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

Yonghong An and Yingyao Hu
Johns Hopkins University

First version: October 2009
This version: November 8, 2009

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 ill-posed. This paper shows that measurement error models for self-reporting data are well-posed, assuming the probability of reporting truthfully is nonzero, which is an observed property in validation studies. This optimistic result suggests that 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^*.

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

1 Introduction

Empirical studies in microeconomics usually involve survey samples, where personal information is reported by the interviewees themselves, and therefore, the corresponding vari-

*The authors can be reached at `yan4@jhu.edu` and `yhu@jhu.edu`. We are grateful to Chris Bollinger, Susanne Schennach, Stephen Shore, Richard Spady, and Tiemen Woutersen for helpful comments or discussions. We also thank Wendy Chi for proofreading the draft. All errors remain our own.

ables 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:

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

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.[1] However, by assuming the probability of reporting truthfully is nonzero, which is an observed property in validation studies, we show that Eq.(1) is a Fredholm equation of the second kind, and therefore, is well-posed.

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). (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, assuming interviewees report truthfully with a nonzero probability. The property can be seen in validation studies by Chen, Hong, and Tarozzi (2008) and Bollinger (1998). This property also distinguishes survey samples used in economics from samples 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. Hence 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 in 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*

[1] According to Hadamard (1923), a well-posed problem should have the following three properties: (i). A solution exists. (ii). The solution is unique. (iii). The solution depends continuously on the data. If any of the three conditions above is violated, then the problem is ill-posed.

(Tikhonov) well-posedness[2] 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 $\mathcal{S}$. Notice that such subsets always exist. 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. We illustrate this implication by associating well-posedness of an inverse problem with the convergence rates of the density estimators.

To our knowledge, we are the first to recognize the implication of the property of self-reporting errors for the well-posedness of the inverse problems in measurement error models. Our findings are important in economic applications in that our results imply the estimation of the latent model f_{X^*} from the observed f_X may not be as technically challenging as previously thought.

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. In section 4, we illustrate the conditional well-posedness for models of classical measurement error. Section 5 concludes. Proofs are in the Appendix.

2 A general setup

We are interested in the 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 $\{X_i\}_{i=1,\dots,n}$ 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). 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 space over the real line, i.e., $L^p(\mathbb{R}), 1 \le p \le +\infty$ with $f_X, f_{X^*} \in L^p$ unless we specify the spaces otherwise.

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

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

[2] The rigorous definition of conditional well-posedness is introduced in the next section.

where the operator $L_{X|X^*} : L^p(\mathbb{R}) \to L^p(\mathbb{R})$ is defined as $\left(L_{X|X^*}h\right)(x) = \int f_{X|X^*}(x|x^*)\, h(x^*)\, dx^*$ for any $h \in L^p(\mathbb{R})$. The well-posedness of the inverse problem (2) is then defined as follows:

Definition 1. *(Carrasco and Florens (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* $L_{X|X^*}^{-1} : L^p(\mathbb{R}) \to L^p(\mathbb{R})$ *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 $f_{X^*}(\cdot)$ so that we make the following assumptions.

Condition 1. $f_{X|X^*}$ *is known and* $L_{X|X^*}$ *is injective.*

This assumption guarantees that the left inverse of $L_{X|X^*}$ exists and f_{X^*} is uniquely identified from Eq. (1).[3] Therefore, we can identify and estimate f_{X^*} as follows:

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

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 the random sample $\{X_i\}$. Let $\widehat{f}$ denote an estimator of f, then the latent model f_{X^*} can be estimated as

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

Since the injectivity of $L_{X|X^*}$ is assumed above, we still need its surjectivity and the continuity of $L_{X|X^*}^{-1}$ 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 $\widehat{f}_{X^*}$, $L_{X|X^*}^{-1}\left(\widehat{f}_X - f_X\right)$, 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 $L_{X|X^*}^{-1}$ is not continuous, then the problem becomes ill-posed and a small estimation error in $\widehat{f}_X$ might cause a huge bias in $\widehat{f}_{X^*}$. As we mentioned before, when the problem is ill-posed on the space L^p, it may still be well-posed on some subsets of L^p, i.e., the problem is conditionally well-posed. We introduce the rigorous definition of conditionally well-posed as follows:

[3] 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.

Definition 2. *(Petrov and Sizikov (2005), p.157) A 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 $L_{X|X^*}^{-1}$ *is continuous on* $\Psi \subset L^p(\mathbb{R})$*.*

As we discussed before, 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 the set Ψ. We may then just focus on solving the equation on the set Ψ, which is well-posed.

3 Measurement error models for self-reporting data

In this section, we show the well-posedness of measurement error models for self-reporting data, which is based on 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.

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 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

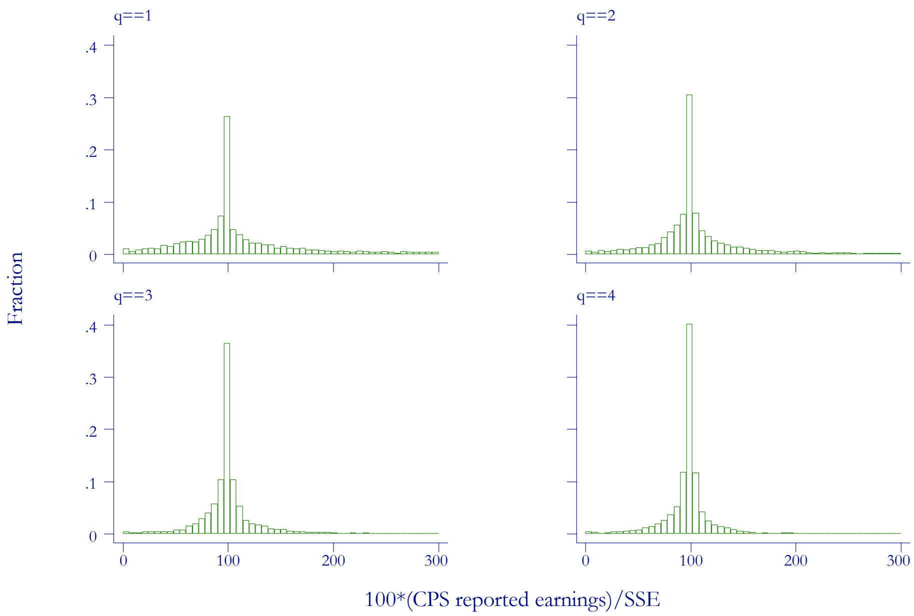

Figure 1: Histograms of measurement error in earnings, by quartile of true (Social Security) earnings. The figure was excerpted from Chen, Hong, and Tarozzi (2008), p.50. The link of the paper is: http://cowles.econ.yale.edu/P/cd/d16a/d1644.pdf.

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.

In fact, Bollinger (1998) provides estimates of the probability of reporting truthfully in CPS. He 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.

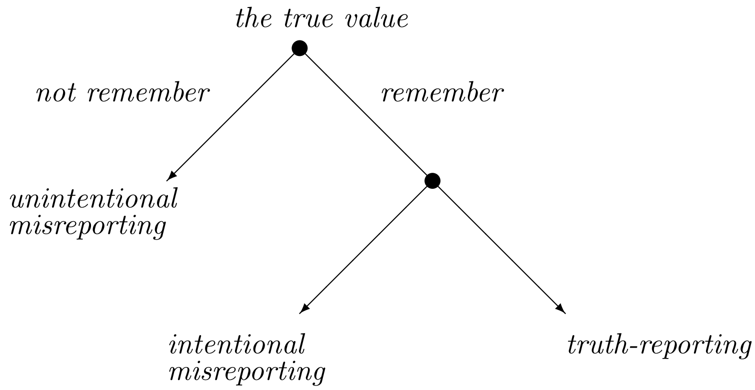

Figure 2: Illustration of Self-reporting

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.

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 2: If he remembers the true value, an interviewee first decides whether to intentionally misreport the truth or not. Empirical evidences above suggest that he would report the truth with a nonzero probability; if he does not remember the true value, he provides an estimate of the true value, which may be considered as unintentionally misreporting. Admittedly, we can't distinguish intentionally misreporting from unintentionally misreporting without further information.

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 nonzero, i.e.*

$$\lambda(x^*) \equiv \Pr\left(X = X^* | X^* = x^*\right) > 0 \;\; for \;\; any \;\; x^*.$$

And therefore, 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.

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

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

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

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

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

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

and the operator $L_g : L^p\left(\mathbb{R}\right) \to L^p\left(\mathbb{R}\right), 1 \leq p \leq \infty$ is defined as

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

Since $0 < \lambda(z)$, 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.*

The sufficient condition for compactness is different in different L^p space. In the commonly used 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.).[4] Hence if we assume

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

then the operator L_g is compact on $L^2(\mathbb{R})$.

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 (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.[5]

4 A further discussion on the classical error case

By further analyzing the relationship between the well-posedness of Eq.(2) and the convergence rate of $\hat{f}_{X^*}$, we illustrate in this section that if some prior information of f_{X^*} is

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

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

available, we usually can narrow the set on which the problem is defined such that the problem is well-posed on the narrowed subset. 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

$$f_{X|X^*}(x|x^*) = f_\epsilon(x - x^*). \tag{8}$$

For the simplicity, we restrict the space on which the problem is defined to all the bounded functions with bounded Fourier transform in L^∞. A result we will repeatedly use in this section is that a linear operator is continuous if and only if it is bounded.[6]

We first analyze the implication of the simplification in Eq. (8) without condition 2. This convolution case has been studied thoroughly so that we only associate the existing results with the ill-posed problem. We will then combine Eq. (8) 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 observed data $\{X_i\}_{i=1,\dots,n}$. A popular estimator for f_X is as follows:

$$\begin{aligned} \hat{f}_X &= \frac{1}{2\pi}\int e^{-itx}\hat{\phi}_X(t)\mathrm{d}t \\ \hat{\phi}_X(t) &= \hat{\phi}_n(t)\phi_K\left(\frac{t}{T_n}\right), \end{aligned} \tag{9}$$

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

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 (2008) suggests that we need

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

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 Ill-posedness with a supersmooth error distribution

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

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

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 as 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* $\tau > 1$.

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

$$\begin{aligned} L_{X|X^*} &: \Upsilon \to \Psi \\ \left(L_{X|X^*}h\right)(x) &= \int f_\epsilon\left(x - x^*\right) h\left(x^*\right) dx^* \end{aligned} \tag{11}$$

where

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

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

and ϕ_f stands for the Fourier transform of function f. Note that we use $O_p\left(|t|^{-\tau}\right)$ instead of $O\left(|t|^{-\tau}\right)$ in order to include empirical densities and empirical ch.f.'s.[7] Given these specifications, we have the following results

Proposition 1. *Suppose conditions 1, 4, and Eq. (8) hold. The operator* $L_{X|X^*} : \Upsilon \to \Psi$ *in (11) 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 (9) is not in* Ψ, *i.e.,* $\hat{f}_X \notin \Psi$.

Proof See Appendix. ■

[7]In this section we need to frequently discuss whether an estimator $\hat{f}_X$ is in L^∞. Notice that $\hat{f}_X$ is a random varaible, there is a difference between $\|\hat{f}_X\|_\infty = O_p(1)$ and $\|\hat{f}_X\|_\infty = O(1)$, i.e., $\hat{f}_X \in L^\infty$. However, such a difference is negligible when the true density f_X is bounded, i.e., $|f_X| < M$. This is because one can always define an alternative random variable $\hat{f}_X = \min\{\tilde{f}_X, M\}$, then $\hat{f}_X$ and $\tilde{f}_X$ have the same asymptotic properties and we have $\|\hat{f}_X\|_\infty = O(1)$, i.e., $\hat{f}_X \in L^\infty$. We thank Tiemen Wouterson for this suggestion.

The result that the usual deconvolution density estimator $\hat{f}_X$ is not in Ψ implies 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 where the operator behaves well. On the other hand, it is also important to realize the 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 Conditional well-posedness with an ordinary smooth error distribution

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

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

where

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

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

Proposition 2. *Suppose conditions 1, 4, and Eq. (8) 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. Moreover, the density estimator* $\hat{f}_X$ *in (9) may be in* Ψ*, i.e.,* $\hat{f}_X \in \Psi$*.*

Proof See Appendix. ∎

This theorem implies that the problem (2) may be conditionally well-posed and the deconvolution estimator $\hat{f}_{X^*}$ 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 an 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 about 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. 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

$$\begin{aligned} f_{X|X^*}(x|x^*) &= f_\epsilon(x-x^*) \\ &= \lambda_n \times \delta(x-x^*) + (1-\lambda_n) \times g_{\tilde{\epsilon}}(x-x^*). \end{aligned} \tag{13}$$

Let $\phi_\epsilon(t)$ and $\phi_{\tilde{\epsilon}}(t)$ denote the characteristic functions of f_ϵ and $g_{\tilde{\epsilon}}$, respectively. Eq.(13) 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 $\phi_{\widetilde{\epsilon}}$ is supersmooth. We then have

Lemma 1. *Suppose condition 5 and Eq. (13) hold. Then* $\phi_\epsilon(t)$ *is ordinary smooth of order* d*, and therefore, the results in Proposition 2 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 $\phi_{\widetilde{\epsilon}}$ 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 $\phi_{\widetilde{\epsilon}}$ 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 1 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 2. 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

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

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 3. *i) Suppose conditions 1, 2, and Eq. (8) hold and the error distribution $g_{\widetilde{\epsilon}}$ satisfies*

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

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 $g_{\widetilde{\epsilon}}$ satisfies

$$\int_{\mathbb{R}} \int_{\mathbb{R}} \left| g_{\widetilde{\epsilon}}(x - x^*) \right|^2 dx\, dx^* < \infty. \tag{14}$$

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

Proof See appendix. ■

Proposition 3(i) shows that the compactness in condition 3 may not be necessary for well-posedness. If the problem is defined on L^2, the the compactness of L_g is satisfied given the error distribution is square integrable. For a general L^p space on which the problem is defined, the compactness of L_g need to be assumed directly as in condition 3.

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 1 and 2. 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.

5 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 desirable properties of estimators of the latent model. 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 supersmooth 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.

Carrasco, M., and J.-P. Florens (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.

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, Mar. 2008.*

Chen, X., and M. Reiss (2007): "On rate optimality for ill-posed inverse problems in econometrics," *Econometric Theory, forthcoming.*

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.

Hu, Y., and G. Ridder (2008): "On Deconvolution as a First Stage Nonparametric Estimator," *Econometric Reviews, forthcoming.*

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

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.

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 $f_X \in L^p$ implies $D_\lambda^{-1} f_X \in L^p$. 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

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

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

Next, we prove the operator $D_\lambda^{-1} L_g \left(I - D_\lambda\right)$ is compact on L^p 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 $X = Y = Z = L^p$, $A = I - D_\lambda$, 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 $\left|\left|(I - D_\lambda) f_{X^*}\right|\right|_p \le M_{1-\lambda}\left|\left|f_{X^*}\right|\right|_p$. 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 $D_\lambda^{-1} L_g \left(I - D_\lambda\right)$ is compact.

To complete the proof, it remains to show that $I + D_\lambda^{-1} L_g \left(I - D_\lambda\right)$ is injective. By condition 1, $L_{X|X^*} = D_\lambda \left(I + D_\lambda^{-1} L_g \left(I - D_\lambda\right)\right)$ is injective. Therefore, for any two distinct functions $f_1, f_2 \in L^p$, 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 drive 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 1. 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 $\left|\frac{\phi_X(t)}{\phi_\epsilon(t)}\right| = O_p\left(|t|^{-\tau}\right)$ with $\tau>1$ for $f_X\in\Psi$.

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

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

where t_0 and M are some positive constants and $\tau>1$. The second inequality holds because $\left|\frac{\phi_X(t)}{\phi_\epsilon(t)}\right| = O_p\left(|t|^{-\tau}\right)$ 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 $L_{X|X^*}^{-1}f_X\in\Upsilon$. Because for any $f_X\in\Psi$, both $||\ L_{X|X^*}^{-1}f_X\ ||_\infty$ and $||\ f_X\ ||_\infty$ are finite, there must exist a constant $N>0$ such that $||\ L_{X|X^*}^{-1}f_X\ ||_\infty < N\ ||\ f_X\ ||_\infty$, i.e., $L_{X|X^*}^{-1}:\Psi\to\Upsilon$ 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. (9) 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:

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

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

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

or equivalently,

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

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

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

Let $T_n = O(\frac{n}{\log n})^{\gamma}, \gamma \in (0, \frac{1}{2})$. According to Lemma 1 in Hu and Ridder (2008), 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 $|\hat{\phi}_X(t) - \phi_X(t)|$ equal 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 $\phi_X(T_n) = O_p\left((\frac{n}{\log n})^{-\tau\gamma}\exp\left(-(n/\log n)^{\beta}/\rho\right)\right)$ as $n \to \infty$. Consequently,

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

Notice that given $\beta, \rho > 0$ the term $(\frac{\log n}{n})^{\frac{1}{2}-(1+\tau)\gamma-\eta}\exp\left((n/\log n)^{\beta}/\rho\right)$ diverges for any

τ, γ, and η. Therefore, the density estimator $\hat{f}_X$ in Eq. (9) 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 2. The proof of the bijectivity of $L_{X|X^*}$ is similar to the proof in Proposition 1, 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 1, it is sufficient to show that

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

holds for $|t| \leq T_n$.

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

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

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. (9) may be in Ψ. ■

Proof of Lemma 1. Eq.(13) 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:

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

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

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

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 2. ■

Proof of Proposition 3. According to the proof of Proposition 1, 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, $\phi_\epsilon(t) = \lambda + (1-\lambda)\,\phi_{\widetilde{\epsilon}}(t)$. Therefore we have

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

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

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

The condition $\int |\phi_{\widetilde{\epsilon}}(t)| \mathrm{d}t < \infty$ implies that $L_{X|X^*}^{-1} f_X \in L_{bc}^\infty$ if $f_X \in L^\infty$, i.e., $L_{X|X^*}^{-1} : L_{bc}^\infty \to L_{bc}^\infty$ is surjective, hence bijective since the injectivity holds by condition 1. Following the argument in proof of Proposition 1, we can also conclude that $L_{X|X^*}^{-1}$ is continuous. This completes the proof of the first part.

In the second part of the proposition, Eq.(14) 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. ■