Closed-Form Estimation of Nonparametric Models with Non-Classical Measurement Errors

Yingyao Hu   Yuya Sasaki*

Johns Hopkins   Johns Hopkins

November 30, 2014

Abstract

This paper proposes closed-form estimators for nonparametric regressions using two measurements with non-classical errors. One (administrative) measurement has location-/scale-normalized errors, but the other (survey) measurement has endogenous errors with arbitrary location and scale. For this setting of data combination, we derive closed-form identification of nonparametric regressions, and practical closed-form estimators that perform well with small samples. Applying this method to NHANES III, we study how obesity explains health care usage. Clinical measurements and self reports of BMI are used as two measurements with normalized errors and endogenous errors, respectively. We robustly find that health care usage increases with obesity.

Keywords: closed form, non-classical measurement errors, nonparametric regressions

JEL Numbers: C14, C21

*Yingyao Hu: Department of Economics, 440 Mergenthaler Hall, 3400 N. Charles St, Baltimore, MD 21218. Email: yhu@jhu.edu. Yuya Sasaki: Department of Economics, 440 Mergenthaler Hall, 3400 N. Charles St, Baltimore, MD 21218. Email: sasaki@jhu.edu. We benefited from useful comments by seminar participants at University of Tokyo and NASM 2013. The usual disclaimer applies.

1 Introduction

For the increasing availability of combined administrative and survey data (Ridder and Moffitt, 2007), econometric methods that can properly handle matched data with measurement errors have become of great practical importance. For econometric methods to be truly useful no matter how complicated a model is, estimators should ideally be given in a closed form explicitly written in terms of observed data, like the OLS. Unfortunately, such convenient characteristics are rarely shared by nonparametric estimators for non-classical measurement errors.

Identification and estimation of regression models with two measurements of explanatory variables are proposed by Li (2002) and Schennach (2004a,b) among others. A limitation with the existing methods is that they require two measurements with classical errors. In practice, empirical data with two measurements often come from matched administrative, imputed, and/or survey data, where particularly survey data are often subject to non-classical errors (e.g., Bound, Brown, and Mathiowetz, 2001; Koijen, Van Nieuwerburgh, and Vestman, 2013). Ignoring the non-classical nature of errors in measurements may lead to inconsistent estimation, as we demonstrate in our simulations. In this paper, we propose closed-form estimators for nonparametric regression models using two measurements with non-classical errors.

Specifically, we explicitly estimate the nonparametric regression function g for the model

$$Y = g(X^*) + U \qquad E\left[U | X^*\right] = 0,$$

where Y is an observed dependent variable, X^* is an unobserved explanatory variable, and U is the regression residual. While the true explanatory variable X^* is not observed, two measurements, X_1 and X_2, are available from matched data. For simplicity, X^* is assumed to be a scalar and continuously distributed. The relationship between the two measurements and the true explanatory variable X^* is modeled as follows.

$$\begin{aligned} X_1 &= \sum_{p=0}^{P} \gamma_p X^{*p} + \mathcal{E}_1 \\ X_2 &= X^* + \mathcal{E}_2 \end{aligned}$$

Unless $\gamma_1 = 1$ and $\gamma_2 = \cdots = \gamma_P = 0$ are true, the first measurement X_1 entails non-classical errors with nonlinearity. Allowing for such non-classical errors is crucial particularly for survey data that are often contaminated by endogenous self-reporting biases. Since the truth X^* is unobserved, the second measurement X_2 is location-/scale-normalized with respect to the unobserved truth X^*. We use alternative independence assumptions on the measurement error $\mathcal{E}_2$ depending on which order P we assume about X_1, but these assumptions are more innocuous than assuming classical errors in any case.

Under assumptions that will be introduced below, we show that the regression function g can be explicitly expressed as a functional of the joint CDF $F_{YX_1X_2}$ in the sense that $g(x^*) = \lambda(x^*|F_{YX_1X_2})$. We provide the concrete expression for this functional $\lambda(x^* \mid \cdot\ )$. In order to construct a sample-counterpart estimator of $g(x^*)$ given this closed-form identifying solution, it suffices to substitute the empirical distribution $\widehat{F}_{YX_1X_2}$ in this known transformation so we get the closed-form estimator $\widehat{g(x^*)} = \lambda(x^* \mid \widehat{F}_{YX_1X_2})$. We present its theoretical large sample properties as well as its small sample performance. Monte Carlo simulations show that the estimator works quite well with $N = 500$, a very small sample size for nonparametrics.

Measurement error models have been extensively studied in both statistics and econometrics. The statistical literature focuses on cases of classical errors, where measurement errors are independent of the true values – see Fuller (1987) and Carroll, Ruppert, Stefanski and Crainiceanu (2006) for reviews. The econometric literature investigates nonlinear models and nonclassical measurement errors – see Chen, Hong and Nekipelov (2011), Bound, Brown and Mathiowetz (2001) and Schennach (2013) for reviews. However, closed-form estimation, nonlinear/nonparametric models, and non-classical measurement errors still remain unsolved, despite their joint practical relevance. Two measurements are known to be useful to correct measurement errors even for external samples if the matched administrative data is known to be true (e.g., Chen, Hong, and Tamer, 2005). The baseline model of our framework was introduced by Li (2002) and Schennach (2004a), where they consider parametric regression models under two measurements with classical errors. Hu and Schennach (2008) provide general identification results for nonseparable and non-classical measurement errors,[1] but their estimator relies on semi-/non-parametric extremal estimator where nuisance functions are approximated by truncated series. [2] Unlike these existing approaches, we develop a closed-form estimator for nonparametric models involving non-classical measurement errors.

Our results share much in common with Schennach (2004b) where she develops a closed-form estimator under the restriction, $\gamma_1 = 1$ and $\gamma_2 = \cdots = \gamma_P = 0$, of a classical-error structure. There are notable differences and thus values added by this paper as well. Our method paves the way for non-classical error structures with high degrees of nonlinearity whereas the existing closed-form estimator can handle only classical errors. To this end, we propose a new method to recover and use the characteristic function of the generated latent variable $\sum_{p=1}^{P} \gamma_p X^{*p}$, instead of just X^*, in the framework of deconvolution approaches. Not surprisingly, as we show through simulations, the classical error assumption $\gamma_1 = 1$ and $\gamma_2 = \cdots = \gamma_P = 0$ can severely bias estimates if the true DGP does not conform with this assumption. In our empirical application, we find that $\gamma_1 \neq 1$ is indeed true when people report their physical characteristics, and hence the existing closed-form estimator that assumes classical errors would likely suffer from biased estimates. The contribution of our method is to overcome these practical limitations of the existing closed-form estimators.

For an empirical illustration, we investigate how obesity measured by the Body Mass Index (BMI) explains the health care usage by using a sample of about 1900 observations extracted from the National Health and Nutrition Examination Survey (NHANES III). This data set

[1] Also see Mahajan (2006), Lewbel (2007), and Hu (2008) for non-/semi-parametric identification and estimation under non-classical measurement errors with discrete variables.

[2] Our model is also closely related to nonparametric regression models with classical measurement errors, which are extensively studied in the rich literature in statistics. When the error distribution is known, the regression function may be estimated by deconvolution – see Fan and Truong (1993) and Carroll, Ruppert, Stefanski and Crainiceanu (2006) for reviews. When the error distribution is unknown, Schennach (2004b) uses Kotlarski's identify (see Rao, 1992) to provide a Nadaraya-Watson-type estimator for the regression function.

uniquely matches self-reports and clinical measurements of the BMI. We allow the former measurement to suffer from endogenous biases with arbitrary location and scale, while the latter measurement is location-/scale-normalized with respect to the true BMI. Our results show a robust upward-sloping tendency of the mean health care usage as a function of the true BMI, controlling for the most important health factors, namely gender and age. This tendency is particularly stronger for females.

2 Closed-Form Identification: A Baseline Model

Our objective is to derive closed-form identifying formulas for the nonparametric regression function g. For the purpose of intuitive exposition, we first focus on the following simple model:

$$\begin{aligned} Y &= g(X^*) + U, & \mathrm{E}[U \mid X^*] = 0 \\ X_1 &= \gamma_1 X^* + \mathcal{E}_1 & \mathrm{E}[\mathcal{E}_1] = \gamma_0 \\ X_2 &= X^* + \mathcal{E}_2, & \mathrm{E}[\mathcal{E}_2] = 0 \end{aligned} \tag{2.1}$$

where we observe the joint distribution of (Y, X_1, X_2). The restriction $\mathrm{E}[U \mid X^*] = 0$ means that $g(X^*)$ is the nonparametric regression of Y on X^*. We do not assume $\mathrm{E}[\mathcal{E}_1]$ to be zero in order to accommodate arbitrary intercept γ_0 for the first measurement X_1. As such, we suppress γ_0 from the equation for X_1, i.e., it is embedded in $\gamma_0 = \mathrm{E}[\mathcal{E}_1]$. On the other hand, the locational normalization $\mathrm{E}[\mathcal{E}_2] = 0$ is imposed on the second measurement X_2. A leading example of (2.1) is the case with $\gamma_1 = 1$ often assumed in related papers in the literature. We do not make such an assumption, and thus our model (2.1) accommodates the possibility that the first measurement X_1 is endogenously biased even if $X^* \perp\!\!\!\perp \mathcal{E}_1$ is assumed, as $\mathrm{E}[X_1 - X^* \mid X^*] = \gamma_0 + (\gamma_1 - 1)X^*$.

We can easily show that γ_1 is identified from the observed data by the closed-form formula

$$\gamma_1 = \frac{\mathrm{Cov}(Y, X_1)}{\mathrm{Cov}(Y, X_2)} \tag{2.2}$$

under the following assumption.

Assumption 1 (Identification of γ_1)**.** *$Cov(\mathcal{E}_1, Y) = Cov(\mathcal{E}_2, Y) = 0$ and $Cov(Y, X_2) \neq 0$.*

The first part of this assumption requires that $\mathcal{E}_1$ and $\mathcal{E}_2$ are uncorrelated with the dependent variable. These zero covariance restrictions can be implied by a lower-level assumption, such as $\mathrm{E}[U \mid X^*, \mathcal{E}_1, \mathcal{E}_2] = 0$, $\mathcal{E}_1 \perp\!\!\!\perp X^*$, and $\mathrm{E}[\mathcal{E}_2 \mid X^*] = 0$, which also imply the additional identifying restrictions presented later (Assumption 3). The second part of Assumption 1 is empirically testable with observed data, and implies a non-zero denominator in the identifying equation (2.2). We state this auxiliary result below for ease of reference.

Lemma 1 (Identification of γ_1)**.** *If Assumption 1 holds, then γ_1 is identified with (2.2).*

In some applications, we may simply assume $\gamma_1 = 1$ from the outset, and Assumption 1 need not be invoked. In any case, we hereafter assume that γ_1 is known either by assumption or by the identifying formula (2.2), and that γ_1 is different from zero.

Assumption 2 (Nonzero γ_1)**.** *$\gamma_1 \neq 0$.*

If this assumption fails, then the observed variable X_1 fails to be an informative signal of X^*. Assumption 2 therefore plays the role of letting X_1 be an effective proxy for the latent variable X^*. To complete our definition of the model (2.1), we impose the following independence restrictions.

Assumption 3 (Restrictions)**.** *(i) $E[U|X_1] = 0$. (ii) $\mathcal{E}_1 \perp\!\!\!\perp X^*$. (iii) $E[\mathcal{E}_2|X_1] = 0$.*

Part (i) states that the residual of the outcome equation is conditional mean independent of the first measurement. A stronger version of part (i) is the mean independence $E[U|X^*, \mathcal{E}_1] = 0$. Part (ii) states that the random error $\mathcal{E}_1$ in X_1 is independent of the true explanatory variable X^*. Notice that the coefficient γ_1 may not equal to one, and therefore the first measurement error defined as $X_1 - X^* = (\gamma_1 - 1)X^* + \mathcal{E}_1$ need not be classical, i.e., the measurement error is not independent of the true value X^*, even under part (ii) of the above assumption. This observation highlights one of the major advantages of our model compared to the existing models which impose $\gamma_1 = 1$. Part (iii) states that the second measurement error $\mathcal{E}_2$ is conditional mean independent of the first measurement X_1. This assumption is different from the classical measurement error assumption that $\mathcal{E}_2$ is independent of X^* and U. The last two parts, (ii) and (iii), can be succinctly implied by the frequently used assumption in the literature that X^*, $\mathcal{E}_1$, and $\mathcal{E}_2$ are mutually independent, but we state the above weaker assumptions for the sake of generality. Plausibility of these independence assumptions will be discussed in the context of a specific empirical application in Section 6.

Let $i = \sqrt{-1}$ denote the unit imaginary number. Define the marginal characteristic functions ϕ_{X_1}, ϕ_{X^*} and $\phi_{\mathcal{E}_1}$ by $\phi_{X_1}(t) = \mathrm{E}\, e^{itX_1}$, $\phi_{X^*}(t) = \mathrm{E}\, e^{itX^*}$ and $\phi_{\mathcal{E}_1}(t) = \mathrm{E}\, e^{it\mathcal{E}_1}$, respectively. Also define the joint characteristic functions $\phi_{X_1X_2}$ and ϕ_{X_1Y} by $\phi_{X_1X_2}(t_1, t_2) = \mathrm{E}\, e^{it_1X_1+it_2X_2}$ and $\phi_{X_1Y}(t_1, s) = \mathrm{E}\, e^{it_1X_1+isY}$, respectively. We let $\mathcal{F}$ denote the transformation defined by $\mathcal{F}f(t) = \int e^{itx} f(x)dx$. With this notation, we state the following assumption for identification of g.

Assumption 4 (Regularity)**.** *(i) ϕ_{X_1} does not vanish on the real line. (ii) f_{X^*} and $\mathcal{F}f_{X^*}$ are continuous and absolutely integrable. (iii) $f_{X^*} \cdot g$ and $\mathcal{F}(f_{X^*} \cdot g)$ are continuous and absolutely integrable.*

Under Assumptions 3 (ii) and 4 (i), the characteristic functions ϕ_{X^*} and $\phi_{\mathcal{E}_1}$ do not vanish on the real line either. This property of non-vanishing characteristic functions is shared by many of the common distribution families, e.g., the normal, chi-squared, Cauchy, gamma, and exponential distributions. In parts of our identifying formula, the characteristic functions appear as denominators, and hence this assumption to rule out zero denominator is crucial. Parts (ii) and (iii) ensure that we can apply the Fourier transform and inversion to those functions. Under this commonly invoked regularity condition together with the independence restrictions in Assumption 3, we can solve relevant integral equations explicitly to obtain the following closed-form identification result.

Theorem 1 (Closed-Form Identification for Affine Models of Endogenous Measurement)**.** *Suppose that Assumptions 1, 2, 3 and 4 hold for the model (2.1). The nonparametric function g evaluated at x^* in the interior of the support of X^* is identified with the closed-form solution:*

$$g(x^*) = \frac{\int_{-\infty}^{+\infty} e^{-itx^*} \ \exp\left(\int_0^t \frac{\left[\frac{\partial}{\partial t_2}\phi_{X_1X_2}(t_1/\gamma_1,t_2)\right]_{t_2=0}}{\phi_{X_1}(t_1/\gamma_1)} dt_1 \right) \ \frac{\left[\frac{\partial}{\partial s}\phi_{X_1Y}(t/\gamma_1,s)\right]_{s=0}}{i \ \phi_{X_1}(t/\gamma_1)} \ dt}{\int_{-\infty}^{+\infty} e^{-itx^*} \exp\left(\int_0^t \frac{\left[\frac{\partial}{\partial t_2}\phi_{X_1X_2}(t_1/\gamma_1,t_2)\right]_{t_2=0}}{\phi_{X_1}(t_1/\gamma_1)} dt_1 \right) dt}, \tag{2.3}$$

where the parameter γ_1 is identified with the closed-form solution (2.2).

A proof is given in Section A.1 in the appendix. Note that every component on the right-hand side of the identifying formula (2.3) is computable directly as a moment of observed data. Replacing the population moments by the corresponding sample moments therefore yields a closed-form estimator of $g(x^*)$.

3 Closed-Form Identification: General Models

In this section, we consider the following generalized extension to the baseline model (2.1):

$$\begin{aligned} Y &= g(X^*) + U, & \mathrm{E}[U \mid X^*] = 0 \\ X_1 &= \sum_{p=1}^{P} \gamma_p X^{*p} + \mathcal{E}_1 & \mathrm{E}[\mathcal{E}_1] = \gamma_0 \\ X_2 &= X^* + \mathcal{E}_2, & \mathrm{E}[\mathcal{E}_2] = 0 \end{aligned} \tag{3.1}$$

where we observe the joint distribution of (Y, X_1, X_2). The first measurement X_1 is systematically biased with an arbitrarily high order of nonlinearity. We demonstrate that a similar closed-form identification result can be obtained for this extended model. To this goal, we impose the following independence restrictions on (3.1).

Assumption 5 (Restrictions for the General Polynomial Model)**.** *(i) $E[U \mid X^*, \mathcal{E}_1, \mathcal{E}_2] = 0$. (ii) $X^* \perp\!\!\!\perp \mathcal{E}_1$. (iii) $(X^*, \mathcal{E}_1) \perp\!\!\!\perp \mathcal{E}_2$.*

Parts (i)–(iii) of this assumption are analogous to the corresponding parts in Assumption 3. We remark that parts (i) and (iii) are stronger than those corresponding parts in Assumption 3, and that we can deal with the higher-order measurement model (3.1) at the cost of this strengthening of the independence assumption. A preliminary step before the closed-form identification of $g(X^*)$ involves identification of the polynomial coefficients $\gamma_0, \cdots, \gamma_P$ and the moments of $\mathcal{E}_2$ up to the P-th order. This preliminary step is presented in Section 3.1. After the preliminary step, we then proceed with closed-form identification of the nonparametric regression function g in Section 3.2.

3.1 A Preliminary Step: Identification of γ_p and $\mathrm{E}[\mathcal{E}_2^p]$

As is the case for the simple affine model of endogenous measurement presented in Section 2 (see (2.2) and Lemma 1), identification of the parameters γ_p and $\sigma_2^p := \mathrm{E}[\mathcal{E}_2^p]$ for the model (3.1) also follows from an appropriate set of moment restrictions. To form such restrictions, one can propose several alternative statistical and mean independence assumptions, and there is not the unique set of identifying restrictions to this goal. One might therefore want to come up with the most convenient set of restriction tailored to specific empirical applications. As a general prescription, we can form restrictions of the form

$$\begin{aligned}
\mathrm{Cov}(YX_2^q, X_1) &= \mathrm{E}\left[Y(X^* + \mathcal{E}_2)^q \left(\sum_{p=1}^{P} \gamma_p X^{*p} + \mathcal{E}_1\right)\right] - \mathrm{E}\left[Y(X^* + \mathcal{E}_2)^q\right] \mathrm{E}\left[\sum_{p=1}^{P} \gamma_p X^{*p} + \mathcal{E}_1\right] \\
&= \sum_{p=0}^{P} \sum_{q'=0}^{q} \gamma_p \sigma_2^{q-q'} \binom{q}{q'} \left(\mathrm{E}[YX^{*(p+q')}] - \mathrm{E}[YX^{*q'}]\,\mathrm{E}[X^{*p}]\right)
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}(YX_2^r, X_2^s) &= \mathrm{E}[Y(X^* + \mathcal{E}_2)^{r+s}] - \mathrm{E}[Y(X^* + \mathcal{E}_2)^r]\,\mathrm{E}[(X^* + \mathcal{E}_2)^s] \\
&= \sum_{r'=0}^{r+s} \sigma_2^{r+s-r'} \binom{r+s}{r'} \mathrm{E}[YX^{*r'}] - \sum_{r'=0}^{r} \sum_{s'=0}^{s} \sigma_2^{r+s-r'-s'} \binom{r}{r'} \binom{s}{s'} \mathrm{E}[YX^{*r'}]\,\mathrm{E}[X^{*s'}]
\end{aligned}$$

for various $q = 0, 1, \cdots, Q - P$, $r = 0, 1, \cdots$ and $s = 1, \cdots$ such that $r + s \leqslant Q$ for some $Q \in \mathbb{N}$. The right-hand sides of the above two equations involve the unknowns, $(\gamma_0, \cdots, \gamma_P)$, $(\sigma_2^2, \cdots, \sigma_2^Q)$, $(\mathrm{E}[X^*], \cdots, \mathrm{E}[X^{*Q}])$, and $(\mathrm{E}[YX^*], \cdots, \mathrm{E}[YX^{*Q}])$, under Assumption 5. As such, we obtain $(Q-P+1)+\frac{Q(Q+1)}{2}$ restrictions for $3Q+P$ unknown parameters, $(\gamma_0,\cdots,\gamma_P)$, $(\sigma_2^2,\cdots,\sigma_2^Q)$, $(\mathrm{E}[X^*],\cdots,\mathrm{E}[X^{*Q}])$, and $(\mathrm{E}[YX^*],\cdots,\mathrm{E}[YX^{*Q}])$. Clearly for any given order P of polynomial, as we increase the number Q, we have sufficiently more number of restrictions than the unknowns to recover the polynomial coefficients $\gamma_0,\cdots,\gamma_P$ and the moments $\sigma_2^2,\cdots,\sigma_2^P$ which we need.

A drawback to the above general prescription is that these moment restrictions may not necessarily lead to a closed-form solution to these parameters. One can make alternative statistical and mean independence assumptions for the goal of obtaining closed-form identification of the polynomial coefficients $\gamma_0,\cdots,\gamma_P$ and the moments $\sigma_2^2,\cdots,\sigma_2^P$. For example, we may show a closed-from solution in the quadratic case, where the endogenous measurement X_1 is modeled with $P=2$ by

$$X_1=\gamma_1 X^*+\gamma_2 X^{*2}+\mathcal{E}_1 \;:\qquad \mathrm{E}[\mathcal{E}_1]=\gamma_0 \tag{3.2}$$

If we assume the homoscedasticity $\mathrm{E}[U^2\mid X^*,\mathcal{E}_1,\mathcal{E}_2]=\mathrm{E}[U^2]$ and the empirically testable rank condition $\mathrm{Cov}(Y,X_2)\cdot\mathrm{Cov}(Y^2,X_2^2)\neq\mathrm{Cov}(Y,X_2^2)\cdot\mathrm{Cov}(Y^2,X_2)$, then we can show that the coefficients γ_1 and γ_2 of the model (3.2) are identified with the closed-form solutions

$$\begin{aligned}\gamma_1 &= \frac{\mathrm{Cov}(Y,X_1)\cdot\mathrm{Cov}(Y^2,X_2^2)-\mathrm{Cov}(Y,X_2^2)\cdot\mathrm{Cov}(Y^2,X_1)}{\mathrm{Cov}(Y,X_2)\cdot\mathrm{Cov}(Y^2,X_2^2)-\mathrm{Cov}(Y,X_2^2)\cdot\mathrm{Cov}(Y^2,X_2)} \qquad \text{and}\\ \gamma_2 &= \frac{\mathrm{Cov}(Y,X_2)\cdot\mathrm{Cov}(Y^2,X_1)-\mathrm{Cov}(Y,X_1)\cdot\mathrm{Cov}(Y^2,X_2)}{\mathrm{Cov}(Y,X_2)\cdot\mathrm{Cov}(Y^2,X_2^2)-\mathrm{Cov}(Y,X_2^2)\cdot\mathrm{Cov}(Y^2,X_2)}.\end{aligned}$$

Furthermore, Assumption 5 also allows us to identify γ_0 and σ_2^2 with the closed-form solution

$$\begin{bmatrix}\gamma_0\\ \sigma_2^2\\ \sigma_2^3\end{bmatrix}=\begin{bmatrix}\mathrm{E}[Y] & -\gamma_2\,\mathrm{E}[Y] & 0\\ \mathrm{E}[X_2] & -\gamma_1-3\gamma_2\,\mathrm{E}[X_2] & -\gamma_2\\ \mathrm{E}[YX_2] & -\gamma_1\,\mathrm{E}[Y]-3\gamma_2\,\mathrm{E}[YX_2] & -\gamma_2\,\mathrm{E}[Y]\end{bmatrix}^{-1}\begin{bmatrix}\mathrm{E}[YX_1]\\ \mathrm{E}[X_1X_2]\\ \mathrm{E}[YX_1X_2]\end{bmatrix},$$

provided the nonsingularity of the inverted matrix. Detailed derivations of these closed-form identifying formulas can be found in Section A.2 in the appendix.

3.2 Identification of Nonparametric Regression g

With the polynomial coefficients $(\gamma_1, \cdots, \gamma_P)$ and the moments $(\sigma_2^2, \cdots, \sigma_2^P)$ for the model (3.1) identified with the methods outlined in Section 3.1, we proceed with closed-form identification of the nonparametric regression function g evaluated at various points x^* in the interior of the support of X^*. To this end, we assume the following rank condition, which is effectively an empirically testable assumption as $(\sigma_2^2, \cdots, \sigma_2^P)$ are identified from observed data $F_{YX_1X_2}$.

Assumption 6 (Empirically Testable Rank Condition)**.** *The following matrix is nonsingular.*

$$\begin{bmatrix} 1 & \binom{P}{P-1}\sigma_2^1 & \cdots & \binom{P}{2}\sigma_2^{P-2} & \binom{P}{1}\sigma_2^{P-1} \\ & 1 & \cdots & \binom{P-1}{2}\sigma_2^{P-3} & \binom{P-1}{1}\sigma_2^{P-2} \\ & & \ddots & \vdots & \vdots \\ & & & 1 & \binom{2}{1}\sigma_2^1 \\ & & & & 1 \end{bmatrix}_{P\times P}$$

Besides its empirical testability, this rank condition is automatically satisfied for the linear case ($P = 1$) and the quadratic case ($P = 2$) due to the normalization $\mathrm{E}[\mathcal{E}_2] = 0$ in (3.1).[3] For convenience of writing, we let Z^* denote the random variable $\sum_{p=1}^P \gamma_p X^{*p}$. The role of Assumption 6 is to identify the distribution of this generated latent variable Z^* in the following manner. Under Assumption 6, we can write the following vector on the left-hand side in terms of the expression on the right-hand side that consists of observed data.

$$\begin{bmatrix} \mu(t,P;\sigma_2^1,\cdots,\sigma_2^P;F_{X_1X_2}) & \cdots & \mu(t,1;\sigma_2^1,\cdots,\sigma_2^P;F_{X_1X_2}) \end{bmatrix}' :=$$
$$\begin{bmatrix} 1 & \binom{P}{P-1}\sigma_2^1 & \cdots & \binom{P}{2}\sigma_2^{P-2} & \binom{P}{1}\sigma_2^{P-1} \\ & 1 & \cdots & \binom{P-1}{2}\sigma_2^{P-3} & \binom{P-1}{1}\sigma_2^{P-2} \\ & & \ddots & \vdots & \vdots \\ & & & 1 & \binom{2}{1}\sigma_2^1 \\ & & & & 1 \end{bmatrix}^{-1} \begin{bmatrix} \mathrm{E}[(X_2^P - \sigma_2^P)e^{itX_1}] \\ \mathrm{E}[(X_2^{P-1} - \sigma_2^{P-1})e^{itX_1}] \\ \vdots \\ \mathrm{E}[(X_2^2 - \sigma_2^2)e^{itX_1}] \\ \mathrm{E}[(X_2 - \sigma_2^1)e^{itX_1}] \end{bmatrix} \tag{3.3}$$

[3]However, when the order of polynomial is $P = 3$ or above, this rank condition can be shown to be unsatisfied, e.g., one can check that $\sigma_2^2 = \frac{1}{3}$ when $P = 3$ fails the assumption.

It is shown in the theorem below that this vector is sufficient to pin down the distribution of the generated latent variable $Z^* = \sum_{p=1}^{P} \gamma_p X^{*p}$, and hence its distribution (equivalently, its characteristic function) can be identified from observed data.

To make use of this auxiliary result to identify the nonparametric regression function g of interest, we next propose the following regularity conditions.

Assumption 7 (Regularity). *(i) ϕ_{X_1} and ϕ_{X_2} do not vanish on the real line. (ii) f_{X^*} and $\mathcal{F} f_{X^*}$ are continuous and absolutely integrable. (iii) f_{Z^*} and $\mathcal{F} f_{Z^*}$ are continuous and absolutely integrable. (iv) $f_{X^*} \cdot g$ and $\mathcal{F}(f_{X^*} \cdot g)$ are continuous and absolutely integrable.*

This assumption plays a similar role to Assumption 4. In parts of our identifying formula, the characteristic functions appear as denominators, and hence part (i) of this assumption rules out zero denominator. This property of non-vanishing characteristic functions is shared by many of the common distribution families, e.g., the normal, chi-squared, Cauchy, gamma, and exponential distributions. Parts (ii) and (iii) ensure that we can apply the Fourier transform and inversion to those functions. The model allows for nonlinear and endogenous errors in the sense of $\mathrm{E}[X_1 \mid X^*] = \sum_{p=0}^{P} \gamma_p X^{*p}$. However, we rule out the case where the report X_1 is decreasing while the truth X^* is increasing. Specifically, we assume the following monotonicity restriction.

Assumption 8 (Monotonicity). *$\sum_{p=0}^{P} \gamma_p x^p$ is non-decreasing in x on the support of X^*.*

This monotonicity assumption is used for the purpose of applying the density transformation formula to derive the density function for the transformed random variable. Polynomial functions do not generally exhibit monotonicity on the entire real line. Note that Assumption 8 only requires the monotonicity to hold on the support of X^*, and hence is not restrictive when the support of X^* is a proper subset of $\mathbb{R}$. For example, many economic variables X^* are innately positive, i.e., $\mathrm{supp}(X^*) \subseteq \mathbb{R}_+$, and the quadratic function $\mathrm{E}[X_1 \mid X^*] = \gamma_2 X^{*2}$, for example, necessarily satisfies Assumption 8 for such variables.

With this set of assumptions, we can still identify the nonparametric function g with a closed-form formula, even if the measurement X_1 is systematically biased with endogeneity and such a high order of nonlinearity. The following theorem states the exact result.

Theorem 2 (Closed-Form Identification for High Order Models of Endogenous Measurement)**.** *Suppose that Assumptions 5, 6, 7 and 8 hold for the model (3.1). The nonparametric function g evaluated at x^* in the interior of the support of X^* is identified with the closed-form solution:*

$$g(x^*) = \frac{\int\int\int e^{-itx^*+itx-it'\left(\sum_{p=1}^P \gamma_p x^p\right)} \left|\sum_{p=1}^P p\gamma_p x^{p-1}\right| \frac{E[Ye^{itX_2}]}{E[e^{itX_2}]}\phi_{Z^*}(t')dt'dxdt}{2\pi\int e^{-ith\left(\sum_{p=1}^P \gamma_p x^{*p}\right)} \left|\sum_{p=1}^P p\gamma_p x^{*(p-1)}\right| \phi_{Z^*}(t)dt},$$

where ϕ_{Z^*} is identified with the closed-form solution

$$\phi_{Z^*}(t) = \exp\left\{\int_0^t \frac{\sum_{p=1}^P \gamma_p \mu(t_1, p\ ; \sigma_2^1, \cdots, \sigma_2^P; F_{X_1X_2})}{E[e^{it_1X_1}]}dt_1\right\}$$

and $\mu(t, p\ ; \sigma_2^1, \cdots, \sigma_2^P; F_{X_1X_2})$ for all $p = 1, \cdots, P$ are given by the closed-form solution (3.3).

A proof is given in Section A.3 in the appendix. Note that this general version of the closed-form identifying formula, involving the triple integral instead of a single integral due to the nonlinear transformation, is qualitatively quite different from the traditional formulas including the one in Theorem 1 as well as that of Schennach (2004b). Theorem 1 may appear to be a special case of this theorem, as the former focuses on affine models and the latter extends to higher order polynomials. Strictly speaking, it is not a special case, because Theorem 1 requires slightly weaker independence assumptions than Theorem 2. As such, we stated Theorem 1 separately in the previous section for the practical importance of parsimonious affine models. Section A.2 in the appendix illustrates how the closed-form identifying formula looks like in the case of quadratic model of measurement, $P = 2$, as an example.

4 Closed-Form Estimator

Given the closed-form identifying formulas of Theorems 1 and 2, one can easily construct a direct sample-counterpart estimator by replacing the population moments by the sample moments for the characteristic functions. As this basic idea is the same across all the cases, we focus on the simplest model (2.1) for simplicity in this section. If γ_1 is known, then the sample-counterpart estimator $\widehat{g(x^*)}$ of the closed-form identifying formula (2.3) is given by

$$\widehat{g(x^*)} = \frac{\int_{-\infty}^{+\infty} e^{-itx^*} \exp\left(i \int_0^t \frac{\sum_{j=1}^n X_{2,j} e^{it_1 X_{1,j}/\gamma_1}}{\sum_{j=1}^n e^{it_1 X_{1,j}/\gamma_1}} dt_1 \right) \frac{\sum_{j=1}^n Y_j e^{itX_{1,j}/\gamma_1}}{\sum_{j=1}^n e^{itX_{1,j}/\gamma_1}} \phi_K(th) dt}{\int_{-\infty}^{+\infty} e^{-itx^*} \exp\left(i \int_0^t \frac{\sum_{j=1}^n X_{2,j} e^{it_1 X_{1,j}/\gamma_1}}{\sum_{j=1}^n e^{it_1 X_{1,j}/\gamma_1}} dt_1 \right) \phi_K(th) dt} \tag{4.1}$$

where ϕ_K denotes the Fourier transform of a kernel function K which we use together with the tuning parameter h for the purpose of regularization.

On the other hand, if γ_1 is not known, we replace γ_1 by its estimate and the estimator thus takes the form

$$\widehat{g(x^*)} = \frac{\int_{-\infty}^{+\infty} e^{-itx^*} \exp\left(i \int_0^t \frac{\sum_{j=1}^n X_{2,j} e^{it_1 X_{1,j}/\hat{\gamma}_1}}{\sum_{j=1}^n e^{it_1 X_{1,j}/\hat{\gamma}_1}} dt_1 \right) \frac{\sum_{j=1}^n Y_j e^{itX_{1,j}/\hat{\gamma}_1}}{\sum_{j=1}^n e^{itX_{1,j}/\hat{\gamma}_1}} \phi_K(th) dt}{\int_{-\infty}^{+\infty} e^{-itx^*} \exp\left(i \int_0^t \frac{\sum_{j=1}^n X_{2,j} e^{it_1 X_{1,j}/\hat{\gamma}_1}}{\sum_{j=1}^n e^{it_1 X_{1,j}/\hat{\gamma}_1}} dt_1 \right) \phi_K(th) dt} \tag{4.2}$$

where $\hat{\gamma}_1$ is computed by the following sample-counterpart of (2.2).

$$\hat{\gamma}_1 = \frac{\frac{1}{n}\sum_{j=1}^n Y_j X_{1,j} - \left(\frac{1}{n}\sum_{j=1}^n Y_j\right)\left(\frac{1}{n}\sum_{j=1}^n X_{1,j}\right)}{\frac{1}{n}\sum_{j=1}^n Y_j X_{2,j} - \left(\frac{1}{n}\sum_{j=1}^n Y_j\right)\left(\frac{1}{n}\sum_{j=1}^n X_{2,j}\right)}.$$

It turns out that the substitution of the estimate $\hat{\gamma}_1$ for the true value of γ_1 does not affect the asymptotic property of $\widehat{g(x^*)}$. We assume the following basic regularity conditions to derive the consistency of $\widehat{g(x^*)}$ in both (4.1) and (4.2).

Assumption 9 (Basic Assumptions for Consistency). *(i)* $\{X^*, \mathcal{E}_1, \mathcal{E}_2, U\}$ *is independently and identically distributed. (ii)* ϕ_K *is symmetric, satisfies* $\phi_K(0) = 1$, *and has integrable second derivatives. (iii)* $E|X_1|^{2+\delta} < \infty$, $E|X_2|^{2+\delta} < \infty$, *and* $E|Y|^{2+\delta} < \infty$ *for some* $\delta > 0$.

In case of using the version (4.2) of the closed-form estimator instead of (4.1), we assume the following bounded fourth moment restriction in addition to part (iii) of Assumption 9.

Assumption 9. *(iii)′* $E|X_1|^4 < \infty$, $E|X_2|^4 < \infty$, *and* $E|Y|^4 < \infty$.

The asymptotic rate of convergence of the closed-form estimators (4.1) and (4.2) depend on the

Hölder exponents of the nonparametric density f_{X^*} and the nonparametric regression g. We therefore introduce the following assumption with index numbers that determine the asymptotic orders.

Assumption 10 (Determinants of the Asymptotic Orders of Biases)**.** *(i) f_{X^*} is twice continuously differentiable at x^*, and the k_1-th derivative of f_{X^*} is k_2-Hölder continuous with Hölder constant bounded by k_0, i.e., $\left|f_{X^*}^{(k_1)}(x) - f_{X^*}^{(k_1)}(x+\delta)\right| \leqslant k_0 \left|\delta\right|^{k_2}$ for all x, δ. (ii) g is twice continuously differentiable at x^*, and the l_1-th derivative of g is l_2-Hölder continuous with Hölder constant bounded by l_0, i.e., $\left|g^{(l_1)}(x) - g^{(l_1)}(x+\delta)\right| \leqslant l_0 \left|\delta\right|^{l_2}$ for all x, δ. Let $k = k_1 + k_2$ and $l = l_1 + l_2$ be the largest numbers satisfying the above properties.*

Since optimal choices of the bandwidth parameter h depend on the shape of the underlying characteristic function, we first state the following auxiliary result of convergence rate under free choice of h.

Lemma 2 (Mean Square Error of the Closed-Form Estimator)**.** *Suppose that Assumptions 2, 3 and 4 hold for the model (2.1). If Assumptions 9 and 10 are satisfied and x^* is in the interior of the support of X^*, then, with any choice of h such that $h \to 0$ and $nh^4 \left|\phi_{X_1}(1/h)\right|^4 \to \infty$ as $n \to \infty$, the mean square error of the closed-form estimator $\widehat{g(x^*)}$ given in (4.1) has the asymptotic order:*

$$\mathcal{O}(h^{2\min\{k,l\}}) + \mathcal{O}\left(\frac{1}{nh^4 \left|\phi_{X_1}(1/h)\right|^4}\right), \tag{4.3}$$

where the first and second terms correspond to the asymptotic orders of the squared bias and the variance, respectively. The same conclusion holds for the closed-form estimator $\widehat{g(x^*)}$ given in (4.2), provided that Assumptions 1, and 9 (iii)$'$ additionally hold.

This lemma implies that the MSE-optimizing choice of h obviously depends on the tail behavior of the characteristic function ϕ_{X_1}, which in turn depends on the characteristic functions ϕ_{X^*} and $\phi_{\mathcal{E}_1}$. Therefore, we branch into the following two cases: (a) at least one of X^* and $\mathcal{E}_1$

has a super-smooth distribution; and (b) both X^* and $\mathcal{E}_1$ have ordinary-smooth distributions. These two cases are precisely stated in the following two separate assumptions.

Assumption 11 (Super-Smooth Distributions)**.** *Assume that (i) the distribution of X^* is super-smooth of order $\beta_1 > 0$, i.e., there exist $\kappa_1 > 0$ such that $|\phi_{X^*}(t)| = \mathcal{O}\left(e^{-|t|^{\beta_1}/\kappa_1}\right)$ as $t \to \pm\infty$, or (ii) the distribution of $\mathcal{E}_1$ is super-smooth of order $\beta_2 > 0$, i.e., there exist $\kappa_2 > 0$ such that $|\phi_{\mathcal{E}_1}(t)| = \mathcal{O}\left(e^{-|t|^{\beta_2}/\kappa_2}\right)$ as $t \to \pm\infty$, OR both (i) and (ii) hold. For convenience of notation, we let $\beta_1 = 0$ (respectively, $\beta_2 = 0$) if the distribution of X^* (respectively, $\mathcal{E}_1$) is not super-smooth.*

Assumption 12 (Ordinary-Smooth Distributions)**.** *Assume that (i) the distribution of X^* is ordinary-smooth of order β_1, i.e., $|\phi_{X^*}(t)| = \mathcal{O}\left(|t|^{-\beta_1}\right)$ as $t \to \pm\infty$, and (ii) the distribution of $\mathcal{E}_1$ is ordinary-smooth of order β_2, i.e., $|\phi_{\mathcal{E}_1}(t)| = \mathcal{O}\left(|t|^{-\beta_2}\right)$ as $t \to \pm\infty$.*

These two smoothness definitions characterized by the tail behavior of the characteristic functions measure the smoothness of the density function. Examples of super-smooth distributions include the normal, Cauchy, and mixed normal distributions. Examples of ordinary-smooth distributions include the gamma, exponential, and uniform distributions. If at least one of X^* and $\mathcal{E}_1$ has a super-smooth distribution in the sense of Assumption 11, then the closed-form estimators follow $\log n$ rates of convergence as follows.

Theorem 3 (Consistency of the Closed-Form Estimator under Super-Smooth Distribution(s))**.** *Suppose that Assumptions 2, 3 and 4 hold for the model (2.1). If Assumptions 9, 10, and 11 are satisfied and x^* is in the interior of the support of X^*, then the closed-form estimator $\widehat{g(x^*)}$ given in (4.1) is consistent with the convergence rate $\left(E\left[\widehat{g(x^*)} - g(x^*)\right]^2\right)^{1/2} = \mathcal{O}\left((\log n)^{\frac{-\min\{k,l\}}{\max\{\beta_1,\beta_2\}}}\right)$ under the choice of the tuning parameter $h \propto (\log n)^{-1/\max\{\beta_1,\beta_2\}}$. The same conclusion holds for the closed-form estimator $\widehat{g(x^*)}$ given in (4.2), provided that Assumptions 1, and 9 (iii)$'$ additionally hold.*

On the other hand, if both X^* and $\mathcal{E}_1$ have ordinary-smooth distributions in the sense of Assumption 12, then the closed-form estimator follow polynomial rates of convergence as follows.

Theorem 4 (Consistency of the Closed-Form Estimator under Ordinary-Smooth Distributions)**.** *Suppose that Assumptions 2, 3 and 4 hold for the model (2.1). If Assumptions 9, 10, and 12 are satisfied and x^* is in the interior of the support of X^*, then the closed-form estimator $\widehat{g(x^*)}$ given in (4.1) is consistent with the convergence rate $\left(E\left[\widehat{g(x^*)} - g(x^*)\right]^2\right)^{1/2} = \mathcal{O}\left(n^{\frac{-\min\{k,l\}}{2(\min\{k,l\}+2(\beta_1+\beta_2+1))}}\right)$ under the choice of the tuning parameter $h \propto n^{-1/2(\min\{k,l\}+2(\beta_1+\beta_2+1))}$. The same conclusion holds for the closed-form estimator $\widehat{g(x^*)}$ given in (4.2), provided that Assumptions 1, and 9 (iii)$'$ additionally hold.*

While the contexts and the setups are different and a direct comparison cannot be made, the two cases covered in our Theorems 3 and 4 can be connected to Cases 2 and 4 of Theorem 2 in Schennach (2004b), respectively.[4] The slow convergence rates in the case of the super-smooth distributions could be improved in theory provided that the mean regression $g(X^*)$ is also super-smooth. However, this improvement requires an infinite order kernel that vanishes the bias faster than any power of the bandwidth parameter, and it may suffer from problems of near zero denominators in practical implementation in finite sample. See Schennach (2004b) for discussions.

5 Monte Carlo Simulations

In this section, we use Monte Carlo simulations to assess the small sample performance of the estimator (4.2) proposed in the previous section.

Each set of simulations constructs data of size $N = 500$ by the following distributional

[4]Specifically, the auxiliary parameters β_v, γ_b and γ_v used in Schennach (2004b) can be reconciled with our regularity parameters through the relations $\beta_v = \max\{\beta_1, \beta_2\}$, $\gamma_b = -\min\{k,l\}$ and $\gamma_v = 2(\beta_1 + \beta_2 + 1)$.

model for the primitives.

$$X^* \sim N(0, 2^2), \qquad \begin{cases} \mathcal{E}_1 \sim N(0, 1^2) \\ \mathcal{E}_2 \sim N(0, 1^2) \end{cases}, \qquad U \sim N(0, 1^2).$$

These four random variables are generated mutually independently. The true X^* has a twice as large variation ($\sigma = 2$) as the noises $\mathcal{E}_1$ and $\mathcal{E}_2$ ($\sigma = 1$). These four latent variables in turn generate the observed random variables, X_1, X_2, and Y through the model (2.1), given a definition of the nonparametric regression function g as well as the coefficients γ_p. We set $\gamma_0 = 0$ and $\gamma_1 = 2$ for the model of endogenous measurement X_1, but the choice of these coefficients does not alter simulation results much unless γ_1 is set arbitrarily close to zero. Notice that this data generating process, with the super-smooth Gaussian distributions, is a worse case scenario in terms of the asymptotic convergence rate (cf. Theorems 3 and 4). In other words, we are not cherry-picking convenient Monte Carlo settings.

Consider the following four function specifications. (i) $g(x^*) = x^*$; (ii) $g(x^*) = (x^* + 1)^2$; (iii) $g(x^*) = \Phi(x^*)$ where Φ is the standard normal cdf; and (iv) $g(x^*) = \sin(x^*)$. For the purpose of checking robustness of the nonparametric closed-form estimator, this list contains two broad classes of functions. The first two functions are polynomial functions, and the latter two functions are transcendental functions. We emphasize that truncated polynomial approximations would not work precisely for the latter class.

We ran 1,000 Monte Carlo iterations for each of the above four function specifications. Figure 1 shows simulation results for the closed-form estimator (4.2), which uses estimates $\hat{\gamma}_1$ for the unknown parameter γ_1. The solid curves represent the true function g. The dashed curves are the 10th, 30th, 50th, 70th, and 90th percentiles of the Monte Carlo distributions. The Monte Carlo quantiles capture the true function in each of the four cases displayed in the figure. Recall that we use only $N = 500$ observations in the simulations. With this very small sample size for nonparametrics, the Monte Carlo distributions are fairly tight with our closed-form estimator (4.2).[5]

For the purpose of comparison, we also ran Monte Carlo iterations with the same setup, but by using the naive version of the closed-form estimator (4.1), where we *wrongly* set $\gamma_1 = 1$ as is the case for the existing methods in the literature that assume classical errors. Figure 2 shows simulation results with this classical error assumption. Unlike the previous results in Figure 1, the Monte Carlo quantiles in Figure 2 fail to capture the true functions well. This failure is particularly the case for the (ii) quadratic and (iv) sine functions, for which the MC quantiles tend to be biased outward from the swinging curves. Even in the absence of curves, the widely spread MC quantiles evidence that this estimator wrongly assuming classical errors suffer from greater variances. Therefore, the estimator assuming classical errors performs poorly both in terms of bias and variance, compared to our estimator (4.2) which allows for non-classical measurement errors.

6 Empirical Illustration: BMI and Health Care Usage

A recently growing body of the health economic literature contains extensive studies of economic causes and economic implications of obesity, including but not limited to the following list of papers. Cohen-Cole and Flecther (2008) and Trogdon et al (2008) study causal and propagation mechanisms of obesity. Cawley, Frisvold, and Meyerthoefer (2013) evaluate preventive programs for obesity. Cawley (2004), Bhattacharya and Bundorf (2009), and Cawley and Maclean (2012) analyze labor and health market implications of obesity. The social cost structure of obesity and its policy implications are discussed by Bhattacharya and Sood (2011) and Cawley and Meyerhoefer (2012). While it should not be regarded as a medical diagnosis, the Body Mass

[5] The results are reasonably robust across alternative values of bandwidth parameters. We refer the readers to Diggle and Hall (1993) for discussions about the choice of tuning parameters for deconvolution estimators based on Fourier transformation.

Index (BMI) is widely used as a measure of obesity. It is defined by the following formula.

$$\text{BMI (kg/m}^2) \;=\; \text{Mass (kg)}/\left(\text{Height (m)}\right)^2$$

This index, or indicators of obesity generated by this index, is used in each of the above list of empirical research papers as well as many others.

Survey data often contain necessary variables to compute the BMI, namely weight and height, but they are usually based on self reports. How accurate are the BMIs constructed by the self-reported body measures? To answer this question, we use National Health and Nutrition Examination Survey (NHANES III) of Center for Disease Control and Prevention, which uniquely combines survey responses and various results of medical examination and laboratory tests. Table 1 shows a summary of variables that we extracted from this source. Using this data set, we can match self-reported body measures and clinically measured body measures. Figure 3 shows a scatter plot of clinically measured BMI on the horizontal axis against self-reported BMI on the vertical axis. It evidences a nontrivial discrepancy between the two measures, showing that self-reports, clinical measures, or both of them have errors.

In this paper, we have shown that the mean regression $g(X^*)$ of an outcome variable Y on the true BMI X^* can be explicitly identified and explicitly estimated using two observed measures of the unobserved truth X^*, where one measurement can be endogenously biased. Applying this econometric method, we analyze how obesity measured by the BMI explains health care usage, taking into account the likely possibility that self reports may be endogenously biased. Specifically, the following list of variables are used for the baseline model (2.1).

$$\begin{aligned}
Y &= \text{Health Care Received (Observed Explained Variable)} \\
X^* &= \text{True BMI (Unobserved Explanatory Variable)} \\
X_1 &= \text{Self-Reported BMI} & \mathcal{E}_1 &= \text{Reporting Error} \\
X_2 &= \text{Clinically Measured BMI} & \mathcal{E}_2 &= \text{Clinical Measurement Error}
\end{aligned}$$

Our model setup requires that the clinical measurement X_2 is location-/scale-normalized with respect to the truth, which is plausible if clinical measurements have only random additive noises with mean zero. On the other hand, the self-reporting can be endogenously biased with arbitrary location and scale, because neither $\gamma_1 = 1$ nor $\gamma_0 = \mathrm{E}[\mathcal{E}_1] = 0$ is assumed. If $\gamma_1 < 1$, for example, then we can accommodate the likely case in which individuals under-report their BMI by $100 \times (1 - \gamma_1)$ percent on average even if $\gamma_0 = \mathrm{E}[\mathcal{E}_1] = 0$ is the case. Similarly, if $\gamma_0 = E[\mathcal{E}_1] < 0$, for example, then we can accommodate the likely case in which individuals under-report their BMI by $|\gamma_0| = |E[\mathcal{E}_1]|$ on average even if $\gamma_1 = 1$ is the case.

We have shown that these parameters γ_0 and γ_1 for the self-reporting model are also identified as a byproduct of our main identification result. Using the estimates, $\hat{\gamma}_0$ and $\hat{\gamma}_1$, we graphically illustrate the mean self-reporting behaviors $\mathrm{E}[X_1 \mid X^*]$ across various gender and age groups. Figure A.6 shows the results, where the dashed lines indicate the 45° line, and the solid lines indicate the estimated regression $\widehat{\mathrm{E}}[X_1 \mid X^*] = \hat{\gamma}_0 + \hat{\gamma}_1 X^*$ of self-reporting patterns. Not surprisingly, these results imply the tendency that actual overweight is associated with under-reporting for all the groups. Formally, we reject the null hypothesis that $\gamma_1 = 1$ for male 50s (at the 10% level) and female 70 or above (at the 5% level). For the entire sample, it is rejected at the 1% level. These results imply that the traditional classical error assumption $\gamma_1 = 1$ is not necessarily innocuous in practice, and hence our estimator proves more relevant to the current empirical problem than the existing closed-form estimators based on classical errors.

Nonparametric estimates of the mean regressions of the health care usage Y with respect to the true BMI X^* are computed. To get a sense of the effects of random sampling, we ran 1,000 bootstrap iterations for each age group for each gender. Figure 5 (respectively, 6) shows 10, 30, 50, 70, and 90-th percentiles of the bootstrap distributions of the estimates based on (4.2) using estimates $\hat{\gamma}_1$ of the unknown parameter γ_1 for male (respectively, female) individuals. The four graphs in each figure illustate results for four age groups. All the curves, except the one for male individuals aged 70 or above, show robust upward-sloping tendency of the mean health care usage with respect to the true BMI. These slopes are steeper particularly for females. Overall, obesity measured by the BMI is a positive explanatory factor for the health care usage, controlling for gender and age groups.

7 Conclusions

This paper provides a closed-form estimator of nonparametric regression models using two measurements with non-classical errors. We allow endogenous biases with arbitrary location and scale for one of two measurements, while the other is location-/scale-normalized with respect to the truth. Two distinct specifications for the models of the two measurements, X_1 and X_2, may be suitable for the common practical setting where two measurements are combined together from different data sources. Because of its closed form like the OLS, our estimator is easily implementable by practitioners. Monte Carlo simulations suggest that the estimator performs well even with a small sample size like $N = 500$. For an illustration, we investigate how obesity explains health care usage by using NHANES III that uniquely match clinical measurement and self-reports of the BMI. While the former measurement is assumed to be location-/scale-normalized with respect to the true BMI, the self-reports are allowed to be endogenously biased. We find robust upward sloping patterns for the health care usage with respect to obesity controlling for gender and age groups. These slopes are steeper especially for females.

References

Bhattacharya, J. and M.K. Bundorf, 2009. The Incidence of the Healthcare Costs of Obesity. Journal of Health Economics 28, 649-658.

Bhattacharya, J. and N. Sood, 2011. Who Pays for Obesity. Journal of Economic Perspectives 25, 139-157.

Bound, J., C. Brown, and N. Mathiowetz, 2001. Measurement Error in Survey Data, in: J.J. Heckman and E. Leamer, (Eds.), Handbook of Econometrics 5, 3705-3843.

Carroll, R.J., D. Ruppert, L.A. Stefanski and C. Crainiceanu, 2006. Measurement Error in Nonlinear Models: A Modern Perspective. Second Edition. Chapman and Hall CRC Press.

Cawley, J., 2004. The Impact of Obesity on Wages. Journal of Human Resources 39, 451474.

Cawley, J., D. Frisvold, and C. Meyerhoefer, 2013. The Impact of Physical Education on Obesity among Elementary School Children. Journal of Health Economics 32, 743-755.

Cawley, J. and J.C. Maclean, 2012. Unfit For Service: The Implications Of Rising Obesity For Us Military Recruitment. Health Economics 21, 1348-1366.

Cawley, J. and C. Meyerhoefer, 2012. The Medical Care Costs of Obesity: An Instrumental Variables Approach. Journal of Health Economics 31, 219-230.

Chen, X., H. Hong, D. Nekipelov, 2011. Nonlinear Models of Measurement Errors. Journal of Economic Literature 49, 901-937.

Chen, X., H. Hong, and E. Tamer, 2006. Measurement Error Models with Auxiliary Data. Review of Economic Studies 72, 343-366.

Cohen-Cole, E. and J.M. Flecther, 2008. Is Obesity Contagious? Social Networks versus Environmental Factors in the Obesity Epidemic. Journal of Health Economics 27, 1382-1387.

Diggle P.J. and P. Hall, 1993. A Fourier Approach to Nonparametric Deconvolution of a Density Estimate. Journal of Royal Statistical Society B 55, 523-531.

Fan, J. and Y.K. Truong, 1993. Nonparametric Regression with Errors in Variables. Annals of Statistics 21, 1900-1925.

Fuller, W.A., 1987. Measurement Error Models. John Wiley & Sons, Inc.

Hu, Y., 2008. Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables: A General Solution. Journal of Econometrics 144, 2761.

Hu, Y. and S. Schennach, 2008. Instrumental Variable Treatment of Nonclassical Measurement Error Models. Econometrica 76, 195-216.

Koijen, R., S. Van Nieuwerburgh, and R. Vestman, 2013. Judging the Quality of Survey Data by Comparison with "Truth" as Measured by Administrative Records: Evidence from Sweden. NBER Working Paper.

Lewbel, A., 2007. Estimation of Average Treatment Effects with Misclassification. Econometrica 75, 537551.

Li, T., 2002. Robust and Consistent Estimation of Nonlinear Errors-in-Variables Models. Journal of Econometrics 110, 1-26.

Mahajan, A., 2006. Identification and Estimation of Regression Models with Misclassification. Econometrica 74, 631665.

Rao, P., 1992. Identifiability in Stochastic Models. Academic Press.

Schennach, S., 2004a. Estimation of Nonlinear Models with Measurement Error. Econometrica 72, 33-75.

Schennach, S., 2004b. Nonparametric Regression in the Presence of Measurement Errors. Econometric Theory 20, 1046-1093.

Schennach, S., 2013. Measurement Error in Nonlinear Models - A Review, in: Advances in Economics and Econometrics, Econometric Society Monographs 3, CH. 8.

Trogdon, J.G., J. Nonnemaker, and J. Paris, 2008. Peer Effects in Adolescent Overweight. Journal of Health Economics 27, 1388-1399.

A Mathematical Appendix

A.1 Proof of Theorem 1

First, note that the coefficient γ_1 is uniquely determined by $\gamma_1 = \operatorname{Cov}(Y,X_1)/\operatorname{Cov}(Y,X_2)$ under Assumption 1 – see Lemma 1. We identify f_{X^*} using Kotlarski's identity (see Rao, 1992) as a preliminary step. Note that the last two equations of the model (2.1) yields $\phi_{X_1X_2}(t_1,t_2) = \mathrm{E}\left[e^{i(\gamma_1 t_1+t_2)X^*+it_1\mathcal{E}_1+it_2\mathcal{E}_2}\right]$. Differentiate this characteristic function with respect to t_2 and evaluate it at $t_2=0$ to obtain

$$\begin{aligned}\frac{\partial}{\partial t_2}\phi_{X_1X_2}(t_1,t_2)\Big|_{t_2=0} &= \mathrm{E}\left[iX^*e^{i\gamma_1t_1X^*+it_1\mathcal{E}_1}\right]+\mathrm{E}\left[i\,\mathcal{E}_2\,e^{iX_1t_1}\right]\\ &= \mathrm{E}\left[iX^*e^{i\gamma_1t_1X^*}\right]\cdot\mathrm{E}\left[e^{it_1\mathcal{E}_1}\right]. \qquad \text{(A.1)}\end{aligned}$$

where the last equality follows from Assumption 3 (ii) and (iii). Given Assumption 3 (ii), we similarly have

$$\phi_{X_1}(t_1)=\mathrm{E}\left[e^{it_1X_1}\right]=\mathrm{E}\left[e^{i\gamma_1t_1X^*}\right]\cdot\mathrm{E}\left[e^{it_1\mathcal{E}_1}\right]. \qquad \text{(A.2)}$$

Assumption 4 (i) allows us to take the ratio of (A.1) to (A.2) to obtain $\frac{\frac{\partial}{\partial t_2}\phi_{X_1X_2}(t_1,t_2)\big|_{t_2=0}}{\phi_{X_1}(t_1)} = \frac{\mathrm{E}\left[iX^*e^{i\gamma_1t_1X^*}\right]}{\mathrm{E}\left[e^{i\gamma_1t_1X^*}\right]} = \frac{\partial}{\partial\tau}\ln\mathrm{E}\left[e^{i\tau X^*}\right]\big|_{\tau=\gamma_1t_1}$. Therefore, it follows that the characteristic function of X^* is given by

$$\phi_{X^*}(t)=\exp\left(\int_0^t\frac{\frac{\partial}{\partial t_2}\phi_{X_1X_2}(t_1/\gamma_1,t_2)\Big|_{t_2=0}}{\phi_{X_1}(t_1/\gamma_1)}dt_1\right). \qquad \text{(A.3)}$$

To solve the model explicitly for $g(x^*)$, we make a similar calculation. For ϕ_{X_1Y} defined by $\phi_{X_1Y}(t_1,s)=\mathrm{E}\left[e^{it_1X_1+isY}\right]$, we have $\frac{\partial}{\partial s}\phi_{X_1Y}(t_1,s)\big|_{s=0} = i\,\mathrm{E}\left[(g(X^*)+U)\ e^{it_1(\gamma_1X^*+\mathcal{E}_1)}\right] = i\,\mathrm{E}\left[g(X^*)\ e^{it_1(\gamma_1X^*+\mathcal{E}_1)}\right]+i\,\mathrm{E}\left[Ue^{it_1X_1}\right] = i\,\mathrm{E}\left[g(X^*)\ e^{it_1\gamma_1X^*}\right]\ \frac{\phi_{X_1}(t_1)}{\phi_{X^*}(t_1\gamma_1)}$, where the last equality is due to 3 (i) and (ii). The last expression makes sense because Assumption 4 (i) implies that the characteristic function ϕ_{X^*} does not vanish on the real line. Rearranging this equality yields $\phi_{X^*}(t_1)\ \frac{\frac{\partial}{\partial s}\phi_{X_1Y}(t_1/\gamma_1,s)\big|_{s=0}}{i\ \phi_{X_1}(t_1/\gamma_1)} = \int e^{it_1x^*}\ g(x^*)\ f_{X^*}(x^*)dx^*$. This is the Fourier inverse of $g\cdot f_{X^*}$, and applying the Fourier transform yields $g(x^*)\ f_{X^*}(x^*)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{-it_1x^*}\phi_{X^*}(t_1)\ \frac{\frac{\partial}{\partial s}\phi_{X_1Y}(t_1/\gamma_1,s)\big|_{s=0}}{i\ \phi_{X_1}(t_1/\gamma_1)}dt_1$

for each point x^* by the Fourier transformation. Therefore, we derive the following closed-form solution to $g(x^*)$.

$$g(x^*) = \frac{\int_{-\infty}^{+\infty} e^{-itx^*} \exp\left(\int_0^t \frac{\frac{\partial}{\partial t_2}\phi_{X_1X_2}(t_1/\gamma_1,t_2)\Big|_{t_2=0}}{\phi_{X_1}(t_1/\gamma_1)} dt_1 \right) \frac{\frac{\partial}{\partial s}\phi_{X_1Y}(t/\gamma_1,s)\Big|_{s=0}}{i\ \phi_{X_1}(t/\gamma_1)}\, dt}{\int_{-\infty}^{+\infty} e^{-itx^*} \exp\left(\int_0^t \frac{\frac{\partial}{\partial t_2}\phi_{X_1X_2}(t_1/\gamma_1,t_2)\Big|_{t_2=0}}{\phi_{X_1}(t_1/\gamma_1)} dt_1 \right) dt}$$

where the first equality uses the Fourier inversion of $\phi_{X^*}(t)$ for f_{X^*} in the denominator and the second equality uses the expression of $\phi_{X^*}(t)$ in equation (A.3).

A.2 Quadratic Model of Measurement

Suppose that the endogenous measurement X_1 is modeled with $P = 2$ by

$$X_1 = \gamma_1 X^* + \gamma_2 X^{*2} + \mathcal{E}_1 \qquad \mathrm{E}[\mathcal{E}_1] = \gamma_0. \tag{A.4}$$

Consider the following homoscedasticity assumption:

$$\mathrm{E}[U^2 \mid X^*, \mathcal{E}_1, \mathcal{E}_2] = \mathrm{E}[U^2]. \tag{A.5}$$

With Assumption 5 and (A.5), if the empirically testable rank condition

$$\mathrm{Cov}(Y, X_2) \cdot \mathrm{Cov}(Y^2, X_2^2) \neq \mathrm{Cov}(Y, X_2^2) \cdot \mathrm{Cov}(Y^2, X_2) \tag{A.6}$$

holds, then we can show that the coefficients γ_1 and γ_2 of the model (A.4) are identified with closed-form solutions as follows.

By Assumption 5, we get $\mathrm{Cov}(Y, X_1) = \gamma_1 \mathrm{Cov}(g(X^*), X^*) + \gamma_2 \mathrm{Cov}(g(X^*), X^{*2})$, $\mathrm{Cov}(Y, X_2) = \mathrm{Cov}(g(X^*), X^*)$ and $\mathrm{Cov}(Y, X_2^2) = \mathrm{Cov}(g(X^*), X^{*2})$. Furthermore, by Assumptions 5 and (A.5), we get we obtain $\mathrm{Cov}(Y^2, X_1) = \gamma_1 \mathrm{Cov}(g(X^*)^2, X^*) + \gamma_2 \mathrm{Cov}(g(X^*)^2, X^{*2})$, $\mathrm{Cov}(Y^2, X_2) = \mathrm{Cov}(g(X^*)^2, X^*)$ and $\mathrm{Cov}(Y^2, X_2^2) = \mathrm{Cov}(g(X^*)^2, X^{*2})$. Combining these six equations yields

$$\begin{bmatrix} \mathrm{Cov}(Y, X_1) \\ \mathrm{Cov}(Y^2, X_1) \end{bmatrix} = \begin{bmatrix} \mathrm{Cov}(Y, X_2) & \mathrm{Cov}(Y, X_2^2) \\ \mathrm{Cov}(Y^2, X_2) & \mathrm{Cov}(Y^2, X_2^2) \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}.$$

Therefore, we identify γ_1 and γ_2 under the rank condition (A.6) with the closed-form formula:

$$\gamma_1 = \frac{\mathrm{Cov}(Y, X_1) \cdot \mathrm{Cov}(Y^2, X_2^2) - \mathrm{Cov}(Y, X_2^2) \cdot \mathrm{Cov}(Y^2, X_1)}{\mathrm{Cov}(Y, X_2) \cdot \mathrm{Cov}(Y^2, X_2^2) - \mathrm{Cov}(Y, X_2^2) \cdot \mathrm{Cov}(Y^2, X_2)} \tag{A.7}$$

$$\gamma_2 = \frac{\mathrm{Cov}(Y, X_2) \cdot \mathrm{Cov}(Y^2, X_1) - \mathrm{Cov}(Y, X_1) \cdot \mathrm{Cov}(Y^2, X_2)}{\mathrm{Cov}(Y, X_2) \cdot \mathrm{Cov}(Y^2, X_2^2) - \mathrm{Cov}(Y, X_2^2) \cdot \mathrm{Cov}(Y^2, X_2)}, \tag{A.8}$$

Furthermore, Assumption 5 also allows us to identify γ_0 and σ_2^2 with closed-form solutions as follows. Again, by using Assumption 5, we obtain $\mathrm{E}[YX_1] = \gamma_1 \mathrm{E}[YX^*] + \gamma_2 \mathrm{E}[YX^{*2}] + \gamma_0 \mathrm{E}[Y]$, $\mathrm{E}[X_1X_2] = \gamma_1 \mathrm{E}[X^{*2}] + \gamma_2 \mathrm{E}[X^{*3}] + \gamma_0 \mathrm{E}[X^*]$, $\mathrm{E}[YX_1X_2] = \gamma_1 \mathrm{E}[YX^{*2}] + \gamma_2 \mathrm{E}[YX^{*3}] + \gamma_0 \mathrm{E}[YX^*]$, $\mathrm{E}[X_2] = \mathrm{E}[X^*]$, $\mathrm{E}[X_2^2] = \mathrm{E}[X^{*2}] + \sigma_2^2$, $\mathrm{E}[X_2^3] = \mathrm{E}[X^{*3}] + 3\sigma_2^2 \mathrm{E}[X^*] + \sigma_2^3$, $\mathrm{E}[X_2] = \mathrm{E}[YX^*]$, $\mathrm{E}[X_2^2] = \mathrm{E}[YX^{*2}] + \sigma_2^2 \mathrm{E}[Y]$, and $\mathrm{E}[X_2^3] = \mathrm{E}[YX^{*3}] + 3\sigma_2^2 \mathrm{E}[YX^*] + \sigma_2^3 \mathrm{E}[Y]$. Substituting the last six equations into the first three equations above, we obtain the following system of three equations $\mathrm{E}[YX_1] = \gamma_1 \mathrm{E}[YX_2] + \gamma_2(\mathrm{E}[YX_2^2] - \sigma_2^2 \mathrm{E}[Y]) + \gamma_0 \mathrm{E}[Y]$, $\mathrm{E}[X_1X_2] = \gamma_1(\mathrm{E}[X_2^2] - \sigma_2^2) + \gamma_2(\mathrm{E}[X_2^2] - 3\sigma_2^2 \mathrm{E}[X_2] - \sigma_2^3) + \gamma_0 \mathrm{E}[X_2]$ and $\mathrm{E}[YX_1X_2] = \gamma_1(\mathrm{E}[YX_2^2] - \sigma_2^2 \mathrm{E}[Y]) + \gamma_2(\mathrm{E}[YX_2^3] - 3\sigma_2^2 \mathrm{E}[YX_2] - \sigma_2^3 \mathrm{E}[Y]) + \gamma_0 \mathrm{E}[YX_2]$. This system can be written as the linear equation:

$$\begin{bmatrix} \mathrm{E}[YX_1] \\ \mathrm{E}[X_1X_2] \\ \mathrm{E}[YX_1X_2] \end{bmatrix} = \begin{bmatrix} \mathrm{E}[Y] & -\gamma_2 \mathrm{E}[Y] & 0 \\ \mathrm{E}[X_2] & -\gamma_1 - 3\gamma_2 \mathrm{E}[X_2] & -\gamma_2 \\ \mathrm{E}[YX_2] & -\gamma_1 \mathrm{E}[Y] - 3\gamma_2 \mathrm{E}[YX_2] & -\gamma_2 \mathrm{E}[Y] \end{bmatrix} \begin{bmatrix} \gamma_0 \\ \sigma_2^2 \\ \sigma_2^3 \end{bmatrix}.$$

If the following empirically testable rank conditions hold, then the above three by three matrix is nonsingular.

$$\begin{aligned} &(i) && \mathrm{Cov}(Y, X_2) \cdot \mathrm{Cov}(Y^2, X_1) \neq \mathrm{Cov}(Y, X_1) \cdot \mathrm{Cov}(Y^2, X_2), \\ &(ii) && \mathrm{Cov}(Y, X_2) \neq 0, \text{and} \\ &(iii) && \mathrm{E}[Y] \neq 0. \end{aligned} \tag{A.9}$$

Therefore, the linear system yields a unique solution to $(\gamma_0, \sigma_2^2, \sigma_2^3)$. In particular, it yields the following closed-form formula for σ_2^2:

$$\sigma_2^2 = \frac{1}{2\gamma_2} \left(\frac{\mathrm{Cov}(Y, X_1X_2)}{\mathrm{Cov}(Y, X_2)} - \frac{\mathrm{E}[YX_1]}{\mathrm{E}[Y]} \right). \tag{A.10}$$

Lastly, recall that $\sigma_2^1 := \mathrm{E}[\mathcal{E}_2^1] = 0$ by the definition of the model (3.1). In summary, we obtain γ_1, γ_2, σ_2^1, and σ_2^2, all as closed-form formulas written in terms of observed data.

Applying the general closed-form identification result of Theorem 2 to this context, we obtain the following specific result. Suppose that Assumptions 5, 6, 7, and 8 hold for the model (A.4). If in addition (A.5) and the empirically testable rank conditions, (A.6) and (A.9), are satisfied, then the nonparametric function g evaluated at x^* in the interior of the support of X^* is identified with the closed-form solution:

$$g(x^*) \;=\; \frac{\int\int\int e^{-itx^*+itx-it'(\gamma_1 x+\gamma_2 x^2)} \left|\gamma_1+2\gamma_2 x\right| \frac{\mathrm{E}[Ye^{itX_2}]}{\mathrm{E}[e^{itX_2}]}\phi_{Z^*}(t')dt'dxdt}{2\pi\int e^{-it(\gamma_1 x^*+\gamma_2 x^{*2})}\left|\gamma_1+2\gamma_2 x^*\right|\phi_{Z^*}(t)dt},$$

where ϕ_{Z^*} is identified with the closed-form solution

$$\phi_{Z^*}(t) \;=\; \exp\left\{\int_0^t \gamma_1\frac{\mathrm{E}[X_2 e^{it_1X_1}]}{\mathrm{E}[e^{it_1X_1}]}+\gamma_2\frac{\mathrm{E}[X_2^2e^{it_1X_1}]-\sigma_2^2\,\mathrm{E}[e^{it_1X_1}]}{\mathrm{E}[e^{it_1X_1}]}dt_1\right\},$$

and γ_1, γ_2 and σ_2^2 are given by the closed-form solutions (A.7), (A.8) and (A.10), respectively.

A.3 Proof of Theorem 2

Proof. Define $Z^* := \sum_{j=1}^J \gamma_j X^{*j}$. Using Assumption 5 (iii), we obtain the equality $\mathrm{E}[X_2^j e^{itX_1}] = \sum_{q=0}^j \binom{j}{q}\sigma_2^{j-q}\,\mathrm{E}[X^{*q}e^{it(Z^*+\mathcal{E}_1)}]$ for all $j=1,\cdots,J$. Hence, we obtain the linear equation

$$\begin{bmatrix} 1 & \binom{J}{J-1}\sigma_2^1 & \cdots & \binom{J}{2}\sigma_2^{J-2} & \binom{J}{1}\sigma_2^{J-1} \\ & 1 & \cdots & \binom{J-1}{2}\sigma_2^{J-3} & \binom{J-1}{1}\sigma_2^{J-2} \\ & & \ddots & \vdots & \vdots \\ & & & 1 & \binom{2}{1}\sigma_2^1 \\ & & & & 1 \end{bmatrix} \begin{bmatrix} \mathrm{E}[X^{*J}e^{it(Z^*+\mathcal{E}_1)}] \\ \mathrm{E}[X^{*(J-1)}e^{it(Z^*+\mathcal{E}_1)}] \\ \vdots \\ \mathrm{E}[X^{*2}e^{it(Z^*+\mathcal{E}_1)}] \\ \mathrm{E}[X^{*}e^{it(Z^*+\mathcal{E}_1)}] \end{bmatrix} = \begin{bmatrix} \mathrm{E}[(X_2^J-\sigma_2^J)e^{itX_1}] \\ \mathrm{E}[(X_2^{J-1}-\sigma_2^{J-1})e^{itX_1}] \\ \vdots \\ \mathrm{E}[(X_2^2-\sigma_2^2)e^{itX_1}] \\ \mathrm{E}[(X_2-\sigma_2^1)e^{itX_1}] \end{bmatrix}.$$

We obtain the closed-form solution $\mathrm{E}[X^{*j}e^{it(Z^*+\mathcal{E}_1)}] = \mu(t,j;\sigma_2^1,\cdots,\sigma_2^J;F_{X_1X_2})$ for $j=1,\cdots,J$ under Assumption 6, where $[\mu(t,J;\sigma_2^1,\cdots,\sigma_2^J;F_{X_1X_2}),\cdots,\mu(t,1;\sigma_2^1,\cdots,\sigma_2^J;F_{X_1X_2})]'$ is explicitly written as

$$\begin{bmatrix} 1 & \binom{J}{J-1}\sigma_2^1 & \cdots & \binom{J}{2}\sigma_2^{J-2} & \binom{J}{1}\sigma_2^{J-1} \\ & 1 & \cdots & \binom{J-1}{2}\sigma_2^{J-3} & \binom{J-1}{1}\sigma_2^{J-2} \\ & & \ddots & \vdots & \vdots \\ & & & 1 & \binom{2}{1}\sigma_2^1 \\ & & & & 1 \end{bmatrix}^{-1} \begin{bmatrix} \mathrm{E}[(X_2^J - \sigma_2^J)e^{itX_1}] \\ \mathrm{E}[(X_2^{J-1} - \sigma_2^{J-1})e^{itX_1}] \\ \vdots \\ \mathrm{E}[(X_2^2 - \sigma_2^2)e^{itX_1}] \\ \mathrm{E}[(X_2 - \sigma_2^1)e^{itX_1}] \end{bmatrix}.$$

Using Assumption 5 (ii), we can write $\mu(t, j; \sigma_2^1, \cdots, \sigma_2^J; F_{X_1X_2}) = \mathrm{E}[X^{*j}e^{itZ^*}]\,\mathrm{E}[e^{it\,\mathcal{E}_1}]$ for each $j = 1, \cdots, J$. Division of this equality by $\mathrm{E}[e^{itX_1}] = \mathrm{E}[e^{itZ^*}]\,\mathrm{E}[e^{it\,\mathcal{E}_1}]$ that also follows by Assumption 5 (ii) yields $\frac{\mu(t,j;\sigma_2^1,\cdots,\sigma_2^J;F_{X_1X_2})}{\mathrm{E}[e^{itX_1}]} = \frac{\mathrm{E}[X^{*j}e^{itZ^*}]}{\mathrm{E}[e^{itZ^*}]}$ for each $j = 1, \cdots, J$. Taking a linear combination of this equality gives $\frac{\sum_{j=1}^J \gamma_j \mu(t,j;\sigma_2^1,\cdots,\sigma_2^J;F_{X_1X_2})}{\mathrm{E}[e^{itX_1}]} = \frac{\mathrm{E}[Z^*e^{itZ^*}]}{\mathrm{E}[e^{itZ^*}]} = \frac{d}{dt}\ln \mathrm{E}[e^{itZ^*}]$. Thus, the characteristic function ϕ_{Z^*} is given by

$$\phi_{Z^*}(t) \;=\; \exp\left\{\int_0^t \frac{\sum_{j=1}^J \gamma_j \mu(t_1, j; \sigma_2^1, \cdots, \sigma_2^J; F_{X_1X_2})}{\mathrm{E}[e^{it_1X_1}]}dt_1\right\}. \tag{A.11}$$

By Assumption 7 (iii), apply the Fourier transform to this characteristic function ϕ_{Z^*} to get $f_{Z^*}(z^*) = \frac{1}{2\pi}\int e^{-itz^*}\phi_{Z^*}(t)dt$. Define the function h by $h(x) = \sum_{j=1}^J \gamma_j x^j$. Note that $Z^* = h(X^*)$ by the definition of Z^*. Then, by Assumption 8, we use the transformation formula to obtain f_{X^*}:

$$f_{X^*}(x^*) = f_{Z^*}(h(x^*))\,|h'(x^*)| = |h'(x^*)|\,\frac{1}{2\pi}\int e^{-ith(x^*)}\phi_{Z^*}(t)dt. \tag{A.12}$$

By Assumption 7 (ii), we apply the Fourier inversion to this f_{X^*} to get

$$\phi_{X^*}(t) = \int e^{itx^*}f_{X^*}(x^*) = \frac{1}{2\pi}\int\int e^{itx^*}e^{-it'h(x^*)}\,|h'(x^*)|\,\phi_{Z^*}(t')dt'dx^*. \tag{A.13}$$

Now, under Assumption 5 (i) and (iii), we have the equalities $\mathrm{E}[Ye^{-itX_2}] = \mathrm{E}[g(X^*)e^{itX^*}]\,\mathrm{E}[e^{it\,\mathcal{E}_2}]$ and $\mathrm{E}[e^{-itX_2}] = \mathrm{E}[e^{itX^*}]\,\mathrm{E}[e^{it\,\mathcal{E}_2}]$. Take the ratio and rearrange the result to obtain $\mathrm{E}[g(X^*)e^{itX^*}] = \frac{\mathrm{E}[Ye^{itX_2}]}{\mathrm{E}[e^{itX_2}]}\phi_{X^*}(t)$. (A.13) yields $\mathrm{E}[g(X^*)e^{itX^*}] = \frac{\mathrm{E}[Ye^{itX_2}]}{\mathrm{E}[e^{itX_2}]}\frac{1}{2\pi}\int\int e^{itx^*}e^{-it'h(x^*)}\,|h'(x^*)|\,\phi_{Z^*}(t')dt'dx^*$. Applying the Fourier transform under Assumption 7 (iv), we obtain the equality $g(x^*)f_{X^*}(x^*) =$

$\frac{1}{4\pi^2}\int\int\int e^{-itx^*+itx-it'h(x)}\frac{\mathrm{E}[Ye^{itX_2}]}{\mathrm{E}[e^{itX_2}]}\left|h'(x)\right|\phi_{Z^*}(t')dt'dxdt$. Finally, divide this equation by (A.12) to identify $g(x^*)$ with the closed-form solution.

$$g(x^*) \;=\; \frac{\int\int\int e^{-itx^*+itx-it'h(x)}\left|h'(x)\right|\frac{\mathrm{E}[Ye^{itX_2}]}{\mathrm{E}[e^{itX_2}]}\phi_{Z^*}(t')dt'dxdt}{2\pi\int e^{-ith(x^*)}\left|h'(x^*)\right|\phi_{Z^*}(t)dt}.$$

Using the closed-form solution (A.11) to ϕ_{Z^*} and the definition of the function h yields the desired result. □

A.4 Proof of Lemma 2

Proof. For compactness of writing, we focus on the case of $\gamma_1 = 1$. Similar lines of argument show that the same conclusion holds for general γ_1. In order to derive the asymptotic distribution of the closed-form estimator $\widehat{g(x^*)}$, we decompose it into the numerator and the denominator given by $\widehat{g(x^*)f_{X^*}}(x^*) \;:=\; \frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{-itx^*}\exp\left(i\int_0^t\frac{\sum_{j=1}^n X_{2,j}e^{it_1X_{1,j}}}{\sum_{j=1}^n e^{it_1X_{1,j}}}dt_1\right)\frac{\sum_{j=1}^n Y_je^{itX_{1,j}}}{\sum_{j=1}^n e^{itX_{1,j}}}\phi_K(th)dt$ and $\widehat{f_{X^*}(x^*)} \;:=\; \frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{-itx^*}\exp\left(i\int_0^t\frac{\sum_{j=1}^n X_{2,j}e^{it_1X_{1,j}}}{\sum_{j=1}^n e^{it_1X_{1,j}}}dt_1\right)\phi_K(th)dt$, respectively.

The absolute bias of $\widehat{f_{X^*}(x^*)}$ is bounded by the sum of two terms:

$$\begin{aligned}\left|\mathrm{E}\,\widehat{f_{X^*}(x^*)} - f_{X^*}(x^*)\right| \;&\leqslant\; \left|\mathrm{E}\,\widehat{f_{X^*}(x^*)} - \frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{-itx^*}\phi_{X^*}(t)\phi_K(th)dt\right| \\ &\quad + \left|\frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{-itx^*}\phi_{X^*}(t)\phi_K(th)dt - f_{X^*}(x^*)\right|.\end{aligned}$$

The first term on the right-hand side is asymptotically bounded by

$$\begin{aligned}&\frac{\|\phi_K\|_\infty\|\phi_{X^*}\|_\infty}{2\pi h}\int_{-1}^{1}\int_0^{t/h}\left(\frac{\mathrm{E}\left|\frac{1}{n}\sum_{j=1}^n X_{2,j}e^{it_1X_{1,j}} - \mathrm{E}\,X_{2,j}e^{it_1X_{1,j}}\right|}{|\phi_{X_1}(t_1)|}\right. \\ &\left.+\frac{\|\phi'_{X^*}\|_\infty\,\mathrm{E}\left|\frac{1}{n}\sum_{j=1}^n e^{it_1X_{1,j}} - \mathrm{E}\,e^{it_1X_{1,j}}\right|}{|\phi_{X_1}(t_1)|} + \xi_f(t_1)\right)dt_1dt \;=\; \mathcal{O}\left(\frac{1}{n^{1/2}h^2\left|\phi_{X_1}(1/h)\right|}\right),\end{aligned}$$

where the higher-order terms ξ_f vanish faster than the leading term uniformly under Assumptions 4 (i) and 9 (iii). On the other hand, the second term is asymptotically

$$\left|\int_{-\infty}^{+\infty}f_{X^*}(x_1)\frac{1}{h}K\left(\frac{x_1-x^*}{h}\right)dx_1 - f_{X^*}(x^*)\right| = \mathcal{O}(h^k),$$

where k is the exponent provided in Assumption 10 (i). Therefore, we obtain the asymptotic order $\left|\mathrm{E}\,\widehat{f_{X^*}(x^*)} - f_{X^*}(x^*)\right| = \mathcal{O}\left(\frac{1}{n^{1/2}h^2\left|\phi_{X_1}(1/h)\right|}\right) + \mathcal{O}(h^k)$ of the absolute bias of $\widehat{f_{X^*}(x^*)}$.

Similarly, the absolute bias of $\widehat{g(x^*)f_{X^*}}(x^*)$ is bounded by the sum of two terms:

$$\left|\mathrm{E}\,\widehat{g(x^*)f_{X^*}}(x^*) - g(x^*)f_{X^*}(x^*)\right| \leqslant \left|\mathrm{E}\,\widehat{g(x^*)f_{X^*}}(x^*) - \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx^*}\phi_{X^*}(t)\frac{\mathrm{E}\,Y_j e^{itX_{1,j}}}{\mathrm{E}\,e^{itX_{1,j}}}\phi_K(th)dt\right| + \left|\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx^*}\phi_{X^*}(t)\frac{\mathrm{E}\,Y_j e^{itX_{1,j}}}{\mathrm{E}\,e^{itX_{1,j}}}\phi_K(th)dt - g(x^*)f_{X^*}(x^*)\right|.$$

The first term on the right-hand side is asymptotically bounded by

$$\frac{\|\phi_K\|_\infty \|\phi_{X^*}\|_\infty \|\phi_{X_1}\|_\infty \mathrm{E}\,|Y_j|}{2\pi h}\int_{-1}^{1}\int_0^{t/h}\left(\frac{\mathrm{E}\left|\frac{1}{n}\sum_{j=1}^n X_{2,j}e^{it_1X_{1,j}} - \mathrm{E}\,X_{2,j}e^{it_1X_{1,j}}\right|}{|\phi_{X_1}(t_1)|} + \right.$$
$$\left.\frac{\|\phi'_{X^*}\|_\infty \mathrm{E}\left|\frac{1}{n}\sum_{j=1}^n e^{it_1X_{1,j}} - \mathrm{E}\,e^{it_1X_{1,j}}\right|}{|\phi_{X_1}(t_1)|} + \xi_f(t_1)\right)\frac{1}{|\phi_{X_1}(t/h)|}dt_1dt +$$
$$\frac{\|\phi_K\|_\infty \|\phi_{X^*}\|_\infty}{2\pi h}\int_{-1}^{1}\left(\frac{\mathrm{E}\left|\frac{1}{n}\sum_{j=1}^n Y_j e^{itX_{1,j}/h} - \mathrm{E}\,Y_j e^{itX_{1,j}/h}\right|}{|\phi_{X_1}(t/h)|} + \right.$$
$$\left.\frac{\mathrm{E}\,|Y_j|\,\|\phi_{X_1}\|_\infty \mathrm{E}\left|\frac{1}{n}\sum_{j=1}^n e^{itX_{1,j}/h} - \mathrm{E}\,e^{itX_{1,j}/h}\right|}{|\phi_{X_1}(t/h)|^2} + \xi_g(t/h)\right)dt = \mathcal{O}\left(\frac{1}{n^{1/2}h^2\,|\phi_{X_1}(1/h)|^2}\right)$$

under Assumption 9 (iii), and the higher-order terms ξ_f and ξ_g vanish faster than the leading terms uniformly. On the other hand, the second term is asymptotically

$$\left|\int_{-\infty}^{+\infty} g(x_1)f_{X^*}(x_1)\frac{1}{h}K\left(\frac{x_1 - x^*}{h}\right)dx_1 - g(x^*)f_{X^*}(x^*)\right| = \mathcal{O}(h^{\min\{k,l\}}),$$

where k is the exponent for f_{X^*} provided in Assumption 10 (i), and l is the exponent for g provided in Assumption 10 (ii). Therefore, we have the following asymptotic order of the absolute bias: $\left|\mathrm{E}\,\widehat{g(x^*)f_{X^*}}(x^*) - g(x^*)f_{X^*}(x^*)\right| = \mathcal{O}\left(\frac{1}{n^{1/2}h^2\left|\phi_{X_1}(1/h)\right|^2}\right) + \mathcal{O}(h^{\min\{k,l\}})$.

Next, the variance of $\widehat{f_{X^*}(x^*)}$ is asymptotically bounded by

$$\frac{\|\phi_*\|_\infty^2 \|\phi_K\|_\infty^2}{4\pi^2h^2}\int_{-1}^{1}\int_{-1}^{1}\int_0^{t/h}\int_0^{\tau/h}\{$$
$$\frac{\left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n X_{2,j}e^{it_1X_{1,j}} - \mathrm{E}\,X_{2,j}e^{it_1X_{1,j}}\right]^2\right)^{\frac{1}{2}}\cdot\left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n X_{2,j}e^{i\tau_1X_{1,j}} - \mathrm{E}\,X_{2,j}e^{i\tau_1X_{1,j}}\right]^2\right)^{\frac{1}{2}}}{|\phi_{X_1}(t_1)|\,|\phi_{X_1}(\tau_1)|} +$$
$$\frac{2\,|\phi'_{X^*}(\tau_1)|\left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n X_{2,j}e^{it_1X_{1,j}} - \mathrm{E}\,X_{2,j}e^{it_1X_{1,j}}\right]^2\right)^{\frac{1}{2}}\cdot\left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n e^{i\tau_1X_{1,j}} - \mathrm{E}\,e^{i\tau_1X_{1,j}}\right]^2\right)^{\frac{1}{2}}}{|\phi_{X_1}(t_1)|\,|\phi_{X_1}(\tau_1)|} +$$
$$\frac{|\phi'_{X^*}(\tau_1)|\,|\phi'_{X^*}(t_1)|\left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n e^{it_1X_{1,j}} - \mathrm{E}\,e^{it_1X_{1,j}}\right]^2\right)^{\frac{1}{2}}\cdot\left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n e^{i\tau_1X_{1,j}} - \mathrm{E}\,e^{i\tau_1X_{1,j}}\right]^2\right)^{\frac{1}{2}}}{|\phi_{X_1}(t_1)|\,|\phi_{X_1}(\tau_1)|} +$$
$$\xi_f(t_1,\tau_1)\;\}\,d\tau_1dt_1dtd\tau = \mathcal{O}\left(\frac{1}{nh^4\,|\phi_{X_1}(1/h)|^2}\right),$$

where the higher-order terms ξ_f vanish faster than the leading terms uniformly under Assumptions 4 (i) and 9 (iii). Similarly, the variance of $\widehat{g(x^*)f_{X^*}}(x^*)$ is asymptotically bounded by

$$\frac{\|\phi_*\|_\infty^2 \|\phi_K\|_\infty^2}{4\pi^2 h^2} \int_{-1}^{1}\int_{-1}^{1}\int_{0}^{t/h}\int_{0}^{\tau/h} I(t,\tau,t_1,\tau_1,h)d\tau_1 dt_1 dt d\tau \ = \ \mathcal{O}\left(\frac{1}{nh^4 \left|\phi_{X_1}(1/h)\right|^4}\right),$$

where $I(t,\tau,t_1,\tau_1,h) =$

$$\frac{(\mathrm{E}\,|Y_j|)^2 \|\phi_{X_1}\|_\infty^2 \left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n X_{2,j}e^{it_1X_{1,j}} - \mathrm{E}\,X_{2,j}e^{it_1X_{1,j}}\right]^2\right)^{\frac{1}{2}} \cdot \left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n X_{2,j}e^{i\tau_1X_{1,j}} - \mathrm{E}\,X_{2,j}e^{i\tau_1X_{1,j}}\right]^2\right)^{\frac{1}{2}}}{|\phi_{X_1}(t_1)|\,|\phi_{X_1}(\tau_1)|\,|\phi_{X_1}(t/h)|\,|\phi_{X_1}(\tau/h)|} +$$

$$\frac{(\mathrm{E}\,|Y_j|)^2 \|\phi_{X_1}\|_\infty^2 \,|\phi'_{X^*}(\tau_1)|\,|\phi'_{X^*}(t_1)| \left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n e^{it_1X_{1,j}} - \mathrm{E}\,e^{it_1X_{1,j}}\right]^2\right)^{\frac{1}{2}} \cdot \left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n e^{i\tau_1X_{1,j}} - \mathrm{E}\,e^{i\tau_1X_{1,j}}\right]^2\right)^{\frac{1}{2}}}{|\phi_{X_1}(t_1)|\,|\phi_{X_1}(\tau_1)|\,|\phi_{X_1}(t/h)|\,|\phi_{X_1}(\tau/h)|} +$$

$$\frac{\left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n Y_je^{itX_{1,j}/h} - \mathrm{E}\,Y_je^{itX_{1,j}/h}\right]^2\right)^{\frac{1}{2}} \cdot \left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n Y_je^{i\tau X_{1,j}/h} - \mathrm{E}\,Y_je^{i\tau X_{1,j}/h}\right]^2\right)^{\frac{1}{2}}}{|\phi_{X_1}(t/h)|\,|\phi_{X_1}(\tau/h)|} +$$

$$\frac{(\mathrm{E}\,|Y_j|)^2 \left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n e^{itX_{1,j}/h} - \mathrm{E}\,e^{itX_{1,j}/h}\right]^2\right)^{\frac{1}{2}} \cdot \left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n e^{i\tau X_{1,j}/h} - \mathrm{E}\,e^{i\tau X_{1,j}/h}\right]^2\right)^{\frac{1}{2}}}{|\phi_{X_1}(t/h)|^2\,|\phi_{X_1}(\tau/h)|^2} +$$

$$\frac{2\,(\mathrm{E}\,|Y_j|)^2 \|\phi_{X_1}\|_\infty^2 \,|\phi'_{X^*}(\tau_1)| \left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n X_{2,j}e^{it_1X_{1,j}} - \mathrm{E}\,X_{2,j}e^{it_1X_{1,j}}\right]^2\right)^{\frac{1}{2}} \cdot \left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n e^{i\tau_1X_{1,j}} - \mathrm{E}\,e^{i\tau_1X_{1,j}}\right]^2\right)^{\frac{1}{2}}}{|\phi_{X_1}(t_1)|\,|\phi_{X_1}(\tau_1)|\,|\phi_{X_1}(t/h)|\,|\phi_{X_1}(\tau/h)|} +$$

$$\frac{2\,(\mathrm{E}\,|Y_j|) \|\phi_{X_1}\|_\infty \left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n X_{2,j}e^{it_1X_{1,j}} - \mathrm{E}\,X_{2,j}e^{it_1X_{1,j}}\right]^2\right)^{\frac{1}{2}} \cdot \left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n Y_je^{i\tau X_{1,j}/h} - \mathrm{E}\,Y_je^{i\tau X_{1,j}/h}\right]^2\right)^{\frac{1}{2}}}{|\phi_{X_1}(t_1)|\,|\phi_{X_1}(t/h)|\,|\phi_{X_1}(\tau/h)|} +$$

$$\frac{2\,(\mathrm{E}\,|Y_j|)^2 \|\phi_{X_1}\|_\infty^2 \left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n X_{2,j}e^{it_1X_{1,j}} - \mathrm{E}\,X_{2,j}e^{it_1X_{1,j}}\right]^2\right)^{\frac{1}{2}} \cdot \left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n e^{i\tau X_{1,j}/h} - \mathrm{E}\,e^{i\tau X_{1,j}/h}\right]^2\right)^{\frac{1}{2}}}{|\phi_{X_1}(t_1)|\,|\phi_{X_1}(t/h)|\,|\phi_{X_1}(\tau/h)|^2} +$$

$$\frac{2\,(\mathrm{E}\,|Y_j|) \|\phi_{X_1}\|_\infty \,|\phi'_{X^*}(\tau_1)| \left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n e^{it_1X_{1,j}} - \mathrm{E}\,e^{it_1X_{1,j}}\right]^2\right)^{\frac{1}{2}} \cdot \left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n Y_je^{i\tau X_{1,j}/h} - \mathrm{E}\,Y_je^{i\tau X_{1,j}/h}\right]^2\right)^{\frac{1}{2}}}{|\phi_{X_1}(t_1)|\,|\phi_{X_1}(t/h)|\,|\phi_{X_1}(\tau/h)|} +$$

$$\frac{2\,(\mathrm{E}\,|Y_j|)^2 \|\phi_{X_1}\|_\infty^2 \,|\phi'_{X^*}(\tau_1)| \left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n e^{it_1X_{1,j}} - \mathrm{E}\,e^{it_1X_{1,j}}\right]^2\right)^{\frac{1}{2}} \cdot \left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n e^{i\tau X_{1,j}/h} - \mathrm{E}\,e^{i\tau X_{1,j}/h}\right]^2\right)^{\frac{1}{2}}}{|\phi_{X_1}(t_1)|\,|\phi_{X_1}(t/h)|\,|\phi_{X_1}(\tau/h)|^2} +$$

$$\frac{2\,(\mathrm{E}\,|Y_j|) \|\phi_{X_1}\|_\infty \left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n Y_je^{itX_{1,j}/h} - \mathrm{E}\,Y_je^{itX_{1,j}/h}\right]^2\right)^{\frac{1}{2}} \cdot \left(\mathrm{E}\left[\frac{1}{n}\sum_{j=1}^n e^{i\tau X_{1,j}/h} - \mathrm{E}\,e^{i\tau X_{1,j}/h}\right]^2\right)^{\frac{1}{2}}}{|\phi_{X_1}(t/h)|\,|\phi_{X_1}(\tau/h)|^2} +$$

$$\xi_f(t_1,\tau_1) + \xi_g(t,\tau),$$

where the higher-order terms ξ_f and ξ_g vanish faster than the leading terms uniformly.

The mean square errors (MSE) of the estimator $\widehat{g(x^*)}$ is asymptotically bounded by $\frac{1}{f_{X^2}(x^*)^2}\cdot$ $\mathrm{MSE}\left(\widehat{g(x^*)f_{X^*}}(x^*)\right) + \frac{g(x^*)^2}{f_{X^*}(x^*)^2}\mathrm{MSE}\left(\widehat{f_{X^*}(x^*)}\right)$ and higher-order terms that vanish faster than these first-order terms. Thus, we have the biases of order $\mathrm{Bias}\left(\widehat{f_{X^*}(x^*)}\right) = \mathcal{O}\left(\frac{1}{n^{1/2}h^2\left|\phi_{X_1}(1/h)\right|}\right)+$

$\mathcal{O}(h^k)$) and $\text{Bias}\left(\widehat{g(x^*)f_{X^*}}(x^*)\right) = \mathcal{O}\left(\frac{1}{n^{1/2}h^2\left|\phi_{X_1}(1/h)\right|^2}\right) + \mathcal{O}(h^{\min\{k,l\}})$, and we have the variances of order $\text{Var}\left(\widehat{f_{X^*}(x^*)}\right) = \mathcal{O}\left(\frac{1}{nh^4\left|\phi_{X_1}(1/h)\right|^2}\right)$ and $\text{Var}\left(\widehat{g(x^*)f_{X^*}}(x^*)\right) = \mathcal{O}\left(\frac{1}{nh^4\left|\phi_{X_1}(1/h)\right|^4}\right)$. Note that the first term in the bias when it is squared gives the same asymptotic order as that of the variance for each of the two components of the estimator. Hence, the MSE go to zero in the order of $\mathcal{O}(h^{2\min\{k,l\}}) + \mathcal{O}\left(\frac{1}{nh^4\left|\phi_{X_1}(1/h)\right|^4}\right)$ with a choice of h such that $h \to 0$ and $nh^4\left|\phi_{X_1}(1/h)\right|^4 \to \infty$ as $n \to \infty$. This completes a proof for the closed-form estimator (4.1).

Lastly, we deal with the case where the closed-form estimator (4.2) is used instead of (4.1). To emphasize on the dependence on γ_1 and $\hat{\gamma}_1$, we let $\widehat{g(x^*,\gamma_1)}$ and $\widehat{g(x^*,\hat{\gamma}_1)}$ denote the closed-form estimators (4.1) and (4.2), respectively. Since $\left(\text{E}\left[\widehat{g(x^*,\hat{\gamma}_1)} - g(x^*)\right]^2\right)^{1/2} \leqslant \left(\text{E}\left[\widehat{g(x^*,\hat{\gamma}_1)} - \widehat{g(x^*,\gamma_1)}\right]^2\right)^{1/2} + \left(\text{E}\left[\widehat{g(x^*,\gamma_1)} - g(x^*)\right]^2\right)^{1/2}$ holds by Minkowski's inequality, it suffices to show $\text{E}\left[\widehat{g(x^*,\hat{\gamma}_1)} - \widehat{g(x^*,\gamma_1)}\right]^2 = \mathcal{O}(h^{2\min\{k,l\}}) + \mathcal{O}\left(\frac{1}{nh^4\left|\phi_{X_1}(1/h)\right|^4}\right)$. By the mean value theorem and Cauchy-Schwarz inequality, we have $\text{E}\left[\widehat{g(x^*,\hat{\gamma}_1)} - \widehat{g(x^*,\gamma_1)}\right]^2 = \text{E}\left[\frac{\partial}{\partial c}\widehat{g(x^*,c)}\Big|_{c=\gamma_1^*} \cdot (\hat{\gamma}_1 - \gamma_1)\right]^2 \leqslant \left(\text{E}\left[\frac{\partial}{\partial c}\widehat{g(x^*,c)}\Big|_{c=\gamma_1^*}\right]^4\right)^{1/2} \cdot \left(\text{E}\left[\hat{\gamma}_1 - \gamma_1\right]^4\right)^{1/2}$ where γ_1^* is between γ_1 and $\hat{\gamma}_1$. Note that $\hat{\gamma}_1 - \gamma_1 = \mathcal{O}\left(n^{-1/2}\right)$ under Assumptions 1 and 9 (iii). The second factor in the last line is $\mathcal{O}\left(n^{-1}\right)$ under Assumptions 1, and 9 with (iii)$'$. On the other hand, the first factor in the last line is $\mathcal{O}\left(h^{-4}\left|\phi_{X_1}(1/h)\right|^{-2}\right)$ under Assumptions 4 (i) and 9 with (iii)$'$ by similar lines of calculations to the ones used to derive the asymptotic order of the variances: $\text{E}\left[\frac{\partial}{\partial c}\widehat{g(x^*,c)}\Big|_{c=\gamma_1^*}\right]^4 \leqslant \frac{\|\phi_*\|_\infty^4\|\phi_K\|_\infty^4}{16\pi^4h^4} \times \int_{-1}^1\int_{-1}^1\int_{-1}^1\int_{-1}^1\int_0^{t/h}\int_0^{\tau/h}\int_0^{\tilde{t}/h}\int_0^{\tilde{\tau}/h} I(t,\tau,t_1,\tau_1,\tilde{t},\tilde{\tau},\tilde{t}_1,\tilde{\tau}_1,h)\, d\tilde{\tau}_1 d\tilde{t}_1 d\tau_1 dt_1 dt d\tau d\tilde{t} d\tilde{\tau} = \mathcal{O}\left(\frac{1}{h^8\left|\phi_{X_1}(1/h)\right|^8}\right)$, where the integrand $I(t,\tau,t_1,\tau_1,\tilde{t},\tilde{\tau},\tilde{t}_1,\tilde{\tau}_1,h)$ consists of first moments of quadratic interactions and higher-order terms that vanish faster than these first-order terms. It follows that $\text{E}\left[\widehat{g(x^*,\hat{\gamma}_1)} - \widehat{g(x^*,\gamma_1)}\right]^2 = \mathcal{O}\left(\frac{1}{nh^4\left|\phi_{X_1}(1/h)\right|^4}\right)$. This part converges at least as fast as the rate $\text{E}\left[\widehat{g(x^*,\hat{\gamma}_1)} - g(x^*)\right]^2 = \mathcal{O}(h^{2\min\{k,l\}}) + \mathcal{O}\left(\frac{1}{nh^4\left|\phi_{X_1}(1/h)\right|^4}\right)$. Therefore, the use of the closed-form estimator (4.2) instead of (4.1) does not alter the asymptotic order of the MSE. $\square$

A.5 Proof of Theorem 3

Proof. Under Assumption 11, Assumption 3 (ii) implies $|\phi_{X_1}(1/h)| = \mathcal{O}\left(e^{-h^{-\max\{\beta_1,\beta_2\}}/\kappa}\right)$ as $h \to 0$ for some $\kappa > 0$. Equating the asymptotic orders of the squared bias and the variance obtained in Lemma 2 with this smoothness condition, we obtain the asymptotic rate $h \sim (\log n)^{-1/\max\{\beta_1,\beta_2\}}$. Substituting this choice of h in the asymptotic order or the squared bias or the variance obtained in Lemma 2, we obtain the asymptotic order $\mathrm{E}\left(\widehat{g(x^*)} - g(x^*)\right)^2 = \mathcal{O}\left((\log n)^{\frac{-2\min\{k,l\}}{\max\{\beta_1,\beta_2\}}}\right)$. □

A.6 Proof of Theorem 4

Proof. Under Assumption 12, Assumption 3 (ii) implies $|\phi_{X_1}(1/h)| = \mathcal{O}\left(h^{\beta_1+\beta_2}\right)$ as $h \to 0$. Equating the asymptotic orders of the squared bias and the variance obtained in Lemma 2 with this smoothness condition, we obtain the asymptotic rate $h \sim n^{\frac{-1}{2(\min\{k,l\}+2(\beta_1+\beta_2+1))}}$. Substituting this choice of h in the asymptotic order or the squared bias or the variance obtained in Lemma 2, we obtain the asymptotic order $\mathrm{E}\left(\widehat{g(x^*)} - g(x^*)\right)^2 = \mathcal{O}\left(n^{\frac{-\min\{k,l\}}{\min\{k,l\}+2(\beta_1+\beta_2+1)}}\right)$. □

Data Folder	Label	Variable Description	Mean (Std. Dev.)	
Demographics	RIAGENDR	Gender		
Demographics	RIDAGEYR	Age		
Data Folder	Label	Variable Description	Male	Female
Examination	BMXHT	Clinically Measured Height (cm)	173.8 (7.9)	159.8 (7.2)
Examination	BMXWT	Clinically Measured Weight (kg)	88.1 (19.6)	76.1 (19.6)
Questionnaire	WHD010	Self-Reported Height (inches)	69.1 (3.4)	63.5 (3.0)
Questionnaire	WHD020	Self-Reported Weight (pounds)	194.3 (41.0)	165.5 (41.0)
Questionnaire	HUQ050	Receive Healthcare	2.16 (1.44)	2.46 (1.33)
Sample of individuals aged 40 or older.			$N = 1,905$	$N = 1,936$

Table 1: NHANES III 2009–2010 variable list and summary statistics.

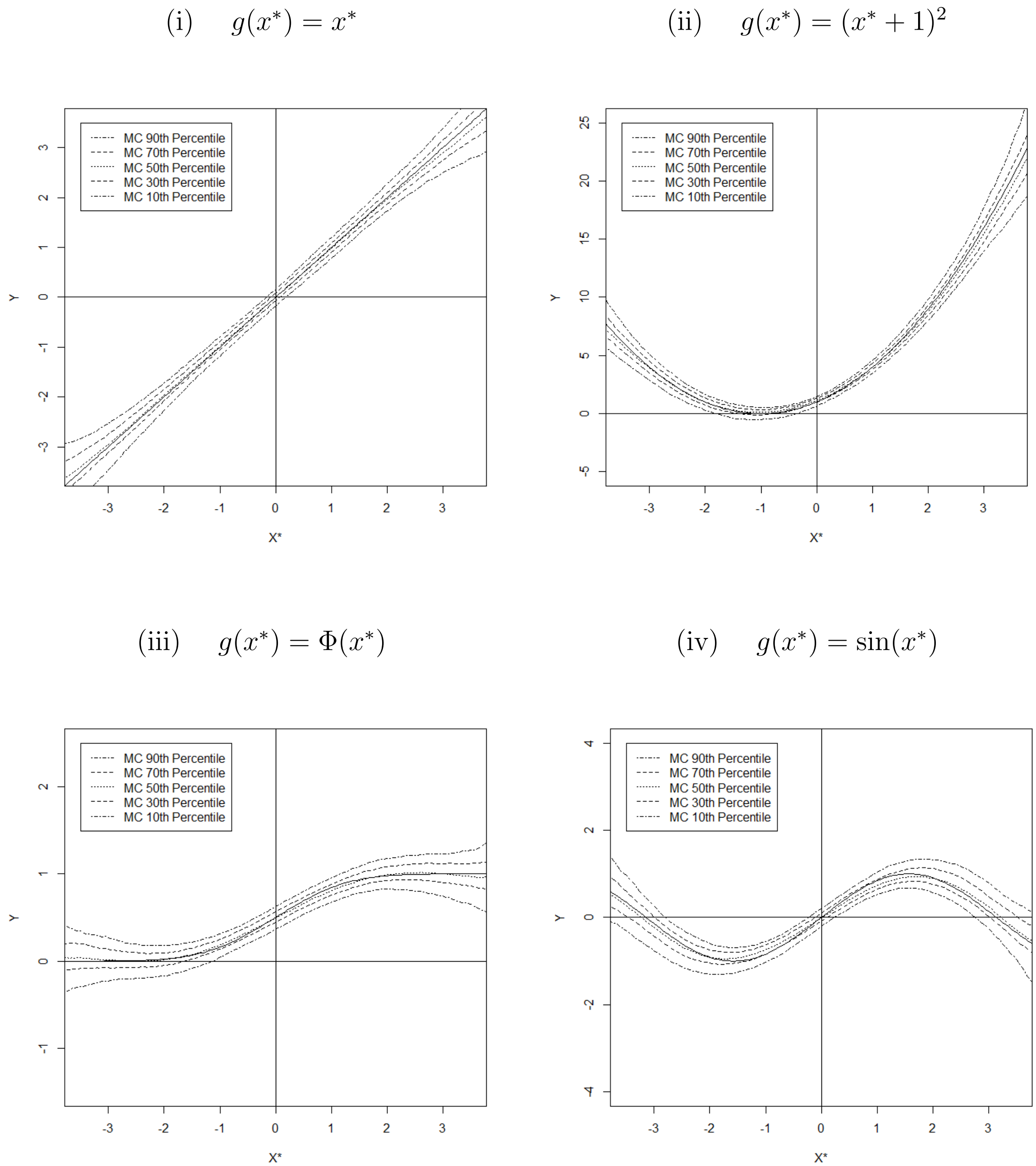

Figure 1: Monte Carlo simulation results with $N = 500$ for the closed-form estimator (4.2) using estimates $\hat{\gamma}_1$. The four function specifications displayed are (i) $g(x^*) = x^*$, (ii) $g(x^*) = (x^*+1)^2$, (iii) $g(x^*) = \Phi(x^*)$, and (iv) $g(x^*) = \sin(x^*)$. The solid curves represent the true function g. The dashed curves are the 10th, 30th, 50th, 70th, and 90th percentiles of the Monte Carlo distributions.

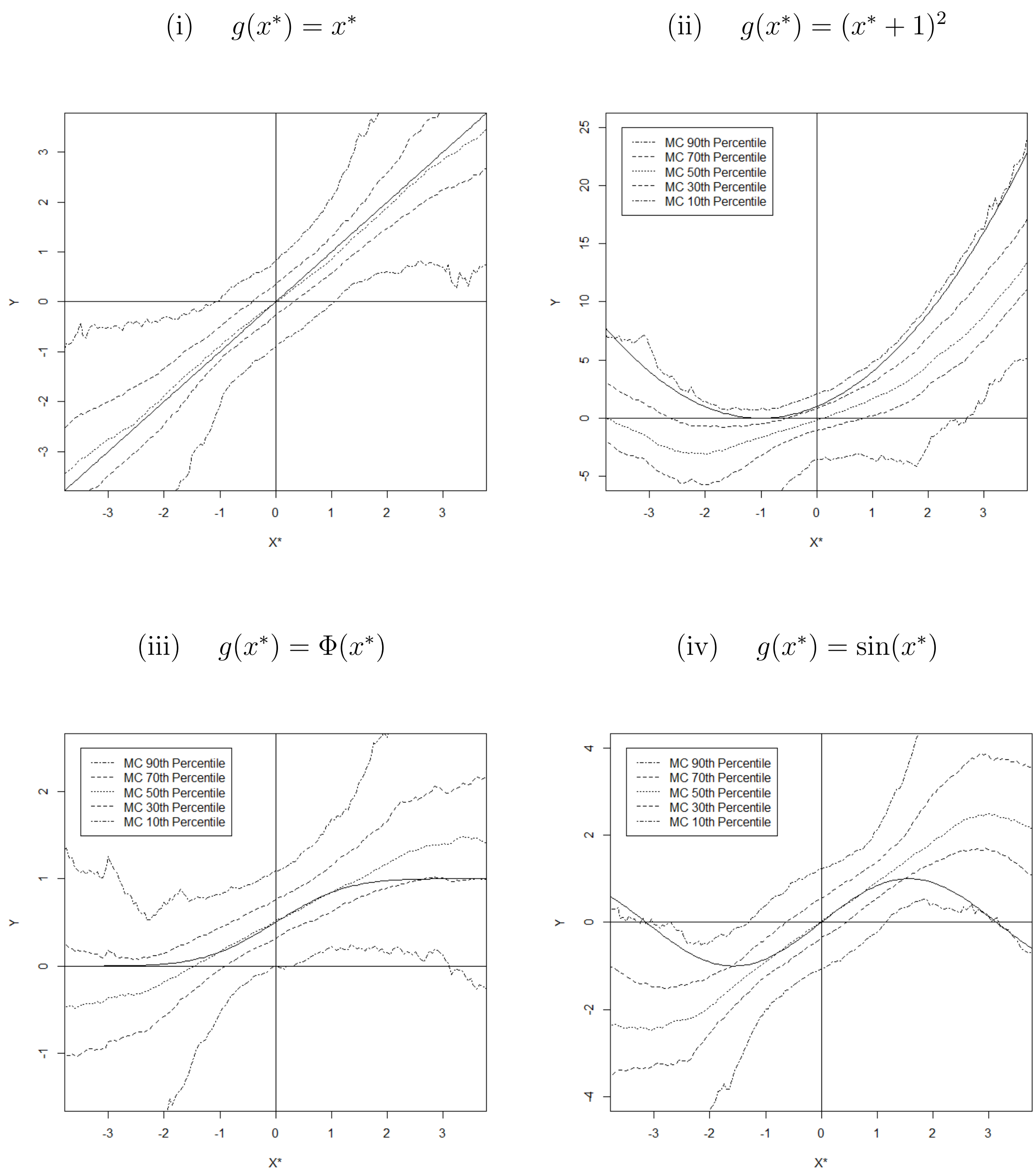

Figure 2: Monte Carlo simulation results with $N = 500$ for the closed-form estimator (4.1) wrongly assuming $\gamma_1 = 1$ is true. The four function specifications displayed are (i) $g(x^*) = x^*$, (ii) $g(x^*) = (x^* + 1)^2$, (iii) $g(x^*) = \Phi(x^*)$, and (iv) $g(x^*) = \sin(x^*)$. The solid curves represent the true function g. The dashed curves are the 10th, 30th, 50th, 70th, and 90th percentiles of the Monte Carlo distributions.

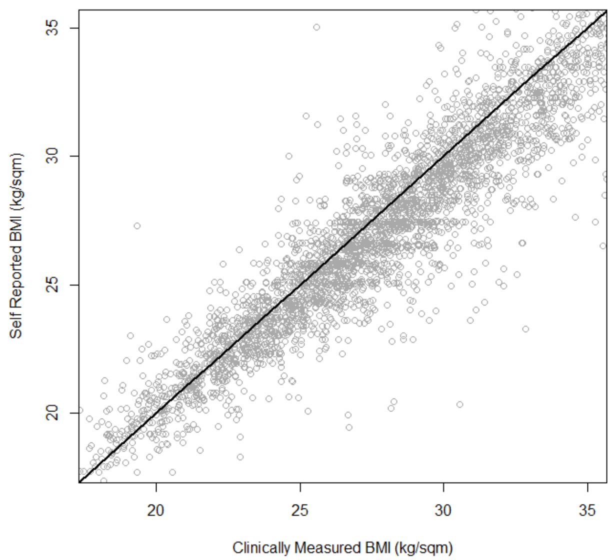

Figure 3: Scatter plot of clinically measured BMI (horizontal axis) against self-reported BMI (vertical axis). 2009–2010 sample of male and female individuals aged 40 or older.

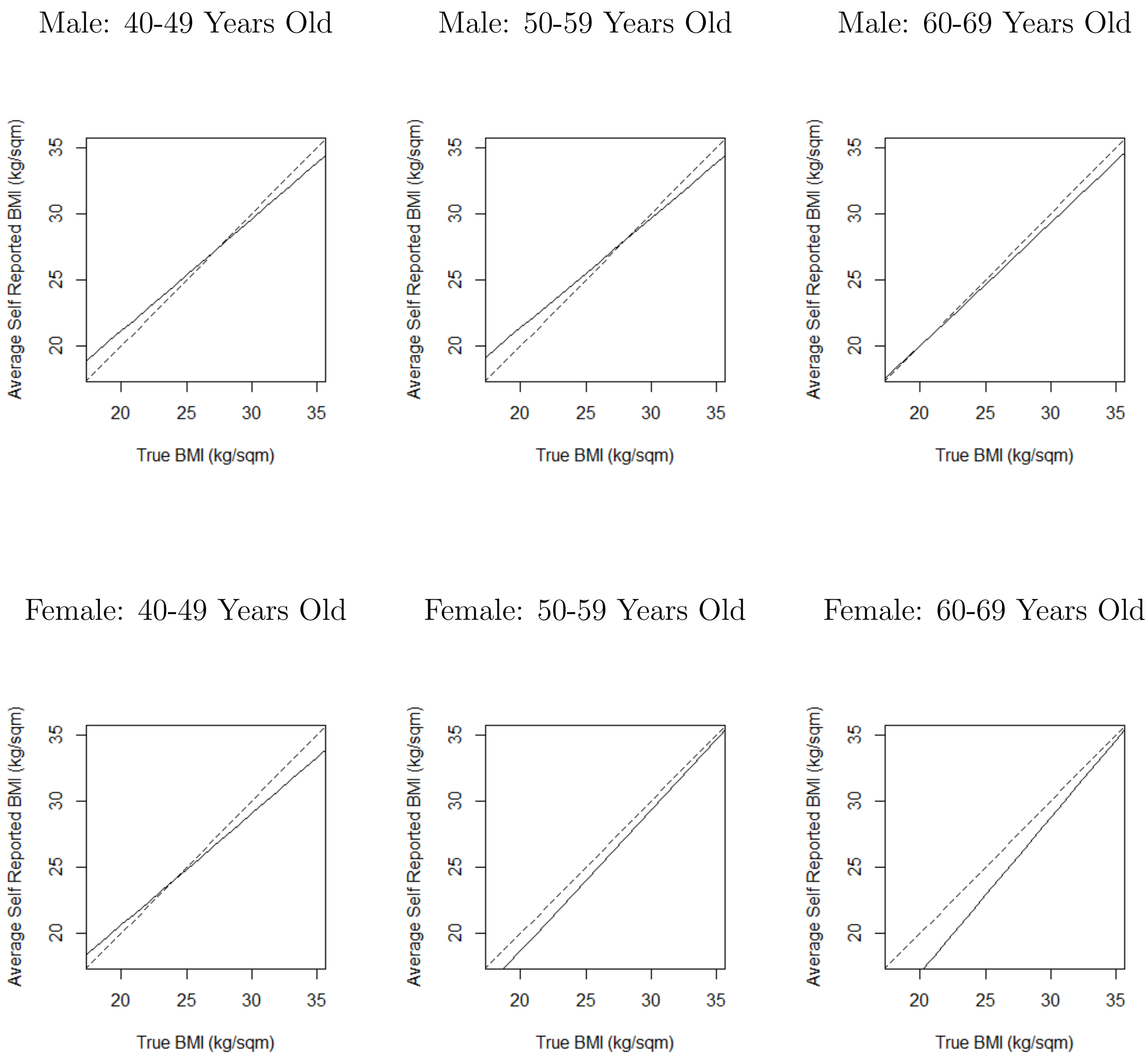

Figure 4: Estimated conditional mean self-reports $\mathrm{E}[X_1 \mid X^*]$ across various sex and age groups. The dashed lines indicate the 45° line, and the solid lines indicate the estimated regression.

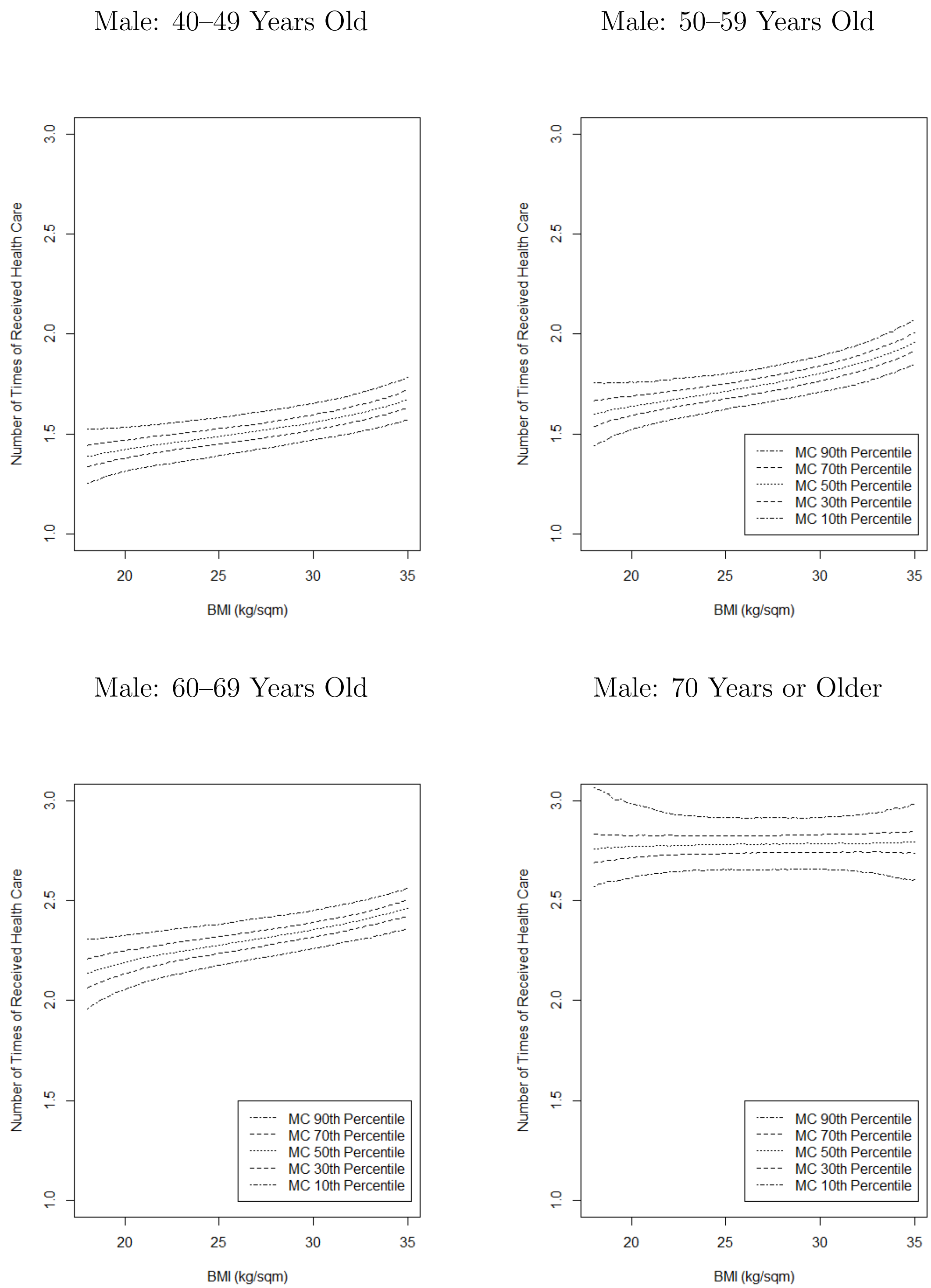

Figure 5: Bootstrap quantiles for average number of times health care is received per year for male individuals by age group and BMI. Estimation is based on (4.2) using $\hat{\gamma}_1$.

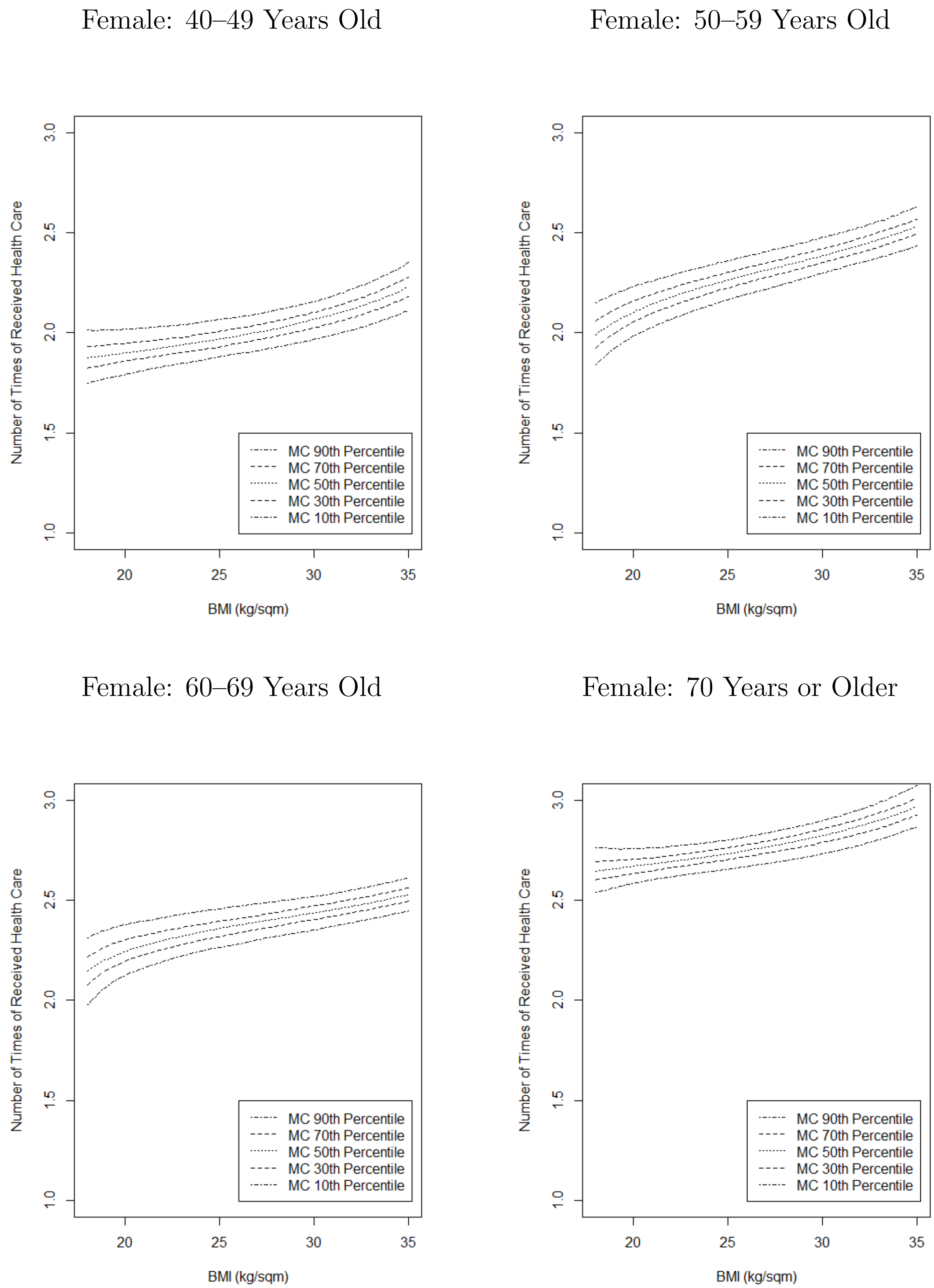

Figure 6: Bootstrap quantiles for average number of times health care is received per year for female individuals by age group and BMI. Estimation is based on (4.2) using $\hat{\gamma}_1$.