# Identification of Nonparametric Monotonic Regression Models with Continuous Nonclassical Measurement Errors*

Yingyao Hu†, Susanne Schennach‡, and Ji-Liang Shiu§

September 9, 2020

**Abstract**

This paper provides sufficient conditions for identification of a nonparametric regression model with an unobserved continuous regressor subject to nonclassical measurement error. The measurement error may be directly correlated with the latent regressor in the model. Our identification strategy does not require the availability of additional data information, such as a secondary measurement, an instrumental variable, or an auxiliary sample. Our main assumptions for nonparametric identification include monotonicity of the regression function, independence of the regression error, and completeness of the measurement error distribution. We also propose a sieve maximum likelihood estimator and investigate its finite sample property through Monte Carlo simulations.

**Keywords**: Nonclassical measurement error, Nonparametric identification, Spectral decomposition, Nonparametric regression, Sieve maximum likelihood estimation.

---

*Susanne Schennach acknowledges support from the US National Science Foundation via Grant SES-1357401. Ji-Liang Shiu acknowledges support from the National Science Council of Taiwan via Grant 99-2410-H-194-032. The authors are solely responsible for any remaining errors.

†Corresponding author. Department of Economics, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218. Email: yhu@jhu.edu.

‡Department of Economics, Brown University. Email: smschenn@brown.edu.

§Institute for Economic and Social Research, Jinan University. Email: jishiu.econ@gmail.com.

# 1. Introduction

This paper considers the following nonparametric errors-in-variables regression model:

$$Y = m_0(X^*) + \eta \tag{1}$$

$$X = X^* + \varepsilon \tag{2}$$

where $m_0$ is an unknown, monotone function, $Y$ is a dependent variable, $\eta$ is the regression error, $X^*$ is an unobserved continuous regressor, $X$ is the observed counterpart of $X^*$, contaminated by a measurement error $\varepsilon$. The main goal is to identify the nonlinear function $m_0$ from the joint distribution of the observed data $(Y, X)$ without a priori knowledge of the distribution of the measurement error $\varepsilon$, while allowing a correlation between $X^*$ and $\varepsilon$, and without relying on the availability of additional side information, such as repeated measurements or instruments.

The nonparametric version of the identification problem in Eqs. (1) and (2) without the monotonicity has only very recently been solved in the case where $\varepsilon$, $\eta$ and $X^*$ are mutually independent (Schennach and Hu (2013)). This paper seeks to relax the independence assumption between $\varepsilon$ and $X^*$ to allow for so-called non-classical measurement error, a topic whose importance is beginning to gather significant attention due to realization that the classical (independent) error may often be violated in applications (Bound, Brown, and Mathiowetz (2001), Hu and Schennach (2008), Bollinger (1998), Bound, Brown, Duncan, and Rodgers (1994)).

The present paper provides a significant step towards generally handling nonclassical error, by allowing flexible correlation between the latent regressor $X^*$ and the nonclassical measurement error $\varepsilon$. We impose the same type of restriction on the nonclassical measurement error as in Hu and Schennach (2008) such as completeness and a location condition of the measurement error distribution but avoid the use of an additional instrumental variable to achieve nonparametric identification. We also avoid the reliance on the information contained in other, correctly measured, regressors (as in, e.g., Ben-Moshe, D'Haultfoeuille, and Lewbel (2016)). Our model is superficially reminiscent of that in Chen, Hu, and Lewbel (2009), where the unobserved regressor $X^*$ and its measurement $X$ share a finite discrete support. Whereas handling the discrete misclassification case could be reduced to solving a finite system of equations, handling the continuous case entails considerable technical challenges, such as requiring the use of advanced operator and Fourier techniques. In addition, we are able to provide primitive conditions for our identification result that are far easier to interpret than those of Chen, Hu, and Lewbel (2009).[1]

We assume non-differential measurement error and that the regression error $\eta$ is independent of the latent regressor $X^*$ and its measurement $X$, that the regression function $m_0$ is monotonic over the support $\mathcal{X}^*$ of $X^*$, and that the measurement error density $f_{X|X^*}$ is *complete* (This can be regarded as a nonparametric rank condition. see, e.g., Mattner (1993) or Andrews (2017)). We show that the regression function $m_0$ and the measurement error distribution $f_{X|X^*}$ are nonparametrically identified by showing that the densities $\left(f_{Y|X^*}, f_{X|X^*}, f_{X^*}\right)$ on the right hand side are uniquely determined from the observed joint density $f_{Y,X}$ on the left hand side of the following integral equation:

$$f_{Y,X}(y,x) = \int_{\mathcal{X}^*} f_{Y|X^*}(y|x^*) f_{X|X^*}(x|x^*) f_{X^*}(x^*) dx^*, \tag{3}$$

obtained after assuming $f_{Y|X^*,X}(y|x^*,x) = f_{Y|X^*}(y|x^*)$.

Based on this result, we propose a sieve maximum likelihood estimator (MLE) for the (possibly infinite-dimensional) parameter of interest $\alpha$ that incorporates the regression function $m_0$ and other nonparametric elements $f_\eta$, $f_{X|X^*}$, and $f_{X^*}$. Sieve estimators represent a powerful and rapidly growing class of estimators (see Ding and Nan (2011), Xue, Miao, and Wu (2010), Chen, Wu, and Yi (2009), Ghosal and Van Der Vaart (2007) for recent examples). Under suitable regularity conditions, one can approximate the unknown functions $m_0$, $f_\eta$, $f_{X|X^*}$, and $f_{X^*}$ by truncated sieve series such as polynomials, Fourier series, or splines and estimate the coefficients of these approximations by maximum likelihood (Chen (2006)). Using techniques from Ai and Chen (2003), we show that the sieve MLE is consistent and, in semiparametric settings, root-$n$ asymptotically normal and efficient, under suitable regularity conditions. We investigate finite sample properties of the proposed sieve maximum likelihood estimator through Monte Carlo simulations.

Measurement error models have been gathering considerable interest in statistics (Chesher (1991), Li and Vuong (1998), Wang (2004), Huang and Wang (2001), Schennach (2013), Carroll, Delaigle, and Hall (2007), Delaigle, Hall, and Meister (2008), Schennach (2004), among many others) and this topic has been the subject of several reviews (e.g., Carroll, Ruppert, Stefanski, and Crainiceanu (2006), Schennach (2016)). The more challenging problem

[1] Our assumptions are all imposed directly on the primitive objects such as $m_0$, $f_{X|X^*}$, etc. The identification condition in Assumption 2.3 of Chen, Hu, and Lewbel (2009) is rather technical and its connections to $m_0$, $f_{X|X^*}$, etc are not straightforward.

of addressing measurement error when side information is unavailable has also been receiving considerable attention, but existing methods have so far focussed on linear models (Geary (1942), Reiersol (1950), Chesher (1998), Pal (1980), Cragg (1997), Lewbel (1997), Lewbel (2012), Dagenais and Dagenais (1997), Erickson and Whited (2000), Erickson and Whited (2002), Erickson, Jiang, and Whited (2014), Ben-Moshe (2014), among others). Examples that combine a classical error on $Y$ and a nonclassical error on $X$ can be found in many fields, from the medical to the economic literatures. For instance, in the study the effect of a specific food intake ($X^*$) on cholesterol levels ($Y$) (e.g., Griffin and Lichtenstein (2013)), it is plausible to maintain a classical error assumption on the laboratory measurement of cholesterol level (Glasziou, Irwig, Heritier, Simes, and Tonkin (2008)) but food intake is unlikely to be contaminated by a simple classical error if it is self-reported, a situation known to induce non-classical errors (Hyslop and Imbens (2001), Bound, Brown, and Mathiowetz (2001)). Another example would be the study of the relation between household income ($X^*$) and children health status ($Y$), as measured by objective quantities (such a body mass index) (e.g., Jin and Jones-Smith (2015)). While the dependent variable is likely to conform to classical assumptions, household income is widely recognized as exhibiting nonclassical error (Bollinger (1998), Bound and Krueger (1991)).

The rest of the paper is organized as follows. Section 2 discusses assumptions for nonparametric identification. Section 3 describes our estimator, and Section 4 presents Monte Carlo simulations. Section 5 concludes. All proofs and auxiliary lemmas are in the Appendix.

## 2. Nonparametric Identification

In this section, we introduce our key assumptions for nonparametric identification of the model and outline the main arguments of the proof in order to give an intuition of the identification result. Let $\mathcal{Y}$, $\mathcal{X}$, and $\mathcal{X}^*$ denote the supports of the distributions of the random variables $Y$, $X$, and $X^*$, respectively. We first assume a boundedness restriction on densities and place some restrictions on the regression error $\eta$.

**Assumption 2.1.** *(Restrictions on densities) The joint distribution of the random variable $X$ and $X^*$ admits a density $f_{X,X^*}$ with respect to the Lebesgue measure and the conditional density of the measurement error $f_{X|X^*}$ and marginal density of the true regressor $f_{X^*}$ are bounded by a constant.*

**Assumption 2.2.** *(Restrictions on regression error) We assume that* *(i) (Independence) the regressor error $\eta$ is independent of the latent true regressor $X^*$,*

*(ii) (Zero conditional mean) $E[\eta|X^*] = 0$,*

*(iii) (Nonvanishing characteristic function) $E[\exp(i\gamma\eta)] \neq 0$ for all $\gamma \in \mathbb{R}$.*

Assumption 2.2(i) effectively imposes an additively separable structure on the regression error $\eta$. This assumption implies that the conditional density $f_{Y|X^*}$ is completely determined by the distribution of the regressor error $\eta$ and the regression function as follows:

$$f_{Y|X^*}(y|x^*) = f_\eta(y - m_0(X^*)).$$

Assumption 2.2(ii) is a standard centering restriction on the model's disturbances.

Let $\mathcal{L}^2(\mathfrak{X}) = \{h : \int_{\mathfrak{X}} |h(x)|^2 dx < \infty\}$. The measurement error satisfies the following:

**Assumption 2.3.** *(Restrictions on Measurement Error) Suppose that*

*(i) (Nondifferential error) the observed measurement $X$ is independent of dependent variable $Y$ conditional on the unobserved regressor $X^*$, i.e., for $\forall(y, x, x^*) \in \mathcal{Y} \times \mathcal{X} \times \mathcal{X}^*$*

$$f_{Y|X^*,X}(y|x^*, x) = f_{Y|X^*}(y|x^*).$$

*(ii) (Invertibility) For any function $h \in \mathcal{L}^2(\mathcal{X}^*)$, $\int f_{X|X^*}(x|x^*)h(x^*)dx^* = 0$ for all $x \in \mathcal{X}$ implies $h(x^*) = 0$ for almost any $x^* \in \mathcal{X}^*$. On the other hand, for any function $h \in \mathcal{L}^2(\mathcal{X})$, $\int f_{X|X^*}(x|x^*)h(x)dx = 0$ for all $x^* \in \mathcal{X}^*$ implies $h(x) = 0$ for almost any $x \in \mathcal{X}$.*

*(iii) (Normalization) There exists a known functional $G$ such that $G\left[f_{X|X^*}(\cdot|x^*)\right] = x^*$ for any $x^* \in \mathcal{X}^*$.*

Assumption 2.3(i) implies that the measurement error is *nondifferential*, that is, $X - X^*$ does not affect the true model, $f_{Y|X^*}$, the distribution of the dependent variable $Y$ conditional on the true value $X^*$. The observed measurement $X$ thus does not provide any more information about $Y$ than the unobserved regressor $X^*$ already does. Such conditional independence restrictions have been extensively used in the recent years.[2] Note that we allow the measurement error $X - X^*$ to be correlated with the true unobserved regressor $X^*$, which reflects the presence of potential nonclassical measurement error.

Assumption 2.3(ii) implies that the conditional density $f_{X|X^*}$ is complete in both $\mathcal{X}$ and $\mathcal{X}^*$. This condition is related to the invertibility of the integral operator with kernel $f_{X|X^*}$.

[2] For example, Altonji and Matzkin (2005), Heckman and Vytlacil (2005), and Hoderlein and Mammen (2007).

Intuitively, assuming completeness of $f_{X|X^*}$ is weaker than assuming independence between $X^*$ and $X - X*$, in the same way the space of invertible matrices is much larger (in terms of dimension) than the space of similarly sized matrices $A$ of the special form $A_{ij} = v_{(j-i)}$ for some vector $v$.[3] Completeness conditions have recently been employed in the nonparametric IV regression models and nonlinear measurement error models and such conditions are often regarded as high level conditions. Canay, Santos, and Shaikh (2013) have shown that the completeness condition is not testable in a nonparametric setting with continuous variables. However, Freyberger (2017) provides a first test for the restricted completeness in a nonparametric instrumental variable model by linking the outcome of the test to consistency of an estimator. Hu, Schennach, and Shiu (2017) rely on known results regarding the Volterra equation to provide sufficient conditions for completeness conditions for densities with compact support with an accessible interpretation and without specific functional form restrictions.[4]

Assumption 2.3(iii) is borrowed from Hu and Schennach (2008), because we also use a spectral decomposition, but with less data information and more restrictions on the regression model. Examples of functional $G$ from Assumption 2.3(iii) include the mean, the mode, median, or the $\tau$-th quantile. It implies that a location of the distribution $f_{X|X^*}(\cdot|x^*)$ reveals the true value $x^*$. This condition also imposes restrictions on the support of $x$, $x^*$, and therefore, the measurement error. Those include that zero is in the support of the measurement error and that the cardinality of the support of $x$ can't be smaller than that of $x^*$. We refer to that paper for further discussion on these conditions.

Finally, we assume the regression function satisfies

**Assumption 2.4.** *(Restrictions on regression function) Suppose that the regression function $m_0$ is continuous, bounded, and strictly monotonic over support $\mathcal{X}^*$.*

The boundedness constraint can be somewhat restrictive and rules out linear functions when the support $\mathcal{X}^*$ is unbounded. However, if the support of $x^*$ is a bounded interval, Assumption 2.4 is a rather mild condition and allows for linear functions.

Our main results is as follows:

**Theorem 2.1.** *Under Assumptions 2.1, 2.3, 2.2, and 2.4, given the observed density $f_{Y,X}(y,x)$,*

[3] This analogy exploits the fact that, in the case of discrete measurement error, the link between the observed distribution of $X$ and the unobserved distribution of $X^*$ can be represented by the multiplication of the vector of unobserved probabilities of the different values of $X^*$ by the mislassification matrix $A$.

[4] More general discussions of completeness can be found in D'Haultfoeuille (2011), Chen, Chernozhukov, Lee, and Newey (2013), Andrews (2017), and Hu and Shiu (2017), Mattner (1993), Newey and Powell (2003) and Blundell, Chen, and Kristensen (2007).

*the equation*

$$f_{Y,X}(y,x) = \int_{\mathcal{X}^*} f_\eta(y - m_0(X^*)) f_{X|X^*}(x|x^*) f_{X^*}(x^*) dx^*$$

*permits a unique solution* $(m_0, f_\eta, f_{X|X^*}, f_{X^*}) \equiv \alpha_0$. *The solution characterizes the nonparametric regression model in Eq. (1).*

The formal proof of this result, reported in the appendix, can be outlined as follows. If one knew the distribution of the model error $\eta$, one could recover the joint distribution of $(m_0(X^*), X)$ by a standard deconvolution argument, thanks to Assumptions 2.2 and 2.3(i). From that distribution, one could then recover $m_0$ and $f_{X|X^*}$ from our assumed normalization restriction (Assumption 2.3(iii)), after exploiting the monotonicity and continuity of $m_0$ (Assumption 2.4).[5] Of course, one does not know, a priori, the distribution of $\eta$, but one can, in principle, consider any possible trial distribution to get various possible trial values of $m_0$ and $f_{X|X^*}$. The key realization is that, whenever the assumed density of $\eta$ is incorrect, this will be detectable by one of the following occurrences: (i) negative densities for the unobserved variables, (ii) violation of Assumption 2.3(ii) (invertibility) or (iii) violation of the boundedness constraint of Assumption 2.4.

The Appendix provides another, completely independent, proof of Theorem 2.1, which delivers a rather different insight into the identification problem. This alternate proof employs operator techniques similar to those used in Hu and Schennach (2008) and can be summarized as follows. The idea is that the integral Equation 2.1 can be cast as a system of operator equivalence relations. Solving this system yields an equivalence between an operator entirely built from observable quantities and a product of unknown operators to be determined. We then show that this factorization can be uniquely determined, because it takes the form of an operator diagonalization identity, i.e., the eigenvalues and eigenfunctions of a known operator yield the different pieces of the product. To ensure uniqueness of this decomposition, we appeal to conditions such as the invertibility and normalization on $f_{X|X^*}$ in Assumptions 2.3(ii)&(iii) and the monotonic restriction on $m_0$ in Assumptions 2.4.

Although the monotonicity is a strong restriction, the condition is applicable to many empirical settings. We provide three examples in different areas of economics where monotonicity is a reasonable assumption. The first example is the estimation of the impact of education $(X^*)$ on wages $(Y)$ in which there could be reporting errors in education level. The higher

[5] In the absence of monotonicity, the measurement error distributions along the $X$ axis for different true values of $X^*$ would mix. As a result, one could not easily identify the measurement error distribution by looking at the distribution of $X$ conditional on the value of $m_0(X^*)$.

education level the higher wage, which implies a monotonic regression function between the wage offer and the true education level. The second empirical example is in estimating the effect of government subsidies ($X^*$) on firm R&D investment ($Y$). The measures of government subsidies may suffer measurement errors because they may be hard to summarize when each firm may receive different types of subsidies. The fact that more government subsidies for firms are likely to increase R&D investments indicates a monotonic relation between them. The third example is the relation between household income ($X^*$, measured with error) and children health status ($Y$). Since wealthier families have more resource to promote children health, higher household income tends to be associated with better children health status. In all these three examples, we can use the mode as the functional $G$ in Assumption 2.3(iii) because people are more likely to tell the truth for their education level, and household income, and firms are more likely to report the true government subsidies.

The point identification result of Theorem 2.1 is not only nonparametric, but also global. This is because we show identification by solving the integral equation directly, in the sense that our identification strategy does not rely on the usual local identification condition that a true parameter value is only distinguishable from those parameters values close to the true one.

Our result is applicable beyond regression settings. In general, we may also consider the observables $(Y, X)$ as two measurements or proxies of the latent variable $X^*$, an observation which is useful, for instance, in factor models. In many empirical applications, the latent variable may represent unobserved heterogeneity or an individual effect. Our result may then allow for flexible relationships between observables and unobservables to achieve nonparametric identification. In addition, our results can also be straightforwardly extended to the case where an additional error-free covariate vector $W$ appears in the regression function, because our assumptions and results can all be restated as conditioned on $W$.

Our results prompt the question of whether it would be possible to further extend the identification proof to cover the case where both the dependent variable and the regressor are contaminated by a nonclassical error. However, this would necessitate a one-to-one mapping between the space of bivariate density $f_{YX}(y, x)$ and the much "larger" space of pairs of bivariate functions $(f_{X,X^*}(x, x^*), f_{Y|X^*}(y|x^*))$, which is a highly unlikely possibility.

## 3. A Sieve Maximum Likelihood Estimator

The nonparametric identification result of Theorem 2.1 allows for a variety of possible parametrization of the model in terms of the parameter $\alpha_0$ that incorporates all the unknown functions of the model. In this section, we assume that the regression function contains a vector of finite dimensional unknown parameters $\theta$ of primary interest and possibly an infinite dimensional unknown function $h$, namely, $m_0(x^*) = m_0(x^*; \widetilde{\theta}_0)$ where $\widetilde{\theta}_0 = (\theta_0, h_0)$.[6] For instance, $\theta_0$ could be a finite-dimensional parameter that represents some average derivative (Härdle and Stoker (1989)) of $m_0(x^*)$, while $h_0$ would be an infinite-dimensional parameter vector allowing the shape of $m_0(x^*)$ to be free of parametric restrictions. Alternatively, the regression function $m_0$ could be parametrically specified as $m_0(x^*; \theta_0)$ for a finite-dimensional parameter $\theta_0$.[7] Hence, practitioners are free to be as parametric or as nonparametric as they wish given the data available.

This approach's underlying motivation is that practitioners often wish to test a specific hypothesis or report a single summary measure $\theta$ of the causal effect of some variable on another variable even when they are unwilling to make parametric restrictions. In this context, the smoothing effect of semiparametric functionals, enables, under suitable regularity conditions, convergence at the parametric rate and a limiting distribution that is normal centered at zero, thus enabling testing in the most natural way and circumventing[8] slow convergence due to the ill-posedness of the problem of inverting integral equations (Schennach (2004)).

The identification of $(\theta_0, f_\eta, f_{X|X^*}, f_{X^*}, h_0)$ in Theorem 2.1 makes use of completeness conditions on $f_{X|X^*}$ and monotonicity of $m_0$ and we show how to accomplish the nontrivial task of integrating these identification conditions into a practical estimation method. This section illustrates the sieve MLE of the nonparametric regression model (1) with the identification restrictions. This constructive estimation method is novel to the literature.

Set $\mathcal{A} = \Theta \times \mathcal{F}_1 \times \mathcal{F}_2 \times \mathcal{F}_3 \times \mathcal{F}_4$, as the parametric space containing the true parameter $\alpha_0 \equiv (\theta_0, \sqrt{f_\eta}, f_{X|X^*}, \sqrt{f_{X^*}}, \sqrt{h_0})$,[9] and $\alpha \equiv (\theta, f_{1s}, f_2, f_{3s}, h_{4s})$, where the lower subscript $s$ indicates the square roots. Note that employing an expansion based on the square root of the densities provides a natural way to ensure that the densities themselves are positive. The true parameter $\alpha_0 \equiv (\theta_0, \sqrt{f_\eta}, f_{X|X^*}, \sqrt{f_{X^*}}, \sqrt{h_0})$ is the solution of the following maximization

[6] While $m$ is identified from Theorem 2.1, the joint identification of $\theta_0$,$h_0$ from $m$ is assumed.

[7] More examples of a partition can be found in Shen (1997).

[8] Of course, this statement is conditional on a number of regularity conditions that may not always hold, see Chen and Liao (2014) and Chen and Pouzo (2015) for examples.

[9] The detail descriptions of $\mathcal{A}$ are provided in Appendix B.

problem:

$$\sup_{\alpha\in\mathcal{A}} E\left[\ln\left(\int_{\mathcal{X}^*} f_1(y-m_0(x^*;\widetilde{\theta}))f_2(x|x^*)f_3(x^*)dx^*\right)\right], \tag{4}$$

where $\widetilde{\theta}=(\theta,h_4)$. An estimator can then be obtained by maximizing the sample analog of Eq. (4) based on the observed sample $\{y_i,x_i\}_{i=1}^n$. Define

$$\widehat{Q}_n(\alpha)=\frac{1}{n}\sum\nolimits_{i=1}^n \ln\left(\int_{\mathcal{X}^*} f_1(y_i-m_0(x^*;\widetilde{\theta}))f_2(x_i|x^*)f_3(x^*)dx^*\right). \tag{5}$$

To obtain a consistent estimator, it is necessary to regularize the optimization procedure by maximizing $\widehat{Q}_n(\alpha)$ over $\mathcal{A}^n\equiv\Theta\times\mathcal{F}_1^n\times\mathcal{F}_2^n\times\mathcal{F}_3^n\times\mathcal{F}_4^n$, a sequence of approximation spaces to $\mathcal{A}$.

### 3.1. Identification Restrictions on Sieve Spaces

In the sieve approximation, we consider each component of the finite-dimensional sieve $\mathcal{A}^n$ spanned by othonormal bases. Let $\{p_i(x):i=1,2,3,...\}$ and $\{p_j(x^*):j=1,2,3,...\}$ be orthonormal base for $L^2(\mathcal{X})$ and $L^2(\mathcal{X}^*)$, respectively. A bivariate basis function for $L^2(\mathcal{X}\times\mathcal{X}^*)$ can be generated by a tensor product construction using $\{p_i(x):i=1,2,3,...\}$ and $\{p_j(x^*):j=1,2,3,...\}$. We thus have a sieve expressions for $f_2$ of the form

$$f_2(x|x^*)=\sum\nolimits_{i=1}^{k_{2,n}}\sum\nolimits_{j=1}^{k_{2,n}}\beta_{2ij}p_i(x)p_j(x^*). \tag{6}$$

With this expansion, the first completeness restriction in Assumption 2.3(ii), projected onto the finite dimensional space of functions spanned by the $p_j$, is equivalent to imposing that the square coefficient matrix $[\beta_{2ij}]_{k_{2,n}\times k_{2,n}}$ is invertible. This follows from the orthogonality of the $p_j$, as shown in more detail in Lemma D.1 of the online appendix. We can similarly find the restriction for the second completeness restrictions and the restriction also requires $[\beta_{2ij}]_{k_{2,n}\times k_{2,n}}$ to be invertible. Incorporating the two other restrictions on $f_2$, namely, the density restriction and normalization in Assumption 2.3(iii), yields the following sieve space:

$$\begin{aligned}\mathcal{F}_2^n =\{ & f_2(\cdot|\cdot) \in \Lambda_c^{\gamma_1,\omega}(\mathcal{X} \times \mathcal{X}^*) : \exists \beta_2 \in \mathbb{R}^{k_{2,n} \times k_{2,n}} \text{ such that} \\ & (i)\ f_2(x|x^*) = \sum_{i,j=1}^{k_{2,n}} \beta_{2ij} p_i(x) p_j(x^*), [\beta_{2ij}]_{k_{2,n}\times k_{2,n}} \text{ is invertible}, \\ & (ii)\ f_2(\cdot|\cdot) \geq 0, \int_{\mathcal{X}} f_2(x|x^*)dx = 1 \text{ for } x^* \in \mathcal{X}^*, \text{ and} \\ & (iii)\ f_2 \text{ satisfies Assumption 2.3}(iii)\}.\end{aligned}$$

Since implementation of some of the restrictions on sieve coefficients are dependent on the specific basis used, we will discuss all restrictions together for a particular orthonormal basis in the Monte Carlo section.

Next, we consider the strict monotonicity condition on $m_0(x^*;\widetilde{\theta})$. Without loss of generality, we assume $m_0(x^*;\widetilde{\theta})$ is strictly increasing. To impose the strictly increasing property, we need to know the semi-parametric structure of the regression function. We divide the semi-parametric structure into three cases, pure parametric cases, pure nonparametric cases, and semi-parametric cases. In the pure parametric case, we can consider $m'(x^*;\theta) > 0$ for all $x^* \in \mathcal{X}^*$. As for the pure nonparametric case $m_0(x^*;\widetilde{\theta}) = m_0(x^*)$, we can consider a sieve expression of a square root of $m'$ and use the following sieve expression

$$m'(x^*) = \beta_{40} + \left(\sum\nolimits_{k=1}^{k_{4,n}-2} \beta_{4k} p_k(x^*)\right)^2 \text{ for some } \beta_{40} > 0. \tag{7}$$

Then, we can use an anti-derivative of the sieve expression in Eq. (7) as a sieve approximation for the regression function,

$$m_0(x^*) = \mu_0 + \beta_{40} x^* + \int_a^{x^*} \left(\sum\nolimits_{k=1}^{k_{4,n}-2} \beta_{4k} p_k(x^*)\right)^2 dt. \tag{8}$$

If $m$ has both parametric ($\theta$) and nonparametric ($h_4$) components, the sieve restriction from the monotonicity condition may depend on the functional form of $m_0(x^*;\widetilde{\theta})$. For example, if $m_0(x^*;\widetilde{\theta}) = H(\theta + h_4(x^*))$, where $H$ is a known function, we may only implement the restriction as $H' > 0$, and $h_4' > 0$, and we can obtain the sieve restriction for $h_4' > 0$ in a similar way as the pure nonparametric case. Therefore, we use the following sieve space for the strictly increasing function $m$

$$\mathcal{F}_4^n = \{\sqrt{m_4'(\cdot;\widetilde{\theta})} \in \mathcal{F}_4 : \exists\,(\beta_{40},\beta_4) \in \mathbb{R}^{1+k_{4,n}} \text{ such that} \tag{9}$$

$$\sqrt{m_4'(\cdot;\widetilde{\theta})} = \beta_{40} + p^{k_{4,n}}(x^*)^T\beta_4\beta_4^T p^{k_{4,n}}(x^*), \beta_{40} > 0\} \tag{10}$$

Because $f_1$, and $f_3$ only have density restrictions, the sieve restrictions for them are easy to impose and their sieve spaces are

$$\mathcal{F}_1^n = \{\sqrt{f_1(\cdot)} \in \mathcal{F}_1 : \exists \beta_1 \in \mathbb{R}^{k_{1,n}} \text{ such that } \sqrt{f_1(\eta)} = p^{k_{1,n}}(\eta)^T\beta_1\}$$
$$\mathcal{F}_3^n = \{\sqrt{f_3(x^*)} \in \mathcal{F}_3 : \exists \beta_3 \in \mathbb{R}^{k_{3,n}} \text{ such that } \sqrt{f_3(x^*)} = p^{k_{3,n}}(x^*)^T\beta_3\}$$

where $p^k(\cdot) = (p_1(\cdot), ..., p_k(\cdot))^T$ is a vector of known univariate basis function.

A consistent sieve MLE $\widehat{\alpha}_n$ is given by

$$\widehat{\alpha}_n = \arg\max_{\alpha\in\mathcal{A}^n} \widehat{Q}_n(\alpha).$$

This estimator is a direct application of the general semi-parametric sieve MLE presented by Shen (1997), Chen and Shen (1998), and Ai and Chen (2003). Ai and Chen (2003) shows that $\widehat{\alpha}_n$ is a consistent estimator, and the parametric component of $\alpha$ has an asymptotically normal distribution. We present all the standard assumptions for consistency of all unknown parameters and root-$n$ normality of the parametric part in the Online Appendix.

# 4. Monte Carlo Study

In this section, we examine the finite sample properties of the estimator via Monte Carlo experiments in a variety of models including parametric and nonparametric regression models.

## 4.1. Parametric Regression Model

In the parametric setting, we focus on three parametric model specifications for two Data Generating Process (DGP) designs with two different types of nonclassical measurement errors. The three different specifications for the monotonic regression function $m_0(x^*;\theta)$ include a polynomial, an exponential function, and a rational fraction. In each experiment, we perform 200 Monte Carlo replications with two sample sizes: 1000, and 2000. For each sample size, we calculate the mean, the median, RMSE and $\text{AIC}_c$ for the estimator across all 200 simulations.

Formal data-driven selection rules for choosing smoothing parameters in sieve maximum likelihood are available in the literature (van der Laan, Dudoit, and Keles (2004), Schennach (2013), Chen, Wu, and Yi (2009)). Here, following Chen, Wu, and Yi (2009), we determine the optimal number of terms based on small sample correction of AIC in Burnham and Anderson (2002) $\mathrm{AIC}_c = -2\widehat{Q}_n(\widehat{\alpha}_n(K_n)) + 2K_n/(n - K_n - 1)$, where $K_n$ is the total number of sieve parameters, with the model with lowest $\mathrm{AIC}_c$ preferred. We report the estimation result using different choices of the order of the sieve coefficients for $f_2(x|x^*)$, $k_{2,n} = 3, 4$, and 5 (larger values of $k_{2,n}$, not reported for conciseness, do not yield improvements in RMSE or $\mathrm{AIC}_c$ values).

The data for the Monte Carlo experiments are generated by the model:

$$y_i = m_0(x_i^*; \theta) + \eta_i, \quad \text{for all } i = 1, ..., N, \tag{11}$$

where, $X^*$ is a standard normal random variable truncated to the interval $[-1, 1]$, and $\eta$ is generated independently by $\eta \sim N(0, 1)$. Consider

$$\text{DGP I: } x = x^* + h(x^*)e, \; e \sim N(0,1) \text{ with } h(x^*) = |x^*|,$$
$$\text{DGP II: } x = x^* + h(x^*)e, \; e \sim N(0,1) \text{ with } h(x^*) = 0.3exp(-x^*).$$

There are three specifications for the parametric regression function

$$\text{Function 1: } m_0(x^*; \theta) = \theta_1 x^* + \theta_2 x^{*2} + x^{*3},$$
$$\text{Function 2: } m_0(x^*; \theta) = \theta_1 x^* + \theta_2 e^{x^*},$$
$$\text{Function 3: } m_0(x^*; \theta) = \theta_1 x^* + \frac{\theta_2 x^*}{2 - x^*}.$$

We use the Hermite orthogonal series as our sieve basis functions for $f_1(\eta)^{1/2}$ and the Legendre polynomial series of $L^2([-1, 1]) = \{h : \int_{-1}^{1} |h(x)|^2 dx < \infty\}$ as our sieve basis functions for $f_2(x|x^*)$, and $f_3(x^*)^{1/2}$ (the observed data is bounded and is trivially scaled to fit the $[-1, 1]$ domain of these series). Denote the Hermite polynomials by $H_1(\eta) = 1, H_2(\eta) = \eta$, $H_3(\eta) = \eta^2 - 1, H_4(\eta) = \eta^3 - 3\eta, ...$ and observe that they form an orthogonal series after multiplication by a Gaussian: $\int_{-\infty}^{\infty} H_n(\eta) H_m(\eta) e^{-\eta^2} d\eta = \sqrt{2\pi} n! \delta_{nm}$, where $\delta_{nm} = 1$ if $n = m$, and $\delta_{nm} = 0$ otherwise. When $k_{1,n} = 4$, $f_1(\eta) = \left(\sum_{k=1}^{4} \beta_{1k} H_k(\eta)\right)^2$. Those sieve coefficients satisfy the following density restriction, $\sqrt{2\pi}(\beta_{11}^2 + \beta_{12}^2 + 2!\beta_{13}^2 + 3!\beta_{14}^2) = 1$.

Denote the Legendre polynomials by $g_1(\eta) = 1, g_2(\eta) = \eta$, $g_3(\eta) = \eta^2 - \frac{1}{3}, g_4(\eta) = \eta^3 -$

$\frac{3}{5}\eta, g_5(\eta) = \eta^4 - \frac{6}{7}\eta^2 + \frac{3}{35}, ...$ and observe that they form an orthogonal series: $\int_{-1}^{1} g_l(\eta) g_m(\eta) d\eta = c_l \delta_{lm}$, where $\delta_{lm} = 1$ if $l = m$, and $\delta_{lm} = 0$ otherwise. Then, we can normalize the orthogonal series to obtain an orthonormal base $\{\phi_i(\eta) : i = 1, 2, 3, ...\}$, where $\phi_i(\eta) = \frac{g_i(\eta)}{\sqrt{c_i}}$.[10] We use the orthonormal base in the sieve approximation series in Eq. (6) for the measurement error probability $f_2(x|x^*)$,

$$f_2(x|x^*) = \sum_{i=1}^{k_{2,n}} \sum_{j=1}^{k_{2,n}} \beta_{2ij} g_i(x) g_j(x^*). \tag{12}$$

As discussed in Section 3.1, the completeness constraint on $f_2(x^*|x)$ to impose on the sieve coefficients is that the square matrix $[\beta_{2ij}]_{k_{2,n} \times k_{2,n}}$ is non-singular. The non-singular property can be imposed by choosing the matrix $[\beta_{2ij}]_{k_{2,n} \times k_{2,n}}$ as a strictly diagonally dominant matrix.[11] This sufficient condition for nonsingularity offers the advantage of allowing for a very straightforward implementation. Alternatively, a standard nonzero determinant conditions could be used, at the expense of a more complex implementation, due to the nonlinearity of the resulting constraint. The diagonal dominance constraint proves convenient, at early stages of the optimization, for efficiently finding an approximate solution. If this constraint turns out to become binding, one can refine the solution using the necessary and sufficient nonzero determinant constraint, a task which is then less numerically challenging because the nonlinear constraint becomes nearly linear in a neighborhood of the solution.

As we use the Legendre orthonormal polynomials as approximation series, the density restrictions $\int_{\mathcal{X}} f_2(x|x^*) dx = 1$, for all $x^*$ can be imposed through $\beta_{211} = 1$ and $\beta_{21j} = 0$ for all $j \neq 1$ and the normalization restrictions $\int_{\mathcal{X}} x f_2(x|x^*) dx = x^*$ can be imposed through $\beta_{222} = 1$ and $\beta_{22j} = 0$ for all $j \neq 2$.[12] We can then choose the following form of sieve coefficients satisfying all three sieve restrictions,

$$\begin{bmatrix} I_{2\times 2} & 0 \\ 0 & D_2 \end{bmatrix},$$

[10] The properties of the Legendre polynomials can be found in Weisstein (2020).

[11] A square matrix is strictly diagonally dominant if the magnitude of the diagonal entry in each row of the matrix is larger than the sum of the magnitudes of all off-diagonal entries in that row. Levy-Desplanques Theorem shows that a strictly diagonally dominant matrix is non-singular. The result can be found as Corollary 5.6.17. in Horn and Johnson (1985). In our case, the condition is $|\beta_{2ii}| > \sum_{j \neq i} |\beta_{2ij}|$ for all $i$.

[12] The density and normalization restrictions stem from the conditions for the Legendre polynomials, $\int_{-1}^{1} g_i(x) dx = 0$ for $i > 1$ and $\int_{-1}^{1} x g_i(x) dx = 0$ for $i \neq 2$. Because of the continuity of $x^*$, these conditions are not just sufficient but also necessary.

where $D_2 = [\beta_{2ij}]_{3 \leq i,j \leq k_{2,n}}$ and $D_2$ is strictly diagonally dominant. Thus, the sieve restrictions may be easily satisfied by using the identity matrix for an initial value for $[\beta_{2ij}]_{k_{2,n} \times k_{2,n}}$. The sieve approximation $f_3(x^*)^{1/2}$ can be constructed in the same manner as $f_1(\eta)^{1/2}$ by using the Legendre orthonormal polynomials.

Three other estimators serve as a basis for comparison. They include: (1) an infeasible estimator based on actually observing $X^*$ (Infeasible with $X^*$), (2) a feasible but biased estimator that ignores the measurement error problem (Biased Estimator), (3) infeasible estimator presented in Eq. (5) using the error-contaminated sample but using knowledge of the distribution of $\eta$, i.e. $f_1$ is assumed to the standard normal density (Infeasible with $\eta$) and $f_2$ and $f_3$ are approximated by the Legendre polynomials. While Tables 1, 3 present the simulation results of the mean, median, and RMSE of the three comparison estimators, Tables 2, 4 report the results for the mean, median, RMSE and AIC$_c$ of the sieve ML estimator. The simulation design contain the three different specifications of the monotonic regression function with two types of DGPs. The Monte Carlo results show that the sieve MLE with different orders of $k_{2,n}$ generally had smaller RMSEs than the Biased Estimator ignoring measurement error. In general, the proposed sieve MLE achieves higher standard deviations than the Infeasible with $\eta$ because the sieve MLE has to estimate the additional unknown function $f_1$. The estimation results of the parameters in all DGPs show small RMSEs and AIC$_c$s for $k_{2,n} = 5$, of $N = 2000$. The RMSEs and AIC$_c$s decrease with the sample size. The means and medians of the estimated parameters are only slightly different in the proposed sieve ML estimator, indicating little skewness in their respective distributions.

### 4.2. Nonparametric Regression Model

The DGPs for $X^*$ and $X$ are the same as the ones in the parametric regression models, but the estimation procedure in this section does not rely on the knowledge of the functional form of the regression function. There are two types of DGPs for the measurement error process which are the same as Section 4.1 and the three regression functions used are:

$$\text{Function 4: } m_0(x^*) = \ln(1.2 + x^* + 0.5x^{*2})$$

$$\text{Function 5: } m_0(x^*) = \begin{cases} 0.8x^* & \text{if } x^* < 0, \\ 1.5x^* & \text{otherwise.} \end{cases}$$

$$\text{Function 6: } m_0(x^*) = \begin{cases} 0.01x^* + 0.01 & \text{if } x^* < 0, \\ x^* + 0.01 & \text{otherwise.} \end{cases}$$

Function 4 is infinitely continuously differentiable, Function 5 has a limited degree of smoothness, and Function 6 is strictly monotone but very close to constant on the interval $[-1, 0]$.

The sieve approximations for $f_1(\eta)^{1/2}$, $f_2(x|x^*)$, and $f_3(x^*)^{1/2}$ are the same as the approximations in the parametric regression models. The identifying restrictions in $f_{X|X^*}$ are imposed through $\beta_{222} = 1$ and $\beta_{22j} = 0$ for all $j \neq 2$ for the normalization and the strictly diagonally dominant matrix $D_2 = [\beta_{2ij}]_{3\leq i,j\leq k_{2,n}}$ for completeness. As for the sieve approximations for the regression function, we also use the Legendre orthonormal polynomials and adopt the series form in Eq. (8) with $k_{4,n} = 4$. The series form in Eq. (8) is an anti-derivative of the sieve expression in Eq. (7) which takes positive values. Thus, this estimation procedure by sampling embeds the monotonicity constraints on $m_0$. The estimator for $m_0(x^*)$ is constructed by sampling the function at the 201 points ranging from -1 to 1 with a 0.01 increment.

Tables 5-7 reports the integrated mean squared errors (IMSE) and AIC$_c$s as a function of $k_{2,n}$ for our sieve MLE estimator $\widehat{m}(x^*)$ for Functions 4, 5, and 6. The results of Table 5 indicate that IMSEs are smaller in DGPs I and II with $k_{2,n} = 4$, and smaller AIC$_c$ for DGP I with $k_{2,n} = 3$ and DGP II with $k_{2,n} = 4$. Table 6 shows that IMSEs are samller for DGP I with $k_{2,n} = 5$ and DGP II with $k_{2,n} = 4$ but AIC$_c$ are samller for DGP I with $k_{2,n} = 4$ and DGP II with $k_{2,n} = 5$. Table 7 shows that IMSEs and AIC$_c$ values point to an optimal choice of $k_{2,n} = 4$ for all DGPs. The estimation results of the sieve MLE for Functions 4, 5, and 6 with $k_{2,n} = 3$, $k_{2,n} = 4$, and 5 are plotted in Figures 1 –3, Figures 4 –6, and Figures 7 –9, respectively. These plots show that the shapes of the estimates (cyan solid lines and black dashed line) are close to the true functions (red dashed line). The black dashed lines represent confidence bands constructed from the 10th and 90th percentile of 1000 curves estimated by sieve MLE. The closeness of the mean plot and the true regression plot in Figures 2, 5, and 8, reflecting small IMSEs. The mean plot and the true regression plot in the Figures are almost inside the black dashed confidence bands.

# 5. Conclusion

This paper investigates identification and estimation of a class of measurement error models without any side information, in which the measurement error may be nonclassical, i.e., correlated with the continuous latent true values. The global nonparametric point identification of the model is proven through two different routes, one exploiting a deep connection between convolutions and completeness for compactly supported densities and the other relying on a spectral decomposition of an integral operator associated with the distribution of observable variables. The main identifying assumptions include restrictions on the range of the regression function and the completeness of the measurement error distribution. Our result allows for a rather flexible structure of regression function and measurement error distribution and thus provides a useful alternative to the existing literature. We also develop a sieve ML estimator for the parameters of interest based on the identification result. We present the asymptotic properties of the sieve MLE and investigate its finite sample properties through a Monte Carlo study and find that it performs satisfactorily.

# A. Proofs

**Lemma A.1.** *For any given probability measure $dA$ supported on a compact interval $[\underline{a},\overline{a}]$ with $\overline{a}>\underline{a}$, the mapping $M:\mathcal{L}^1\left(\left[\underline{b}-\overline{a},\overline{b}-\underline{a}\right]\right)\mapsto\mathcal{L}^1\left(\left[\underline{b},\overline{b}\right]\right)$ with $-\infty<\underline{b}<\overline{b}<\infty$ defined by:*

$$[Mf](x)=\int_{\underline{a}}^{\overline{a}}f(x-u)\,dA(u)\ \text{ for } x\in\left[\underline{b},\overline{b}\right]$$

*is not injective (even if the characteristic function of $A$ is nonvanishing on the real line). The same conclusion holds for any compactly supported measure $dA$ whose Fourier transform has a zero somewhere in the complex plane.*

**Proof** Without loss of generality, we consider $A$ supported on $[-l,l]$ with $l>0$, since the problem can always be reduced to this case by eliminating a trivial translation by $(\underline{a}+\overline{a})/2$ from the mapping $M$. For the same reason, we can also assume that $\underline{b}<0<\bar{b}$. Since $A$ has compact support, by Theorem 7.2.3 and Remark 4 in Lukacs (1970), its characteristic function $\alpha(\xi)$ has infinitely many zeros in the complex plane. (If the characteristic function of $A$ is nonvanishing on the real line, these zeros lie elsewhere in the complex plane.) Let

$\omega + i\rho$ denote one of these zeros. Then, consider the function

$$f(x) = \exp(\rho x)\cos(\omega x)\, 1\left(x \in \left[\underline{b} - l, \overline{b} + l\right]\right)$$

and observe that, for $x \in \left[\underline{b}, \overline{b}\right]$

$$\begin{aligned}
[Mf](x) &= \int_{-l}^{l} \exp(-\rho(x-u))\cos(-\omega(x-u))\, 1\left(x \in \left[\underline{b} - l, \overline{b} + l\right]\right) dA(u) \\
&= \operatorname{Re}\left\{\int_{-l}^{l} \exp(-(\rho + i\omega)(x-u))\, 1\left(x \in \left[\underline{b} - l, \overline{b} + l\right]\right) dA(u)\right\} \\
&= \operatorname{Re}\left\{\int_{-l}^{l} \exp(-(\rho + i\omega)(x-u))\, dA(u)\right\} \\
&= \operatorname{Re}\left\{\exp(-(\rho + i\omega)x)\int_{-l}^{l} \exp((\rho + i\omega)u)\, dA(u)\right\} \\
&= \operatorname{Re}\left\{\exp(-(\rho + i\omega)x)\int_{-\infty}^{\infty} \exp((\rho + i\omega)u)\, dA(u)\right\} \\
&= \operatorname{Re}\left\{\exp(-(\rho + i\omega)x)\,\alpha(\omega + i\rho)\right\} = 0
\end{aligned}$$

Hence we have found a nonzero function supported on $\left[\underline{b} - l, \overline{b} + l\right]$ that is mapped onto the zero function on $\left[\underline{b}, \overline{b}\right]$ and the mapping $M$ is thus not injective. The same construction obviously holds for any measure whose Fourier transform vanishes at some point $\omega + i\rho$ in the complex plane.

**Proof of Theorem 2.1** For a given random variable $V$, we let $\phi_V(\nu) \equiv E[e^{i\nu V}]$ denote its characteristic function and given another random variable $W$ assumed to have a density $f_W(w)$, we also define the "partial" characteristic function $\phi_{V;W}(\nu; w) \equiv E[e^{i\nu V}|W = w]f_W(w)$. We then have,

$$\begin{aligned}
\phi_{Y;X}(\gamma; x) &\equiv E\left[e^{i\gamma Y}|X = x\right] f_X(x) = E\left[E\left[e^{i\gamma Y}|X^*, X = x\right]|X = x\right] f_X(x) \\
&= E\left[E\left[e^{i\gamma Y}|X^*\right]|X = x\right] f_X(x) = E\left[E\left[e^{i\gamma(m(X^*)+\eta)}|X^*\right]|X = x\right] f_X(x) \\
&= E\left[e^{i\gamma m(X^*)}E\left[e^{i\gamma\eta}|X^*\right]|X = x\right] f_X(x) = E\left[e^{i\gamma m(X^*)}E\left[e^{i\gamma\eta}\right]|X = x\right] f_X(x) \\
&= E\left[e^{i\gamma\eta}\right] E\left[e^{i\gamma m(X^*)}|X = x\right] f_X(x) = E\left[e^{i\gamma\eta}\right]\ E\left[e^{i\gamma Y^*}|X = x\right] f_X(x) \\
&\equiv \phi_\eta(\gamma)\,\phi_{Y^*;X}(\gamma; x) \qquad (13)
\end{aligned}$$

where we have used, in turn, the definition of $\phi_{Y;X}(\gamma; x)$, iterated expectations, Assumptions 2.3(i), the definition of $Y$, Assumption 2.2(i), the definition $Y^* \equiv m(X^*)$ and the definition of

$\phi_{Y^*;X}(\gamma;x)$. Observe that Equation (13) states, in Fourier space, that the joint density of $Y$ and $X$ is the convolution of the density of $\eta$ and the probability density of $(X,Y^*)$. Hence, if one knew the distribution of $\eta$, one could determine the joint distribution of $Y^*$ and $X$ from the observed joint distribution of $Y$ and $X$ through the relation $\phi_{Y^*;X}(\gamma;x) = \phi_{Y;X}(\gamma;x)/\phi_\eta(\gamma)$.

As one does not know, a priori, the distribution of $\eta$, we consider some trial zero-mean density of $\eta$ denoted $\tilde{f}_\eta(\eta)$. (The zero-mean constraint is needed to meet the requirement of Assumption 2.2(ii).) To this $\tilde{f}_\eta(\eta)$ corresponds a trial value of all other unobservable quantities (also denoted with tildes). In particular, the trial value of $f_{Y^*,X}(y^*,x)$, denoted $\tilde{f}_{Y^*,X}(y^*,x)$ can be obtained, thanks to Assumptions 2.2 and 2.3(i), through a standard deconvolution procedure:

$$\tilde{\phi}_{Y^*;X}(\gamma;x) = \frac{\phi_{Y;X}(\gamma;x)}{\tilde{\phi}_\eta(\gamma)} \tag{14}$$

To this trial value of $f_{Y^*,X}$ also corresponds a trial regression function $\tilde{m}_0(x^*)$ that can be identified as follows. Note that any valid trial $\tilde{m}_0(x^*)$ must be strictly monotonic and continuous by Assumption 2.4, hence $\tilde{m}_0^{-1}(y^*)$ exists. Furthermore, conditioning on $x^*$ or $y^* \equiv \tilde{m}_0(x^*)$ is equivalent. We can then use the centering restriction (Assumption 2.3(iii)) $G\left[f_{X|X^*}(\cdot|x^*)\right] = x^*$ to write

$$\tilde{m}_0^{-1}(y^*) = G\left[\tilde{f}_{X|Y^*}(\cdot|y^*)\right] \tag{15}$$

where $\tilde{f}_{X|Y^*}(\cdot|y^*) = \tilde{f}_{Y^*X}(y^*,x)/\int \tilde{f}_{Y^*X}(y^*,x)\,dx$, since $G\left[\tilde{f}_{X|Y^*}(\cdot|y^*)\right] = G\left[\tilde{f}_{X|Y^*}(\cdot|\tilde{m}(x^*))\right] = G\left[\tilde{f}_{X|X^*}(\cdot|x^*)\right] = x^* = \tilde{m}_0^{-1}(y^*)$.

In principle, once can compute (14) for any trial density $\tilde{f}_\eta(\eta)$, however, if $\tilde{f}_\eta(\eta)$ is not the true density of $\eta$, this will be detectable in one of the following ways:

1. If $\tilde{f}_\eta(\eta)$ is not a factor[13] of $f_{Y|X}(y|x)$ for some $x$, then $\phi_{Y^*;X}(\gamma;x)/\tilde{\phi}_\eta(\gamma)$ will not be a valid characteristic function for some $x$ (i.e. the inverse Fourier Transform of $\phi_{Y;X}(\gamma;x)/\tilde{\phi}_\eta(\gamma)$ is either taking on negative values or is diverging in such a way that the result is not a probability measure).

2. If $\tilde{f}_\eta(\eta)$ is a factor of $f_{Y|X}(y|x)$ for all $x$, but yields a $\tilde{f}_{Y^*X}(y^*,x)$ with a support that is not compact along $y^*$ ($\sup\{|y^*| : (y^*,x) \in \operatorname{supp} f_{Y^*,X}\} = \infty$), then the resulting $\tilde{m}_0(x^*)$

[13] A probability measure $\mu_A$ is said to be a factor of another probability measure $\mu_B$ if there exists a third probability measure $\mu_C$ such that $\mu_B$ is equal to the convolution of $\mu_A$ and $\mu_C$ (Lukacs (1970)). This definition specializes to the case where $\mu_A$ and $\mu_B$ can be represented by densities and for conditional measures.

(from Equation (15)) will not be bounded, contrary to Assumption 2.4.

3. Next, consider the case where $\tilde{f}_\eta(\eta)$ is a factor of $f_{Y|X}(y|x)$ for all $x$ and yields a $\tilde{f}_{Y^*X}(y^*,x)$ with a compact support along $y^*$. However, $\tilde{f}_\eta(\eta)$ is not the true $f_\eta(\eta)$ but a factor of it. In this case, $\tilde{f}_{Y^*,X}$ can then be written as the convolution, along $y^*$, of the true $f_{Y^*,X}$ with a compactly supported probability measure $a(y^*)$, whose characteristic function is equal to $\alpha(\gamma) \equiv \phi_\eta(\gamma)/\tilde{\phi}_\eta(\gamma)$, since we must have

$$\phi_{Y;X}(\gamma;x) = \tilde{\phi}_{Y^*;X}(\gamma;x)\,\tilde{\phi}_\eta(\gamma) = \phi_{Y^*;X}(\gamma;x)\,\phi_\eta(\gamma)$$

or

$$\tilde{\phi}_{Y^*;X}(\gamma;x) = \phi_{Y^*;X}(\gamma;x)\frac{\phi_\eta(\gamma)}{\tilde{\phi}_\eta(\gamma)} = \phi_{Y^*;X}(\gamma;x)\,\alpha(\gamma)\,. \tag{16}$$

Equation (16) states, in Fourier representation, that $\tilde{f}_{Y^*X}(y^*,x)$ is the convolution of $f_{Y^*X}(y^*,x)$ (for fixed $x$) with the probability measure $a(y^*)$. The measure $a(y^*)$ must be compactly supported, because $\tilde{f}_{Y^*X}(y^*,x)$ is (for a given $x$). But then, the integral operator associated with $\tilde{f}_{Y^*X}(y^*,x)$ can be written as the composition of two operators: the integral operator associated with $f_{Y^*X}(y^*,x)$ and a convolution with $a(y^*)$. But by Lemma A.1, the latter operator is not injective because $a(y^*)$ has compact support. It follows that the operator associated with $\tilde{f}_{Y^*X}(y^*,x)$ cannot be injective either. The same conclusion carries over to the operator associated with $\tilde{f}_{XX^*}(x,x^*)$), since $x^*$ and $y*$ are simply related by a one-to-one mapping, due to the assumed monotonicity of any valid trial $\tilde{m}$. This lack of injectivity contradicts Assumption 2.3(ii).

4. Finally, consider the case where $\tilde{f}_\eta(\eta)$ is a factor $f_{Y|X}(y|x)$ for all $x$ and yields a $\tilde{f}_{Y^*X}(y^*,x)$ with a compact support along $y^*$, and $\tilde{f}_\eta(\eta)$ is neither the true $f_\eta(\eta)$ nor a factor of it. We show by contradiction that this is not possible. Consider the function $\alpha(\gamma) = \phi_\eta(\gamma)/\tilde{\phi}_\eta(\gamma)$ (whose inverse Fourier transform is necessarily not a valid probability measure). By construction, as in case 3, we have the equality $\tilde{\phi}_{Y^*;X}(\gamma;x) = \phi_{Y^*;X}(\gamma;x)\,\alpha(\gamma)$ or

$$\alpha(\gamma) = \frac{\tilde{\phi}_{Y^*;X}(\gamma;x)}{\phi_{Y^*;X}(\gamma;x)}$$

wherever the denominator is not vanishing. Since $\tilde{\phi}_{Y^*;X}(\gamma;x)$ for a given $x$ is the characteristic function of a compactly supported probability measure, by Theorem 7.2.3

and Remark 4 in Lukacs (1970), $\tilde{\phi}_{Y^*;X}(\gamma;x)$ has infinitely many zeros in the complex plane. Next, we make use of a number well-known results in the theory of entire functions (see Boas (1954)). Compact support of $Y^*$ implies that $\tilde{\phi}_{Y^*;X}(\gamma;x)$ is entire. The same conclusion applies to $\phi_{Y^*;X}(\gamma;x)$ since $f_{Y^*X}(y^*,x)$ has compact support along $y^*$. There are three cases to consider. (a) At least one zero of $\tilde{\phi}_{Y^*;X}(\gamma;x)$ does not coincide with a zero of $\phi_{Y^*;X}(\gamma;x)$, or at least one zero does coincide but its multiplicity in $\tilde{\phi}_{Y^*;X}(\gamma;x)$ is higher than that of $\phi_{Y^*;X}(\gamma;x)$. In this case $\alpha(\gamma)$ has a zero somewhere in the complex plane and thus, by the second conclusion of Lemma A.1, the operator associated with $\tilde{f}_{Y^*X}(y^*,x)$ would not be injective, leading to a violation of Assumption 2.3(ii), as in case 3. (b) The case described in (a) holds with the roles of $\tilde{\phi}_{Y^*;X}(\gamma;x)$ and $\phi_{Y^*;X}(\gamma;x)$ reversed, leading to a similar conclusion. (c) Each zero of $\tilde{\phi}_{Y^*;X}(\gamma;x)$ coincides with a zero of $\phi_{Y^*;X}(\gamma;x)$ and these zeros have the same multiplicity. In that case the function $\alpha(\gamma)$ has no zero anywhere in the complex plane and the function $1/\alpha(\gamma)$ has thus no singularity anywhere in the complex plane. We can eliminate the case of zeros "at infinity" by permuting the role of the two alternative models if necessary. Thus $1/\alpha(\gamma)$ has no singularity in the extended complex plane and is bounded. Yet, $\alpha(\gamma)$ is meromorphic because it is the ratio of two entire functions (see Lang (2003), Chapter XIII). This, combined with its lack of singularities, implies that $\alpha(\gamma)$ is also entire.[14] By Liouville's Theorem (e.g., Theorem 7.5 in Lang (2003)), $1/\alpha(\gamma)$ must then be constant. That constant must be 1, since $\alpha(\gamma)$ is the ratio $\phi_\eta(\gamma)/\tilde{\phi}_\eta(\gamma)$ of two characteristic functions (that are necessarily such that $\phi_\eta(0)=\phi_{\tilde{\eta}}(0)=1$). This, in turn, implies that $\phi_\eta(\gamma)/\tilde{\phi}_\eta(\gamma)=1$ or $\tilde{\phi}_\eta(\gamma)=\phi_\eta(\gamma)$ and thus $\tilde{f}_\eta(\eta)$ would in fact be the true distribution of $\eta$.

## B. Alternative Proof of Theorem 2.1

We provide an alternative proof of the main nonparametric identification result in Theorem 2.1.

We first derive the basic integral equation that needs to be solved. Combining Assumption 2.2(i) and 2.3(i), we can obtain the relationship between the observed density and the

[14] This can be shown by noting that a meromorphic function satisfies the Cauchy-Riemann equations (Lang (2003), Section I.6) everywhere, except perhaps at isolated singularities. But if there are no singularities in the extended complex plane, those conditions are everywhere satisfied and the function is thus entire.

unobserved ones:

$$
\begin{aligned}
f_{Y,X}(y,x) &= \int_{\mathcal{X}^*} f_{Y|X^*}(y|x^*) f_{X,X^*}(x,x^*) dx^* \\
&= \int_{\mathcal{X}^*} f_\eta(y - m_0(x^*)) f_{X|X^*}(x|x^*) f_{X^*}(x^*) dx^*.
\end{aligned}
$$

Since a characteristic function of any random variable completely determines its probability distribution, the above equation is equivalent to

$$
\begin{aligned}
\phi_{f_{Y,X=x}}(t) &\equiv \int_{\mathcal{Y}} e^{ity} f_{Y,X}(y,x) dy \\
&= \phi_\eta(t) \int_{\mathcal{X}^*} e^{itm_0(x^*)} f_{X|X^*}(x|x^*) f_{X^*}(x^*) dx^*, \\
&= |\phi_\eta(t)| \int_{\mathcal{X}^*} e^{i(tm_0(x^*)+e(t))} f_{X|X^*}(x|x^*) f_{X^*}(x^*) dx^*,
\end{aligned} \tag{17}
$$

for all real-valued $t$, where $\phi_\eta(t) = \int_\eta e^{it\eta} f_\eta(\eta) d\eta$ and we define $e(t)$ such that the following holds $\phi_\eta(t) \equiv |\phi_\eta(t)| e^{ie(t)}$ and $e(t)$ is the phase of the function. Then Eq. (17) can expressed in terms of two real equations:

$$
Re\phi_{f_{Y,X=x}}(t) = |\phi_\eta(t)| \int_{\mathcal{X}^*} \cos(tm_0(x^*) + e(t)) f_{X|X^*}(x|x^*) f_{X^*}(x^*) dx^*, \tag{18}
$$

$$
Im\phi_{f_{Y,X=x}}(t) = |\phi_\eta(t)| \int_{\mathcal{X}^*} \sin(tm_0(x^*) + e(t)) f_{X|X^*}(x|x^*) f_{X^*}(x^*) dx^*. \tag{19}
$$

Without loss of generality, we can make the following assumption:

**Assumption B.1.** *(Locally symmetric range) The range of the regression function* $\{m_0(x^*) : x^* \in \mathcal{X}^*\}$ *has an open subset containing zero.*

Assumption B.1 is not restrictive because one may always shift the mean of the dependent variable $Y$ and redefine the regression function accordingly. Also, the range of the regression is never reduced to a point, by the strict monotonicity imposed by Assumption 2.4.

Using Assumptions B.1 and 2.4 we can rescale the range of the regression function such that the range is equal to the interval $[-c, d]$ for positive constants $c, d$ and $c + d < \pi$. Because $|\phi_\eta(t)|$ is continuous at 0 (a property of any characteristic function) and $|\phi(0)| = 1$, we can find a $\bar{t} \leq \pi$ such that $0 < |\phi_\eta(t)| < b_1$ for all $t$ in $[0, \bar{t}]$ and a constant $b_1$. Denote the variance of the regression error as $\sigma_\eta^2$. Choose a constant $t_u$ such that

$$0 < t_u < \min\left\{\bar{t}, \sqrt{\frac{2}{\sigma_\eta^2}}\right\}.$$

Use Eq. (18) to derive an operator equivalence relationship as following: for an arbitrary $h \in \mathcal{L}^2([0, t_u])$

$$
\begin{aligned}
(20) \quad (L_{Re\phi_{f_{Y,X}}} h)(x) &= \int Re\phi_{f_{Y,X=x}}(t)h(t)dt, \\
&= \int |\phi_\eta(t)| \int_{\mathcal{X}^*} \cos(tm_0(X^*) + e(t)) f_{X|X^*}(x|x^*) f_{X^*}(x^*) dx^* h(t) dt \\
&= \int_{\mathcal{X}^*} f_{X|X^*}(x|x^*) f_{X^*}(x^*) \left( \int \cos(tm_0(x^*) + e(t)) |\phi_\eta(t)| h(t) dt \right) dx^* \\
&= \int_{\mathcal{X}^*} f_{X|X^*}(x|x^*) f_{X^*}(x^*) \left( \int \cos(tm_0(x^*) + e(t)) (\Delta_{|\phi_\eta|} h)(t) dt \right) dx^* \\
&= \int_{\mathcal{X}^*} f_{X|X^*}(x|x^*) f_{X^*}(x^*) \left( L_{\cos_{m_0,e}} \Delta_{|\phi_\eta|} h)(x^*) \right) dx^* \\
&= \left( L_{f_{X|X^*}} \Delta_{f_{X^*}} L_{\cos_{m_0,e}} \Delta_{|\phi_\eta|} h \right)(x),
\end{aligned}
$$

where we have used (i) Eq. (18), (ii) an interchange of the order of integration (justified by Fubini's theorem), (iii) the definition of $\Delta_{|\phi_\eta|}$ , (iv) the definition of $L_{\cos_{m_0,e}}$ operating on the function $\Delta_{|\phi_\eta|} h$, and (v) the definition of $L_{f_{X|X^*}} \Delta_{f_{X^*}}$ operating on the function $L_{\cos_{m_0,e}} \Delta_{|\phi_\eta|} h$. Thus, we obtain

$$(21) \qquad L_{Re\phi_{f_{Y,X}}} = L_{f_{X|X^*}} \Delta_{f_{X^*}} L_{\cos_{m_0,e}} \Delta_{|\phi_\eta|} \equiv L_{f_{X|X^*}} \Delta_{f_{X^*}} L_{Re\phi_{f_{Y|X^*}}},$$

We can also express Eq. (19) as the following operator equivalence relationships:

$$(22) \qquad L_{Im\phi_{f_{Y,X}}} = L_{f_{X|X^*}} \Delta_{f_{X^*}} L_{\sin_{m_0,e}} \Delta_{|\phi_\eta|} \equiv L_{f_{X|X^*}} \Delta_{f_{X^*}} L_{Im\phi_{f_{Y|X^*}}}.$$

Both $L_{Re\phi_{f_{Y|X^*}}}$ and $L_{Im\phi_{f_{Y|X^*}}}$ are bounded linear operators from $\mathcal{L}^2([0, t_u])$ to $\mathcal{L}^2(\mathcal{X}^*)$ because operators in the right hand side are all bounded by Assumptions 2.1 and continuity of characteristic functions.

Our identification technique is to derive a spectral decomposition of an observed integral operator and show the uniqueness of the decomposition under our assumptions. We can derive some primitive conditions for the invertibility of the operators $L_{Re\phi_{f_{Y,X}}}$, and $L_{Im\phi_{f_{Y,X}}}$ which are related to the invertibility of the operator $L_{f_{X|X^*}}$ and the invertibility of the operators

$L_{Re\phi_{f_{Y|X^*}}}$ and $L_{Im\phi_{f_{Y|X^*}}}$.

**Lemma B.1.** *Assumptions 2.1 and 2.3(ii), $L^{-1}_{f_{X|X^*}}$ exists and is densely defined over $\mathcal{L}^2(\mathcal{X})$.*

See the Online Appendix for the proof. This result shows $L_{f_{X|X^*}}$ is onto and the injectivity of the operators $L_{f_{X|X^*}}$ is directly assumed from the first part of Assumption 2.3(ii). Therefore, $L^{-1}_{f_{X|X^*}}$ exists and $L^{-1}_{f_{X|X^*}} L_{f_{X|X^*}} = L_{f_{X|X^*}} L^{-1}_{f_{X|X^*}} = I$ where $I$ is the identity map from $\mathcal{L}^2(\mathcal{X}^*)$ to itself. The discussion hereafter focuses on the conditions for the completeness of $L_{\cos_{m_0,e}}$, and $L_{\sin_{m_0,e}}$. Define $ns(t) = 1 - \cos(e(t)) = \frac{|\phi_\eta| - Re(\phi_\eta)}{|\phi_\eta|}$ as a measure of degree of non-symmetry. If the distribution of the error term $\eta$ is symmetric then $\phi_\eta(t)$ is real-valued and $ns(t) = 0$ for $t \in [0, t_u]$. Continuity of characteristic functions and Assumption B.1 are sufficient conditions for the invertibility of the operators $L_{\cos_{m_0,e}}$, and $L_{\sin_{m_0,e}}$. We have

**Lemma B.2.** *If Assumption B.1 holds, then each of systems, $\{\cos(tm_0(x^*) + e(t))|\phi_\eta(t)| : x^* \in \mathcal{X}^*\}$ and $\{\sin(tm_0(x^*) + e(t))|\phi_\eta(t)| : x^* \in \mathcal{X}^*\}$, is complete over $\mathcal{L}^2([0, t_u])$. This implies the operators $L_{Re\phi_{f_{Y|X^*}}}$ and $L_{Im\phi_{f_{Y|X^*}}}$ are both injective from $\mathcal{L}^2([0, t_u])$ to $\mathcal{L}^2(\mathcal{X}^*)$.*

The injectivity implies the inverses of $L_{Re\phi_{f_{Y|X^*}}}$ and $L_{Im\phi_{f_{Y|X^*}}}$ exist and can defined over the range of the operators. To show this primitive conditions for the invertibility, we utilize results from Fourier analysis. We provide the following result of the trigonometric system.[15]

**Lemma B.3.** *If $1 < p < \infty$ and $\lambda_k$ is a sequence of distinct real or complex numbers for which $|\lambda_k| \leq k + \frac{1}{2p}, k = 1, 2, 3, ...,$ then the sequence $\{e^{it\lambda_k}\}_{k=1}^{\infty}$ is complete in $\mathcal{L}^p([-\pi, \pi])$.*

We can directly use this neat result to establish the following completeness.

**Lemma B.4.** *If the range of the regression function $\{m_0(x^*) : x^* \in \mathcal{X}^*\}$ contains a sequence of distinct numbers $\{\lambda_1, \lambda_2, \lambda_3, ...\}$ such that $|\lambda_k| \leq k + \frac{1}{4}, k = 1, 2, 3, ...,$ then the family of the functions $\{e^{itm_0(x^*)} : x^* \in \mathcal{X}^*\}$ is complete in $\mathcal{L}^2([-\pi, \pi])$.*

Next, we establish the completeness of the two systems: $\{\cos(tm_0(x^*)) : x^* \in \mathcal{X}^*\}$ and $\{\sin(tm_0(x^*)) : x^* \in \mathcal{X}^*\}$ over $\mathcal{L}^2([0, t_u])$.

**Lemma B.5.** *If the range of the regression function $\{m_0(x^*) : x^* \in \mathcal{X}^*\}$ contains a sequence of distinct numbers $\{\lambda_1, \lambda_2, \lambda_3, ...\}$ such that $|\lambda_k| \leq k + \frac{1}{4}, k = 1, 2, 3, ...,$ then the families of the functions $\{\cos(tm_0(x^*)) : x^* \in \mathcal{X}^*\}$ and $\{\sin(tm_0(x^*)) : x^* \in \mathcal{X}^*\}$ are complete in $\mathcal{L}^2([0, t_u])$ for any $t_u \leq \pi$.*

[15] See Theorem 4 of page 119 in Young (1980).

See the Online Appendix for the proof. This result gives the invertibility of the operators $L_{\cos_{m_0,e}}$, and $L_{\sin_{m_0,e}}$ under the symmetric distribution of the regression error $\eta$, i.e., $e(t) = 0$ $\forall t$. Next, we try to generalize the invertibility or completeness of the symmetric case to a non-symmetric case. Comparing the function in the symmetric case $\cos(tm_0(x^*))$ with the function in the non-symmetric case $\cos(tm_0(x^*) + e(t))$ suggests that we can look into a situation where $e(t)$ is under "small" perturbations around zero (symmetry; $e(t) = 0$ $\forall t$) and investigate what restrictions on the range of $e(t)$ leads to invertibility of operators. In this way, questions about "small" perturbations can be regarded as questions about the stability of completeness because we have already provided a sufficient condition for the symmetric case in Lemma B.5. We will adopt a stability criterion to study completeness. The following result can be found in Young (1980).[16]

**Lemma B.6.** *Let $\{b_k\}$ be a complete sequence for a Hilbert space $(H, \|\cdot\|)$, and suppose that $\{f_k\}$ is sequence of elements of $H$ such that*

$$\|\sum_{k=1}^{n} c_k(b_k - f_k)\| \leq \lambda \|\sum_{k=1}^{n} c_k b_k\|$$

*for some constant $0 \leq \lambda < 1$, and all choices of the scalar $\{c_1, c_2, c_3, ..., c_n\}$, Then $\{f_k\}$ is complete for $H$.*

Lemma B.6 is based on the fact that a bounded linear operator $T$ on a Banach space is invertible whenever $\|I - T\| < 1$ because the inverse operator of $T$ can exist by the formula $T^{-1} = \sum_{k=0}^{\infty}(I - T)^k$.[17] Define $ns(t) = 1 - \cos(e(t)) = \frac{|\phi_\eta| - Re(\phi_\eta)}{|\phi_\eta|}$ as a measure of degree of non-symmetry. If the distribution of the error term $\eta$ is symmetric then $\phi_\eta(t)$ is real-valued and $ns(t) = 0$. The following result provides an upper bound on the absolute values of $ns(t)$ and it will be used to prove Lemma B.2.

**Lemma B.7.** *For $t \in [0, t_u]$, $ns(t)$ is nonnegative and its maximum is less than 1 .*

See the Online Appendix for the proof. Applying the stability criterion and Lemma B.7 to Lemma B.5 under Assumptions B.1 and 2.4, we can prove Lemma B.2. See the Online Appendix for details.

[16] See Problem 2 in page 41. The result is stated for a Banach space and the dense property. Here we adopt Hilbert space version by an important consequence of the Hahn-Banach theorem and the Riesz representation theorem that the dense property is equivalent to the completeness in a Hilbert space.

[17] The result is like ordinary numbers: if $|1-t| < 1$, then $t^{-1}$ exists. More discussions can be found in Young (1980).

In order to provide the onto property of the operators $L_{Re\phi_{f_{Y|X^*}}}$ and $L_{Im\phi_{f_{Y|X^*}}}$, we need a variant of the stability result as in Lemma B.6. We introduce the following notations and statements. Any function $f$ in a Hilbert space can be expressed as a linear combination of the basis function with a unique sequence of scalars $\{c_1, c_2, c_3, ...\}$. Therefore, we can consider $c_n$ as a function of $f$. In fact, $c_n(\cdot)$ is the so-called coefficient functional.[18]

**Definition B.1.** *If $\{f_1, f_2, f_3, ...\}$ is a basis in a Hilbert space $\mathcal{H}$, then every function $f$ in $\mathcal{H}$ has a unique series $\{c_1, c_2, c_3, ...\}$ such that $f = \sum_{n=1}^{\infty} c_n(f) f_n$. Each $c_n$ is a function of $f$. The functionals $c_n$ $(n = 1, 2, 3, ...)$ are called the coefficient functionals associated with the basis $\{f_1, f_2, f_3, ...\}$. Because $c_n$ is a coefficient functional from $\mathcal{H}$ to $\mathbb{R}$. Define its norm by $\|c_n\| = \sup\{|c_n(f)| : f \in \mathcal{H}, \quad \|f\| \leq 1\}$.*

The following results regarding the coefficient functionals are from Theorem 3 in section 6 in Young (1980).

**Lemma B.8.** *If $\{f_1, f_2, f_3, ...\}$ is a basis in a Hilbert space $\mathcal{H}$. Define $c_n$ as coefficient functionals associated with the basis. Then, there exists a constant $M$ such that $1 \leq \|f_n\| \cdot \|c_n\| \leq M$, for all $n$.*

Based on this result, we show, in the Online Appendix, that

**Lemma B.9.** *Denote $H$ as a Hilbert space. Suppose that*

*i) the sequence $\{e_k(\cdot) : k = 1, 2, ...\}$ is a basis in a Hilbert space $\mathcal{H}$;*

*ii) the sequence $\{f_k(\cdot) : k = 1, 2, ...\}$ in $\mathcal{H}$ is $\omega-$independent;*

*iii) $\sum_{n=1}^{\infty} \frac{\|f_k(\cdot) - e_k(\cdot)\|}{\|e_k(\cdot)\|} < \infty$.*

*Then, the sequence $\{f_k(\cdot) : k = 1, 2, ...\}$ is a basis in $\mathcal{H}$.*

Applying this stability result, we have

**Lemma B.10.** *If Assumptions B.1 and 2.4 hold, then each of systems, $\{\cos(tm_0(x^*) + e(t)) : t \in [0, t_u]\}$ and $\{\sin(tm_0(x^*) + e(t)) : t \in [0, t_u]\}$, is complete over $\mathcal{L}^2(\mathcal{X}^*)$. This implies that the inverse operators $L^{-1}_{Re\phi_{f_{Y|X^*}}}$ and $L^{-1}_{Im\phi_{f_{Y|X^*}}}$ exist and are densely defined over $\mathcal{L}^2(\mathcal{X}^*)$.*

See the Online Appendix for the proof.

The completeness results in Lemma B.2 imply the injectivity of $L_{Re\phi_{f_{Y|X^*}}}$ and $L_{Im\phi_{f_{Y|X^*}}}$ while Lemma B.10 gives the onto property of these operators. Therefore, the operators

[18] The introduction of coefficient functional can be found in the page 22 of Young (1980).

invertible with $L^{-1}_{Re\phi_{f_{Y|X^*}}} L_{Re\phi_{f_{Y|X^*}}} = L_{Re\phi_{f_{Y|X^*}}} L^{-1}_{Re\phi_{f_{Y|X^*}}} = I$ and $L^{-1}_{Im\phi_{f_{Y|X^*}}} L_{Im\phi_{f_{Y|X^*}}} = L_{Im\phi_{f_{Y|X^*}}} L^{-1}_{Im\phi_{f_{Y|X^*}}} = I$, where $I$ is the identity map from $\mathcal{L}^2([0,t_u])$ to itself.

Define $L_{K_1}$ as

$$L_{K_1} = L^{-1}_{Re\phi_{f_{Y|X^*}}} L_{Im\phi_{f_{Y|X^*}}}$$

by the existence of $L^{-1}_{Re\phi_{f_{Y|X^*}}}$ over $\mathcal{L}^2(\mathcal{X}^*)$ by Lemma B.10. We can elicit simpler representations of the operator $L_{K_1}$ under Assumption B.1. Furthermore, this simpler representation of $L_{K_1}$ implies the angle function $e(t)$ is identified.

**Lemma B.11.** *If Assumption B.1 holds, then $L_{K_1}$ is a multiplier operator such that $(L_{K_1}h)(t) = \tan(e(t))h(t)$ or $(L_{K_1}h)(t) = \frac{Im\phi_\eta(t)}{Re\phi_\eta(t)}h(t)$ for $t \in [0, t_u]$.*

See the Online Appendix for the proof. We now are ready to prove the main theorem.

**Alternate proof of Theorem 2.1.** We start with the operator equivalence relationships in Eqs. (21) and (22):

$$\begin{aligned}
L_{Re\phi_{f_{Y,X}}} &= L_{f_{X|X^*}} \Delta_{f_{X^*}} L_{\cos_{m_0,e}} \Delta_{|\phi_\eta|} \equiv L_{f_{X|X^*}} \Delta_{f_{X^*}} L_{Re\phi_{f_{Y|X^*}}}, \\
L_{Im\phi_{f_{Y,X}}} &= L_{f_{X|X^*}} \Delta_{f_{X^*}} L_{\sin_{m_0,e}} \Delta_{|\phi_\eta|} \equiv L_{f_{X|X^*}} \Delta_{f_{X^*}} L_{Im\phi_{f_{Y|X^*}}},
\end{aligned}$$

Those operator equivalence relationships may not provide enough information to derive the spectral decomposition of the operator of interest. In order to solicit more useful operator equivalence relationships, we take derivative with respect to $t$ in Eq. (17). It gives that

$$\begin{aligned}
\text{(23)} \qquad \frac{\partial}{\partial t}\phi_{f_{Y,X=x}}(t) = &\left(\frac{\partial}{\partial t}|\phi_\eta(t)|\right) \int_{\mathcal{X}^*} e^{i(tm_0(x^*)+e(t))} f_{X|X^*}(x|x^*) f_{X^*}(x^*) dx^* \\
&+ i\left(\frac{\partial}{\partial t}e(t)\right)|\phi_\eta(t)| \int_{\mathcal{X}^*} e^{i(tm_0(x^*)+e(t))} f_{X|X^*}(x|x^*) f_{X^*}(x^*) dx^* \\
&+ i|\phi_\eta(t)| \int_{\mathcal{X}^*} e^{i(tm_0(x^*)+e(t))} m_0(x^*) f_{X|X^*}(x|x^*) f_{X^*}(x^*) dx^*.
\end{aligned}$$

We split Eq. (23) into a real part and an imaginary part:

$$
\begin{aligned}
(24)\quad Re\frac{\partial}{\partial t}\phi_{f_{Y,X=x}}(t) = &\left(\frac{\partial}{\partial t}|\phi_\eta(t)|\right)\int_{\mathcal{X}^*}\cos(tm_0(x^*)+e(t))f_{X|X^*}(x|x^*)f_{X^*}(x^*)dx^* \\
&-\left(\frac{\partial}{\partial t}e(t)\right)|\phi_\eta(t)|\int_{\mathcal{X}^*}\sin(tm_0(x^*)+e(t))f_{X|X^*}(x|x^*)f_{X^*}(x^*)dx^* \\
&-|\phi_\eta(t)|\int_{\mathcal{X}^*}\sin(tm_0(x^*)+e(t))m_0(x^*)f_{X|X^*}(x|x^*)f_{X^*}(x^*)dx^*,
\end{aligned}
$$

(25)

$$
\begin{aligned}
Im\frac{\partial}{\partial t}\phi_{f_{Y,X=x}}(t) = &\left(\frac{\partial}{\partial t}|\phi_\eta(t)|\right)\int_{\mathcal{X}^*}\sin(tm_0(x^*)+e(t))f_{X|X^*}(x|x^*)f_{X^*}(x^*)dx^* \\
&+\left(\frac{\partial}{\partial t}e(t)\right)|\phi_\eta(t)|\int_{\mathcal{X}^*}\cos(tm_0(x^*)+e(t))f_{X|X^*}(x|x^*)f_{X^*}(x^*)dx^* \\
&+|\phi_\eta(t)|\int_{\mathcal{X}^*}\cos(tm_0(x^*)+e(t))m_0(x^*)f_{X|X^*}(x|x^*)f_{X^*}(x^*)dx^*.
\end{aligned}
$$

We define operators as follows:

$$(26)\quad L_{Re\frac{\partial}{\partial t}\phi_{f_{Y,X}}}:\mathcal{L}^2([0,t_u])\to\mathcal{L}^2(\mathcal{X}) \text{ with } (L_{Re\frac{\partial}{\partial t}\phi_{f_{Y,X}}}h)(x)=\int Re\frac{\partial}{\partial t}\phi_{f_{Y,X=x}}(t)h(t)dt,$$

$$(27)\quad L_{Im\frac{\partial}{\partial t}\phi_{f_{Y,X}}}:\mathcal{L}^2([0,t_u])\to\mathcal{L}^2(\mathcal{X}) \text{ with } (L_{Im\frac{\partial}{\partial t}\phi_{f_{Y,X}}}h)(x)=\int Im\frac{\partial}{\partial t}\phi_{f_{Y,X=x}}(t)h(t)dt,$$

$$(28)\quad \Delta_{\partial|\phi_\eta|}:\mathcal{L}^2([0,t_u])\to\mathcal{L}^2([0,t_u]) \text{ with } (\Delta_{\partial|\phi_\eta|}h)(t)=\left(\frac{\partial}{\partial t}|\phi_\eta(t)|\right)h(t),$$

$$(29)\quad \Delta_{\partial e}:\mathcal{L}^2([0,t_u])\to\mathcal{L}^2([0,t_u]) \text{ with } (\Delta_{\partial e}h)(t)=\left(\frac{\partial}{\partial t}e(t)\right)h(t),$$

$$(30)\quad \Delta_{m_0}:\mathcal{L}^2(\mathcal{X}^*)\to\mathcal{L}^2(\mathcal{X}^*) \text{ with } (\Delta_{m_0}h)(x^*)=m_0(x^*)h(x^*).$$

Similarly to the derivation in Eq. (18), we can obtain operator equivalence relationships to Eqs. (24) and (25) as the following:

$$
\begin{aligned}
(31)\qquad L_{Re\frac{\partial}{\partial t}\phi_{f_{Y,X}}} &= L_{f_{X|X^*}}\Delta_{f_{X^*}}L_{\cos_{m_0,e}}\Delta_{\partial|\phi_\eta|}-L_{f_{X|X^*}}\Delta_{f_{X^*}}L_{\sin_{m_0,e}}\Delta_{|\phi_\eta|}\Delta_{\partial e} \\
&\quad -L_{f_{X|X^*}}\Delta_{f_{X^*}}\Delta_{m_0}L_{\sin_{m_0,e}}\Delta_{|\phi_\eta|},
\end{aligned}
$$

$$
\begin{aligned}
(32)\qquad L_{Im\frac{\partial}{\partial t}\phi_{f_{Y,X}}} &= L_{f_{X|X^*}}\Delta_{f_{X^*}}L_{\sin_{m_0,e}}\Delta_{\partial|\phi_\eta|}+L_{f_{X|X^*}}\Delta_{f_{X^*}}L_{\cos_{m_0,e}}\Delta_{|\phi_\eta|}\Delta_{\partial e} \\
&\quad +L_{f_{X|X^*}}\Delta_{f_{X^*}}\Delta_{m_0}L_{\cos_{m_0,e}}\Delta_{|\phi_\eta|}.
\end{aligned}
$$

Define $\Delta_{\partial\ln|\phi_\eta|}:\mathcal{L}^2([0,t_u])\to\mathcal{L}^2([0,t_u])$ with $(\Delta_{\partial\ln|\phi_\eta|}h)(t)=\left(\frac{\frac{\partial}{\partial t}|\phi_\eta(t)|}{|\phi_\eta(t)|}\right)h(t)$. The following derivation is dedicated to the identification of

$$L_A = L^{-1}_{Re\phi_{f_{Y|X^*}}} \Delta_{m_0} L_{Re\phi_{f_{Y|X^*}}},$$

where $L^{-1}_{Re\phi_{f_{Y|X^*}}}$ exists and is densely defined over $\mathcal{L}^2(\mathcal{X})$ by Lemma B.10. We will show $L_A$ is identified and use it to construct a spectral decomposition. Note that the invertibility of the operators $L_{Re\phi_{f_{Y,X}}}$ and $L_{Im\phi_{f_{Y,X}}}$ is equivalent to the invertibility of operators, $L_{f_{X|X^*}}$, $L_{Re\phi_{f_{Y|X^*}}}$, and $L_{Im\phi_{f_{Y|X^*}}}$ and the boundedness of $f_{X^*}$. While Assumption 2.3(ii) and Lemma B.1.1 permits the invertibility of $L_{f_{X|X^*}}$, Lemma B.2, and Lemma B.10 guarantee the invertibility of $L_{Re\phi_{f_{Y|X^*}}}$, and $L_{Im\phi_{f_{Y|X^*}}}$. The boundedness is ensured by Assumption 2.1. Post-multiplying $L^{-1}_{Re\phi_{f_{Y|X^*}}}$ to Eq. (21) yields

$$L_{Re\phi_{f_{Y,X}}} L^{-1}_{Re\phi_{f_{Y|X^*}}} = L_{f_{X|X^*}} \Delta_{f_{X^*}},$$

which is justified by Lemma B.10. Use this relation to rewrite Eq. (31) as

$$\begin{aligned}
L_{Re\frac{\partial}{\partial t}\phi_{f_{Y,X}}} &= L_{f_{X|X^*}} \Delta_{f_{X^*}} L_{\cos_{m_0,e}} \Delta_{\partial|\phi_\eta|} - L_{f_{X|X^*}} \Delta_{f_{X^*}} L_{\sin_{m_0,e}} \Delta_{|\phi_\eta|} \Delta_{\partial e} \\
&\quad - L_{f_{X|X^*}} \Delta_{f_{X^*}} \Delta_{m_0} L_{\sin_{m_0,e}} \Delta_{|\phi_\eta|}, \\
&= L_{Re\phi_{f_{Y,X}}} \Big[ L^{-1}_{Re\phi_{f_{Y|X^*}}} L_{\cos_{m_0,e}} \Delta_{\partial|\phi_\eta|} - L^{-1}_{Re\phi_{f_{Y|X^*}}} L_{\sin_{m_0,e}} \Delta_{|\phi_\eta|} \Delta_{\partial e} \\
&\quad - L^{-1}_{Re\phi_{f_{Y|X^*}}} \Delta_{m_0} L_{\sin_{m_0,e}} \Delta_{|\phi_\eta|} \Big]
\end{aligned}$$

Because $L_{Re\phi_{f_{Y,X}}}$ is injective by the injectivity of operators, $L_{f_{X|X^*}}$, $L_{Re\phi_{f_{Y|X^*}}}$, and $f_{X^*}$, $L^{-1}_{Re\phi_{f_{Y,X}}} L_{Re\phi_{f_{Y,X}}} = I$. This implies

$$\begin{aligned}
L_{B_1} &\equiv L^{-1}_{Re\phi_{f_{Y,X}}} L_{Re\frac{\partial}{\partial t}\phi_{f_{Y,X}}} \\
&= \left(L_{\cos_{m_0,e}} \Delta_{|\phi_\eta|}\right)^{-1} L_{\cos_{m_0,e}} \Delta_{\partial|\phi_\eta|} - \left( L^{-1}_{Re\phi_{f_{Y|X^*}}} L_{\sin_{m_0,e}} \Delta_{|\phi_\eta|} \right) \Delta_{\partial e} \\
&\quad - \left( L^{-1}_{Re\phi_{f_{Y|X^*}}} \Delta_{m_0} L_{Re\phi_{f_{Y|X^*}}} \right) \left( L^{-1}_{Re\phi_{f_{Y|X^*}}} L_{\sin_{m_0,e}} \Delta_{|\phi_\eta|} \right) \\
&= \Delta_{\partial \ln|\phi_\eta|} - L_{K_1} \Delta_{\partial e} - L_A L_{K_1},
\end{aligned} \tag{33}$$

where we have used $L_{Re\phi_{f_{Y|X^*}}} L^{-1}_{Re\phi_{f_{Y|X^*}}} = I$. Similar, using Eqs. (22) and (32), we obtain

$$\begin{aligned} L_{B_2} &\equiv L^{-1}_{Im\phi_{f_{Y,X}}} L_{Im\frac{\partial}{\partial t}\phi_{f_{Y,X}}} \\ &= \Delta_{\partial \ln|\phi_\eta|} + L^{-1}_{K_1}\Delta_{\partial e} + L^{-1}_{K_1} L_A. \end{aligned} \tag{34}$$

We eliminate the operator $L_A$ in Eqs. (33) and (34) by applying $L_{K_1}$ to the left and right sides of Eq. (34) and then adding with Eq. (33):

$$\begin{aligned} L_C &= L_{B_1} + L_{K_1} L_{B_2} L_{K_1} \\ &= \Delta_{\partial \ln|\phi_\eta|} - L_{K_1}\Delta_{\partial e} + L_{K_1}\Delta_{\partial \ln|\phi_\eta|} L_{K_1} + \Delta_{\partial e} L_{K_1} \\ &= \Delta_{\partial \ln|\phi_\eta|} + L_{K_1}\Delta_{\partial \ln|\phi_\eta|} L_{K_1}, \end{aligned} \tag{35}$$

where we have used $L_{K_1}\Delta_{\partial e} = \Delta_{\partial e} L_{K_1}$ which is justified by Lemma B.11. Note that LHS are observable and $\Delta_{\partial \ln|\phi_\eta|}$ is the unobservable operators in RHS. Applying the observed operator $L_C$ in Eq. (35) to the constant function $\mathbf{1}$ and using Lemma B.11 yields

$$\begin{aligned} (L_C\mathbf{1})(t) &= \frac{\frac{\partial}{\partial t}|\phi_\eta(t)|}{|\phi_\eta(t)|} + \tan(e(t))^2 \frac{\frac{\partial}{\partial t}|\phi_\eta(t)|}{|\phi_\eta(t)|} \\ &= (1+\tan(e(t))^2)\frac{\frac{\partial}{\partial t}|\phi_\eta(t)|}{|\phi_\eta(t)|}. \end{aligned} \tag{36}$$

Because $L_{K_1}$, and therefore $e(t)$, are identified, this implies that both $\frac{\frac{\partial}{\partial t}|\phi_\eta(t)|}{|\phi_\eta(t)|}$ is identified. It follows that $L_A$ is identified from Eq. (34) as follows:

$$L_A = L_{K_1}\left(L_{B_2} - \Delta_{\partial \ln|\phi_\eta|}\right) - \Delta_{\partial e}.$$

Pre-multiplying the operator $L_{f_{X|X^*}}\Delta_{f_{X^*}}$ to the both sides of the equation $L_{Re\phi_{f_{Y|X^*}}} L_A = \Delta_{m_0} L_{Re\phi_{f_{Y|X^*}}}$, we have

$$L_{Re\phi_{f_{Y,X}}} L_A \quad = \quad L_{f_{X|X^*}}\Delta_{f_{X^*}}\Delta_{m_0} L_{Re\phi_{f_{Y|X^*}}}. \tag{37}$$

Post-multiplying the operator $L^{-1}_{Re\phi_{f_{Y|X^*}}}$ to the both sides of Eq. (37) (justified by Lemma B.10) yields

$$L_{Re\phi_{f_{Y,X}}} L_A L^{-1}_{Re\phi_{f_{Y|X^*}}} \quad = \quad L_{f_{X|X^*}}\Delta_{f_{X^*}}\Delta_{m_0}. \tag{38}$$

Because $\Delta_{f_{X^*}}^{-1}$ and $L_{f_{X|X^*}}^{-1}$ both defined over a dense subset of their domain spaces (Assumption 2.1 and Lemma B.1.1 ), we post-multiply these operators to Eq. (38) to obtain

$$\begin{aligned}
\underbrace{L_{Re\phi_{f_{Y,X}}} L_A L_{Re\phi_{f_{Y,X}}}^{-1}}_{\text{Identified}} &= \left(L_{Re\phi_{f_{Y,X}}} L_A L_{Re\phi_{f_{Y|X^*}}}^{-1}\right) \Delta_{f_{X^*}}^{-1} L_{f_{X|X^*}}^{-1} \\
&= L_{f_{X|X^*}} \Delta_{f_{X^*}} \Delta_{m_0} \Delta_{f_{X^*}}^{-1} L_{f_{X|X^*}}^{-1} \\
&= L_{f_{X|X^*}} \Delta_{m_0} L_{f_{X|X^*}}^{-1}.
\end{aligned} \tag{39}$$

The above operator to be diagonalized is defined in terms of observable operators, while the resulting eigenvalues $m_0(x^*)$ and eigenfunctions $f_{X|X^*}(\cdot|x^*)$ (both indexed by $x^*$) provide the unobserved function of interest including the regression function and the joint distribution of the joint distribution of the unobserved regressor $x^*$ and the observed regressor $x$. Assumptions 2.3(iii) and 2.4 ensure the uniqueness of the spectral decomposition of the observed operator Eq. (37). Similarly, we have $f_{Y,X}(y,x) = \int_{\mathcal{X}^*} f_{Y,X^*}(y,x^*) f_{X|X^*}(x|x^*)dx^*$ and it implies that for any $y \in \mathcal{Y}$, $(L_{f_{X|X^*}} f_{Y,X^*})(x) = f_{Y,X}(y,x)$. Thus the identification of $f_{X|X^*}$ induces the identification of $f_{Y,X^*}$ as follow, for any $y \in \mathcal{Y}$,

$$f_{Y,X^*}(y,x^*) = (L_{f_{X|X^*}}^{-1} f_{Y,X})(x^*),$$

where the inverse is justified by the first part of 2.3(ii). Therefore, the densities $f_{Y|X^*}$ and $f_{X^*}$ are identified and so is the regression error distribution $f_\eta$. We have reached our main result. QED.

# References


Ai, C., and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71(6), 1795–1843.

Altonji, J., and R. Matzkin (2005): "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 73(4), 1053–1102.

Andrews, D. (2017): "Examples of $L^2$-complete and Boundedly-complete Distributions," *Journal of Econometrics*, 199(2), 213–220.

Ben-Moshe, D. (2014): "Identification of Dependent Nonparametric Distributions in a System of Linear Equations," Working Paper, The Hebrew University of Jerusalem.

Ben-Moshe, D., X. D'Haultfoeuille, and A. Lewbel (2016): "Identification of Additive and Polynomial Models of Mismeasured Regressors Without Instruments," Working Paper, Hebrew University of Jerusalem.

Blundell, R., X. Chen, and D. Kristensen (2007): "Semi-nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75(6), 1613.

Boas, R. P. (1954): *Entire Functions*. Academic Press.

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, G. Duncan, and W. Rodgers (1994): "Evidence on the Validity of Cross-sectional and Longitudinal Labor Market Data," *Journal of Labor Economics*, 12(3), 345–368.

Bound, J., C. Brown, and N. Mathiowetz (2001): "Measurement Error in Survey Data," *Handbook of econometrics*, pp. 3705–3843.

Bound, J., and A. B. Krueger (1991): "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right," *Journal of Labor Economics*, 9, 1–24.

Burnham, K., and D. Anderson (2002): *Model Selection and Multimodel Inference*. Springer, New York.

Canay, I., A. Santos, and A. Shaikh (2013): "On the Testability of Identification in Some Nonparametric Models with Endogeneity," *Econometrica*, 81(6), 2535–2559.

Carroll, R. J., A. Delaigle, and P. Hall (2007): "Nonparametric regression estimation from data contaminated by a mixture of Berkson and classical errors," *Journal of the Royal Statistical Society B*, 69, 859–878.

Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu (2006): *Measurement Error in Nonlinear Models*. New York: Chapman & Hall.

Chen, X. (2006): "Large Sample Sieve Estimation of Semi-Nonparametric Models," *Handbook of Econometrics*, 6.

Chen, X., V. Chernozhukov, S. Lee, and W. K. Newey (2013): "Local Identification of Nonparametric and Semiparametric Models," Working paper, Yale University.

Chen, X., Y. Hu, and A. Lewbel (2009): "Nonparametric Identification and Estimation of Nonclassical Errors-In-Variables Models without Additional Information," *Statistica Sinica*, 19, 949–968.

Chen, X., and Z. Liao (2014): "Sieve M Inference on Irregular Parameters," *Journal of Econometrics*, 182, 70–86.

Chen, X., and D. Pouzo (2015): "Sieve Wald and QLR Inferences On Semi/Nonparametric Conditional Moment Models," *Econometrica*, 83, 1013–1079.

Chen, X., and X. Shen (1998): "Sieve Extremum Estimates for Weakly Dependent Data," *Econometrica*, 66(2), 289–314.

Chen, X., W. B. Wu, and Y. Yi (2009): "Efficient Estimation of Copula-based Semiparametric Markov Models," *The Annals of Statistics*, 37(6B), 4214–4253.

Chesher, A. (1991): "The Effect of Measurement Error," *Biometrika*, 78, 451.

——— (1998): "Polynomial Regression with Normal Covariate Measurement Error," Discussion Paper 98/448, University of Bristol.

Cragg, J. C. (1997): "Using Higher Moments to Estimate the Simple Errors-in-variables Model," *Rand Journal of Economics*, 28, S71–S91.

Dagenais, M. G., and D. L. Dagenais (1997): "Higher Moment Estimators for Linear Regression Models with Errors in Variables," *Journal of Econometrics*, 76, 193–221.

Delaigle, A., P. Hall, and A. Meister (2008): "On Deconvolution with Repeated Measurements," *Annals of Statistics*, 36, 665–685.

D'Haultfoeuille, X. (2011): "On the Completeness Condition in Nonparametric Instrumental Problems," *Econometric Theory*, 1, 1–12.

Ding, Y., and B. Nan (2011): "A Sieve M-theorem for Bundled Parameters in Semiparametric Models, with Application to the Efficient Estimation in a Linear Model for Censored Data," *Annals of Statistics*, 39(6), 3032–3061.

Erickson, T., C. H. Jiang, and T. M. Whited (2014): "Minimum Distance Estimation of the Errors-in-variables Model Using Linear Cumulant Equations," *Journal of Econometrics*, 183.

Erickson, T., and T. M. Whited (2000): "Measurement Error and the Relationship between Investment and "q"," *Journal of Political Economy*, 108, 1027–1057.

——— (2002): "Two-step GMM Estimation of the Errors-in-variables Model Using High-order Moments," *Econometric Theory*, 18, 776–799.

Freyberger, J. (2017): "On Completeness and Consistency in Nonparametric Instrumental Variable Models," *Econometrica*, 85, 1629–1644.

Geary, R. C. (1942): "Inherent Relations Between Random Variables," *Proceedings of the Royal Irish Academy*, 47A, 63–76.

Ghosal, S., and A. W. Van Der Vaart (2007): "Posterior Convergence Rates of Dirichlet Mixtures at Smooth Densities," *The Annals of Statistics*, 35(2), 697–723.

Glasziou, P., L. Irwig, S. Heritier, R. J. Simes, and A. Tonkin (2008): "Monitoring cholesterol levels: measurement error or true change?," *Annals of Internal Medecine*, 148, 656–661.

Griffin, J. D., and A. H. Lichtenstein (2013): "Dietary Cholesterol and Plasma Lipoprotein Profiles: Randomized-Controlled Trials," *Current Nutrition Reports*, 2, 274–282.

Heckman, J., and E. Vytlacil (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, pp. 669–738.

Hoderlein, S., and E. Mammen (2007): "Identification of Marginal Effects in Nonseparable Models without Monotonicity," *Econometrica*, 75(5), 1513–1518.

Horn, R., and C. Johnson (1985): *Matrix Analysis*. Cambridge University Press.

Härdle, W., and T. Stoker (1989): "Investigating Smooth Multiple Regression by the Method of Average Derivatives," *Journal of the American Statistical Association*, 84, 986–995.

Hu, Y., and S. Schennach (2008): "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 76(1), 195–216.

Hu, Y., S. Schennach, and J. Shiu (2017): "Injectivity of a Class of Integral Operators with Compactly Supported Kernels," *Journal of Econometrics, forthcoming.*

Hu, Y., and J. Shiu (2017): "Nonparametric Identification Using Instrumental Variables: Sufficient Conditions For Completeness," *Econometric Theory, forthcoming.*

Huang, Y., and C. Y. Wang (2001): "Consistent Functional Methods for Logistic Regression with Errors in Covariates," *Journal of the American Statistical Association*, 96, 1469–1482.

Hyslop, D. R., and G. W. Imbens (2001): "Bias from classical and other forms of measurement error," *Journal of Business & Economic Statistics*, 19, 475–481.

Jin, Y., and J. C. Jones-Smith (2015): "Associations Between Family Income and Children's Physical Fitness and Obesity in California, 2010-2012," *Preventing Chronic Diseases*, 12, 140392.

Lang, S. (2003): *Complex Analysis*. Springer.

Lewbel, A. (1997): "Constructing Instruments for Regressions with Measurement Error when no Additional Data are Available, with an Application to Patents and R&D," *Econometrica*, 65(5), 1201–1213.

——— (2012): "Using Heteroskedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models," *Journal of Business & Economic Statistics*, 30, 67–80.

Li, T., and Q. Vuong (1998): "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators," *Journal of Multivariate Analysis*, 65, 139–165.

Lukacs, E. (1970): *Characteristic Functions*. Griffin, London, second edn.

Mattner, L. (1993): "Some Incomplete but Boundedly Complete Location Families," *The Annals of Statistics*, 21(4), 2158–2162.

Newey, W., and J. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71(5), 1565–1578.

Pal, M. (1980): "Consistent Moment Estimators of Regression-coefficients in the Presence of Errors in Variables," *Journal of Econometrics*, 14, 349–364.

Reiersol, O. (1950): "Identifiability of a Linear Relation between Variables Which Are Subject to Error," *Econometrica*, 18, 375–389.

Schennach, S., and Y. Hu (2013): "Nonparametric Identification and Semiparametric Estimation of Classical Measurement Error Models without Side Information," *Journal of the American Statistical Association*, 108(501), 177–186.

Schennach, S. M. (2004): "Estimation of Nonlinear Models with Measurement Error," *Econometrica*, 72, 33–75.

——— (2013): "Instrumental variable treatment of the nonparametric Berkson measurement error model," *Annals of Statistics*, 41, 1642–1668.

——— (2016): "Recent Advances in the Measurement Error Literature," *Annual Reviews of Economics*, 8, 341–377.

Shen, X. (1997): "On Methods of Sieves and Penalization," *Annals of Statistics*, 25, 2555–2591.

van der Laan, M. J., S. Dudoit, and S. Keles (2004): "Asymptotic optimality of likelihood-based cross-validation," *Statistical Applications in Genetics and Molecular Biology*, 3, 4.

Wang, L. (2004): "Estimation of nonlinear models with Berkson measurement errors.," *Annals of Statistics*, 32, 2559–2579.

Weisstein, E. W. (2020): "Legendre Polynomial," *From MathWorld–A Wolfram Web Resource.*

Xue, H., H. Miao, and H. Wu (2010): "Sieve Estimation of Constant and Time-varying Coefficients in Nonlinear Ordinary Differential Equation Models by Considering Both Numerical Error and Measurement Error," *Annals of Statistics*, 38(4), 2351.

Young, R. (1980): *An Introduction to Nonharmonic Fourier Series.* Academic Press.

Table 1: Simulation Results of the Comparison Estimators in Section 4.1 (n=1000)

| | Infeasible with $X^*$ | | Biased Estimator | | Infeasible with $\eta$ | |
|---|---|---|---|---|---|---|
| | $\theta_1 = 0.5$ | $\theta_2 = 0.5$ | $\theta_1 = 0.5$ | $\theta_2 = 0.5$ | $\theta_1 = 0.5$ | $\theta_2 = 0.5$ |
| DGP I: | Function 1: $m_0(x^*;\theta) = \theta_1 x^* + \theta_2 x^{*2} + x^{*3}$ | | | | | |
| Mean | 0.498 | 0.501 | 0.070 | 0.311 | 0.416 | 0.415 |
| Median | 0.498 | 0.505 | 0.069 | 0.311 | 0.405 | 0.416 |
| RMSE | 0.060 | 0.078 | 0.573 | 0.201 | 0.139 | 0.134 |
| | Function 2: $m_0(x^*;\theta) = \theta_1 x^* + \theta_2 e^{x^*}$ | | | | | |
| Mean | 0.497 | 0.501 | 0.164 | 0.483 | 0.419 | 0.385 |
| Median | 0.498 | 0.501 | 0.162 | 0.482 | 0.415 | 0.388 |
| RMSE | 0.065 | 0.027 | 0.342 | 0.033 | 0.134 | 0.148 |
| | Function 3: $m_0(x^*;\theta) = \theta_1 x^* + \frac{\theta_2 x^*}{2-x^*}$ | | | | | |
| Mean | 0.497 | 0.501 | 0.364 | 0.276 | 0.407 | 0.417 |
| Median | 0.497 | 0.501 | 0.365 | 0.274 | 0.407 | 0.413 |
| RMSE | 0.154 | 0.250 | 0.188 | 0.298 | 0.132 | 0.135 |
| DGP II: | Function 1: $m_0(x^*;\theta) = \theta_1 x^* + \theta_2 x^{*2} + x^{*3}$ | | | | | |
| Mean | 0.498 | 0.501 | 0.181 | 0.539 | 0.354 | 0.417 |
| Median | 0.498 | 0.505 | 0.181 | 0.539 | 0.319 | 0.416 |
| RMSE | 0.060 | 0.078 | 0.324 | 0.081 | 0.183 | 0.125 |
| | Function 2: $m_0(x^*;\theta) = \theta_1 x^* + \theta_2 e^{x^*}$ | | | | | |
| Mean | 0.497 | 0.501 | 0.304 | 0.472 | 0.432 | 0.367 |
| Median | 0.498 | 0.501 | 0.303 | 0.472 | 0.431 | 0.329 |
| RMSE | 0.065 | 0.027 | 0.206 | 0.039 | 0.110 | 0.167 |
| | Function 3: $m_0(x^*;\theta) = \theta_1 x^* + \frac{\theta_2 x^*}{2-x^*}$ | | | | | |
| Mean | 0.497 | 0.501 | 0.310 | 0.516 | 0.368 | 0.415 |
| Median | 0.497 | 0.501 | 0.306 | 0.519 | 0.359 | 0.417 |
| RMSE | 0.154 | 0.250 | 0.233 | 0.222 | 0.151 | 0.130 |

Note: The mean, the median and the Root Mean Square Error (RMSE) of the parameters are computed by the estimates over 1000 replications. DGP I and DGP II are referred to the DGPs for the measurement error process in Section 4.1. The orders of $f_2$ and $f_3$ in the estimator, Infeasible with $\eta$, is $k_{2,n} = 4$ and $k_{3,n} = 4$, respectively.

Table 2: Simulation Results of the Sieve MLE in Section 4.1 (n=1000)

| | $k_{2,n}=3$ | | $k_{2,n}=4$ | | $k_{2,n}=5$ | |
|---|---|---|---|---|---|---|
| | $\theta_1=0.5$ | $\theta_2=0.5$ | $\theta_1=0.5$ | $\theta_2=0.5$ | $\theta_1=0.5$ | $\theta_2=0.5$ |
| DGP I: | Function 1: $m_0(x^*;\theta)=\theta_1 x^*+\theta_2 x^{*2}+x^{*3}$ | | | | | |
| Mean | 0.395 | 0.423 | 0.425 | 0.418 | 0.379 | 0.418 |
| Median | 0.387 | 0.427 | 0.425 | 0.416 | 0.381 | 0.423 |
| RMSE | 0.164 | 0.146 | 0.122 | 0.124 | 0.189 | 0.166 |
| AIC$_c$ | 0.230 | | 0.299 | | 0.146 | |
| | Function 2: $m_0(x^*;\theta)=\theta_1 x^*+\theta_2 e^{x^*}$ | | | | | |
| Mean | 0.391 | 0.402 | 0.432 | 0.423 | 0.433 | 0.451 |
| Median | 0.395 | 0.414 | 0.429 | 0.419 | 0.438 | 0.462 |
| RMSE | 0.155 | 0.148 | 0.119 | 0.124 | 0.156 | 0.142 |
| AIC$_c$ | 0.127 | | 0.135 | | 0.135 | |
| | Function 3: $m_0(x^*;\theta)=\theta_1 x^*+\frac{\theta_2 x^*}{2-x^*}$ | | | | | |
| Mean | 0.403 | 0.412 | 0.411 | 0.421 | 0.432 | 0.396 |
| Median | 0.406 | 0.402 | 0.415 | 0.422 | 0.447 | 0.396 |
| RMSE | 0.156 | 0.159 | 0.139 | 0.133 | 0.164 | 0.186 |
| AIC$_c$ | 0.166 | | 0.128 | | 0.136 | |
| DGP II: | Function 1: $m_0(x^*;\theta)=\theta_1 x^*+\theta_2 x^{*2}+x^{*3}$ | | | | | |
| Mean | 0.409 | 0.436 | 0.438 | 0.458 | 0.402 | 0.421 |
| Median | 0.404 | 0.444 | 0.417 | 0.428 | 0.401 | 0.424 |
| RMSE | 0.152 | 0.123 | 0.146 | 0.148 | 0.166 | 0.147 |
| AIC$_c$ | 0.218 | | 0.268 | | 0.194 | |
| | Function 2: $m_0(x^*;\theta)=\theta_1 x^*+\theta_2 e^{x^*}$ | | | | | |
| Mean | 0.417 | 0.408 | 0.422 | 0.402 | 0.429 | 0.456 |
| Median | 0.414 | 0.423 | 0.424 | 0.405 | 0.432 | 0.467 |
| RMSE | 0.146 | 0.145 | 0.131 | 0.137 | 0.148 | 0.124 |
| AIC$_c$ | 0.165 | | 0.193 | | 0.205 | |
| | Function 3: $m_0(x^*;\theta)=\theta_1 x^*+\frac{\theta_2 x^*}{2-x^*}$ | | | | | |
| Mean | 0.396 | 0.407 | 0.407 | 0.421 | 0.428 | 0.422 |
| Median | 0.385 | 0.400 | 0.409 | 0.420 | 0.441 | 0.416 |
| RMSE | 0.160 | 0.153 | 0.137 | 0.132 | 0.153 | 0.165 |
| AIC$_c$ | 0.268 | | 0.220 | | 0.194 | |

Note: The mean, the median and the Root Mean Square Error (RMSE) of the parameters are computed by the estimates over 1000 replications. DGP I and DGP II are referred to the DGPs for the measurement error process in Section 4.1. The orders of the sieve approximations in the sieve MLE are $k_{1,n}=4$, $k_{3,n}=4$, and $k_{4,n}=4$.

Table 3: Simulation Results of the Comparison Estimators in Section 4.1 (n=2000)

| | Infeasible with $X^*$ | | Biased Estimator | | Infeasible with $\eta$ | |
|---|---|---|---|---|---|---|
| | $\theta_1 = 0.5$ | $\theta_2 = 0.5$ | $\theta_1 = 0.5$ | $\theta_2 = 0.5$ | $\theta_1 = 0.5$ | $\theta_2 = 0.5$ |
| DGP I: | Function 1: $m_0(x^*;\theta) = \theta_1 x^* + \theta_2 x^{*2} + x^{*3}$ | | | | | |
| Mean | 0.502 | 0.499 | 0.067 | 0.310 | 0.461 | 0.484 |
| Median | 0.503 | 0.502 | 0.064 | 0.310 | 0.461 | 0.485 |
| RMSE | 0.042 | 0.055 | 0.569 | 0.197 | 0.101 | 0.099 |
| | Function 2: $m_0(x^*;\theta) = \theta_1 x^* + \theta_2 e^{x^*}$ | | | | | |
| Mean | 0.502 | 0.500 | 0.169 | 0.482 | 0.477 | 0.463 |
| Median | 0.503 | 0.499 | 0.170 | 0.481 | 0.478 | 0.467 |
| RMSE | 0.047 | 0.019 | 0.335 | 0.027 | 0.101 | 0.092 |
| | Function 3: $m_0(x^*;\theta) = \theta_1 x^* + \frac{\theta_2 x^*}{2-x^*}$ | | | | | |
| Mean | 0.505 | 0.495 | 0.169 | 0.482 | 0.468 | 0.490 |
| Median | 0.504 | 0.502 | 0.170 | 0.481 | 0.469 | 0.487 |
| RMSE | 0.115 | 0.180 | 0.335 | 0.027 | 0.098 | 0.101 |
| DGP II: | Function 1: $m_0(x^*;\theta) = \theta_1 x^* + \theta_2 x^{*2} + x^{*3}$ | | | | | |
| Mean | 0.502 | 0.499 | 0.183 | 0.538 | 0.444 | 0.473 |
| Median | 0.503 | 0.502 | 0.184 | 0.537 | 0.442 | 0.473 |
| RMSE | 0.042 | 0.055 | 0.320 | 0.062 | 0.105 | 0.093 |
| | Function 2: $m_0(x^*;\theta) = \theta_1 x^* + \theta_2 e^{x^*}$ | | | | | |
| Mean | 0.502 | 0.500 | 0.307 | 0.471 | 0.461 | 0.450 |
| Median | 0.503 | 0.499 | 0.308 | 0.471 | 0.461 | 0.448 |
| RMSE | 0.047 | 0.019 | 0.198 | 0.035 | 0.096 | 0.097 |
| | Function 3: $m_0(x^*;\theta) = \theta_1 x^* + \frac{\theta_2 x^*}{2-x^*}$ | | | | | |
| Mean | 0.505 | 0.495 | 0.313 | 0.516 | 0.448 | 0.470 |
| Median | 0.504 | 0.502 | 0.315 | 0.512 | 0.448 | 0.472 |
| RMSE | 0.115 | 0.180 | 0.212 | 0.156 | 0.097 | 0.096 |

Note: The mean, the median and the Root Mean Square Error (RMSE) of the parameters are computed by the estimates over 1000 replications. DGP I and DGP II are referred to the DGPs for the measurement error process in Section 4.1. The orders of $f_2$ and $f_3$ in the estimator, Infeasible with $\eta$, is $k_{2,n} = 4$ and $k_{3,n} = 4$, respectively.

Table 4: Simulation Results of the Sieve MLE in Section 4.1 (n=2000)

| | $k_{2,n}=3$ | | $k_{2,n}=4$ | | $k_{2,n}=5$ | |
|---|---|---|---|---|---|---|
| | $\theta_1=0.5$ | $\theta_2=0.5$ | $\theta_1=0.5$ | $\theta_2=0.5$ | $\theta_1=0.5$ | $\theta_2=0.5$ |
| DGP I: | Function 1: $m_0(x^*;\theta)=\theta_1 x^*+\theta_2 x^{*2}+x^{*3}$ | | | | | |
| Mean | 0.388 | 0.434 | 0.430 | 0.422 | 0.481 | 0.505 |
| Median | 0.382 | 0.442 | 0.431 | 0.414 | 0.483 | 0.508 |
| RMSE | 0.164 | 0.142 | 0.114 | 0.120 | 0.098 | 0.094 |
| AIC$_c$ | 0.197 | | 0.314 | | 0.062 | |
| | Function 2: $m_0(x^*;\theta)=\theta_1 x^*+\theta_2 e^{x^*}$ | | | | | |
| Mean | 0.387 | 0.410 | 0.428 | 0.424 | 0.488 | 0.499 |
| Median | 0.392 | 0.417 | 0.428 | 0.421 | 0.487 | 0.516 |
| RMSE | 0.150 | 0.137 | 0.121 | 0.117 | 0.100 | 0.109 |
| AIC$_c$ | 0.087 | | 0.051 | | 0.034 | |
| | Function 3: $m_0(x^*;\theta)=\theta_1 x^*+\frac{\theta_2 x^*}{2-x^*}$ | | | | | |
| Mean | 0.399 | 0.417 | 0.413 | 0.419 | 0.490 | 0.518 |
| Median | 0.407 | 0.413 | 0.425 | 0.426 | 0.483 | 0.514 |
| RMSE | 0.157 | 0.163 | 0.136 | 0.137 | 0.101 | 0.111 |
| AIC$_c$ | 0.199 | | 0.042 | | 0.036 | |
| DGP II: | Function 1: $m_0(x^*;\theta)=\theta_1 x^*+\theta_2 x^{*2}+x^{*3}$ | | | | | |
| Mean | 0.404 | 0.447 | 0.440 | 0.468 | 0.476 | 0.511 |
| Median | 0.386 | 0.457 | 0.414 | 0.422 | 0.474 | 0.515 |
| RMSE | 0.154 | 0.113 | 0.149 | 0.154 | 0.095 | 0.098 |
| AIC$_c$ | 0.191 | | 0.267 | | 0.048 | |
| | Function 2: $m_0(x^*;\theta)=\theta_1 x^*+\theta_2 e^{x^*}$ | | | | | |
| Mean | 0.418 | 0.421 | 0.423 | 0.405 | 0.488 | 0.500 |
| Median | 0.424 | 0.437 | 0.421 | 0.409 | 0.490 | 0.502 |
| RMSE | 0.138 | 0.136 | 0.125 | 0.129 | 0.099 | 0.096 |
| AIC$_c$ | 0.196 | | 0.176 | | 0.040 | |
| | Function 3: $m_0(x^*;\theta)=\theta_1 x^*+\frac{\theta_2 x^*}{2-x^*}$ | | | | | |
| Mean | 0.380 | 0.411 | 0.402 | 0.419 | 0.492 | 0.503 |
| Median | 0.359 | 0.403 | 0.402 | 0.419 | 0.486 | 0.502 |
| RMSE | 0.169 | 0.150 | 0.135 | 0.130 | 0.102 | 0.111 |
| AIC$_c$ | 0.341 | | 0.210 | | 0.024 | |

Note: The mean, the median and the Root Mean Square Error (RMSE) of the parameters are computed by the estimates over 1000 replications. DGP I and DGP II are referred to the DGPs for the measurement error process in Section 4.1. The orders of the sieve approximations in the sieve MLE are $k_{1,n}=4$, $k_{3,n}=4$, and $k_{4,n}=4$.

Table 5: The IMSEs in the Estimation of Function 4 in Section 4.2

| | $k_{2,n}=3$ | | $k_{2,n}=4$ | | $k_{2,n}=5$ | |
|---|---|---|---|---|---|---|
| | DGP I | DGP II | DGP I | DGP II | DGP I | DGP II |
| N=1000: | 0.531 | 0.273 | 0.218 | 0.203 | 0.336 | 0.476 |
| $\text{AIC}_c$ | 0.289 | 0.538 | 0.019 | 0.573 | 0.448 | 0.655 |
| N=2000: | 0.581 | 0.259 | 0.242 | 0.202 | 0.321 | 0.559 |
| $\text{AIC}_c$ | 0.337 | 0.513 | 0.014 | 0.887 | 0.395 | 0.650 |

Note: The IMSEs are defined by $\int[m(x^*) - m_0(x^*)]^2 dx^*$. DGP I and DGP II are referred to the DGPs for the measurement error process in Section 4.1.

Table 6: The IMSEs in the Estimation of Function 5 in Section 4.2

| | $k_{2,n}=3$ | | $k_{2,n}=4$ | | $k_{2,n}=5$ | |
|---|---|---|---|---|---|---|
| | DGP I | DGP II | DGP I | DGP II | DGP I | DGP II |
| N=1000: | 0.606 | 0.366 | 0.242 | 0.209 | 0.201 | 0.258 |
| $\text{AIC}_c$ | 0.182 | 0.175 | 0.008 | 0.259 | 0.187 | 0.016 |
| N=2000: | 0.588 | 0.316 | 0.287 | 0.183 | 0.199 | 0.245 |
| $\text{AIC}_c$ | 0.158 | 0.207 | 0.014 | 0.326 | 0.126 | 0.045 |

Note: The IMSEs are defined by $\int[m(x^*) - m_0(x^*)]^2 dx^*$. DGP I and DGP II are referred to the DGPs for the measurement error process in Section 4.1.

Table 7: The IMSEs in the Estimation of Function 6 in Section 4.2

| | $k_{2,n}=3$ | | $k_{2,n}=4$ | | $k_{2,n}=5$ | |
|---|---|---|---|---|---|---|
| | DGP I | DGP II | DGP I | DGP II | DGP I | DGP II |
| N=1000: | 0.409 | 0.307 | 0.263 | 0.196 | 0.546 | 0.482 |
| $\text{AIC}_c$ | 0.218 | 0.474 | 0.011 | 0.293 | 0.667 | 0.806 |
| N=2000: | 0.437 | 0.377 | 0.249 | 0.213 | 0.527 | 0.503 |
| $\text{AIC}_c$ | 0.166 | 0.502 | 0.009 | 0.330 | 0.626 | 0.889 |

Note: The IMSEs are defined by $\int[m(x^*) - m_0(x^*)]^2 dx^*$. DGP I and DGP II are referred to the DGPs for the measurement error process in Section 4.1.

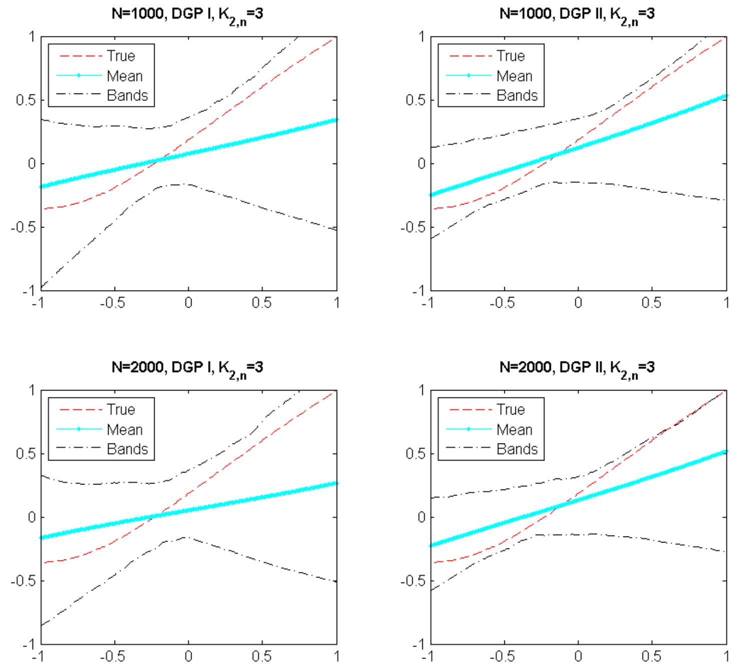


Figure 1: The Estimated Function 4 in Section 4.2 Using $k_{2,n} = 3$ for $f_2(x|x^*)$

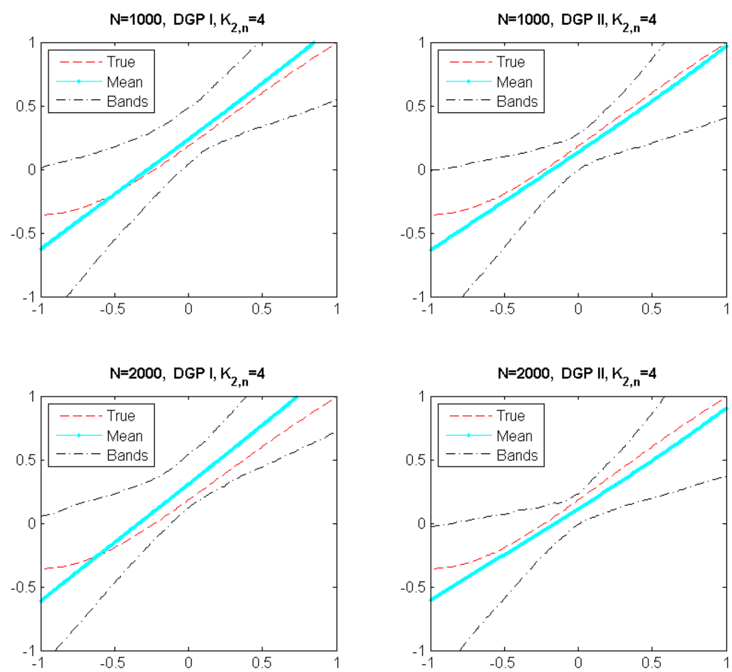


Figure 2: The Estimated Function 4 in Section 4.2 Using $k_{2,n} = 4$ for $f_2(x|x^*)$

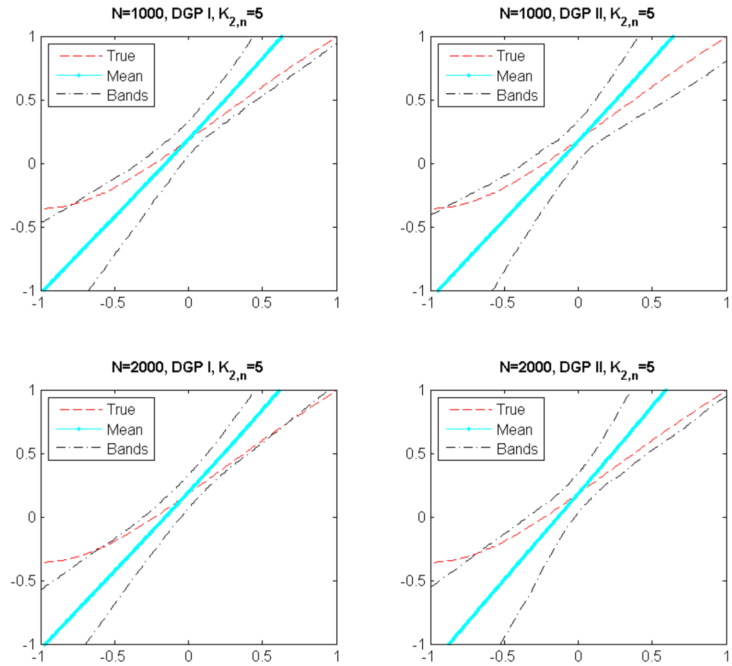


Figure 3: The Estimated Function 4 in Section 4.2 Using $k_{2,n} = 5$ for $f_2(x|x^*)$

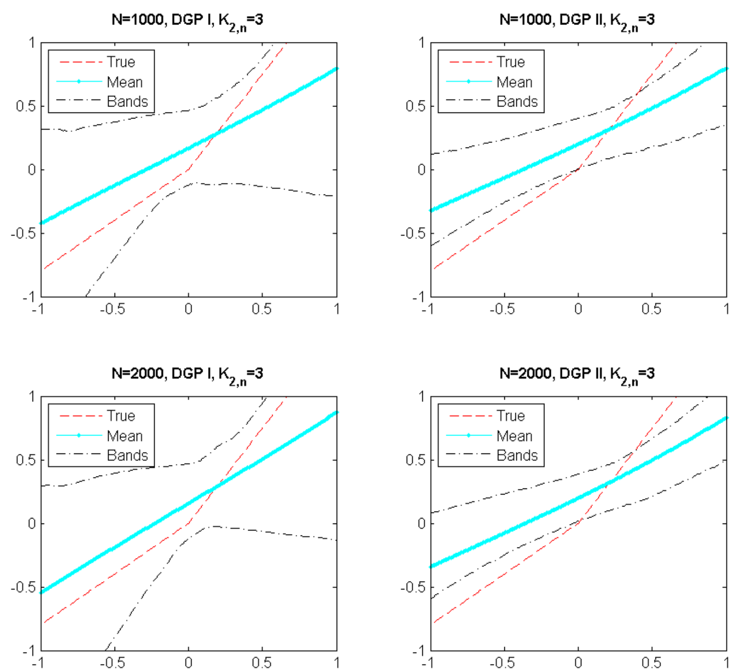


Figure 4: The Estimated Function 5 in Section 4.2 Using $k_{2,n} = 3$ for $f_2(x|x^*)$

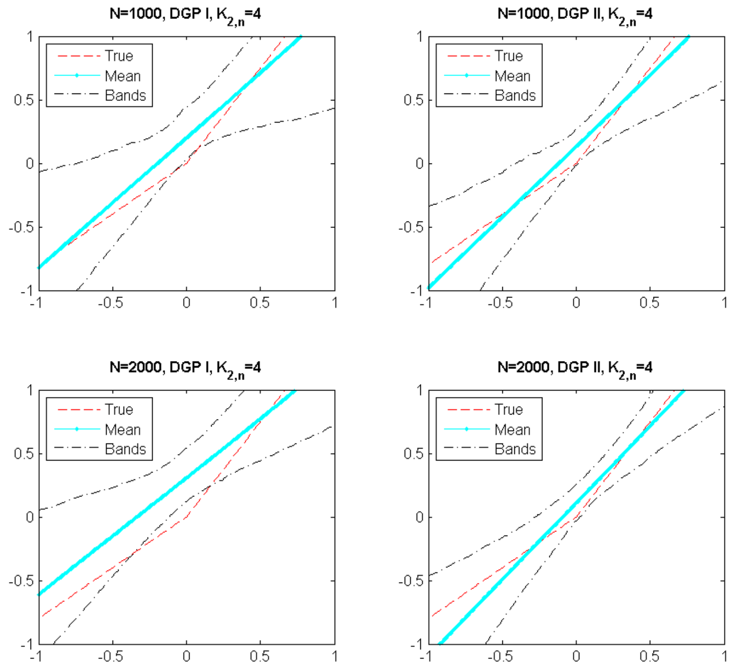


Figure 5: The Estimated Function 5 in Section 4.2 Using $k_{2,n} = 4$ for $f_2(x|x^*)$

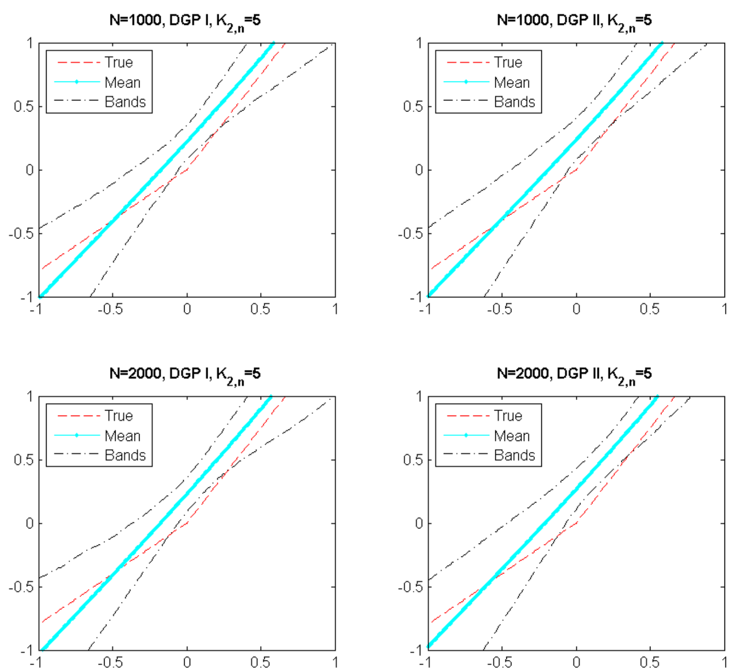


Figure 6: The Estimated Function 5 in Section 4.2 Using $k_{2,n} = 5$ for $f_2(x|x^*)$

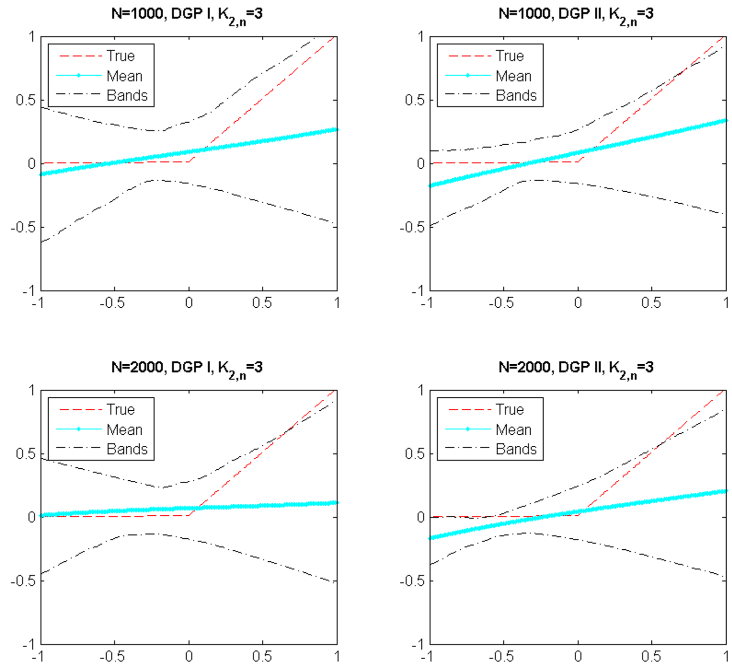


Figure 7: The Estimated Function 6 in Section 4.2 Using $k_{2,n} = 3$ for $f_2(x|x^*)$

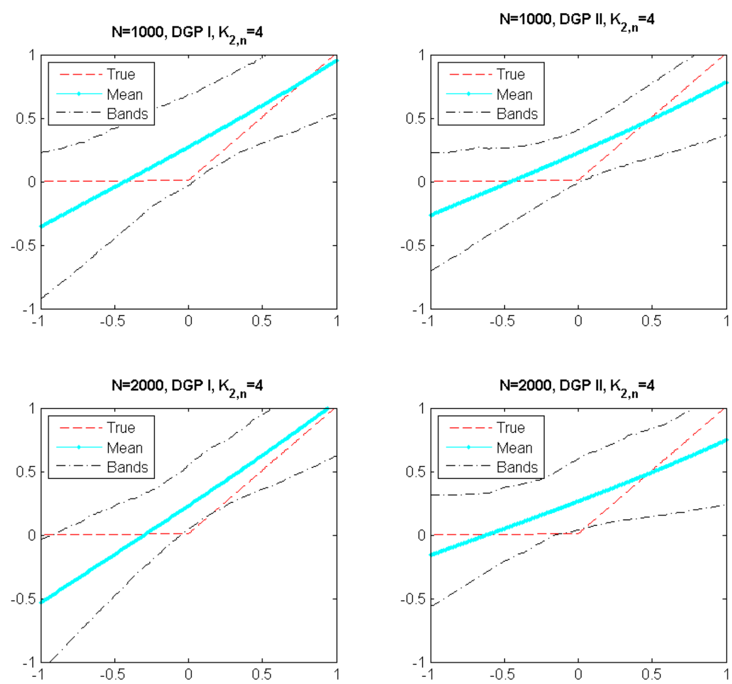


Figure 8: The Estimated Function 6 in Section 4.2 Using $k_{2,n} = 4$ for $f_2(x|x^*)$

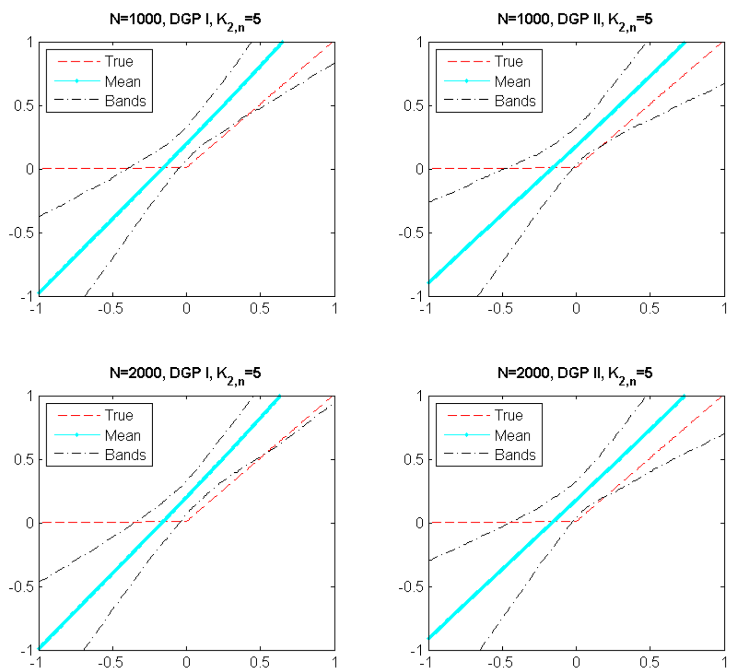


Figure 9: The Estimated Function 6 in Section 4.2 Using $k_{2,n} = 5$ for $f_2(x|x^*)$