Partial Identification of Nonlinear Models with Misclassification Error: A Perturbation Approach*

Yingyao Hu† Ji-Liang Shiu‡

April 25, 2024

Abstract

In this paper, we derive partially identified sets for various nonlinear models with misclassification errors using sup-norm deviations to relax the conditional independence assumptions required for point identification. We express these deviations as a perturbation matrix between an observable matrix and an unobserved eigenvalue-eigenvector decomposition. The perturbation theory of the eigenvalues of a diagonalizable matrix then provides bounds indexed by the upper bound of misreporting probabilities and deviations. As the deviations approach zero, the nonparametric partial identification of the nonlinear models with misclassification error becomes the point identification. We propose a systematic sensitivity analysis to construct the identified sets incorporating more practical information to determine the upper bounds of deviations. Our simulations imply that the identified sets with the recommended upper bounds can cover the true parameters of interest, and conclusions may apply locally rather than globally. Then we illustrate the partial identification approach by investigating the impact of misreported schooling on wages.

Keywords: *general misclassified covariate, partial independence, systematic sensitivity analysis, nonparametric identification.*

---

*Helpful comments by Francesca Molinari, Alexandre Poirier, Ruli Xiao, and Li Zheng are acknowledged. We also thank participants at CEMP 2021 Workshop in Jinan University, Penn State-Cornell Econometrics and IO Conference, 2021 and 2021 Conference on Econometrics and Statistics in Xiamen University for helpful comments and discussions. The authors are solely responsible for any remaining errors.

†Department of Economics, Johns Hopkins University. Email: yhu@jhu.edu.

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

1. Introduction

This paper focuses on nonparametric nonlinear models where a discrete covariate X^* is measured with an error as X, which is also discrete. The problem of nonlinear models with misclassified covariates has been analyzed in a few studies, including Kane et al. (1999), Walter and Hui (1980), Ramalho (2002), Bollinger (1996), and Aigner (1973). Under some ideal conditions, one can nonparametrically point-identify models with misclassification errors. Mahajan (2006) studies the problem of nonparametrically identifying and estimating regression models with a misclassified binary regressor using a binary instrument. Lewbel (2007) uses instruments to solve misclassification in binary treatment effect models. Hu (2008) proposes using instrumental variables to identify and estimate nonlinear models with misclassification errors in a general discrete explanatory variable. Hu (2008) formalizes the identification problems using matrix notation for the general discrete case, fully exploiting an analogy between a discrete repeated measurement model and matrix diagonalization. Chen et al. (2009) also adopt the matrix notation to identify the nonparametric regression model containing misclassified discrete regressor without relying on any additional variable (such as instruments). Their result relies on the monotonicity of the regression function and a testable rank condition.

Point identification of econometric models often relies on conditional independence assumptions, including for models with misclassification errors. The type of conditional independence assumed has important implications for whether we can achieve point identification. This paper addresses the identification problem of various nonlinear models with misclassification errors by gradually relaxing the baseline conditional independence assumptions underlying point identification.

We adopt Masten and Poirier (2018)'s approach of using sup-norm deviations to relax conditional independence assumptions into conditional partial independence assumptions. The sup-norm difference represents nonparametric deviations from conditional independence. We express deviations as a perturbation matrix relating an observable matrix to an unobserved eigenvalue-eigenvector decomposition using the matrix diagonalization technique in Hu (2008) and then apply Bauer-Fike perturbation theory of the eigenvalue of a diagonalizable matrix to analyze identification. We obtain identified sets for nonlinear models with misclassification errors, including conditional density and regression functions. The width of the identified sets depends on three factors: the condition number of the matrix of misclassification probabilities, the matrix norms of the observed data, and the sup-norm of the deviations. Thus, we can index the identified sets by the upper bound of misreporting probabilities and the deviations for an observed sample.

Since a reasonable value for the upper bound of misreporting probabilities may come from validation studies or economic theories, we regard this value as prior information and incorporate it as baseline information into constructing the identified sets. Given a known value for the upper bound of misreporting probabilities, we propose a systematic sensitivity analysis with two steps: (1) Impose the identification assumptions on the sensitivity parameter and the restrictions from probability theory, such as density restrictions. (2) Specify additional desired conclusions to conduct breakdown frontier analysis. We choose the upper bound for the deviation or sensitivity parameter as the minimum of the two upper bounds from these two steps to have strong identifying power. Our simulations imply that the outer identified set with the recommended upper bound covers the true parameters of interest. The partial identification results rely on the unique solution of a matrix diagonalization technique, which requires: (1) A distinct eigenvalue restriction on the latent models; and (2) A higher chance of reporting truthfully.

We conduct simulations to compare partial identification using the proposed approach to point identification under varying deviations from conditional independence. The nonparametric bounds show moderate values and cover the truth in most designs, recovering identifying power lost to deviations and giving reasonably accurate estimates for observed data-however, inferences far beyond that risk misleading findings. The empirical illustration explores the sensitivity of the conditional mean log wage to deviations from conditional independence, given the true schooling levels. The identified sets strictly contain the 95% CI of point estimates, showing the latter underestimates wage differences by education level due to assuming conditional independence. By systematically exploring sensitivity to deviations from model assumptions, we gain insight into where conclusions could be drawn from observed data. The proposed approach provides a nonparametric characterization of neighborhoods yielding inferences reasonably robust to failures of conditional independence assumptions. However, the approach can not achieve the global outer identified set, limiting generalizability.

This paper examines conditions under which nonparametric partial identification of nonlinear models with misclassification error approaches point identification as sup-norm deviations from conditional independence decrease toward zero. As deviations shrink, partial identification increasingly approximates point identification. Hu (2008) imposed common restrictions on misclassification probabilities to point-identify model parameters of interest. However, some data may contain further information beyond his assumptions, which cannot be readily used within a point-identified framework. Because the identified sets in our proposed approach are indexed by upper bounds on misreporting probabilities and deviations, we can directly incorporate any prior information about misreporting patterns into the analysis. By systematically exploring sensitivity over a range of plausible bounds, we gain insight into where inferences may reasonably apply even if the overall outer identified set is unrecoverable.

Our bounding approach relates to and extends the existing set identification literature (e.g. Manski (2003), Imbens and Manski (2004), Chernozhukov et al. (2007), Magnac and Maurin (2008), Bontemps et al. (2012), Chandrasekhar et al. (2012), Chesher (2013), Masten and Poirier (2016), Masten and Poirier (2018), Chen et al. (2018), Chen et al. (2021)). Masten and Poirier (2016) consider three classes of weaker exogeneity assumptions deviating from full statistical independence. They study identification in nonseparable models under these three different classes of deviations. The first deviation is based on quantile independence, while the second and third are based on a distance-from-independence metric using either a conditional CDF or propensity score. Masten and Poirier (2018) derive identified sets for various treatment effect parameters using the sup-norm of the deviations from conditional independence of treatment assignment from potential outcomes. Their work develops fully nonparametric methods for sensitivity analysis using only one sensitivity parameter. In contrast, we can conduct a breakdown frontier analysis because our identification sets are represented with a two-dimensional index. The literature on the breakdown frontier analysis goes back to Horowitz and Manski (1995), in which they study a measurement error model and index identification sets by an upper bound of misreporting probabilities. The extensive literature on this includes Imbens (2003); Manski (2009); Stoye (2010); Gundersen et al. (2012); Kline and Santos (2013); Manski and Pepper (2018); and Masten and Poirier (2020). Kline and Santos (2013) study a class of relaxations of the missing-at-random assumption in a missing data model, index the magnitude of the relaxation, define an identification breakdown point and develop a weighted bootstrap procedure to conduct inference on the breakdown point. Masten and Poirier (2020) study a potential outcomes model with a binary treatment where a two-dimensional index parameterizes relaxations of baseline assumptions. They derive the breakdown frontier for two kinds of assumptions for a specific conclusion and propose an inference on the multidimensional breakdown frontier. We apply Masten and Poirier (2020)'s breakdown frontier analysis to select the upper bounds of deviations and construct identified sets. An economic study similar in spirit to our own is Aguiar and Serrano (2017) in which they propose an approach to measure departures from rationality and classify departures corresponding to three anomalies: inattentiveness to changes in purchasing power, money illusion, and violations of the compensated law of demand.

The source of set identification problems can often be traced to imposing prior restrictions on measurement errors. Examples of such bounding techniques in discrete latent variable models that our approach builds upon include Klepper (1988); Horowitz and Manski (1995); Black et al. (2000); Molinari (2008); and Adams (2016). Horowitz and Manski (1995) introduce fully nonparametric methods to obtain informative bounds on the distribution of a contaminated random variable, assuming upper bounds on probabilities of misreport. Molinari (2008) introduces the direct misclassification approach based on the matrix of misclassification probabilities. This approach incorporates any prior information into the analysis through sets of restrictions on the misclassification probabilities matrix to derive identification regions.

The work of this paper is closely related to Molinari (2003, 2008), which also employs matrix notation to analyze the misclassification of discrete explanatory variables. Molinari (2003, 2008) shows that when the probability of correct report exceeds one-half, bounds on the mean regression identification region can be estimated. The bounding method differs from the proposed approach because the source of our set identification comes from the continuous relaxation of the conditional independence assumptions, and we show how partial identification becomes point identification.

Section 2 presents partial identification results for misclassification models, including conditional density functions and regression functions under sup-norm deviations. Section 3 develops $\sqrt{N}-$consistent nonparametric estimators for the lower bound of the identification sets in Section 2. Section 4 proposes a systematic sensitivity analysis and describes a two-step selection procedure to calculate the upper bounds of the sensitivity parameter for identification regions. Section 5 illustrates identification regions and the systematic sensitivity analysis in two numerical examples. Section 6 empirically illustrates the estimators for bounds of the mean log wage across three education categories. Section 7 concludes. The appendix contains the proofs for all propositions.

2. Partial Identification of Misclassification Models

2.1. Partial Identification of the Latent Density

Let Y represent an observed outcome variable. We define X^* as an unobserved latent categorical variable subject to misclassification. Assume X^* takes on the values $\{1,2,3,\ldots,K\}$ for some known positive integer K and denote the support of X^* as $\mathcal{K}=\{1,\ldots,K\}$. Define the conditional density of the observed outcome Y on X^* as

$$f_{Y|X^*}(y|x^*). \tag{1}$$

We directly observe only the proxy measure X, which provides imperfect information on the latent variable X^* subject to misclassification error. The misclassification error may be correlated with X^* and assume there exists an instrumental variable Z. Without loss of generality, we assume X, Z, and X^* have the same support $\mathcal{K} = \{1,...,K\}$.[1] We have $\{Y,X,Z\}$ for an observable i.i.d. sample. Lowercase letters will be used to describe particular values of the corresponding uppercase random variables. To keep things simple, we are not including any extra accurately measured regressors W in the conditioning set. However, our proposed method can easily be expanded to take W into account by conditioning on its values. Hu (2008) demonstrates that the conditional density $f_{Y|X^*}$ is nonparametrically identified under following assumption:

Condition 2.1. *(Conditional Independence) The variable (Y,X,Z,X^*) satisfies a conditional independence assumption as follows*

$$f_{YXZ|X^*}\left(Y,X,Z|X^*\right) = f_{Y|X^*}\left(Y|X^*\right) f_{X|X^*}\left(X|X^*\right) f_{Z|X^*}\left(Z|X^*\right), \tag{2}$$

or

$$Y \perp X \perp Z \Big| X^* \tag{3}$$

We relax the conditional independence assumption rather than imposing it directly. Specifically, we weaken this assumption in the following way:

Assumption 2.1. *(Conditional Partial Independence) We define the perturbation or deviation term from the conditional independence in Equation (2), denoted as $d_0(Y,X,Z,X^*)$, as follows:*

$$d_0(Y,X,Z,X^*) \equiv f_{YXZ|X^*}\left(Y,X,Z|X^*\right) - f_{Y|X^*}\left(Y|X^*\right) f_{X|X^*}\left(X|X^*\right) f_{Z|X^*}\left(Z|X^*\right). \tag{4}$$

There exists a constant $h \geq 0$ such that $|d_0(Y,X,Z,X^*)| \leq h$ and $\int_y |d_0(y,X,Z,X^*)|\,dy \leq h$.

[1]There are two measurements X and Z of the latent variable X^* and as discussed in Hu (2017), one of the requirements for them is that the cardinality of the supports of X and Z are greater than or equal to that of the latent variable X^*. We rule out the case that X or Z takes fewer values than X^*. When X or Z is continuous, we can discretize X or Z to generate new measurements taking the same numbers of possible values as the latent variable X^*. When X or Z is discrete, and their numbers of possible values are more than X^*, we can regroup them to share the same number of the support of the latent variable X^*.

If Y is discrete, the condition $\int_y |d_0(y,X,Z,X^*)|\,dy \le h$ implies $|d_0(Y,X,Z,X^*)| \le h$. Conditional partial independence allows the conditional joint probability $f_{YXZ|X^*}(Y,X,Z|X^*)$ to deviate from the product of the conditional marginal probabilities $f_{Y|X^*}(Y|X^*)$, $f_{X|X^*}(X|X^*)$, and $f_{Z|X^*}(Z|X^*)$. Note that $\int_y d_0(y,X,Z,X^*)dy = f_{X,Z|X^*}(X,Z|X^*) - f_{X|X^*}(X|X^*)f_{Z|X^*}(Z|X^*)$. This term $\int_y d_0(y,X,Z,X^*)dy$ represents the deviation term from the conditional independence of X and Z given X^* in which the conditional independence also holds under the conditional independence in Equation (2). We refer to the parameter h as the sensitivity parameter. When this sensitivity parameter h is zero, conditional partial independence reduces to the conditional independence condition in Condition 2.1. When the value of the sensitivity parameter h is greater than $f_{Y|X^*}(Y|X^*)f_{X|X^*}(X|X^*)f_{Z|X^*}(Z|X^*)$, and $1 - f_{Y|X^*}(Y|X^*)f_{X|X^*}(X|X^*)f_{Z|X^*}(Z|X^*)$, the condition $|d_0(Y,X,Z,X^*)| \le h$ does not restrict the conditional probability $f_{YXZ|X^*}(Y,X,Z|X^*)$ in any way.[2] As discussed by Masten and Poirier (2018), we interpret a binding value of our sensitivity parameter h as imposing a restriction on latent structures or selection on unobservables, similar to their approach.

Consider the perturbation term from the conditional independence in Assumption 2.1 as

$$
\begin{aligned}
(5)\qquad d_0(Y,X,Z,X^*) = {} & \Big(f_{Y|XZX^*}\left(Y|X,Z,X^*\right) f_{X|ZX^*}\left(X|Z,X^*\right) f_{Z|X^*}\left(Z|X^*\right) \\
& \quad - f_{Y|XZX^*}\left(Y|X,Z,X^*\right) f_{X|X^*}\left(X|X^*\right) f_{Z|X^*}\left(Z|X^*\right)\Big) \\
& + \Big(f_{Y|XZX^*}\left(Y|X,Z,X^*\right) f_{X|X^*}\left(X|X^*\right) f_{Z|X^*}\left(Z|X^*\right) \\
& \quad - f_{Y|X^*}\left(Y|X^*\right) f_{X|X^*}\left(X|X^*\right) f_{Z|X^*}\left(Z|X^*\right)\Big) \\
= {} & f_{Y|XZX^*}\left(Y|X,Z,X^*\right)\left(f_{X|ZX^*}\left(X|Z,X^*\right) - f_{X|X^*}\left(X|X^*\right)\right) f_{Z|X^*}\left(Z|X^*\right) \\
& + \left(f_{Y|XZX^*}\left(Y|X,Z,X^*\right) - f_{Y|X^*}\left(Y|X^*\right)\right) f_{X|X^*}\left(X|X^*\right) f_{Z|X^*}\left(Z|X^*\right).
\end{aligned}
$$

The conditions (i) $f_{Y|XZX^*}(Y|X,Z,X^*) = f_{Y|X^*}(Y|X^*)$ and (ii) $f_{X|ZX^*}(X|Z,X^*) = f_{X|X^*}(X|X^*)$ imply the zero deviation and the conditional independence holds. The condition (i) implies that the measurement error is nondifferential, that is, $X - X^*$ does not affect the distribution of the dependent variable Y conditional on the true value X^*. Relaxing the restriction allows a possible correlation between the measurement error $X - X^*$ and the dependent variable Y. As for the condition (ii), it implies that the misclassification error in X is independent of the instrumental variable Z conditional on X^*. When the instrument Z is a repeated measurement

[2] The condition $|d_0(Y,X,Z,X^*)| \le h$ implies that $-h < f_{YXZ|X^*}(Y,X,Z|X^*) - f_{Y|X^*}(Y|X^*)f_{X|X^*}(X|X^*)f_{Z|X^*}(Z|X^*) < h$ which is equivalent to $-h + f_{Y|X^*}(Y|X^*)f_{X|X^*}(X|X^*)f_{Z|X^*}(Z|X^*) < f_{YXZ|X^*}(Y,X,Z|X^*) < h + f_{Y|X^*}(Y|X^*)f_{X|X^*}(X|X^*)f_{Z|X^*}(Z|X^*)$. When h takes these values, $-h + f_{Y|X^*}(Y|X^*)f_{X|X^*}(X|X^*)f_{Z|X^*}(Z|X^*) < 0$ and $h + f_{Y|X^*}(Y|X^*)f_{X|X^*}(X|X^*)f_{Z|X^*}(Z|X^*) > 1$, the restriction from h is not binding for $f_{YXZ|X^*}(Y,X,Z|X^*)$.

of X^*, relaxing the restriction allows a possible correlation between the measurement error $X - X^*$ and the measurement error $Z - X^*$. Thus, we can decompose the perturbation term into two components: (1) Deviation due to differential measurement error (relaxing condition (i)), and (2) Deviation due to dependent misclassification error (relaxing condition (ii)). Relaxing the two conditions allows modeling differential measurement error and dependent misclassification, which are important relaxations for practical applications.

Next, we introduce matrix notations to express the deviations as a perturbation matrix relating an observable matrix to an unobserved eigenvalue-eigenvector decomposition. Multiplying Equation (4) by $f_{X^*}(x^*)$ and summing over the support of X^* yields the following.

$$\sum_{x^*} d_0(Y,X,Z,x^*) f_{X^*}\left(x^*\right) = f_{YXZ}(Y,X,Z) - \sum_{x^*} f_{Y|X^*}\left(Y|X^*\right) f_{X|X^*}\left(X|X^*\right) f\left(Z,x^*\right). \tag{6}$$

Define $d_1(Y,X,Z) = \sum_{x^*} d_0(Y,X,Z,x^*) f_{X^*}(x^*)$ and for each y, define the $K \times K$ matrix $M_d(y)$ as follows

$$M_d(y) = \begin{bmatrix} d_1(y,1,1) & \cdots & d_1(y,K,1) \\ \vdots & & \\ d_1(y,1,K) & \cdots & d_1(y,K,K) \end{bmatrix}_{K\times K}. \tag{7}$$

Because $|d_1(Y,X,Z)| \leq \sum_{x^*} |d_0(Y,X,Z,x^*)| f_{X^*}(x^*) \leq h$ by Assumption 2.1, every entry of the matrix $M_d(y)$ is less than or equal to h.

Given y, define the following matrices:

$$M_{yXZ} = \begin{bmatrix} f_{YXZ}(y,1,1) & \cdots & f_{YXZ}(y,K,1) \\ \vdots & \cdots & \vdots \\ f_{YXZ}(y,1,K) & \cdots & f_{YXZ}(y,K,K) \end{bmatrix}_{K\times K}, \tag{8}$$

$$M_{X^*Z} = \begin{bmatrix} f_{X^*Z}(1,1) & \cdots & f_{X^*Z}(K,1) \\ \vdots & \cdots & \vdots \\ f_{X^*Z}(1,K) & \cdots & f_{X^*Z}(K,K) \end{bmatrix}_{K\times K}, \tag{9}$$

$$D_{y|X^*} = \begin{bmatrix} f_{Y|X^*}(y|1) & 0 & \cdots & 0 \\ \vdots & f_{Y|X^*}(y|k) & \vdots & 0 \\ 0 & \cdots & 0 & f_{Y|X^*}(y|K) \end{bmatrix}_{K\times K}, \tag{10}$$

$$
(11) \qquad M_{X|X^*} = \begin{bmatrix} f_{X|X^*}(1|1) & \cdots & f_{X|X^*}(K|1) \\ \vdots & \cdots & \vdots \\ f_{X|X^*}(1|K) & \cdots & f_{X|X^*}(K|K) \end{bmatrix}_{K\times K},
$$

$$
(12) \qquad M_{XZ} = \begin{bmatrix} f_{XZ}(1,1) & \cdots & f_{XZ}(K,1) \\ \vdots & \cdots & \vdots \\ f_{XZ}(1,K) & \cdots & f_{XZ}(K,K) \end{bmatrix}_{K\times K}.
$$

With the above notation, given each y we can write out Equation (6) by the matrix expression

$$
(13) \qquad M_d(y) = M_{yXZ} - M_{X^*Z} D_{y|X^*} M_{X|X^*}.
$$

Summing over all y, we have

$$
(14) \qquad M_d = M_{XZ} - M_{X^*Z} M_{X|X^*},
$$

where $M_d = \int_y M_d(y)dy$. The (i,j) entry of the matrix M_d is $\int_y d_1(y,i,j)dy$ which is bounded by $|\int_y \sum_{x^*} d_0(y,i,j,x^*) f_{X^*}(x^*)dy| \leq \sum_{x^*}\int_y |d_0(y,i,j,x^*)| dy f_{X^*}(x^*) \leq h$ by Assumption 2.1. Thus, every matrix entry M_d is less than or equal to h.

Rewrite Equations (13) and (14) as

$$
(15) \qquad M_{X^*Z} D_{y|X^*} M_{X|X^*} = M_{yXZ} - M_d(y),
$$

$$
(16) \qquad M_{X^*Z} M_{X|X^*} = M_{XZ} - M_d.
$$

To relate the magnitude of deviation matrices to our sensitivity parameter h, we introduce and define a matrix norm.[3] The set of all $K \times K$ square matrices over $\mathbb{R}$ is denoted by M_K. Denote $\|x\|_\infty = \max\{|x_1|,\ldots,|x_K|\}$ as the max norm ($l_\infty-$norm) for a vector x in $\mathbb{R}^K$. Any matrix in M_K induces a linear operator from $\mathbb{R}^K$ to $\mathbb{R}^K$ for a basis in $\mathbb{R}^K$, and for $A \in M_K$ define a corresponding operator norm or a matrix norm as follows:

$$
(17) \qquad |||A||| = \sup\left\{\frac{\|Ax\|_\infty}{\|x\|_\infty} : x \in \mathbb{R}^K \text{ with } x \neq 0\right\}.
$$

The above matrix norm $|||\cdot|||$ is induced by the $l_\infty-$norm on $\mathbb{R}^K$ and has the following important

[3]For reference, our definition of the matrix norm in this section follows the treatment in Chapter 5 of Matrix Analysis by Horn and Johnson (2013).

expression:

$$\text{(18)} \qquad |||A||| = \max_{1\le i\le K} \sum_{j=1}^{K} |a_{ij}|.$$

where $A = [a_{ij}]_{1\le i,j\le K}$ and this is the maximum row sum matrix norm for $A \in M_K$.[4] Since every entry of the matrix $M_d(y)$ is less than h, applying the result in Equation (18) to $M_d(y)$ yields

$$\text{(19)} \qquad |||M_d(y)||| \le Kh.$$

Similarly, we obtain

$$\text{(20)} \qquad |||M_d||| \le Kh.$$

To ensure the invertibility of M_{X^*Z} through Equation (16), we will make the following assumptions and restrict the sensitivity parameter h.

Assumption 2.2. *Assume that (i) M_{XZ} is invertible and $h < \frac{1}{K|||M_{XZ}^{-1}|||} \equiv h_I$;*
(ii) $M_{X|X^*}$ is invertible.

Combining Assumption 2.2(i) with Equation (20) yields $|||M_d||| < \frac{1}{|||M_{XZ}^{-1}|||}$ and $M_{XZ} - M_d$ is invertible under the result in Lemma A.1 for the inverse of a perturbed matrix. The invertibility of M_{XZ} in Assumption 2.2(i) is testable and can be verified empirically. Applying these assumptions, the matrix relationship described in Equations (15) and (16) gives rise to the following eigenvalue-eigenvector decomposition:

$$\text{(21)} \qquad M_{X|X^*}^{-1} D_{y|X^*} M_{X|X^*} = (M_{XZ} - M_d)^{-1} \left(M_{yXZ} - M_d(y)\right).$$

Applying Lemma A.1 to $M_{XZ} - M_d$, we have

$$\text{(22)} \qquad (M_{XZ} - M_d)^{-1} \equiv M_{XZ}^{-1} + B_d,$$

where $B_d = \sum_{i=1}^{\infty} (M_{XZ}^{-1} M_d)^n M_{XZ}^{-1}$. Use the notation to rewrite Equation (21) as

$$\text{(23)} \qquad \underbrace{M_{X|X^*}^{-1} D_{y|X^*} M_{X|X^*}}_{\text{a diagonal structure}} = \underbrace{M_{XZ}^{-1} M_{yXZ}}_{\text{observable}} + P_y^h(d),$$

[4]The result is in Example 5.6.5 of Horn and Johnson (2013).

where $P_y^h(d) = -M_{XZ}^{-1}M_d(y) + B_d\left(M_{yXZ} - M_d(y)\right)$. Notice the perturbation matrix $P_y^h(d)$ related to the deviation matrices $M_d(y)$ and M_d and the matrices $M_d(y)$ and M_d are relatively small if h is small. Our partial identification strategy incorporates observed information to determine properties of the latent distribution without imposing conditional independence assumptions. While the left-hand side of Equation (23) connects to the latent model, the right-hand side of Equation (23) relates to the matrices of the observed densities and the deviation matrices. Identification of the properties of our latent model relies on conditions imposed on the latent eigenvalue-eigenvector structure $M_{X|X^*}^{-1}D_{y|X^*}M_{X|X^*}$. Specifically, these conditions require: (1) An assumption ruling out eigenvalues of $M_{X|X^*}^{-1}D_{y|X^*}M_{X|X^*}$ with multiplicity greater than 1. (2) An assumption ensuring an ordering of the eigenvectors of $M_{X|X^*}^{-1}D_{y|X^*}M_{X|X^*}$.

Assumption 2.3. *(Distinct Eigenvalues) For each y, $f_{Y|X^*}(y|x^* = i) \neq f_{Y|X^*}(y|x^* = j)$ for all $i \neq j$.*

Assumption 2.4. *(Ordering of Eigenvectors) For some constant $0 \leq \lambda < \frac{1}{2}$, define the upper bound on the probabilities of misreport*

$$\sum_{j \neq i} f_{X|X^*}(j|i) \leq \lambda \quad \forall i, \tag{24}$$

which is equivalent to the lower bound on the probabilities of correct report

$$f_{X|X^*}(i|i) \geq 1 - \lambda > 0 \quad \forall i. \tag{25}$$

Assumption 2.3 suffices to ensure distinct eigenvalues,[5] while Assumption 2.4 determines the ordering of the corresponding eigenvectors. Therefore, the eigenvalue-eigenvector decomposition in the latent structure $M_{X|X^*}^{-1}D_{y|X^*}M_{X|X^*}$ is identified. Assumption 2.4 requires that eigenvectors or row vectors of $M_{X|X^*}$ is identified by the largest entry in that row. If the largest entry of an eigenvector of $M_{X|X^*}$ is the jth entry, that eigenvector is equal to the jth row of the matrix $M_{X|X^*}$. Assumption 2.4 implies that $M_{X|X^*}$ is strictly diagonally dominant and invertible. Thus, we do not need Assumption 2.2(ii) when Assumption 2.4 is assumed. In other words, Assumption 2.4 restricts that the probabilities of truthful reporting always exceed

[5] A sufficient condition more general than Assumption 2.3 is that there exists a function $\omega(\cdot)$, such that $E\left[\omega(y)|x^* = i\right] \neq E\left[\omega(y)|x^* = j\right]$ for all $i \neq j$. Multiplying both sides of Equation (6) by $\omega(Y)$ and integrating over Y yields an equation analogous to Equation (13), $M_d(\omega(y)) = M_{\omega(y)XZ} - M_{X^*Z}D_{E[\omega(y)|X^*]}M_{X|X^*}$, where we replace each term with its weighted integral counterpart. Proceeding as in the derivation of Equation (23), we obtain the eigenvalue-eigenvector decomposition $M_{X|X^*}^{-1}D_{E[\omega(y)|X^*]}M_{X|X^*} = M_{XZ}^{-1}M_{\omega(y)XZ} + E_{\omega(y)}(d)$, where $E_{\omega(y)}(d)$ represents deviations. In this decomposition, $E\left[\omega(y)|X^* = k\right]$ is an eigenvalue, but the eigenvectors in $M_{X|X^*}$ are unchanged from Equation (23). Therefore, according to Assumption 2.3, the eigenvalues $E\left[\omega(y)|X^* = k\right]$ of this decomposition are distinct.

misreporting probabilities. This assumption is credible and consistent with various validation studies. For example, misclassification probabilities of the self-reported employment status in Poterba and Summers (1995) and Bound et al. (2001), and misclassification probabilities of the self-reported education attainment in Kane et al. (1999) satisfy this assumption. Horowitz and Manski (1995) and Molinari (2003, 2008) also adopt the assumption and assume a known lower bound on the probabilities of correct report is available.

Equation (23) represents deviations as a perturbation matrix relating an observable matrix to an unobserved eigenvalue-eigenvector decomposition. This framework aligns with the perturbation theory of eigenvalues for diagonalizable matrices, allowing us to approximate unobserved eigenvalues using observed, solvable information. Specifically, we will apply the Bauer-Fike theorem[6] to derive bounds for deviations of eigenvalues of the perturbed matrix in Equation (23) from appropriately chosen eigenvalues of the observable matrix.

Theorem 2.1. *(Bauer-Fike Perturbation Theorem) Let $A \in M_K$ be diagonalizable with $A = S\Lambda S^{-1}$ and $\Lambda = diag(\rho_{A1},...,\rho_{AK})$. Suppose $E \in M_K$. If ρ is an eigenvalue of $A+E$, then there is some eigenvalue ρ_{Ak} of A for which*

$$|\rho - \rho_{Ak}| \leq |||S||| \cdot |||S^{-1}||| \cdot |||E||| = \kappa(S)|||E|||,$$

where $\kappa(\cdot)$ is the condition number with respect to the matrix norm $|||\cdot|||$.

For a given y, define

$$A = M^{-1}_{X|X^*} D_{y|X^*} M_{X|X^*}, \tag{26}$$

$$E = -P^h_y(d). \tag{27}$$

Then, $A+E = A - P^h_y(d) = M^{-1}_{XZ} M_{yXZ}$, and the corresponding matrices S, Λ, and the condition number $\kappa(S)$ are

$$S = M^{-1}_{X|X^*}, \tag{28}$$

$$\Lambda = D_{y|X^*}, \tag{29}$$

$$\kappa(S) = |||M^{-1}_{X|X^*}||| \cdot |||M_{X|X^*}|||. \tag{30}$$

We apply Theorem 2.1 to provide perturbation bounds for the eigenvalues of the matrix A in Equation (26), based on the condition number in Equation (30) and perturbation matrix

[6] The result comes from Theorem 6.3.2 in Horn and Johnson (2013).

$-P_y^h(d)$. Specifically, $\kappa(S)$ quantifies the degree of misclassification, with larger values indicating more severe misclassification. Additionally, $|||P_y^h(d)|||$ signifies how well the conditional partial independence approximates the conditional independence characterized by h, with smaller values denoting a closer approximation. Therefore, if the probability of truthful reporting is high (λ close to 0), then $\kappa(S)$ approaches 1. In this case, small perturbations in $P_y^h(d)$ will result in only minor deviations in the eigenvalues. In order to analyze the bounds of $\kappa(S)$, we adopt Assumption 2.4 and obtain the specific restriction for $\kappa(S)$ in Proposition A.1. Denote $d_{Ey} = \min_{i,j} \frac{|f_{Y|X^*}(y|x^*=i) - f_{Y|X^*}(y|x^*=j)| - \epsilon}{2} > 0$ for some $\epsilon > 0$. The bounded ranges of key parameters are summarized in the following Proposition.

Proposition 2.1. *Under the assumption of a discrete, finite support for Y and Assumptions 2.1, 2.2(i), 2.3, and 2.4, we analyze the effects of a small perturbation $P_y^h(d)$ in Equation (23) on the diagonal structure using the eigenvalues of $M_{XZ}^{-1}M_{yXZ}$. Specifically, if a small perturbation $P_y^h(d)$ satisfies the bound $|||P_y^h(d)||| \le \frac{d_{Ey}}{\overline{\kappa_\lambda}}$, then there exists distinct perturbed eigenvalues $\rho_1, ..., \rho_K$ of $M_{XZ}^{-1}M_{yXZ}$, with ρ_k for $k = 1, ..., K$ satisfying:*

$$f_{Y|X^*}\left(y|x^* = k\right) \in \left(\max\left\{\rho_k - \overline{\kappa_\lambda} \cdot \overline{P_y^h}, 0\right\}, \min\left\{1, \rho_k + \overline{\kappa_\lambda} \cdot \overline{P_y^h}\right\}\right), \tag{31}$$

where $\overline{\kappa_\lambda} = \frac{1}{(1-\lambda)(1-2\lambda)}$ and $\overline{P_y^h} = |||M_{XZ}^{-1}|||Kh + \frac{|||M_{XZ}^{-1}|||^2 Kh}{1-|||M_{XZ}^{-1}|||Kh}\left(|||M_{yXZ}||| + Kh\right)$ is an upper bound for $|||P_y^h(d)|||$.

Bauer-Fike theorem in Theorem 2.1 also postulates that the eigenvalues of $A+E$ lie in the union of the discs D_k where each D_k is centered at ρ_{Ak} and possesses a radius of $\kappa(S)|||E|||$ for $k = 1, ..., K$. The result does not preclude the possibility that several eigenvalues of $A+E$ lie in some disc D_{k0} for a particular $k0$. If the radius of each disc is strictly less than one-half of the minimum distances between these eigenvalues of A, then only one eigenvalue of $A+E$ lie in one of the discs. In order to correctly identify the model, it is essential to ensure that the degree of deviation is not greater than the distance, and this means that we must avoid mixing the perturbed eigenvalues inside one of the discs. Otherwise, the perturbed eigenvalues cannot be uniquely associated with their original eigenvalues. We apply the discussion to a perturbation matrix relation in Equation (23) to acquire the condition $|||P_y^h(d)||| \le \frac{d_{Ey}}{\overline{\kappa_\lambda}}$ and this helps to separate $f_{Y|X^*}(y|x^*)$ for different true values of x^*. As a result, one could easily identify $f_{Y|X^*}(y|x^*)$ at each x^* by examining at the location of the perturbed eigenvalues of $M_{XZ}^{-1}M_{yXZ}$. Define the sup value of the separation restriction as $h_S \equiv \sup\left\{h : \overline{P_y^h} < \frac{\min_{i,j}\left|f_{Y|X^*}(y|x^*=i) - f_{Y|X^*}(y|x^*=j)\right|}{2\overline{\kappa_\lambda}}\right\}$. If $h < h_S$, then $|||P_y^h(d)||| \le \overline{P_y^h} < \frac{\min_{i,j}\left|f_{Y|X^*}(y|x^*=i) - f_{Y|X^*}(y|x^*=j)\right|}{2\overline{\kappa_\lambda}} = \frac{1}{\overline{\kappa_\lambda}} \frac{\min_{i,j}\left|f_{Y|X^*}(y|x^*=i) - f_{Y|X^*}(y|x^*=j)\right|}{2}$ and this implies that $|||P_y^h(d)||| \leq \frac{d_{Ey}}{\overline{\kappa_\lambda}}$ for some d_{Ey}. Therefore, our proposed bound requires that the sensitivity parameter h is strictly smaller than h_S, $h < h_S$, and the separation restriction could be slight because it results from the matrix perturbation theory. This indicates the localized nature of the bounds. In the simulation section, the bounds under the separation restriction still suffice to cover the true parameter values, and they are limited in how precisely they could represent the population. Their localization around the truth suggests our bounds may apply in a neighborhood of conditional independence, but caution is needed in extrapolating findings far from that neighborhood.

Remark 2.1. *We impose the restriction $|||P_y^h(d)||| \leq \frac{d_{Ey}}{\overline{\kappa_\lambda}}$ to make h sufficiently small, thereby ensuring that each $f_{Y|X^*}(y|x^*=k)$ uniquely corresponds to the interval of the perturbed eigenvalue ρ_k without overlaps in the intervals $\left(\rho_k - \overline{\kappa_\lambda}\cdot\overline{P_y^h}, \rho_k + \overline{\kappa_\lambda}\cdot\overline{P_y^h}\right)$. If $f_{Y|X^*}(y|x^*)$ is strictly increasing in x^* for each y, we can mitigate the potential overlap problem by considering the sequential dependence of the following intervals:*

$$
\begin{aligned}
f_{Y|X^*}(y|x^*=1) &\in \left(\max\left\{\lambda_1 - \overline{\kappa_\lambda}\cdot\overline{P_y^h}, 0\right\}, \lambda_1 + \overline{\kappa_\lambda}\cdot\overline{P_y^h}\right) \\
f_{Y|X^*}(y|x^*=2) &\in \left(\max\left(f_{Y|X^*}(y|x^*=1), \lambda_2 - \overline{\kappa_\lambda}\cdot\overline{P_y^h}\right), \lambda_2 + \overline{\kappa_\lambda}\cdot\overline{P_y^h}\right) \\
&\vdots \\
f_{Y|X^*}(y|x^*=K) &\in \left(\max\left(f_{Y|X^*}(y|x^*=K-1), \lambda_K - \overline{\kappa_\lambda}\cdot\overline{P_y^h}\right), \min\left\{1, \lambda_K + \overline{\kappa_\lambda}\cdot\overline{P_y^h}\right\}\right)
\end{aligned}
$$

The condition that $f_{Y|X^*}(y|\cdot)$ increases for each y can be substituted with the condition that the expected value $E[Y|x^*]$ is strictly increasing in x^*. However, the notation $f_{Y|X^*}(y|x^*=1)$ within the bounds for $f_{Y|X^*}(y|x^*=2)$ is not conventionally valid because it tries to use the result of a set interval as a direct variable. We prefer a direct mathematical one without referencing the actual density bounds from previous intervals.

Conversely, we can adjust the handling of overlaps in eigenvalue intervals so that overlaps are included in the interval of a lower eigenvalue. This approach may be appropriate in scenarios where specific modeling choices or assumptions support that lower x^* values should have broader intervals. Assuming that $E[Y|x^*]$ is strictly increasing in x^*, we adapt the bounds on the densities to accommodate overlaps at lower x^* values as follows:

$$
\begin{aligned}
f_{Y|X^*}(y|x^*=1) &\in \left(\max\left\{\lambda_1 - \overline{\kappa_\lambda}\cdot\overline{P_y^h}, 0\right\}, \lambda_1 + \overline{\kappa_\lambda}\cdot\overline{P_y^h}\right) \\
f_{Y|X^*}(y|x^*=2) &\in \left(\max\left(\lambda_1 + \overline{\kappa_\lambda}\cdot\overline{P_y^h}, \lambda_2 - \overline{\kappa_\lambda}\cdot\overline{P_y^h}\right), \lambda_2 + \overline{\kappa_\lambda}\cdot\overline{P_y^h}\right) \\
&\vdots
\end{aligned}
$$

$$f_{Y|X^*}(y|x^* = K) \in \left(\max\left(\lambda_{K-1} + \overline{\kappa_\lambda} \cdot \overline{P_y^h}, \lambda_K - \overline{\kappa_\lambda} \cdot \overline{P_y^h}\right), \min\left\{1, \lambda_K + \overline{\kappa_\lambda} \cdot \overline{P_y^h}\right\}\right)$$

This formulation ensures that each x^* value's density bounds effectively encapsulate potential overlaps from preceding intervals, aligning with the increasing trend of $f_{Y|X^*}(y|x^*)$.

Remark 2.2. *Proposition 2.1 shows that the bounds center at distinct eigenvalues, $\rho_1, ..., \rho_K$ and the width of the bounds is close to $2\overline{\kappa_\lambda} \cdot \overline{P_y^h}$. Given a fixed upper bound on the probabilities of misreport λ_v, define the max value of the choices of these sensitivity parameters as $h_E \equiv \max\left\{h : \overline{P_y^h} \leq \frac{\min_{i,j}\left|\rho_i - \rho_j\right|}{2\overline{\kappa_{\lambda_v}}}\right\}$. For any sensitivity parameter h between 0 and h_E, the bounds $\left(\max\left\{\rho_{x^*} - \overline{\kappa_{\lambda_v}} \cdot \overline{P_y^h}, 0\right\}, \min\left\{1, \rho_{x^*} + \overline{\kappa_{\lambda_v}} \cdot \overline{P_y^h}\right\}\right)$ does not intersect the bounds of ρ_j for $j \neq x^*$. The bounds in Equations (31) using λ_v and h strictly less than h_E are disjoint. When h exceeds h_E, some of the bounds using λ_v and $\overline{P_y^h}$ may overlap with other bounds and this is a conflict with the distinct eigenvalue condition in Assumption 2.3.*

Remark 2.3. *The width of bounds is influenced by $\overline{\kappa_\lambda}$ and $\overline{P_y^h}$. The term $\overline{\kappa_\lambda}$, which changes in response to the misclassification parameter λ, and $\overline{P_y^h}$, which varies with the sensitivity parameter h associated with conditional partial independence, impact the bounds differently. $\overline{\kappa_\lambda}$ has a lower bound, while $\overline{P_y^h}$ diminishes to zero as h approaches zero. Furthermore, since the derivatives of $\overline{\kappa_\lambda}$ and $\overline{P_y^h}$ with respect to λ and h are positive across their respective domains, $\overline{\kappa_\lambda}$ increases when $0 \leq \lambda < 1/2$, and $\overline{P_y^h}$ increases when $h \in [0, h_I)$. Setting $\lambda = 0$ eliminates measurement error, resulting in $X^* = X$, $\overline{\kappa_\lambda} = 1$, and point identification of the model, regardless of h's value. The deviation term in the Assumption 2.1 would reduce to $d_0(Y, X, Z) \equiv f_{YZ|X}(Y, Z|X) - f_{Y|X}(Y|X) f_{Z|X}(Z|X)$. However, even in this scenario, the bounds described in Proposition 2.1 become $\left(\max\left\{\rho_k - \overline{P_y^h}, 0\right\}, \min\left\{1, \rho_k + \overline{P_y^h}\right\}\right)$ for each k which vary with h and do not reduce to a singleton. This contrasts with the behavior of h, which can lead to a singleton bound if set to $h = 0$.*

Remark 2.4. *The analysis using partial identification could be more precise. The accuracy of the pivotal Equation (23) depends on the degree to which conditional independence is violated. Under Assumptions 2.3 and 2.4, the conditional density function $f_{Y|X^*}(y|x^* = k)$ corresponds to the k-th largest eigenvalue of the combined observed and perturbation matrices as described in the right-hand side of Equation (23). To refine the partial identification, we can analyze the spectrum of eigenvalues for $M_{XZ}^{-1} M_{yXZ} + P_y^h(d)$. Define the set of all possible deviation terms as $D_h = d_0(Y, X, Z, X^) : \left|d_0(Y, X, Z, X^)\right| \leq h, \int_y |d_0(y, X, Z, X^*)| dy \leq h$. The procedure for achieving a sharp partial identification involves: (i) Selecting a deviation d from D_h and calculating the* *eigenvalues of $M_{XZ}^{-1}M_{yXZ}+P_y^h(d)$ for this d. (ii) Iterating this process for each element in D_h. (iii) Aggregating the results by taking the union of all calculated eigenvalues across $d \in D_h$, which establishes sharp bounds for the density $f_{Y|X^*}(y|x^*)$. Consequently, the bounds depend on the set of all potential values from D_h and the matrix $M_{X|X^*}$. However, the outer identification sets specified in Equations (31) hinge solely on the parameters h and λ.*

There are two potential reasons why the bounds in Proposition 2.1 may not be sharp: (a) applying the Bauer-Fike theorem may not yield sharp results.[7] (b) the bounds $\overline{\kappa_\lambda}$ and $\overline{P_y^h}$ are estimated for the condition number $\kappa(M_{X|X^*}^{-1})$ and the norm $|||P_y^h(d)|||$ respectively.

Next, we present partial identification results for the conditional cumulative distribution function (CDF) of continuous or discrete outcomes given the latent variable. We focus on the continuous case here, with the discrete case addressed similarly. Let $F_{\widetilde{Y}|X^*}(\widetilde{y}|x^*) = \int_{-\infty}^{\widetilde{y}} f_{Y|X^*}(y|x^*)dy$. Integrating out Y in Equation (6) over the domain $(-\infty, \widetilde{y})$ yields

$$\sum_{x^*}\left(\int_{-\infty}^{\widetilde{y}} d_0(y,X,Z,x^*)dy\right) f_{X^*}\left(x^*\right) = \left(\int_{-\infty}^{\widetilde{y}} f_{YXZ}(Y,X,Z)dy\right) - \sum_{x^*}\left(\int_{-\infty}^{\widetilde{y}} f_{Y|X^*}\left(y|x^*\right)dy\right) f_{X|X^*}\left(X|X^*\right) f_{ZX^*}\left(Z,x^*\right). \tag{32}$$

Define $d_{F1}(\widetilde{y},X,Z) = \sum_{x^*}\left(\int_{-\infty}^{\widetilde{y}} d_0(y,X,Z,x^*)dy\right) f_{X^*}(x^*)$. Given $\widetilde{y}$, define the following matrices:

$$M_{dF}(\widetilde{y}) = \begin{bmatrix} d_{F1}(\widetilde{y},1,1) & \cdots & d_{F1}(\widetilde{y},K,1) \\ \vdots & & \\ d_{F1}(\widetilde{y},1,K) & \cdots & d_{F1}(\widetilde{y},K,K) \end{bmatrix}_{K\times K},$$

$$M_{F_{\widetilde{y}}XZ} = \begin{bmatrix} \int_{-\infty}^{\widetilde{y}} f_{YXZ}(y,1,1)dy & \cdots & \int_{-\infty}^{\widetilde{y}} f_{YXZ}(y,K,1)dy \\ \vdots & \cdots & \vdots \\ \int_{-\infty}^{\widetilde{y}} f_{YXZ}(y,1,K)dy & \cdots & \int_{-\infty}^{\widetilde{y}} f_{YXZ}(y,K,K)dy \end{bmatrix}_{K\times K},$$

$$D_{F_{\widetilde{y}|X^*}} = \begin{bmatrix} \int_{-\infty}^{\widetilde{y}} f_{Y|X^*}(y|x^*=1)dy & \cdots & 0 \\ \vdots & \vdots & 0 \\ 0 & 0 & \int_{-\infty}^{\widetilde{y}} f_{Y|X^*}(y|x^*=K)dy \end{bmatrix}_{K\times K}.$$

We have $|||M_{dF}(\widetilde{y})||| \le Kh$ because the (i,j) entry of the matrix $M_{dF}(\widetilde{y})$ is bounded through $|\sum_{x^*}\left(\int_{-\infty}^{\widetilde{y}} d_0(y,j,i,x^*)dy\right) f_{X^*}(x^*)| \le \sum_{x^*}\int_y |d_0(y,i,j,x^*)|dy f_{X^*}(x^*) \le h$ by Assumption 2.1. With

[7] Eigenvalues can be less sensitive to perturbations than the bound suggests, mainly when the perturbations are structured in a way that the theorem does not account for specifically. In practice, this means the actual change in eigenvalues due to perturbations might be significantly smaller than the upper bound predicted by Bauer-Fike.

these matrix notations, we can write Equation (32) as

$$M_{X^*Z} D_{F_{\tilde{y}}|X^*} M_{X|X^*} = M_{F_{\tilde{y}}XZ} - M_{dF}(\tilde{y}), \tag{33}$$

which is similar to the matrix equation in Equation (15). We also have the matrix equation (16). By applying the invertibility of $M_{XZ} - M_d$ and Equation (22) to the matrix relationship in Equation (33), we obtain the following eigenvalue-eigenvector decomposition:

$$\begin{aligned}\underbrace{M_{X|X^*}^{-1} D_{F_{\tilde{y}}|X^*} M_{X|X^*}}_{\text{a diagonal structure}} &= (M_{XZ} - M_d)^{-1} \left(M_{F_{\tilde{y}}XZ} - M_{dF}(\tilde{y})\right) \\ &= \underbrace{M_{XZ}^{-1} M_{F_{\tilde{y}}XZ}}_{\text{observable}} + P_{F_{\tilde{y}}}^h(d),\end{aligned} \tag{34}$$

where $P_{F_{\tilde{y}}}^h(d) = -M_{XZ}^{-1} M_{dF}(\tilde{y}) + B_d \left(M_{F_{\tilde{y}}XZ} - M_{dF}(\tilde{y})\right)$. Therefore, we obtain a matrix structure similar to Equation (23) and can apply the partial identification method in Proposition 2.1 using the Bauer-Fike perturbation result (Theorem 2.1).

Assumption 2.5. *(Distinct Eigenvalues) Assume $F_{\tilde{Y}|X^*}(\tilde{y}|x^* = i) \neq F_{\tilde{Y}|X^*}(\tilde{y}|x^* = j)$ for all $i \neq j$ and $\tilde{y}$.*

Let $dF_{\tilde{y}} = \min_{i,j} \frac{|F_{\tilde{Y}|X^*}(\tilde{y}|x^*=i) - F_{\tilde{Y}|X^*}(\tilde{y}|x^*=j)| - \epsilon}{2} > 0$ for some $\epsilon > 0$.

Proposition 2.2. *Under Assumptions 2.1, 2.2(i), 2.4, and 2.5, for given $\tilde{y}$, if a small perturbation $P_{F_{\tilde{y}}}^h(d)$ satisfying $|||P_{F_{\tilde{y}}}^h(d)||| \leq \frac{dF_{\tilde{y}}}{\overline{\kappa_\lambda}}$, then there exist distinct eigenvalues $\rho_{F1}, \ldots, \rho_{FK}$ of $M_{XZ}^{-1} M_{F_{\tilde{y}}XZ}$ satisfying: for each $k = 1, \ldots, K$,*

$$F_{\tilde{Y}|X^*}(\tilde{y}|x^* = k) \in \left(\max\left\{\rho_{Fk} - \overline{\kappa_\lambda} \cdot \overline{P_{F_{\tilde{y}}}^h}, 0\right\}, \min\left\{1, \rho_{Fk} + \overline{\kappa_\lambda} \cdot \overline{P_{F_{\tilde{y}}}^h}\right\}\right), \tag{35}$$

where $\overline{\kappa_\lambda} = \frac{1}{(1-\lambda)(1-2\lambda)}$ and $\overline{P_{F_{\tilde{y}}}^h} = |||M_{XZ}^{-1}|||Kh + \frac{|||M_{XZ}^{-1}|||^2 Kh}{1 - |||M_{XZ}^{-1}|||Kh} \left(|||M_{F_{\tilde{y}}XZ}||| + Kh\right)$ is an upper bound for $|||P_{F_{\tilde{y}}}^h(d)|||$.

Remark 2.5. *Given a user-specified function $G(\cdot)$, multiplying Equation (6) by $G(y)$ and summing over the support of Y yields*

$$\begin{aligned}&\sum_{x^*} \left(\int G(y) d_0(y, X, Z, x^*) dy\right) f_{X^*}\left(x^*\right) \\ &= \left(\int G(y) f_{YXZ}(Y, X, Z) dy\right) - \sum_{x^*} \left(\int G(y) f_{Y|X^*}\left(y|x^*\right) dy\right) f_{X|X^*}\left(X|X^*\right) f_{ZX^*}\left(Z, x^*\right).\end{aligned} \tag{36}$$

The above equation can be expressed using matrix notation similar to Equation (33). The matrix expression enables us to generalize Proposition 2.2 and investigate partial identification for $E(G(y)|x^*)$. By doing so, we can provide the bounds for $E(G(y)|x^*)$.

2.2. Partial Identification of Regression Functions

Within the model framework outlined in Subsection 2.1, the conditional mean outcome given X^* for the variables Y, X, Z, and X^* can be expressed as:

$$E(Y|X^*). \tag{37}$$

For point identification of non-linear regression models with misclassification, Mahajan (2006)[8] assumes the following condition:

Condition 2.2. *(Conditional Independence between X and Z) The variable (X,Z,X^*) satisfies a conditional independence assumption as follows*

$$f_{X,Z|X^*}\left(X,Z|X^*\right) = f_{X|X^*}\left(X|X^*\right) f_{Z|X^*}\left(Z|X^*\right), \tag{38}$$

or

$$X \perp Z \Big| X^* \tag{39}$$

We relax the abovementioned condition and consider the following assumption, similar to Assumption 2.1.

Assumption 2.6. *(Conditional Partial Independence Between X and Z) The perturbation or deviation from conditional independence in Equation (38) can be expressed as:*

$$\widetilde{d_0}(X,Z,X^*) \equiv f_{X,Z|X^*}\left(X,Z|X^*\right) - f_{X|X^*}\left(X|X^*\right) f_{Z|X^*}\left(Z|X^*\right). \tag{40}$$

There exists a constant $h \geq 0$ satisfying $\left|\widetilde{d_0}(X,Z,X^*)\right| \leq h$.

We can replicate the derivation employed for partially identifying the latent density in Subsection 2.1 to obtain partial identification of the conditional functions.

---

[8]Mahajan (2006) studies the identification and estimation in nonparametric regression models with a misclassified binary regressor, but in this subsection, we still focus on a general misclassified discrete regressor.

Assumption 2.7. *(Conditionally Mean Independent of the Outcome)*

$$E(Y|X,Z,X^*) = E(Y|X^*). \tag{41}$$

Assumption 2.8. *Assume that M_{XZ} is invertible and $h < \frac{1}{K|||M_{XZ}^{-1}|||}$.*

Assumption 2.9. *(Distinct Eigenvalues) Assume $E[Y|x^*=i] \neq E[Y|x^*=j]$ for all $i \neq j$.*

Define

$$M_{\bar{y}XZ} = \begin{bmatrix} E(Y|1,1)f_{XZ}(1,1) & \cdots & E(Y|K,1)f_{XZ}(K,1) \\ \vdots & \cdots & \vdots \\ E(Y|1,K)f_{XZ}(1,K) & \cdots & E(Y|K,K)f_{XZ}(K,K) \end{bmatrix}_{K\times K}. \tag{42}$$

Using the matrix notation, we can express the conditional partial independence in Assumption 2.6 similar to Equations (15) and (16) for Proposition 2.1. Consider

$$\begin{aligned} M_{X^*Z} D_{\bar{y}|X^*} M_{X|X^*} &= M_{\bar{y}XZ} - \widetilde{M}_d \\ M_{X^*Z} M_{X|X^*} &= M_{XZ} - \widetilde{M}_{d_2}, \end{aligned}$$

where the definition of the matrices here can be found in the proof of Proposition 2.3. Given this matrix expression, we can now utilize it to examine the following:

$$\begin{aligned} M_{X|X^*}^{-1} D_{\bar{y}|X^*} M_{X|X^*} &= \left(M_{XZ} - \widetilde{M}_{d_2}\right)^{-1}\left(M_{\bar{y}XZ} - \widetilde{M_d}\right) \\ &\equiv M_{XZ}^{-1} M_{\bar{y}XZ} + P_{\bar{y}}^h(d), \end{aligned}$$

where $\left(M_{XZ} - \widetilde{M}_{d_2}\right)^{-1} \equiv M_{XZ}^{-1} + \widetilde{B}_d$, and $P_{\bar{y}}^h(d) = -M_{XZ}^{-1}\widetilde{M_d} + \widetilde{B}_d\left(M_{\bar{y}XZ} - \widetilde{M_d}\right)$. This formulation displays an analogous matrix structure as Equation (23); consequently, the Bauer-Fike perturbation result (Theorem 2.1) can be exploited accordingly. Let $\widetilde{d} \equiv \min_{i,j} \frac{|E[Y|x^*=i]-E[Y|x^*=j]|-\epsilon}{2} > 0$ for some $\epsilon > 0$.

Proposition 2.3. *Under Assumptions 2.4, 2.6, 2.7, 2.8, and 2.9, if a small perturbation $P_{\bar{y}}^h(d)$ satisfying $|||P_{\bar{y}}^h(d)||| \leq \frac{\widetilde{d}}{\overline{\kappa_\lambda}}$, then there exist distinct eigenvalues $\widetilde{\rho}_1,\ldots,\widetilde{\rho}_K$ of the matrix $M_{XZ}^{-1}M_{\bar{y}XZ}$ satisfying: for each $k = 1,\ldots,K$,*

$$E\left[Y|x^*=k\right] \in \left(\widetilde{\rho}_k - \overline{\kappa_\lambda}\cdot\overline{P_{\bar{y}}^h}, \widetilde{\rho}_k + \overline{\kappa_\lambda}\cdot\overline{P_{\bar{y}}^h}\right), \tag{43}$$

where $\overline{\kappa_\lambda} = \frac{1}{(1-\lambda)(1-2\lambda)}$ and $\overline{P_{\bar{y}}^h} = |||M_{XZ}^{-1}|||Kh + \frac{|||M_{XZ}^{-1}|||^2 Kh}{1-|||M_{XZ}^{-1}|||Kh}\left(|||M_{\bar{y}XZ}||| + K^2h|E(Y)|\right)$ is an upper *bound for* $|||P^h_{\overline{y}}(d)|||$.

3. Estimation

This paper focuses on partial identification. We briefly discuss estimating the lower and upper bounds of the identification regions in Section 2. These regions have closed-form solutions and center at distinct eigenvalues of a combination of observed density matrices. Their dispersion depends on the upper bound of misclassification probabilities, λ, and a sensitivity parameter, h. We will describe an estimation method to calculate the identification regions in Proposition 2.1. An analogous approach prevails for additional identification regions in Section 2.

For a particular y, Proposition 2.1 provides the following identification regions:

$$\left(\max\left\{\rho_k - \overline{\kappa_\lambda}\cdot\overline{P^h_y},0\right\},\min\left\{1,\rho_k+\overline{\kappa_\lambda}\cdot\overline{P^h_y}\right\}\right) \text{ for } k=1,\dots,K. \tag{44}$$

The lower and upper bounds of the identification regions are known functionals of the terms, ρ_k, and $\overline{\kappa_\lambda}\cdot\overline{P^h_y}$. We propose nonparametric estimators for these two terms and then derive their $\sqrt{N}$-consistency and asymptotic results. We then apply a delta method for directionally differentiable functionals in Fang and Santos (2019) to provide a bootstrap procedure to construct asymptotic confidence bands for the bounds.

The lower and upper bounds constitute a specified function of the matrices of the joint distributions, M_{yXZ} and M_{XZ}, with a defined prior λ and h. Define $\gamma_0 = \left(\text{vec}(M_{yXZ})^T,\text{vec}(M_{XZ})^T\right)^T \in \mathbb{R}^{2K^2}$ for a given y, where vec(A) signifies the vector encompassing the entries of matrix A in a column vector. As delineated in Andrew et al. (1993), the distinct eigenvalues of $M^{-1}_{XZ}M_{yXZ}$, $\rho_1,\dots,\rho_K$, can be explicitly denoted as $\rho_k=\phi^e_y\left(x^*=k,\gamma_0\right)$ for $k=1,\dots,K$, where $\phi^e_y(\cdot,\cdot)$ constitutes a specified well-behaved analytic function. Subsequently, denote $\phi^d_y(\lambda,h,\gamma_0)=\overline{\kappa_\lambda}\cdot\overline{P^h_y}$. Since the matrix norms in $\overline{P^h_y}$ are applied to non-negative densities, $\phi^d_y(\lambda,h,\gamma_0)$ is well-defined for sufficiently small h. Let $\phi^L_y\left(x^*,\gamma_0;\lambda,h\right)\equiv\phi^e_y\left(x^*,\gamma_0\right)-\phi^d_y(\lambda,h,\gamma_0)$ and $\phi^U_y\left(x^*,\gamma_0;\lambda,h\right)\equiv\phi^e_y\left(x^*,\gamma_0\right)+\phi^d_y(\lambda,h,\gamma_0)$. Then, for each $k=1,\dots,K$, the lower and upper bounds of the identification region can be written as follows:

$$\max\left\{\rho_k-\overline{\kappa_\lambda}\cdot\overline{P^h_y},0\right\}=\max\left\{\phi^L_y\left(x^*=k,\gamma_0;\lambda,h\right),0\right\}=\phi_{\max}\left(\phi^L_y\left(x^*=k,\gamma_0;\lambda,h\right)\right), \tag{45}$$

$$\min\left\{1,\rho_k+\overline{\kappa_\lambda}\cdot\overline{P^h_y}\right\}=\min\left\{1,\phi^U_y\left(x^*=k,\gamma_0;\lambda,h\right)\right\}=\phi_{\min}\left(\phi^U_y\left(x^*=k,\gamma_0;\lambda,h\right)\right), \tag{46}$$

where $\phi_{\max}(\theta)=\max\{\theta,0\}$, and $\phi_{\min}(\theta)=\min\{1,\theta\}$. The lower and upper bounds constitute compositions of the max and min operators, $\phi_{\max}$ and $\phi_{\min}$, in conjunction with ϕ_y^L and ϕ_y^U respectively. The bounds do not constitute typical Hadamard differentiable mappings of γ_0 because the min and max operators are not Hadamard differentiable.

We propose nonparametric plug-in estimators to estimate these bounds. We first employ a nonparametric technique to evaluate the observed densities f_{YXZ} and f_{XZ} and concentrate on the case where the variable Y is discrete. Suppose that $\mathbf{1}(\cdot)$ signifies an indicator function and the function $K(\cdot)$ constitutes a defined kernel function with bandwidth h_b. Given an observed i.i.d. sample $\{y_i, x_i, z_i\}_{i=1}^n$, construct the following density estimators:

$$\widehat{f}_{YXZ}(y,x,z) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(y_i = y)\mathbf{1}(x_i = x)\mathbf{1}(z_i = z), \tag{47}$$

$$\widehat{f}_{XZ}(x,z) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(x_i = x)\mathbf{1}(z_i = z). \tag{48}$$

Let $v = (y,x,z)$. For a given y, the infinite-dimensional parameter γ_0 can be estimated nonparametrically applying $\widehat{\gamma}_n(v) = \left(\text{vec}(\widehat{M}_{yXZ})^T, \text{vec}(\widehat{M}_{XZ})^T\right)^T$, where $\widehat{M}_{yXZ}$ and $\widehat{M}_{XZ}$ constitute estimators for M_{yXZ} and M_{XZ} respectively utilizing the nonparametric density estimators $\widehat{f}_{YXZ}$ and $\widehat{f}_{XZ}$ for their entries. Therefore, we assess these bounds by plug-in estimators $\phi_{\max}\big(\phi_y^L(x^*, \widehat{\gamma}_n; \lambda, h)\big)$ and $\phi_{\min}\big(\phi_y^U(x^*, \widehat{\gamma}_n; \lambda, h)\big)$. In the remainder of this section, we will be examining the asymptotic properties of $\phi_{\max}\big(\phi_y^L(x^*, \widehat{\gamma}_n; \lambda, h)\big)$. It is worth noting that the same approach applies to $\phi_{\min}\big(\phi_y^U(x^*, \widehat{\gamma}_n; \lambda, h)\big)$ as well. We will first derive the limiting distribution of $\phi_y^L(x^*, \widehat{\gamma}_n; \lambda, h)$ and then implement the functional delta method in Fang and Santos (2019) for the max operator $\phi_{\max}$ to show convergence in distribution of our estimator $\phi_{\max}\big(\phi_y^L(x^*, \widehat{\gamma}_n; \lambda, h)\big)$.

Each element of $\widehat{\gamma}_n$ constitutes a nonparametric frequency-based estimator at some point. We can directly apply the outcomes of the asymptotic normality of the nonparametric frequency density estimator in Li and Racine (2003) and Li and Racine (2007) to reach the asymptotic normality of the element of $\widehat{\gamma}_n$ in a pointwise sense.

Lemma 3.1. *Let γ_{0q} be the $q-$th element of γ_0 and $\widehat{\gamma}_{nq}$ is the corresponding estimator in $\widehat{\gamma}_n$. Suppose the density γ_{0q} has three times bounded continuous derivatives. Given an observed i.i.d. sample $\{y_i, x_i, z_i\}_{i=1}^n$ with the support of v is finite, we have*

$$\sqrt{n}(\widehat{\gamma}_{nq}(\tilde{v}) - \gamma_{0q}(\tilde{v})) \xrightarrow{d} N(0, \gamma_{0q}(\tilde{v})(1 - \gamma_{0q}(\tilde{v}))), \tag{49}$$

where $\tilde{v}$ is an interior point of density.

Assumption 3.1. *For any vector $c \in \mathbb{R}^{2K^2}$, the fixed linear combination of the coordinates of $\widehat{\gamma}_n(v) - \gamma_0(v)$ converges to a normal distribution, i.e., $\sqrt{n}c'(\widehat{\gamma}_n(v) - \gamma_0(v)) \xrightarrow{d} N(0, c'\Omega c)$, where $\Omega = n \lim_{n\to\infty} E\left[(\widehat{\gamma}_n - E[\widehat{\gamma}_n])(\widehat{\gamma}_n - E[\widehat{\gamma}_n])^T\right]$.*

Under Assumption 3.1, we can utilize the Cramer-Wold theorem to deduce the subsequent asymptotic distribution of $\widehat{\gamma}_n(v) - \gamma_0(v)$.

Lemma 3.2. *Under Assumption 3.1 and the conditions in Lemma 3.1, we have*

$$\sqrt{n}(\widehat{\gamma}_n(v) - \gamma_0(v)) \xrightarrow{d} N(0, \Omega), \tag{50}$$

where $\Omega = n \lim_{n\to\infty} E\left[(\widehat{\gamma}_n - E[\widehat{\gamma}_n])(\widehat{\gamma}_n - E[\widehat{\gamma}_n])^T\right]$.

Next, we define the pathwise derivative of $\phi_y^L(x^*, \gamma; \lambda, h)$ at the direction $[\gamma - \gamma_0]$ evaluated at γ_0, and this can be defined as

$$\frac{d\phi_y^L(x^*, \gamma_0; \lambda, h)}{d\gamma}[\gamma - \gamma_0] \equiv \left.\frac{d\phi_y^L(x^*, \gamma_0 + \tau[\gamma - \gamma_0]; \lambda, h)}{d\tau}\right|_{\tau=0} \quad \text{a.s.}$$

The pathwise derivative can be denoted as the ordinary derivative through

$$\frac{d\phi_y^L(x^*, \gamma_0 + \tau[\gamma - \gamma_0]; \lambda, h)}{d\tau} = \frac{\partial\phi_y^L(x^*, \gamma_0 + \tau[\gamma - \gamma_0]; \lambda, h)}{\partial\gamma^T} \times (\gamma - \gamma_0).$$

Applying the delta method to the asymptotic outcome in Lemma 3.2, we attain an asymptotic normality for $\phi_y^L(x^*, \widehat{\gamma}_n; \lambda, h)$.

Proposition 3.1. *Under Assumption 3.1 and the prerequisites in Lemma 3.1, we have*

$$\sqrt{n}\left(\phi_y^L(x^*, \widehat{\gamma}_n; \lambda, h) - \phi_y^L(x^*, \gamma_0; \lambda, h)\right) \xrightarrow{d} N(0, \widetilde{\Omega}), \tag{51}$$

where $\widetilde{\Omega} = E\left[\left(\frac{\partial\phi_y^L(x^*, \gamma_0; \lambda, h)}{\partial\gamma^T}\right)\Omega\left(\frac{\partial\phi_y^L(x^*, \gamma_0; \lambda, h)}{\partial\gamma^T}\right)^T\right]$.

We can regard the lower bound estimator $\phi_{\max}(\phi_y^L(x^*, \widehat{\gamma}_n; \lambda, h))$ as a Hadamard directional differentiability of the mapping linking the term $\phi_y^L(x^*, \widehat{\gamma}_n; \lambda, h)$. As discussed in Fang and Santos (2019), the limiting distribution of $\phi_{\max}(\phi_y^L(x^*, \widehat{\gamma}_n; \lambda, h))$ depends only on two key conditions: (i) The directional derivative $\phi_{\max}$, and (ii) The limiting distribution of $\phi_y^L(x^*, \widehat{\gamma}_n; \lambda, h)$. The map $\phi_{\max}$ is Hadamard directionally differentiable at θ_0 by the existence of the directional derivative of $\phi_{\max}$ as

(52)
$$\phi'_{\max;\theta_0}(h) = \begin{cases} h & \text{if } \theta_0 > 0 \\ \max\{h, 0\} & \text{if } \theta_0 = 0 \\ 0 & \text{if } \theta_0 < 0 \end{cases}$$

for any $h \in \mathbb{R}$. On the other hand, the asymptotic distribution of $\phi_y^L(x^*, \widehat{\gamma}_n; \lambda, h)$ is presented in Proposition 3.1. Therefore, Assumptions 1 and 2 in Fang and Santos (2019) hold and we can apply the functional delta method for Hadamard directionally differentiable mappings in Theorem 2.1 in Fang and Santos (2019) to show convergence in distribution of the lower bound estimator.

Proposition 3.2. *Under Assumption 3.1 and the conditions in Lemma 3.1, we obtain*

(53)
$$\sqrt{n}\Big(\phi_{\max}\big(\phi_y^L(x^*, \widehat{\gamma}_n; \lambda, h)\big) - \phi_{\max}\big(\phi_y^L(x^*, \gamma_0; \lambda, h)\big)\Big) \xrightarrow{d} \phi'_{\max;\theta_0}(N(0, \widetilde{\Omega}))$$

, where

(54)
$$\phi'_{\max;\theta_0}(N(0, \widetilde{\Omega})) = \begin{cases} N(0, \widetilde{\Omega}) & \textit{if } \phi_y^L(x^*, \gamma_0; \lambda, h) > 0 \\ \max\{N(0, \widetilde{\Omega}), 0\} & \textit{if } \phi_y^L(x^*, \gamma_0; \lambda, h) = 0 \\ 0 & \textit{if } \phi_y^L(x^*, \gamma_0; \lambda, h) < 0 \end{cases}.$$

Although we obtain the asymptotic distribution of the lower bound, we do not utilize this distribution for making inferences because it would be difficult to obtain analytical asymptotic confidence intervals. We propose a bootstrap procedure to make inferences on the estimated bound. This approach is straightforward to approximate the asymptotic distribution. Let $\widehat{\gamma}_n^*$ denote a bootstrap estimate drawn from the nonparametric bootstrap distribution of $\widehat{\gamma}_n$. Since $\phi_{\max}(\phi_y^L(x^*, \gamma; \lambda, h)$ is Hadamard differentiable, by Theorem 3.1 of Fang and Santos (2019), the limiting distribution of $\sqrt{n}\Big(\phi_y^L(x^*, \widehat{\gamma}_n; \lambda, h) - \phi_y^L(x^*, \gamma_0; \lambda, h)\Big)$ can be consistently estimated by the nonparametric bootstrap

(55)
$$\sqrt{n}\Big(\phi_y^L(x^*, \widehat{\gamma}_n^*; \lambda, h) - \phi_y^L(x^*, \widehat{\gamma}_n; \lambda, h)\Big).$$

Then, by Theorem 3.2 of Fang and Santos (2019), we employ

(56)
$$\phi'_{\max;\theta_0}(\sqrt{n}\Big(\phi_y^L(x^*, \widehat{\gamma}_n^*; \lambda, h) - \phi_y^L(x^*, \widehat{\gamma}_n; \lambda, h)\Big))$$

to estimate the distribution of $\phi'_{\max;\theta_0}(N(0,\widetilde{\Omega}))$ in Proposition 3.2.

4. Systematic Sensitivity Analysis

To empirically examine the partial failure of conditional independence, we must ascertain an appropriate upper bound for h and evaluate the value utilizing empirical data. In this section, we provide a detailed delineation of an algorithm that summarizes the steps in calculating an upper bound for the sensitivity parameter h whose identification region would encompass the true parameters of interest, including the true density or the regression function in Section 2. Although the width of the identification regions is indexed by λ and h, the upper bound on the probabilities of misreport λ often has constraints from many possible sources such as validation studies, economic theory, and social behaviors, etc. Molinari (2008) assumes that the researcher has a known λ and regards it as baseline information. We adopt a similar approach regarding λ as an input furnished by validation studies to calculate identification regions.

For a minimal set of premises, the derivation of the upper bound for the sensitivity parameter comprises two stages. The first step involves imposing the identification assumptions on the sensitivity parameter h, and the constraints arising from probability theory. In the second step, we follow the approach of Masten and Poirier (2020) to execute breakdown frontier analysis. The breakdown frontier analysis necessitates specifying additional desired conclusions. For simplicity, we only contemplate one additional conclusion and derive the weakest combinations of the identification assumptions indexed by λ and h, which conduce to the conclusion. The breakdown frontier is a curve of λ and h, signifying the weakest combinations. After solving for h in the weakest combinations, we arrive at an expression for the breakdown frontier in terms of λ. Since λ is available from validation studies, this does impose some constraints on the sensitivity parameter h. We find that the range of the sensitivity parameter is the intersection of the two ranges from the first step and the second step, and the upper bound for the range is determined by taking the minimum of the two upper bounds from these two steps.

4.1. Breakdown Frontier Analysis

This section comprises a breakdown frontier analysis for the identification sets of the conditional CDF in Proposition 2.2 and the regression functions in Proposition 2.3. In both cases, we relax the conditional independence assumptions and deduce the identification sets as a function of the parameter λ and the sensitivity parameter h or h. Thus, there may exist trade-offs between the upper bound of the misclassification probabilities and relaxations of the conditional independence assumptions in drawing a particular conclusion.

For a fixed $\widetilde{y}_m$ in the support of Y and a fixed $\underline{p} \in [0,1]$, we consider the conclusion that the proportion of units whose outcomes greater than $\widetilde{y}_m$ is at least $\underline{p}$,

$$P(y \geq \widetilde{y}_m | x^*) \geq \underline{p}. \tag{57}$$

Given observed data, we investigate the weakest assumptions that allow us to obtain the conclusion. Applying the relation $P(y \geq \widetilde{y}_m|x^*) = 1 - P(y < \widetilde{y}_m|x^*) = 1 - F_{\widetilde{Y}|X^*}(\widetilde{y}_m|x^*)$ to the partial identification result in Proposition 2.2, we obtain the following bounds

$$P(y \geq \widetilde{y}_m | x^* = k) \in \left(1 - \min\left\{1, \rho_{Fk} + \overline{\kappa_\lambda} \cdot \overline{P^h_{F_{\widetilde{y}_m}}}\right\}, 1 - \max\left\{\rho_{Fk} - \overline{\kappa_\lambda} \cdot \overline{P^h_{F_{\widetilde{y}_m}}}, 0\right\}\right), \tag{58}$$

where ρ_{Fk} is an eigenvalue of $M^{-1}_{XZ} M_{F_{\widetilde{y}_m} XZ}$. When $\rho_{Fk} + \overline{\kappa_\lambda} \cdot \overline{P^h_{F_{\widetilde{y}_m}}} \geq 1$, the lower bound of the above outer identified set is not informative, and we just set it equal to 0 in this case. Thus, we consider $\rho_{Fk} + \overline{\kappa_\lambda} \cdot \overline{P^h_{F_{\widetilde{y}_m}}} < 1$ and the requirement related to the conclusion is

$$1 - \rho_{Fk} - \overline{\kappa_\lambda} \cdot \overline{P^h_{F_{\widetilde{y}_m}}} \geq \underline{p}. \tag{59}$$

Denote $\widetilde{I} = [0, h_I)$, where $h_I = \frac{1}{K|||M^{-1}_{XZ}|||}$. The robust region for the conclusion is

$$RR_{\widetilde{y}_m}(x^* = k, \underline{p}) = \left\{(\lambda, h) \in [0, 1/2) \times \widetilde{I} : 1 - \rho_{Fk} - \underline{p} \geq \overline{\kappa_\lambda} \cdot \overline{P^h_{F_{\widetilde{y}_m}}}\right\}.$$

The breakdown frontier is defined as the boundary of the robust region, and then the breakdown frontier for the conclusion is

$$BF_{\widetilde{y}_m}(x^* = k, \underline{p}) = \left\{(\lambda, h) \in [0, 1/2) \times \widetilde{I} : 1 - \rho_{Fk} - \underline{p} = \overline{\kappa_\lambda} \cdot \overline{P^h_{F_{\widetilde{y}_m}}}\right\}.$$

In this case, the breakdown frontier is the locus of various points showing different combinations of (λ, h) providing an equal value for a fixed $\underline{p}$. Since $\overline{\kappa_\lambda}$ is increasing with λ and $\overline{P^h_{F_{\widetilde{y}_m}}}$ is also increasing with h, this implies that any relaxation of the upper bounds on the misreporting probability requires strengthening the conditional independence assumption in order to maintain our specific conclusion.

Denote $\kappa(x^* = k, \underline{p}, \lambda) = \frac{1 - \rho_{Fk} - p}{\overline{\kappa_\lambda}}$, a positive variable does not depend on h. Solving for h in the inequality for the robust region $1 - \rho_{Fk} - \underline{p} \geq \overline{\kappa_\lambda} \cdot \overline{P^h_{F_{\widetilde{y}_m}}}$ with the definition of $\overline{P^h_{F_{\widetilde{y}_m}}}$ in

Proposition 2.2 yields

$$\frac{\kappa(x^* = k, \underline{p}, \lambda)}{K|||M_{XZ}^{-1}|||\left(1+\kappa(x^* = k, \underline{p}, \lambda) + |||M_{XZ}^{-1}||| \cdot |||M_{F_{\tilde{y}_m}XZ}|||\right)} \geq h. \tag{60}$$

For $0 \leq \lambda < \frac{1}{2}$, the analytical expression for the breakdown frontier:

$$bf_{\tilde{y}_m}(\lambda; x^*, \underline{p}) = \frac{\kappa(x^*, \underline{p}, \lambda)}{K|||M_{XZ}^{-1}|||\left(1+\kappa(x^*, \underline{p}, \lambda) + |||M_{XZ}^{-1}||| \cdot |||M_{F_{\tilde{y}_m}XZ}|||\right)}. \tag{61}$$

Given $(\tilde{y}_m, x^*, \underline{p})$, $bf_{\tilde{y}_m}(\lambda; x^*, \underline{p})$ is a continuously differentiable function of λ for $0 \leq \lambda < \frac{1}{2}$. With the expression in Equation (61), we obtain

$$BF_{\tilde{y}_m}(x^*, \underline{p}) = \left\{ \left(\lambda, bf_{\tilde{y}_m}(\lambda; x^*, \underline{p})\right) : \lambda \in [0, 1/2) \right\}. \tag{62}$$

This curve provides the weakest combinations of the two types of assumptions which lead to the considered conclusion. The shape of the curve allows us to quantify the trade-off between the upper bound of the misclassification probabilities and the conditional independence assumption relaxations.

Next consider the estimation of the breakdown frontier $bf_{\tilde{y}_m}(\lambda; x^* = k, \underline{p})$ constituted of ρ_{Fk}, M_{XZ} and $M_{F_{\tilde{y}_m}XZ}$. For a given $\tilde{y}_m$, define $\gamma_{F0} = \left(\text{vec}(M_{F_{\tilde{y}_m}XZ})^T, \text{vec}(M_{XZ})^T\right)^T \in \mathbb{R}^{2K^2}$. Similar to the discussion in Section 3, for each $k = 1, ..., K$, ρ_{Fk}, can be explicit expressed as $\rho_{Fk} = \phi^e_{\tilde{y}_m}\left(x^* = k, \gamma_{F0}\right)$, where $\phi^e_{\tilde{y}_m}(\cdot, \cdot)$ is a known well-behaved analytic function. Therefore, the breakdown frontier in Equation (61) can be expressed as

$$bf_{\tilde{y}_m}(\lambda; x^*, \underline{p}) = \phi^{bf}_{\tilde{y}_m}(\lambda, \gamma_{F0}; x^*, \underline{p}) = q_d(q_\kappa(\phi^e_{\tilde{y}_m}\left(x^*, \gamma_{F0}\right), \lambda; \underline{p}), \gamma_{F0}), \tag{63}$$

where $q_d(\zeta, \gamma_{F0}) = \frac{\zeta}{K|||M_{XZ}^{-1}|||\left(1+\zeta+|||M_{XZ}^{-1}|||\cdot|||M_{F_{\tilde{y}_m}XZ}|||\right)}$ and $q_\kappa(\rho, \lambda; \underline{p}) = \frac{1-\rho-\underline{p}}{\kappa_\lambda}$. This implies that the breakdown frontier is the functional $\phi^{bf}_{\tilde{y}_m}(\lambda, \gamma_F; x^*, \underline{p})$ evaluated at γ_{F0} and we estimate it by a plug-in estimator $\phi^{bf}_{\tilde{y}_m}(\lambda, \widehat{\gamma}_F; x^*, \underline{p})$ with an estimator $\widehat{\gamma}_F$ for γ_{F0}. We use the following nonparametric density estimator for an element of $M_{F_{\tilde{y}_m}XZ}$,

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(y_i \leq \tilde{y}_m)\mathbf{1}(x_i = x)\mathbf{1}(z_i = z),$$

and denote $\widehat{M}_{F_{\tilde{y}_m}XZ}$ as an estimator for $M_{F_{\tilde{y}_m}XZ}$ using the nonparametric estimator. For a given $\tilde{y}_m$, we use $\widehat{\gamma}_{Fn} = \left(\text{vec}(\widehat{M}_{F_{\tilde{y}_m}XZ})^T, \text{vec}(\widehat{M}_{XZ})^T\right)^T$ as an estimator for γ_{F0}. Similar to the asymptotic normality result in Lemma 3.2, we can obtain a pointwise asymptotic normality result for $\widehat{\gamma}_{Fn}$ converging to γ_{F0}. Since it is straightforward to see the functional $\phi^{bf}_{\widetilde{y}_m}(\lambda, \gamma_F; x^*, \underline{p})$ is continuous pathwise differential in a neighborhood of γ_{F0}, the pointwise asymptotic normality result of $\widehat{\gamma}_{Fn}$ will carry over to the functional by the delta method. The limiting distribution of the plug-in estimator $\phi^{bf}_{\widetilde{y}_m}(\lambda, \widehat{\gamma}_F; x^*, \underline{p})$ has a Gaussian limiting process as Proposition 3.1.

As for the breakdown frontier analysis for the identification sets of the regression functions in Proposition 2.3, we consider a specific conclusion that $E[Y|x^* = k] \geq \underline{\mu}$ for a given value of $x^* = k$ and for some $\underline{\mu} \in \mathbb{R}$. When the conditional independence between X and Z holds ($h = 0$) with an upper bound on the misreporting probability ($0 \leq \lambda < \frac{1}{2}$), the identification set in Proposition 2.3 collapses to a single point in which we reach point identification. The robust region of this conclusion is a collection of all values of λ and h such that the lower bound of the outer identified set $E[Y|x^* = k]$ exceeds $\underline{\mu}$. That is

$$
\begin{aligned}
RR(x^* = k, \underline{\mu}) &= \left\{(\lambda, h) \in [0, 1/2) \times \widetilde{I} : \widetilde{\rho}_k - \overline{\kappa_\lambda} \cdot \overline{P^h_{\overline{y}}} \geq \underline{\mu}\right\} \\
&= \left\{(\lambda, h) \in [0, 1/2) \times \widetilde{I} : \widetilde{\rho}_k - \underline{\mu} \geq \overline{\kappa_\lambda} \cdot \overline{P^h_{\overline{y}}}\right\}.
\end{aligned}
$$

Since $\overline{\kappa_\lambda}$ and $\overline{P^h_{\overline{y}}}$ are both positive, the robust region will be empty if $\widetilde{\rho}_k < \underline{\mu}$, and nonempty if $\widetilde{\rho}_k > \underline{\mu}$, where $\widetilde{\rho}_k$ is a value of $E[Y|x^* = k]$ under the point identifying assumptions. Therefore, we only focus on $\widetilde{\rho}_k > \underline{\mu}$. The breakdown frontier for the conclusion that $E[Y|x^* = k] \geq \underline{\mu}$ is

$$
BF(x^* = k, \underline{\mu}) = \left\{(\lambda, h) \in [0, 1/2) \times \widetilde{I} : \widetilde{\rho}_k - \underline{\mu} = \overline{\kappa_\lambda} \cdot \overline{P^h_{\overline{y}}}\right\}.
$$

Denote $\kappa(x^* = k, \underline{\mu}, \lambda) = \frac{\widetilde{\rho}_k - \underline{\mu}}{\overline{\kappa_\lambda}}$. Similar to the previous derivation for $BF_{\widetilde{y}_m}(x^*, \underline{p})$ with the expression in Equations (62), we have

$$
BF(x^*, \underline{\mu}) = \left\{\left(\lambda, bf(\lambda; x^*, \underline{\mu})\right) : \lambda \in [0, 1/2)\right\},
$$

with the breakdown frontier

$$
bf(\lambda; x^*, \underline{\mu}) = \frac{\kappa(x^*, \underline{\mu}, \lambda)}{K |||M^{-1}_{XZ}||| \left(1 + \kappa(x^*, \underline{\mu}, \lambda) + |||M^{-1}_{XZ}||| \cdot |||M_{\overline{y}XZ}|||\right)}. \tag{64}
$$

We can estimate the breakdown frontier $bf(\lambda; x^*, \underline{\mu})$ using a plug-in estimator as we do for the breakdown frontier $bf_{\widetilde{y}_m}(\lambda; x^* = k, \underline{p})$ for the conditional CDF case in Proposition 2.2 and the estimator may possess a similar pointwise asymptotic normality result.

4.2. Two Step Selection Procedure

This section provides a two-step procedure to select possible values for the upper bound of the sensitivity parameter h. Since validation studies can provide a potential value on the upper probability of misreporting, we can assume that λ is a known value. For simplicity, we focus on the sensitivity analysis for the identification sets of the conditional probability $f_{Y|X^*}(y|x^*)$ in Proposition 2.1.

The First Step:

To construct bounds empirically, given an upper bound of probabilities of misreporting λ_v we recommend using a minimum value of three restrictions on the sensitivity parameter in the derivation of Proposition 2.1, a restriction from obtaining the inverse of a perturbed matrix $h_I = \frac{1}{K|||M_{XZ}^{-1}|||}$, a restriction from the distinct eigenvalue condition $h_E = \max\left\{h : \overline{P_y^h} \leq \frac{\min_{i,j}\left|\rho_i - \rho_j\right|}{2\overline{\kappa_{\lambda_v}(S)}}\right\}$, and a restriction from a density restriction $h_{Dx^*} \equiv \max\left\{h : \overline{P_y^h} \leq \frac{\rho_{x^*}}{\overline{\kappa_{\lambda_v}}}, \text{ and } \overline{P_y^h} \leq \frac{1-\rho_{x^*}}{\overline{\kappa_{\lambda_v}}}\right\}$. We have considered three measures of h, h_I, h_E, and h_{Dk} and the first step selection for the upper bound of h is $h_{R1x^*} \equiv \min\left\{h_I, h_E, h_{Dx^*}\right\}$ for all x^*.

The Second Step:

As for the breakdown frontier analysis for the identification sets of the probability in Proposition 2.1, we consider a specific conclusion that $f_{Y|X^*}(y|x^*) \geq \underline{p}$ for a given value of (y, x^*) and for some $\underline{p} \in (0,1)$. Assume that $\rho_k - \underline{p} > 0$. The breakdown frontier for the conclusion is

$$BF_y(x^* = k, \underline{p}) = \left\{(\lambda, h) \in [0, 1/2) \times \widetilde{I} : \rho_k - \underline{p} = \overline{\kappa_\lambda} \cdot \overline{P_y^h}\right\}.$$

Denote $\kappa(x^* = k, \underline{p}, \lambda) = \frac{\rho_k - \underline{p}}{\overline{\kappa_\lambda}}$. Similar to the previous breakdown frontier analysis for the identification sets of the regression functions in Proposition 2.3, we obtain

$$BF_y(x^*, \underline{p}) = \left\{\left(\lambda, bf_y(\lambda; x^*, \underline{p})\right) : \lambda \in [0, 1/2)\right\},$$

with the breakdown frontier

$$bf_y(\lambda; x^*, \underline{p}) = \frac{\kappa(x^*, \underline{p}, \lambda)}{K|||M_{XZ}^{-1}|||\left(1 + \kappa(x^*, \underline{p}, \lambda) + |||M_{XZ}^{-1}||| \cdot |||M_{yXZ}|||\right)}. \tag{65}$$

Next, we consider the breakdown frontier $bf_y(\lambda; x^*, \underline{p})$ at λ_v and construct its one-sided lower confidence interval by a bootstrap method. Denote $\phi_y^{bf}(\lambda, \gamma_0; x^*, \underline{p}) = bf_y(\lambda; x^*, \underline{p})$, where $\gamma_0 =$

$\left(\text{vec}(M_{yXZ})^T,\text{vec}(M_{XZ})^T\right)^T \in \mathbb{R}^{2K^2}$. Suppose $\widehat{\gamma}_n(v) = \left(\text{vec}(\widehat{M}_{yXZ})^T,\text{vec}(\widehat{M}_{XZ})^T\right)^T$ is an nonparametric estimator for γ_0 by the sample $\{y_i,x_i,z_i\}_{i=1}^n$. The plug-in estimator $\phi_y^{bf}(\lambda_v,\widehat{\gamma}_n;x^*,\underline{p})$ is an estimator for the breakdown frontier $bf_y(\lambda;x^*,\underline{p})$ at λ_v.

As for the bootstrap one-sided lower confidence interval, first we generate n observations randomly with replacement from $\{y_i,x_i,z_i\}_{i=1}^n$ to obtain a bootstrap data set $\{y_i^*,x_i^*,z_i^*\}_{i=1}^n$, and then calculate bootstrap estimates $\widehat{\gamma}_{nj}^*$ from the resample $\{y_i^*,x_i^*,z_i^*\}_{i=1}^n$ for $j = 1,...,B$, where B is the number of bootstrap samples. Next, we compute $\phi_y^{bf}(\lambda_v,\widehat{\gamma}_{nj}^*;x^*,\underline{p})$ for $j = 1,...,B$. For these bootstrap estimators, we sort them into order and use the $100\alpha_s$th percentile of $\phi_y^{bf}(\lambda_v,\widehat{\gamma}_{nj}^*;x^*,\underline{p})$ as a lower bound of the $1-\alpha_s$ confidence interval and denote it as $Lb_y(\lambda_v;x^*,\underline{p})$. Thus, the one-sided lower confidence interval of $\phi_y^{bf}(\lambda_v,\widehat{\gamma}_n;x^*,\underline{p})$ with $1-\alpha_s$ coverage is

$$\left(Lb_y(\lambda_v;x^*,\underline{p}),\infty\right). \tag{66}$$

When the sensitivity parameter h is lower than the lower bound $Lb_y(\lambda_v;x^*,\underline{p})$, this implies approximately $100(1-\alpha_s)\%$ of the time, the assumptions under the pair (λ_v,h) leads us to draw the conclusion that $f_{Y|X^*}(y|x^*=k) \geq \underline{p}$. Hence $Lb_y(\lambda_v;x^*,\underline{p})$ is a measure of the robustness of our conclusion to failure of the point identifying assumptions with λ_v. The second step selection for the upper bound of h is $h_{R2x^*} = Lb_y(\lambda_v;x^*,\underline{p})$.

When the identification assumptions are maintained, we have used h_{R1x^*} in the first step selection as the upper bound of the sensitivity parameter. To incorporate the implications of the additional conclusion for the identification regions, we consider the case that the upper bound of the sensitivity parameter h is smaller than the lower bound $Lb_y(\lambda_v;x^*,\underline{p})$ and h_{R1x^*}. Then, the upper bound of the sensitivity parameter is the minimum of the two upper bounds from these two steps, i.e. $h_{Rx^*} \equiv \min\left\{h_{R1x^*},h_{R2x^*}\right\}$ for all x^*. When $Lb_y(\lambda_v;x^*,\underline{p}) \geq h_{R1x^*}$, and the upper bound of the sensitivity parameter from the additional conclusion gives an identification region greater than the one from the identification assumptions, there is no identifying power of the extra conclusion. The simulation study in Section 5 shows that the bounds using λ_v and h_{Rx^*} can cover the true values in simulation designs.

5. Simulation Illustration

This section presents two simulation studies exploring the bounds and breakdown frontiers of the latent density and the regression model under misclassification. Specifically, we consider the following setup: We simulate an outcome variable Y and a latent regressor variable $X^* \in \{0,1\}$. However, X^* is not observed directly. Rather, we observe two proxy variables, X and Z, measured with error as observations of X^*. Given this setup with mismeasured regressors, our simulation studies investigate the following: (i) The bounds of the conditional density $f_{Y|X^*}$ and the conditional means of the outcome $E[Y|X^*]$ that can be recovered from the observed data on (Y,X,Z). (ii) The breakdown frontier analysis for these identification regions. We explore how the bounds and breakdown frontiers change by varying aspects of the data-generating process - such as the deviation from the conditional independence - across simulations. Through these simulation studies, we can understand the effectiveness of our proposed estimation methods.

5.1. Density Case

Consider the outer identified set developed in the density case to a binary choice model with a mismeasured 0-1 dichotomous explanatory variable X^*. We begin with simulated data for the binary variables X^* and Z generated as follows:

$$p(X^* = 1) = 0.5,$$

$$\begin{bmatrix} f_{Z|X^*}(0|0) & f_{Z|X^*}(1|0) \\ f_{Z|X^*}(0|1) & f_{Z|X^*}(1|1) \end{bmatrix}_{2\times 2} = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}_{2\times 2}.$$

The data-generating process for the binary outcome Y and mismeasured regressor X is:

$$X = 1\left(-0.25 + X^* + \gamma Z + u_x \geq 0\right), \tag{67}$$

$$Y = 1\left(-1.2 + 0.6X^* + \alpha X + u_y \geq 0\right), \tag{68}$$

with $u_x \sim N(0, 0.2^2), u_y \sim U(0,1)$, and $u_x \perp (X^*, Z)$ and $u_y \perp (X, X^*, Z)$. We consider 10 values of the parameters (γ, α):

DGP I: $(\gamma, \alpha) = (0,0)$,

DGP II: $(\gamma, \alpha) = (0, 0.15)$,

DGP III: $(\gamma, \alpha) = (0.15, 0)$,

DGP IV: $(\gamma, \alpha) = (0, 0.3)$,

DGP V: $(\gamma, \alpha) = (0.3, 0)$,

DGP VI: $(\gamma, \alpha) = (0.3, 0.3)$,

$$\text{DGP VII: } (\gamma, \alpha) = (0.4, 0.4),$$

$$\text{DGP VIII: } (\gamma, \alpha) = (0.5, 0.5),$$

$$\text{DGP IX: } (\gamma, \alpha) = (0.55, 0.55),$$

$$\text{DGP X: } (\gamma, \alpha) = (0.6, 0.6).$$

When $(\gamma, \alpha) = (0,0)$, we have $Y \perp X \perp Z | X^*$, satisfying the conditional independence assumption. For $\gamma \neq 0$, $f_{X|ZX^*}(X|Z, X^*) \neq f_{X|X^*}(X|X^*)$, and for $\alpha \neq 0$, $f_{Y|XX^*}(Y|X, X^*) \neq f_{Y|X^*}(Y|X^*)$. Therefore, deviating (γ, α) from $(0,0)$ represents a failure of the conditional independence assumption. To measure the deviation from conditional independence in each DGP, we define h_M as the maximum difference between the true joint density $f_{YXZ|X^*}(Y, X, Z|X^*)$ and the product $f_{Y|X^*}(Y|X^*) f_{X|X^*}(X|X^*) f_{Z|X^*}(Z|X^*)$:

$$h_M = \max\{h : h = |f_{YXZ|X^*}(Y, X, Z|X^*) - f_{Y|X^*}(Y|X^*) f_{X|X^*}(X|X^*) f_{Z|X^*}(Z|X^*)|\}.$$

Estimating h_M for each DGP, Table 1 shows that h_M increases with the parameters (γ, α) that directly induce deviation. h_M increases more rapidly with γ due to the stronger effect of Z on X. The bounds depends on λ and h. Since X^* is observed in the simulations, we estimate $\widehat{\lambda} = \max_{1 \le i \le K} \sum_{j \neq i} \widehat{f}_{X|X^*}(j|i)$, where $\widehat{f}_{X|X^*}$ is a nonparametric frequency estimator of $f_{X|X^*}$. We find $\widehat{\lambda}$ increases with γ but not α. To derive bounds, we set $\lambda = 0.1$ for tractability. We use the first step of systematic sensitivity analysis to determine appropriate upper bounds h_{R1x^*} on h for each DGP. For DGPs I-VI, the distinct eigenvalue condition limits $h_{R1x^*} = h_E$. For DGPs VII-X, the density restriction limits $h_{R1x^*} = h_{D_{x^*}}$. We estimate h_S, the largest deviation from conditional independence that satisfies the separation restriction in Proposition 2.1. h_S is larger than h_{R1x^*} in all simulations, so the separation restriction does not bind in the first step selection. The estimated h_S is much smaller than h_M for all DGPs. Since h_S represents the largest perturbation the separation restriction allows, our bounds could only locally capture how the data distribution informs the parameters of interest. The bounds do not fully reflect the extent of deviation from conditional independence, as measured by h_M. We define h_0 as the smallest h that yields an outer identified set covering the true density $f_{Y|X^*}(1|1)$. Since X^* is observed, we can compute h_0. We find $h_{R1x^*} \geq h_0$ for all DGPs, so the bounds cover the truth. In summary, by increasing (γ, α) across DGPs, we find: (1) h_M increases, indicating greater deviation from conditional independence; (2) either distinct eigenvalue or density restrictions determine h bounds; and (3) the resulting bounds still cover the true densities. Our results guide us on when our bound method can be applied despite failures of conditional independence.

To conduct a breakdown frontier analysis in the second step selection, we impose an additional assumption that the conditional mean $E[Y|X^* = 1] \geq \underline{\mu} = 0.15$. Proposition 2.1 implies the identification set for $f_{Y|X^*}(1|1) = E[Y|X^* = 1]$ is:

$$\left(\max\left\{\rho_2 - \overline{\kappa_\lambda} \cdot \overline{P_1^h}, 0\right\}, \min\left\{1, \rho_2 + \overline{\kappa_\lambda} \cdot \overline{P_1^h}\right\}\right),$$

where ρ_2 is a point estimation for $f_{Y|X^*}(1|1)$ and $\overline{P_1^h}$ is the misclassification matrix evaluated at $y = 1$. An informative lower bound is:

$$\rho_2 - \overline{\kappa_\lambda} \cdot \overline{P_1^h} \geq \underline{\mu}.$$

It follows that the breakdown frontier for a given λ is:

$$bf(\lambda; \underline{\mu}) = \frac{\kappa(\underline{\mu}, \lambda)}{K||M_{XZ}^{-1}||(1 + \kappa(\underline{\mu}, \lambda) + ||M_{XZ}^{-1}||||M_{\overline{y}XZ}||)},$$

where $\kappa(\underline{\mu}, \lambda) \equiv \frac{\rho_2 - \underline{\mu}}{\overline{\kappa_\lambda}}$.

In Table 2, we estimate the breakdown frontier $bf(\lambda; \underline{\mu})$ at $\lambda = 0.1$, the breakdown frontier $bf(\lambda; \underline{\mu})$ at $\lambda = 0$, the 95% one-sided lower confidence interval of $bf(\lambda = 0.1; \underline{\mu})$, the recommended upper bound h_{Rx^*}, etc. Only in DGPs I, III, and V, the upper bounds for the sensitivity parameter from the breakdown frontier analysis (using the lower bound of the 95% one-sided lower confidence interval of $bf(\lambda = 0.1; \underline{\mu})$) is less than h_{R1x^*} and this implies that there is identifying power of the conclusion to construct bounds. For most DGPs, $h_{Rx^*} > h_0$, so bounds at h_{Rx^*} covered the truth, suggesting the bounds are quite robust to failures of conditional independence despite their lack of sharpness.

Figure 1 shows the breakdown frontier analysis results for DGPs I-VI. The horizontal axis indicates λ, the upper bound on misclassification probabilities. The vertical axis indicates h, the degree of deviation from conditional independence. The breakdown frontiers slope negatively, indicating a diminishing rate of substitution between relaxing the assumptions on λ and h. Greater measurement error λ implies less deviation h from conditional independence is needed to conclude, and vice versa. The upper bound h_{BF0} is the maximum h that allows the conclusion when $\lambda = 0$ (no measurement error). h_{Rx^*}, the recommended upper bound on h, equal $h_{R2_x^*}$ for DGPs I, III and V but $h_{R1_x^*}$ for DGPs II, IV and VI. The breakdown frontier analysis illuminates how the identifying power comes from jointly relaxing both assumptions. There are trade-offs in how much one can perturb the model before losing the ability to draw a meaningful conclusion. The specific trade-off depends on the data-generating process. Comparing h_{BF0} to h_M and $h_{Rx^*} \geq h_0$ suggests the approach can identify a neighborhood of the outer identified set, recovering part of the truth even moderately far from conditional independence. However, extrapolating too far risks misleading inferences.

Figure 2 shows estimated bounds for $f_{Y|X^*}(1|1)$ using the bounds in Proposition 2.1 with $\lambda = 0.1$ and $h \in [0, h_{Rx^*}]$, under conditional partial independence. We examine how the bounds change with increasing failure of conditional independence across DGPs I-VI by varying (γ, α). The cyan stars show point estimates that assume conditional independence, while the blue stars show the true $f_{Y|X^*}(1|1)$ values. As expected, the stars separate more with greater departure from conditional independence. For DGPs I, III, and V, the distances between the stars increase with γ, rising from 0.025 to 0.071 as γ increased from 0 to 0.3. For DGPs I, II, and IV, the stars remained close as α increases from 0 to 0.3, with $\gamma = 0$. For all DGPs, the bounds using $\lambda = 0.1$ and h_{Rx^*} cover the true $f_{Y|X^*}(1|1)$ values. Despite failures of conditional independence, our recommended nonparametric bounds yield reasonable estimates of the truth. The results show that the approach can recover part of the identifying power lost by deviations from conditional independence. The bounds expand as needed to continue covering the truth. However, the stars also separate more, highlighting the bias from relying on point estimates that falsely assume conditional independence. This tension suggests standard methods may apply locally, but extrapolating too far risks misleading inferences – especially if deviations are moderate to large. Comparing the bounds to h_M for each DGP provides a sense of how much the approach can recover while yielding valid conclusions for that population. The bounds reflect the data sufficiently to represent part of the truth, giving useful information even when conditional independence fails.

5.2. Regression Case

The DGP for the proxy X is:

$$X = 1\left(-0.25 + 0.5X^* + \gamma Z + u_x \geq 0\right),$$

where $u_x \sim N(0, 0.2^2)$, $u_x \perp (X^*, Z)$, $p(X^* = 1) = 0.5$, and the conditional probability matrix for Z given X^* is:

$$\begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix}.$$

The DGP for the outcome Y is:

$$Y = -4.5 + 5X^* + \varepsilon,$$

where $\varepsilon \sim U(-1/2, 1/2)$. The model satisfies Assumption 2.7, and $\gamma \neq 0$ represents failure of conditional independence, i.e. $f_{X|ZX^*}(X|Z,X^*) \neq f_{X|X^*}(X|X^*)$. We consider nine values of γ:

DGP I: $\gamma = 0$,

DGP II: $\gamma = 0.2$,

DGP III: $\gamma = 0.3$,

DGP IV: $\gamma = 0.4$,

DGP V: $\gamma = 0.45$,

DGP VI: $\gamma = 0.5$,

DGP VII: $\gamma = 0.55$,

DGP VIII: $\gamma = 0.6$,

DGP IX: $\gamma = 0.625$.

The deviation from conditional independence is measured by:

$$\tilde{h}_M = \max\left\{h : h = |f_{XZ|X^*}(X,Z|X^*) - f_{X|X^*}(X|X^*)f_{Z|X^*}(Z|X^*)|\right\}.$$

Table 3 reports the estimated $\tilde{h}_M$ and $\widehat{\lambda}$ for each DGP. As expected, both increase with γ. We set $\lambda = 0.1$ as in the density case. Table 3 shows the first estimates h_I, h_E, h_S, and h_{R1x^*}. For all DGPs, $h_E \leq h_S$ and the distinct eigenvalue restriction limits $h_{R1x^*} = h_E$. In summary, this simulation design systematically varies the degree of deviation from conditional independence by increasing γ across DGPs. We find that the measured deviation $\tilde{h}_M$ and misclassification probability λ rose with greater departure from the model assumption. However, the distinct eigenvalue restriction still determines the first upper bound h_{R1x^*} for all DGPs.

We consider the additional conclusion for the second selection step that $E[Y|X^* = 0] \geq \underline{\mu} = -6$ to evaluate breakdown frontiers. Table 4 gives the results. Only for DGP I, the breakdown frontier analysis yield a smaller h bound (h_{Rx^*}) than h_{R1x^*}. For all DGPs, $h_{Rx^*} > h_0$, so the bounds cover the true conditional means. Figure 3 shows the breakdown frontier plots, similar to Figure 1. The frontiers slope negatively, indicating decreasing marginal rates of substitution between relaxing λ and h. More measurement error required less deviation from conditional independence, and vice versa, to conclude. For DGP I, the additional conclusion provided meaningful identifying power by yielding a h_{Rx^*} bound smaller than the h_{R1x^*} bound. For all DGPs, while $h_{Rx^*} > h_0$ ensures the conclusions remain valid, the differences between h_{Rx^*} and h_M highlight a lack of sharpness. The breakdown frontier of the regression case exhibits similar patterns to the density case.

Figure 4 shows identified sets for $E[Y|X^* = 0]$ under Assumption 2.6 (conditional partial independence) using the bounds in Proposition 2.3. The cyan stars indicate point estimates assuming full conditional independence, while the blue stars show the true $E[Y|X^* = 0]$ values. For DGPs II, III, and IV, the stars are far apart when $\gamma \neq 0$, indicating bias from relying on estimates that falsely assume conditional independence. The bounds begin covering the truth once $h_{Rx^*} > h_0$. Since $h_{Rx^*} > h_0$ for all DGPs, the bounds span the true $E[Y|X^* = 0]$ values. The results show that our nonparametric bounds also perform well for the regression case. However, the separation between the stars highlights the risk of misleading inferences from the point identification method that does not account for such deviations. The differences between h_{Rx^*} and $\tilde{h}_M$ suggest the bounds are not fully sharp, only locally capturing the data-generating process. While yielding reasonably accurate sets, caution is needed in interpreting findings too far from the values used in the analysis. The bounds provide useful information on $E[Y|X^* = 0]$ for neighborhoods of the data distribution utilized but may not represent the entire outer identified set globally.

6. Empirical Application

This section applies the developed estimator to a regression model to investigate the impact of schooling on wages in which the true education level of each individual is subject to misreporting error.[9] We use the dataset in Kane et al. (1999), which contains wages and education levels from the National Longitudinal Study of 1972 (NLS-72) and the Postsecondary Education Transcript Study (PETS). Consider a measurement model of a wage equation as follows:

$$Y = E(Y|X^*) + U, \tag{69}$$

where Y represents log wage, X^* symbolizes the true education level, and U constitutes a regression error. We observe two measurements of the true education level X^*, namely the self-

[9]We are grateful to Professor Ruli Xiao at Indiana University for providing some MATLAB codes and other materials in this section.

reported education attainment X and the transcript-recorded education attainment Z.[10] Individuals are more likely to report the same educational attainment as the one in the transcript. Table 5 reports summary statistics for the three education categories in transcript-recorded and self-reported education attainments. When analyzing the data, we have categorized it into three different education categories: *High School*, *Some College*, and *Bachelor*. Individuals with more schooling in transcript-recorded or self-reported education attainments appear to have higher wages.

Our approach to analyzing the relationship between wages and schooling is to consider the partial identification for the regression model under misclassification in Proposition 2.3. The advantage of the approach is relaxing conditional independence by using conditional partial independence between the self-reported education attainment X and the transcript-recorded education attainment Z in Assumption 2.6. Rewrite the perturbation equation (40) in this empirical application as

$$\widetilde{d_0}(X,Z,X^*) \equiv \left(f_{X|ZX^*}\left(X|Z,X^*\right) - f_{X|X^*}\left(X|X^*\right)\right) f_{Z|X^*}\left(Z|X^*\right). \tag{70}$$

If $f_{Z|X^*}(Z|X^*) > 0$, the variation of the conditional partial independence between X and Z is reduced to the variation of $f_{X|ZX^*}(X|Z,X^*) - f_{X|X^*}(X|X^*)$ which reflects the marginal impact of adding the extra conditioning variable Z into the conditional probability $f_{X|X^*}(X|X^*)$. That is, given (X,X^*), $f_{X|ZX^*}(X|Z,X^*)$ depends on the transcript-recorded education attainment Z. More precisely, reporting behavior of education attainment conditional on the actual education level fluctuates when there is additional information on transcript-recorded education attainment. For instance, consider the probabilities of reporting for high-school individuals at another transcript-recorded education attainment, such as $f(X|Z = \textit{High School}, X^* = \textit{High School})$ and $f(X|Z = \textit{Bachelor}, X^* = \textit{High School})$. If $f(X = \textit{Bachelor}|Z = \textit{Bachelor}, X^* = \textit{High School}) > f(X = \textit{Bachelor}|Z = \textit{High School}, X^* = \textit{High School})$, this indicates there is a tendency for overreporting when there is a discrepancy in the transcript-recorded education attainment. Therefore, our approach permits the variation to integrate information on the transcript-recorded education attainment on probabilities of misreporting. Assumption 2.4 in which people are more willing, conditional on their education, to tell the truth rather than lie. It can be observed that Assumption 2.4 is in line with the findings of Kane et al. (1999) (Table 5, p. 18). According to the research, individuals with various education levels are likely to provide

[10] The wage in Kane et al. (1999) are limited to observations reporting wages between \$1.50 and \$80.00 per hour in 1986 and there may exist measurement error problems in it. However, we focus on the case of misreporting in a categorical variable with more than two measures and assume the dataset provides accurate wage measures.

accurate self-reports of their educational attainment, with probabilities greater than or close to 0.92. Based on the research, the probability of accurate self-reports of educational attainment is high among individuals with various education levels. Therefore, the upper bound of misreporting educational attainments is reasonable, given that the sum of conditional probabilities for misreporting is less than or equal to 0.08. We will specify $\lambda_v = 0.08$ for the sensitivity analysis in Section 4 to compute identification regions. Assumption 2.7 asserts that the conditional mean of log wage given that the true education level is unaffected by the self-reported education attainment X and the transcript-recorded education attainment Z. Assumption 2.9 guarantees that the conditional means of log wage given the true education levels differ across varying levels.

Table 6 reports point estimates of mean log wage on the three education categories and their standard errors using the estimator in Hu (2008). The maintained conditional independence assumption of the point estimates is that reporting behavior of educational attainment does not depend on the transcript-recorded education attainment when the true education level is known. The results in Table 6 show that the mean log wage of *High School* is less than the mean log wage of *Some College* by about $2.208 - 2.025 = 0.183$ while the mean log wage of *Some College* is less than mean log wage of *Bachelor* about $2.446 - 2.208 = 0.238$. The wage differences between education categories increase with the education level.

Table 7 presents the estimation results of the bounds of mean log wage across the three education categories at the verified probability of misreporting $\lambda_v = 0.08$. For the given misclassification probability, we merely conduct the first step selection in subsection 4.2 and do not impose any additional conclusion for the second step selection because Assumption 2.9 already provides the desired constraints on the conditional mean of log wage given the true education level. In the regression model, there is no density limitation, and the range of the sensitivity parameter from the first step selection is $h_{R1} = h_E$. The empirical recommendation for the sensitivity parameter h of the bounds in Table 7 is 8.549×10^{-4}. The selection of the sensitivity parameters generally leads to the bounds in Table 7 being disjoint. Although the recommended value is small, the corresponding bounds and their 95% confidence intervals in Table 7 across different education attainments are significantly *wider* than the 95% confidence intervals of the point estimator in Table 6. The wider intervals imply that the recommended value is statistically significantly different from zero and the corresponding bounds provide meaningful ranges of mean log wage when the assumption of conditional independence is not holding. It is crucial to understand that h primarily measures the extent to which the conditional independence assumption is relaxed within the model. Although the small magnitude of h might initially suggest a minimal deviation from conditional independence, its fundamental role in this empirical context is to signify whether the assumption holds rather than quantify its degree of relaxation. The modest size of h in this instance complicates its interpretation as a marker of varying degrees of assumption relaxation. Therefore, while h is instrumental in indicating the presence of deviation from conditional independence, it does not provide a granular measure of the extent of this deviation.

While precision is lost relative to point estimates under ideal assumptions, partial identification provides sensible ranges reflecting uncertainty from model deviations. Incorporating sensitivity analysis provides more realistic and informative bounds rather than relying on implausible assumptions for mathematical convenience. This allows better-calibrated predictions and decisions than misleading point estimates under flawed assumptions. There are the worst-case bounds for the wage differentials between the education categories which represent the widest possible range of values for wage difference by educational attainment level. While the worst wage differential from *High School* to *Some College* is 2.329-1.906=0.423, the worst wage differentials from *Some College* to *Bachelor* is 2.559-2.102=0.457. On the other hand, the worst wage differentials from *High School* to *Some College* and from *Some College* to *Bachelor* using the 95% confidence intervals of the point estimator are 2.230-2.005=0.225 and 2.464-2.186=0.278 respectively.

The point estimator underestimates the wage difference by educational attainment levels. In all estimation results, wage difference by educational attainment level increases with the education level. Figure 5 plots the 95% confidence intervals of the point estimator, the bounds, and their 95% confidence intervals at the education categories. Both 95% confidence intervals of the point estimator and the bounds are symmetric intervals centered at the point estimator for all educational attainments. However, the centers of the 95% confidence interval of the bounds at *High School*, and *Some College*, are 2.017 (smaller than the point estimator 2.025), and 2.216 (larger than the point estimator 2.208), respectively. While the mean log wage of individuals in *High School* has a more extended spread below the point estimator, the mean log wage spread of individuals in *Some College* has a longer spread above the point estimator. The indicates little asymmetry in the distributions of the bounds at *High School*, and *Some College*.

7. Conclusion

In this paper, we relax the baseline conditional independence assumptions of the point identifications to derive identified sets for various nonlinear models when a discrete explanatory variable is subject to misclassification error. We use sup-norm deviations to weaken the conditional independence assumptions, and the deviations can be expressed as a perturbation matrix between an observable matrix and an eigenvalue-eigenvector decomposition. Then, we apply the perturbation theory of the eigenvalue of a diagonalizable matrix to provide the bounds of the nonlinear models with misclassification error, and the upper bound of misreporting probabilities and deviations can index the bounds. Treating the upper bound of misreporting probabilities as baseline information and input provided by validation studies, we propose a systematic sensitivity analysis to select the upper bounds of the deviation to construct identified sets. The selection procedure requires two steps. The first step involves imposing the identification assumptions on the sensitivity parameter and the restrictions coming from probability theory. The second step is specifying additional desired conclusions for a breakdown frontier analysis. The upper bound for the deviation is the minimum of the two upper bounds from these two steps. Our simulation results suggest that the identified sets using the recommended upper bounds of the sensitivity parameter can cover the true parameters of interest. The empirical illustration provides meaningful ranges for the potential values of the mean log wage parameters. Though wider, the bounds likely contain the true values and are informative for prediction/decision tasks.

The partial identification results rely on those distinct restrictions on the latent data-generating process and on misreporting probabilities (higher chance of reporting truthfully). As the deviations approach zero, the nonparametric partial identification of the nonlinear models with misclassification error becomes the point identification. It may be useful for researchers to consider any prior information they have regarding the misreporting pattern when analyzing data. By doing so, they can incorporate this information directly into their partial identification analysis, which is indexed by the upper bound of misreporting probabilities and deviations. This approach can improve the analysis's accuracy and reliability, leading to more insightful and meaningful results. The systematic sensitivity analysis uses prior information and observed data to determine an upper bound on deviations ensuring reasonably accurate conclusions, at least for neighborhoods of values utilized. However, caution is needed to interpret too far beyond that, as the identified set remains unknown.

A. Proofs

It is worth mentioning that the condition number of the misclassification probability matrix is defined as $\kappa(M_{X|X^*}) = |||M_{X|X^*}^{-1}||| \cdot |||M_{X|X^*}|||$, and this information plays a crucial role in determining the perturbation bound. Our research has revealed that the condition number of the misclassification probability matrix depends on the specific restriction outlined in Assumption 2.4.

Proposition A.1. *(Bounds for $\kappa(M_{X|X^*})$) Under Assumption 2.4, the possible values for the condition number $\kappa(M_{X|X^*})$ is*

$$1 \leq \kappa(M_{X|X^*}) \leq \frac{1}{(1-\lambda)(1-2\lambda)} \equiv \overline{\kappa_\lambda}, \tag{71}$$

where $0 \leq \lambda < \frac{1}{2}$ is a constant defined in Assumption 2.4.

Proof of Proposition A.1: Assumption 2.4 implies the following inequality

$$f_{X|X^*}(i|i) > \sum_{j\neq i} f_{X|X^*}(j|i) \quad \forall i. \tag{72}$$

This ensures that $M_{X|X^*}$ is strictly diagonally dominant and every main diagonal entry of $M_{X|X^*}$ is nonzero. Denote I_K as the $K \times K$ identity matrix and the $K \times K$ diagonal matrix $D_{X|X^*} = diag(f_{X|X^*}(1|1), ..., f_{X|X^*}(K|K))$. Consider a row scaling away matrix $D^{-1}_{X|X^*} M_{X|X^*}$ which has all 1s on its main diagonal and its (i,j) element is $\frac{f_{X|X^*}(j|i)}{f_{X|X^*}(i|i)}$ for $j \neq i$. Then, the matrix

$$B \equiv [b_{ijw}] = I_K - D^{-1}_{X|X^*} M_{X|X^*} \tag{73}$$

has all 0s on the main diagonal, and $b_{ijw} = -\frac{f_{X|X^*}(j|i)}{f_{X|X^*}(i|i)}$. Consider the maximum row sum matrix norm of B,

$$|||B||| = \max_{1\leq i\leq K} \frac{\sum_{j\neq i} f_{X|X^*}(j|i)}{f_{X|X^*}(i|i)}. \tag{74}$$

Applying the triangle inequality to Equation (73) yields

$$|||D^{-1}_{X|X^*} M_{X|X^*}||| = |||I_K - (I_K - D^{-1}_{X|X^*} M_{X|X^*})||| \leq 1 + |||B|||. \tag{75}$$

Assumption 2.4 implies $|||B||| \leq \frac{\lambda}{1-\lambda} < 1$ for $\lambda \in [0, 1/2)$. This implies that $I_K - B = D^{-1}_{X|X^*} M_{X|X^*}$

is nonsingular with[11]

$$M^{-1}_{X|X^*} D_{X|X^*} = (I_K - B)^{-1} = \sum_{j=0}^{\infty} B^j. \tag{76}$$

It follows that

$$|||M^{-1}_{X|X^*} D_{X|X^*}||| \leq \sum_{j=0}^{\infty} |||B|||^j = \frac{1}{1 - |||B|||}. \tag{77}$$

Combining the inequalities in Equations (75) and (77) with the submultiplicativity of the matrix norm yields the upper bound of $\kappa(M_{X|X^*})$

$$\begin{aligned}
\kappa(M_{X|X^*}) &= |||M^{-1}_{X|X^*}||| \cdot |||M_{X|X^*}||| && (78)\\
&= |||\left(M^{-1}_{X|X^*} D_{X|X^*}\right) D^{-1}_{X|X^*}||| \cdot |||D_{X|X^*}\left(D^{-1}_{X|X^*} M_{X|X^*}\right)||| && (79)\\
&\leq |||M^{-1}_{X|X^*} D_{X|X^*}||| \cdot |||D^{-1}_{X|X^*}||| \cdot |||D_{X|X^*}||| \cdot |||D^{-1}_{X|X^*} M_{X|X^*}||| && (80)\\
&\leq |||D^{-1}_{X|X^*}||| \cdot |||D_{X|X^*}||| \cdot \frac{1 + |||B|||}{1 - |||B|||} && (81)\\
&\leq \frac{\max_i f_{X|X^*}(i|i)}{\min_i f_{X|X^*}(i|i)} \cdot \frac{1 + |||B|||}{1 - |||B|||} && (82)\\
&\leq \frac{1}{(1-\lambda)(1-2\lambda)}, && (83)
\end{aligned}$$

where we have used that $|||B||| < \frac{\lambda}{1-\lambda}$ and lower bounds on the probabilities of correct report in Equation (25), i.e., $\min_i f_{X|X^*}(i|i) \geq 1 - \lambda$.

As for the lower bound of $\kappa(M_{X|X^*})$, consider

$$1 = |||I_K||| = |||M^{-1}_{X|X^*} M_{X|X^*}||| \leq |||M^{-1}_{X|X^*}||| \cdot |||M_{X|X^*}||| = \kappa(M_{X|X^*}). \tag{84}$$

Q.E.D.

Proof of Proposition 2.1: We have formalized the conditional partial independence in Assumption 2.1 using matrix notation in Equations (15) and (16). We first show that Assumption 2.2(i) implies that $M_{XZ} - M_d$ is invertible. To find the existence of the inverse term, we adopt the following result of approximating the inverse of a perturbed matrix from Horn and Johnson (2013).[12]

[11] Corollary 5.6.16 in page 351 of Horn and Johnson (2013): If $|||\cdot|||$ is a matrix norm, and if $|||A||| < 1$, then $I_K - A$ is nonsingular and $\left(I_K - A\right)^{-1} = \sum_{j=0}^{\infty} A^j$.

[12] The result can be found on page 381 of Horn and Johnson (2013).

Lemma A.1. *Let $|||\cdot|||$ be a matrix norm. Consider an invertible matrix A, subject to a small perturbation ΔA. If $|||A^{-1}\Delta A||| < 1$, then $A+\Delta A$ is invertible and*

$$(A+\Delta A)^{-1} = A^{-1} + B, \tag{85}$$

where $B = \sum_{i=1}^{\infty}(-A^{-1}\Delta A)^n A^{-1}$, *and* $|||B||| \le \frac{|||A^{-1}|||\cdot|||A^{-1}|||\cdot|||\Delta A|||}{1-|||A^{-1}\Delta A|||}$.

Because the matrix norm is a submultiplicative, we have $|||M_{XZ}^{-1}M_d||| \le |||M_{XZ}^{-1}|||\cdot|||M_d||| \le |||M_{XZ}^{-1}|||Kh$ by Equation (20). Under Assumption 2.2(i), $|||M_{XZ}^{-1}M_d||| < 1$. Then, the inverse matrix $(M_{XZ}-M_d)^{-1}$ exists by Lemma A.1 and this implies M_{X^*Z} is invertible by Equation (16) and the invertibility of $M_{X|X^*}$ in Assumption 2.4. In addition, $(M_{XZ}-M_d)^{-1} = M_{XZ}^{-1}+B_d$, where

$$|||B_d||| \le \frac{|||M_{XZ}^{-1}|||\cdot|||M_{XZ}^{-1}|||\cdot|||M_d|||}{1-|||M_{XZ}^{-1}M_d|||} \le \frac{|||M_{XZ}^{-1}|||^2 Kh}{1-|||M_{XZ}^{-1}|||Kh}. \tag{86}$$

The perturbation matrix $P_y^h(d)$ is expressed in terms of the observed matrices M_{XZ}, M_{yXZ} and the perturbation matrices $M_d(y)$ and M_d. Since the deviation matrices $M_d(y)$ and M_d are relatively small when h is small, the perturbation matrix $P_y^h(d)$ is also relatively small for a small h. The perturbation matrix $P_y^h(d)$ can be bounded by

$$\begin{aligned} |||P_y^h(d)||| &\le |||-M_{XZ}^{-1}M_d(y)+B_d\left(M_{yXZ}-M_d(y)\right)||| \\ &\le |||M_{XZ}^{-1}M_d(y)|||+|||B_d|||\left(|||M_{yXZ}|||+|||M_d(y)|||\right) \\ &\le |||M_{XZ}^{-1}|||Kh+\frac{|||M_{XZ}^{-1}|||^2Kh}{1-|||M_{XZ}^{-1}|||Kh}\left(|||M_{yXZ}|||+Kh\right) \equiv \overline{P_y^h}, \end{aligned} \tag{87}$$

where we have used the inequalities in Equations (19) and (86). Notice that $P_y^h(d) = 0$ if $h = 0$. The equation (23) implies that the eigenvalue-eigenvector decomposition $M_{X|X^*}^{-1}D_{y|X^*}M_{X|X^*}$ is close to the observable matrix $M_{XZ}^{-1}M_{yXZ}$ under Assumption 2.1.

Next, we apply the Bauer-Fike theorem in Theorem 2.1 to the matrix equation (23) using the notations from Equations (26)-(30). Under Assumptions 2.3, and 2.4, $M_{X|X^*}^{-1}D_{y|X^*}M_{X|X^*}$ has a unique eigenvalue-eigenvector decomposition with distinct eigenvalues $\left\{f_{Y|X^*}(y|x^*=1),\ldots,f_{Y|X^*}(y|x^*=K)\right\}$ for given y. If ρ is an eigenvalue of $M_{XZ}^{-1}M_{yXZ}$ for given y, then for given y we claim that there exists a unique eigenvalue $\rho^* \in \left\{f_{Y|X^*}(y|x^*=1),\ldots,f_{Y|X^*}(y|x^*=K)\right\}$ for which

$$|\rho-\rho^*| \le \overline{\kappa_\lambda}\cdot\overline{P_y^h}. \tag{88}$$

Suppose the claim is false. Under the perturbation result in Theorem 2.1, there would exist $\rho_1^* \neq \rho_2^* \in \left\{f_{Y|X^*}(y|x^*=1),...,f_{Y|X^*}(y|x^*=K)\right\}$ such that

$$|\rho-\rho_1^*| \leq \overline{\kappa_\lambda}\cdot\overline{P_y^h} \leq d_{Ey}, \tag{89}$$

$$|\rho-\rho_2^*| \leq \overline{\kappa_\lambda}\cdot\overline{P_y^h} \leq d_{Ey}. \tag{90}$$

This implies

$$|\rho_1^*-\rho_2^*| \leq |\rho-\rho_1^*|+|\rho-\rho_2^*| \leq 2d_{Ey}, \tag{91}$$

which contradicts the definition of d_{Ey}. We have reached the claim, and we can label the eigenvalue of $M_{XZ}^{-1}M_{yXZ}$ corresponding to the eigenvalue $f_{Y|X^*}(y|x^*=k)$ as ρ_k and $|\rho_k - f_{Y|X^*}(y|x^*=k)| \leq d_{Ey}$. As for the distinctness of $\rho_k, k=1,...,K$, given y, consider for $k_1 \neq k_2 \in \left\{1,...,K\right\}$

$$\begin{aligned}
&\left|f_{Y|X^*}\left(y|x^*=k_1\right)-f_{Y|X^*}\left(y|x^*=k_2\right)\right| \\
&=\left|f_{Y|X^*}\left(y|x^*=k_1\right)-\rho_{k_1}+\rho_{k_1}-\rho_{k_2}+\rho_{k_2}-f_{Y|X^*}\left(y|x^*=k_2\right)\right| \\
&\leq\left|f_{Y|X^*}\left(y|x^*=k_1\right)-\rho_{k_1}\right|+\left|\rho_{k_1}-\rho_{k_2}\right|+\left|\rho_{k_2}-f_{Y|X^*}\left(y|x^*=k_2\right)\right|.
\end{aligned}$$

This implies that

$$\begin{aligned}
\left|\rho_{k_1}-\rho_{k_2}\right| \geq{}& \left|f_{Y|X^*}\left(y|x^*=k_1\right)-f_{Y|X^*}\left(y|x^*=k_2\right)\right| \\
&-\left|f_{Y|X^*}\left(y|x^*=k_1\right)-\rho_{k_1}\right|-\left|\rho_{k_2}-f_{Y|X^*}\left(y|x^*=k_2\right)\right| \\
>{}& 2d_{Ey}-d_{Ey}-d_{Ey}=0,
\end{aligned}$$

where we have used the definition of d_{Ey}. *Q.E.D.*

Proof of Proposition 2.2: By synthesizing the matrix equations in Equations (16) and (33), the eigenvalue-eigenvector decomposition in Equation (34) is attained, $M_{X|X^*}^{-1}D_{F_{\tilde{y}}|X^*}M_{X|X^*} = (M_{XZ}-M_d)^{-1}\left(M_{F_{\tilde{y}}XZ}-M_{dF}(\tilde{y})\right) = M_{XZ}^{-1}M_{F_{\tilde{y}}XZ}+P_{F_{\tilde{y}}}^h(d)$. By replicating the derivation employed in the proof of Proposition 2.1, an upper bound on $P_{F_{\tilde{y}}}^h(d)$ emerges, as follows:

$$|||P_{F_{\tilde{y}}}^h(d)||| = |||-M_{XZ}^{-1}M_{dF}(\tilde{y})+B_d\left(M_{F_{\tilde{y}}XZ}-M_{dF}(\tilde{y})\right)||| \tag{92}$$

$$\leq |||M_{XZ}^{-1}|||Kh+\frac{|||M_{XZ}^{-1}|||^2Kh}{1-|||M_{XZ}^{-1}|||Kh}\left(|||M_{F_{\tilde{y}}XZ}|||+Kh\right)\equiv\overline{P_{F_{\tilde{y}}}^h}. \tag{93}$$

Therefore, the conclusions of Proposition 2.2 can be attained by employing the proof strategy of Proposition 2.1. *Q.E.D.*

Proof of Proposition 2.3: Let us begin by considering the conditional partial independence assumption stated in Assumption 2.6. We can multiply Equation (40) by $E(Y|X^*)c$ and sum over the support of X^* and this yields

$$\begin{aligned}
\text{(94)}\qquad & \sum_{x^*} \widetilde{d_0}(X,Z,x^*)E(Y|x^*)f_{X^*}\left(x^*\right) \\
&= \sum_{x^*} E(Y|x^*)f_{XZX^*}\left(X,Z,x^*\right) - \sum_{x^*} E\left(Y|x^*\right) f_{X|X^*}\left(X|X^*\right) f_{ZX^*}\left(Z,x^*\right) \\
&= \sum_{x^*} E(Y|X,Z,x^*)f_{XZX^*}\left(X,Z,x^*\right) - \sum_{x^*} E\left(Y|x^*\right) f_{X|X^*}\left(X|X^*\right) f_{ZX^*}\left(Z,x^*\right) \\
&= E(Y|X,Z)f_{XZ}(X,Z) - \sum_{x^*} E\left(Y|x^*\right) f_{X|X^*}\left(X|X^*\right) f_{ZX^*}\left(Z,x^*\right),
\end{aligned}$$

where we have used Equation (41) in Assumption 2.7. We can express the equation in matrix notation by defining the $K \times K$ matrix $\widetilde{M}_d$ as follows:

$$\text{(95)}\qquad \widetilde{M}_d = \begin{bmatrix} \widetilde{d_1}(1,1) & \cdots & \widetilde{d_1}(K,1) \\ \vdots & & \\ \widetilde{d_1}(1,K) & \cdots & \widetilde{d_1}(K,K) \end{bmatrix}_{K\times K},$$

where $\widetilde{d_1}(X,Z) \equiv \sum_{x^*} \widetilde{d_0}(X,Z,x^*)E(Y|x^*)f_{X^*}(x^*)$. Based on Assumption 2.6, we can infer that $|\widetilde{d_1}(X,Z)| \le \sum_{x^*} |\widetilde{d_0}(X,Z,x^*)||\int_y y f_{YX^*}(y,x^*)dy| \le Kh|E(Y)|$, each element of the matrix $\widetilde{M}_d$ is less than $Kh|E(Y)|$. By the definition of the matrix norm in Equation (18), we obtain

$$\text{(96)}\qquad |||\widetilde{M}_d||| \le K^2 h|E(Y)|.$$

Define the following matrix:

$$\text{(97)}\qquad D_{\overline{y}|X^*} = \begin{bmatrix} E(Y|x^*=1) & 0 & \cdots & 0 \\ \vdots & E(Y|x^*=k) & \vdots & 0 \\ 0 & \cdots & 0 & E(Y|x^*=K) \end{bmatrix}_{K\times K}.$$

The matrix expression of Equation (94) is

$$\text{(98)}\qquad \widetilde{M}_d = M_{\overline{y}XZ} - M_{X^*Z} D_{\overline{y}|X^*} M_{X|X^*}.$$

After analyzing Equation (40) and multiplying it by $f_{X^*}(X^*)$, and then summing over the support of X^*, we have

$$\sum_{x^*} \widetilde{d_0}(X,Z,x^*) f_{X^*}\left(x^*\right) = f(X,Z) - \sum_{x^*} f_{X|X^*}\left(X|X^*\right) f_{ZX^*}\left(Z,x^*\right). \tag{99}$$

Define $\widetilde{d_2}(X,Z) = \sum_{x^*} \widetilde{d_0}(X,Z,x^*) f_{X^*}(x^*)$ and define the $K \times K$ matrix $\widetilde{M}_{d_2}$ as follows

$$\widetilde{M}_{d_2} = \begin{bmatrix} \widetilde{d_2}(1,1) & \cdots & \widetilde{d_2}(K,1) \\ \vdots & & \\ \widetilde{d_2}(1,K) & \cdots & \widetilde{d_2}(K,K) \end{bmatrix}_{K \times K}. \tag{100}$$

The matrix expression of Equation (99) is

$$\widetilde{M}_{d_2} = M_{XZ} - M_{X^*Z} M_{X|X^*}. \tag{101}$$

By Assumption 2.6, $|\widetilde{d_2}(X,Z)| \le \sum_{x^*} |\widetilde{d_0}(X,Z,x^*)| f_{X^*}(x^*) \le h \sum_{x^*} f_{X^*}(x^*) = h$, every entry of the matrix $\widetilde{M}_{d_2}$ is less than h. Similar to the inequality in Equation (96), we have

$$|||\widetilde{M}_{d_2}||| \le Kh. \tag{102}$$

We apply Lemma A.1 with Assumption 2.8 to address the existence of the inverse matrix $\left(M_{XZ} - \widetilde{M}_{d_2}\right)^{-1}$ and this yields

$$\left(M_{XZ} - \widetilde{M}_{d_2}\right)^{-1} = M_{XZ}^{-1} + \widetilde{B}_d, \tag{103}$$

where $|||\widetilde{B}_d||| \le \frac{|||M_{XZ}^{-1}||| \cdot |||M_{XZ}^{-1}||| \cdot |||\widetilde{M}_{d_2}|||}{1 - |||M_{XZ}^{-1}\widetilde{M}_{d_2}|||}$. Similar to the derivation of the bounds for B_d in Equation (86), we obtain $|||M_{XZ}^{-1}\widetilde{M}_{d_2}||| \le |||M_{XZ}^{-1}||| \cdot |||\widetilde{M}_{d_2}||| \le |||M_{XZ}^{-1}|||Kh$ and then

$$|||\widetilde{B}_d||| \le \frac{|||M_{XZ}^{-1}|||^2 Kh}{1 - |||M_{XZ}^{-1}|||Kh}. \tag{104}$$

As explicated in the proof of Proposition 2.1, the invertibility of $M_{XZ} - \widetilde{M}_{d_2}$ and the invertibility of $M_{X|X^*}$ in Assumption 2.4 ensures M_{X^*Z} is invertible. Thus, we can combine the matrix relationship in Equations (98) and (101) as

$$M_{X|X^*}^{-1} D_{\overline{y}|X^*} M_{X|X^*} = \left(M_{XZ} - \widetilde{M}_{d_2}\right)^{-1} \left(M_{\overline{y}XZ} - \widetilde{M_d}\right). \tag{105}$$

Put the relation in Equation (103) back to Equation (105),

$$
\begin{aligned}
(106) \qquad M_{X|X^*}^{-1} D_{\overline{y}|X^*} M_{X|X^*} &= \left(M_{XZ}^{-1} + \widetilde{B}_d\right)\left(M_{\overline{y}XZ} - \widetilde{M_d}\right) \\
&= M_{XZ}^{-1} M_{\overline{y}XZ} - M_{XZ}^{-1}\widetilde{M_d} + \widetilde{B}_d\left(M_{\overline{y}XZ} - \widetilde{M_d}\right).
\end{aligned}
$$

Define the perturbation between the observable matrix $M_{XZ}^{-1} M_{\overline{y}XZ}$ and the eigenvalue-eigenvector decomposition $M_{X|X^*}^{-1} D_{\overline{y}|X^*} M_{X|X^*}$ to be

$$
(107) \qquad P_{\overline{y}}^{h}(d) = \underbrace{M_{XZ}^{-1} M_{\overline{y}XZ}}_{\text{observable}} - \underbrace{M_{X|X^*}^{-1} D_{\overline{y}|X^*} M_{X|X^*}}_{\text{a diagonal structure}}
$$

$$
(108) \qquad = -M_{XZ}^{-1}\widetilde{M_d} + \widetilde{B}_d\left(M_{\overline{y}XZ} - \widetilde{M_d}\right).
$$

The perturbation matrix is bounded by

$$
\begin{aligned}
(109) \qquad |||P_{\overline{y}}^{h}(d)||| &\leq |||M_{XZ}^{-1}\widetilde{M_d} + \widetilde{B}_d\left(M_{\overline{y}XZ} - \widetilde{M_d}\right)||| \\
&\leq |||M_{XZ}^{-1}\widetilde{M_d}||| + |||\widetilde{B}_d|||\left(|||M_{\overline{y}XZ}||| + |||\widetilde{M_d}|||\right) \\
&\leq |||M_{XZ}^{-1}|||Kh + \frac{|||M_{XZ}^{-1}|||^2 Kh}{1 - |||M_{XZ}^{-1}|||Kh}\left(|||M_{\overline{y}XZ}||| + K^2 h|E(Y)|\right) \equiv \overline{P_{\overline{y}}^{h}}.
\end{aligned}
$$

Assume the condition in Assumption 2.9 to avoid repeated eigenvalues, and maintain the ordering of the eigenvectors by Assumption 2.4 to identify the eigenvalue-eigenvector decomposition in $M_{X|X^*}^{-1} D_{\overline{y}|X^*} M_{X|X^*}$. To apply Theorem 2.1 to the regression models with misclassification error, we use

$$
(110) \qquad \widetilde{A} = M_{X|X^*}^{-1} D_{\overline{y}|X^*} M_{X|X^*},
$$

$$
(111) \qquad \widetilde{A} + P_{\overline{y}}^{h}(d) = M_{XZ}^{-1} M_{\overline{y}XZ}.
$$

In this application, the corresponding diagonal matrix is $D_{\overline{y}|X^*}$ and the matrix of eigenvectors is the misclassification probability matrix $M_{X|X^*}$. The condition number is the same as the condition number $\kappa(M_{X|X^*})$ which is bounded by

$$
(112) \qquad 1 \leq \kappa(M_{X|X^*}) \leq \overline{\kappa_\lambda},
$$

where the upper bound $\overline{\kappa_\lambda}$ is defined in Equation (71) for Theorem A.1. *Q.E.D.*

References

Adams, C. P. (2016). Finite mixture models with one exclusion restriction. *Econometrics Journal*, 19(2):150–165.

Aguiar, V. H. and Serrano, R. (2017). Slutsky matrix norms: The size, classification, and comparative statics of bounded rationality. *Journal of Economic Theory*, 172:163–201.

Aigner, D. (1973). Regression with a binary independent variable subject to errors of observation. *Journal of Econometrics*, 1(1):49–59.

Andrew, A. L., Chu, K.-W. E., and Lancaster, P. (1993). Derivatives of eigenvalues and eigenvectors of matrix functions. *SIAM Journal on Matrix Analysis and Applications*, 14(4):903–926.

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

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

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

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

Chandrasekhar, A., Chernozhukov, V., Molinari, F., and Schrimpf, P. (2012). Inference for best linear approximations to set identified functions. *CeMMAP Working Papers*.

Chen, X., Christensen, T. M., and Tamer, E. (2018). Monte carlo confidence sets for identified sets. *Econometrica*, 86(6):1965–2018.

Chen, X., Hansen, L. P., and Hansen, P. (2021). Robust inference for moment condition models without rational expectations. *Working Paper*.

Chen, X., Hu, Y., and Lewbel, A. (2009). Nonparametric identification and estimation of nonclassical errors-in-variables models without additional information. *Statistica Sinica*, 19:949–968.

Chernozhukov, V., Han, H., and Tamer, E. (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, 75(5):1243–1284.

Chesher, A. (2013). Semiparametric structural models of binary response: Shape restrictions and partial identification. *Econometric Theory*, 29(02):231–266.

Fang, Z. and Santos, A. (2019). Inference on directionally differentiable functions. *The Review of Economic Studies*, 86(1):377–412.

Gundersen, C., Kreider, B., and Pepper, J. (2012). The impact of the national school lunch program on child health: A nonparametric bounds analysis. *Journal of Econometrics*, 166(1):79–91.

Horn, R. and Johnson, C. (2013). *Matrix Analysis, second edition*. Cambridge University Press.

Horowitz, J. and Manski, C. (1995). Identification and robustness with contaminated and corrupted data. *Econometrica*, 68:281–302.

Hu, Y. (2008). Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *Journal of Econometrics*, 144(1):27–61.

Hu, Y. (2017). The econometrics of unobservables: Applications of measurement error models in empirical industrial organization and labor economics. *Journal of Econometrics*, 200(2):154–168.

Imbens, G. and Manski, C. (2004). Confidence intervals for partially identified parameters. *Econometrica*, 72:1845–1857.

Imbens, G. W. (2003). Sensitivity to exogeneity assumptions in program evaluation. *American Economic Review*, 93(2):126–132.

Kane, T. J., Rouse, C. E., and Staiger, D. (1999). Estimating returns to schooling when schooling is misreported. *NBER Working Paper #7235*.

Klepper, S. (1988). Bounding the effects of measurement error in regressions involving dichotomous variables. *Journal of Econometrics*, 37(3):343–359.

Kline, P. and Santos, A. (2013). Sensitivity to missing data assumptions: Theory and an evaluation of the us wage structure. *Quantitative Economics*, 4(2):231–267.

Lewbel, A. (2007). Estimation of average treatment effects with misclassification. *Econometrica*, 75(2):537–551.

Li, Q. and Racine, J. (2003). Nonparametric estimation of distributions with categorical and continuous data. *Journal of Multivariate Analysis*, 86(2):266–292.

Li, Q. and Racine, J. S. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Magnac, T. and Maurin, E. (2008). Partial identification in monotone binary models: Discrete regressors and interval data. *Review of Economic Studies*, (75):835–864.

Mahajan, A. (2006). Identification and estimation of regression models with misclassification. *Econometrica*, 74(3):631–665.

Manski, C. (2003). *Partial Identification of Probability Distributions*. Springer-Verlag, New York.

Manski, C. F. (2009). *Identification for Prediction and Decision*. Harvard University Press.

Manski, C. F. and Pepper, J. V. (2018). How do right-to-carry laws affect crime rates? coping with ambiguity using bounded-variation assumptions. *Review of Economics and Statistics*, 100(2):232–244.

Masten, M. and Poirier, A. (2016). Partial independence in nonseparable models. *CeMMAP Working Papers*.

Masten, M. and Poirier, A. (2018). Identification of treatment effects under conditional partial independence. *Econometrica*, 86(1):317–351.

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

Molinari, F. (2003). Contaminated, corrupted, and missing data. *Ph.D. Dissertation, Northwestern University.*

Molinari, F. (2008). Partial identification of probability distributions with misclassified data. *Journal of Econometrics*, 144:81–117.

Poterba, J. M. and Summers, L. H. (1995). Unemployment benefits and labor market transitions: A multinomial logit model with errors in classification. *The Review of Economics and Statistics*, 77(2):207–216.

Ramalho, E. (2002). Regression models for choice-based samples with misclassification in the response variable. *Journal of Econometrics*, 106(1):171–201.

Stoye, J. (2010). Partial identification of spread parameters. *Quantitative Economics*, 1(2):323–357.

Walter, S. and Hui, S. (1980). Estimating the error rates of diagnostic tests. *Biometrics*, 36(1):167–171.

Table 1: Measures of Sensitivity Parameter in Binary Choice Models

$n = 1000$	h_M	$\widehat{\lambda}$	h_I	h_E	h_S	h_{Dx^*}	h_{R1x^*}	h_0
DGP I: $(\gamma,\alpha) = (0,0)$	0.016	0.100	2.382	0.008	0.008	0.011	0.008	0.001
DGP II: $(\gamma,\alpha) = (0,0.15)$	0.021	0.100	2.382	0.009	0.009	0.011	0.009	0.001
DGP III: $(\gamma,\alpha) = (0.15,0)$	0.039	0.161	2.230	0.008	0.009	0.011	0.008	0.003
DGP IV: $(\gamma,\alpha) = (0,0.3)$	0.016	0.100	2.382	0.010	0.010	0.010	0.010	0.001
DGP V: $(\gamma,\alpha) = (0.3,0)$	0.076	0.209	2.131	0.008	0.009	0.012	0.008	0.004
DGP VI: $(\gamma,\alpha) = (0.3,0.3)$	0.087	0.209	2.131	0.011	0.012	0.012	0.011	0.004
DGP VII: $(\gamma,\alpha) = (0.4,0.4)$	0.137	0.256	2.050	0.012	0.012	0.012	0.012	0.004
DGP VIII: $(\gamma,\alpha) = (0.5,0.5)$	0.184	0.291	1.997	0.012	0.013	0.008	0.008	0.004
DGP IX: $(\gamma,\alpha) = (0.55,0.55)$	0.207	0.311	1.969	0.013	0.013	0.007	0.007	0.005
DGP X: $(\gamma,\alpha) = (0.6,0.6)$	0.221	0.313	1.966	0.013	0.013	0.004	0.004	0.004

Note: The definitions of the measures can be found in subsection 5.1.

Table 2: Measures of Sensitivity Parameter From Breakdown Frontier Analysis

$n = 1000$	$bf(\lambda = 0.1; \underline{\mu})$	h_{BF0}	C.I. of $bf(\lambda = 0.1; \underline{\mu})$	h_{Rx^*}	$h_{Rx^*} > h_0$?
DGP I: $(\gamma,\alpha) = (0,0)$	0.009	0.012	$(0.007,\infty)$	0.007	Yes
DGP II: $(\gamma,\alpha) = (0,0.15)$	0.012	0.016	$(0.010,\infty)$	0.009	Yes
DGP III: $(\gamma,\alpha) = (0.15,0)$	0.008	0.011	$(0.006,\infty)$	0.006	Yes
DGP IV: $(\gamma,\alpha) = (0,0.3)$	0.015	0.020	$(0.013,\infty)$	0.010	Yes
DGP V: $(\gamma,\alpha) = (0.3,0)$	0.008	0.011	$(0.006,\infty)$	0.006	Yes
DGP VI: $(\gamma,\alpha) = (0.3,0.3)$	0.015	0.020	$(0.013,\infty)$	0.011	Yes
DGP VII: $(\gamma,\alpha) = (0.4,0.4)$	0.017	0.022	$(0.015,\infty)$	0.012	Yes
DGP VIII: $(\gamma,\alpha) = (0.5,0.5)$	0.019	0.025	$(0.016,\infty)$	0.008	Yes
DGP IX: $(\gamma,\alpha) = (0.55,0.55)$	0.019	0.026	$(0.017,\infty)$	0.007	Yes
DGP X: $(\gamma,\alpha) = (0.6,0.6)$	0.020	0.027	$(0.018,\infty)$	0.004	No

Note: The 95% confidence intervals of the breakdown frontiers at $\lambda = 0.1$ are computed by the bootstrap estimates across 1000 simulations. Denote h_{BF0} as the breakdown frontier $bf(\lambda; \underline{\mu})$ at $\lambda = 0$.

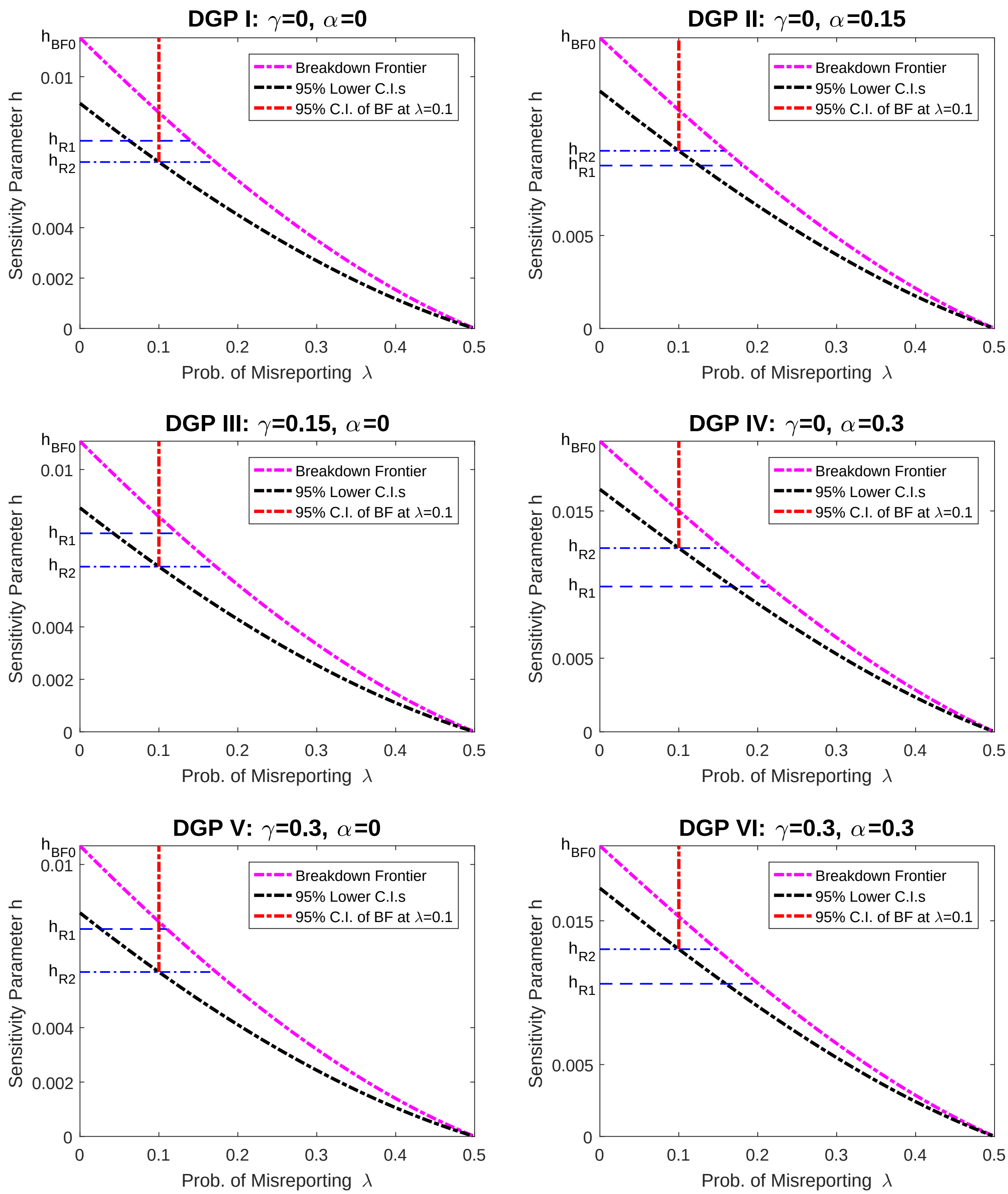

Figure 1: The Illustration of the Breakdown Frontiers of Binary Choice Models

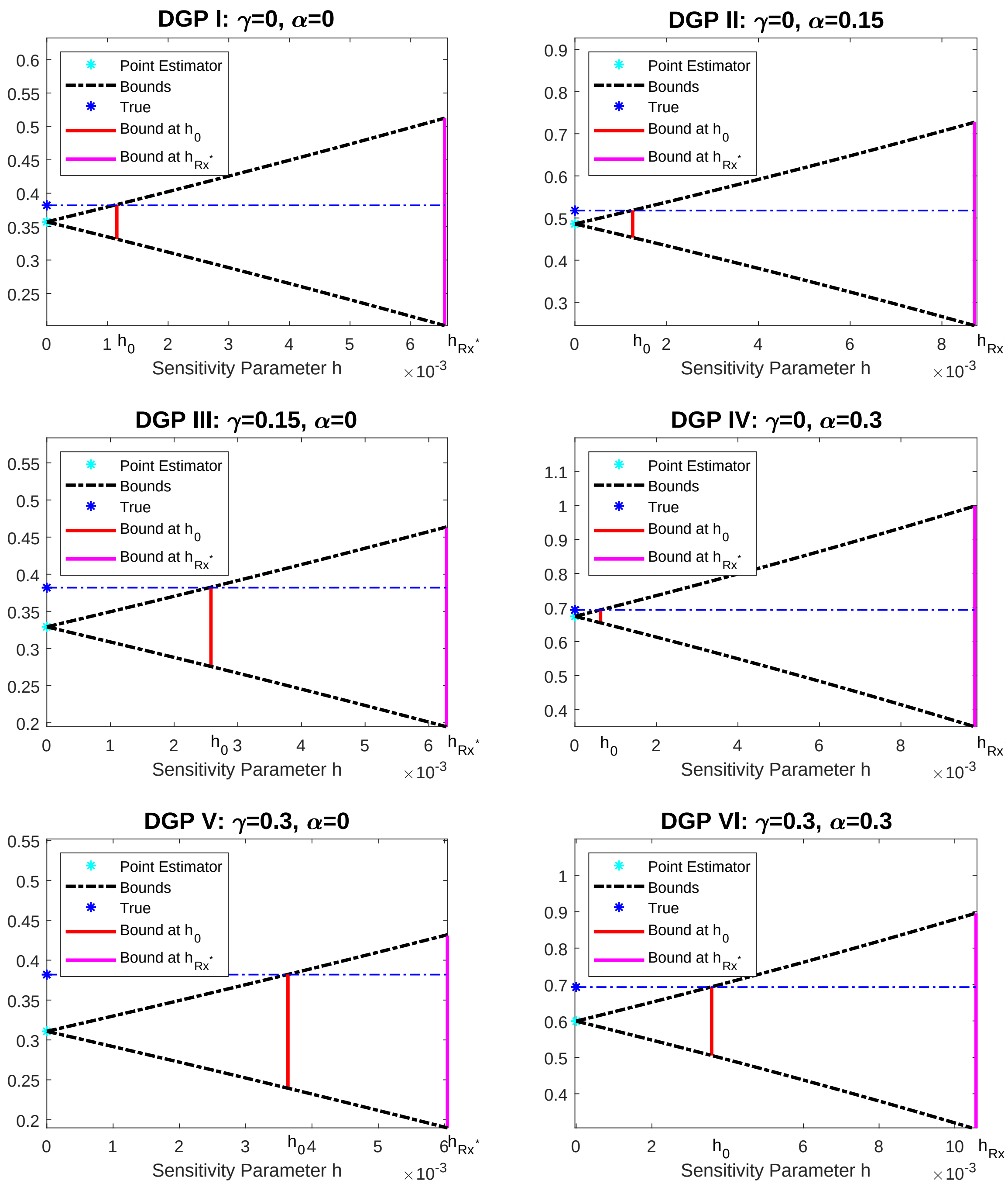

Figure 2: The Illustration of the Bounds of Binary Choice Models

Table 3: Measures of Sensitivity Parameter in Regression Models

$n = 2000$	$\tilde{h}_M$	$\widehat{\lambda}$	h_I	h_E	h_S	h_{R1x^*}	h_0
DGP I: $\gamma = 0$	0.004	0.113	5.144	0.005	0.005	0.005	0.001
DGP II: $\gamma = 0.2$	0.079	0.228	3.127	0.009	0.013	0.009	0.003
DGP III: $\gamma = 0.3$	0.126	0.305	2.747	0.011	0.016	0.011	0.004
DGP IV: $\gamma = 0.4$	0.169	0.374	2.519	0.012	0.019	0.012	0.005
DGP V: $\gamma = 0.45$	0.184	0.398	2.453	0.012	0.020	0.012	0.005
DGP VI: $\gamma = 0.5$	0.194	0.415	2.412	0.012	0.021	0.012	0.005
DGP VII: $\gamma = 0.55$	0.202	0.428	2.380	0.013	0.021	0.013	0.005
DGP VIII: $\gamma = 0.6$	0.206	0.434	2.366	0.013	0.021	0.013	0.005
DGP IX: $\gamma = 0.625$	0.206	0.434	2.366	0.013	0.021	0.013	0.005

Note: The definitions of the measures can be found in subsection 5.1.

Table 4: Measures of Sensitivity Parameter From Breakdown Frontier Analysis

$n = 2000$	$bf(\lambda = 0.1; \underline{\mu})$	h_{BF0}	C.I. of $bf(\lambda = 0.1; \underline{\mu})$	h_{Rx^*}	$h_{Rx^*} > h_0$?
DGP I: $\gamma = 0$	0.004	0.005	(0.002,0.004)	0.002	Yes
DGP II: $\gamma = 0.2$	0.011	0.014	(0.009,0.011)	0.009	Yes
DGP III: $\gamma = 0.3$	0.014	0.018	(0.012,0.014)	0.011	Yes
DGP IV: $\gamma = 0.4$	0.016	0.021	(0.014,0.016)	0.012	Yes
DGP V: $\gamma = 0.45$	0.016	0.021	(0.015,0.016)	0.012	Yes
DGP VI: $\gamma = 0.5$	0.017	0.022	(0.015,0.017)	0.012	Yes
DGP VII: $\gamma = 0.55$	0.017	0.022	(0.016,0.017)	0.013	Yes
DGP VIII: $\gamma = 0.6$	0.017	0.023	(0.016,0.017)	0.013	Yes
DGP IX: $\gamma = 0.625$	0.017	0.023	(0.016,0.017)	0.013	Yes

Note: The 95% confidence intervals of the breakdown frontiers at $\lambda = 0.1$ are computed by the bootstrap estimates across 1000 simulations. Denote h_{BF0} as the breakdown frontier $bf(\lambda; \underline{\mu})$ at $\lambda = 0$.

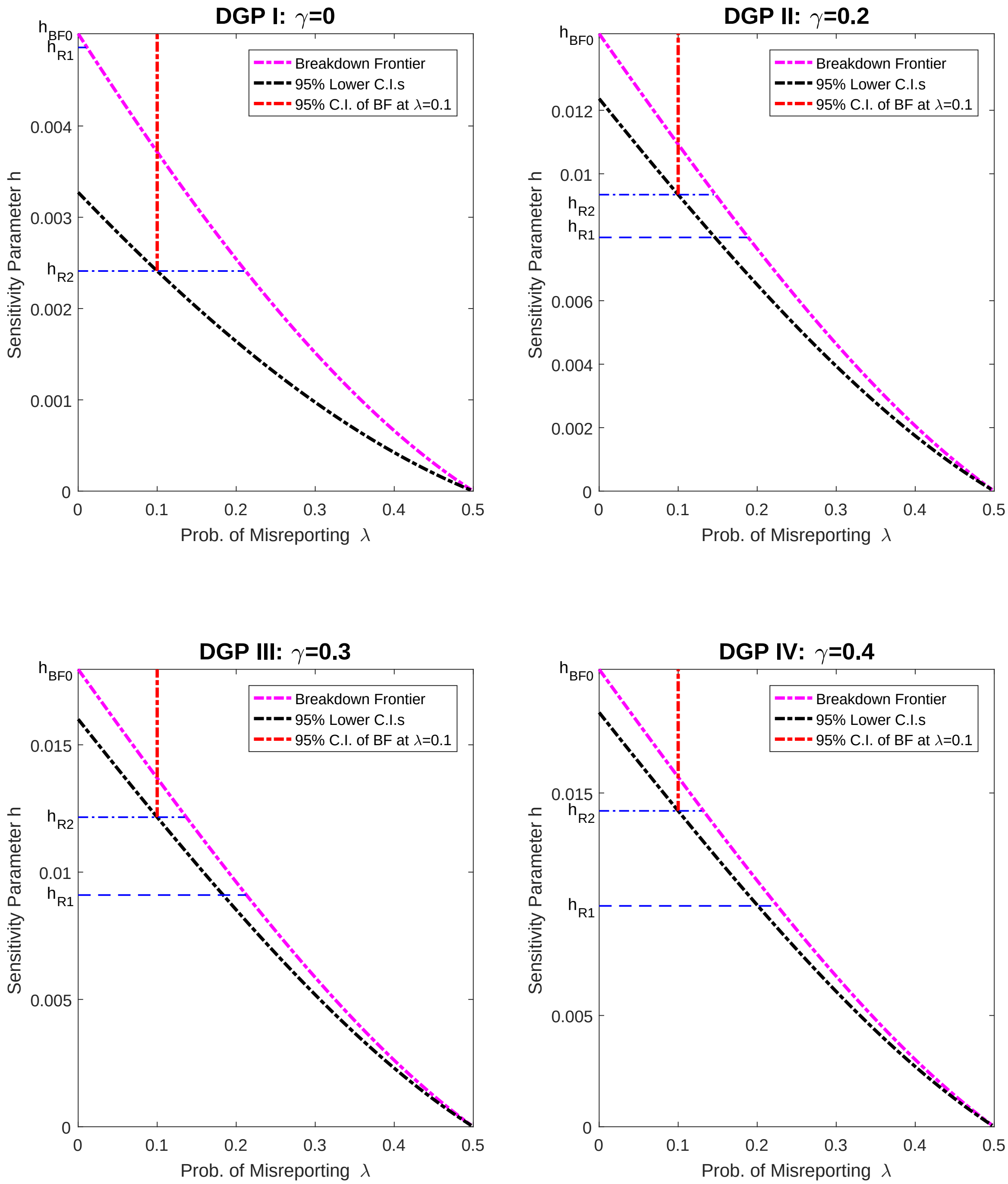

Figure 3: The Illustration of the Breakdown Frontiers of Regression Models

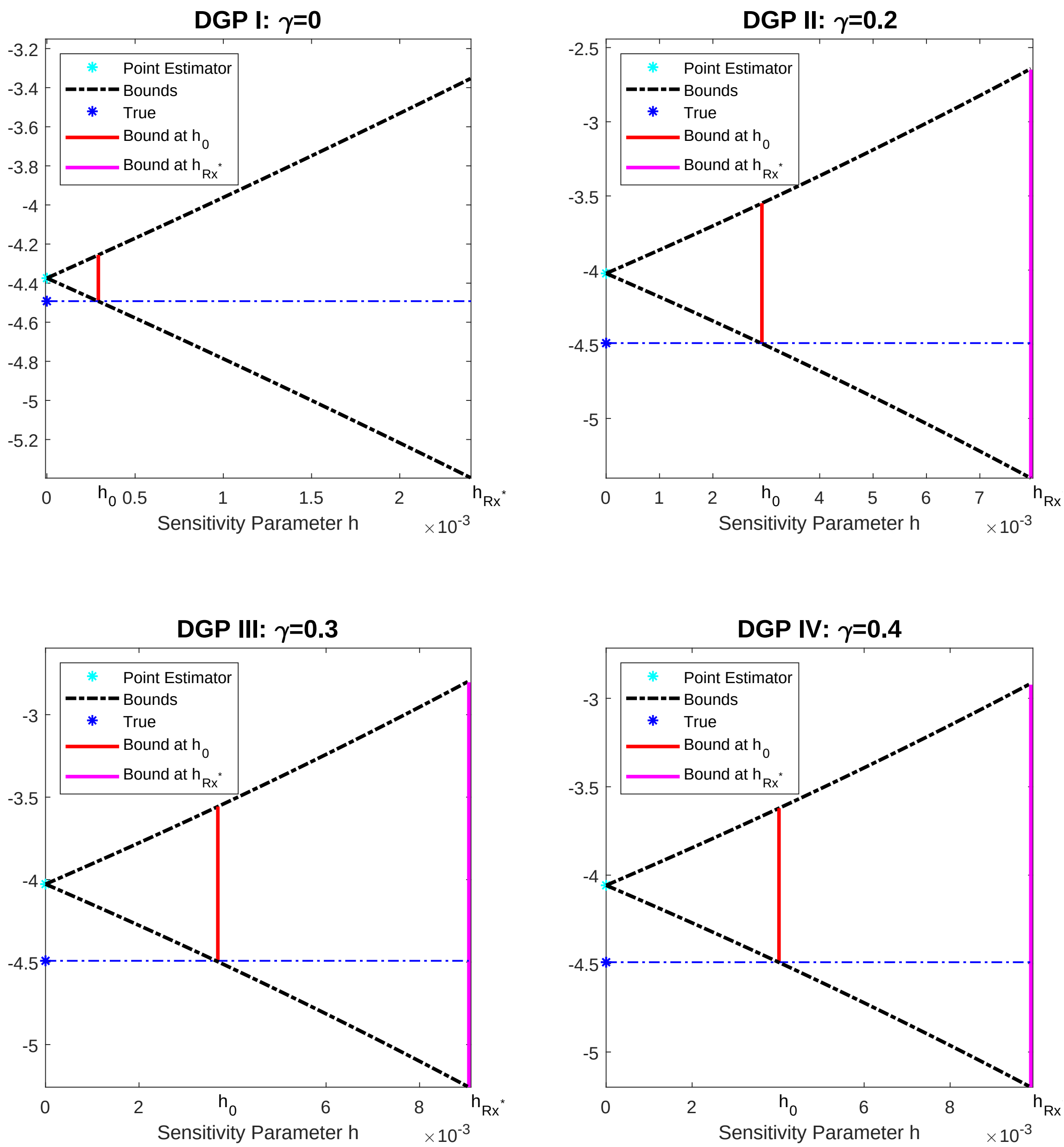

Figure 4: The Illustration of the Bounds of Regression Models

Table 5: Sample Proportions and Mean Log Wages in 1986

Sample Proportions:				
	Self-Reported Schooling:			
Transcript-recorded	*High School*	*Some College*	*Bachelor*	Row Total
High School	0.288	0.060	0.015	0.362
Some College	0.034	0.250	0.025	0.309
Bachelor	0.000	0.005	0.324	0.329
Column Total	0.322	0.332	0.363	1.000

Mean Log Wages in 1986:				
	Self-Reported Schooling:			
Transcript-recorded	*High School*	*Some College*	*Bachelor*	Row Total
High School	2.026	2.104	2.330	2.051
	(0.497)	(0.494)	(0.609)	(0.506)
Some College	2.118	2.206	2.375	2.210
	(0.453)	(0.488)	(0.516)	(0.492)
Bachelor	2.460	2.310	2.446	2.444
	(0.350)	(0.417)	(0.495)	(0.494)
Column Total	2.036	2.188	2.436	2.229
	(0.495)	(0.490)	(0.502)	(0.524)

1. Educational attainment was measured in 1979, and average log hourly wages were observed in 1986. The sample size is 9261. 2. Source: NLS-72 and PETS.

Table 6: Estimation of Mean Log Wage

	High School	*Some College*	*Bachelor*
Point Estimation $E(Y\|X^*)$	2.025^{***}	2.208^{***}	2.446^{***}
	(0.010)	(0.011)	(0.009)
95% Confidence Interval	(2.005,2.045)	(2.186,2.230)	(2.428,2.464)

1. Standard errors (calculated by bootstrap) are in parentheses. 2. Symbols *** indicate that the test is significant at a level of 1%.

Table 7: The Bounds of Mean Log Wage of True Education

	High School	*Some College*	*Bachelor*
Bound	(1.934,2.116)	(2.116,2.299)	(2.355,2.537)
95% C.I. of Bound	(1.906,2.129)	(2.102,2.329)	(2.334,2.559)
Sensitivity Parameter h_{R1}		8.568×10^{-4}	

Note: Bootstrap estimates compute the 95% confidence intervals of the bounds across 1000 simulations.

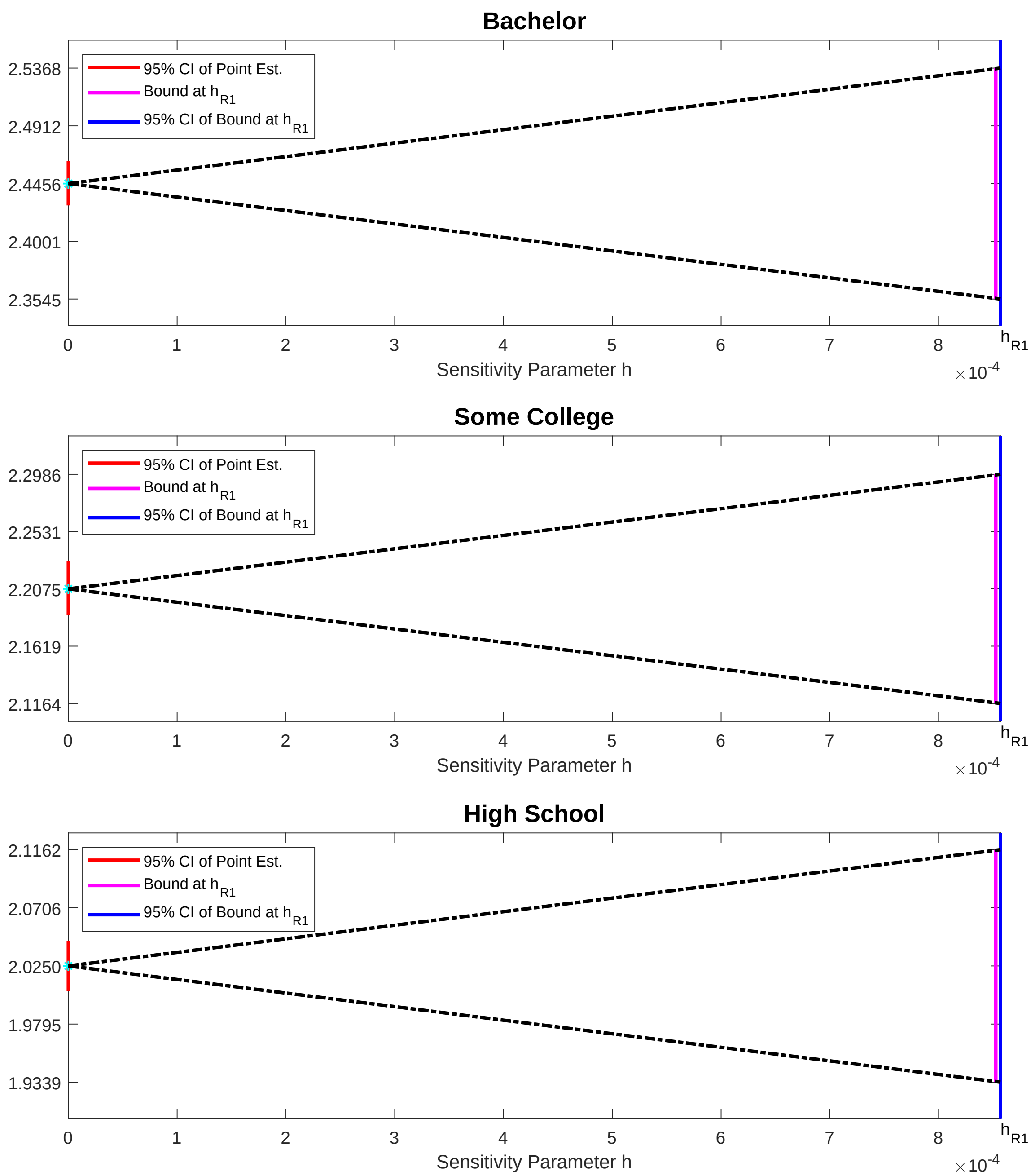

Figure 5: The Illustration of the Bounds of Mean Log Wage