Well-posedness of Measurement Error Models for Self-Reported Data*

Yonghong An and Yingyao Hu

Johns Hopkins University

First version: October 2009

This version: December 2010

Abstract

It is widely admitted that the inverse problem of estimating the distribution of a latent variable X^* from an observed sample of X, a contaminated measurement of X^*, is a Fredholm integral equation of the first kind and ill-posed. This paper shows that such measurement error models for self-reporting data are Fredholm integral equations of the second kind and well-posed. The condition for the well-posedness is that the probability of reporting truthfully is nonzero, which is an observed property in validation studies. Comparing with ill-posedness, well-posedness generally can be translated into faster rates of convergence for the nonparametric density estimators of X^*. Therefore, our optimistic result on well-posedness is of importance in economic applications, and one should not ignore the point mass at zero in the error distribution when modeling measurement errors in self-reported data. We also illustrate that the classical measurement error models may in fact be conditionally well-posed given prior information on the distribution of the latent variable X^*. 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 distribution of X^*.

Keywords: *well-posed, conditionally well-posed, ill-posed, inverse problem, Fredholm integral equation, deconvolution, rate of convergence, measurement error model, self-reported data, survey data.*

JEL classification codes: C14

*The authors can be reached at `yan4@jhu.edu` and `yhu@jhu.edu`. 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.

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 sample 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]

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

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 the 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, Chen, and Kristensen (2007), Chen and Reiss (2007), 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 a nonzero probability. The property of truthful-reporting can be observed from validation studies by Bollinger (1998) and Chen, Hong, and Tarozzi (2008). This property also distinguishes survey samples used in economics from sam-

[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 above is violated, then the problem is ill-posed.

ples usually used in statistical literature, where data are generated from certain measurement equipment. Based on this property, we prove that Eq.(1) described earlier is in fact a *Fredholm integral equation of the second kind*, which is generally well-posed. We further employ the existing results in 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 postive 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 errors models with a generally ill-posed setup, such as Eq.(1).

We also discuss the well-known classical measurement error case, where the error structure $f_{X|X^*}(x|x^*)$ may be reduced to $f_\epsilon(x - x^*)$. We refer to the concept of *conditional (Tikhonov) well-posedness* to discuss the relationship between the error distribution f_ϵ and the property of ill-posedness. Basically, an inverse problem is conditionally well-posed if it is ill-posed on a function space $\mathcal{S}$, but still well-posed on some subsets of $\mathcal{S}$. Based on this concept, another point we make in this paper is that it is important to find such subsets of $\mathcal{S}$ that is large enough to contain the usual density estimator $\widehat{f}_X$ of f_X. If we find such a subset containing $\widehat{f}_X$, the inverse problem in the measurement error models can be treated as well-posed. Consequently, a nicely behaved nonparametric estimator for distribution of X^* can be obtained on the subsets.

This paper points out that the property of self-reporting errors implies the well-posedness of the the inverse problems in measurement error models. Our findings are important in economic applications in that our 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 the sense that the achievable rates of convergence for estimating f_{X^*} may be much faster than what available in the literature. 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]

The paper is organized 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 illustrate the conditional well-posedness for models of classical measurement error, and compare the rates of convergence for $\widehat{f}_{X^*}$ in well-posed, conditional well-posed, and ill-posed cases. In section 5, we provide Monte Carlo evidence of the improvement 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 $\{X_i\}_{i=1,...,n}$ 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^2 space over the real line, i.e., $L^2(\mathbb{R})$, with $f_X, f_{X^*} \in L^2$ unless we specify the space otherwise.[4]

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

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

where the operator $L_{X|X^*} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is defined as

$$\left(L_{X|X^*} h\right)(x) = \int f_{X|X^*}(x|x^*)\, h(x^*)\, dx^*, \forall h \in L^2(\mathbb{R}).$$

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

[3] Also see Carrasco, Florens, and Renault (2007) for more interesting examples.

[4] Our results apply to L^p space for $1 \leq p \leq \infty$.

Definition 1. *(Carrasco, Florens, and Renault (2007), p.5670)*
The equation $L_{X|X^*} f_{X^*} = f_X$ $(f_{X^*}, f_X \in L^2)$ *is well-posed if* $L_{X|X^*}$ *is bijective and the inverse operator* $L^{-1}_{X|X^*} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ *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. $f_{X|X^*}$ *is known and* $L_{X|X^*}$ *is injective.*[5]

This assumption guarantees that the left inverse of $L_{X|X^*}$ exists and f_{X^*} is uniquely identified from Eq. (1).[6] 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

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

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 $\widehat{f}_{X^*}$ on the data of X, i.e., the bias in $\widehat{f}_{X^*}$, $L^{-1}_{X|X^*} \left(\widehat{f}_X - f_X \right)$, 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

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

[6] 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.

and a small estimation error in $\widehat{f}_X$ might cause a huge bias in $\widehat{f}_{X^*}$. As we mentioned before, when the problem is ill-posed on the space L^2, it may still be well-posed on some subsets of L^2 if some prior information of f_{X^*} is available, i.e., the problem is conditionally well-posed. We introduce the rigorous definition of conditionally well-posed as follows:

Definition 2. *(Petrov and Sizikov (2005), p.157) An operator equation*

$$L_{X|X^*} f_{X^*} = f_X$$

with $f_{X^*}, f_X \in L^2(\mathbb{R})$ 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 $\Upsilon \subset L^2(\mathbb{R})$;
(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 $\Psi \subset L^2(\mathbb{R})$.

In general, it is not difficult to find such subsets Υ and Ψ. But it is crucial to find a set Ψ such that a density estimator $\widehat{f}_X$ is in it. Given the sets Υ and Ψ, we may then just focus on solving the inverse problem defined from Υ to Ψ, which is well-posed. We take Theorem 8.2.1 in Lebedev, Vorovich, and Gladwell (2002) as an illustrating example of the existence of Υ and Ψ. Suppose the problem of solving f_{X^*} from $L_{X|X^*} f_{X^*} = f_X$ is ill-posed on $L^2(\mathbb{R})$ and the operator $L_{X|X^*}$ is a continuous one-to-one operator on $L^2(\mathbb{R})$. The theorem states that if we know that f_{X^*} is in a compact set $\Upsilon \in L^2(\mathbb{R})$, then the problem is well-posed on Υ to Ψ, where $\Psi \equiv L_{X|X^*}\Upsilon$. This result is also employed by Newey and Powell (2003) to overcome the ill-posedness of estimating a nonparametric structural function.[7]

3 Measurement error models for self-reporting data

In this section, we show the well-posedness of measurement error models for self-reporting data. We first present a property observed in validation studies that in-

[7] See p.1569 in Newey and Powell (2003) for details.

dividuals 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.

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.

A key property of self-reporting errors is that it has a nonzero probability of being equal to zero. This can be seen from a validation study by Chen, Hong, and Tarozzi (2008), which provides 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 Figure 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.[8]

In fact, Bollinger (1998) provides estimates of the probability of reporting truthfully

[8]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.

in CPS. That 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, Brown, and Mathiowetz (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 cited above, we plot histogram of social security earnings X^* for those X^* that are exactly equal to X, the self-reported earnings in Figure 2. The histogram shows that people report truthfully almost at every earning level, which implies 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 Figure 3: if she remembers the true value, an interviewee 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 value, which may be considered as unintentional misreporting.[9] Admittedly, we can't distinguish intentional misreporting from unintentional misreporting without further information. The

[9]This conjecture distinguishes self-reported data in economics from self-reported data in other fields, say biostatistics. According to Carroll, Ruppert, Stefanski, and Crainiceanu (2006), validation data are also available in biostatistics. However, the existence of validation data in biostatistics does not imply the error distribution has a point mass. The reason is that in biostatistics, interviewees hardly know exactly the value they need to report. For instance, a person hardly knows his weight as accurate as measured by weight scales. Consequently, there is no way for interviewees report the true value with positive probability.

example on self-reported occupations we mentioned 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.[10]

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 bonded away from nonzero, i.e.*

$$\lambda(x^*) \equiv \Pr\left(X = X^* | X^* = x^*\right) \geq c > 0 \ \ for \ \ any \ \ x^*.^{11}$$

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

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

where $\delta(\cdot)$ is a Dirac delta function and $g\left(x|x^*\right)$ is the conditional density corresponding to misreporting errors.[12]

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

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

[10]However, we do have the risk that the zero 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, Hong, and Tamer (2005)).

[11]The probability $\lambda(x^*)$ is a positive number that is independent of the sample size. We will consider the case where $\lambda(x^*)$ may depend on the sample size in Section 4.

[12]The misreporting error density $g\left(x|x^*\right)$ is corresponding to both unintentional and intentional misreporting in Figure 3, and the two sources are indistinguishable without further information.

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

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

where I is an identity operator defined on L^2, $D_\lambda : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is the multiplication operator defined as

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

and the operator $L_g : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is defined as

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

Since $\lambda(z) \geq c > 0$, this operator equation can be written as

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

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 in Eq.(6) is compact.*

Since the compactness of L_g is corresponding to certain properties of the density $g(\cdot|\cdot)$, the condition is less abstract as it looks. 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.).[13] Hence if we assume

$$\left|\left| g(\cdot|\cdot) \right|\right|_2 < \infty,$$

then the operator L_g is compact on $L^2(\mathbb{R})$, i.e., a sufficient condition for compactness of the operator L_g is that the error density $g(\cdot|\cdot)$ is square integrable.

We summarize the well-posedness of problem (4) in the following theorem.

Theorem 1. *Under Conditions 1, 2, and 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 reverse the pessimistic perspective on this inverse problem. Theorem 1 implies that the estimator of f_{X^*} based on equation (2) with self-reported data should perform well in general because 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. Our results 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.[14]

[13] 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., $||k||_2 = \int_I \int_I |k(s,t)|^2 dsdt < \infty$.

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

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.[15] 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 kinds of problems, 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 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 $\widehat{\omega}_n$ denote the estimates of K and ω, respectively, where n is the sample size. Carrasco, Florens, and Renault (2007) give the rate of convergence in estimating f_{X^*}, we restate the result in the following theorem.[16]

Proposition 1. *(Carrasco, Florens, and Renault, 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* $n \to \infty$.

Thus far, we assume both the probability $\lambda(x^*)$, and the error density $f_{X|X^*}$ are known. If this is the case, the results above show that the convergence rate for the estimator $\widehat{f}_{X^*}$ is the same as that for the estimator $\widehat{f}_X$, 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 is feasible according to Shen (1997). In many economic applications, the error density $f_{X|X^*}$ and probability $\lambda(x^*)$ may be unknown and need to be estimated. The impacts

[15] Given Condition 2 is satisfied, the asymptotic properties of $\widehat{f}_{X^*}$ does not change in 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 $\widehat{f}_{X^*}$.

[16] $\| \cdot \|$ is L^2-norm. For a linear operator K, $\| K \| \equiv \sup_{\|\phi\|=1} \| K\phi \|$.

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 $\widehat{\lambda}(\cdot)$ are not slower than that of $\widehat{f}_X$, then estimating $f_{X|X^*}$ and $\lambda(\cdot)$ 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 $\widehat{\lambda}(\cdot)$. The impacts of estimating error density 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 measurement error and the true data, Chen, Hong, and Tamer (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 the nonparametric estimation $\widehat{f}_{X^*}$ 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 $\widehat{f}_{X^*}$.[17] 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 one 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 (2007) for details). Nevertheless, the slow rate can be well illustrated by the deconvolution case: 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.[18]

[17] If regularization schemes are employed to approximate the solution of an ill-posed problem, the rate of convergence can be fast under some additional assumptions, please see Carrasco, Florens, and Renault (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.

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

4 A further discussion on the classical error case

In this section, we illustrate that if some prior information of f_{X^*} is available, we usually can narrow the sets on which the problem is defined such that the problem is well-posed on the narrowed subsets. In other words, the original problem is conditionally well-posed. Moreover, we argue that conditional well-posedness rather than well-posedness is sufficient in many economic applications.

In order to conduct our analysis, we assume in this section that the error is classical, i.e., $X = X^* + \epsilon$, where the true value X^* is independent of the measurement error ϵ. Therefore, we have

$$f_{X|X^*}(x|x^*) = f_\epsilon(x - x^*). \tag{9}$$

For simplicity, we restrict the space on which the problem is defined to all the bounded functions with bounded Fourier transform in L^∞, but the basic idea does not rely on this simplification, and the results can apply to general L^p space, where $1 \leq p \leq \infty$. A result we will repeatedly use in this section is that a linear operator is continuous if and only if it is bounded.[19]

We first analyze the implication of the simplification in Eq. (9) without Condition 2. Then we combine Eq. (9) and Condition 2 to show the well-posedness in the classical error case.

If X^* is independent of ϵ, then it is known that the characteristic functions of f_X, f_{X^*}, and f_ϵ (denoted by ϕ_X, ϕ_{X^*}, and ϕ_ϵ, respectively) have the following relation:

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

Assumption.1 guarantees that $\phi_\epsilon(t) \neq 0$ for any real t. Therefore, the density f_{X^*} can be recovered from its characteristic function $\phi_{X^*}(t) = \phi_X(t)/\phi_\epsilon(t)$ through $\frac{1}{2\pi}\int e^{-itx}\frac{\phi_X(t)}{\phi_\epsilon(t)}\mathrm{d}t = \frac{1}{2\pi}\int e^{-itx}\phi_{X^*}(t)\mathrm{d}t$. Hence the deconvolution here is well-defined.

In empirical applications, however, the density f_X needs to be estimated by using the

[19] See Theorem 2.5. in Kress (1999).

observed data $\{X_i\}_{i=1,\ldots,n}$. A popular estimator for f_X is as follows:

$$\begin{aligned} \hat{f}_X &= \frac{1}{2\pi}\int e^{-itx}\hat{\phi}_X(t)\mathrm{d}t \\ \hat{\phi}_X(t) &= \hat{\phi}_n(t)\phi_K(\frac{t}{T_n}), \end{aligned} \tag{10}$$

where $\hat{\phi}_n(t)$ is the empirical characteristic function defined by

$$\hat{\phi}_n(t) = \frac{1}{n}\sum_{i=1}^{n} e^{itX_i},$$

and $\phi_K(\frac{t}{T_n})$ is the Fourier transform of the kernel function K 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 such that $\phi_K(t) = 0$ for $|t| > 1$. In order to let $\hat{\phi}_X(t)$ 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) suggests that we need

$$T_n = O\left(\frac{n}{\log n}\right)^{\gamma} \text{ for } \gamma \in \left(0, \frac{1}{2}\right). \tag{11}$$

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

$$\begin{aligned} \hat{f}_{X^*}(x^*) &= \frac{1}{2\pi}\int e^{-itx^*}\frac{\hat{\phi}_X(t)}{\phi_\epsilon(t)}\mathrm{d}t \\ &= f_{X^*}(x^*) + \frac{1}{2\pi}\int e^{-itx^*}\frac{\hat{\phi}_X(t)-\phi_X(t)}{\phi_\epsilon(t)}\mathrm{d}t. \end{aligned}$$

The equation shows that we need to focus on the second term of the last line when we analyze the well-posedness of the inverse problem. In the remaining part of this section, we explore the well-posedness of the problem for three categories of error distributions.

4.1 The case with supersmooth error distributions

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

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

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. For simplicity of our analysis, we assume $d = d_1$.

Intuitively, since $\phi_\epsilon(t)$ converges to zero at an exponential rate, which is much faster than $\hat{\phi}_X(t) - \phi_X(t)$ does when $t \to \infty$, it must be true that either the integral

$$bias\left(\hat{f}_{X^*}(x)\right) = \frac{1}{2\pi}\int e^{-itx^*}\frac{\hat{\phi}_X(t) - \phi_X(t)}{\phi_\epsilon(t)}\mathrm{d}t$$

does not exist, or a small bias of $\hat{\phi}_X(t)$ causes a huge bias of $\hat{f}_{X^*}$. In either cases, the problem is ill-posed on L^∞. We show in the following proposition that the problem might be well-posed on some subsets of L^∞, i.e., the problem might be conditionally well-posed, given certain information on the latent density f_{X^*}. The prior information we need is as follows:

Condition 4. $|\phi_{X^*}(t)| = O\big(|t|^{-\tau}\big)$ *as* $|t| \to \infty$ *for some constants* $\tau > 1$.

This condition requires that the degree of smoothness τ is known. According to Hu and Ridder (2010), the degree of smoothness of a distribution function is related to the "rang-restricted" properties of the function which has a bounded support. In some applications, the assumption of known degree of smoothness is strong. We employ this prior information to exemplify that well-posedness is achievable for given prior information of the unknown density f_{X^*}. Naturally, different prior information may imply different subsets on which the problem is well-posed.

In order to show the conditional well-posedness, we define the operator

$$\begin{aligned} L_{X|X^*} &: \Upsilon \to \Psi \\ \left(L_{X|X^*}h\right)(x) &= \int f_\epsilon\left(x - x^*\right) h\left(x^*\right) dx^* \end{aligned} \tag{12}$$

where

$$\Upsilon = \left\{ f \in L^\infty(\mathbb{R}) : \sup_{t\in\mathbb{R}} |\phi_f(t)| < \infty \text{ and } |\phi_f(t)| = O\left(|t|^{-\tau}\right) \text{as } t \to \infty \text{ for } \tau > 1 \right\},$$

$$\Psi = \left\{ f \in L^\infty(\mathbb{R}) : \sup_{t\in\mathbb{R}} |\phi_f(t)| < \infty \text{ and } |\phi_f(t)| = O\left(|t|^{-\tau}\exp(-|t|^\beta/\rho)\right) \text{as } t \to \infty \text{ for } \tau > 1 + d \right\}.$$

and ϕ_f stands for the Fourier transform of function f. Given these specifications, we have the following results.

Proposition 2. *Suppose conditions 1, 4, and Eq. (9) hold. The operator $L_{X|X^*} : \Upsilon \to \Psi$ in (12) is bijective, and its inverse $L^{-1}_{X|X^*} : \Psi \to \Upsilon$ is continuous. Thus, problem (2) is conditionally well-posed. However, the density estimator $\hat{f}_X$ in (10) is not in Ψ in the sense that $\phi_{\hat{f}}(T_n) = O_p(T_n^{-r_1})$ as $T_n = O(n^{r_2})$ for some positive constants r_1 and r_2.*

Proof See Appendix. ■

The result that the usual kernel density estimator $\hat{f}_X$ is not in Ψ implies that it is not enough for empirical applications to just find spaces Υ and Ψ because the well-posedness over Ψ does not help back out the latent density f_{X^*}. On the one hand, it is interesting to find the spaces on which the operator behaves well. On the other hand, it is also important to realize that the empirical density has to be in the space Ψ so that the theoretical results on well-posedness may be useful for empirical research.

4.2 The case with ordinary smooth error distributions

Fan (1991) defines that an ordinary smooth distribution of ϵ satisfies

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

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

If the distribution of ϵ is ordinary smooth, then $|\hat{\phi}_X(t) - \phi_X(t)|$ may converge to zero faster than $\phi_\epsilon(t)$ does as $t \to \infty$, i.e., $\frac{\hat{\phi}_X(t)-\phi_X(t)}{\phi_\epsilon(t)}$ tends to zero as $t \to \infty$, thus the left inverse $L^{-1}_{X|X^*}$ may be continuous over certain subspace of L^∞. We formalize this intuition in the following proposition. Define the operator

$$\begin{aligned} L_{X|X^*} &: \Upsilon \to \Psi \\ \left(L_{X|X^*}h\right)(x) &= \int f_\epsilon\left(x - x^*\right) h\left(x^*\right) dx^* \end{aligned} \tag{13}$$

where

$$\Upsilon = \left\{ f \in L^\infty(\mathbb{R}) : \sup_{t\in\mathbb{R}} |\phi_f(t)| < \infty \text{ and } |\phi_f(t)| = O\left(|t|^{-\tau}\right) \text{as } t \to \infty \text{ for } \tau > 1 \right\},$$

$$\Psi = \left\{ f \in L^\infty(\mathbb{R}) : \sup_{t\in\mathbb{R}} |\phi_f(t)| < \infty \text{ and } |\phi_f(t)| = O\left(|t|^{-\tau}\right) \text{as } t \to \infty \text{ for } \tau > 1 + d \right\}.$$

Proposition 3. *Suppose conditions 1, 4, and Eq. (9) hold. The operator $L_{X|X^*} : \Upsilon \to \Psi$ in (13) is bijective, and its inverse $L^{-1}_{X|X^*} : \Psi \to \Upsilon$ is continuous. Thus, problem (2) is conditionally well-posed. Moreover, the density estimator $\hat{f}_X$ in (10) may be in Ψ in the sense that $\phi_{\hat{f}}(T_n) = O_p(T_n^{-r_1})$ as $T_n = O(n^{r_2})$ for some positive constants r_1 and r_2.*

Proof See Appendix. ■

In the proof, we also show that the estimator $\widehat{f_X}$ can be in the set Ψ if $\gamma \leq \frac{1}{2(1+\tau)} \in (0, 1/4)$, where γ is the parameter in $T_n = O\left(\frac{n}{\log n}\right)^\gamma$. In this case, the conditional well-posedness guarantees the deconvolution estimator $\widehat{f_{X^*}}$ has a rate described in Proposition 1.

This proposition implies that the problem (2) may be conditionally well-posed and the deconvolution estimator $\hat{f}_{X^*} = \frac{1}{2\pi}\int e^{-itx^*}\frac{\hat{\phi}_X(t)}{\phi_\epsilon(t)}\mathrm{d}t$ is well-defined when the error term has an ordinary smooth distribution. In order to obtain a well-behaved estimator for f_{X^*}, what we really need is whether the operator $L_{X|X^*}$ has a continuous left inverse

over a space containing the estimator $\hat{f}_X$ for some T_n. In other words, the problem may be treated as a well-posed one given a suitable set Ψ. In this sense, many ill-posed problems in economic literature may be solved as well-posed ones if some prior information on f_{X^*} is available.

4.3 Well-posedness under condition 2

Having shown in Section 3 that measurement error models of self-reporting data are well-posed, we further explore the implications of condition 2 on the estimation of f_{X^*} when the error is classical.

In this section, we assume that $\lambda(x^*) = \lambda$ is a constant for simplicity.[20] Our discussion can be extended to the general case straightforwardly. On the other hand, we start the discussion with the case where the probability of truth-reporting $\lambda = \lambda(n)$ converges to zero as the sample size n goes to infinity. Denote the probability by $\lambda_n \equiv \lambda(n)$. Notice that this is a relaxation of condition 2. The condition is assumed to be true at the population level, hence when sample size n goes to infinity, the probability of truth-reporting is still strictly positive under this condition. However, we relax this condition in the sense that we allow the probability to converge to zero as sample size increases. This generalization of the probability λ_n indicates that the proportion of people who report truthfully shrinks with the increase of the sample size n. Accordingly, the error distribution is

$$\begin{aligned} f_{X|X^*}(x|x^*) &= f_\epsilon(x - x^*) \\ &= \lambda_n \times \delta(x - x^*) + (1 - \lambda_n) \times g_{\tilde{\epsilon}}(x - x^*). \end{aligned} \tag{14}$$

In the expression of the density above, both λ_n and $1 - \lambda_n$ depend on sample size n while $\delta(x - x^*)$ and $g_{\tilde{\epsilon}}(x - x^*)$ are from population. As we discussed above, when $n \to \infty$ this density goes to $g_{\tilde{\epsilon}}(x - x^*)$ which does not have a point mass.

Let $\phi_\epsilon(t)$ and $\phi_{\tilde{\epsilon}}(t)$ denote the characteristic functions of f_ϵ and $g_{\tilde{\epsilon}}$, respectively.

[20] 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.

Eq.(14) then implies that

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

Next, we show that $\phi_\epsilon(t)$ is ordinary smooth under the following condition:

Condition 5. *i)* $\phi_{\tilde{\epsilon}}(t) = o(|t|^{-\beta})$ *with* $\beta > 0$, *as* $|t| \to \infty$.

ii) $\lambda_n = O(T_n^{-d})$ *for any* $\beta \geq d > 0$, *where* $T_n \to \infty$ *as* $n \to \infty$.

Assumption 5(i) implies that the error ϵ is either ordinary smooth of order lower than β or supersmooth. And assumption 5(ii) implies that the probability λ_n may converge to zero at the rate of $O(T_n^{-d})$ as $T_n \to \infty$. The requirement $\beta \geq d$ implies that $\phi_\epsilon(T_n) = O(\lambda_n)$, and therefore, $\phi_\epsilon(t)$ is ordinary smooth of order d. Notice that β and d may be any finite constant, i.e., $\beta < \infty$ and $d < \infty$, when $\phi_{\tilde{\epsilon}}$ is supersmooth. We then have

Lemma 1. *Suppose condition 5 and Eq. (14) hold. Then* $\phi_\epsilon(t)$ *is ordinary smooth of order d, and therefore, the results in Proposition 3 hold.*

Proof See appendix. ■

The probability of truth-telling λ_n may be interpreted as the proportion of the error-free sample in the whole sample, i.e., $\lambda_n = n_v/n$, where n is the total sample size while n_v the size of an error-free sample. When combining an error-free sample of a fixed size with a sample containing classical errors, we require $\lambda_n = O\left(\frac{1}{n}\right)$ due to the fixed n_v. This is feasible when $\phi_{\tilde{\epsilon}}$ is supersmooth. Let $\lambda_n = O(T_n^{-d})$ with $T_n = (n)^\gamma$ and $\gamma \in (0, 1/2)$, which implies that $\lambda_n = O\left(n^{-d\gamma}\right)$. Notice that d may be any finite constant when $\phi_{\tilde{\epsilon}}$ is supersmooth, which implies that we may have $\lambda_n = O\left(\frac{1}{n}\right)$. This result implies that the model with a supersmooth classical error may be ill-posed by Proposition 2 but we may transform the problem to a conditionally well-posed one by combining an error-free sample of a fixed size according to Proposition 3. An interesting implication is that an error-free sample may make the problem conditionally well-posed even if its sample size is relatively small compared with the error-ridden sample.

Next, we discuss the well-posedness under Condition 2. If the probability of truth-reporting $\lambda > 0$ is fixed and does not change as sample size n increases, it is readily to show 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

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

We have the following results:

Proposition 4. *i) Suppose conditions 1, 2, and Eq. (9) hold and the error distribution $g_{\tilde{\epsilon}}$ satisfies*

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

Then problem (2) is well-posed with $L_{X|X^*} : L_{bc}^{\infty} \to L_{bc}^{\infty}$.
ii) Suppose conditions 1 and 2 hold and the error distribution $g_{\tilde{\epsilon}}$ satisfies

$$\int_{\mathbb{R}}\int_{\mathbb{R}} \left|g_{\tilde{\epsilon}}(x-x^*)\right|^2 dx\, dx^* < \infty. \tag{15}$$

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

Proof See appendix. ■

Notice that we do not need prior information on f_{X^*} when the problem is well-posed. The restrictions imposed on the error distribution is also weak compared to Propositions 2 and 3. The reason is that if λ is fixed, the corollary is just a specific case of Theorem 1. Even though it is not as general as Theorem 1, the corollary might be very useful in applications since it assures us to solve a consistent estimator of f_{X^*} with a desirable convergence rate from the sample $\{X_i\}$ for a very general error distribution.

4.4 Rates of convergence for deconvolution estimators

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

Hesse (1995) demonstrates that in L^∞-norm the best uniform rate for convergence of $\hat{f}_{X^*}$ to f_{X^*} is $(\log n/n)^{2/5}$ under mild conditions if the observations are "partially contaminated".[21] 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 what Hesse (1995) gets. To restate the result, we first specify the conditions.[22]

A1. *f_{X^*}, f'_{X^*}, and f''_{X^*} are uniformly absolutely bounded.*

A2. *For any $x > 0$, there exists some $\rho > 0$ such that $P(|X_i| > x) \leq x^{-\rho}$.*

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 $\inf_t |\phi_{X^*}(t)| \geq \lambda$.*

Proposition 5. *(Hesse 1995) Under condition A1-A4 and the choice of optimal bandwidth $\frac{1}{T_n} = c(\frac{\log n}{n})^{1/5}$, we have*

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

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

According to Proposition 5, the rate $(\log n/n)^{2/5}$ is general and achievable for any

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

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

distribution of measurement error. Hence the result permits us to demonstrate how well-posedness can be translated into faster rates of convergence than that of ill-posed problems in general cases. Based on our discussions earlier in this section, when the distribution of measurement error is super-smooth, the inverse problem Eq.(2) is ill-posed, 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, we demonstrated that the ill-posed problem Eq.(2) may be conditionally well-posed. For some specific conditionally well-posed problems, the rate $(\log n/n)^{2/5}$ may be slower than the rate for a classical deconvolution estimator. 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 $(\log n/n)^{2/5}$.[23]

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.

[23]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 $n \to \infty$.

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

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

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 the normalized sinc function:

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

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

$$\phi_K(t) = \begin{cases} 0 & \text{if } |t| > \frac{1}{2} \\ \frac{1}{2} & \text{if } |t| = \frac{1}{2}. \\ 1 & \text{if } |t| < \frac{1}{2}. \end{cases} \tag{16}$$

We present simulation results for sample size $n = 1000$ in Figure 4, 5 and 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 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 given 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, Hong, and Tarozzi (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, Hong, and Tamer (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, Hong, and Tarozzi (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 \$16500 to reduce bias may caused by the topcoding.[24] Following the literature we introduced above, we assume that

[24]See Chen, Hong, and Tarozzi (2008) for detailed description of the topcoding.

the error ϵ, which is defined as $\epsilon = \log X - \log X^*$ is distributed according to the density[25]

$$f_\epsilon(\epsilon) = \lambda\delta(\epsilon) + (1-\lambda)\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\epsilon-\mu)^2}{2\sigma^2}}. \tag{17}$$

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.[26]

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 Figure 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 $\hat{f}_{X^*}$.

[25] 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 $X^* = 16500$, too).

[26] Standard errors of estimated parameters are computed by bootstrap resampling (200 times).

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. Moreover, we illustrate that the ill-posedness of models for classical measurement errors may be fixed and the models may actually be conditionally well-posed, which is sufficient enough for many economic applications. An interesting result is that an error-free sample may make the classical error model, especially with a super-smooth error distribution, conditionally well-posed even if its sample size is relatively small compared to the error-ridden sample. Furthermore, 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.

References

Blundell, R., X. Chen, and D. Kristensen (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., C. Brown, and N. Mathiowetz (2001): "Measurement error in survey data," *Handbook of Econometrics*, Vol. 5, by J.Heckman and E. Leamer. Amsterdam: North-Holland, 3705–3843.

Bound, J., and A. Krueger (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., J.-P. Florens, and E. Renault (2007): "Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization," *Handbook of Econometrics*, Vol. 6B, ed. by J.Heckman and E. Leamer. Amsterdam: North-Holland.

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

Carroll, R., D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu (2006): *Measurement error in nonlinear models: a modern perspective (2nd edition)*. Chapman & Hall/CRC.

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

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

Chen, X., and M. Reiss (2007): "On rate optimality for ill-posed inverse problems in econometrics," *Cowles Foundation Discussion Paper No. 1626*.

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., and J. Horowitz (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., and G. Ridder (2010): "On Deconvolution as a First Stage Nonparametric Estimator," *Econometric Reviews, forthcoming*, 29(4), 1–32.

Kress, R. (1999): *Linear Integral Equations (2nd edition)*. Springer-Verlag.

Lebedev, L., I. Vorovich, and G. Gladwell (2002): *Functional Analysis: Applications in Mechanics and Inverse Problems (2nd Edition)*. Kluwer Academic Publishers.

Lewbel, A., and O. Linton (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., and Q. Vuong (1998): "Nonparametric estimation of the measurement error model using multiple indicators," *Journal of Multivariate Analysis*, 65(2), 139–165.

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

Newey, W., and J. Powell (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., and V. Sizikov (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.

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). First we show $f_X \in L^2$ implies $D_\lambda^{-1} f_X \in L^2$. According to the definition of D_λ^{-1}, we have

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

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

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

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 $\left|\left|(I - D_\lambda) f_{X^*}\right|\right|_2 \leq M_{1-\lambda} \left|\left|f_{X^*}\right|\right|_2$, 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 $\left|\left|(I - D_\lambda) f_{X^*}\right|\right|_2 \leq M_{1-\lambda} \left|\left|f_{X^*}\right|\right|_2$. 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, $L_{X|X^*} = D_\lambda \left(I + D_\lambda^{-1} L_g (I - D_\lambda)\right)$ is injective. Therefore, for any

two distinct functions $f_1, f_2 \in L^2$, we have $L_{X|X^*} f_1 \neq L_{X|X^*} f_2$. Because of the boundedness of the operator D_λ^{-1}, we can derive that $D_\lambda^{-1} L_{X|X^*} f_1 \neq D_\lambda^{-1} L_{X|X^*} f_2$, or equivalently $\left(I + D_\lambda^{-1} L_g \left(I - D_\lambda\right)\right) f_1 \neq \left(I + D_\lambda^{-1} L_g \left(I - D_\lambda\right)\right) f_2$. The result means $I + D_\lambda^{-1} L_g \left(I - D_\lambda\right)$ is injective.

Now, let the operator C in Kress's Theorem 3.4 be $-D_\lambda^{-1} L_g \left(I - D_\lambda\right)$. 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

$$\left(L_{X|X^*} f_{X^*}\right)(x) = \int f_\epsilon \left(x - x^*\right) f_{X^*}\left(x^*\right) \mathrm{d}x^*,$$

and

$$\begin{aligned} \left(L_{X|X^*}^{-1} f_X\right)(x^*) &= \frac{1}{2\pi} \int e^{-itx^*} \frac{\phi_X(t)}{\phi_\epsilon(t)} \mathrm{d}t \\ &= \int \left(\frac{1}{2\pi} \int \frac{e^{it(x-x^*)}}{\phi_\epsilon(t)} \mathrm{d}t\right) f_X(x)\, \mathrm{d}x. \end{aligned}$$

By condition 1, the operator $L_{X|X^*} : \Upsilon \to \Psi$ 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., $L_{X|X^*}^{-1} f_X \in \Upsilon$ for any $f_X \in \Psi$. Recall that

$$\left(L_{X|X^*}^{-1} f_X\right)(x^*) = \frac{1}{2\pi} \int e^{-itx^*} \frac{\phi_X(t)}{\phi_\epsilon(t)} \mathrm{d}t.$$

Then the Fourier transform, i.e., the ch.f. of $L_{X|X^*}^{-1} f_X$ is $\frac{\phi_X(t)}{\phi_\epsilon(t)}$. Notice that condition 1 guarantees that $\phi_\epsilon(t)$ is bounded away from zero, and therefore, $\left|\frac{\phi_X(t)}{\phi_\epsilon(t)}\right|$ is finite. As $|t| \to \infty$, we have $\left|\frac{\phi_X(t)}{\phi_\epsilon(t)}\right| = O\left(|t|^{-\tau}\right)$ with $\tau > 1$ for $f_X \in \Psi$.

We now examine $\left|\left| L^{-1}_{X|X^*} f_X \right|\right|_\infty$.

$$\begin{aligned}
\left|\left| L^{-1}_{X|X^*} f_X \right|\right|_\infty &= \sup_{x^*} \left| \frac{1}{2\pi} \int e^{-itx^*} \frac{\phi_X(t)}{\phi_\epsilon(t)} \mathrm{d}t \right| \\
&\leq \int \left| \frac{1}{2\pi} \frac{\phi_X(t)}{\phi_\epsilon(t)} \right| \mathrm{d}t \\
&\leq \int_{-t_0}^{t_0} \left| \frac{1}{2\pi} \frac{\phi_X(t)}{\phi_\epsilon(t)} \right| \mathrm{d}t + \int_{t_0}^{\infty} \frac{2}{2\pi} M|t|^{-\tau} \mathrm{d}t \\
&< \infty,
\end{aligned}$$

where t_0 and M are some positive constants and $\tau > 1$. The second inequality holds because $\left|\frac{\phi_X(t)}{\phi_\epsilon(t)}\right| = O\left(|t|^{-\tau}\right)$ implies that there exist some positive t_0 and M such that $\left|\frac{\phi_X(t)}{\phi_\epsilon(t)}\right| \leq M|t|^{-\tau}$ when $t > t_0$.

Thus, we conclude that $L^{-1}_{X|X^*} f_X \in \Upsilon$. Because for any $f_X \in \Psi$, both $\left|\left| L^{-1}_{X|X^*} f_X \right|\right|_\infty$ and $\left|\left| f_X \right|\right|_\infty$ are finite, there must exist a constant $N > 0$ such that $\left|\left| L^{-1}_{X|X^*} f_X \right|\right|_\infty < N \left|\left| f_X \right|\right|_\infty$, i.e., $L^{-1}_{X|X^*} : \Psi \to \Upsilon$ is bounded and continuous on Ψ. The first part of our proposition is now proved.

We then consider the estimator $\hat{f}_X$ of f_X in Eq. (10) with ch.f.

$$\hat{\phi}_X(t) = \hat{\phi}_n(t)\phi_K(\frac{t}{T_n}).$$

Since $\hat{\phi}_X(t)$ is associated with $\phi_X(t)$ according to the relationship as follows:

$$|\phi_{\hat{X}}(t)| = |\phi_X(t)| \left[1 + O_p\left(\frac{|\phi_{\hat{X}}(t) - \phi_X(t)|}{|\phi_X(t)|}\right)\right],$$

a sufficient and necessary condition for $\hat{f}_X \in \Psi$ is that

$$|\phi_{\hat{X}}(t)| = O_p\left(|\phi_X(t)|\right),$$

or equivalently,

$$O_p\left(\frac{|\phi_{\hat{X}}(t) - \phi_X(t)|}{|\phi_X(t)|}\right) = O_p(1).$$

Recall that $\hat{\phi}_X(t) = \hat{\phi}_n(t)\phi_K(\frac{t}{T_n})$. It follows that for any $|t| > T_n$, $\hat{\phi}_X(t) = 0$ so

that the condition above holds. However, we demonstrate that when $|t| \leq T_n$, the condition above can't hold. For this purpose, we examine

$$O_p\left(\frac{|\phi_{\hat{X}}(t) - \phi_X(t)|}{|\phi_X(t)|}\right), |t| \leq T_n.$$

Let $T_n = O(\frac{n}{\log n})^{\gamma}, \gamma \in (0, \frac{1}{2})$. According to Lemma 1 in Hu and Ridder (2010), the rate of convergence for $|\hat{\phi}_X(t) - \phi_X(t)|$ is at most $(\frac{\log n}{n})^{\frac{1}{2}-\gamma}$ for $|t| \leq T_n$. This result suggests a geometric convergence rate of $|\hat{\phi}_X(t) - \phi_X(t)|$ equals to $(\frac{\log n}{n})^{\frac{1}{2}-\gamma-\eta}$ for an arbitrary $\eta > 0$.

Recall that $\phi_X(t) = O_p(|t|^{-\tau}\exp(-|t|^{\beta}/\rho))$. By employing $T_n = O(\frac{n}{\log n})^{\gamma}, \gamma \in (0, \frac{1}{2})$, we have $\phi_X(T_n) = O_p\left((\frac{n}{\log n})^{-\tau\gamma}\exp\left(-\left(n/\log n\right)^{\beta}/\rho\right)\right)$ as $n \to \infty$. Consequently,

$$\begin{aligned} O_p\left(\frac{|\phi_{\hat{X}}(T_n) - \phi_X(T_n)|}{|\phi_X(T_n)|}\right) &= O_p\left(\frac{(\frac{\log n}{n})^{\frac{1}{2}-\gamma-\eta}}{(\frac{n}{\log n})^{-\tau\gamma}\exp\left(-\left(n/\log n\right)^{\beta}/\rho\right)}\right) \\ &= O_p\left((\frac{\log n}{n})^{\frac{1}{2}-(1+\tau)\gamma-\eta}\exp\left(\left(n/\log n\right)^{\beta}/\rho\right)\right). \end{aligned}$$

Notice that given $\beta, \rho > 0$ the term $(\frac{\log n}{n})^{\frac{1}{2}-(1+\tau)\gamma-\eta}\exp\left(\left(n/\log n\right)^{\beta}/\rho\right)$ diverges for any τ, γ, and η. Therefore, the density estimator $\hat{f}_X$ in Eq. (10) is not in Ψ. Notice that it is possible to make $\hat{f}_X$ in Ψ by taking $T_n = O((\log n)^{\delta})$ for some suitable δ. But such an estimator $\hat{f}_X$ is not useful for most empirical applications because it converges very slowly to f_X. ■

Proof of Proposition 3. The proof of the bijectivity of $L_{X|X^*}$ is similar to the proof in Proposition 2, we omit it here. It remains to show the existence of an estimator $\hat{f}_X \in \Psi$ for f_X. According to the argument in proof Proposition 2, it is sufficient to show that

$$O_p\left(\frac{|\phi_{\hat{X}}(t) - \phi_X(t)|}{|\phi_X(t)|}\right) = o_p(1)$$

holds for $|t| \leq T_n$.

Follow what we did previously, let $T_n = O(\frac{n}{\log n})^{\gamma}, \gamma \in (0, \frac{1}{2})$,

$$\begin{aligned} & O_p\left(\frac{|\phi_{\hat{X}}(T_n) - \phi_X(T_n)|}{|\phi_X(T_n)|}\right) \\ = \; & o_p\left(\frac{(\frac{\log n}{n})^{\frac{1}{2}-\gamma}}{(\frac{n}{\log n})^{-\tau\gamma}}\right) \\ = \; & o_p\left((\frac{\log n}{n})^{\frac{1}{2}-(1+\tau)\gamma}\right). \end{aligned}$$

In order for $o_p\left((\frac{\log n}{n})^{\frac{1}{2}-(1+\tau)\gamma}\right)$ to be equal to $O_p(1)$, we may take

$$\gamma \leq \frac{1}{2(1+\tau)} \in (0, 1/4).$$

Therefore, the density estimator $\hat{f}_X$ in Eq. (10) may be in Ψ. ■

Proof of Lemma 1. Eq.(14) implies that

$$\begin{aligned} \phi_\epsilon(t) &= \int f_\epsilon(x) e^{it(x)} \mathrm{d}x \\ &= \lambda_n \int \delta(x) e^{itx} \mathrm{d}x + (1-\lambda_n) \int g_{\tilde{\epsilon}}(x) e^{itx} \mathrm{d}x \\ &= \lambda_n + (1-\lambda_n)\, \phi_{\tilde{\epsilon}}(t). \end{aligned}$$

Then, $\phi_\epsilon(T_n)$ satisfies the inequality:

$$\Big|\, \lambda_n - (1-\lambda_n)|\phi_{\tilde{\epsilon}}(T_n)| \,\Big| \leq |\phi_\epsilon(T_n)| \leq \lambda_n + (1-\lambda_n)|\phi_{\tilde{\epsilon}}(T_n)|.$$

Since $(1-\lambda_n)$ is bounded as $n \to \infty$, we have $(1-\lambda_n)|\phi_{\tilde{\epsilon}}(T_n)| = o(|T_n|^{-\beta})$. Condition 4 implies that $|\phi_{\tilde{\epsilon}}(T_n)|$ is dominated by λ_n, i.e.,

$$O\left(\Big|\, \lambda_n - (1-\lambda_n)|\phi_{\tilde{\epsilon}}(T_n)| \,\Big|\right) = O\Big(\lambda_n + (1-\lambda_n)|\phi_{\tilde{\epsilon}}(T_n)|\Big) = O(\lambda_n),$$

which leads to the relationship $|\phi_\epsilon(T_n)| = O(\lambda_n) = O\left(|T_n|^{-d}\right)$. Therefore, $\phi_\epsilon(t)$ is ordinary smooth of order d. The results then directly follow from Proposition 3. ■

Proof of Proposition 4. According to the proof of Proposition 2, we know that ch.f. of $L_{X|X^*}f_X$ is $\phi_X(t)/\phi_\epsilon(t)$. 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

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

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

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

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. Following the argument in proof of Proposition 2, we can also conclude that $L^{-1}_{X|X^*}$ is continuous. This completes the proof of the first part.

In the second part of the proposition, Eq.(15) 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. ■

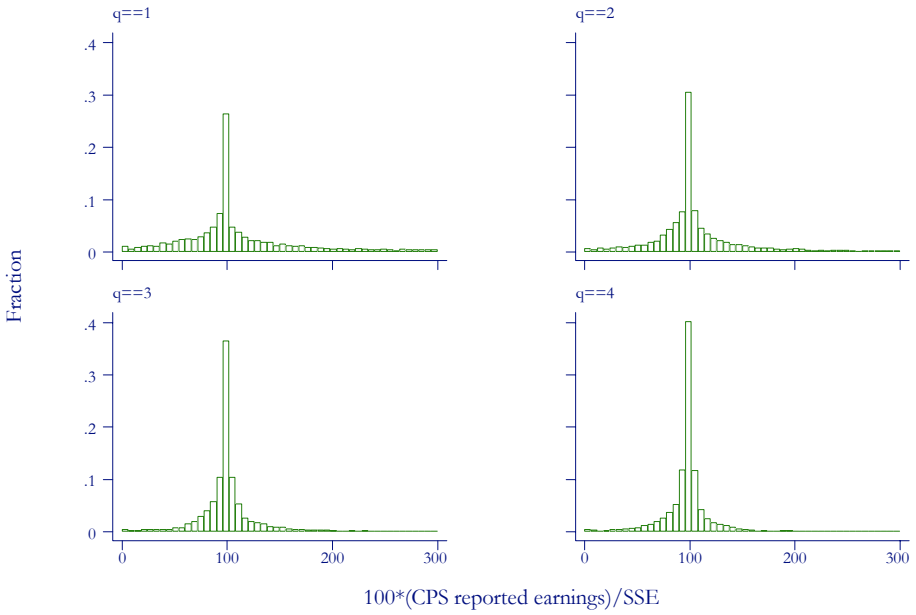

Figure 1: Histograms of measurement error in earnings, by quartile of true (Social Security) earnings. The figure was excerpted from Chen, Hong, and Tarozzi (2008), p.50. The link of the paper is: http://cowles.econ.yale.edu/P/cd/d16a/d1644.pdf.

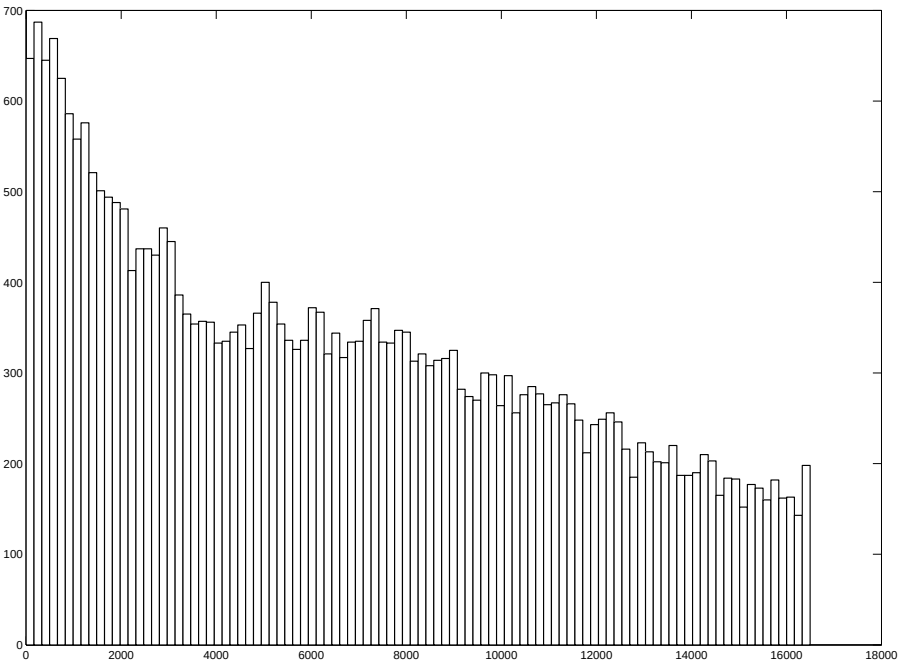

Figure 2: Histogram of X^* given $X^* = X$

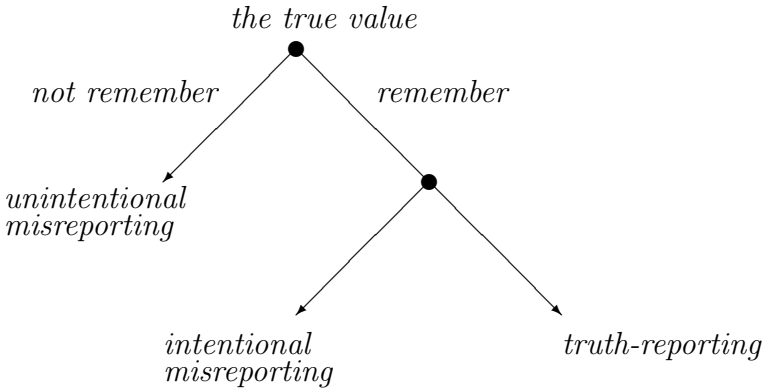

Figure 3: Illustration of Self-reporting

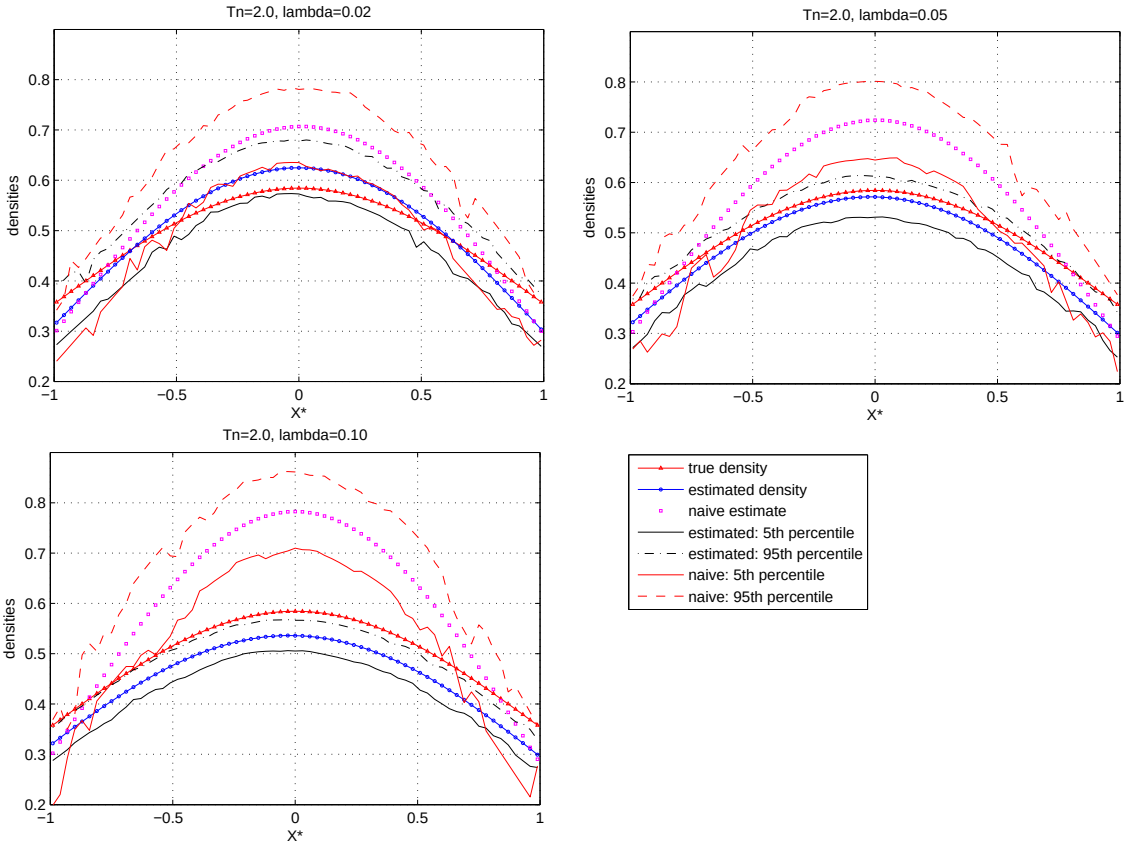

Figure 4: Simulation results: $T_n = 2.0$

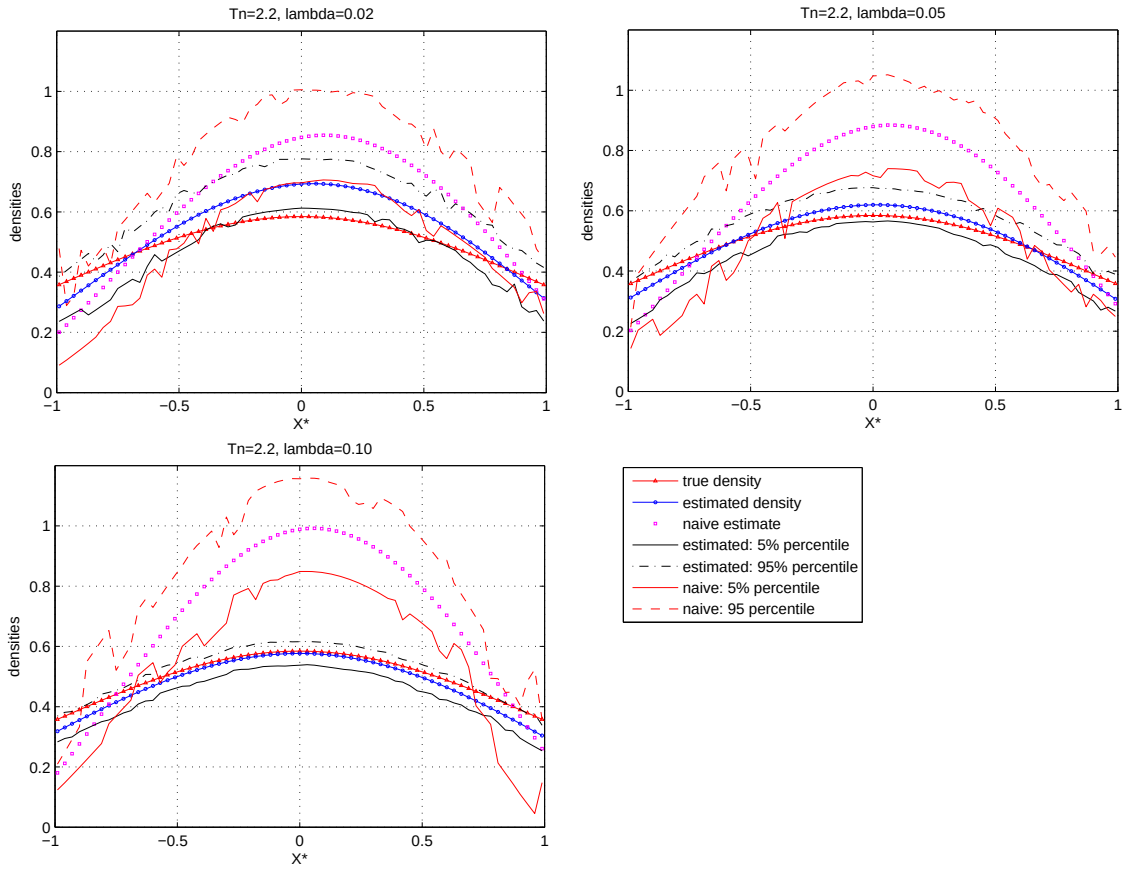

Figure 5: Simulation results: $T_n = 2.2$

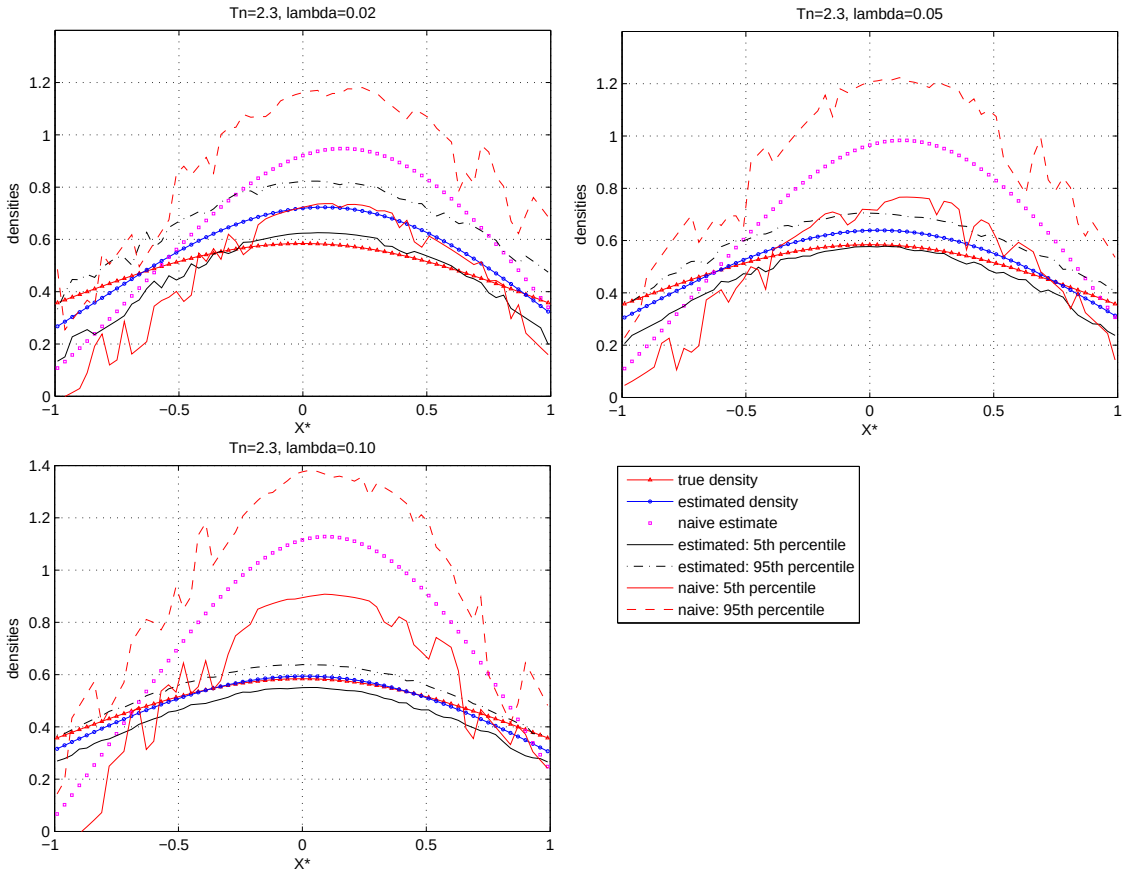

Figure 6: Simulation results: $Tn = 2.3$

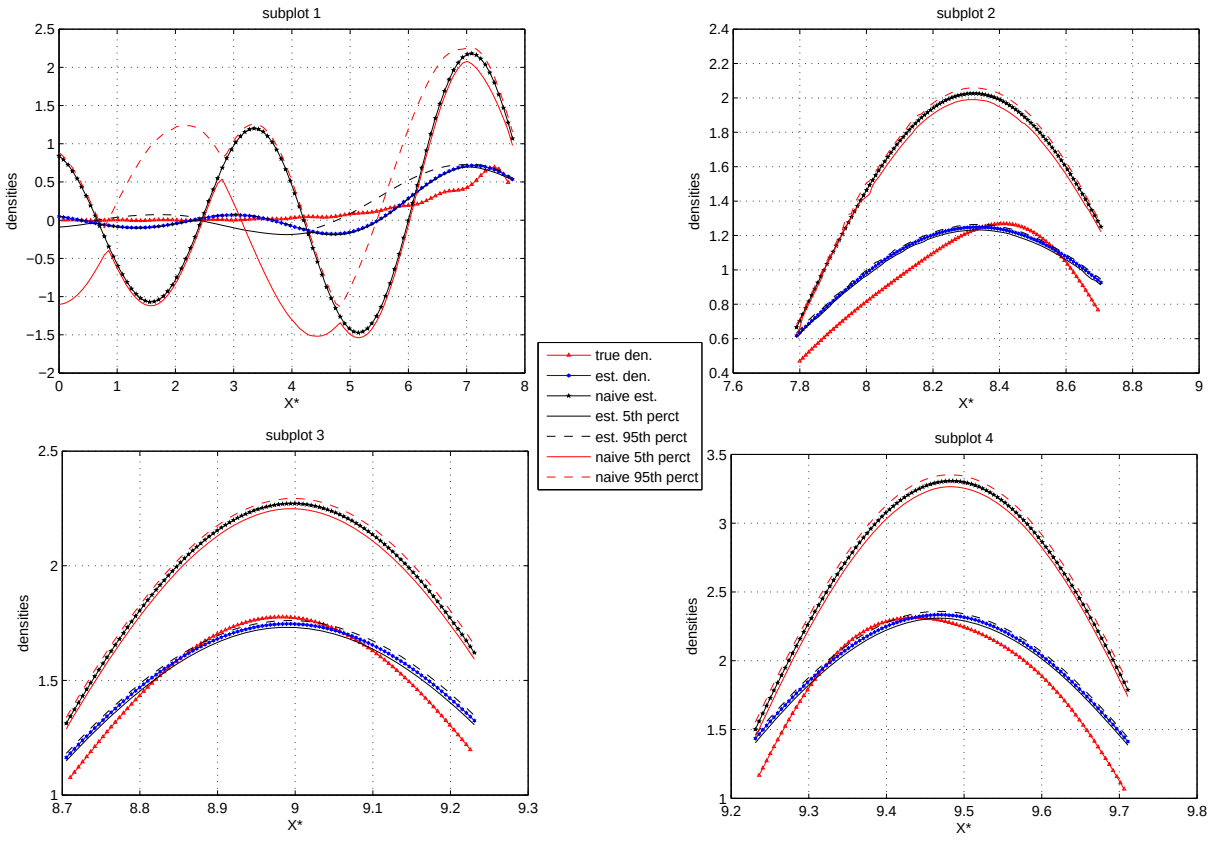

Figure 7: Estimation results: densities

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)	—