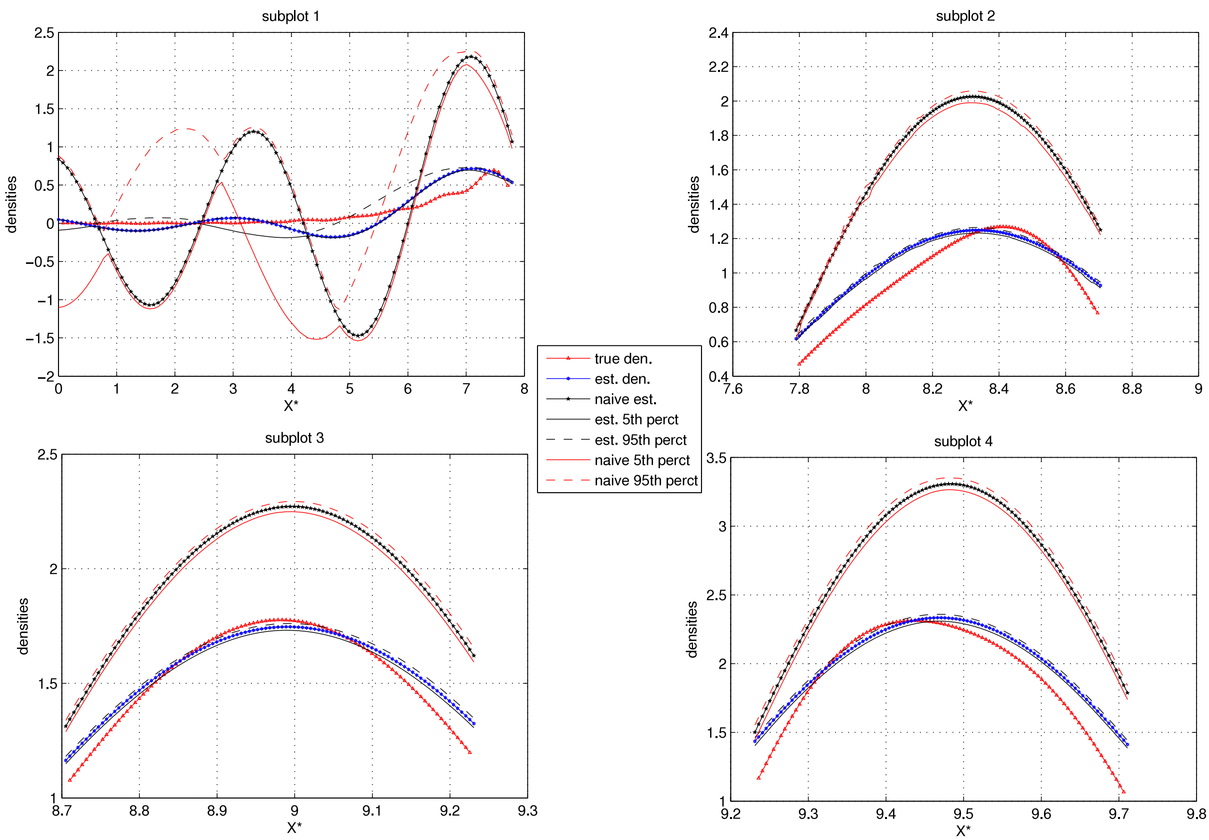

Fig. 7. Estimation results: densities.

Appendix

Proof of Theorem 1. The result is an application of Theorem 3.4 in Kress (1999). The theorem states that if $C : \Phi \to \Phi$ is a compact operator defined on a normed space Φ, and $(I-C)$ is injective, then the inverse operator $(I-C)^{-1} : \Phi \to \Phi$ exists and is bounded, i.e., the problem $(I-C)\phi = f$, for all $f \in \Phi$ is well-posed.

To prove our theorem using this result, we work on Eq. (7).[24] First we show that $f_X \in L^2$ implies $D_\lambda^{-1} f_X \in L^2$. According to the definition of D_λ^{-1}, we have

$$(D_\lambda^{-1} f_X)(x) = \frac{f_X(x)}{\lambda(x)}.$$

Recall that $\lambda(x)$ is bounded below, then $1/\lambda(x)$ has an upper bound, denoted by M_λ. Therefore we have

$$\begin{aligned}\|D_\lambda^{-1} f_X\|_2 &= \left(\int_{-\infty}^{+\infty} \left|\frac{f_X(x)}{\lambda(x)}\right|^2 dx\right)^{\frac{1}{2}} \\ &\le M_\lambda \left(\int_{-\infty}^{+\infty} |f_X(x)|^2 dx\right)^{\frac{1}{2}} \\ &= M_\lambda \|f_X(x)\|_2 \\ &< \infty,\end{aligned}$$

where in the last step we use the fact that $f_X \in L^2$. The inequality implies that $D_\lambda^{-1} f_X \in L^2$, and the operator D_λ^{-1} is bounded. Similarly, it is readily to prove $\|(I - D_\lambda) f_{X^*}\|_2 \le M_{1-\lambda} \|f_{X^*}\|_2$, where $M_{1-\lambda}$ is the upper bound of $1 - \lambda(x)$. Consequently, $(I - D_\lambda) f_{X^*} \in L^2$.

Next, we prove the operator $D_\lambda^{-1} L_g (I - D_\lambda)$ is compact on L^2 under Condition 3. The proof is a direct application of Theorem 2.16 in Kress (1999). This theorem states that if two operators $A : X \to Y$ and $B : Y \to Z$ are both bounded and linear, and one of the operators is compact, then $BA : X \to Z$ is compact. Let $X = Y = Z = L^2$, $A = I - D_\lambda$, and $B = L_g$, then L_g is compact by assumption and hence bounded. Moreover, we conclude that $(I - D_\lambda)$ is also bounded from the result $\|(I - D_\lambda) f_{X^*}\|_2 \le M_{1-\lambda} \|f_{X^*}\|_2$. Therefore, Theorem 2.16 applies and we know that $L_g(I - D_\lambda)$ is compact. If we apply the theorem again by letting $A = L_g(I - D_\lambda)$ and $B = D_\lambda^{-1}$, we can show that $D_\lambda^{-1} L_g (I - D_\lambda)$ is compact.

To complete the proof, it remains to show that $I + D_\lambda^{-1} L_g (I - D_\lambda)$ is injective. By Condition 1, $L_{X|X^*} = D_\lambda (I + D_\lambda^{-1} L_g (I - D_\lambda))$ is injective. Therefore, for any two distinct functions $f_1, f_2 \in L^2$, we have $L_{X|X^*} f_1 \ne L_{X|X^*} f_2$. Because of the boundedness of the operator D_λ^{-1}, we can derive that $D_\lambda^{-1} L_{X|X^*} f_1 \ne D_\lambda^{-1} L_{X|X^*} f_2$, or equivalently $(I + D_\lambda^{-1} L_g (I - D_\lambda)) f_1 \ne (I + D_\lambda^{-1} L_g (I - D_\lambda)) f_2$. The result means $I + D_\lambda^{-1} L_g (I - D_\lambda)$ is injective.

Now, let the operator C in Kress's Theorem 3.4 be $-D_\lambda^{-1} L_g (I - D_\lambda)$. Then all our arguments before in this proof hold, hence we demonstrated that C is compact and $I - C$ is injective. This completes our proof. □

Proof of Proposition 2. First, we specify the operator $L_{X|X^*}$ and $L_{X|X^*}^{-1}$ in the deconvolution case

$$(L_{X|X^*} f_{X^*})(x) = \int f_\epsilon (x - x^*) f_{X^*}(x^*) dx^*,$$

[24] Without loss of generality, we prove the theorem in L^2 space. The proof can be easily extended to L^p space for $1 \le p \le \infty$.

and

$$(L^{-1}_{X|X^*}f_X)(x^*) = \frac{1}{2\pi}\int e^{-itx^*}\frac{\phi_X(t)}{\phi_\epsilon(t)}dt$$
$$= \int\left(\frac{1}{2\pi}\int\frac{e^{it(x-x^*)}}{\phi_\epsilon(t)}dt\right)f_X(x)dx.$$

By Condition 1, the operator $L_{X|X^*}: L^\infty_{bc} \to L^\infty_{bc}$ is injective. Thus, in order to prove the bijectivity of the operator, it is sufficient to show $L_{X|X^*}$ is also surjective, i.e., $L^{-1}_{X|X^*}f_X \in L^\infty_{bc}$ for any $f_X \in L^\infty_{bc}$. Recall that

$$(L^{-1}_{X|X^*}f_X)(x^*) = \frac{1}{2\pi}\int e^{-itx^*}\frac{\phi_X(t)}{\phi_\epsilon(t)}dt.$$

Then the Fourier transform, i.e., the ch.f. of $L^{-1}_{X|X^*}f_X$ is $\frac{\phi_X(t)}{\phi_\epsilon(t)}$. Notice that the injectivity in Condition 1 implies that the ch.f. $\phi_\epsilon(t)$ is bounded away from zero. Therefore, $\phi_X(t)/\phi_\epsilon(t)$ is bounded if $\phi_X(t)$ is bounded for all t. Furthermore, $\phi_\epsilon(t) = \lambda + (1-\lambda)\phi_{\tilde{\epsilon}}(t)$. Therefore we have

$$\begin{aligned}
\|(L^{-1}_{X|X^*}f_X)\|_\infty &= \sup_{x^*}\left|\frac{1}{2\pi}\int e^{-itx^*}\frac{\phi_X(t)}{\phi_\epsilon(t)}dt\right| \\
&\le \sup_{x^*}\frac{1}{\lambda}\left|\frac{1}{2\pi}\int e^{-itx^*}\phi_X(t)dt\right| \\
&\quad + \sup_{x^*}\left|\frac{1}{2\pi}\int e^{-itx^*}\left(\frac{\phi_X(t)}{\lambda+(1-\lambda)\phi_{\tilde{\epsilon}}(t)} - \frac{\phi_X(t)}{\lambda}\right)dt\right| \\
&\le O(\|f_X\|_\infty) + O\left(\int\left|\phi_X(t)\left(\frac{\frac{1-\lambda}{\lambda}\phi_{\tilde{\epsilon}}(t)}{\lambda+(1-\lambda)\phi_{\tilde{\epsilon}}(t)}\right)\right|dt\right) \\
&= O(\|f_X\|_\infty) + O\left(\int|\phi_X(t)|\,|\phi_{\tilde{\epsilon}}(t)|dt\right).
\end{aligned}$$

Since $|\phi_X(t)|$ is always bounded in L^∞_{bc}, we have

$$\|(L^{-1}_{X|X^*}f_X)\|_\infty = O(\|f_X\|_\infty) + O\left(\int|\phi_{\tilde{\epsilon}}(t)|dt\right).$$

The condition $\int|\phi_{\tilde{\epsilon}}(t)|\mathrm{d}\,t < \infty$ implies that $L^{-1}_{X|X^*}f_X \in L^\infty_{bc}$ if $f_X \in L^\infty$, i.e., $L^{-1}_{X|X^*}: L^\infty_{bc} \to L^\infty_{bc}$ is surjective, hence bijective since the injectivity holds by Condition 1.

Because for any $f_X \in L^\infty_{bc}$, both $\|L^{-1}_{X|X^*}f_X\|_\infty$ and $\|f_X\|_\infty$ are finite, there must exist a constant $M > 0$ such that $\|L^{-1}_{X|X^*}f_X\|_\infty < M\|f_X\|_\infty$, i.e., $L^{-1}_{X|X^*}: L^\infty_{bc} \to L^\infty_{bc}$ is bounded and therefore continuous on L^∞_{bc}. This completes the proof of the first part.

In the second part of the proposition, Eq. (11) implies that the operator L_g with the kernel $g_{\tilde{\epsilon}}(x-x^*)$ is a Hilbert–Schmidt operator, and it is compact. A direct application of Theorem 1 completes the proof of this part. □

References

Blundell, R., Chen, X., Kristensen, D., 2007. Semi-nonparametric IV estimation of shape-invariant Engel curves. Econometrica 75 (6), 1613–1669.

Bollinger, C., 1998. Measurement error in the current population survey: a nonparametric look. Journal of Labor Economics 16 (3), 576–594.

Bound, J., Brown, C., Mathiowetz, N., 2001. In: Heckman, J., Leamer., E. (Eds.), Measurement Error in Survey Data. In: Handbook of Econometrics, vol. 5. North-Holland, Amsterdam, pp. 3705–3843.

Bound, J., Krueger, A., 1991. The extent of measurement error in longitudinal earnings data: do two wrongs make a right? Journal of Labor Economics 9 (1), 1–24.

Carrasco, M., Florens, J.-P., Renault, E., 2007. In: Heckman, J., Leamer., E. (Eds.), Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization. In: Handbook of Econometrics, vol. 6B. North-Holland, Amsterdam.

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

Carroll, R., Ruppert, D., Stefanski, L.A., Crainiceanu, C.M., 2006. Measurement Error in Nonlinear Models: a Modern Perspective, second ed. Chapman & Hall/CRC.

Chen, X., Hong, H., Tamer, E., 2005. Measurement error models with auxiliary data. Review of Economic Studies 72 (2), 343–366.

Chen, X., Hong, H., Tarozzi, A., 2008. Semiparametric efficiency in GMM models of nonclassical measurement errors, missing data and treatment effects, Cowles Foundation Discussion Paper No. 1644.

Chen, X., Reiss, M., 2011. On rate optimality for ill-posed inverse problems in econometrics. Econometric Theory 27 (3), 497–521.

Fan, J., 1991. On the optimal rates of convergence for nonparametric deconvolution problems. Annals of Statistics 19 (3), 1257–1272.

Hadamard, J., 1923. Lectures on Cauchy's Problem in Linear Partial Differential Equations. Yale University Press, New Heaven.

Hall, P., Horowitz, J., 2005. Nonparametric methods for inference in the presence of instrumental variables. Annals of Statistics 33 (6), 2904–2929.

Hesse, C., 1995. Deconvolving a density from partially contaminated observations. Journal of Multivariate Analysis 55 (2), 246–260.

Hu, Y., Ridder, G., 2010. On Deconvolution as a First Stage Nonparametric Estimator. Econometric Reviews 29 (4), 1–32.

Kress, R., 1999. Linear Integral Equations, second ed. Springer-Verlag.

Lewbel, A., Linton, O., 2010. Nonparametric Euler equation identification and estimation, Boston College Department of Economics Working Paper No. 757.

Li, T., 2002. Robust and consistent estimation of nonlinear errors-in-variables models. Journal of Econometrics 110 (1), 1–26.

Li, T., Vuong, Q., 1998. Nonparametric estimation of the measurement error model using multiple indicators. Journal of Multivariate Analysis 65 (2), 139–165.

Naylor, A.W., Sell, G.R., 2000. Linear Operator Theory in Engineering and Sciences. Springer-Verlag.

Newey, W., Powell, J., 2003. Instrumental variable estimation of nonparametric models. Econometrica 71 (5), 1565–1578.

Pedersen, M., 1999. Functional Analysis in Applied Mathematics and Engineering. CRC Press.

Petrov, Y.P., Sizikov, V., 2005. Well-posed, ill-posed, and intermediate problems with applications. V.S.P. Intl Science.

Shen, X., 1997. On methods of sieves and penalization. Annals of Statistics 25 (6), 2555–2591.

Tikhonov, A., 1963. On the solution of incorrectly formulated problems and the regularization method. Soviet Math. Doklady 4 (4), 1035–1038.