Bounding a Class of Parameters in Measurement Error Models under Data Combination*

Angda Li†

October 29, 2025

Abstract

This paper studies measurement error models where the measurement and the true variable are observed in two different datasets that cannot be matched. A common example arises when a noisy measurement is available in survey data, while the true variable is recorded in administrative data. We consider a class of parameters that characterize the structure of measurement error and derive bounds for them by solving linear programming problems. Our framework accommodates a range of identifying assumptions, allowing for flexible correlation between the measurement error and the true variable. We apply our method to examine patterns of underreporting using two datasets on welfare benefits—one containing reported benefits and the other administrative records of actual benefits.

Keywords: Systematic error; Data combination; Linear programming; Partial identification

*I am especially indebted to Yonghong An, Jackson Bunting, and Qi Li for their guidance and support. I am also grateful to Christopher R. Bollinger, Marco Castillo, Li Gan, Jorge Garcia, Manu Navjeevan, Ragan Petrie, Han Xu, and Anna Ziff for their helpful comments.

†Department of Economics, Texas A&M University, 2935 Research Pkwy, College Station, TX 77845. Email: angdali97@tamu.edu. Website: https://angdali97.github.io.

1 Introduction

In this paper, we study measurement error models where the measurement and the true variable are observed in two different datasets that cannot be matched. A common example arises when a noisy measurement is available in survey data, while the true variable is recorded in administrative data, which are typically more accurate and serve as benchmarks for assessing measurement error (see, e.g., Bound et al., 2001). Such datasets may lack unique identifiers for matching, and even when identifiers exist, researchers may face attrition or imperfect linkage during the matching process[1].

Our analysis focuses on a class of parameters that characterize the structure of measurement error. These include measures of systematic misreporting, which capture the average tendency of individuals to overreport or underreport within specific groups—a feature commonly observed in self-reported data (see, e.g., Bollinger, 1998; Blattman et al., 2017). Such parameters provide insight into systematic bias in reporting behavior. Another key parameter is the covariance between the measured and true variables, which reflects the reliability of the measurement. This covariance also plays a crucial role in identifying the ordinary least squares (OLS) coefficient on the true variable. For example, it can be used to estimate the causal effect of the true variable on an outcome variable that is also observed in the survey data, as demonstrated in Section 7.

Since the joint distribution of the measurement and the true variable cannot be recovered from separate datasets, these target parameters are generally not point-identified. Therefore, we develop a framework to characterize and compute their bounds. Our method builds on the general decomposition of measurement error into random and systematic errors, as developed in Li (2025). We introduce a relative measure of systematic error, referred to as the *slope function*, defined as the ratio of the conditional expectation of the measurement, given the true variable, to the true value itself. The slope function is central to our analysis because the target parameter can be expressed as a linear functional of this object, and it captures how the measurement systematically deviates from the true value.

[1]For instance, Bollinger et al. (2019) found that when records from the Current Population Survey are matched with administrative earnings data, unmatched individuals are more likely to be foreign-born, have lower educational attainment, and report lower earnings. This suggests that the attrition during the matching process may be endogenous.

We then show how information about the slope function can be extracted from two observed datasets. Drawing on results related to the mean-preserving spread in D'Haultfoeuille et al. (2021), we derive a set of moment equalities and inequalities that impose linear restrictions on the slope function. Using these restrictions, we formulate linear programming problems—where the slope function serves as the decision variable—that yield bounds on the target parameter. In addition, we provide conditions under which these bounds are sharp, meaning that they are the tightest possible under the maintained assumptions.

We introduce several additional assumptions on the slope function that can tighten the bounds on the target parameters. These include shape restrictions, such as boundedness and monotonicity, as well as restrictions on the functional form. These assumptions are fully nonparametric and intuitively interpretable, which makes them well-suited for empirical applications. Importantly, our framework allows these restrictions to be incorporated directly into the linear programming formulation without requiring new analytical derivations of the bounds.

We adopt the methodology outlined by Mogstad et al. (2018) to develop consistent estimators for our bounds, which can be computed using linear programming. Additionally, we follow their working paper (Mogstad et al., 2017) to develop a corresponding inference procedure. This procedure not only facilitates the construction of valid confidence intervals for the target parameter, but also provides an approach to test various model specifications.

Our method is illustrated using two unmatched datasets on welfare benefits, as in Hu and Ridder (2012): one containing self-reported benefits and the other administrative records of actual benefits. The analysis assumes that individuals tend to underreport the benefits they receive. We find that ignoring the systematic error may lead to an overestimation of the correlation between reported and actual benefit levels. Moreover, when either of two additional shape restrictions on the slope function is imposed, the results suggest that underreporting is prevalent in both the lower and upper tails of the true benefit distribution.

The main contribution of this paper is to develop a new framework for bounding a class of parameters using only the marginal information of the measurement and the true variable. This framework accommodates a wide range of identification assumptions and imposes fewer data requirements than validation datasets that contain both variables for the same units (e.g., Carroll and Wand, 1991; Bound et al., 1994; Lee and Sepanski, 1995; Chen et al., 2005). While Hu and Ridder (2012) also relied on marginal information, that approach assumed independence between the measurement error and the true variable. In contrast, our framework allows the measurement error to be flexibly correlated with the true variable.

Our paper also contributes to the recent literature on data combination in various settings, including linear models (e.g., Pacini, 2019a,b; Hwang, 2023; D'Haultfœuille et al., 2025) and potential outcome models (e.g., Fan and Park, 2010; Fan et al., 2014; Russell, 2021; Fan et al., 2023; Gechter, 2024). The most closely related study is D'Haultfoeuille et al. (2021), who developed a test of rational expectations using only the marginal distributions of realizations and subjective beliefs. Applied to our setting, their approach can be used to test the assumption of "no systematic error," which we discuss in Section 3.2. Our method extends their work by enabling the testing of additional model specifications, such as shape restrictions on the slope function.

Finally, we add to the literature that employs linear programming methods to characterize identified sets, including Torgovitsky (2019), Tebaldi et al. (2023), Kamat (2024), and Kamat and Norris (2025). Our approach extends this line of work by incorporating economically meaningful assumptions that restrict the slope function, which have intuitive interpretations in the context of measurement error. Similar to Mogstad et al. (2018) and Han and Yang (2024), we use sieve approximations to transform the infinite-dimensional problem into a tractable finite-dimensional linear program.

Notation

We use $F_{A|B}(\cdot|b)$ to denote the distribution function of the random variable A conditional on random variable $B = b$. Define conditional quantile function $Q_A(\tau|B = b) = \inf\{a \in \mathbb{R} : F_{A|B}(a|b) \geq \tau\}$. For random vectors C and D, define $C^{\perp D} = C - \mathbb{E}[CD'] \cdot (\mathbb{E}[DD'])^{-1} \cdot D$, which is the residual from a linear projection C on D.

2 Model

2.1 Basic Setup

We consider two available datasets. In the primary dataset, we observe a noisy measurement, denoted by X. In the auxiliary dataset, we observe the true variable, denoted by X^*. In practice, both datasets may also contain additional covariates $Z \in \mathbb{R}^{d_Z}$ in both datasets. We assume that Z includes a constant term. When the primary and auxiliary datasets do not share any common covariates, Z reduces to a constant, i.e., $d_Z = 1$.

Assumption 1. (i) The joint distribution of (X, Z) is identified from the primary dataset. (ii) The joint distribution of (X^*, Z) is identified from the auxiliary dataset.

Assumption 1 (ii) differs from the validation data setting as described in Chen et al. (2005), which consists of pairs (X, X^*) that allow us to recover their joint distribution. In contrast, our primary and auxiliary datasets cannot be matched, which means we can only identify the marginal distributions of X and X^*. In many applications, the auxiliary dataset originates from administrative records; see Hu and Ridder (2012) for several illustrative examples.

We allow both X and X^* to be continuous or discrete, without requiring them to be of the same type. Define measurement error

$$u = X - X^*. \tag{1}$$

The concept of classical measurement error is widely used in the literature. We follow the *weakly classical* definition in Schennach (2022), which states the measurement error u is mean-zero and mean-independent of the true variable X^*. If not, we refer to the measurement error as non-classical, indicating a correlation between u and X^*.

Throughout our analysis, we impose the following regularity condition.

Assumption 2. The joint distribution of (X, X^*, Z) has finite second moments.

2.2 Target Parameters

This paper aims to bound a class of parameters by optimization. Accordingly, the focus is on the scalar parameters of the form[2]:

$$\gamma = \mathbb{E}[X \cdot h(X^*, Z)], \tag{2}$$

where h is an identified (or known) function that is measurable and has a finite second moment. Below are several examples that are crucial for characterizing the structure of measurement error.

Example 1. Covariance between X and X^*

$$\gamma = \mathrm{Cov}(X, X^*) = \mathbb{E}[X \cdot (X^* - \mathbb{E}[X^*])], \tag{3}$$

which fits the form (2) with $h(x^*, z) = x^* - \mathbb{E}[X^*]$. We will revisit this parameter in the linear regression context in Section 7. Similarly, the correlation between X and X^* can be written as

$$\mathrm{Corr}(X, X^*) = \mathbb{E}\left[X \cdot \frac{X^* - \mathbb{E}[X^*]}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(X^*)}}\right],$$

which also fits the form (2) with $h(x^*, z) = (x^* - \mathbb{E}[X^*]) \Big/ \sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(X^*)}$.

The next two examples measure the amount of systematic error in the measurement (see Section 3.1 for a formal definition of systematic error), and we explain them in the context of self-reported data.

Example 2. The conditional average reporting ratio, given that the true value X^* falls within $\mathcal{X}$ and the covariates Z fall within $\mathcal{Z}$

$$\gamma = \frac{\mathbb{E}[X \mid X^* \in \mathcal{X}^*, Z \in \mathcal{Z}]}{\mathbb{E}[X^* \mid X^* \in \mathcal{X}^*, Z \in \mathcal{Z}]} = \mathbb{E}\left[X \cdot \left(\frac{\mathbf{1}\{X^* \in \mathcal{X}^*, Z \in \mathcal{Z}\}}{\mathbb{E}[X^* \cdot \mathbf{1}\{X^* \in \mathcal{X}^*, Z \in \mathcal{Z}\}]}\right)\right], \tag{4}$$

[2]Since the conditional distribution $F_{X^*|X,Z}$ is identified within the framework of either Chen et al. (2005) or Hu and Ridder (2012), they allow for the point identification of more general parameters such as $\mathbb{E}[\tilde{h}(X, X^*, Z)]$, where $\tilde{h}(\cdot)$ is a known (or identified) function.

which fits the form (2) with $h(x^*, z) = \frac{\mathbf{1}\{x^* \in \mathcal{X}^*, z \in \mathcal{Z}\}}{\mathbb{E}[X^* \mathbf{1}\{X^* \in \mathcal{X}^*, Z \in \mathcal{Z}\}]}$. This parameter reflects the average level of overreporting or underreporting within the conditional group defined by $\mathcal{X}$ and $\mathcal{Z}$. A value greater than 1 indicates an average of overreporting, while a value less than 1 indicates an average of underreporting within the specified subpopulation.

Example 3. The expected individual-level reporting ratio, given that the true value X^* falls within $\mathcal{X}$ and the covariates Z fall within $\mathcal{Z}$

$$\gamma = \mathbb{E}\left[\frac{X}{X^*}\bigg| X^* \in \mathcal{X}^*, Z \in \mathcal{Z}\right] = \mathbb{E}\left[X \cdot \frac{\mathbf{1}\{X^* \in \mathcal{X}^*, Z \in \mathcal{Z}\}}{X^* \cdot \Pr(X^* \in \mathcal{X}^*, Z \in \mathcal{Z})}\right], \tag{5}$$

which fits the form (2) with $h(x^*, z) = \frac{\mathbf{1}\{x^* \in \mathcal{X}^*, z \in \mathcal{Z}\}}{x^* \cdot \Pr(X^* \in \mathcal{X}^*, Z \in \mathcal{Z})}$. This parameter reflects the expected degree of overreporting or underreporting within the conditional group defined by $\mathcal{X}$ and $\mathcal{Z}$.

2.3 Bounds Using Only Marginal Distributions

Our primary dataset and auxiliary dataset do not provide enough information to point identify the target parameter $\gamma = \mathbb{E}[X \cdot h(X^*, Z)]$, since the joint distribution of (X, X^*, Z) is not identified. However, even without any assumptions on measurement error, γ is still constrained by the data. Using monotone rearrangement inequality[3], we provide its bounds in the following proposition.

Proposition 1. Suppose Assumption 1 and 2 hold, we have sharp bounds:

$$\mathbb{E}\left[\int_0^1 Q_X(1-\tau|Z) Q_H(\tau|Z) d\tau\right] \leq \mathbb{E}[X \cdot h(X^*, Z)] \leq \mathbb{E}\left[\int_0^1 Q_X(\tau|Z) Q_H(\tau|Z) d\tau\right], \tag{6}$$

where the random variable $H = h(X^*, Z)$. These bounds are finite.

Bounds in Proposition 1 are sharp when only the marginal distributions of (X, Z) and (X^*, Z) are known. The upper bound attains if X and H are perfectly positively assortatively matched conditional on Z, while X and H are perfectly negatively assortatively matched conditional on Z. As a result, these bounds can be pretty wide in practice. Consider Example

[3]For the application of the monotone rearrangement inequality in data combination, see Fan et al. (2014) and Hwang (2023).

1 ($H = X^* - \mathbb{E}[X^*]$), since X is likely to be positively correlated with X^*, the lower bound in (6) will not be informative. On the other hand, the upper bound in (6) will be large since it ignores the noise in the measurement X. Tighter bounds can be established by incorporating assumptions on the measurement error, such as those discussed in Section 3.5.

3 Computing Bounds with General Measurement Error

3.1 Systematic Error and Slope Function

Our framework considers general measurement error, allowing for flexible correlation between measurement error u and (X^*, Z). To achieve this, we decompose the measurement error into two components:

$$u = \underbrace{\Big(X - \mathbb{E}[X \mid X^*, Z]\Big)}_{\text{random error } \tilde{u}} + \underbrace{\Big(\mathbb{E}[X \mid X^*, Z] - X^*\Big)}_{\text{systematic error}}. \tag{7}$$

This decomposition is unique and does not depend on any assumptions regarding measurement error; see Li (2025) for a related formulation in linear IV models. By construction, the random error $\tilde{u}$ is mean independent of (X^*, Z) and typically unavoidable. Hence, taking multiple measurements can reduce the effect of random error. On the other hand, systematic error, which reflects the deviations from the true variable X^* in the conditional mean of the measurement given (X^*, Z), can cause the measurement to be consistently higher or lower than the true value. This type of error is usually difficult to assess and identify.

Under the strong assumption of "No systematic error" ($\mathbb{E}[X \mid X^*, Z] = X^*$), the target parameter is point-identified as

$$\gamma = \mathbb{E}[\mathbb{E}[X \mid X^*, Z] \cdot h(X^*, Z)] = \mathbb{E}[X^* h(X^*, Z)],$$

which implies that the information from auxiliary data alone suffices to identify the target parameter. It is worth noting that the assumption of "no systematic error " implies that measurement error u is classical and uncorrelated with covariates Z. However, these conditions are strong and increasingly questioned in empirical research. In practice, systematic misreporting is often prevalent in self-reported data, resulting in non-classical measurement error (see, e.g., Bollinger, 1998; Blattman et al., 2017; Abay et al., 2019; Bick et al., 2022). Moreover, empirical evidence shows that measurement error can be correlated with demographic variables such as education, age, or income (see, e.g., Bound and Krueger, 1991; Bound et al., 1994; Angel et al., 2019). These concerns underscore the importance of accounting for systematic error in empirical studies.

We then introduce a relative measure of the systematic error. Let Ω denote the support of (X^*, Z). Define *slope function* on $\Omega_0 = \{(x^*, z) \in \Omega : x^* \neq 0\}$:

$$\lambda(x^*, z) = \frac{\mathbb{E}[X \mid X^* = x^*, Z = z]}{x^*}, \tag{8}$$

which measures how the conditional expectation of measurement X deviates from the true value X^*, conditional on covariates Z. No systematic error means that $\lambda(x^*, z) = 1$. Therefore, we consider the presence of systematic error by allowing the slope function to differ from one. In self-reported data, the slope function reflects the average rate of underreporting or overreporting among individuals with a given true value x^* (and covariates z). Overreporting implies that $\lambda(x^*, z) > 1$, while underreporting indicates that $\lambda(x^*, z) < 1$.

Assumption 3. One of the following holds: (i) $\Pr(X^* = 0) = 0$. (ii) $\mathbb{E}[X \mid X^* = 0, Z] = 0$.

Assumption 3 implies there is no point mass or no systematic error at $X^* = 0$, which ensures a key equation holds almost surely[4]:

$$X = \lambda(X^*, Z) \cdot X^* + \tilde{u}, \tag{9}$$

where $\tilde{u}$ is the random error defined in (7). Equation (9) connects the measurement with the true variable through the slope function. By the law of iterated expectation, our target

[4]For $x^* = 0$, $\lambda(x^*, z)$ can take any positive number.

parameter can be expressed as a linear functional of the slope function:

$$\gamma = \mathbb{E}[\mathbb{E}[X \mid X^*, Z] \cdot h(X^*, Z)] = \mathbb{E}[\lambda(X^*, Z) \cdot X^* h(X^*, Z)]. \tag{10}$$

In Appendix C.1, we present a straightforward modification of equation (9) and the corresponding bound calculations for the case in which Assumption 3 does not hold.

3.2 What We Know about the Slope Function

We now show how to extract information about the slope function from two datasets. For example, consider the expectation from both sides in equation (9),

$$\mathbb{E}[\lambda(X^*, Z) \cdot X^*] = \mathbb{E}[X],$$

which is a linear constraint for the slope function. In the context of self-reported data, it restricts the extent of overreporting or underreporting so that two moments derived from primary and auxiliary datasets should be equal. Following proposition 2 in D'Haultfoeuille et al. (2021), the next lemma provides more information about the slope function.

Lemma 1. Suppose Assumptions 1, 2 and 3 hold, we have

$$\mathbb{E}\Big[\lambda(X^*, Z) \cdot X^* \mid Z\Big] = \mathbb{E}[X \mid Z], \tag{11}$$

$$\mathbb{E}\left[(x - \lambda(X^*, Z) \cdot X^*)^+ \mid Z\right] \leq \mathbb{E}[(x - X)^+ \mid Z], \quad \forall x \in \mathbb{R}, \tag{12}$$

where $a^+ = \max(0, a)$.

The proof of Lemma 1 uses the fact that the distribution of random variable X is a mean-preserving spread of the distribution of $\mathbb{E}[X \mid X^*, Z] = \lambda(X^*, Z) \cdot X^*$, as implied by equation (9). Intuitively, the measurement X has the same mean but greater dispersion due to additional noise. The statistical test proposed by D'Haultfoeuille et al. (2021) enables us to test both moment equality (11) and moment inequalities (12) for any specified slope function. In particular, the assumption of "no systematic error" can be tested with $\lambda(x^*, z) = 1$.

According to Lemma 1, the following proposition presents unconditional moment conditions for the slope function. It is important to note that both the moment equality (13) and moment inequality (14) are linear in the slope function, which facilitates the computation later.

Proposition 2. Suppose Assumptions 1, 2 and 3 hold. Suppose that $t \in \mathbb{R}$ and $s : \mathbb{R}^{dz} \to \mathbb{R}^+ \cup \{0\}$ is an identified (or known) function that is measurable and has a finite second moment. Define indicator function $d_t(u) = \mathbf{1}\{u \leq t\}$. Then, we have

$$\mathbb{E}\Big[\lambda(X^*, Z) \cdot s(Z)X^*\Big] = \mu_s, \tag{13}$$

$$\mathbb{E}\Big[\lambda(X^*, Z) \cdot s(Z)X^* \cdot d_t(X^*)\Big] \geq \delta_{s,t}, \tag{14}$$

where $\mu_s = \mathbb{E}[s(Z)X]$ and $\delta_{s,t} = \sup_{x \in \mathbb{R}} \mathbb{E}\left[s(Z)\Big(x \cdot d_t(X^*) - (x - X)^+\Big)\right]$ are point-identified.

Moment equality (13) follows directly from (11) by using $s(Z)$ as an instrument. We explain how to derive linear moment inequality (14) from nonlinear moment inequalities (12) with $s(z) = 1$ below. Consider the expectation of (12), we have (for a specific $t \in \mathbb{R}$):

$$\begin{aligned} \mathbb{E}[(x - X)^+] &\geq \mathbb{E}\Big[(x - \lambda(X^*, Z) \cdot X^*)^+\Big] \\ &\geq \mathbb{E}\Big[(x - \lambda(X^*, Z) \cdot X^*) \cdot d_t(X^*)\Big], \end{aligned} \tag{15}$$

where the second step uses the fact that $a^+ \geq a \cdot d$ for any real number a and $d \in \{0, 1\}$. Simple algebra yields

$$\mathbb{E}\Big[\lambda(X^*, Z) \cdot X^* \cdot d_t(X^*)\Big] \geq \mathbb{E}\left[\Big(x \cdot d_t(X^*) - (x - X)^+\Big)\right]. \tag{16}$$

Note that the left side does not contain x; we can take the supremum of the right side over $x \in \mathbb{R}$ and thus verify linear moment inequality (14) with $s(z) = 1$. As a result, we generate many linear moment inequalities for different thresholds $t \in \mathbb{R}$.

3.3 Identified Set of the Target Parameter

We assume researchers have restricted the slope function to lie in some parameter space $\mathcal{L}$. Researchers can incorporate shape or functional form restrictions into the parameter space (as discussed in Section 3.5). We aim to establish the bounds on the target parameter γ that could be generated by slope functions $\lambda \in \mathcal{L}$ that also satisfy a collection of conditions in Proposition 2.

Let $\mathcal{S}$ denote a collection of function s described in Proposition 2 and $\mathcal{T} \subseteq \mathbb{R}$. As a result, slope function λ must lie in the set

$$\mathcal{L}_{\mathcal{S},\mathcal{T}} = \{\lambda \in \mathcal{L} : \text{ (13) and (14) hold for all } s \in \mathcal{S} \text{ and } t \in \mathcal{T}\}. \tag{17}$$

Recall that the target parameter γ can be expressed as an identified linear map of the slope function. From (10), we define this linear map as $\Gamma^\star : \mathcal{L} \to \mathbb{R}$, with

$$\Gamma^\star(\lambda) \equiv \mathbb{E}[\lambda(X^*, Z) \cdot h(X^*, Z)X^*]. \tag{18}$$

Then, the target parameter must belong to the identified set

$$\mathcal{G}_{\mathcal{S},\mathcal{T}} = \{\gamma \in \mathbb{R} : \ \gamma = \Gamma^\star(\lambda) \text{ for some } \lambda \in \mathcal{L}_{\mathcal{S},\mathcal{T}}\}.$$

The following result shows that if $\mathcal{L}$ is convex, then $\mathcal{G}_{\mathcal{S},\mathcal{T}}$ is an interval whose endpoints can be characterized by solving two convex optimization problems.

Proposition 3. Suppose Assumptions 1, 2 and 3 hold. Suppose that $\mathcal{L}$ is convex and $\mathcal{L}_{\mathcal{S},\mathcal{T}}$ is non-empty. Then, the closure of $\mathcal{G}_{\mathcal{S},\mathcal{T}}$ is an interval $[\underline{\gamma}^\star, \overline{\gamma}^\star]$, where

$$\underline{\gamma}^\star = \inf_{\lambda \in \mathcal{L}_{\mathcal{S},\mathcal{T}}} \Gamma^\star(\lambda) \quad \text{and} \quad \overline{\gamma}^\star = \sup_{\lambda \in \mathcal{L}_{\mathcal{S},\mathcal{T}}} \Gamma^\star(\lambda). \tag{19}$$

3.4 Information and Point Identification

The set $\mathcal{L}_{\mathcal{S},\mathcal{T}}$ consists of all slope functions in $\mathcal{L}$ that are consistent with conditions in Proposition 2 chosen by researchers. However, $\mathcal{L}_{\mathcal{S},\mathcal{T}}$ may not exhaust all of the information available in the data. We examine the issue by considering the sharp identified set for the slope function below:

$$\mathcal{L}_{\text{IS}} = \Big\{ \lambda \in \mathcal{L} : \text{there exists random variables } (\tilde{X}, \tilde{X}^*, \tilde{Z}) \text{ such that} \\ (\tilde{X}, \tilde{Z}) \overset{d}{=} (X, Z),\ (\tilde{X}^*, \tilde{Z}) \overset{d}{=} (X^*, Z),\ \lambda(x^*, z) = \frac{\mathbb{E}[\tilde{X} \mid \tilde{X}^* = x^*, \tilde{Z} = z]}{x^*} \Big\}, \tag{20}$$

which include all slope functions that are compatible with marginal distributions of (X, Z) and (X^*, Z) and the researcher's assumptions.

The following proposition shows, under a monotonic condition, conditions in Lemma 1 characterize the sharp identified set $\mathcal{L}_{\text{IS}}$. Moreover, we provide conditions under which $\mathcal{L}_{\mathcal{S},\mathcal{T}}$ will include only slope functions that are compatible with the observed data.

Assumption 4. For all z in its support, $\mathbb{E}[X \mid X^* = x^*, Z = z]$ is strictly increasing in x^*.

Proposition 4. Suppose Assumptions 1, 2, 3 and 4 hold. Suppose that every $\lambda \in \mathcal{L}$ satisfies $\mathbb{E}[\lambda(X^*, Z)X^*]^2 < \infty$. Then,

$$\mathcal{L}_{\text{IS}} = \{\lambda \in \mathcal{L} :\ \lambda \text{ satisfies (11) and (12)}\} \subseteq \mathcal{L}_{\mathcal{S},\mathcal{T}}, \tag{21}$$

where $\mathcal{S}$ include functions that satisfy conditions of Proposition 2 and $\mathcal{T} \subseteq \mathbb{R}$. Moreover, suppose that $\mathcal{T}$ is dense in $\text{supp}(X^*)$ and includes all the discontinuities of $F_{X^*}(\cdot)$, and either of these two conditions hold:

(i) Z has a discrete support $\{z_1, z_2, \cdots, z_m\}$ and $\mathcal{S} = \{\mathbf{1}\{z = z_j\} : 1 \leq j \leq m\}$.

(ii) $\mathbb{E}[X \mid X^*, Z] = \mathbb{E}[X \mid X^*]$ and the linear span of $\mathcal{S}$ is norm dense in $L^2(Z) \equiv \{f : \mathbb{R}^{d_Z} \to \mathbb{R} :\ \mathbb{E}[f(Z)^2] < \infty\}$.

Then, $\mathcal{L}_{\text{IS}} = \mathcal{L}_{\mathcal{S},\mathcal{T}}$.

Assumption 4 is natural for self-reported data, as it requires only that individuals with higher true values tend to report higher values on average. Even if reports are biased or noisy, this weak monotonicity ensures that the expected report preserves the ordering of the true values, a reasonable condition in many empirical settings.

Under Assumption 4, Proposition 4 shows that if functions $\mathcal{S}$ and combinations $\mathcal{T}$ are sufficiently rich, $\mathcal{L}_{\mathcal{S},\mathcal{T}}$ will exhaust all the information in the primary and auxiliary datasets when Z is discrete. Consequently, $\mathcal{G}_{\mathcal{S},\mathcal{T}}$ is the sharp identified set for values of the target parameter that are consistent with marginal distributions of (X, Z) and (X^*, Z) and the assumptions of the model. When Z is continuous, the same conclusion holds under the assumption of conditional mean independence. The class of functions $\mathcal{S}$ can be the set of half-spaces $\{\mathbf{1}\{z \leq z^\dagger\} : z^\dagger \in \mathbb{R}^{d_Z}\}$. When Assumption 4 does not hold, it is possible to add additional moment linear inequalities derived from (12) to extract more information from the observed data, which we discuss in Appendix C.2.

While we view partial identification as the general case, our analysis does not preclude point identification. Proposition 5 below presents two cases under which the slope function can be identified, which also implies the point identification of the target parameter.

Proposition 5. Suppose Assumptions 1, 2 and 3 hold. Suppose either of these two conditions holds:

(i) $\lambda(x^*, z) = \lambda_0(z)$ and $\mathbb{E}[X^* \mid Z = z] \neq 0$ for all $(x^*, z) \in \Omega$.

(ii) $\lambda(x^*, z) = \sum_{m=1}^{M} \alpha_m \cdot \rho_m(x^*, z)$ where $\{\rho_m(\cdot)\}_{m=1}^{M}$ is a set of known functions defined on $\mathbb{R}^{1+d_Z}$ and $\{\alpha_m\}_{m=1}^{M}$ are unknown coefficients. Also, there exists $\{z_1, \cdots, z_M\} \in \text{supp}(Z)$ such that $M \times M$ matrix F with entries $\mathbb{E}[\rho_i(X^*, Z)X^* \mid Z = z_j]$ is full rank.

Then, the slope function and the target parameter are point-identified.

The first case of Proposition 5 assumes that the slope function relies solely on common covariates Z, and the point identification result follows from moment equality (11):

$$\lambda_0(z) = \frac{\mathbb{E}[X \mid Z = z]}{\mathbb{E}[X^* \mid Z = z]}.$$

The second case restricts the slope function to be a linear combination of known functions (e.g., polynomials) and requires sufficient variations of Z to identify those linear coefficients. For example, if Z is discrete, it requires that the support of Z is no smaller than M (number of basis functions). In these two cases of point identification, we can apply the statistical test proposed by D'Haultfoeuille et al. (2021) to test the model specification, as explained in Section 3.2.

3.5 Identifying Assumptions of the Slope Function

In this subsection, we propose identifying assumptions that can be imposed on the slope function to narrow bounds for the target parameter. These assumptions are implemented by including restrictions on the parameter space $\mathcal{L}$ and then applying Proposition 3.

3.5.1 Bounds of the Slope Function

We restrict the extent of systematic error in the following assumption.

Assumption 5. There exists known parameters $\lambda_u \geq \lambda_l > 0$ such that

$$\lambda_l \leq \lambda(x^*, z) \leq \lambda_u, \text{ for all } (x^*, z) \in \Omega.$$

Assumption 5 states that the lower and upper bounds of the slope function (λ_l and λ_u) are known to researchers. Since $\lambda_l = \lambda_u = 1$ implies no systematic error ($\mathbb{E}[X \mid X^*, Z] = X^*$), this assumption allows for systematic error by allowing λ_l, λ_u to differ from 1, but not too different. In self-reported data, we can interpret (λ_l, λ_u) based on the extent of underreporting and overreporting. If subjects tend to overreport the true outcomes, we can set $\lambda_l = 1$ and utilize prior knowledge to calibrate λ_u. In general, if subjects report at most $\psi_u\%$ above true outcomes and at most $\psi_l\%$ below true outcomes on average, we set $\lambda_u = 1 + \psi_u\%$ and $\lambda_l = 1 - \psi_l\%$.

When researchers are unsure about (λ_l, λ_u), Proposition 3 can be used to perform sensitivity analysis regarding the extent of systematic error. By considering (λ_l, λ_u) as sensitivity parameters, we can explore how the bounds of the target parameter change with them.

3.5.2 Monotonicity and Concavity

Researchers can impose common shape restrictions on the slope function, including monotonicity and concavity.

Assumption 6. For all z in its support, $\lambda(x^*, z)$ is weakly increasing in x^*.

Assumption 7. For all z in its support, $\lambda(x^*, z)$ is weakly concave in x^*.

In the context of overreporting, Assumption 6 indicates that the average degree of overreporting increases with the true variable. Angel et al. (2019) discovered that respondents at the lower end of the wage distribution tend to overreport their earnings, while those at the upper end tend to underreport. This pattern may be consistent with a decreasing slope function. Additionally, Assumption 7 indicates that the rate of overreporting increases at a slower pace as the true values rise.

Assumption 4 is also a shape restriction on the slope function, which is weaker than Assumption 6. If $\lambda(x^*, z)$ is weakly decreasing in x^*, Assumption 4 constrains the rate of decreasing in such a way that $\mathbb{E}[X \mid X^* = x^*, Z = z] = \lambda(x^*, z) \cdot x^*$ remains an increasing function of x^*.

3.5.3 Functional Form Restrictions

Conditional mean independence ($\mathbb{E}[X \mid X^*, Z] = \mathbb{E}[X \mid X^*]$) is a common and convenient assumption in measurement error models. It implies that, once the true variable X^* is given, the covariates Z provide no additional information about the expected value of the mismeasured variable X. This condition requires the slope function to satisfy the following equation:

$$\lambda(x^*, z) = \tilde{\lambda}(x^*), \tag{22}$$

where $\tilde{\lambda}(\cdot)$ is a function that depends only on the true variable X^*.

Another type of functional form restriction is separability between X^* and Z, where the slope function can be decomposed as

$$\lambda(x^*, z) = \lambda_1(x^*) + \lambda_2(z), \tag{23}$$

where λ_1 and λ_2 are functions that can themselves satisfy some shape restrictions. This type of separability implies that the derivatives of the slope function with respect to x^* do not vary with z. Specifications (22) and (23) can also be used to mitigate the curse of dimensionality.

4 Implementation

4.1 Sieve Approximation

Directly solving the optimization problem (19) is often infeasible in practice because the function space $\mathcal{L}_{\mathcal{S},\mathcal{T}}$ is potentially infinite-dimensional. To address this challenge, we approximate (19) by finite-dimensional problems, using a sieve approximation of the slope function $\lambda(x^*, z)$ (see also Mogstad et al., 2018; Han and Yang, 2024).

Suppose that we approximate the slope function with a finite basis:

$$\lambda(x^*, z) \approx \sum_{k=1}^{K} \theta_k b_k(x^*, z), \tag{24}$$

where $\{b_k(x^*, z)\}_{k=1}^{K}$ are known basis functions and $\theta = (\theta_1, \cdots, \theta_K)$ is a finite-dimensional parameter. The parameter space $\mathcal{L}$ generate a parameter space

$$\Theta = \left\{ \theta \in \mathbb{R}^K : \sum_{k=1}^{K} \theta_k b_k(x^*, z) \in \mathcal{L} \right\}.$$

Then, we can formulate the following finite-dimensional optimization problem that corresponds to the upper bound in (19):

$$\begin{aligned}
\overline{\gamma}^{\star}_{\text{fd}} \equiv \sup_{\theta \in \Theta} \; & \sum_{k=1}^{K} \theta_k \cdot \mathbb{E}\Big[b_k(X^*, Z) \cdot h(X^*, Z) X^* \Big] \\
\text{s. t.} \; & \sum_{k=1}^{K} \theta_k \cdot \mathbb{E}\Big[b_k(X^*, Z) \cdot s(Z) X^* \Big] = \mu_s, \;\; \forall s \in \mathcal{S} \\
& \sum_{k=1}^{K} \theta_k \cdot \mathbb{E}\Big[b_k(X^*, Z) \cdot s(Z) X^* \cdot d_t(X^*) \Big] \geq \delta_{s,t}, \;\; \forall s \in \mathcal{S} \text{ and } t \in \mathcal{T},
\end{aligned} \tag{25}$$

where $\underline{\gamma}^{\star}_{\text{fd}}$ can be defined analogously. If Θ is a polyhedral set, then (25) is a finite-dimensional linear program. When using Bernstein polynomials as the sieve basis, shape restrictions on $\mathcal{L}$ can easily translate into Θ; see Appendix B.1 for more detials.

4.2 Estimation

So far we have assumed that researchers know the joint distributions of (X, Z) and (X^*, Z). In practice, researchers observe two random samples $\{X_i, Z_{p,i}\}_{i=1}^{n_p}$ and $\{X_i^*, Z_{a,i}\}_{i=1}^{n_a}$. Let $n = 2n_p n_a/(n_p + n_a)$.

We follow Mogstad et al. (2018) to construct consistent estimators for $\underline{\gamma}^\star$ and $\overline{\gamma}^\star$. For any fixed $s \in \mathcal{S}$, define a map $\Gamma_{1,s}(\cdot) : \mathcal{L} \to \mathbb{R}$

$$\Gamma_{1,s}(\lambda) \equiv \mathbb{E}\Big[\lambda(X^*, Z) \cdot s(Z)X^*\Big].$$

For any fixed $s \in \mathcal{S}$ and $t \in \mathcal{T}$, define map $\Gamma_{2,s,t}(\cdot) : \mathcal{L} \to \mathbb{R}$:

$$\Gamma_{2,s,t}(\lambda) \equiv \mathbb{E}\Big[\lambda(X^*, Z) \cdot s(Z)X^* \cdot d_t(X^*)\Big].$$

Suppose we have consistent estimators $\hat{\mu}_s$ for μ_s, $\hat{\delta}_{s,t}$ for $\delta_{s,t}$, $\hat{\Gamma}_{1,s}$ for $\Gamma_{1,s}$, $\hat{\Gamma}_{2,s,t}$ for $\Gamma_{2,s,t}$ and $\hat{\Gamma}^\star$ for $\Gamma^\star$. We discuss the construction of these estimators in Appendix B.2. We approximation $\mathcal{L}$ with a finite-dimensional subset $\mathcal{L}_n$. We introduce nonnegative slackness variables $\{\eta_{s,t}\}_{(s,t)\in\mathcal{S}\times\mathcal{T}}$, each satisfying $\eta_{s,t} \geq 0$. Define

$$\mathcal{E} = \big\{\eta : \mathcal{S} \times \mathcal{T} \to [0, \infty) \mid \sup_{(s,t)\in\mathcal{S}\times\mathcal{T}} \eta_{s,t} < \infty\big\}.$$

Then, we can redefine the set $\mathcal{L}_{\mathcal{S},\mathcal{T}}$ using only moment equalities:

$$\mathcal{R}_{\mathcal{S},\mathcal{T}} = \{(\lambda, \eta) \in \mathcal{L} \times \mathcal{E} : \Gamma_{1,s}(\lambda) = \mu_s,\ \Gamma_{2,s,t}(\lambda) - \eta_{s,t} = \delta_{s,t}, \forall s \in \mathcal{S}, t \in \mathcal{T}\}.$$

Define

$$Q_n(\lambda, \eta) = \sum_{s\in\mathcal{S}} |\hat{\Gamma}_{1,s}(\lambda) - \hat{\mu}_s| + \sum_{(s,t)\in\mathcal{S}\times\mathcal{T}} |\hat{\Gamma}_{2,s,t}(\lambda) - \eta_{s,t} - \hat{\delta}_{s,t}|$$

and

$$\hat{\mathcal{R}}_{\mathcal{S},\mathcal{T}} = \{(\lambda, \eta) \in \mathcal{L}_n \times \mathcal{E} : Q_n(\lambda, \eta) \leq \inf_{(\lambda,\eta)\in\mathcal{L}_n\times\mathcal{E}} Q_n(\lambda, \eta) + \kappa_n\}, \tag{26}$$

where $\kappa_n > 0$ is a tuning parameter that converges to zero. Define estimators for $\underline{\gamma}^\star$ and $\overline{\gamma}^\star$:

$$\hat{\underline{\gamma}}^\star \equiv \inf_{(\lambda,\eta)\in\hat{\mathcal{R}}_{\mathcal{S},\mathcal{T}}} \hat{\Gamma}^\star(\lambda) \text{ and } \hat{\overline{\gamma}}^\star \equiv \sup_{(\lambda,\eta)\in\hat{\mathcal{R}}_{\mathcal{S},\mathcal{T}}} \hat{\Gamma}^\star(\lambda).$$

We account for statistical error from estimation to ensure that these two modified optimization problems always have a feasible solution. As a result, the estimators $\hat{\underline{\gamma}}^\star$ and $\hat{\overline{\gamma}}^\star$ always exist. Theorem 1 in Mogstad et al. (2018) provided conditions under which $\hat{\underline{\gamma}}^\star$ and $\hat{\overline{\gamma}}^\star$ are consistent estimators for $\underline{\gamma}^\star$ and $\overline{\gamma}^\star$, which we discuss in Appendix B.3.

4.3 Inference

Consider the following hypotheses:

$$H_0 : \lambda \in \mathcal{L}_0, \quad H_1 : \lambda \in \mathcal{L}/\mathcal{L}_0,$$

where the set $\mathcal{L}_0$ is defined as

$$\mathcal{L}_0 = \{\lambda \in \mathcal{L}_{\mathcal{S},\mathcal{T}} : \ \Gamma^\star(\lambda) = \gamma^\star\}. \tag{27}$$

The set $\mathcal{L}_0$ consists of slope functions that satisfy a collection of conditions in Proposition 2 and generate the target parameter $\gamma^\star$. Then, a minimum distance test statistic can be constructed as:

$$T_n \equiv \inf_{(\lambda,\eta)\in\mathcal{L}_n\times\mathcal{E}} \sqrt{n}\left(|\hat{\Gamma}^\star(\lambda) - \gamma^\star| + \sum_{s\in\mathcal{S}} |\hat{\Gamma}_{1,s}(\lambda) - \hat{\mu}_s| + \sum_{(s,t)\in\mathcal{S}\times\mathcal{T}} |\hat{\Gamma}_{2,s,t}(\lambda) - \eta_{s,t} - \hat{\delta}_{s,t}|\right). \tag{28}$$

As demonstrated by Mogstad et al. (2017), T_n is the solution to an optimization problem that can be reformulated as a linear program. We reject the null hypothesis if $T_n > \hat{c}_{1-\alpha}$ where α is the chosen significance level and $\hat{c}_{1-\alpha}$ is the critical value for the test. We follow the bootstrap procedure in Mogstad et al. (2017) to calculate $\hat{c}_{1-\alpha}$; see the Appendix B.4 for details.

The above analysis can further be used to conduct specification tests. Specifically, a rejection of the null hypothesis $H_0 : \mathcal{L}_{\mathcal{S},\mathcal{T}} \neq \emptyset$ indicates the model is misspecified. A natural statistic for such a test is

$$T_n^{\dagger} \equiv \inf_{(\lambda,\eta)\in\mathcal{L}_n\times\mathcal{E}} \sqrt{n}\left(\sum_{s\in\mathcal{S}} |\hat{\Gamma}_{1,s}(\lambda) - \hat{\mu}_s| + \sum_{(s,t)\in\mathcal{S}\times\mathcal{T}} |\hat{\Gamma}_{2,s,t}(\lambda) - \eta_{s,t} - \hat{\delta}_{s,t}|\right). \tag{29}$$

5 Numerical Illustration

5.1 Data Generating Process

We illustrate the previous results by considering the following model without covariates

$$X = \lambda(X^*) \cdot X^* + \tilde{u}, \quad \tilde{u} \mid X^* \sim \mathcal{N}(0, 0.3^2). \tag{30}$$

The true variable X^* is supposed to follow a uniform distribution on $[2, 3]$, i.e., $X^* \sim \text{Unif}[2, 3]$. The slope function is given by

$$\lambda(x^*) = -0.1(3 - x^*)^2 + 1.1.$$

As shown in Figure 1, this slope function ranges from 1 to 1.1 and is both strictly increasing and concave. This indicates a scenario of overreporting, where the average degree of overreporting rises as the true variable x^* increases. We generate two independent samples of (X, X^*), each containing 10^5 observations. In the first sample, we drop X, and in the second sample, we drop X^* instead.

5.2 Bounds on Target Parameters

We consider two target parameters:

$$\gamma_1 = \text{Corr}(X, X^*), \tag{31}$$

$$\gamma_2 = \frac{\mathbb{E}[X \mid 2 \leq X^* \leq 2.5]}{\mathbb{E}[X^* \mid 2 \leq X^* \leq 2.5]}. \tag{32}$$

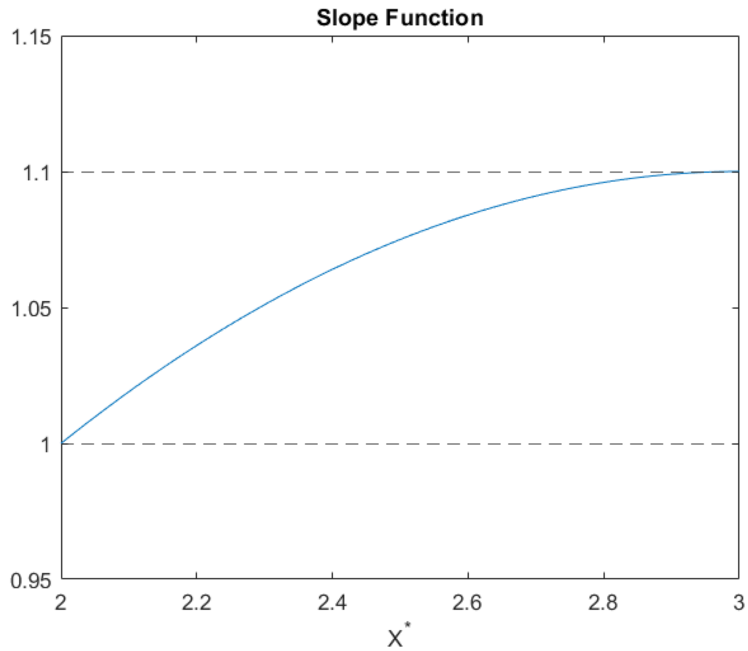

Figure 1: Slope function $\lambda(x^*)$ in the numerical example

Table 1 contains bounds on the correlation coefficient γ_1 under different assumptions. We approximate the slope function with Bernstein polynomials of order $K = 20$, and the set $\mathcal{T}$ includes 100 equally spaced points ranging from 2 to 3. Without any assumption on measurement error, we obtain bounds $[-0.988, 0.988]$ according to Proposition 1, which is too loose since the correlation coefficient takes values from $[-1, 1]$. Next, we assume that subjects tend to overreport and impose Assumption 5 with different (λ_l, λ_u). Consider $(\lambda_l, \lambda_u) = (1, 1.1)$ which corresponds to the range of the slope function, bounds of γ_1 become $[0.461, 0.823]$, which is narrower than without any assumption on the measurement error. Under Assumption 4 ($\mathbb{E}[X \mid X^*]$ is strictly increasing in X^*), we obtain similar bounds $[0.462, 0.823]$.

Next, we impose monotonicity and concavity on the slope function. Under Assumption 6 ($\lambda(\cdot)$ is weakly increasing), bounds become $[0.638, 0.825]$, which is much narrower than only using bounds of the slope function. The upper bound is further tightened to 0.795 after adding Assumption 7 ($\lambda(\cdot)$ is weakly concave). We find that imposing shape restrictions on the slope function effectively yields tight bounds, even when we assume wider bounds on the slope function ($\lambda_u = 1.2$).

Table 2 presents the bounds on γ_2 based on various assumptions. Even without accounting for measurement error, the bounds, which range from $[1.011, 1.364]$, indicate a tendency to overreport. When considering only the bounds of the slope function with $(\lambda_l, \lambda_u) = (1, 1.1)$, the bounds narrow to $[1.032, 1.100]$. By applying the assumptions of monotonicity and concavity, we achieve even tighter bounds, reducing the upper bound to 1.069. Increasing λ_u to 1.2 results in slightly larger bounds for γ_2, but they remain informative.

Additionally, we report estimated bounds for these two target parameters with $(\lambda_l, \lambda_u) = (0.9, 1.1)$ and $(0.9, 1.2)$ in Appendix C.3. These wider bounds on the slope function lead to broader bounds for γ_1 and γ_2; however, incorporating the assumption of monotonicity significantly tightens these bounds.

Table 1: Bounds on γ_1 under various assumptions

Assumptions	$(\lambda_l, \lambda_u) = (1, 1.1)$	$(\lambda_l, \lambda_u) = (1, 1.2)$	True value
No assumption	$[-0.988, 0.988]$	$[-0.988, 0.988]$	0.783
Only Bounds of $\lambda(\cdot)$	$[0.461, 0.823]$	$[0.248, 0.936]$	0.783
$\mathbb{E}[X \mid X^*]$ increasing	$[0.462, 0.823]$	$[0.264, 0.936]$	0.783
$\lambda(\cdot)$ increasing	$[0.638, 0.823]$	$[0.638, 0.935]$	0.783
$\lambda(\cdot)$ increasing+concave	$[0.638, 0.795]$	$[0.638, 0.828]$	0.783

Notes: Bounds under "no assumption" refer to that in Proposition 1 using monotone rearrangement inequality. Bounds in the remaining four rows are calculated by linear programming with the order of Bernstein polynomials $K = 20$, where we assume the slope function is bounded between λ_l and λ_u. The set $\mathcal{T}$ includes 100 equally spaced points ranging from 2 to 3.

Table 2: Bounds on γ_2 under various assumptions

Assumptions	$(\lambda_l, \lambda_u) = (1, 1.1)$	$(\lambda_l, \lambda_u) = (1, 1.2)$	True value
No assumption	$[1.011, 1.364]$	$[1.011, 1.364]$	1.043
Only Bounds of $\lambda(\cdot)$	$[1.032, 1.100]$	$[1.011, 1.149]$	1.043
$\mathbb{E}[X \mid X^*]$ increasing	$[1.032, 1.100]$	$[1.011, 1.144]$	1.043
$\lambda(\cdot)$ increasing	$[1.032, 1.069]$	$[1.012, 1.069]$	1.043
$\lambda(\cdot)$ increasing +concave	$[1.038, 1.069]$	$[1.034, 1.069]$	1.043

Notes: See that in Table 1.

6 Empirical Application

6.1 Data Description

In this section, we revisit two unmatched datasets on welfare benefits used in Hu and Ridder (2012)[5]. The primary dataset is sourced from the Survey of Income and Program Participation (SIPP) and includes the reported benefit level. The auxiliary dataset comes from the Aid to Families with Dependent Children Quality Control system (AFDC QC) and contains the actual benefit level. We follow Hu and Ridder (2012) and focus on all single mothers aged 18 to 64 who entered the program between October 1991 and December 1995.

Hu and Ridder (2012) analyzed the relationship between welfare benefits and the duration of welfare spells using a duration model, assuming that measurement error is independent of the true variable[6]. By applying our framework of systematic error, we allow for more flexible forms of measurement error and seek to understand what we can learn about specific parameters from these two unmatched datasets.

Let X represent the reported benefit level and X^* the actual benefit level, both of which are nominal variables. In this application, we do not include any covariates. Figure 2 shows the estimated density, and Panel A in Table 3 presents the summary statistics for these variables. A comparison of the mean and median suggests that benefit levels are likely being underreported, particularly for individuals with lower benefit levels, as indicated by the means of each bin. Furthermore, the variance of X is greater than that of X^*, which confirms the presence of noise in the reported benefit levels. Panel B in Table 3 presents the results of specification tests conducted by D'Haultfoeuille et al. (2021). We reject the assumption of "no systematic error" ($\lambda(x^*) = 1$) at the 10% significance level. However, we do not reject the assumption that the slope function is constant. Specifically, the estimate of this constant is given by $\hat{\lambda}_0 = 0.928$, which implies an average degree of underreporting of 7.2% across all benefit levels.

[5]Data source: `http://qed.econ.queensu.ca/jae/datasets/hu002/`.

[6]They examined the logarithmic benefit level: $\log X = \log X^* + u^\dagger$, where the error term $u^\dagger$ (which may not be mean-zero) is independent of X^*.

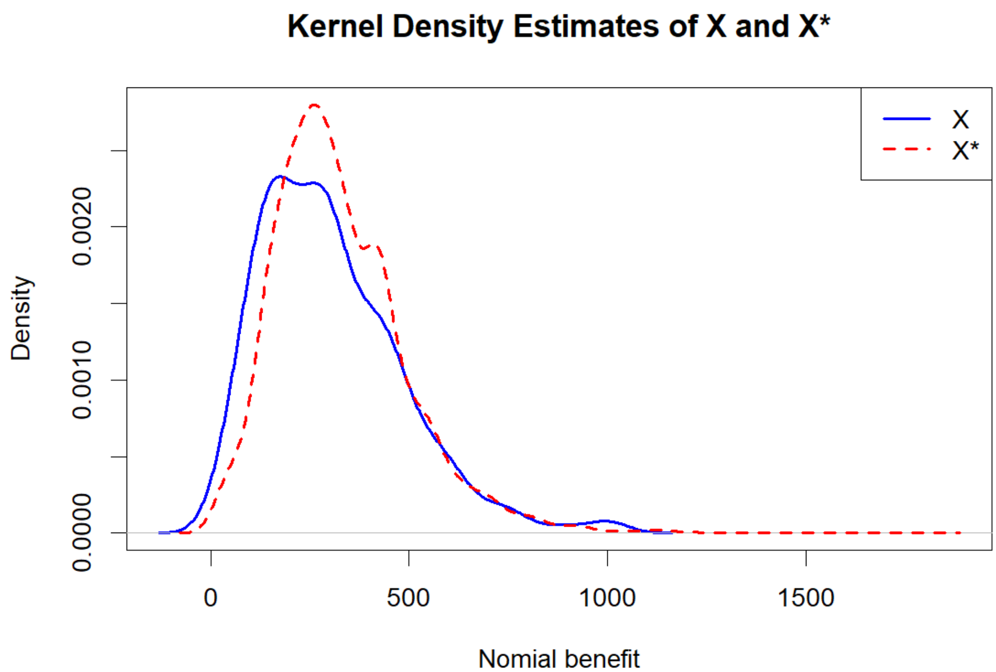

Figure 2: Density estimates of nominal welfare benefits in SIPP and QC samples

6.2 Empirical Results

As shown in Examples 1 and 2, we examine the following three parameters:

$$\begin{aligned}\gamma_1 &= \mathrm{Corr}(X, X^*),\\ \gamma_2 &= \frac{\mathbb{E}[X \mid X^* \le Q_{X^*}(0.25)]}{\mathbb{E}[X^* \mid X^* \le Q_{X^*}(0.25)]},\\ \gamma_3 &= \frac{\mathbb{E}[X \mid X^* \ge Q_{X^*}(0.75)]}{\mathbb{E}[X^* \mid X^* \ge Q_{X^*}(0.75)]}.\end{aligned} \tag{33}$$

Table 4 presents the estimated bounds for these parameters based on different assumptions. We assume that the slope function is bounded above $\lambda_u = 1$, which reflects a scenario of underreporting. In Panels A and B, we consider two different lower bounds for the slope function, specifically set at 0.8 and 0.9. Column (2) further imposes that the slope function is monotonically increasing, which means that the average degree of underreporting decreases as the actual benefit level rises. In addition, Column (3) sets a stricter lower bound for the slope function when the actual benefit level exceeds its median. We set $K = 10$ in this

Table 3: Comparison of nominal welfare benefits in SIPP and QC samples

	X	X^*	Ratio
Panel A: Summary Statistics			
Sample size	520	3318	0.157
Mean	304.2	327.7	0.928
Median	280	294	0.952
SD	180.9	169.4	1.068
Mean in [0, 200)	127.3	139.1	0.915
Mean in [200, 400)	288.0	289.6	0.994
Mean in [400, 600)	473.5	472.0	1.003
Mean in [600, 800)	672.8	678.1	0.992
Mean in [800, ∞)	938.6	965.8	0.972
Panel B: Specification Test			
$\lambda(x^*) = 1$	p-value= 0.09		
$\lambda(x^*) = \lambda_0$	p-value= 1.00 with $\hat{\lambda}_0 = 0.928$		

Notes: The ratio in Panel A refers to that of X compared to X^*. The specification test in Panel B refers to the method developed by D'Haultfoeuille et al. (2021), and we utilize their accompanying R package, RationalExp.

application and choose tuning parameter $\kappa_n = 0.1$ (defined in (26)) to account for statistical errors.

Assuming there is no systematic error ($\lambda(x^*) = 1$), the correlation coefficient γ_1 is point-identified and estimated to be 0.936.[7] When using only bounds of the slope function with $\lambda_l = 0.8$, the bound on γ_1 is $[0.767, 0.955]$. Adding two additional shape restrictions further tightens the bounds, with lower bounds of 0.869 and 0.862, respectively. These lower bounds are notably below the estimate obtained under the assumption of no systematic error, suggesting that neglecting systematic error may lead to overestimation of the correlation

[7] Assuming there are no systematic errors, we have $\mathrm{Cov}(X, X^*) = \mathrm{Var}(X^*)$. Therefore,

$$\mathrm{Corr}(X, X^*) = \frac{\mathrm{Cov}(X, X^*)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(X^*)}} = \sqrt{\frac{\mathrm{Var}(X^*)}{\mathrm{Var}(X)}}.$$

The estimator for this correlation directly follows from the estimators of the variances.

coefficient—and hence to an overly optimistic assessment of the reliability of the reported values. Moreover, the point estimate of 0.936 lies outside the tighter bounds derived under $\lambda_l = 0.9$ reported in Panel B, indicating that the classical measurement error assumption can yield materially misleading conclusions

Next, we consider γ_2 and γ_3, which reflect the average degree of underreporting in the lower and upper tails of the true benefit distribution. Specifically, γ_2 focuses on individuals with X^* below the 25th percentile, while γ_3 considers those above the 75th percentile. These measures help assess heterogeneity in reporting behavior across different population segments.

In column (2) of panel A, using monotonicity of the slope function improves the bounds of γ_2 and γ_3 compared to column (1) without it. The upper bounds indicate that the average degree of underreporting is at least 7.1% $(1-92.9\%)$ for individuals below the 25th percentile and 2.9% $(1-97.1\%)$ for those above the 75th percentile. When applying a stricter lower bound of the slope function in column (3) of panel A, these numbers become more similar, at 6.5% $(1-93.5\%)$ and 6.1% $(1-93.9\%)$, respectively. These findings, together with those in Panel B, indicate that underreporting is prevalent in both subpopulations, with stronger evidence when either of the two additional shape restrictions is imposed.

7 Linear Regression with Mismeasured Regressor under Data Combination

7.1 Model Setup

Let Y be the outcome variable. Let W_0 be a set of covariates and $W = (1, W_0) \in \mathbb{R}^{d_W}$. Consider the linear regression:

$$Y = \beta X^* + \pi' W + \epsilon, \tag{34}$$

where $\epsilon = Y^{\perp (X^*, W)}$ is mean-zero and uncorrelated with (X^*, W_0) by construction. Our parameter of interest is the coefficient β, which may be interpreted as the causal effects of X^* on the outcome variable Y. Unfortunately, we usually observe the mismeasured variable

Table 4: Estimated Bounds under different assumptions

	(1) Only Bounds of $\lambda(\cdot)$	(2) $\lambda(\cdot)$ increasing	(3) $\lambda(x^*) \geq \lambda_0$ if $x^* \geq \mathrm{Med}(X^*)$
Panel A: $(\lambda_l, \lambda_u) = (0.8, 1)$			
γ_1	[0.767, 0.955]	[0.869, 0.955]	[0.862, 0.898]
γ_2	[0.847, 0.991]	[0.847, 0.929]	[0.884, 0.935]
γ_3	[0.880, 0.971]	[0.928, 0.971]	[0.927, 0.939]
Panel B: $(\lambda_l, \lambda_u) = (0.9, 1)$			
γ_1	[0.833, 0.917]	[0.869, 0.917]	[0.867, 0.884]
γ_2	[0.903, 0.965]	[0.903, 0.929]	[0.914, 0.931]
γ_3	[0.911, 0.950]	[0.928, 0.950]	[0.928, 0.932]

Notes: This table presents the estimated bounds of γ_i (for $1 \leq i \leq 3$), as defined in (33). The set $\mathcal{T}$ is selected as the empirical quantiles of X^* at the levels $0, 1/49, 2/49, \ldots, 1$. The bounds are calculated with a Bernstein polynomial order of $K = 10$ and a tuning parameter $\kappa_n = 0.1$, as specified in (26). In this analysis, we assume that the slope function is constrained between the bounds λ_l and λ_u.

X rather than the true variable X^* in the primary dataset. To help identify the coefficient, we rely on an auxiliary dataset that contains the true variable X^*.

Assumption 8. (i) The distribution of (Y, X, W, Z) is identified from the primary dataset; (ii) The distribution of (X^*, Z) is identified from the auxiliary dataset; (iii) The variance matrix of (Y, X, X^*, W) is finite and positive definite; (iv) The distribution of Z has finite second moments.

Assumption 9. (i) $\mathrm{Cov}(X, \epsilon) = 0$; (ii) $\mathrm{Cov}(X^{\perp W}, (X^*)^{\perp W}) > 0$.

The data environment in Assumption 8 is slightly different from that in Assumption 1, since we also observe (Y, W) in the primary dataset. We allow that some components of W may also appear in Z, while others may be excluded; that is, some components of W may be observed in two datasets, while others may only be observed in the primary dataset. Assumption 8 (iii) ensures the coefficient β is well-defined.

Assumption 9 (i) states the exclusion condition for X, which implies that the effect of X

on the outcome variable Y is only through the effect of X^*. Assumption 9 (ii) states that X and X^* are positively correlated after removing parts explained by W.

Let β^{OLS} be the coefficient on X in the OLS estimand of Y on (X, W). When u is uncorrelated with X^* and W, β^{OLS} suffers from the familiar attenuation bias:

$$\beta^{\text{OLS}} = \beta \cdot \left[1 - \frac{\text{Var}(u)}{\text{Var}(X^{\perp W})}\right]. \tag{35}$$

With the auxiliary dataset, we can quantify the extent of attenuation bias and thus identify β. This is due to the fact that $\text{Var}(u)$ can be identified as $\text{Var}(X) - \text{Var}(X^*)$ when using two datasets. Proposition 6 below summarizes this finding.

Proposition 6. Suppose Assumptions 8 and 9 hold. If u is uncorrelated with X^* and W, β is point-identified.

7.2 Bounds of the Coefficient with General Measurement Error

Now we study the identification of the coefficient β while allowing for flexible correlations between u and (X^*, W). Assumption 9 ensures that we can instrument X^* with X in the regression (34):

$$\beta = \frac{\text{Cov}(X^{\perp W}, Y^{\perp W})}{\text{Cov}(X^{\perp W}, (X^*)^{\perp W})}. \tag{36}$$

The numerator in (36) is identified from primary dataset, and it remains to derive bounds of the denominator. Consider two linear projections:

$$X = \pi_X' W + X^{\perp W}, \tag{37}$$

$$W = \Pi \cdot Z + W^{\perp Z}, \tag{38}$$

where π_X is a $d_W \times 1$ vector and Π is a $d_W \times d_Z$ matrix. Since we observe (X, W, Z) in the primary dataset, linear projection coefficients π_X and Π are point-identified. Then, the denominator in (36) can be written as

$$\text{Cov}(X^{\perp W}, X^*) = \text{Cov}(X, X^*) - \text{Cov}(\pi_X' W, X^*)$$

$$= \mathrm{Cov}(X, X^*) - \mathrm{Cov}(\pi_X' \Pi \cdot Z, X^*) - \mathrm{Cov}(\pi_X' W^{\perp Z}, X^*). \tag{39}$$

The first term in (39) is the target parameter shown in Example 1 and can be bounded by Proposition 3 under different restrictions that researchers wish to maintain on the slope function. The second term in (39) is point-identified using information from the two datasets. It is easy to see that the last term in (39) vanishes when W is included in common covariates Z. In cases where some components of W are not included in Z, we need to impose the following assumption.

Assumption 10. (i) $\Pr(X^* = 0) = 0$; (ii) there exists $\lambda_u^\dagger \geq \lambda_l^\dagger > 0$ such that the following holds almost sure,

$$\lambda_l^\dagger \leq \frac{\mathbb{E}[X \mid X^* = x^*, W = w, Z = z]}{x^*} \leq \lambda_u^\dagger.$$

Assumption 10 bounds the slope function with the additional effect of W. Therefore, $[\lambda_l^\dagger, \lambda_u^\dagger]$ should contain the bounds defined in Assumption 5. Proposition 7 derives bounds for the coefficient β. We use the convention $1/0 = \infty$.

Proposition 7. Suppose Assumption 8, 9 and 10 hold. Normalize $\mathrm{Cov}(X^{\perp W}, Y^{\perp W}) > 0$. We have

$$\frac{\mathrm{Cov}(X^{\perp W}, Y^{\perp W})}{\overline{\gamma}^\star - \mathrm{Cov}(\pi_X' \Pi \cdot Z, X^*) - \underline{A}} \leq \beta \leq \frac{\mathrm{Cov}(X^{\perp W}, Y^{\perp W})}{\max\{\underline{\gamma}^\star - \mathrm{Cov}(\pi_X' \Pi \cdot Z, X^*) - \overline{A}, 0\}},$$

where $\overline{\gamma}^\star$ and $\underline{\gamma}^\star$ are bounds of $\gamma = \mathrm{Cov}(X, X^*)$ obtained from Proposition 3 and

$$\underline{A} = \max\left\{\mathbb{E}\left[\frac{\pi_X' W^{\perp Z} X}{\lambda_l^\dagger + \mathbf{1}\{\pi_X' W^{\perp Z} X > 0\}(\lambda_u^\dagger - \lambda_l^\dagger)}\right], \mathbb{E}\left[\int_0^1 Q_{\pi_X' W^{\perp Z}}(1-\tau|Z) Q_{X^*}(\tau|Z) d\tau\right]\right\},$$
$$\overline{A} = \min\left\{\mathbb{E}\left[\frac{\pi_X' W^{\perp Z} X}{\lambda_l^\dagger + \mathbf{1}\{\pi_X' W^{\perp Z} X < 0\}(\lambda_u^\dagger - \lambda_l^\dagger)}\right], \mathbb{E}\left[\int_0^1 Q_{\pi_X' W^{\perp Z}}(\tau|Z) Q_{X^*}(\tau|Z) d\tau\right]\right\}.$$

The interval $[\underline{A}, \overline{A}]$ represents the intersection of two distinct sets of bounds derived for $\mathrm{Cov}(\pi_X' W^{\perp Z}, X^*)$ (the final term in equation (39)), as established in the proof of Proposition 7. The first set of bounds is constructed based on Assumption 10, which varies continuously with the sensitivity parameters $(\lambda_l^\dagger, \lambda_u^\dagger)$. The second set of bounds uses the monotone rearrangement inequality, since the marginal distributions of $\pi_X' W^{\perp Z}$ and X^* are identified from two datasets. Deriving sharp bounds for the coefficient β is challenging, and we will leave it for future studies.

8 Conclusion

In this paper, we study measurement error models when measurements and true values are recorded in two separate datasets that cannot be matched. Our analysis centers on parameters that describe the structure of measurement error, and the framework readily extends to OLS coefficients. The framework allows researchers to derive computationally tractable bounds under a flexible set of identifying assumptions and provides a practical tool for assessing measurement error when matched data are unavailable. One potential extension is to explore other target parameters that can be expressed as a quadratic form of the slope function, such as the variance of random error. Furthermore, we hope to apply our approach to other contexts involving data combination.

References

Abay, K. A., Abate, G. T., Barrett, C. B., and Bernard, T. (2019). Correlated non-classical measurement errors,'second best'policy inference, and the inverse size-productivity relationship in agriculture. *Journal of Development Economics*, 139:171–184.

Angel, S., Disslbacher, F., Humer, S., and Schnetzer, M. (2019). What did you really earn last year?: explaining measurement error in survey income data. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 182(4):1411–1437.

Bick, A., Blandin, A., and Rogerson, R. (2022). Hours and wages. *The Quarterly Journal of Economics*, 137(3):1901–1962.

Blattman, C., Jamison, J. C., and Sheridan, M. (2017). Reducing crime and violence:

Experimental evidence from cognitive behavioral therapy in liberia. *American Economic Review*, 107(4):1165–1206.

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

Bollinger, C. R., Hirsch, B. T., Hokayem, C. M., and Ziliak, J. P. (2019). Trouble in the tails? what we know about earnings nonresponse 30 years after lillard, smith, and welch. *Journal of Political Economy*, 127(5):2143–2185.

Bound, J., Brown, C., Duncan, G. J., and Rodgers, W. L. (1994). Evidence on the validity of cross-sectional and longitudinal labor market data. *Journal of Labor Economics*, 12(3):345–368.

Bound, J., Brown, C., and Mathiowetz, N. (2001). Measurement error in survey data. In *Handbook of econometrics*, volume 5, pages 3705–3843. Elsevier.

Bound, J. and Krueger, A. B. (1991). The extent of measurement error in longitudinal earnings data: Do two wrongs make a right? *Journal of labor economics*, 9(1):1–24.

Carroll, R. J. and Wand, M. P. (1991). Semiparametric estimation in logistic measurement error models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3):573–585.

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

D'Haultfoeuille, X., Gaillac, C., and Maurel, A. (2021). Rationalizing rational expectations: Characterizations and tests. *Quantitative Economics*, 12(3):817–842.

D'Haultfœuille, X., Gaillac, C., and Maurel, A. (2025). Partially linear models under data combination. *Review of Economic Studies*, 92(1):238–267.

Fan, Y. and Park, S. S. (2010). Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory*, 26(3):931–951.

Fan, Y., Sherman, R., and Shum, M. (2014). Identifying treatment effects under data combination. *Econometrica*, 82(2):811–822.

Fan, Y., Shi, X., and Tao, J. (2023). Partial identification and inference in moment models with incomplete data. *Journal of Econometrics*, 235(2):418–443.

Gechter, M. (2024). Generalizing the results from social experiments: theory and evidence from india. *Journal of Business & Economic Statistics*, 42(2):801–811.

Han, S. and Yang, S. (2024). A computational approach to identification of treatment effects for policy evaluation. *Journal of Econometrics*, 240(1):105680.

Hu, Y. and Ridder, G. (2012). Estimation of nonlinear models with mismeasured regressors using marginal information. *Journal of Applied Econometrics*, 27(3):347–385.

Hwang, Y. (2023). Bounding omitted variable bias using auxiliary data: With an application to estimate neighborhood effects. *Available at SSRN 3866876*.

Kamat, V. (2024). Identifying the effects of a program offer with an application to head start. *Journal of Econometrics*, 240(1):105679.

Kamat, V. and Norris, S. (2025). Estimating welfare effects in a non-parametric choice model: The case of school vouchers. *The Review of Economic Studies*, page rdaf064.

Lee, L.-f. and Sepanski, J. H. (1995). Estimation of linear and nonlinear errors-in-variables models using validation data. *Journal of the American Statistical Association*, 90(429):130–140.

Li, A. (2025). Assessing measurement error in linear instrumental variables models. Working Paper.

Mogstad, M., Santos, A., and Torgovitsky, A. (2017). Using instrumental variables for inference about policy relevant treatment effects. *National Bureau of Economic Research Working Paper No. 23568*. Available at https://www.nber.org/papers/w23568.

Mogstad, M., Santos, A., and Torgovitsky, A. (2018). Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica*, 86(5):1589–1619.

Pacini, D. (2019a). Two-sample least squares projection. *Econometric Reviews*, 38(1):95–123.

Pacini, D. (2019b). The two-sample linear regression model with interval-censored covariates. *Journal of Applied Econometrics*, 34(1):66–81.

Russell, T. M. (2021). Sharp bounds on functionals of the joint distribution in the analysis of treatment effects. *Journal of Business & Economic Statistics*, 39(2):532–546.

Schennach, S. (2022). Measurement systems. *Journal of Economic Literature*, 60(4):1223–1263.

Tebaldi, P., Torgovitsky, A., and Yang, H. (2023). Nonparametric estimates of demand in the california health insurance exchange. *Econometrica*, 91(1):107–146.

Torgovitsky, A. (2019). Nonparametric inference on state dependence in unemployment. *Econometrica*, 87(5):1475–1505.

A Proofs

Proof of Proposition 1. It follows from Lemma 2 and Corollary 1 in Hwang (2023). *Q.E.D.*

Proof of Lemma 1. Define random variable $X^*_\lambda = \lambda(X^*, Z) \cdot X^*$. By Assumption 3, we have $X = X^*_\lambda + \tilde{u}$, where random error $\tilde{u}$ is defined in (7). Then, we have $\mathbb{E}[\tilde{u} \mid X^*_\lambda, Z] = 0$ since $\mathbb{E}[\tilde{u} \mid X^*, Z] = 0$. This implies that the distribution of random variable X is a mean-preserving spread of the distribution of X^*_λ, and our results follow from proposition 2 in D'Haultfoeuille et al. (2021). *Q.E.D.*

Proof of Proposition 2. Moment equality (13) follows directly from (11) by using $s(Z)$ as an instrument. It remains to show (14). Using $s(Z)$ as an instrument in (12), we have (since $s(Z) \geq 0$)

$$\begin{aligned}\mathbb{E}\Big[(x - X)^+ \cdot s(Z)\Big] &\geq \mathbb{E}\Big[(x - \lambda(X^*, Z) \cdot X)^+ \cdot s(Z)\Big] \\ &\geq \mathbb{E}[(x - \lambda(X^*, Z) \cdot X) \cdot d_t(X^*) \cdot s(Z)],\end{aligned} \tag{40}$$

where the second step uses the fact that $a^+ \geq a \cdot d$ for any real number a and $d \in \{0, 1\}$. This leads to

$$\mathbb{E}\Big[\lambda(X^*, Z) \cdot s(Z) X^* \cdot d_t(X^*)\Big] \geq \mathbb{E}\left[s(Z)\Big(x \cdot d_t(X^*) - (x - X)^+\Big)\right].$$

Note that the left side does not depend on x, we have

$$\mathbb{E}\Big[\lambda(X^*, Z) \cdot s(Z) X^* \cdot d_t(X^*)\Big] \geq \sup_{x \in \mathbb{R}} \mathbb{E}\left[s(Z)\Big(x \cdot d_t(X^*) - (x - X)^+\Big)\right] = \delta_{s,t}.$$

Thus, we complete the proof. *Q.E.D.*

Proof of Proposition 3. Since $\mathcal{L}$ is convex and constraints in $\mathcal{L}_{\mathcal{S},\mathcal{T}}$ are linear, we have $\mathcal{L}_{\mathcal{S},\mathcal{T}}$ is convex. Note that $\mathcal{L}_{\mathcal{S},\mathcal{T}}$ is non-empty and the map $\Gamma^\star : \mathcal{L} \to \mathbb{R}$ is linear, we further obtain that $\mathcal{G}_{\mathcal{S},\mathcal{T}}$ is a convex set, and so its closure is $[\inf_{\lambda \in \mathcal{L}_{\mathcal{S},\mathcal{T}}} \Gamma^*(\lambda), \sup_{\lambda \in \mathcal{L}_{\mathcal{S},\mathcal{T}}} \Gamma^*(\lambda)] \equiv [\underline{\gamma}^\star, \overline{\gamma}^\star]$. *Q.E.D.*

Proof of Proposition 4. For notational simplicity

$$\mathcal{L}_{\text{is}} = \{\lambda \in \mathcal{L} : \ \lambda \text{ satisfies (11) and (12)}\}.$$

It is easy to see that $\mathcal{L}_{\text{is}} \subseteq \mathcal{L}_{\mathcal{S},\mathcal{T}}$. We separate this proof into three parts.

We firstly show $\mathcal{L}_{\text{is}} = \mathcal{L}_{\mathcal{S},\mathcal{T}}$ under Assumption 4. Since Lemma 1 implies that $\mathcal{L}_{\text{IS}} \subseteq \mathcal{L}_{\text{is}}$, it suffices to show $\mathcal{L}_{\text{is}} \subseteq \mathcal{L}_{\text{IS}}$. For any $\lambda \in \mathcal{L}_{\text{is}}$, define random variable $X^*_\lambda = \lambda(X^*, Z)X^*$. Then, we have

$$\mathbb{E}\Big[X^*_\lambda \mid Z\Big] = \mathbb{E}[X \mid Z],$$
$$\mathbb{E}\left[(x - X^*_\lambda)^+ \mid Z\right] \leq \mathbb{E}[(x - X)^+ \mid Z], \quad \forall x \in \mathbb{R}.$$

By the equivalence result in proposition 2 of D'Haultfoeuille et al. (2021), there exist random variables $(\tilde{X}, \tilde{X}^*_\lambda, \tilde{Z})$ such that $(\tilde{X}, \tilde{Z}) \overset{d}{=} (X, Z)$, $(\tilde{X}^*_\lambda, \tilde{Z}) \overset{d}{=} (X^*_\lambda, Z)$ and $\mathbb{E}[\tilde{X} \mid \tilde{X}^*_\lambda = x^*_\lambda, \tilde{Z}] = x^*_\lambda$. Note that $X^*_\lambda = \mathbb{E}[X \mid X^*, Z]$, there exists a one-to-one mapping between X^* and X^*_λ given Z by Assumption 4. Then we can write down $X^* = \phi(X^*_\lambda, Z)$ for some well-defined function $\phi(\cdot)$. Let $\tilde{X}^* = \phi(\tilde{X}^*_\lambda, \tilde{Z})$, we have

$$\mathbb{E}[\tilde{X} \mid \tilde{X}^* = x^*, \tilde{Z} = z] = \mathbb{E}[\tilde{X} \mid \tilde{X}^*_\lambda = \lambda(x^*, z)x^*, \tilde{Z} = z] = \lambda(x^*, z)x^*,$$

which shows that $\lambda \in \mathcal{L}_{\text{IS}}$. Thus, we verify that $\mathcal{L}_{\text{is}} = \mathcal{L}_{\mathcal{S},\mathcal{T}}$.

Next, suppose that $\mathcal{T}$ is dense in $\text{supp}(X^*)$ and includes all the discontinuities of $F_{X^*}(\cdot)$, and condition (i) hold; that is, Z has a discrete support $\{z_1, z_2, \cdots, z_m\}$ and $\mathcal{S} = \{\mathbf{1}\{z = z_j\} : 1 \leq j \leq m\}$. To prove $\mathcal{L}_{\mathcal{S},\mathcal{T}} = \mathcal{L}_{\text{is}}$, it suffices to show that $\mathcal{L}_{\mathcal{S},\mathcal{T}} \subseteq \mathcal{L}_{\text{is}}$. Note that for any $\lambda \in \mathcal{L}_{\mathcal{S},\mathcal{T}}$ and $s(z) = \mathbf{1}\{z = z_j\}(1 \leq j \leq m)$, we obtain (by (13))

$$\mathbb{E}[\lambda(X^*, Z) \cdot X^* \cdot \mathbf{1}\{z = z_j\}] = \mathbb{E}[X \cdot \mathbf{1}\{z = z_j\}],$$

which leads to $\mathbb{E}[\lambda(X^*, Z) \cdot X^* \mid Z = z_j] = \mathbb{E}[X \mid Z = z_j]$ for all $1 \leq j \leq m$. It remains to show that for any $\lambda \in \mathcal{L}_{\mathcal{S},\mathcal{T}}$, moment inequalities (12) hold, which is equivalent to

$$\mathbb{E}\left[(x-X)^{+}\mathbf{1}\{Z=z_j\}\right] \leq \mathbb{E}\Big[(x-\lambda(X^*,Z)\cdot X)^{+}\cdot \mathbf{1}\{Z=z_j\}\Big] \tag{41}$$

for $1 \leq j \leq m$. By Assumption 4, the function $G(x^*) = x - \lambda(x^*, z_j)x^*$ is strictly decreasing for any given $x \in \mathbb{R}$ and $1 \leq j \leq m$. Define $t_0 = \inf\{x^* \in \mathrm{Supp}(X^*) : G(x^*) \leq 0\}$. Since $\mathrm{Supp}(X^*)$ is closed, we have $t_0 \in \mathrm{Supp}(X^*)$. For $s(z) = \mathbf{1}\{z = z_j\}(1 \leq j \leq m)$ and $t \in \mathcal{T}$, we obtain (by (14))

$$\mathbb{E}[\lambda(X^*,Z)\cdot \mathbf{1}\{Z=z_j\}X^*\cdot d_t(X^*)] \geq \mathbb{E}\left[\mathbf{1}\{Z=z_j\}\Big(x\cdot d_t(X^*)-(x-X)^{+}\Big)\right].$$

Note that $\mathbb{E}[\lambda(X^*,Z)\cdot \mathbf{1}\{Z=z_j\}X^*]^2 \leq \mathbb{E}[\lambda(X^*,Z)\cdot X^*]^2 < \infty$. By Lemma 2, we have

$$\mathbb{E}[\lambda(X^*,Z)\cdot \mathbf{1}\{Z=z_j\}X^*\cdot d_{t_0}(X^*)] \geq \mathbb{E}\left[\mathbf{1}\{Z=z_j\}\Big(x\cdot d_{t_0}(X^*)-(x-X)^{+}\Big)\right],$$

which leads to

$$\begin{aligned}\mathbb{E}\left[|x-X|\mathbf{1}\{Z=z_j\}\right] &\geq \mathbb{E}\Big[(x-\lambda(X^*,Z)\cdot X)\cdot d_{t_0}(X^*)\mathbf{1}\{Z=z_j\}\Big]\\ &= \mathbb{E}\Big[|x-\lambda(X^*,Z)\cdot X|\cdot \mathbf{1}\{Z=z_j\}\Big],\end{aligned} \tag{42}$$

where the last step uses the definition of t_0. Thus, we verify (41).

Finally, suppose that $\mathcal{T}$ is dense in $\mathrm{supp}(X^*)$ and includes all the discontinuities of $F_{X^*}(\cdot)$, and condition (ii) hold; that is, $\mathbb{E}[X \mid X^*, Z] = \mathbb{E}[X \mid X^*]$ and the linear span of $\mathcal{S}$ is norm dense in $L^2(Z) \equiv \{f : \mathbb{R}^{d_Z} \to \mathbb{R} : \ \mathbb{E}[f(Z)^2] < \infty\}$. Then we can write $\lambda(x^*, z) = \widetilde{\lambda}(x^*)$ for function $\widetilde{\lambda}(\cdot)$. Note that (13) implies that

$$\mathbb{E}[(\widetilde{\lambda}(X^*)X^*-X)\cdot s(Z)] = 0$$

for all $s \in \mathcal{S}$. Since $\mathcal{S}$ is norm dense in $L^2(Z)$, we conclude that $\mathbb{E}[\lambda(X^*,Z)X^* - X \mid Z] = 0$ almost sure. It remains to verify the inequalities for all $s \in \mathcal{S}$:

$$\mathbb{E}\left[(x-X)^{+}s(Z)\right] \leq \mathbb{E}\Big[(x-\widetilde{\lambda}(X^*)\cdot X)^{+}\cdot s(Z)\Big]. \tag{43}$$

Define $t^* = \inf\{x^* \in \text{Supp}(X^*) : x - \tilde{\lambda}(x^*)x^* \leq 0\}$. We can follow the arguments leading to (42) to verify (43); thus, we complete the proof. *Q.E.D.*

Lemma 2. Suppose that M is a constant and $g(X^*, Z)$ is a function such that $\mathbb{E}[g^2(X^*, Z)] < \infty$. Suppose that $\mathcal{T}$ is dense in $\text{supp}(X^*)$ and includes all the discontinuities of $F_{X^*}(\cdot)$. Then, if $\mathbb{E}[g(X^*, Z) \cdot d_t(X^*)] \geq M$ holds for all $t \in \mathcal{T}$, we have $\mathbb{E}[g(X^*, Z) \cdot d_t(X^*)] \geq M$ holds for all $t \in \text{supp}(X^*)$.

Proof of Lemma 2. Suppose $t_c \in \text{Supp}(X^*)$ and $F_{X^*}(\cdot)$ is continuous at t. Since $\mathcal{T}$ is dense in $\text{Supp}(X^*)$, there exists a sequence $\{t_k\}_{k\leq 1} \subseteq \mathcal{T}$ such that $\lim_{k\to\infty} t_k = t_c$. Note that $\mathbb{E}[d_{t_1}(X^*) - d_{t_2}(X^*)]^2 = |F_{X^*}(t_1) - F_{X^*}(t_2)|$ for any t_1, t_2, we have by Cauchy-Schwarz inequality

$$\begin{aligned}
&\Big|\mathbb{E}[g(X^*, Z) \cdot d_{t_k}(X^*)] - \mathbb{E}[g(X^*, Z) \cdot d_{t_c}(X^*)]\Big| \\
&\leq \big(\mathbb{E}[g(X^*, Z)^2]\big)^{1/2} \cdot \big(\mathbb{E}[d_{t_k}(X^*) - d_{t_c}(X^*)]^2\big)^{1/2} \\
&= \big(\mathbb{E}[g(X^*, Z)^2]\big)^{1/2} \cdot |F_{X^*}(t_k) - F_{X^*}(t_c)|^{1/2} \to 0, \quad \text{as } k \to \infty,
\end{aligned} \tag{44}$$

where we use the continuity of $F_{X^*}(\cdot)$ at t_c. Then we have $\mathbb{E}[g(X^*, Z) \cdot d_{t_c}(X^*)] \geq M$ since $\mathbb{E}[g(X^*, Z) \cdot d_{t_k}(X^*)] \geq M$ for all $k \geq 1$. This proves $\mathbb{E}[g(X^*, Z) \cdot d_t(X^*)] \geq M$ holds for $t \in \text{Supp}(X^*)$ such that $F_{X^*}(\cdot)$ is continuous at t. Since $\mathcal{T}$ includes all the discontinuities of $F_{X^*}(\cdot)$, we complete the proof. *Q.E.D.*

Proof of Proposition 5. In case (i), the slope function is identified from (11):

$$\lambda_0(z) = \frac{\mathbb{E}[X \mid Z = z]}{\mathbb{E}[X^* \mid Z = z]}.$$

In the case (ii), we follow from (11):

$$\mathbb{E}[X|Z = z_j] = \sum_{m=1}^{M} \mathbb{E}[\rho_m(X^*, Z)X^*|Z = z_j] \cdot \alpha_m$$

for $1 \leq j \leq M$. Then, $\{\alpha_m\}_{m=1}^M$ is identified from the full-rank condition, which implies that the slope function is identified. Finally, by (10), the point identification of the target parameter follows from the point identification of the slope function. Thus, we finish the proof. *Q.E.D.*

Proof of Proposition 6. By (34), we have $Y^{\perp W} = \beta (X^*)^{\perp W} + \epsilon$. Then,

$$
\begin{aligned}
\beta^{\text{OLS}} &= \frac{\text{Cov}(X^{\perp W}, Y^{\perp W})}{\text{Var}(X^{\perp W})} \\
&= \frac{\text{Cov}(X, \beta (X^*)^{\perp W} + \epsilon)}{\text{Var}(X^{\perp W})} \\
&= \beta \cdot \frac{\text{Cov}(X, (X^*)^{\perp W})}{\text{Var}(X^{\perp W})},
\end{aligned}
\tag{45}
$$

where we use $\text{Cov}(X, \epsilon) = 0$ in Assumption 9. Since u is uncorrelated with (X^*, W), we have $X^{\perp W} = (X^*)^{\perp W} + u$ and $\text{Var}(u) = \text{Var}(X^{\perp W}) - \text{Var}((X^*)^{\perp W})$. Then,

$$
\text{Cov}(X, (X^*)^{\perp W}) = \text{Cov}(X^{\perp W}, (X^*)^{\perp W}) = \text{Var}((X^*)^{\perp W}) = \text{Var}(X^{\perp W}) - \text{Var}(u). \tag{46}
$$

Combined (46) with (45), we obtain

$$
\beta^{\text{OLS}} = \beta \cdot \left[1 - \frac{\text{Var}(u)}{\text{Var}(X^{\perp W})} \right].
$$

The point identification of β follows from that $\text{Var}(u) = \text{Var}(X) - \text{Var}(X^*)$ is identified form two datasets. *Q.E.D.*

Proof of Proposition 7. By (36) and (39), it is suffices to show $\text{Cov}(\pi_X' W^{\perp Z}, X^*) = \mathbb{E}[\pi_X' W^{\perp Z} X^*] \in [\underline{A}, \overline{A}]$. By the Monotone rearrangement inequality,

$$
\begin{aligned}
\mathbb{E}\left[\int_0^1 Q_{\pi_X' W^{\perp Z}}(1-\tau|Z) Q_{X^*}(\tau|Z) d\tau \right] &\leq \mathbb{E}[\pi_X' W^{\perp Z} X^*] \\
&\leq \mathbb{E}\left[\int_0^1 Q_{\pi_X' W^{\perp Z}}(\tau|Z) Q_{X^*}(\tau|Z) d\tau \right].
\end{aligned}
\tag{47}
$$

On the other side, define

$$
\lambda^\dagger(x^*, w, z) = \frac{\mathbb{E}[X \mid X^* = x^*, W = w, Z = z]}{x^*}.
$$

Denote

$$\tilde{u}^\dagger = X - \mathbb{E}[X \mid X^*, W, Z].$$

We have $\mathbb{E}[\tilde{u}^\dagger \mid X^*, W, Z] = 0$. By Assumption 10 (i),

$$X = \lambda^\dagger(X^*, W, Z) \cdot X^* + \tilde{u}^\dagger, \tag{48}$$

which leads to

$$\mathbb{E}[\pi_X' W^{\perp Z} X^*] = \mathbb{E}\left[\pi_X' W^{\perp Z} \cdot \frac{X - \tilde{u}^\dagger}{\lambda^\dagger(X^*, W, Z)}\right] = \mathbb{E}\left[\pi_X' W^{\perp Z} \cdot \frac{X}{\lambda^\dagger(X^*, W, Z)}\right].$$

By Assumption 10 (ii), we have $\lambda_l^\dagger \le \lambda^\dagger(x^*, w, z) \le \lambda_u^\dagger$ almost surely, which implies that

$$\begin{aligned} \mathbb{E}\left[\frac{\pi_X' W^{\perp Z} X}{\lambda_l^\dagger + \mathbf{1}\{\pi_X' W^{\perp Z} X > 0\}(\lambda_u^\dagger - \lambda_l^\dagger)}\right] &\le \mathbb{E}[\pi_X' W^{\perp Z} X^*] \\ &\le \mathbb{E}\left[\frac{\pi_X' W^{\perp Z} X}{\lambda_l^\dagger + \mathbf{1}\{\pi_X' W^{\perp Z} X < 0\}(\lambda_u^\dagger - \lambda_l^\dagger)}\right]. \end{aligned} \tag{49}$$

Combined (47) with (49), we prove that $\mathbb{E}[\pi_X' W^{\perp Z} X^*] \in [\underline{A}, \overline{A}]$. *Q.E.D.*

B More Details on Estimation and Inference

B.1 Bernstein Polynomials

An example of a basis is the Bernstein basis, which can approximate any continuous function defined on $[0, 1]$. The kth Berstein polynomial of degree K is a function $b_k^K\colon [0, 1] \to \mathbb{R}$:

$$b_k^K(t) = C_K^k t^k (1 - t)^{K-k}.$$

We approximate a continuous function $f(\cdot)$ by

$$f(t) = \sum_{k=0}^{K} \theta_k b_k^K(t).$$

Shape restrictions of $f(\cdot)$ can be constrained by imposing linear restrictions on θ:

- Bounded below by λ_l and above by λ_u: $\lambda_l \leq \theta_k \leq \lambda_u$ for all k.
- Monotonically increasing: $\theta_0 \leq \theta_1 \leq \cdots \leq \theta_K$.

B.2 First Step Estimators

Suppose we have two random samples $\{X_i, Z_{p,i}\}_{i=1}^{n_p}$ and $\{X_i^*, Z_{a,i}\}_{i=1}^{n_a}$. Estimators for mappings $\Gamma^\star, \Gamma_{1,s}$ and $\Gamma_{2,s,t}$ are

$$\begin{aligned}
\hat{\Gamma}^\star(\lambda) &\equiv \frac{1}{n_a}\sum_{i=1}^{n_a} \lambda(X_i^*, Z_{a,i}) \cdot h(X_i^*, Z_{a,i}) X_i^*, \\
\hat{\Gamma}_{1,s}(\lambda) &\equiv \frac{1}{n_a}\sum_{i=1}^{n_a} \lambda(X_i^*, Z_{a,i}) \cdot s(Z_{a,i}) X_i^*, \\
\hat{\Gamma}_{2,s,t}(\lambda) &\equiv \frac{1}{n_a}\sum_{i=1}^{n_a} \lambda(X_i^*, Z_{a,i}) \cdot s(Z_{a,i}) X_i^* \cdot d_t(X_i^*).
\end{aligned}$$

For μ_s and $\delta_{s,t}$, natural estimators are

$$\begin{aligned}
\hat{\mu}_s &\equiv \frac{1}{n_p}\sum_{i=1}^{n_p} s(Z_{p,i}) X_i, \\
\hat{\delta}_{s,t} &\equiv \sup_{x \in [x_l, x_u]} \left\{ x \cdot \frac{1}{n_a}\sum_{i=1}^{n_a} s(Z_{a,i}) d_t(X_i^*) - \frac{1}{n_p}\sum_{j=1}^{n_p} s(Z_{p,j})(x - X_j)^+ \right\}.
\end{aligned}$$

where $x_l = \min_{1 \leq i \leq n_p} X_i$ and $x_u = \max_{1 \leq i \leq n_p} X_i$.

B.3 Consistency

We follow assumption 1 and theorem 1 in Mogstad et al. (2018) to establish the consistency of estimators $\underline{\hat{\gamma}}^\star$ and $\hat{\bar{\gamma}}^\star$. Some additional notations are required. We will view $s \mapsto \mu_s$ as a function of $s \in \mathcal{S}$ and we denote this function as μ. We assume that μ is uniformly bounded over $\mathcal{S}$, that is, $\mu \in \ell^\infty(\mathcal{S})$. Similarly, δ is a function of $(s,t) \in \mathcal{S} \times \mathcal{T}$, and we assume $\delta \in \ell^\infty(\mathcal{S} \times \mathcal{T})$.

For any fixed $\lambda \in \mathcal{L}$, we view $s \mapsto \Gamma_{1,s}(\lambda)$ as a function of $s \in \mathcal{S}$, which we denote as $\Gamma_{1,s}(\lambda)$ and assume to be an element of $\ell^\infty(\mathcal{S})$. Similarly, we view $(s,t) \mapsto \Gamma_{2,s,t}(\lambda)$ as a function of $(s,t) \in \mathcal{S} \times \mathcal{T}$, which we denote as $\Gamma_2(\lambda)$ and assume to be an element of $\ell^\infty(\mathcal{S} \times \mathcal{T})$. We also assume that $\mathcal{L}$ is a subset of a Banach space $\mathbf{L}$ and let $\Gamma_1 : \mathbf{L} \to \ell^\infty(\mathcal{S})$ and $\Gamma_2 : \mathbf{L} \to \ell^\infty(\mathcal{S} \times \mathcal{T})$ that returns $\Gamma_1(\lambda)$ and $\Gamma_2(\lambda)$ for each $\lambda \in \mathbf{L}$. Without loss of generality, we assume that the sample sizes for the two datasets are the same, denoted as n; that is, $n_p = n_a = n$.

Assumption 11. (i) $\mathcal{L}_{\mathcal{S},\mathcal{T}}$ defined in (17) is not empty.

(ii) $\Gamma^\star : \mathbf{L} \to \mathbb{R}$, $\Gamma_1 : \mathbf{L} \to \ell^\infty(\mathcal{S})$ and $\Gamma_2 : \mathbf{L} \to \ell^\infty(\mathcal{S} \times \mathcal{T})$ are continuous. The set $\mathcal{L} \in \mathbf{L}$ is compact in the weak topology.

(iii) There exists tight stochastic processes $(\mathbb{G}_{\Gamma_1}, \mathbb{G}_{\Gamma_2}, \mathbb{G}_\mu, \mathbb{G}_\delta)$ such that as $n \to \infty$,

$$(\sqrt{n}(\hat{\Gamma}_1 - \Gamma_1), \sqrt{n}(\hat{\Gamma}_2 - \Gamma_2), \sqrt{n}(\hat{\mu} - \mu), \sqrt{n}(\hat{\delta} - \delta)) \xrightarrow{d} (\mathbb{G}_{\Gamma_1}, \mathbb{G}_{\Gamma_2}, \mathbb{G}_\mu, \mathbb{G}_\delta)$$

in $\ell^\infty(\mathcal{L} \times \mathcal{S}) \times \ell^\infty(\mathcal{L} \times \mathcal{S} \times \mathcal{T}) \times \ell^\infty(\mathcal{S}) \times \ell^\infty(\mathcal{S} \times \mathcal{T})$. Also, $\hat{\Gamma}^\star \xrightarrow{p} \Gamma^\star$ in $\ell^\infty(\mathcal{L})$.

(iv) $\mathcal{L}_n \subseteq \mathcal{L}$ and there exists $\Pi_n : \mathcal{L} \to \mathcal{L}_n$ such that $\Gamma^\star \circ \Pi_n = \Gamma^\star + o(1)$ in $\ell^\infty(\mathcal{L})$, $\sqrt{n}(\hat{\Gamma}_1 \circ \Pi_n - \hat{\Gamma}_1) = o_p(1)$ in $\ell^\infty(\mathcal{L} \times \mathcal{S})$, and $\sqrt{n}(\hat{\Gamma}_2 \circ \Pi_n - \hat{\Gamma}_2) = o_p(1)$ in $\ell^\infty(\mathcal{L} \times \mathcal{S} \times \mathcal{T})$.

(v) $\sqrt{n}\kappa_n \uparrow \infty$ and $\kappa_n \to 0$.

Assumption 11(iii) places high-level conditions on our first-step estimators, which can be interpreted as requiring the central limit theorem to be applied uniformly. Assumption 11 (iv) requires the approximation from employing $\mathcal{L}_n$ instead of $\mathcal{L}$ to vanish sufficiently quickly.

Proposition 8. Suppose Assumption 11 holds. Then, $\hat{\underline{\gamma}}^\star \xrightarrow{p} \underline{\gamma}^\star$ and $\hat{\overline{\gamma}}^\star \xrightarrow{p} \overline{\gamma}^\star$.

The proof of Proposition 8 follows directly from Theorem 1 in Mogstad et al. (2018) and is thus omitted.

B.4 The Bootstrap Statistic

We follow Mogstad et al. (2017) to construct the bootstrap statistics. Recall that we approximate $\mathcal{L}$ with a finite-dimensional space $\mathcal{L}_n$. We also assume that $|\mathcal{S}|$ and $|\mathcal{T}|$ are finite. Then, we can regard μ and $\Gamma_1(\lambda)$ (for a fixed λ) as vectors of length $|\mathcal{S}|$. Similarly, we can view δ and $\Gamma_2(\lambda)$ (for a fixed λ) as vectors of length $|\mathcal{S}| \times |\mathcal{T}|$. Denote

$$\beta = \begin{pmatrix} \gamma^\star \\ \mu \\ \delta \end{pmatrix}, \quad \Gamma(\lambda, \eta) = \begin{pmatrix} \Gamma^\star(\lambda) \\ \Gamma_1(\lambda) \\ \Gamma_2(\lambda) - \eta \end{pmatrix}.$$

Assume $\beta \in \mathbb{R}^{d_\beta}$ and define $1-$norm on $\mathbb{R}^{d_\beta}$: $\|\beta\|_1 = \sum_{l=1}^{d_\beta} |\beta_l|$. Then, our test statistic T_n defined in (28) can be written as

$$T_n = \inf_{(\lambda \times \eta) \in \mathcal{L}_n \times \mathcal{E}} \|\hat{\beta} - \hat{\Gamma}(\lambda, \eta)\|_1,$$

where $\hat{\beta}$ and $\hat{\Gamma}$ are corresponding estimators for β and Γ. Define bootstrap analogs $\hat{\beta}^{\text{bs}}$ and $\hat{\Gamma}^{\text{bs}}$ for $\hat{\beta}$ and $\hat{\Gamma}$. Given them, define $\hat{\mathbb{G}}_\beta = \sqrt{n}(\hat{\beta}^{\text{bs}} - \hat{\beta})$ and $\hat{\mathbb{G}}_\Gamma = \sqrt{n}(\hat{\Gamma}^{\text{bs}} - \hat{\Gamma})$.

Some additional notations are required. Define the unit ball in $\mathbb{R}^{d_\beta}$ as $\mathcal{D} = \{b^* \in \mathbb{R}^{d_\beta} : \|b^*\|_\infty \leq 1\}$, where $\|b^*\|_\infty$ is the max-norm on $\mathbb{R}^{d_\beta}$. For any convex set $\mathcal{C} \in \mathbb{R}^{d_\beta}$, define its support function $\nu(\cdot, \mathcal{C}) : \mathcal{D} \mapsto \mathbb{R}$ as $\nu(b^*, \mathcal{C}) = \sup_{b \in \mathcal{C}} (b^*)'b$. Define

$$\hat{\mathcal{D}}_n = \{b^* \in \mathcal{D} : (b^*)'\hat{\beta} - \nu(b^*, \hat{\Gamma}(\mathcal{L}_n \times \mathcal{E})) \geq -\hat{\kappa}_n^u\},$$

where $\hat{\kappa}_n^u$ is a bandwidth chosen by the researcher. Let $(\mathcal{L} \times \mathcal{E})^*$ denote the dual space of $\mathcal{L} \times \mathcal{E}$. For any $b^* \in \mathbb{R}^{d_\beta}$, define

$$\hat{\mathcal{G}}_n(b^*) = \{g \in (\mathcal{L} \times \mathcal{E})^* : |g(v) - (b^*)'\hat{\Gamma}(v)| \leq \hat{\kappa}_n^g, \text{ for all } v \in \mathcal{V}_n\},$$

where $\hat{\kappa}_n^g \downarrow 0$ and $\mathcal{V}_n \subseteq \mathbf{L}$ are chosen by the researcher.

The bootstrap statistic is defined as

$$T_n^{\text{bs}} = \sup_{b^* \in \hat{\mathcal{D}}_n} \sup_{g \in \hat{\mathcal{G}}_n(b^*), (\tilde{\lambda}, \tilde{\eta}) \in \hat{\mathcal{R}}_n} \inf_{(\lambda, \eta) \in \hat{\mathcal{R}}_n} \Big\{ (b^*)' \left[\hat{\mathbb{G}}_\beta - \hat{\mathbb{G}}_\Gamma(\lambda, \eta) \right] \\ \text{s.t. } g(\lambda - \tilde{\lambda}, \eta - \tilde{\eta}) \geq 0 \Big\}, \tag{50}$$

where

$$\hat{\mathcal{R}}_n = \{(\lambda, \eta) \in \mathcal{L}_n \times \mathcal{E} : \|\hat{\beta} - \hat{\Gamma}(\lambda, \eta)\|_1 \leq \hat{\kappa}_n^m\}.$$

Here, $\hat{\kappa}_n^m$ is another tuning variable. As shown in Mogstad et al. (2017), T_n^{bs} can be reformulated as the solution of a bilinear program. Mogstad et al. (2017) also suggest data-driven choices for $\hat{\kappa}_n^u$ and $\hat{\kappa}_n^g$ by solving a number of mixed integer linear problems, as well as $\hat{\kappa}_n^m = \hat{\kappa}_n^u + T_n$; see Mogstad et al. (2017) for more details. We use Gurobi to solve these programs. The critical value $\hat{c}_{1-\alpha}$ is given by $(1-\alpha)$th quantile of the bootstrap sample $\{T_{n,i}^{\text{bs}}\}_{i=1}^B$.

C Additional Results

C.1 Systematic Error at $X^* = 0$

In this subsection, we extend our framework to scenarios when Assumption 3 fails. Denote the systematic error at $X^* = 0$ as a function below:

$$\alpha_0(z) = \mathbb{E}[X \mid X^* = 0, Z = z]. \tag{51}$$

Then, we can write

$$\mathbb{E}[X \mid X^* = x^*, Z = z] = \lambda(x^*, z) \cdot x^* + \alpha_0(z) \cdot \mathbf{1}\{x^* = 0\}. \tag{52}$$

We assume that $(\lambda, \alpha_0) \in \mathcal{L} \times \mathcal{A}$ for two known sets $\mathcal{L}$ and $\mathcal{A}$. By law of iterated expectation, the target parameter can be expressed as

$$\gamma = \mathbb{E}[X \cdot h(X^*, Z)] = \mathbb{E}\left[\lambda(X^*, Z) \cdot X^* h(X^*, Z) + \alpha_0(Z) \cdot \mathbf{1}\{X^* = 0\} \cdot h(X^*, Z)\right], \tag{53}$$

which is linear in (λ, α). Similarly, we can modify linear moment conditions in Proposition 2 as

$$\mathbb{E}\Big[\lambda(X^*, Z) \cdot s(Z)X^* + \alpha_0(Z) \cdot s(Z)\mathbf{1}\{X^* = 0\}\Big] = \mu_s, \tag{54}$$

$$\mathbb{E}\Big[\lambda(X^*, Z) \cdot s(Z)X^* \cdot d_t(X^*) + \alpha_0(Z) \cdot s(Z)\mathbf{1}\{X^* = 0\} \cdot d_t(X^*)\Big] \geq \delta_{s,t}. \tag{55}$$

From (53),define target parameter as $\gamma = \Gamma^\star(\lambda, \alpha)$, where $\Gamma^\star : \mathcal{L} \times \mathcal{A} \to \mathbb{R}$ is a linear map. Bounds of the target parameter γ are characterized as

$$\underline{\gamma}^\star = \inf_{(\lambda,\alpha_0)\in\mathcal{L}\times\mathcal{A}} \Gamma^\star(\lambda, \alpha_0), \quad \text{s.t. (54) and (55) hold for all } s \in \mathcal{S} \text{ and } t \in \mathcal{T},$$

$$\overline{\gamma}^\star = \inf_{(\lambda,\alpha_0)\in\mathcal{L}\times\mathcal{A}} \Gamma^\star(\lambda, \alpha_0), \quad \text{s.t. (54) and (55) hold for all } s \in \mathcal{S} \text{ and } t \in \mathcal{T},$$

where $\mathcal{S}$ denote a collection of function s described in Proposition 2 and $\mathcal{T} \subseteq \mathbb{R}$.

C.2 More Linear Moment Inequalities

In the following proposition, we demonstrate how to derive more general linear moment inequalities from inequality (12).

Proposition 9. Suppose Assumptions 1, 2 and 3 hold. Suppose that $U \subseteq \mathbb{R}$ and $s : \mathbb{R}^{d_Z} \to \mathbb{R}^+ \cup \{0\}$ is an identified (or known) function that is measurable and has a finite second moment. Define indicator function $\mathbf{d}_U(u) = \mathbf{1}\{u \in U\}$. Then, we have

$$\mathbb{E}\Big[\lambda(X^*, Z) \cdot s(Z)X^* \cdot \mathbf{d}_U(X^*)\Big] \geq \boldsymbol{\delta}_{s,U}, \tag{56}$$

where $\boldsymbol{\delta}_{s,U} = \sup_{x\in\mathbb{R}} \mathbb{E}\left[s(Z)\Big(x \cdot \mathbf{d}_U(X^*) - (x - X)^+\Big)\right]$ is point-identified.

Proof of Proposition 9. Following the proof of Proposition 2, we can obtain

$$\mathbb{E}\Big[\lambda(X^*, Z) \cdot s(Z)X^* \cdot \mathbf{d}_U(X^*)\Big] \geq \mathbb{E}\left[s(Z)\Big(x \cdot \mathbf{d}_U(X^*) - (x - X)^+\Big)\right].$$

Note that the left side does not depend on x, we finish the proof. *Q.E.D.*

Proposition 9 shows that we can generate many linear moment inequalities by varying the choice of the subset $U \subseteq \mathbb{R}$. In particular, the moment inequality (14) in Proposition 2 corresponds the special case $U = (-\infty, t]$ for some threshold t. Under Assumption 4, Proposition 4 further shows that it is possible to select the collection $\mathcal{U} = \{(-\infty, t] : t \in \text{Supp}(X^*)\}$ to exhaust all the information from the observed data.

When Assumption 4 does not hold, it may be necessary to consider a richer class of subsets $U \subseteq \mathbb{R}$ to extract more identifying information from the data. For example, if X^* is a discrete random variable and there are no covariates, Proposition 2 yields $|\text{Supp}(X^*)|$ moment inequalities by choosing $t \in \mathcal{T} = \text{Supp}(X^*)$. In contrast, Proposition 9 allows for arbitrary subsets U, generating up to $2^{|\text{Supp}(X^*)|}$ moment inequalities, and thereby offering a much richer set of identification conditions.

C.3 Additional Tables in the Numerical Illustration

Table 5: Bounds on γ_1 under various assumptions

Assumptions	$(\lambda_l, \lambda_u) = (0.9, 1.1)$	$(\lambda_l, \lambda_u) = (0.9, 1.2)$	True value
Only bounds of $\lambda(\cdot)$	[0.426, 0.868]	[0.020, 0.982]	0.783
$\mathbb{E}[X \mid X^*]$ increasing	[0.453, 0.868]	[0.128, 0.982]	0.783
$\lambda(\cdot)$ increasing	[0.638, 0.868]	[0.638, 0.981]	0.783
$\lambda(\cdot)$ increasing + concave	[0.638, 0.858]	[0.638, 0.960]	0.783

Notes: See that in Table 1.

Table 6: Bounds on γ_2 under various assumptions

Assumptions	$(\lambda_l, \lambda_u) = (0.9, 1.1)$	$(\lambda_l, \lambda_u) = (0.9, 1.2)$	True value
Only bounds of $\lambda(\cdot)$	[1.031, 1.100]	[1.011, 1.194]	1.043
$\mathbb{E}[X \mid X^*]$ increasing	[1.031, 1.100]	[1.011, 1.168]	1.043
$\lambda(\cdot)$ increasing	[1.031, 1.069]	[1.011, 1.069]	1.043
$\lambda(\cdot)$ increasing + concave	[1.031, 1.069]	[1.011, 1.069]	1.043

Notes: See that in Table 1.