Assessing Measurement Error in Linear Instrumental Variables Models*

Angda Li

Department of Economics, Texas A&M University

August 16, 2025

Abstract

In linear regression analysis, it is common to use instruments to address measurement error in the regressor. However, bias can still arise if the measurement error correlates with either the true variable, other regressors, or the instrument. This paper develops a sensitivity analysis framework for linear instrumental variables (IV) models that accounts for such concerns. We establish bounds for the parameter of interest using a set of sensitivity parameters that restrict the consistent deviations of the measurement from the true variable. We illustrate our methods in an empirical study that uses twins data to analyze the effect of schooling level on wages.

Keywords: Linear regression; Systematic error; Sensitivity analysis; Partial identification

*I am especially indebted to Yonghong An, Jackson Bunting, and Qi Li for their guidance and support. I am also grateful to Marco Castillo and Ragan Petrie for their helpful comments.

1 Introduction

Concerns about measurement error in linear models are prevalent in economics. When the regressor is contaminated by measurement error, the usual result is that the ordinary least squares (OLS) estimator for the parameter of interest is subject to attenuation bias. A straightforward approach to address this issue is to use instrumental variables (IV) (e.g., Farber et al., 2021; Hoffman and Tadelis, 2021; Rogall, 2021).

However, the use of IV estimators for addressing measurement error has limitations. First, non-classical measurement error, which indicates a correlation between measurement error and the true value, can lead to bias in IV estimators (Black et al., 2000). Such concerns are common in self-reported data due to systematic misreporting (e.g., Bollinger, 1998; Blattman et al., 2017; Abay et al., 2019; Bick et al., 2022). Second, Measurement error in self-reported data can correlate with demographic variables (e.g., Bound and Krueger, 1991; Bound et al., 1994; Angel et al., 2019), which may result in biased IV estimators if these variables are included as controls in the regression. Finally, when the instrument is a second measurement, bias can arise due to correlated measurement errors (e.g., Ashenfelter and Krueger, 1994; Giglio et al., 2021). Therefore, it is natural to question how to assess the bias in IV estimators given these concerns.

Without additional information (e.g., validation data), little is known about the nature of measurement error. A possible approach to address these concerns is to explore the robustness of empirical results to alternative measures (e.g., Allcott et al., 2016; Bastos et al., 2018; Dobbie et al., 2018; Macchiavello and Morjaria, 2021). If the coefficient remains stable under alternative measures, then the bias in IV estimators due to measurement error may be limited. However, it is important to note that there is no formal econometric theory supporting such conclusions[1].

The main contribution of this paper is to propose a framework for assessing sensitivity to general measurement error in linear IV models. This framework allows the measurement

[1]The robustness check approach may lead to incorrect conclusions. See Section 4.2 for discussion.

error to be flexibly correlated with the true variable, other covariates, and the instrument. We derive bounds for the parameter of interest using a set of sensitivity parameters that impose that the measurement does not consistently deviate "too much" from the true value. Applied to self-reported data, our key sensitivity parameters can be interpreted in terms of the rates of underreporting and overreporting.

More concretely, the parameter of interest is β, the coefficient on an unobserved variable X^* in the regression of outcome Y on X^* and other covariates W. Since X^* is unobserved, we use a noisy measurement X and an instrument Z. The IV estimand β^{IV} is the coefficient on X from regressing Y on X and W, instrumenting X with Z. In this linear IV model, measurement error u can be decomposed as

$$u = \underbrace{\Big(X - \mathbb{E}[X \mid X^*, W, Z]\Big)}_{\text{random error}} + \underbrace{\Big(\mathbb{E}[X \mid X^*, W, Z] - X^*\Big)}_{\text{systematic error}} \tag{1}$$

where we borrow notions *random error* and *systematic error* from experimental physics[2]. We begin by considering a baseline assumption of "no systematic error", which implies that IV estimand β^{IV} equals β. Hence, any bias in IV estimand β^{IV} (i.e., $\beta^{\text{IV}} \neq \beta$) can be attributed to the systematic error.

To relax the baseline assumption, we propose a relative measure of systematic error called *slope function*. Our identification results use the lower and upper bounds of this function as sensitivity parameters (λ_l and λ_u), which restrict the extent of systematic error in the measurement. Our first and main result provides closed-form bounds for β consistent with restrictions imposed by (λ_l, λ_u). This result requires no systematic error at $X^* = 0$, but it does not impose conditional mean independence or parametric assumptions on the measurement error. As discussed in Section 3.2, these bounds involve only one sensitivity parameter after reparameterization in several special cases (e.g., λ_l and λ_u are symmetric

[2] Chapter 4 in Taylor (1997) stated that "Experimental uncertainties that can be revealed by repeating the measurements are called random errors; those that cannot be revealed in this way are called systematic errors."

across 1).

Our main result can be used in two ways. First, we show that validation data can help identify sensitivity parameters (λ_l, λ_u) under mild assumptions and then characterize bounds for β consistent with data. Second, researchers can assess the sensitivity of their empirical conclusions to the extent of systematic error by varying sensitivity parameters. When there is only one sensitivity parameter, we provide simple expressions for the breakdown point, defined as the largest deviation we can allow while maintaining a specific baseline finding according to our bounds.

We provide two extensions of the main result. Firstly, if the true variable has a point mass at zero, we generalize our bounds by incorporating a sensitivity parameter to restrict systematic error at $X^* = 0$. Secondly, in scenarios with multiple measurements, we characterize the bounds as intersection bounds and demonstrate that the model is falsifiable.

We apply our framework to the twins data from Ashenfelter and Krueger (1994), who studied the effect of schooling level on wages. We show that their high estimates of the return to education are robust to the presence of systematic error. This indicates that their unusual results might stem from sampling error, as Rouse (1999) concluded.

After discussing the related literature, we introduce the model and the decomposition of general measurement error in Section 2. Section 3 presents the main identification results and discusses their practical application and interpretation. In Section 4, we propose two extensions. Section 5 considers an empirical application. Section 6 concludes. All proofs and additional supporting results can be found in the appendix.

Related Literature

This paper relates to the literature on the impact of (non-classical) measurement error on OLS and IV estimators; see, for example, Ashenfelter and Krueger (1994), Black et al. (2000), Hyslop and Imbens (2001), and Abay et al. (2019). These studies focus on the direction of bias but do not develop a formal sensitivity analysis. In contrast, our approach introduces a different set of assumptions: instead of restricting the correlations between measurement error and (X^*, W, Z), we bound the extent of systematic error, which is easy to interpret and permits flexible dependence on (X^*, W, Z). For example, Black et al. (2000) assumes that the measurement error is negatively correlated with X^* and does not correlate with covariates W; these conditions are not necessary in our approach.

Several other works have conducted sensitivity analysis of measurement error in linear models. For the binary mismeasured regressor, restrictions are typically imposed through misclassification probabilities (e.g., Bollinger, 1996; Kreider and Pepper, 2007; Jiang and Ding, 2020). While these methods are straightforward, it remains unclear how to extend them to the continuous regressor. Some approaches that allow for measurement error in continuous variables include Klepper and Leamer (1984), Erickson (1993), Bollinger (2003), Chalak and Kim (2020) and DiTraglia and García-Jimeno (2021). Unlike these papers, which assume measurement error is uncorrelated with the true variable, our framework allows for more flexible correlations.

Our paper also fits into recent literature on sensitivity analysis of various endogeneity in linear models, including omitted variable bias (e.g., Oster, 2019; Cinelli and Hazlett, 2020; Diegert et al., 2023; Masten and Poirier, 2024) and invalid IV (e.g., Small, 2007; Ashley, 2009; Masten and Poirier, 2021; Cinelli and Hazlett, 2025). Compared with these papers, our method is specific to the context of measurement error.

Finally, our work is complementary to the literature on the nonparametric identification of models with non-classical measurement error (e.g., Hu and Schennach, 2008; Hu et al., 2022). While these papers concentrate on the point identification of nonparametric structural functions, our aim is to characterize the bounds for a specific regression coefficient of interest. Furthermore, our framework does not depend on conditional independence assumptions, nor does it require the normalization of the latent variable.

Notation Remark

For random variable A and random $(k \times 1)$ vector B , define $A^{\perp B} = A - B' \cdot (\mathbb{E}[BB'])^{-1}\mathbb{E}[BA]$, which is the residual from a linear projection A on B. Note by definition $A^{\perp B}$ is a random variable and $\mathbb{E}[A^{\perp B}B] = 0$. When B includes a constant, we have $\mathbb{E}[A^{\perp B}] = 0$.

2 Measurement Error in Linear IV Models

We study a classical linear IV model with a mismeasured regressor. In this section, we set up notations and define a decomposition for general measurement error. We then state a baseline assumption, which point identifies the parameter of interest.

2.1 Basic Setup

Let Y, X^*, X, and Z be scalar variables. Let W_0 be a vector of observed covariates of dimension d and $W = (1, W_0)$. The following assumption ensures the parameter we consider is well-defined.

Assumption 1. The variance matrix of (Y, X^*, W, Z, X) is finite and positive definite.

We can write

$$Y = \beta X^* + \gamma' W + \epsilon \tag{2}$$

where $\epsilon = Y^{\perp(X^*,W)}$ and is uncorrelated with (X^*, W) by construction. We do not require $\mathbb{E}[\epsilon \mid X^*, W] = 0$, which allows for possible misspecification in the linear model. Suppose the parameter of interest is the coefficient β.

The regressor X^* is not observed; instead, we observe a measurement X. Define measurement error

$$u = X - X^* \tag{3}$$

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 this is not the case, we refer to the measurement error as non-classical, indicating a correlation between u and X^*.

The OLS estimator of β is biased towards zero in traditional errors-in-variables models where u is uncorrelated with (X^*, W, ϵ) (see Chapter 9.4 in Wooldridge, 2013). To address the bias, we use an instrument variable Z, which is possibly the second measurement.

Assumption 2. (i) $\text{Cov}(Z, \epsilon) = 0$. (ii) $\text{Cov}(Z^{\perp W}, X^{\perp W}) \neq 0$.

Assumption 2 states the exclusion and relevance of instrument variable Z. To keep the discussion simple, we assume $\text{Cov}(Z^{\perp W}, X^{\perp W}) > 0$ throughout the rest of the paper.

2.2 Bias in IV Estimand

Let β^{IV} denote the coefficient on X in the IV estimand of Y on (X, W) with the instrument variable Z:

$$\beta^{\text{IV}} = \frac{\text{Cov}(Z^{\perp W}, Y^{\perp W})}{\text{Cov}(Z^{\perp W}, X^{\perp W})} \tag{4}$$

In the following analysis, we explain how measurement error is likely to yield bias in β^{IV} ($\beta^{\text{IV}} \neq \beta$). Consider the OLS estimand of Z on W, we can write

$$Z = \pi' W + Z^{\perp W}$$

where each component of W is uncorrelated with $Z^{\perp W}$ by construction. By equation (4), we have[3]

$$\beta^{\text{IV}} - \beta = -\beta \cdot \frac{\text{Cov}(Z^{\perp W}, u)}{\text{Cov}(Z^{\perp W}, X)} = -\beta \cdot \frac{\text{Cov}(Z - \pi' W, u)}{\text{Cov}(Z^{\perp W}, X)} \tag{5}$$

which means the bias arises from the correlation between the measurement error u and Z or W. Although not always the case, non-classical measurement error could be a reason behind this since u is likely to correlate with Z due to the correlation between X^* and Z.

We discuss three examples in the empirical studies below:

[3] See equation (26) in Appendix A for detailed derivation.

Example 1. Bick et al. (2022) showed evidence that workers who work for long hours (more than 45 hours per week) tend to overreport their hours. Let X^* = true working hours, X = reporting working hours. It is possible to have

$$\text{Cov}(X^*, u) > 0 \Rightarrow \text{Cov}(Z, u) > 0 \Rightarrow \text{Cov}(Z^{\perp W}, u) \neq 0$$

We can raise similar concerns for self-reported income or expenditure data.

Example 2. Angel et al. (2019) found that males tend to overreport wages, pensions, and unemployment benefits more than females, driven by a desire to showcase social status. If we include the dummy variable W (1 for male and 0 for female) in the regression, we have:

$$\text{Cov}(W, u) \neq 0 \Rightarrow \text{Cov}(Z^{\perp W}, u) \neq 0$$

Example 3. Gillen et al. (2019) suggested using multiple elicitations to address the measurement error in experimental studies. They choose another measurement for X^* as the instrument variable; that is

$$Z = X^* + \xi$$

When employing this method, Giglio et al. (2021) raised the concern that errors may be positively correlated across the two elicitations, which means $\text{Cov}(\xi, u) > 0$. In this case, if both u and ξ are independent of X^*, we have

$$\text{Cov}(Z, u) = \text{Cov}(\xi, u) \neq 0 \Rightarrow \text{Cov}(Z^{\perp W}, u) \neq 0$$

The three examples provided illustrate different ways in which the error u could be correlated with $Z^{\perp W}$, potentially leading to bias in β^{IV}. Additionally, multiple channels may exist, which complicates the analysis of measurement error in linear IV models.

2.3 Random Error and Systematic Error

To further analyze the bias in IV estimand, we decompose the measurement error into two components[4]:

$$\begin{aligned} u &= \underbrace{\Big(X - \mathbb{E}[X \mid X^*, W, Z]\Big)}_{\text{random error } \tilde{u}} + \underbrace{\Big(\mathbb{E}[X \mid X^*, W, Z] - X^*\Big)}_{\text{systematic error}} \\ &= \tilde{u} + \mathbb{E}[u \mid X^*, W, Z] \end{aligned} \tag{6}$$

This decomposition is unique and does not rely on any assumption of the measurement error. By construction, the random error $\tilde{u}$ is mean independent of (X^*, W, 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^*, W, 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.

For instance, self-reported data can be affected by both random and systematic errors[5]. If individuals can not accurately recall events, errors may be random since people may sometimes overestimate and sometimes underestimate. Conversely, if individuals intentionally misreport the true outcome (Abay et al., 2019), the error is systematic.

We introduce a natural baseline assumption below, which implies that the measurement error u is mean independent of the true variable X^*, covariates W, and instrument Z. Denote Ω as the support of random vector (X^*, W, Z).

[4] Hyslop and Imbens (2001) studied the effect of Berkson error ($\mathbb{E}[u \mid X] = 0$) on coefficients. Their decomposition of measurement error is similar to us:

$$u = \Big(X - \mathbb{E}[X \mid X^*]\Big) + \Big(-X^* + \mathbb{E}[X^* \mid X]\Big) + \Big(\mathbb{E}[X \mid X^*] - \mathbb{E}[X^* \mid X]\Big)$$

If X is mean independent of (W, Z) given X^*, our random error becomes their first term, and systematic error is the sum of their remaining two terms.

[5] See Kirkpatrick (2024) for a detailed introduction of random and systematic errors in self-reported dietary intake data.

Assumption 3 (No systematic error)**.** $\mathbb{E}[u \mid X^* = x^*, W = w, Z = z] = 0$ for all $(x^*, w, z) \in \Omega$.

It is evident that Assumption 3 does not hold in examples 1, 2, and 3, as their u is correlated with one of (X^*, W, Z) respectively. We have the following equivalent characterization.

Proposition 1. *Assumption 3 holds if and only if for all* $(x^*, w, z) \in \Omega$

(i) $\mathbb{E}[X \mid X^* = x^*] = x^*$.

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

Our baseline assumption is stronger than the *weakly classical* in Schennach (2022), which is shown in condition (i). Additionally, we include that the measurement X is mean independent of (W, Z) conditional on X^* in condition (ii). With no systematic error, we have the following result.

Proposition 2. *Suppose the joint distribution of* (Y, X, W, Z) *is known. Suppose Assumptions 1, 2 and 3 hold. Then* $\beta^{\mathrm{IV}} = \beta$. *Consequently,* β *is point identified.*

The point identification result of β with a valid instrument Z is widely used in the literature. We highlight its sufficient condition[6] in Proposition 2. Therefore, the bias in β^{IV} can be attributed to systematic error.

3 Sensitivity Analysis to the Systematic Error

In this section, we assess the impact of the systematic error on IV estimands. Specifically, we allow both conditions (i) and (ii) in Proposition 1 to be violated.

[6]A sufficient and necessary condition is that $\mathrm{Cov}(Z^{\perp W}, u) = 0$, which is weaker than our baseline assumption but not as easily understood. Nevertheless, we suggest that researchers carefully examine the systematic error when using IV estimators.

Table 1: Slope Functions under Different Assumptions of Measurement Error

Assumptions on Measurement Error	Characterizations of Slope Function
Baseline Assumption	$\lambda(x^*, w, z) = 1$
Mean independence	$\lambda(x^*, w, z) = \tilde{\lambda}(x^*)$
Overreporting	$\lambda(x^*, w, z) > 1$
Underreporting	$\lambda(x^*, w, z) < 1$

Notes: Mean independence refers to condition (ii) in Proposition 1. If X is mean independent of Z conditional on (X^*, W), we have $\lambda(x^*, w, z) = \tilde{\lambda}(x^*, w)$. Overreporting refers to $X > X^*$, while underreporting means that $X < X^*$.

3.1 A Relative Measure of the Systematic Error

Recall that the systematic error is $\mathbb{E}[X \mid X^*, W, Z] - X^*$. We begin with a relative measure of systematic error. Let $\Omega_0 = \{(x^*, w, z) \in \Omega : x^* \neq 0\}$. Define *slope function* on Ω_0

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

which can be interpreted as a "measure of location" (Hu and Schennach, 2008). In particular, our baseline assumption of "no systematic error" implies that $\lambda(x^*, w, z) = 1$. Table 1 provides equivalent characterizations of the slope function under different assumptions of measurement error.

We will reinterpret Examples 1-3 using the slope function below.

Example 1 (Continued)**.** Recall that workers tend to overreport their hours when they work more than 45 hours per week. For $x^* > 45$, we have $\lambda(x^*, w, z) > 1$.

Example 2 (Continued)**.** Recall that males ($W = 1$) tend to overreport income more than females ($W = 0$). Then, we have $\lambda(x^*, w = 1, z) > \lambda(x^*, w = 0, z)$.

Example 3 (Continued)**.** We can regard X and Z as two measurements for X^*. Suppose their measurement errors u and ξ are independent of (X^*, W, Z) and follow the joint normal distribution:

$$\begin{pmatrix} u \\ \xi \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right], \text{ where } \sigma_1, \sigma_2, \rho > 0$$

Then, we have

$$\mathbb{E}[X \mid X^* = x^*, W = w, Z = z] = x^* + \mathbb{E}[u \mid \xi = z - x^*] = x^* + \rho \cdot \frac{\sigma_1}{\sigma_2} \cdot (z - x^*)$$

which implies $\lambda(x^*, w, z) > 1$ when $z > x^*$ and $\lambda(x^*, w, z) < 1$ when $z < x^*$.

We relax the baseline assumption below.

Assumption 4. (i) $\Pr(X^* = 0) = 0$ or $\mathbb{E}[X \mid X^* = 0, W, Z] = 0$. (ii) (**bounds of the slope function**) There exists known parameters $\lambda_u \geq \lambda_l > 0$ such that

$$\lambda_l \leq \lambda(x^*, w, z) \leq \lambda_u, \ \forall (x^*, w, z) \in \Omega_0$$

Assumption 4 (i) implies there is no point mass or no systematic error at $X^* = 0$, which ensures a key equation holds almost surely[7]:

$$X = \lambda(X^*, W, Z) \cdot X^* + \tilde{u} \tag{7}$$

where $\tilde{u}$ is the random error defined in (6). This equation connects the measurement with the true variable through the slope function. Assumption 4 (ii) states that the lower and upper bounds of the slope function (λ_l and λ_u) are known to researchers. This assumption includes the baseline assumption as a special case where $\lambda_l = \lambda_u = 1$. Moreover, it allows for deviations from the baseline assumption by allowing λ_l, λ_u to differ from 1.

Assumption 4 has two important features: First, it does not impose mean independence

[7]For $x^* = 0$, $\lambda(x^*, w, z)$ can take any number between λ_l and λ_u.

restrictions on the slope function, as described in Table 1. This means that we allow for flexible correlations between measurement error and (X^*, W, Z). Second, the slope function does not require a particular functional form. Researchers only need to reason about the bounds of the slope function.

One notable case occurs when the slope function degenerates to a constant $\lambda \neq 1$. In this situation, we can express the relationship as follows[8]:

$$X = \lambda X^* + \tilde{u} \tag{8}$$

where $\tilde{u}$ is the random error. We discuss its properties in Appendix B.

Interpreting Bounds of the Slope Function

Our sensitivity parameters (λ_l, λ_u) specify the largest deviations from the baseline assumption. We make several remarks on how to interpret them in practice.

First, we can interpret and calibrate (λ_l, λ_u) based on the deviation rate of the conditional mean of the measurement from the true value. Take self-reported data, for example; the slope function represents the average rates of underreporting or overreporting within specific subgroups, which rules out the effect of random error on the measurement. If subjects tend to overreport the true value, we can set $\lambda_l = 1$ and utilize our prior knowledge to calibrate λ_u. In general, if the average of self-reports deviates by no more than $\psi_u\%$ above or $\psi_l\%$ below the true value, we set $\lambda_u = 1 + \psi_u\%$ and $\lambda_l = 1 - \psi_l\%$.

Second, we need to evaluate how the variables (W, Z) affect the slope function when X^* is fixed. If the error u is mean independent of (W, Z) given X^*, there is no additional effect after controlling for X^*. Otherwise, we need to assess the degree of underreporting and overreporting in certain subgroups characterized by (W, Z). Likewise, if the error u is mean independent of Z given (X^*, W), we only need to consider the extra effects of W.

[8] Haider and Solon (2006) and An et al. (2022) modeled the relationship between observed income (denoted as X) and permanent income (denoted as X^*) using parametric expression (8), with slightly different restrictions on $\tilde{u}$.

Finally, the analysis mentioned above does not require the true variable to be continuous or discrete. While the baseline assumption is less likely to hold for discrete variables (Black et al., 2000), we can still apply the slope function to ordinal outcomes, such as years of schooling in our empirical application.

3.2 Identification Using bounds of the Slope Function

Recall from Section 2 that our parameter of interest is β, the OLS coefficient on X^* in the regression of Y on (X^*, W). Since X^* is not observed, we cannnot compute this regression from the data. Instead, we rely on the IV estimand β^{IV}, defined in equation (4). Our main result, Theorem 1 below, characterizes the bounds of β using bounds of the slope function in Assumption 4.

Theorem 1. *Define*

$$\alpha = \frac{\mathbb{E}(|X \cdot Z^{\perp W}|) - \text{Cov}(Z^{\perp W}, X)}{2\text{Cov}(Z^{\perp W}, X)} \geq 0$$

Suppose the joint distribution of (Y, X, W, Z) is known. Suppose Assumptions 1, 2, 4 and $\lambda_u/\lambda_l < 1 + \alpha^{-1}$ are satisfied. Normalize $\beta^{\text{IV}} > 0$. Then, $\beta > 0$ and

$$\underline{B}(\lambda_l, \lambda_u) \leq \beta \leq \overline{B}(\lambda_l, \lambda_u)$$

where

$$\underline{B}(\lambda_l, \lambda_u) = \beta^{\text{IV}} \Big/ \left(\frac{1+\alpha}{\lambda_l} - \frac{\alpha}{\lambda_u} \right), \quad \overline{B}(\lambda_l, \lambda_u) = \beta^{\text{IV}} \Big/ \left(\frac{1+\alpha}{\lambda_u} - \frac{\alpha}{\lambda_l} \right).$$

We first sketch the proof and then discuss its implications. First, we consider the ratio between β^{IV} and β:

$$\frac{\beta^{\text{IV}}}{\beta} = \frac{\text{Cov}(Z^{\perp W}, X^*)}{\text{Cov}(Z^{\perp W}, X)}.$$

The second step is to notice that bounding $\text{Cov}(Z^{\perp W}, X^*)$ is similar to that for interval data in Bontemps et al. (2012). To see the intuition, consider the extreme case $\tilde{u} = 0$ (no random error), restrictions in Assumption 4 (ii) become $X/X^* \in [\lambda_l, \lambda_u]$. This allows us to derive its bounds based on (λ_l, λ_u). Finally, we obtain

$$\frac{1}{\lambda_u} - (\frac{1}{\lambda_l} - \frac{1}{\lambda_u}) \cdot \alpha \leq \frac{\beta^{\text{IV}}}{\beta} \leq \frac{1}{\lambda_l} + (\frac{1}{\lambda_l} - \frac{1}{\lambda_u}) \cdot \alpha \tag{9}$$

The length of the bounds for β^{IV}/β is given by $(1 + 2\alpha) \cdot \left(\frac{1}{\lambda_l} - \frac{1}{\lambda_u}\right)$. It is important to note that α is identified from observed data. A larger value of α results in wider bounds for β, which makes it more sensitive to systematic errors. Additionally, a stronger correlation between $X^{\perp W}$ and $Z^{\perp W}$ typically leads to a smaller value of α.

Since Theorem 1 provides simple and closed-form expressions for the lower and upper bounds of β, we can immediately derive some properties. First, when $\lambda_l = \lambda_u = 1$, the bounds collapse to β^{IV}, the point estimand from the baseline model with no systematic error. Second, as sensitivity parameters change, the bounds may not be a singleton, but they change continuously with λ_l and λ_u. Both lower and upper bounds can be derived from the second moments of (Y, X, W, Z) and the sensitivity parameters. Finally, these bounds are positive only if $\lambda_u/\lambda_l < 1 + \alpha^{-1}$. This condition ensures that β^{IV} and β take the same sign[9].

We discuss three special cases of Theorem 1 below, each involving only one sensitivity parameter.

Example 4. (λ_l and λ_u are symmetric across 1) Let $\lambda_l = 1 - \psi$ and $\lambda_u = 1 + \psi$ for some $\psi \in [0, \frac{1}{2\alpha+1})$, where ψ denotes the largest deviation rate of the conditional mean of the measurement from the true variable. The bounds become

$$\begin{aligned} &[\underline{B}(1-\psi, 1+\psi), \quad \overline{B}(1-\psi, 1+\psi)] \\ &= \left[\frac{1-\psi^2}{1+(1+2\alpha)\psi} \cdot \beta^{\text{IV}}, \quad \frac{1-\psi^2}{1-(1+2\alpha)\psi} \cdot \beta^{\text{IV}}\right] \end{aligned} \tag{10}$$

[9]If $\lambda_u/\lambda_l > 1 + \alpha^{-1}$, we have $\overline{B}(\lambda_l, \lambda_u) < 0$ and $\beta \in (-\infty, \overline{B}(\lambda_l, \lambda_u)] \cup [\underline{B}(\lambda_l, \lambda_u), +\infty)$ by bound (9).

Example 5. (Case of overreporting) Let $\lambda_l = 1$ and $\lambda_u = 1 + \psi$ for some $\psi \in [0, \alpha^{-1})$, where ψ denotes the largest average rate of overreporting. The bounds become

$$\begin{aligned} &[\underline{B}(1, 1+\psi), \quad \overline{B}(1, 1+\psi)] \\ &= \left[\frac{1+\psi}{1+(1+\alpha)\psi} \cdot \beta^{\text{IV}}, \quad \frac{1+\psi}{1-\alpha\psi} \cdot \beta^{\text{IV}} \right] \end{aligned} \tag{11}$$

Note that the lower bound is smaller than β^{IV}, so overreporting does not necessarily underestimate β under general measurement error.

Example 6. (Case of underreporting) Let $\lambda_l = 1 - \psi$ and $\lambda_u = 1$ for some $\psi \in [0, \frac{1}{\alpha+1})$, where ψ denotes the largest average rate of underreporting. The bounds become

$$\begin{aligned} &[\underline{B}(1-\psi, 1), \quad \overline{B}(1-\psi, 1)] \\ &= \left[\frac{1-\psi}{1+\alpha\psi} \cdot \beta^{\text{IV}}, \quad \frac{1-\psi}{1-(1+\alpha)\psi} \cdot \beta^{\text{IV}} \right] \end{aligned} \tag{12}$$

Numerical Illustration

We conclude this subsection with a brief numerical illustration. The outcome equation is $Y = 1 + 2X^* + \epsilon_y$ ($\beta = 2$), and the instrument is generated by $Z = (X^*)^2 - \log(X^*) + \epsilon_z$, where $X^* \sim \text{Unif}[1, 3]$, $\epsilon_y \sim \mathcal{N}(0, 0.6^2)$ and $\epsilon_z \sim \mathcal{N}(0, 1)$. Consider two measurements with linear slope functions:

$$X_1 = \underbrace{(0.95 + 0.05X^*)}_{\lambda_1(x^*)} \cdot X^* + \tilde{u}_1, \quad X_2 = \underbrace{(1.15 - 0.05X^*)}_{\lambda_2(x^*)} \cdot X^* + \tilde{u}_2$$

where random error $\tilde{u}_i \sim \mathcal{N}(0, 0.3^2)$ ($i = 1, 2$). We plot these two slope functions on the left of Figure 1, which satisfy Assumption 4 with $(\lambda_l, \lambda_u) = (1, 1.1)$. The average degree of overreporting is increasing in x^* for measurement X_1, while it is decreasing in x^* for measurement X_2.

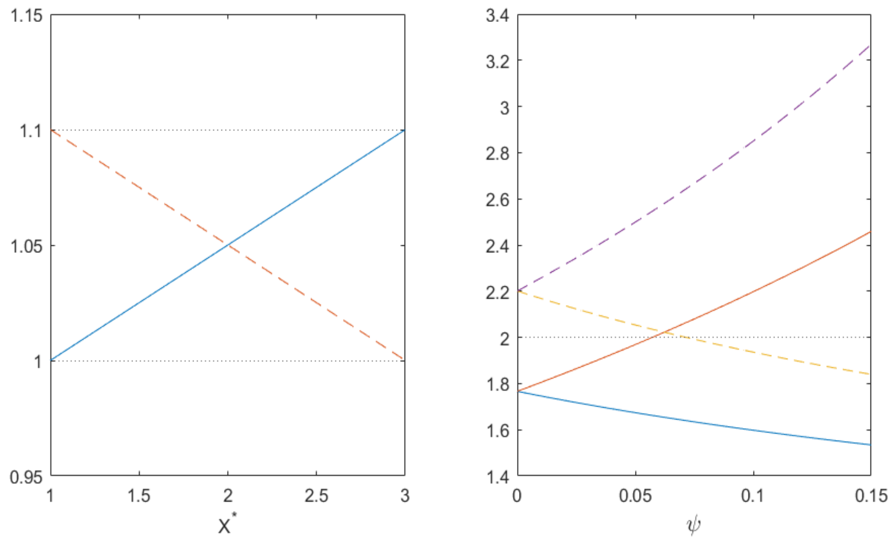

Figure 1: Numerical illustration. Left: solid line and dashed line are slope functions for measurement X_1 and X_2 respectively, and dotted lines are bounds of these two slope functions. Right: solid lines and dashed lines are bounds of β using measurement X_1 and X_2, respectively, and the dotted line represents the true value $\beta = 2$.

The right of Figure 1 presents bounds of β using (11) for two measurements with different deviations ψ. At $\psi = 0$, these bounds are singletons, corresponding to IV estimators: 1.77 for measurement X_1 and 2.20 for measurement X_2. We see that both bounds change continuously and become wider as ψ increases. Moreover, when $\psi = 0.1$ (the largest deviation in two slope functions), our bounds include the true value $\beta = 2$.

3.3 Discussion

Sharpness of Lower and Upper Bounds

Let $\mathcal{B}(\lambda_l, \lambda_u)$ denote the identified set for β under conditions in Theorem 1. Given the joint distribution of (Y, X, W, Z), it includes all β such that there exists a joint distribution of (Y, X, W, Z, X^*) consistent with Assumptions 1,2 and 4 , where

$$\beta = \frac{\mathrm{Cov}(Y^{\perp W}, X^*)}{\mathrm{Var}(X^{*\perp W})}.$$

For simplicity, we focus on the smallest and largest elements in this set. Unfortunately, it is possible that $\underline{B}(\lambda_l, \lambda_u)$ and $\overline{B}(\lambda_l, \lambda_u)$ might not represent these elements. In the Appendix C, we prove the sharpness of these bounds under additional restrictions. Nevertheless, bounds in Theorem 1 contain the sharp identified set of β as long as our assumptions hold.

Identification of λ_l and λ_u with Validation Data

Validation data provides an additional sample containing the measurement and the true value of the variable. It has been a useful approach to address the measurement error (e.g., Chen et al., 2005; Blattman et al., 2017). In our model, it helps to identify λ_l and λ_u and characterize bounds for β. We make mild assumptions about the validation data below.

Assumption 5. The validation data $\{X_{0i}, X^*_{0i}\}_{i=1}^{n_0}$ satisfies:

(i) Let $g_0(x^*) = \mathbb{E}(X_{0i} \mid X^*_{0i} = x^*)$ and $\lambda_0(x^*) = g_0(x^*)/x^*$. There exists a compact set $S \subseteq \mathbb{R}$ known to researchers such that $0 \notin S$ and

$$\lambda_l = \inf_{x^* \in S} \lambda_0(x^*),\ \ \lambda_u = \sup_{x^* \in S} \lambda_0(x^*)$$

(ii) There exists an estimator $\hat{g}_0(\cdot)$ such that $\sup_{x^* \in S} |\hat{g}_0(x^*) - g_0(x^*)| \xrightarrow{p} 0$ as $n_0 \to \infty$.

Assumption 5 (i) states that λ_l and λ_u are identified as the minimum and maximum values of the slope function from the validation data on a known set S. Condition (ii) assumes the uniform consistency for the estimator $\hat{g}(\cdot)$, which holds for local constant or local linear estimators under certain conditions (Li and Racine, 2007).

We estimate λ_l and λ_u by

$$\hat{\lambda}_l = \inf_{x^* \in S} \frac{\hat{g}_0(x^*)}{x^*},\ \ \hat{\lambda}_u = \sup_{x^* \in S} \frac{\hat{g}_0(x^*)}{x^*}$$

The next proposition shows that these estimators are consistent.

Proposition 3. *Suppose Assumption 5 hold, then* $\hat{\lambda}_l \xrightarrow{p} \lambda_l$ *and* $\hat{\lambda}_u \xrightarrow{p} \lambda_u$ *as* $n_0 \to \infty$.

Given the small validation sample size, we may adopt the method in Chernozhukov et al. (2013) to address the finite sample bias. We leave a detailed analysis for these bias-corrected estimators and inference to future work.

Breakdown Analysis

Without validation data, researchers may utilize prior knowledge to calibrate λ_l and λ_u based on the interpretation of the slope function, as discussed in Section 3.1. To check the robustness formally, we can perform the breakdown analysis described in Masten and Poirier (2020). Suppose the baseline model we find that $\beta^{\mathrm{IV}} \geq \beta_0$ for some $\beta_0 > 0$. Define *robust region*:

$$\mathrm{RR}(\beta_0) = \{\lambda_u \geq \lambda_l > 0 : \beta \geq \beta_0 \quad \forall \beta \in \mathcal{B}(\lambda_l, \lambda_u)\} \tag{13}$$

which contain all sensitivity parameters that allow us to draw our desired conclusion: $\beta \geq \beta_0$.

Corollary 1. *Let* $\Lambda = \{(\lambda_l, \lambda_u) \in \mathbb{R}_+^2 : 1 \leq \lambda_u/\lambda_l < 1 + \alpha^{-1}\}$. *Suppose the assumptions of Theorem 1 hold. For any* $\beta_0 \in (0, \beta^{\mathrm{IV}}]$, *define inner robust region:*

$$\begin{aligned} \mathrm{RR}^{\mathrm{I}}(\beta_0) &= \{(\lambda_l, \lambda_u) \in \Lambda : \underline{B}(\lambda_l, \lambda_u) \geq \beta_0\} \\ &= \{(\lambda_l, \lambda_u) \in \Lambda : \frac{1+\alpha}{\lambda_l} - \frac{\alpha}{\lambda_u} \leq \frac{\beta^{\mathrm{IV}}}{\beta_0}\} \end{aligned} \tag{14}$$

Then $\mathrm{RR}^{\mathrm{I}}(\beta_0) \subseteq \mathrm{RR}(\beta_0)$.

The inner robust region is the set of all (λ_l, λ_u) that provides a lower bound in Theorem 1 that is greater than or equal to β_0. It may be more conservative than the robust region but has an explicit expression. In particular, the inner robust region shrinks as α becomes larger. We only consider the case where the conclusion ($\beta \geq \beta_0$) holds under the baseline assumption, ensuring $\mathrm{RR}^{\mathrm{I}}(\beta_0) \neq \emptyset$.

Example 4 (Continued)**.** Recall that λ_l and λ_u are symmetric across 1 in this example, so there is only a one-dimensional sensitivity parameter ψ. Redefine *robust region* and *inner robust region* in terms of ψ:

$$\begin{aligned} \mathrm{RR}(\beta_0) &= \{\psi \geq 0 : \beta \geq \beta_0 \quad \forall \beta \in \mathcal{B}(1-\psi, 1+\psi)\} \\ \mathrm{RR}^{\mathrm{I}}(\beta_0) &= \{\psi \in \left[0, \frac{1}{1+2\alpha}\right) : \underline{B}(1-\psi, 1+\psi) \geq \beta_0\} \end{aligned} \tag{15}$$

We have $\mathrm{RR}^{\mathrm{I}}(\beta_0) \subseteq \mathrm{RR}(\beta_0)$ by Corollary 1. Moreover, the inner robust region becomes a line at the point

$$\psi^{\mathrm{bp}}(\beta_0) = \sup\left\{\psi \in \left[0, \frac{1}{1+2\alpha}\right) : \frac{1+(1+2\alpha)\psi}{1-\psi^2} \leq \frac{\beta^{\mathrm{IV}}}{\beta_0}\right\} \tag{16}$$

which is a breakdown point for the conclusion that $\beta \geq \beta_0$[10]. It is the largest deviation rate of the conditional mean of the measurement from the true variable we can allow for while still concluding $\beta \geq \beta_0$ based on Theorem 1. The analytical expression for the breakdown point is

$$\psi^{\mathrm{bp}}(\beta_0) = \min\left\{\frac{1}{1+2\alpha}, \frac{-(1+2\alpha)\beta_0 + \sqrt{(1+2\alpha)^2\beta_0^2 + 4(\beta^{\mathrm{IV}} - \beta_0)\beta^{\mathrm{IV}}}}{2\beta^{\mathrm{IV}}}\right\}$$

Example 5 (Continued)**.** In the case of overreporting, redefine *robust region* and *inner robust region* in terms of ψ:

$$\begin{aligned} \mathrm{RR}(\beta_0) &= \{\psi \geq 0 : \beta \geq \beta_0 \quad \forall \beta \in \mathcal{B}(1, 1+\psi)\} \\ \mathrm{RR}^{\mathrm{I}}(\beta_0) &= \{\psi \in [0, \alpha^{-1}) : \underline{B}(1, 1+\psi) \geq \beta_0\} \end{aligned}$$

[10] Masten and Poirier (2020) defined the breakdown point using the robust region. Our breakdown point is defined using the inner robust region so that it is more conservative.

The breakdown point is

$$\psi^{\mathrm{bp}}(\beta_0) = \sup\left\{\psi \in [0,\alpha^{-1}) : \frac{1+(1+\alpha)\psi}{1+\psi} \leq \frac{\beta^{\mathrm{IV}}}{\beta_0}\right\}$$
$$= \begin{cases} \alpha^{-1} & \text{if } (1+\alpha)\beta_0 \leq \beta^{\mathrm{IV}} \\ \min\left\{\alpha^{-1}, \frac{\beta^{\mathrm{IV}}-\beta_0}{(1+\alpha)\beta_0-\beta^{\mathrm{IV}}}\right\} & \text{otherwise} \end{cases}$$

which describes the largest average rate of overreporting that we can allow while still concluding $\beta \geq \beta_0$ based on Theorem 1.

4 Extensions

In this section, we provide bounds for two extensions on the model in Section 3. We also discuss an extension with omitted variables (i.e., the exclusion condition in Assumption 2 (i) does not hold) in Appendix D.

4.1 Systematic Error at $X^* = 0$

We relax condition (i) in Assumption 4 in this subsection by introducing another sensitivity parameter to bound the systematic error at $X^* = 0$.

Assumption 6. There exists $A_0 \geq 0$ such that $\forall (0, w, z) \in \Omega$,

$$\Big|\mathbb{E}[X \mid X^* = 0, W = w, Z = z]\Big| \leq A_0.$$

Theorem 2. *Suppose Assumptions 1, 2, 4 (ii) and 6 hold. Then,*

$$\frac{1+\alpha}{\lambda_u} - \frac{\alpha}{\lambda_l} - D(A_0, \lambda_u) \leq \frac{\beta^{\mathrm{IV}}}{\beta} \leq \frac{1+\alpha}{\lambda_l} - \frac{\alpha}{\lambda_u} + D(A_0, \lambda_u) \tag{17}$$

where

$$D(A_0, \lambda_u) = \frac{A_0 \cdot \mathbb{E}[|Z^{\perp W}| \cdot 1\{X^* = 0\}]}{\lambda_u \text{Cov}(Z^{\perp W}, X)}.$$

Compared with bounds (9) in Section 3, Theorem 2 includes an additional term $D(A_0, \lambda_u)$. In practice, since $D(A_0, \lambda_u)$ is not identified from data, we need to use prior knowledge to construct its upper bound. For example, there may exist a set $S_0 \in \mathbb{R}$ such that $X^* = 0$ only if $X \in S_0$, then we have

$$D(A_0, \lambda_u) \leq \overline{D}(A_0, \lambda_u) \equiv \frac{A_0 \cdot \mathbb{E}[|Z^{\perp W}| \cdot 1\{X \in S_0\}]}{\lambda_u \text{Cov}(Z^{\perp W}, X)}.$$

where $\overline{D}(A_0, \lambda_u)$ can be identified from data.

Example 4 (Continued)**.** Suppose we are interested in the robustness of the conclusion that $\beta \geq \beta_0$ for some known scalar A_0. By Theorem 2 (let $\lambda_l = 1 - \psi$ and $\lambda_u = 1 + \psi$),

$$\frac{1 - (1 + 2\alpha)\psi}{1 - \psi^2} - \overline{D}(A_0, 1 + \psi) \leq \frac{\beta^{\text{IV}}}{\beta} \leq \frac{1 + (1 + 2\alpha)\psi}{1 - \psi^2} + \overline{D}(A_0, 1 + \psi) \tag{18}$$

where $\overline{D}(A_0, 1 + \psi) \geq D(A_0, 1 + \psi)$ can be identified from data. Let

$$\Psi(A_0) = \left\{ \psi \in [0, 1) : \frac{1 - (1 + 2\alpha)\psi}{1 - \psi^2} - \overline{D}(A_0, 1 + \psi) > 0 \right\}$$

It is easy to see that β and β^{IV} takes the same sign if $\psi \in \Psi(A_0)$. Similarly, define the breakdown point

$$\psi^{\text{bp}}(\beta_0, A_0) = \sup \left\{ \psi \in \Psi(A_0) : \frac{1 + (1 + 2\alpha)\psi}{1 - \psi^2} + \overline{D}(A_0, 1 + \psi) \leq \frac{\beta^{\text{IV}}}{\beta_0} \right\} \tag{19}$$

which is the largest deviation rate of the conditional mean of the measurement from the true variable we can allow for while still concluding $\beta \geq \beta_0$ based on bounds (18) and A_0.

4.2 Multiple Measurements

When multiple measurements are available, a possible approach is to compare these IV estimators to explore the robustness of empirical results (e.g., Allcott et al., 2016; Bastos et al., 2018). For example, Dobbie et al. (2018) stated: "We explore the robustness of our results to alternative measures of pretrial release…and find very similar results." The intuition behind this approach is that the concerns over IV estimators due to measurement error may be limited if a coefficient remains stable after using other measurements. However, it is important to note that there is no formal econometric theory supporting such conclusions. In particular, if multiple measurements are subject to similar forms of measurement error, the resulting IV estimators may be consistently biased in the same direction—leading to stable, but still biased, coefficients.

As a corollary of Theorem 1, we can characterize the bounds of β as intersection bounds with multiple measurements. In the numerical example presented in Section 3.2, when the value of ψ is set to 0.1, the bounds for β are $[1.60,\ 2.20]$ using measurement X_1 and $[1.94,\ 2.85]$ using measurement X_2. While these bounds are relatively wide, we can tighten them to $[1.94,\ 2.20]$ if both measurements X_1 and X_2 are available.

Corollary 2. *Let $X_1, X_2, \cdots X_m$ be measurements ($m \geq 2$). Suppose assumptions in Theorem 1 hold for (Y, X_i, W, Z) with sensitivity parameters $(\lambda_{l,i}, \lambda_{u,i})$ ($1 \leq i \leq m$). Then, we have*

$$\max_{1\leq i\leq m} \underline{B}_i(\lambda_{l,i}, \lambda_{u,i}) \leq \beta \leq \min_{1\leq i\leq m} \overline{B}_i(\lambda_{l,i}, \lambda_{u,i}).$$

where bounds $[\underline{B}_i(\lambda_{l,i}, \lambda_{u,i}), \overline{B}_i(\lambda_{l,i}, \lambda_{u,i})]$ are obtained by applying Theorem 1 on (Y, X_i, W, Z). Moreover, the model is falsified if

$$\min_{1\leq i\leq m} \overline{B}(\lambda_{l,i}, \lambda_{u,i}) < \max_{1\leq i\leq m} \underline{B}(\lambda_{l,i}, \lambda_{u,i}).$$

Corollary 2 implies that the model is falsifiable with multiple measurements. To salvage the falsified model, we can allow for larger deviations or consider the falsification adaptive set described in Masten and Poirier (2021). A detailed analysis will be deferred to future work.

5 Empirical Application: the Returns to Education

We apply our method to the twins data from Ashenfelter and Krueger (1994), which estimates the economic returns to schooling. They compared the wage rates of identical twins and used the first difference to address the omitted ability bias. The idea is that identical twins, having the same genetics and family background, are expected to have similar abilities.

Data was collected at the Annual Twins Day Festival in Twinsburg, Ohio, in August 1991[11]. Ashenfelter and Krueger (1994) utilized a questionnaire based on the instrument of the Current Population Survey (CPS). They asked each twin about their wage rates, years of schooling, and common demographic variables. After dropping missing values, data in our analysis consists of 147 twins.

5.1 Baseline Model Results

Denote random variables Y_1 and Y_2 by the logarithms of the wage rates of the first and second twins in the pair[12]. Let S_m^* and W_m $(m = 1, 2)$ represent the true schooling level and other covariates of the mth twin.

Denote random variable μ by an unobservable component that varies by family. Ashenfelter and Krueger (1994) specified wage rates as

$$\begin{aligned} Y_1 &= \beta S_1^* + \gamma' W_1 + \mu + \epsilon_1 \\ Y_2 &= \beta S_2^* + \gamma' W_2 + \mu + \epsilon_2 \end{aligned} \tag{20}$$

[11] Data source: https://dataspace.princeton.edu/handle/88435/dsp012801pg35n, which contains a public use file with 183 twins and selected variables from the 1991 twins data Ashenfelter and Krueger (1994) used.

[12] Ashenfelter and Krueger (1994) randomly selected one twin as the first in each pair.

where ϵ_m $(m = 1, 2)$ are unobservable individual components. Our outcome equation is the difference between two equations in (20)

$$\underbrace{Y_1 - Y_2}_{\Delta Y} = \beta \cdot \underbrace{(S_1^* - S_2^*)}_{X^*} + \gamma' \underbrace{(W_1 - W_2)}_{W} + (\epsilon_1 - \epsilon_2) \tag{21}$$

In our study, the true variable X^* is the difference of true schooling levels in a pair, and the parameter of interest β measures the effect of schooling level on wages.

Let S_m^n $(m, n = 1, 2)$ refer to the education level of the mth twin as reported by nth twin. We have two measurements for X^*:

$$X = S_1^1 - S_2^2, \quad Z = S_1^2 - S_2^1 \tag{22}$$

which represent the difference in self-reports and sibling-reports, respectively. Define measurement errors

$$u = X - X^*, \quad \xi = Z - X^* \tag{23}$$

In columns (i) and (iii) of Table 2, we replicate AK's OLS and IV estimates of the outcome equation (use X as the independent variable and Z as the instrument). In column (ii), We also run OLS using Z as the independent variable. In column (iv), we use Z as the independent variable and X as the instrument. These point estimates are statistically significant at conventional levels. Moreover, IV estimators are much larger than OLS estimators.

5.2 Extension and Discussion for Baseline Estimates

Measurement error in self-reported schooling levels can be serious, and differencing data may increase the severity of measurement error (Griliches, 1979). Ashenfelter and Krueger (1994) adjusted for correlated measurement errors between X and Z and indicated that an additional year of schooling increases wages by 12-16 percent, a higher estimate than that in prior and later work. These results are based on assumptions that measurement errors (u, ξ)

Table 2: Estimates of the Outcome Equation (First Difference)

Panel A	OLS Own Report (i)	OLS Sibling's Report (ii)	IV Own Report (iii)	IV Sibling's Report (iv)
Years of schooling	0.091	0.098	0.179	0.158
	(0.022)	(0.021)	(0.041)	(0.038)
Sample size	147	147	147	147
Panel B. Sensitivity Analysis				
$\hat{\alpha} \times 100\%$			27.8%	28.4%
$\hat{\psi}^{bp} \times 100\%$ ($\beta_0 = 0.1$)			35.8%	28.7%

Notes: Panel A is the same as table 2 in Black et al. (2000), which uses Ashenfelter and Krueger (1994) data. The dependent variable is the logarithmic differences in the wages of identical twins. Each equation also contains differences in tenure, martial status and union coverage as controls. Numbers in parentheses are estimated standard errors. Panel B is new and calculates the estimates for α and breakdown point.

follow a joint normal distribution and are uncorrelated with the true variable X^*. Using a larger sample of twins, Rouse (1999) estimated that the return to schooling among identical twins is about 10 percent and concluded that the unusual results in Ashenfelter and Krueger (1994) were likely due to sampling error.

Bounds in Black et al.(2000)

Black et al. (2000) provided bounds for the coefficient in the presence of non-classical measurement error with an application of Ashenfelter and Krueger (1994) data. Based on certain assumptions, they demonstrated that the coefficient falls between the OLS estimand and the IV estimand. Specifically, their analysis revealed that the estimate in AK data ranged from 13% to 16%.

Their lower bound comes from the OLS regression in a subsample of twins who agree about the magnitude of the difference in education ($X = Z$). Its validity builds on the assumption that X is independent of Z conditional on X^*. However, it may not hold if their

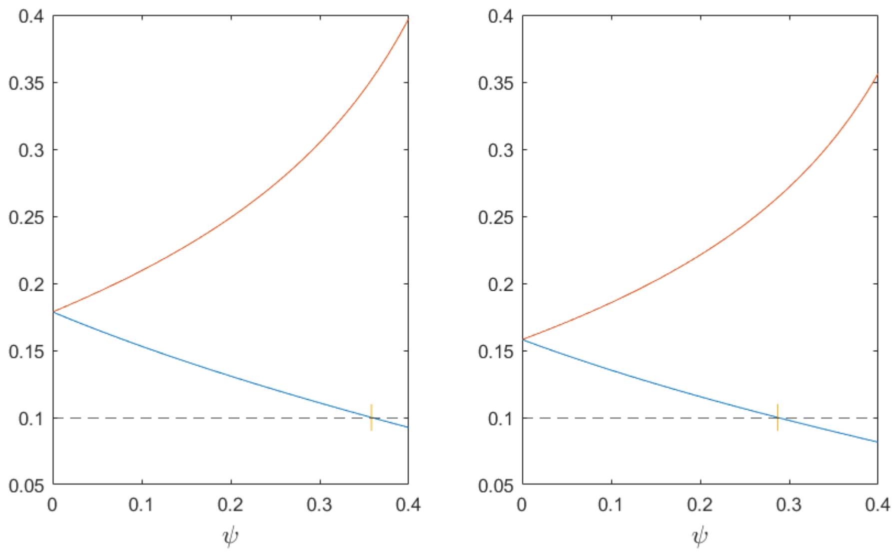

Figure 2: Sensitivity analysis under Theorem 1 in the empirical application (Left side: IV own report; Right side: IV sibling's report)

reporting behavior is highly correlated. Additionally, Black et al. (2000) require that u is uncorrelated with other covariates in multivariate regression, but this is not necessary in our framework.

5.3 Sensitivity Analysis Based on the Slope Function

The baseline results in Table 2 (iii) and (iv) rely on the assumption of "no systematic error". In particular, column (iii) requires that the measurement error in X is mean independent of (X^*, W, Z). Next, we use our framework to assess the sensitivity of high estimates ($\beta \geq 0.1$) in Ashenfelter and Krueger (1994) to systematic error. Since X^* can be zero, we first use Theorem 1 by assuming no systematic error at $X^* = 0$ for two measurements and then relax it.

We consider two specifications: IV own report and IV sibling's report, as shown in columns (iii) and (iv) in Table 2. In columns (iii) and (iv) of Panel B, the sample-analog estimators for α are 27.8% and 28.4%, respectively. This suggests that the two specifications are similarly sensitive to the presence of systematic errors.

Breakdown Analysis For simplicity, we assume sensitivity parameters λ_l and λ_u are symmetric across 1, as specified in Example 4. Figure 2 presents the results for two specifications, with the single parameter ψ indicating the deviation from the baseline assumption. Two solid lines show the estimated bounds for β as a function of ψ, as described by bounds (10). The dashed horizontal line represents the estimate in Rouse (1999): $\beta_0 = 0.1$, whose interactions with the solid line refer to the breakdown point ψ^{bp}: the minimum deviation needed to overturn the result $\beta \geq \beta_0$. As reported in Panel B of Table 2, the estimated breakdown points $\hat{\psi}^{bp}$ are 35.8% for IV own report and 28.7% for IV sibling's report. For example, for the specification of IV own report, we can conclude $\beta \geq 0.1$ as long as the measurement X deviates consistently at most 35.8% from the true variable.

Calibrated ψ We provide a data-drive method to calibrate ψ in Appendix E. The main idea is to split the sample into two and compare their IV estimators. Table 3 presents the estimates of the calibration parameter ψ using various methods for splitting the sample. Consider column (iii), calibrated ψ is smaller than the breakdown point estimate of 35.8% for all but the upper quantile of random spilt and the split using years of tenure, suggesting that the conclusion ($\beta \geq 0.1$) is quite robust to systematic error. In column (iv), calibrated ψ is much smaller and confirms the robustness.

Allow for Systematic Error at $X^* = 0$ To apply Theorem 2, we calibrate A_0 as

$$\bar{A} = \max\left\{\Big|\mathbb{E}_n[X \mid Z = 0]\Big|,\ \Big|\mathbb{E}_n[Z \mid X = 0]\Big|\right\} \tag{24}$$

where $\mathbb{E}_n$ refers to sample-based empirical expectation. To establish the bound on $D(\bar{A}, 1+\psi)$ in Theorem 2, we assume that if either of the two measurements exceeds M (which is either 3 or 4), then X^* is unlikely to be zero. Then, we have

$$\mathbb{E}\left[|Z^{\perp W}| \cdot 1\{X^* = 0\}\right] \leq \mathbb{E}\left[|Z^{\perp W}| \cdot 1\{X \leq M\} \cdot 1\{Z \leq M\}\right] \tag{25}$$

Table 3: Calibrating ψ: Compare IV Estimators in Subgroups

	IV Own Report (iii)	IV Sibling's Report (iv)
Random split		
25th	14.9%	9.8%
Median	29.0%	19.7%
75th	47.0%	32.1%
Split by Variables		
Age	25.4%	6.9%
Diff in Years of Tenure	47.0%	2.0%
White	27.3%	15.2%
Male	6.3%	8.2%

Notes: Random split is chosen by $K = 2$ with 500 times. We also split sample by variables. For continuous variables (age and diff in years of tenure), we split two subsamples based on their means. For binary variables (white and male), we split samples based on their values.

As shown in Panel A of Table 2, estimated breakdown points become smaller when A_0 increases but are still larger than most calibrated values.

Overall, our sensitivity analysis examines the robustness of the effect of education on wages under more general measurement errors compared to previous research. We find that the unusually high estimates in Ashenfelter and Krueger (1994) are robust to the presence of systematic error. The reason may be sampling error, as Rouse (1999) suggested.

Table 4: Sensitivity Analysis Based on ψ and A_0

A_0	IV Own Report (iii)		IV Sibling's Report (iv)	
	$M = 3$	$M = 4$	$M = 3$	$M = 4$
0	35.8%	35.8%	28.7%	28.7%
$\bar{A}$	34.7%	34.3%	27.3%	26.9%
$1.3\bar{A}$	34.3%	33.8%	26.9%	26.3%

Notes: This table reports estimates of the breakdown point ($\beta_0 = 0.1$) based on equation (19).

6 Conclusion

In this paper, we develop a framework to assess sensitivity to systematic error in linear IV models. Relative to existing literature, we allow for measurement error that is correlated with the true variable, other covariates, and the instrument in a flexible manner. We propose a relative measure of systematic error, called the slope function, which is intuitive to interpret. Using the bounds of the slope function, we derive simple bounds for the parameter of interest. One possibility of extension is to tighten our bounds by adding more assumptions on the slope function, such as monotonicity and functional form restrictions (e.g., conditional mean independence). Additionally, we hope that our approach can be applied to explore the sensitivity to the systematic error in other related models.

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.

Allcott, H., Collard-Wexler, A., and O'Connell, S. D. (2016). How do electricity shortages affect industry? evidence from india. *American Economic Review*, 106(3):587–624.

An, Y., Wang, L., and Xiao, R. (2022). A nonparametric nonclassical measurement error approach to estimating intergenerational mobility elasticities. *Journal of Business & Economic Statistics*, 40(1):169–185.

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.

Ashenfelter, O. and Krueger, A. (1994). Estimates of the economic return to schooling from a new sample of twins. *The American economic review*, pages 1157–1173.

Ashley, R. (2009). Assessing the credibility of instrumental variables inference with imperfect instruments via sensitivity analysis. *Journal of Applied Econometrics*, 24(2):325–337.

Bastos, P., Silva, J., and Verhoogen, E. (2018). Export destinations and input prices. *American Economic Review*, 108(2):353–392.

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

Black, D. A., Berger, M. C., and Scott, F. A. (2000). Bounding parameter estimates with non-classical measurement error. *Journal of the American Statistical Association*, 95(451):739–748.

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. (1996). Bounding mean regressions when a binary regressor is mismeasured. *Journal of Econometrics*, 73(2):387–399.

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. (2003). Measurement error in human capital and the black-white wage gap. *Review of Economics and Statistics*, 85(3):578–585.

Bontemps, C., Magnac, T., and Maurin, E. (2012). Set identified linear models. *Econometrica*, 80(3):1129–1155.

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. 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.

Chalak, K. and Kim, D. (2020). Measurement error in multiple equations: Tobin'sq and corporate investment, saving, and debt. *Journal of Econometrics*, 214(2):413–432.

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

Chernozhukov, V., Lee, S., and Rosen, A. M. (2013). Intersection bounds: Estimation and inference. *Econometrica*, 81(2):667–737.

Cinelli, C. and Hazlett, C. (2020). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(1):39–67.

Cinelli, C. and Hazlett, C. (2025). An omitted variable bias framework for sensitivity analysis of instrumental variables. *Biometrika*, page asaf004.

Diegert, P., Masten, M. A., and Poirier, A. (2023). Assessing omitted variable bias when the controls are endogenous. *arXiv preprint arXiv:2206.02303*.

DiTraglia, F. J. and García-Jimeno, C. (2021). A framework for eliciting, incorporating, and disciplining identification beliefs in linear models. *Journal of Business & Economic Statistics*, 39(4):1038–1053.

Dobbie, W., Goldin, J., and Yang, C. S. (2018). The effects of pre-trial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review*, 108(2):201–240.

Erickson, T. (1993). Restricting regression slopes in the errors-in-variables model by bounding the error correlation. *Econometrica: Journal of the Econometric Society*, pages 959–969.

Farber, H. S., Herbst, D., Kuziemko, I., and Naidu, S. (2021). Unions and inequality over the twentieth century: New evidence from survey data. *The Quarterly Journal of Economics*, 136(3):1325–1385.

Giglio, S., Maggiori, M., Stroebel, J., and Utkus, S. (2021). Five facts about beliefs and portfolios. *American Economic Review*, 111(5):1481–1522.

Gillen, B., Snowberg, E., and Yariv, L. (2019). Experimenting with measurement error: Techniques with applications to the caltech cohort study. *Journal of Political Economy*, 127(4):1826–1863.

Griliches, Z. (1979). Sibling models and data in economics: Beginnings of a survey. *Journal of political Economy*, 87(5, Part 2):S37–S64.

Haider, S. and Solon, G. (2006). Life-cycle variation in the association between current and lifetime earnings. *American economic review*, 96(4):1308–1320.

Hoffman, M. and Tadelis, S. (2021). People management skills, employee attrition, and manager rewards: An empirical analysis. *Journal of political economy*, 129(1):243–285.

Hu, Y., Schennach, S., and Shiu, J.-L. (2022). Identification of nonparametric monotonic regression models with continuous nonclassical measurement errors. *Journal of Econometrics*, 226(2):269–294.

Hu, Y. and Schennach, S. M. (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica*, 76(1):195–216.

Hyslop, D. R. and Imbens, G. W. (2001). Bias from classical and other forms of measurement error. *Journal of Business & Economic Statistics*, 19(4):475–481.

Jiang, Z. and Ding, P. (2020). Measurement errors in the binary instrumental variable model. *Biometrika*, 107(1):238–245.

Kirkpatrick, S. (2024). Principles of nutritional assessment: Measurement error in dietary assessment. Accessed on: Dec 26, 2024.

Klepper, S. and Leamer, E. E. (1984). Consistent sets of estimates for regressions with errors in all variables. *Econometrica: Journal of the Econometric Society*, pages 163–183.

Kreider, B. and Pepper, J. V. (2007). Disability and employment: Reevaluating the evidence in light of reporting errors. *Journal of the American Statistical Association*, 102(478):432–441.

Li, Q. and Racine, J. S. (2007). *Nonparametric econometrics: theory and practice*. Princeton University Press.

Macchiavello, R. and Morjaria, A. (2021). Competition and relational contracts in the rwanda coffee chain. *The Quarterly Journal of Economics*, 136(2):1089–1143.

Masten, M. A. and Poirier, A. (2020). Inference on breakdown frontiers. *Quantitative Economics*, 11(1):41–111.

Masten, M. A. and Poirier, A. (2021). Salvaging falsified instrumental variable models. *Econometrica*, 89(3):1449–1469.

Masten, M. A. and Poirier, A. (2024). The effect of omitted variables on the sign of regression coefficients. *arXiv preprint arXiv:2208.00552*.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2):187–204.

Rogall, T. (2021). Mobilizing the masses for genocide. *American economic review*, 111(1):41–72.

Rouse, C. E. (1999). Further estimates of the economic return to schooling from a new sample of twins. *Economics of Education Review*, 18(2):149–157.

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

Small, D. S. (2007). Sensitivity analysis for instrumental variables regression with overidentifying restrictions. *Journal of the American Statistical Association*, 102(479):1049–1058.

Taylor, J. R. (1997). *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books, Sausalito, CA.

Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach*. Cengage Learning, 5th edition.

The Online Supplement to "Assessing Measurement Error in Linear Instrumental Variables Models"

Appendix A contains the proofs of the results presented in the main body. Appendix B examines a simple case of non-classical measurement error. Appendix C outlines the conditions related to the sharpness of our bounds. Appendix D presents an extension that addresses omitted variables. Appendix E offers a data-driven method to calibrate λ_u/λ_l. Finally, Appendix F provides proofs for the results discussed in the appendices.

A Proofs of Main Results

A.1 Proofs for Section 2

Proof of Proposition 1. Assumption 3 is equivalent to $\forall(x^*, w, z) \in \Omega$

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

We can easily see its equivalence with two conditions in Proposition 1. *Q.E.D.*

Proof of Proposition 2. Since both $Y - Y^{\perp W}$ and $X - X^{\perp W}$ are linear combinations of W, we have

$$\mathrm{Cov}(Z^{\perp W}, Y - Y^{\perp W}) = 0, \quad \mathrm{Cov}(Z^{\perp W}, X - X^{\perp W}) = 0.$$

Then,

$$\begin{aligned}
\beta^{\mathrm{IV}} &= \frac{\mathrm{Cov}(Z^{\perp W}, Y^{\perp W})}{\mathrm{Cov}(Z^{\perp W}, X^{\perp W})} \\
&= \frac{\mathrm{Cov}(Z^{\perp W}, Y)}{\mathrm{Cov}(Z^{\perp W}, X)} \\
&= \frac{\mathrm{Cov}(Z^{\perp W}, \beta X^* + \gamma' W + \epsilon)}{\mathrm{Cov}(Z^{\perp W}, X)}
\end{aligned}$$

By the construction of ϵ, we have $\mathrm{Cov}(W, \epsilon) = 0$. We further obtain $\mathrm{Cov}(Z^{\perp W}, \epsilon) = 0$ by

Assumption 2 (i). Combined it with $\mathrm{Cov}(Z^{\perp W}, W) = 0$, we have

$$\beta^{\mathrm{IV}} = \beta \cdot \frac{\mathrm{Cov}(Z^{\perp W}, X^*)}{\mathrm{Cov}(Z^{\perp W}, X)} \tag{26}$$

Since Assumption 3 implies $\mathrm{Cov}(Z^{\perp W}, u) = 0$, we have

$$\beta^{\mathrm{IV}} = \beta - \beta \cdot \frac{\mathrm{Cov}(Z^{\perp W}, u)}{\mathrm{Cov}(Z^{\perp W}, X)} = \beta$$

Finally, since β^{IV} is identified from the joint distribution of (Y, X, W, Z), β is point identified. *Q.E.D.*

A.2 Proofs for Section 3

Proof of Theorem 1. From equation (26), we have $\beta \neq 0$ and

$$\frac{\beta^{\mathrm{IV}}}{\beta} = \frac{\mathrm{Cov}(Z^{\perp W}, X^*)}{\mathrm{Cov}(Z^{\perp W}, X)} \tag{27}$$

Then we only need to focus on the bounds of $\mathrm{Cov}(Z^{\perp W}, X^*)$. Since W includes a constant, we have $\mathbb{E}[Z^{\perp W}] = 0$ and $\mathrm{Cov}(Z^{\perp W}, X^*) = \mathbb{E}[Z^{\perp W} X^*]$.

By Assumption 4 (ii), the following equation holds almost surely:

$$X = \lambda(X^*, W, Z) \cdot X^* + \tilde{u} \tag{28}$$

where $\tilde{u}$ is the random error. Then

$$\mathrm{Cov}(Z^{\perp W}, X^*) = \mathbb{E}\left[Z^{\perp W} \cdot \frac{X - \tilde{u}}{\lambda(X^*, W, Z)}\right] = \mathbb{E}\left[Z^{\perp W} \cdot \frac{X}{\lambda(X^*, W, Z)}\right] \tag{29}$$

where we use $\mathbb{E}[\tilde{u} \mid X^*, W, Z] = 0$.

Let

$$p^+ = \Pr(Z^{\perp W} X \geq 0), \quad p^- = \Pr(Z^{\perp W} X < 0).$$

By equation (29), we have

$$\begin{aligned}\mathrm{Cov}(Z^{\perp W}, X^*) \geq & \mathbb{E}\left[Z^{\perp W} \cdot \frac{X}{\lambda_l + \mathbf{1}\{Z^{\perp W}X \geq 0\}(\lambda_u - \lambda_l)}\right] \\ = & \frac{1}{\lambda_u} \cdot \mathbb{E}\left[Z^{\perp W}X \mid Z^{\perp W}X \geq 0\right] \cdot p^+ + \frac{1}{\lambda_l} \cdot \mathbb{E}\left[Z^{\perp W}X \mid Z^{\perp W}X < 0\right] \cdot p^- \\ = & \frac{1}{\lambda_u} \cdot \mathbb{E}[Z^{\perp W}X] - \frac{1}{2} \cdot (\frac{1}{\lambda_l} - \frac{1}{\lambda_u}) \cdot (\mathbb{E}(|Z^{\perp W}X|) - \mathbb{E}[Z^{\perp W}X])\end{aligned} \tag{30}$$

where we use

$$\begin{aligned}\mathbb{E}\left[Z^{\perp W}X \mid Z^{\perp W}X \geq 0\right] \cdot p^+ &= \frac{1}{2}\Big(\mathbb{E}(|Z^{\perp W}X|) + \mathbb{E}[Z^{\perp W}X]\Big) \\ \mathbb{E}\left[Z^{\perp W}X \mid Z^{\perp W}X < 0\right] \cdot p^- &= -\frac{1}{2}\Big(\mathbb{E}(|Z^{\perp W}X|) - \mathbb{E}[Z^{\perp W}X]\Big)\end{aligned}$$

Similarly, we have

$$\mathrm{Cov}(Z^{\perp W}, X^*) \leq \frac{1}{\lambda_l} \cdot \mathbb{E}[Z^{\perp W}X] + \frac{1}{2} \cdot (\frac{1}{\lambda_l} - \frac{1}{\lambda_u}) \cdot (\mathbb{E}(|Z^{\perp W}X|) - \mathbb{E}[Z^{\perp W}X]) \tag{31}$$

By equation (27) and bounds (30), (31), we obtain

$$\frac{1}{\lambda_u} - (\frac{1}{\lambda_l} - \frac{1}{\lambda_u}) \cdot \alpha \leq \frac{\beta^{\mathrm{IV}}}{\beta} \leq \frac{1}{\lambda_l} + (\frac{1}{\lambda_l} - \frac{1}{\lambda_u}) \cdot \alpha \tag{32}$$

Then $\beta > 0$ since $\beta^{\mathrm{IV}} > 0$ and $\lambda_u/\lambda_l < 1 + \alpha^{-1}$. Our bounds of β follow from the bounds above. *Q.E.D.*

Proof of Proposition 3. Let $s = \min_{x^* \in S} |x^*|$, then we have $s > 0$ since S is compact and $0 \notin S$.

By Assumption 5 (ii), we have

$$\begin{aligned}\hat{\lambda}_l &= \inf_{x^* \in S} \frac{\hat{g}_0(x^*)}{x^*} \\ &\geq \inf_{x^* \in S} \frac{g_0(x^*)}{x^*} - \frac{\sup_{x^* \in S} |\hat{g}_0(x^*) - g_0(x^*)|}{s} \\ &= \lambda_l - o_p(1)\end{aligned} \tag{33}$$

On the other hand, since S is compact, there exists $x_l^* \in S$ such that $\lambda_0(x_l^*) = \lambda_l$, then

$$\begin{aligned}\hat{\lambda}_l &\leq \frac{\hat{g}_0(x_l^*)}{x_l^*} \\ &\leq \lambda_0(x_l^*) + \frac{\sup_{x^* \in S} |\hat{g}_0(x^*) - g_0(x^*)|}{s} \\ &= \lambda_l + o_p(1)\end{aligned} \tag{34}$$

Thus we conclude $\hat{\lambda}_l \xrightarrow{p} \lambda_l$ as $n_0 \to \infty$.The consistency of $\hat{\lambda}_u$ can be similarly proved and therefore omitted. *Q.E.D.*

Proof of Corollary 1. Note that Theorem 1 implies $\mathcal{B}(\lambda_l, \lambda_u) \in [\underline{B}(\lambda_l, \lambda_u), \overline{B}(\lambda_l, \lambda_u)]$ if $(\lambda_l, \lambda_u) \in \Lambda$. Then we have $\mathrm{RR}^{\mathrm{I}}(\beta_0) \subseteq \mathrm{RR}(\beta_0)$ by the definition of the robust region in (13). *Q.E.D.*

A.3 Proofs for Section 4

Proof of Theorem 2. Define function:

$$g(w, z) = \mathbb{E}[X \mid X^* = 0, W = w, Z = z]$$

Then we have $|g(W, Z)| \leq A_0$ by Assumption 6.

For $(0, w, z) \in \Omega$, let $\lambda(0, w, z) = \lambda_u$. Write down

$$X = g(W, Z) \cdot 1\{X^* = 0\} + \lambda(X^*, W, Z) \cdot X^* + \tilde{u} \tag{35}$$

where $\tilde{u}$ is the random error. Then,

$$\begin{aligned}\mathbb{E}[Z^{\perp W} X^*] &= \mathbb{E}\left[Z^{\perp W} \cdot \frac{X - g(W,Z)\cdot 1\{X^*=0\} - \tilde{u}}{\lambda(X^*,W,Z)}\right] \\ &= \mathbb{E}\left[Z^{\perp W} \cdot \frac{X}{\lambda(X^*,W,Z)}\right] - \mathbb{E}\left[Z^{\perp W} \cdot \frac{g(W,Z)\cdot 1\{X^*=0\}}{\lambda(X^*,W,Z)}\right]\end{aligned} \tag{36}$$

where we use $\mathbb{E}[\tilde{u} \mid X^*, W, Z] = 0$. The absolute value of the second part is bounded by

$$\mathbb{E}\left[|Z^{\perp W}| \cdot \frac{A_0 1\{X^*=0\}}{\lambda(0,W,Z)}\right] = \mathbb{E}\left[|Z^{\perp W}| \cdot \frac{A_0 1\{X^*=0\}}{\lambda_u}\right] \tag{37}$$

In the proof of Theorem 1, we have shown

$$\frac{1+\alpha}{\lambda_u} - \frac{\alpha}{\lambda_l} \leq \frac{\mathbb{E}\left[Z^{\perp W} \cdot \frac{X}{\lambda(X^*,W,Z)}\right]}{\mathrm{Cov}(Z^{\perp W}, X)} \leq \frac{1+\alpha}{\lambda_l} - \frac{\alpha}{\lambda_u} \tag{38}$$

Finally, by (27) and (36)

$$\begin{aligned}\frac{\beta^{\mathrm{IV}}}{\beta} &= \frac{\mathbb{E}\left[Z^{\perp W} \cdot \frac{X}{\lambda(X^*,W,Z)}\right]}{\mathrm{Cov}(Z^{\perp W}, X)} - \frac{\mathbb{E}\left[Z^{\perp W} \cdot \frac{g(W,Z)\cdot 1\{X^*=0\}}{\lambda(X^*,W,Z)}\right]}{\mathrm{Cov}(Z^{\perp W}, X)} \\ &\in \left[\frac{1+\alpha}{\lambda_u} - \frac{\alpha}{\lambda_l} - D(A_0, \lambda_u), \quad \frac{1+\alpha}{\lambda_l} - \frac{\alpha}{\lambda_u} + D(A_0, \lambda_u)\right]\end{aligned} \tag{39}$$

Q.E.D.

Proof of Corollary 2. It follows directly from Theorem 1. *Q.E.D.*

B A Simple Case of Non-classical Measurement Error

In this section, we consider a simple case of non-classical measurement error under the following assumption.

Assumption B.1. $\mathbb{E}[X \mid X^* = x^*, W, Z] = \lambda x^*$ for some positive $\lambda \neq 1$.

Assumption B.1 states that the conditional mean $\mathbb{E}[X \mid X^* = x^*]$ is proportional to x^*, but the slope is not 1. It violates condition (i) but still satisfies condition (ii) in Proposition 1. In self-reported data, we can understand it as subjects tending to overreport ($\lambda > 1$) or underreport ($\lambda < 1$) the true value, while the average degree remains the same across all levels of the true value.

Under Assumption B.1, a more familiar expression is

$$X = \lambda X^* + \tilde{u} \tag{40}$$

where $\tilde{u}$ is the random error. Then, we have the decomposition of the measurement error

$$u = X - X^* = \tilde{u} + \underbrace{(\lambda - 1)X^*}_{\text{systematic error}} \tag{41}$$

which implies that $\lambda > 1$ ($\lambda < 1$) leads to the positive (negative) correlation between X^* and u. Let β^{OLS} be the coefficient on X in the OLS estimand of Y on (X, W). The following Proposition gives the expressions of OLS and IV estimands.

Proposition B.1. *Suppose Assumptions 1, 2 and B.1 hold, we have* $\beta^{\text{IV}} = \frac{\beta}{\lambda}$*. Moreover, if* $\text{Cov}(\epsilon, \tilde{u}) = 0$*, then*

$$\beta^{\text{OLS}} = \frac{\beta}{\lambda} \cdot \underbrace{\left[\frac{\text{Var}(X^{*\perp W})}{\text{Var}(X^{*\perp W}) + \lambda^{-2}\,\text{Var}(\tilde{u})} \right]}_{\leq 1}.$$

There are two kinds of bias in β^{OLS}: attenuation bias from the random error $\tilde{u}$ and bias from systematic error ($\lambda \neq 1$). While β^{IV} can avoid the attenuation bias, it is unable to address the bias from systematic error. An implication of the IV estimand is that overreporting ($\lambda > 1$) underestimates β, while underreporting ($\lambda < 1$) overestimates β.

Proposition B.1 implies that the bias in the OLS estimand tends to be more complicated than that in the IV estimand. Overreporting ($\lambda > 1$) leads to a downward bias in OLS estimand, but the direction of bias under underreporting ($\lambda < 1$) is ambiguous, depending on the ratio of λ and attenuation bias.

We can conduct a sensitivity analysis if Assumption B.1 holds for some λ within the interval $[\lambda_l, \lambda_u]$, where $\lambda_u \geq \lambda_l > 0$ are known parameters. Proposition B.1 provides bounds on β as follows:

$$\frac{1}{\lambda_u} \leq \frac{\beta^{\text{IV}}}{\beta} \leq \frac{1}{\lambda_l}$$

These bounds are narrower than those in bounds (9), which account for general measurement error.

C Sharpness of Lower and Upper Bounds

We first state Assumption C.1, which ensures the sharpness of bounds in Theorem 1. Given $\lambda_1, \lambda_2 > 0$, define random variable

$$X(\lambda_1, \lambda_2) = \frac{X}{\lambda_1} \cdot \mathbf{1}\{(Z - \pi' W)X \geq 0\} + \frac{X}{\lambda_2} \cdot \mathbf{1}\{(Z - \pi' W)X < 0\} \tag{42}$$

Define the sign function of a real number x as:

$$\text{sgn}(x) := \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}$$

Assumption C.1. The following hold.

(i) The variance matrix of (Y, X, W, Z) is finite and positive definite.

(ii) $\text{Cov}(Z^{\perp W}, X) > 0$ and $\text{Cov}(Z^{\perp W}, Y) > 0$.

(iii) $\lambda_u / \lambda_l < 1 + \alpha^{-1}$.

(iv) The following inequalities hold for $\widetilde{X} = X(\lambda_l, \lambda_u)$ and $X(\lambda_u, \lambda_l)$

$$\frac{\mathrm{Cov}(Y^{\perp W}, \widetilde{X})}{\mathrm{Var}(\widetilde{X}^{\perp W})} \leq \frac{\mathrm{Cov}(Z^{\perp W}, Y)}{\mathrm{Cov}(Z^{\perp W}, \widetilde{X})} \leq \frac{\mathrm{Cov}(Y^{\perp W}, \mathbb{E}[\widetilde{X}|Z, W, \mathrm{sgn}(X)])}{\mathrm{Var}(\mathbb{E}[\widetilde{X}^{\perp W}|Z, W, \mathrm{sgn}(X)])}$$

where these denominators are bounded away from zero.

Theorem C.1. *Suppose the joint distribution of* (Y, X, W, Z) *satisfies Assumption C.1. Then, bounds in Theorem 1 are sharp; that is*

$$\underline{B}(\lambda_l, \lambda_u) = \inf \mathcal{B}(\lambda_l, \lambda_u),\ \overline{B}(\lambda_l, \lambda_u) = \sup \mathcal{B}(\lambda_l, \lambda_u).$$

The key assumption is the condition (iv) in Assumption C.1. We illustrate it under $\lambda_l = \lambda_u = 1$:

$$\underbrace{\frac{\mathrm{Cov}(Y^{\perp W}, X)}{\mathrm{Var}(X^{\perp W})} \leq \frac{\mathrm{Cov}(Z^{\perp W}, Y)}{\mathrm{Cov}(Z^{\perp W}, X)}}_{\beta^{\mathrm{OLS}} \leq \beta^{\mathrm{IV}}} \leq \frac{\mathrm{Cov}(Y^{\perp W}, \mathbb{E}[X|V])}{\mathrm{Var}(\mathbb{E}[X^{\perp W}|V])}$$

for $V = (Z, W, \mathrm{sgn}(X))$. The first inequality states $\beta^{\mathrm{OLS}} \leq \beta^{\mathrm{IV}} = \beta$, which is known as attenuation bias. The second inequality suggests that $\mathrm{Var}(\mathbb{E}[X \mid V])$ can not be too large, indicating there is "enough noise" in X after being explained by (Z, W). These inequalities are testable and provide sufficient conditions for the sharpness.

D An Extension with Omitted Variables

In empirical studies, both measurement error and omitted variables can be sources of endogeneity. Therefore, researchers may intend to use IV to address them together. A natural question arises regarding adjusting our bounds in the presence of omitted variables.

Consider the linear regression with omitted variables D:

$$Y = \beta X^* + \gamma' W + \underbrace{\rho' D + Y^{\perp(X^*, W, D)}}_{\text{error term } \epsilon}$$

where the coefficient β is our parameter of interest. Endogeneity arises if X^* or W is correlated with any of the omitted variables. Let

$$\epsilon = \chi_1 Z + \chi_2' W + \epsilon^{\perp(Z,W)}$$

We allow ϵ to be correlated with Z, but put the restriction below.

Assumption D.1. $|\chi_1| \leq \kappa$ for known parameter $\kappa \geq 0$.

Proposition D.1. *Suppose Assumption 1, 2 (ii), 4 and D.1 hold. Suppose* $\lambda_u/\lambda_l < 1 + \alpha^{-1}$ *and* $\mathrm{Var}(Z^{\perp W}) \cdot \kappa \leq |\mathrm{Cov}(Z^{\perp W}, Y)|$ *hold. Normalize* $\beta^{\mathrm{IV}} > 0$*. Then,* $\beta > 0$ *and*

$$\underline{B}(\lambda_l, \lambda_u)\left(1 - \frac{\mathrm{Var}(Z^{\perp W})}{\mathrm{Cov}(Z^{\perp W}, Y)} \cdot \kappa\right) \leq \beta \leq \overline{B}(\lambda_l, \lambda_u)\left(1 + \frac{\mathrm{Var}(Z^{\perp W})}{\mathrm{Cov}(Z^{\perp W}, Y)} \cdot \kappa\right)$$

Proposition D.1 shows that the adjustment term is proportional to κ. If the exclusion condition holds ($\kappa = 0$), the adjustment term becomes zero.

E Calibrating the magnitude of λ_u/λ_l

We provide a data-dependent approach to calibrate the magnitude of the ratio $r = \lambda_u/\lambda_l$. To see the intuition, suppose the slope function only takes two values: λ_1 and λ_2 ($\lambda_1 < \lambda_2$), and data is divided into two groups. In group k ($k = 1, 2$), the measurement error satisfies Assumption B.1 with $\lambda = \lambda_k$. Let β_k^{IV} denote the IV estimand in group k, by Proposition B.1, we have

$$\beta_1^{IV} = \frac{\beta}{\lambda_1},\ \beta_2^{IV} = \frac{\beta}{\lambda_2} \Longrightarrow r = \frac{\lambda_2}{\lambda_1} = \frac{\beta_1^{IV}}{\beta_2^{IV}}$$

which implies that r is identified if we can identify these two groups. This example suggests comparing IV estimators in different subgroups to calibrate r, which motivates the following method:

Step 1. Randomly divide data into K groups, where $K \geq 2$ is a predetermined integer.

Step 2. Obtain IV estimators within groups: $\hat{\beta}_1^{IV}, \hat{\beta}_2^{IV}, \cdots, \hat{\beta}_K^{IV}$.

Step 3. Compute

$$\hat{r} = \max_{1\leq k\leq K} |\hat{\beta}_k^{IV}| / \min_{1\leq k\leq K} |\hat{\beta}_k^{IV}|$$

Step 4. Repeat *Step 1-3* for N times, yielding $\hat{r}_1, \hat{r}_2, \cdots, \hat{r}_N$.

Step 5. Calibrate r as the median of $\{\hat{r}_j\}_{j=1}^N$.

This approach can help us calibrate λ_l and λ_u in special cases. Taking example 4 ($\lambda_l = 1-\psi, \lambda_u = 1+\psi$), after calibrating r, we can solve for ψ from

$$\frac{1+\psi}{1-\psi} = r$$

We can calibrate the parameter ψ in examples 5 and 6 similarly.

F Proofs for Appendix

F.1 Proofs for Appendix B

Proof of Proposition B.1. Recall that random error $\tilde{u} = X - \lambda X^*$, we have $\text{Cov}(Z^{\perp W}, \tilde{u}) = 0$. By equation (26),

$$\beta^{\text{IV}} = \beta \cdot \frac{\text{Cov}(Z^{\perp W}, X^*)}{\text{Cov}(Z^{\perp W}, \lambda X^* + \tilde{u})} = \frac{\beta}{\lambda}$$

For OLS estimand, we have $\text{Cov}(X, \epsilon) = \text{Cov}(\lambda X^* + \tilde{u}, \epsilon) = 0$. Since $X^{\perp W}$ is a linear combination of X and W, we obtain $\text{Cov}(X^{\perp W}, \epsilon) = 0$. Then,

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

Note that $\tilde{u}^{\perp W} = \tilde{u}$ since $\tilde{u}$ is mean independent of W. Then,

$$\begin{aligned}\mathrm{Cov}(X^{\perp W}, \beta X^*) &= \mathrm{Cov}(X^{\perp W}, \beta X/\lambda) - \mathrm{Cov}(X^{\perp W}, \beta \tilde{u}/\lambda) \\ &= \frac{\beta}{\lambda}\mathrm{Var}(X^{\perp W}) - \mathrm{Cov}(X^{*\perp W} + \tilde{u}, \beta \tilde{u}/\lambda) \\ &= \frac{\beta}{\lambda}\left[\mathrm{Var}(X^{\perp W}) - \mathrm{Var}(\tilde{u})\right]\end{aligned}$$

where we use $\mathrm{Cov}(X^{*\perp W}, \tilde{u}) = 0$ in the last step. We can conclude

$$\beta^{\mathrm{OLS}} = \frac{\beta}{\lambda}\left[1 - \frac{\mathrm{Var}(\tilde{u})}{\mathrm{Var}(X^{\perp W})}\right] = \frac{\beta}{\lambda} \cdot \left[\frac{\mathrm{Var}(X^{*\perp W})}{\mathrm{Var}(X^{*\perp W}) + \lambda^{-2}\,\mathrm{Var}(\tilde{u})}\right]$$

where we use $\mathrm{Var}(X^{\perp W}) = \lambda^2\,\mathrm{Var}(X^{*\perp W}) + \mathrm{Var}(\tilde{u})$. *Q.E.D.*

F.2 Proofs of Appendix C

Before proving Theorem C.1, we present three useful lemmas. For two random vectors A and B with the same distribution, we denote $A \overset{d}{=} B$.

Lemma 1. *Let U be a random variable and V be a random vector. Suppose the variance matrix (U, V) is finite and positive definite. Then there exists random variables $\{\check{U}_a\}_{0\le a\le 1}$ such that*

(i) $\check{U}_1 = U$ *and* $(\check{U}_a, V) \overset{d}{=} (U, V)$, *for all* $0 \le a \le 1$.

(ii) $\mathbb{E}[U \mid V, \check{U}_a = u] = au + (1-a)\mathbb{E}[U \mid V]$.

(iii) $\mathbb{E}[(\check{U}_b - \check{U}_a)^2] \le 4(b-a)\cdot\mathbb{E}[U^2]$, $\forall 0 \le a \le b \le 1$.

Proof of Lemma 1. Firstly, we can find random variable $\check{U}$ such that

$$(\check{U}, V) \overset{d}{=} (U, V) \text{ and } U \perp \check{U} \mid V$$

Then we can find random variable $\xi \sim U[0,1]$ and $\xi \perp (U, \check{U}, V)$. $\forall a \in [0,1]$, let

$$\check{U}_a = \begin{cases} U & \text{if } \xi \leq a \\ \check{U}, & \text{if } \xi > a \end{cases}$$

It is obvious that $\check{U}_1 = U$. Denote $F_{A|B}(\cdot)$ as c.d.f of random variable A conditional on B. We have

$$F_{\check{U}_a|V}(u) = aF_{U|V}(u) + (1-a)F_{\check{U}|V}(u) = F_{U|V}(u)$$

since $(\check{U}, V) \overset{d}{=} (U, V)$. Thus we verify that $(\check{U}_a, V) \overset{d}{=} (U, V)$ for all $0 \leq a \leq 1$.

As for (ii), by law of iterated expectation,

$$\begin{aligned} & \mathbb{E}[U \mid V, \check{U}_a = u] \\ =& \Pr(\xi \leq a) \cdot \mathbb{E}[U \mid V, \check{U}_a = u, \xi \leq a] + \Pr(\xi > a) \cdot \mathbb{E}[U \mid V, \check{U}_a = u, \xi > a] \\ =& a \cdot \mathbb{E}[U \mid V, U = u, \xi \leq a] + (1-a) \cdot \mathbb{E}[U \mid V, \check{U} = u, \xi > a] \\ =& au + (1-a)\mathbb{E}[U \mid V] \end{aligned}$$

where we use $U \perp \check{U} \mid V$ and $\xi \perp (U, \check{U}, V)$ in the last step.

It remains to verify result (iii). Note that $\check{U}_a = \check{U}_b$ if $\xi \notin [a, b]$, then

$$\begin{aligned} \mathbb{E}[(\check{U}_b - \check{U}_a)^2] &= \Pr(\xi \in [a,b]) \cdot \mathbb{E}\left[(\check{U}_b - \check{U}_a)^2 | \xi \in [a,b]\right] \\ &= (b-a) \cdot \mathbb{E}\left[(U - \check{U})^2 | \xi \in [a,b]\right] \\ &= (b-a) \cdot \mathbb{E}[(U - \check{U})^2] \\ &\leq (b-a) \cdot \mathbb{E}[2(U^2 + \check{U}^2)] \\ &= 4(b-a) \cdot \mathbb{E}[U^2] \end{aligned}$$

where we use $\xi \perp (U, \check{U})$ and $U \overset{d}{=} \check{U}$. *Q.E.D.*

Lemma 2. *Theorem C.1 holds when* $\lambda_l = \lambda_u = 1$*. Moreover, the constructed true variable* X^* *satisfies that* $\text{sgn}(X^*) = \text{sgn}(X)$*.*

Proof of Lemma 2. When $\lambda_u = \lambda_l = 1$, we have $X(1,1) = X$. We can write condition (iv) in Assumption C.1

$$\frac{\text{Cov}(Y^{\perp W}, X)}{\text{Var}(X^{\perp W})} \leq \frac{\text{Cov}(Z^{\perp W}, Y)}{\text{Cov}(Z^{\perp W}, X)} \leq \frac{\text{Cov}(Y^{\perp W}, \mathbb{E}[X|V])}{\text{Var}(\mathbb{E}[X^{\perp W}|V])} \tag{43}$$

for $V = (Z, W, \text{sgn}(X))$. We need to find random variable X^* such that (Y, X, W, Z, X^*) satisfies Assumption 1, 2 and 3.

Step 1 (Construction of X_a^*) : By Lemma 1 (let $U = X$), there exists random variables $\{\check{X}_a\}_{0 \leq a \leq 1}$ such that

(i) $\check{X}_1 = X$ and $(\check{X}_a, V) \overset{d}{=} (X, V)$, for all $0 \leq a \leq 1$.

(ii) $\mathbb{E}[X \mid V, \check{X}_a = u] = au + (1-a)\mathbb{E}[X \mid V]$.

(iii) $\mathbb{E}[(\check{X}_b - \check{X}_a)^2] \leq 4(b-a) \cdot \mathbb{E}[X^2]$, $\forall 0 \leq a \leq b \leq 1$.

Let $X_a^* = \mathbb{E}[X \mid V, \check{X}_a]$ and $\tilde{u}_a = X - X_a^*$, then we have $\mathbb{E}[\tilde{u}_a \mid V, \check{X}_a] = 0$. Since X_a^* is a function of $(V, \check{X}_a)$, we verify the baseline assumption for all $0 \leq a \leq 1$:

$$\mathbb{E}[\tilde{u}_a \mid X_a^*, W, Z] = 0$$

It is also easy to show $\text{sgn}(X_a^*) = \text{sgn}(X)$.

The next two steps show there exists $0 \leq a \leq 1$ such that the exclusion condition of Z holds. That is,

$$\frac{\text{Cov}(Y^{\perp W}, X_a^*)}{\text{Var}(X_a^{*\perp W})} = \frac{\text{Cov}(Z^{\perp W}, Y)}{\text{Cov}(Z^{\perp W}, X)}$$

Step 2 (Continuity of OLS coefficient): Let

$$H(a) = \frac{\text{Cov}(Y^{\perp W}, X_a^*)}{\text{Var}(X_a^{*\perp W})} = \frac{H_1(a)}{H_2(a)}$$

We need to show both the $H_1(\cdot)$ and $H_2(\cdot)$ are continuous in $a \in [0,1]$.

By result (ii) in Lemma 1, we have

$$H_1(a) = a\text{Cov}(Y, \check{X}_a) + (1-a)\text{Cov}(Y, \mathbb{E}[X \mid V])$$

By Cauchy-Schwarz inequality and result (iii) in Lemma 1, we have $\forall 0 \leq a \leq b \leq 1$,

$$|\text{Cov}(Y, \check{X}_a) - \text{Cov}(Y, \check{X}_b)| \leq 4(b-a) \cdot \mathbb{E}[X^2] \cdot \mathbb{E}[Y^2]$$

which implies that the function $\text{Cov}(Y, \check{X}_a)$ is continuous in $a \in [0,1]$. Thus, $H_1(\cdot)$ is continuous.

Denote η as the OLS estimand of X on W. Then we have $X^{\perp W} = X - W'\eta$. Since $\text{Cov}(X - X_a^*, W) = 0$, we have $X_a^{*\perp W} = X_a^* - W'\eta$ and it is easy to show $\text{Var}(X_a^*) \geq \text{Var}(\mathbb{E}[X \mid V])$ for all $a \in [0,1]$, then

$$\begin{aligned} H_2(a) =& \text{Var}(X_a^*) - \text{Var}(W'\eta) \\ \geq& \text{Var}(\mathbb{E}[X \mid V]) - \text{Var}(W'\eta) = \text{Var}(\mathbb{E}[X^{\perp W} \mid V]) > 0 \end{aligned}$$

Denote $F_{V,X}(\cdot)$ as the c.d.f. of (V, X), since $(X, V) \overset{d}{=} (\check{X}_a, V)$, we have

$$\begin{aligned} H_2(a) =& \mathbb{E}[X_a^{*2}] - (\mathbb{E}[X])^2 - \text{Var}(W'\eta) \text{ (since } \mathbb{E}[X_a^*] = \mathbb{E}[X]) \\ =& \mathbb{E}\left\{\mathbb{E}[X \mid V, \check{X}_a]^2\right\} - (\mathbb{E}[X])^2 - \text{Var}(W'\eta) \\ =& \int \mathbb{E}[X \mid V = v, \check{X}_a = x]^2 dF_{V,X}(v,x) - (\mathbb{E}[X])^2 - \text{Var}(W'\eta) \\ =& \int \left\{ax + (1-a)\mathbb{E}[X \mid V = v]\right\}^2 dF_{V,X}(v,x) - (\mathbb{E}[X])^2 - \text{Var}(W'\eta) \end{aligned}$$

which is continuous in a. Thus, we verify the continuity of $H(\cdot)$.

Step 3 (Mean-value theorem): Note that

$$H(0) = \frac{\text{Cov}(Y^{\perp W}, \mathbb{E}[X|V])}{\text{Var}(\mathbb{E}[X^{\perp W}|V])}$$

and using $\check{X}_1 = X$

$$H(1) = \frac{\text{Cov}(Y^{\perp W}, \mathbb{E}[X|V, X])}{\text{Var}(\mathbb{E}[X|V, X]^{\perp W})} = \frac{\text{Cov}(Y^{\perp W}, X)}{\text{Var}(X^{\perp W})}$$

Then, we can write inequalities (43) as

$$H(0) \leq \frac{\text{Cov}(Z^{\perp W}, Y)}{\text{Cov}(Z^{\perp W}, X)} \leq H(1)$$

By mean value theorem, there exists $a^* \in [0, 1]$ such that

$$H(a^*) = \frac{\text{Cov}(Z^{\perp W}, Y)}{\text{Cov}(Z^{\perp W}, X)}$$

Hence, we show that the exclusion condition holds under the true variable is $X^*_{a^*}$ and finish the proof. *Q.E.D.*

Lemma 3 specifies conditions under which the lower and upper bounds in Theorem 1 are attained. These requirements on the slope function may not be unique if $\Pr(X^* = 0) \neq 0$. We also require that X^* and X have the same sign almost everywhere.

Lemma 3. *Suppose* $\text{sgn}(X^*) = \text{sgn}(X)$ *hold almost surely. Suppose the assumptions of Theorem 1 hold. Then, the upper bound* $\overline{B}(\lambda_l, \lambda_u)$ *is attained if the slope function satisfies*

$$\lambda(x^*, w, z) = \begin{cases} \lambda_u & \textit{if } (z - \pi' w)x^* \geq 0 \\ \lambda_l & \textit{otherwise} \end{cases}$$

Moreover, the lower bound $\underline{B}(\lambda_l, \lambda_u)$ *is attained if*

$$\lambda(x^*, w, z) = \begin{cases} \lambda_l & \textit{if } (z - \pi' w)x^* \geq 0 \\ \lambda_u & \textit{otherwise} \end{cases}$$

Proof of Lemma 3. Consider the slope function:

$$\lambda(x^*, w, z) = \begin{cases} \lambda_a & \text{if } (z - \pi' w)x^* \geq 0 \\ \lambda_b, & \text{otherwise} \end{cases}$$

for some $\lambda_a, \lambda_b \in [\lambda_l, \lambda_u]$. Note that $\text{sgn}(X^*) = \text{sgn}(X)$ implies $\text{sgn}(Z^{\perp W} X^*) = \text{sgn}(Z^{\perp W} X)$. By equation (29), we have

$$\begin{aligned} \text{Cov}(Z^{\perp W}, X^*) &= \mathbb{E}\left[Z^{\perp W} \cdot \frac{X}{\lambda(X^*, W, Z)}\right] \\ &= \mathbb{E}\left[Z^{\perp W} \cdot \frac{X}{\lambda_b + (\lambda_a - \lambda_b)\mathbf{1}\{Z^{\perp W} X^* \geq 0\}}\right] \\ &= \mathbb{E}\left[Z^{\perp W} \cdot \frac{X}{\lambda_b + (\lambda_a - \lambda_b)\mathbf{1}\{Z^{\perp W} X \geq 0\}}\right] \\ &= \frac{1}{\lambda_a} \cdot \mathbb{E}[Z^{\perp W} X] - \frac{1}{2} \cdot (\frac{1}{\lambda_b} - \frac{1}{\lambda_a}) \cdot (\mathbb{E}(|Z^{\perp W} X|) - \mathbb{E}[Z^{\perp W} X]) \end{aligned}$$

Then,

$$\frac{\beta^{\text{IV}}}{\beta} = \frac{\text{Cov}(Z^{\perp W}, X^*)}{\text{Cov}(Z^{\perp W}, X)} = \frac{1}{\lambda_a} - (\frac{1}{\lambda_b} - \frac{1}{\lambda_a}) \cdot \alpha = \frac{1 + \alpha}{\lambda_a} - \frac{1}{\lambda_b}$$

Let $(\lambda_a, \lambda_b) = (\lambda_u, \lambda_l)$ and (λ_l, λ_u), we finish the proof. *Q.E.D.*

We are now ready to prove Theorem C.1.

Proof of Theorem C.1. Consider the upper bound, let $\tilde{X} = X(\lambda_u, \lambda_l)$. By definition (42), we have $\text{sgn}(X) = \text{sgn}(\tilde{X})$. By Lemma 2, there exists random variable $\tilde{X}^*$ such that $\text{sgn}(\tilde{X}^*) = \text{sgn}(\tilde{X})$ and

$$\mathbb{E}[\tilde{X} - \tilde{X}^* \mid \tilde{X}^*, W, Z] = 0$$

By the construction of $\tilde{X} = X(\lambda_u, \lambda_l)$ (defined in (42)), the slope function of measurement X is $\lambda(\tilde{x}^*, w, z) = \lambda_l + (\lambda_u - \lambda_l)\mathbf{1}\{(z - \pi' w)\tilde{x}^* \geq 0\}$, where we treat $\tilde{X}^*$ as the true variable. Then by Lemma 3, the upper bound in Theorem 1 is achieved for $(Y, X, W, Z, \tilde{X}^*)$. The proof of the lower bound is similar; thus, it is omitted. *Q.E.D.*

F.3 Proofs for Appendix D

Proof of Proposition D.1. Note that

$$\beta^{\text{IV}} = \frac{\text{Cov}(Z^{\perp W}, Y)}{\text{Cov}(Z^{\perp W}, X)} = \beta \cdot \frac{\text{Cov}(Z^{\perp W}, X^*)}{\text{Cov}(Z^{\perp W}, X)} + \frac{\text{Cov}(Z^{\perp W}, \epsilon)}{\text{Cov}(Z^{\perp W}, X)}$$

By Assumption D.1,we have

$$\frac{|\text{Cov}(Z^{\perp W}, \epsilon)|}{\text{Var}(Z^{\perp W})} = |\chi_1| \leq \kappa$$

Then,

$$\beta^{\text{IV}} \left(1 - \frac{\text{Var}(Z^{\perp W})}{\text{Cov}(Z^{\perp W}, Y)} \cdot \kappa\right) \leq \beta \cdot \frac{\text{Cov}(Z^{\perp W}, X^*)}{\text{Cov}(Z^{\perp W}, X)} \leq \beta^{\text{IV}} \left(1 + \frac{\text{Var}(Z^{\perp W})}{\text{Cov}(Z^{\perp W}, Y)} \cdot \kappa\right) \tag{44}$$

where both upper and lower bounds are positive since $\text{Var}(Z^{\perp W}) \cdot \kappa \leq |\text{Cov}(Z^{\perp W}, Y)|$. Following the proof of Theorem 1, we have

$$\frac{1+\alpha}{\lambda_u} - \frac{\alpha}{\lambda_l} \leq \frac{\text{Cov}(Z^{\perp W}, X^*)}{\text{Cov}(Z^{\perp W}, X)} \leq \frac{1+\alpha}{\lambda_l} - \frac{\alpha}{\lambda_u} \tag{45}$$

Combined equation (44) with (45), we finish the proof. *Q.E.D.*