Chalak, K., & Kim, D. (2019). Measurement Error without the Proxy Exclusion Restriction. *Journal of Business & Economic Statistics*, 1-44.

https://doi.org/10.1080/07350015.2019.1617156

Measurement Error without the Proxy Exclusion Restriction

Karim Chalak*†
University of Virginia

Daniel Kim
BI Norwegian Business School

April 26, 2019

Abstract

This paper studies the identification of the coefficients in a linear equation when data on the outcome, covariates, and an error-laden proxy for a latent variable are available. We maintain that the measurement error in the proxy is classical and relax the assumption that the proxy is excluded from the outcome equation. This enables the proxy to directly affect the outcome and allows for differential measurement error. Without the proxy exclusion restriction, we first show that the effects of the latent variable, the proxy, and the covariates are not identified. We then derive the sharp identification regions for these effects under any configuration of three auxiliary assumptions. The first weakens the assumption of no measurement error by imposing an upper bound on the noise to signal ratio. The second imposes an upper bound on the outcome equation coefficient of determination that would obtain had there been no measurement error. The third weakens the proxy exclusion restriction by specifying whether the latent variable and its proxy affect the outcome in the same or the opposite direction, if at all. Using the College Scorecard aggregate data, we illustrate our framework by studying the financial returns to college selectivity and characteristics and student characteristics when the average SAT score at an institution may directly affect earnings and serves as a proxy for the average ability of the student cohort.

JEL codes: C21, I23.

Keywords: *college selectivity, college characteristics, endogeneity, exclusion restriction, differential measurement error, partial identification, proxy, sensitivity analysis.*

---

*Karim Chalak (corresponding author), Department of Economics, University of Virginia, chalak@virginia.edu. Daniel Kim, The Wharton School, University of Pennsylvania, kimdanie@wharton.upenn.edu.

†Acknowledgments: We thank the editor, associate editor, and three referees for their constructive comments and suggestions. We also thank the participants in the New York Camp Econometrics XII, the 10^{th} West Coast Experiments Conference, 2017 North American Summer Meeting of the Econometric Society, 70^{th} European Meeting of the Econometric Society, 27^{th} Annual Meeting of the Midwest Econometrics Group, 87^{th} Annual Meeting of the Southern Economic Association, 2017 Triangle Econometrics Conference, and the seminars at Boston College, University of Virginia, SUNY Binghamton, Emory, and Oxford as well as Jennifer Jones and John Pepper for their helpful feedback. Karim Chalak acknowledges the support from the Bankard Fund for Political Economy. Any errors are the responsibility of the authors.

1 Introduction

In many studies, the researchers do not observe a key explanatory variable U and employ a useful proxy W in its place. For example, when estimating wage or earnings equations, a test score is often used as a proxy for the unobserved individual "ability" (see e.g. Neal and Johnson, 1996). When the equation for the outcome Y is linear and the proxy W suffers from classical measurement error, a regression of Y on W and the correctly measured covariates X does not point identify the effects of U or X on Y, except in special cases. In particular, the regression estimand for the effect of U suffers from "attenuation bias." Nevertheless, a quintessential result establishes sharp bounds for the coefficients on U and X (e.g. Gini, 1921; Frisch, 1934; Klepper and Leamer, 1984; Bollinger, 2003). These bounds can be informative in several empirical contexts. For example, Bollinger (2003) reports bounds on the black-white wage gap when a test score serves as an error-laden proxy for ability.

These standard bounds are valid under the assumption that, unlike U, the proxy W is excluded from the equation for Y. This is similar to the standard assumption that an instrumental variable is excluded from the outcome equation. As illustrated below, sometimes the proxy for the latent variable may directly affect the outcome. What can be learned about the effects of U, W, and X on Y if the available proxy W for the latent explanatory variable U also affects the outcome Y? To address this question, the paper characterizes the sharp bounds on these effects under assumptions that allow W to be included in the outcome equation. This is akin to studying the consequences of weakening the exclusion restriction imposed on an instrumental variable (e.g. Conley, Hansen, and Rossi, 2012).

Without the proxy exclusion restriction, the measurement error is "differential" since the proxy may help predict the outcome even after conditioning on the latent variable. Several key identification results in the literature maintain that the measurement error is "non-differential" (see e.g. Chesher (1991) and Chen, Hong, and Nekipelov (2011, assumption 2.1)). This assumption posits that the distributions of $Y|(W, U, X)$ and $Y|(U, X)$ (or e.g. their means) coincide. We relax this assumption in the context of a linear specification that allows W to directly enter the Y equation. In this sense, the paper puts forward partial identification results that enable inference in a leading setting for differential measurement error that "occurs when W is not merely a mismeasured version of [U], but is a separate variable acting as a type of proxy for [U]" (Carroll, Ruppert, Stefanski, and Crainiceanu, 2006, p. 36), as occurs in the examples below.

In particular, we characterize the joint sharp identification region for the direct and total (direct and mediated via W) effect of U on Y, the direct effects of W and X on Y, and the (net-of-X) "signal to total variance ratio" (the ratio of the (net-of-X) variances of U and W). We then show that none of these parameters are separately identified since projecting the joint identification region onto the supports of each of its components yields the full support. This demonstrates the crucial role that the proxy exclusion restriction plays in ensuring the validity of the standard bounds discussed above. To proceed, we derive the joint and projected sharp identification regions under any configuration of three auxiliary assumptions. The first weakens the benchmark assumption of "no measurement error" by imposing an upper bound on the (net-of-X) "noise to signal" ratio (the ratio of the (net-of-X) variances of the measurement error and U). The second imposes an upper bound on the outcome equation coefficient of determination that would obtain had W measured U without error. The third weakens the proxy exclusion restriction by specifying whether the latent variable U and its proxy W affect the outcome in the same or the opposite direction, if at all. We do not require particular auxiliary assumptions; rather, we establish the mapping from each configuration of these assumptions to the sharp identification regions. By varying the two upper bounds and the sign restriction in these three auxiliary assumptions, a researcher can conduct a sensitivity analysis of how the measurement error in the proxy, the fit of the model, and the proxy exclusion restriction affect the sharp identification regions.

After discussing estimation and inference, we illustrate our results by studying the financial returns to the college and student characteristics. Specifically, we analyze the recently released College Scorecard (CS) data which reports information on postsecondary institutions in the US. CS is aggregated at the institution level and includes information on the institution, students, affordability, admission and academic attributes, and earnings outcomes. While CS has some limitations that are partly due to data aggregation, it is "the first nationally comprehensive data on students' post-enrollment earnings, measured for a consistently defined set of students at nearly all post-secondary institutions in the United States" (Council of Economic Advisors, 2015). We use a parsimonious specification for a student's earnings as a function of his or her individual characteristics, the college's characteristics, including its selectivity measured by the average SAT (equivalent) score of the student's cohort, and the student's unobserved scholastic "ability." We allow a student's ability to freely depend on his or her observed characteristics as well as the characteristics of the college he or she attended. We then study the consequences of deviating from the "selection on observables" assumption by allowing a student's SAT (equivalent) score to serve as an error-laden proxy for his or her ability, with classical measurement error. Because CS reports only aggregate data, we average the earnings equation across students in each college. The average SAT score now serves as an error-laden proxy for the average latent ability and is included in the average earnings equation, thereby violating the proxy exclusion restriction. We apply the paper's framework and obtain informative bounds on the earnings equation coefficients and study their sensitivity to the three auxiliary assumptions on the extent of the measurement error in the average SAT score, the fit of the model, and the signs of the returns to the student's ability and the selectivity of the college he or she attended.

More broadly, the paper's results are useful in any setting where one suspects that the proxy for the latent variable may directly affect the outcome. For example, consumers may not fully observe the quality of a product (e.g. financial asset or movie) and their demand may be influenced by a product rating (e.g. asset rating or movie score (e.g. Rotten Tomatoes)) that the econometrician uses as a proxy for quality. Also, a medical test result that serves as a proxy for the unobserved health status of a patient may directly affect the patient's behavior (e.g. a worker may work shorter hours if he or she is incorrectly prescribed rest). Another example occurs in studies of state-level corruption in the US, when convictions of public officials for past corruption serves as a proxy for current corruption and can affect e.g. growth or environmental policy by entailing a change in the bureaucracy. Last, aggregate variables (e.g. income in a neighborhood) that serve as proxies for socioeconomic individual variables (e.g. individual income) can directly affect the outcome (e.g. individual health) if there are "contextual effects" (see e.g. Geronimus, Bound, and Neidert (1996) and Bound, Brown, and Mathiowetz (2001, footnote 8)).

The paper is organized as follows. Section 2 specifies the assumptions and notation. Section 3 characterizes the sharp identification regions when none, some, or all of the auxiliary assumptions are imposed. Section 4 provides a numerical example. Section 5 discusses estimation and inference. Section 6 contains the empirical application. Section 7 concludes.

The Supplement gathers the mathematical proofs and additional results.

2 Data Generation and Assumptions

We consider the following data generating structural system.

Assumption $\mathbf{A_1}$ *Data Generation: (i) Let* $(\underset{k\times 1}{X}', \underset{1\times 1}{W}, \underset{1\times 1}{Y})'$ *be a random vector with a finite variance. (ii) Let a structural system, with constant slope coefficients, generate the random vector* X *and variables* η, ε, U, W, *and* Y *such that*

$$Y = X'\beta + W\phi + U\delta + \eta \qquad \textit{and} \qquad W = U + \varepsilon. \tag{1}$$

The researcher observes realizations of $(X', W, Y)'$ *but not of* (η, ε, U).

We maintain two standard assumptions on the unobservables η and ε. A_2 assumes that the "disturbance" η is uncorrelated with $(X', U)'$.

Assumption $\mathbf{A_2}$ *Uncorrelated Disturbance:* $Cov[\eta, (X', U)'] = 0$.

Further, A_1 decomposes the proxy W into the "signal" component U and the "noise" or error ε and A_3 assumes that the measurement error ε is uncorrelated with (X', U, η).

Assumption $\mathbf{A_3}$ *Uncorrelated measurement error:* $Cov[\varepsilon, (X', U, \eta)'] = 0$.

We are interested in identifying ϕ, δ, $\phi + \delta$, and β. The slope coefficients ϕ, δ, and β are the ceteris paribus causal effects of the proxy W, the latent variable U, and the covariates X on the outcome Y respectively. The sum $\phi + \delta$ is the total (direct and mediated via W) effect of U on Y. One identification challenge is due to U being unobserved and correlated with W and possibly X. Moreover, we only require the uncorrelation assumptions A_2 and A_3 and do not impose stronger (e.g. mean) independence assumptions. In particular, A_2 suffices for a linear regression of Y on (X, U) to point identify $\phi + \delta$ and β had U been observed without error (i.e. $\varepsilon = 0$) and A_3 relaxes the assumption that $\varepsilon = 0$.

A_1 allows, but does not require, the proxy W to directly affect Y. When $\phi = 0$, A_1-A_3 are the classical error-in-variables assumptions (see e.g. Wooldridge, 2002, p. 80). We relax these benchmark assumptions by studying the consequences of deviating from the exclusion restriction $\phi = 0$ on the identification of ϕ, δ, $\phi + \delta$, and β. Relaxing $\phi = 0$ leads to a second identification challenge. In particular, it is widely assumed in the literature that the measurement error is "nondifferential" so that $E(Y|X,W,U) = E(Y|X,U)$ (see e.g. Bollinger, 1996; Mahajan, 2006; Lewbel, 2007; Hu, 2008; Wooldridge (2002, p. 79) refers to this as the "redundancy condition"). Incorrectly assuming that the measurement error is nondifferential may result in misleading inference on δ and β. Bound, Brown, and Mathiowetz (2001, p. 3717) discuss several examples that "highlight the potential importance of differential measurement error." Here, we have

$$E(Y|X,W,U) - E(Y|X,U) = [\varepsilon - E(\varepsilon|X,U)]\phi + E(\eta|X,W,U) - E(\eta|X,U)$$

so that, even when $E(\eta|X,W,U) = E(\eta|X,U)$ and $E(\varepsilon|X,U) = 0$, $E(Y|X,W,U)$ differs from $E(Y|X,U)$ by $\varepsilon\phi$ and the measurement error is differential.

Last, we briefly comment on some related papers that modify A_2-A_3. Under $\phi = 0$, Erickson (1993) weakens A_3 by imposing bounds on $Corr(\varepsilon, \eta)$, Hyslop and Imbens (2001) replace A_3 with the assumption that W is an optimal prediction of U so that ε is uncorrelated with W and correlated with U, and DiTraglia and Garcia-Jimeno (2016) weaken A_2 to allow $Cov(U, \eta)$ to be nonzero. In contrast, Lewbel (1997) and Erickson and Whited (2002) maintain $\phi = 0$ and strengthen A_2-A_3 by imposing restrictions on the higher order moments of η, ε, U, and X that can point identify $(\beta', \delta)'$. Last, recall that $Cov(\varepsilon, U) = 0$ in A_3 generally rules out that U and W are binary variables. Imai and Yamamoto (2010) study bounding the average effect of a binary misclassified treatment on a binary outcome under alternative assumptions on the differential measurement error.

2.1 Notation and Linear Projection

To shorten the notation, for generic random vectors A and B, we write:

$$\sigma_A^2 \equiv Var(A) \quad \text{and} \quad \sigma_{A,B} \equiv Cov(A, B).$$

Further, we use a concise notation for the linear regression estimand and residual

$$b_{A.B} \equiv \sigma_B^{-2}\sigma_{B,A} \qquad \text{and} \qquad \epsilon'_{A.B} \equiv [A - E(A)]' - [B - E(B)]' b_{A.B}$$

so that by construction $E(\epsilon_{A.B}) = 0$ and $Cov(B, \epsilon_{A.B}) = 0$. For example, $b_{Y.X}$ is the vector of slope coefficients associated with X in a linear regression of Y on $(1, X')'$. Last, for a scalar A, we let $R^2_{A.B} \equiv \sigma_A^{-2}(\sigma_{A,B}\sigma_B^{-2}\sigma_{B,A})$ denote the population coefficient of determination (R-squared) from a regression of A on B (if $\sigma^2_A = 0$ set $R^2_{A.B} \equiv 0$).

Under A_1-A_3, $Cov[(\eta, \varepsilon)', X] = 0$. Thus, provided σ^2_X is nonsingular, by substituting for $U = W - \varepsilon$ in the Y equation we obtain

$$b_{Y.X} = \beta + b_{W.X}(\phi + \delta). \tag{2}$$

Using the shorthand notation $\tilde{A} \equiv \epsilon_{A.X}$ for the residuals from a regression of A on $(1, X')'$, we employ the convenient system of projected linear equations:

$$\tilde{Y} = \tilde{W}\phi + \tilde{U}\delta + \tilde{\eta} \quad \text{and} \quad \tilde{W} = \tilde{U} + \tilde{\varepsilon}, \tag{3}$$

in order to study the identification of ϕ, δ, and $\phi + \delta$. The identification region for β then obtains from the identification region for $\phi + \delta$ using equation (2).

2.2 Auxiliary Assumptions

We also study the identification gain that results from imposing any configuration of three auxiliary assumptions A_4-A_6. The first weakens the standard "no measurement error" assumption $\sigma^2_\varepsilon = 0$ by imposing an upper bound κ on the net-of-X noise to signal ratio $\frac{\sigma^2_\varepsilon}{\sigma^2_{\tilde{U}}}$.

Assumption A_4 *Bounded Net-of-X Noise to Signal Ratio: $\sigma^2_\varepsilon \leq \kappa\sigma^2_{\tilde{U}}$ where $0 \leq \kappa$.*

For example, setting $\kappa = 0$ yields the no measurement error assumption $\sigma^2_\varepsilon = 0$ and setting $\kappa = 1$ assumes that the variance of the measurement error is at most as large as the variance of $\tilde{U}$, $\sigma^2_\varepsilon \leq \sigma^2_{\tilde{U}}$. By A_1-A_3, we have $\sigma^2_{\tilde{W}} = \sigma^2_{\tilde{U}} + \sigma^2_\varepsilon$. It follows that A_4 sets a lower bound $\frac{1}{1+\kappa}$ on ρ, the net-of-X "signal to total variance ratio":

$$\frac{1}{1+\kappa} \leq \rho \equiv \frac{\sigma^2_{\tilde{U}}}{\sigma^2_{\tilde{W}}} = \frac{\sigma^2_{\tilde{U}}}{\sigma^2_{\tilde{U}} + \sigma^2_\varepsilon}.$$

Since $\rho \equiv \frac{\sigma^2_{\tilde{U}}}{\sigma^2_{\tilde{W}}} = \frac{R^2_{W.U} - R^2_{W.X}}{1 - R^2_{W.X}}$ (e.g. Dale and Krueger (2002, p. 1514) and DiTraglia and Garcia-Jimeno (2016, eq. (20))), A_4 imposes a lower bound κ' on the "reliability ratio," $\kappa' \equiv \frac{1+\kappa R^2_{W.X}}{1+\kappa} \leq R^2_{W.U}$ where $R^2_{W.X} \leq \kappa'$. A researcher can resort to any of these equivalent interpretations of A_4.

Let $\tilde{R}^2_* \equiv 1 - \frac{\sigma^2_\eta}{\sigma^2_{\tilde{Y}}}$ be the coefficient of determination that would obtain in equations (3) had W measured U without error. By A_1-A_3 and Lemma 1 in the Supplement, $R^2_{\tilde{Y}.\tilde{W}} \leq \tilde{R}^2_*$. The second assumption imposes a bound τ on how large can $\tilde{R}^2_*$ be.

Assumption $\mathbf{A_5}$ *Bounded Net-of-X Coefficient of Determination:* $\tilde{R}^2_* \leq \tau$ *where* $0 < \tau$ *and* $R^2_{\tilde{Y}.\tilde{W}} \leq \tau \leq 1$.

Since $R^2_{A.(X',B)'} = \frac{\sigma^2_{\tilde{A}}}{\sigma^2_A}(R^2_{\tilde{A}.\tilde{B}} - 1) + 1$, A_5 imposes an upper bound $\tau' \equiv \frac{\sigma^2_{\tilde{Y}}}{\sigma^2_Y}(\tau - 1) + 1$ on $R^2_* \equiv 1 - \frac{\sigma^2_\eta}{\sigma^2_Y}$ which would obtain in equations (1) had W measured U without error.

Klepper and Leamer (1984), Bekker, Kapteyn, and Wansbeek (1987), and Klepper (1988) use restrictions similar to A_4 and A_5 when $\phi = 0$. We vary κ and τ in A_4 and A_5 to conduct a sensitivity analysis that weakens the no measurement error assumption $\kappa = 0$ (or $\tau = R^2_{\tilde{Y}.\tilde{W}}$ in A_5) or/and controls the fit of the model ($\tilde{R}^2_* \leq \tau$). Conversely, we study for what value of κ or τ does the identification region admit a plausible value or range for e.g. δ or β. To keep the exposition concise, we impose A_4 and A_5 throughout the analysis and treat the results when A_4 or A_5 is not binding as a special case in which $\kappa \to +\infty$ or $\tau = 1$.

The last auxiliary assumption weakens the proxy exclusion restriction $\phi = 0$ (A_6^0) by specifying whether ϕ and δ have the same or the opposite (weak) sign.

Assumption $\mathbf{A_6}$ *Coefficient Sign Restriction:* $\phi\delta \geq 0$ *(A_6^+),* $\phi\delta \leq 0$ *(A_6^-), or* $\phi = 0$ *(A_6^0).*

Under A_6^+ (A_6^-), U and W affect Y in the same (opposite) direction. For instance, A_6^+ assumes that the average SAT score W (the college selectivity) and the average student ability U affect the mean earnings Y in the same direction. Similarly, a rating W of a financial asset (movie) and the asset's (movie's) quality may affect the demand for the asset (movie) in the same direction. On the other hand, A_6^- assumes that a diabetic patient with a high blood sugar level (U) may feel fatigued and exercise (Y) less ($\delta \leq 0$) but that receiving a high blood sugar test result (W) may affect the patient's exercising positively ($\phi \geq 0$). We note that A_4 and A_6 resemble the assumptions of a maximum misclassification rate and a monotone treatment response used in e.g. Kreider , Pepper, Gundersen, and Jolliffe (2012) and Gundersen, Kreider, and Pepper (2012) to bound the average effect of a binary treatment.

3 Identification

We characterize the sharp identification regions for ϕ, δ, and $\phi + \delta$, and thus $\beta = b_{Y.X} - b_{W.X}(\phi + \delta)$, under the sequentially stronger assumptions A_1-A_5, A_1-A_6^+ or A_1-A_6^-, and A_1-A_6^0. From the proof of Theorem 3.1 below, we can express the moments in $Var[(\tilde{Y}, \tilde{W})']$

under A_1-A_3 by

$$\sigma^2_{\tilde{W}} = \sigma^2_{\tilde{U}} + \sigma^2_{\varepsilon}, \quad \sigma_{\tilde{W},\tilde{Y}} = (\phi+\delta)\sigma^2_{\tilde{U}} + \phi\sigma^2_{\varepsilon}, \quad \text{and } \sigma^2_{\tilde{Y}} = (\phi+\delta)^2\sigma^2_{\tilde{U}} + \phi^2\sigma^2_{\varepsilon} + \sigma^2_{\eta}.$$

Dividing $\sigma_{\tilde{W},\tilde{Y}}$ by $\sigma^2_{\tilde{W}} \neq 0$, gives that $b_{\tilde{Y}.\tilde{W}}$ is a weighted average of ϕ and $\phi+\delta$:

$$b_{\tilde{Y}.\tilde{W}} = \phi(1-\rho) + (\phi+\delta)\rho \qquad \text{where } \rho \equiv \frac{\sigma^2_{\tilde{U}}}{\sigma^2_{\tilde{W}}} = \frac{\sigma^2_{\tilde{U}}}{\sigma^2_{\tilde{U}} + \sigma^2_{\varepsilon}}. \tag{4}$$

Clearly, $0 \leq \rho \leq 1$. If there is no measurement error ($\sigma^2_{\varepsilon} = 0$) then $\rho = 1$ and $b_{\tilde{Y}.\tilde{W}} = \phi + \delta$ whereas if $\tilde{U}$ is degenerate ($\sigma^2_{\tilde{U}} = 0$ and U and X are perfectly collinear) then $\rho = 0$ and $b_{\tilde{Y}.\tilde{W}} = \phi$. Similarly, normalizing $\sigma^2_{\tilde{Y}}$ by $\sigma^2_{\tilde{W}}$, we have that

$$\frac{\sigma^2_{\tilde{Y}}}{\sigma^2_{\tilde{W}}} = \phi^2(1-\rho) + (\phi+\delta)^2\rho + \frac{\sigma^2_{\eta}}{\sigma^2_{\tilde{W}}}, \tag{5}$$

where, by definition, we have the inequality

$$0 \leq \xi^2 \equiv \frac{\sigma^2_{\eta}}{\sigma^2_{\tilde{W}}}. \tag{6}$$

As we demonstrate, the nonlinear system of moment (in)equalities (2) and (4-6) exhausts the information on $(\rho, \phi, \delta, \phi+\delta, \beta)$ implied by A_1-A_3. A_4 adds the constraints $\frac{1}{1+\kappa} \leq \rho \leq 1$, A_5 tightens the lower bound in (6) to $(1-\tau)\frac{\sigma^2_{\tilde{Y}}}{\sigma^2_{\tilde{W}}} \leq \xi^2$, and A_6 specifies whether $0 \leq \phi\delta$, $\phi\delta \leq 0$, or $\phi = 0$.

When U and X are not perfectly collinear ($\rho \neq 0$), Theorem 3.1 employs equations (2, 4, 5) to express δ, $\phi+\delta$, β, and ξ^2 as functions D, G, B, and C^2 of (ρ, ϕ). This mapping enables characterizing the sharp identification region for $(\rho, \phi, \delta, \phi+\delta, \beta)$ in terms of restrictions on (ρ, ϕ) only. It facilitates studying the consequences of deviating from the benchmark no measurement error assumption ($\rho = 1$) or the proxy exclusion restriction ($\phi = 0$).

Theorem 3.1 *Assume A_1-A_3 and let $Var[(X', U)']$ be nonsingular so that $0 < \rho \leq 1$. Then*

$$\begin{aligned}
\delta = D(\rho,\phi) &\equiv \frac{1}{\rho}(b_{\tilde{Y}.\tilde{W}} - \phi) \\
\phi + \delta = G(\rho,\phi) &\equiv \frac{1}{\rho}[b_{\tilde{Y}.\tilde{W}} - \phi(1-\rho)] \\
\beta = B(\rho,\phi) &\equiv b_{Y.X} - b_{W.X}\frac{1}{\rho}[b_{\tilde{Y}.\tilde{W}} - \phi(1-\rho)], \textit{ and} \\
\xi^2 = C^2(\rho,\phi) &\equiv \frac{\sigma^2_{\tilde{Y}}}{\sigma^2_{\tilde{W}}} - \frac{(1-\rho)}{\rho}(\phi - b_{\tilde{Y}.\tilde{W}})^2 - b^2_{\tilde{Y}.\tilde{W}}.
\end{aligned}$$

Theorem 3.1 shows that if $\rho = 1$ then $\phi + \delta$, β, and ξ^2 are point identified. Further, if $R^2_{\tilde{W}.\tilde{Y}} = 1$ then $\frac{\sigma^2_{\tilde{Y}}}{\sigma^2_{\tilde{W}}} = b^2_{\tilde{Y}.\tilde{W}}$ and it follows from $0 \leq \xi^2$ that either $\rho = 1$ or $\delta = 0$ and, therefore, that $\phi + \delta = b_{\tilde{Y}.\tilde{W}}$ and $\beta = b_{Y.X} - b_{W.X} b_{\tilde{Y}.\tilde{W}}$. Last, if $b_{W.X} = 0$ then $b_{Y.X} = \beta$.

3.1 Identification Regions under A_1-A_5

Corollary 3.2 uses (in)equalities (2) and (4-6) and the mappings in Theorem 3.1 to characterize the sharp identification region for $(\rho, \phi, \delta, \phi + \delta, \beta)$ under A_1-A_5.

Corollary 3.2 *Under the conditions of Theorem 3.1, A_4, and A_5, $(\rho, \phi, \delta, \phi+\delta, \beta)$ is partially identified in the sharp set*

$$\mathcal{S}_{\kappa,\tau} \equiv \left\{ (r, f, D(r,f), G(r,f), B(r,f)) : \frac{1}{1+\kappa} \leq r \leq 1 \;\; \textit{and} \;\; (1-\tau)\frac{\sigma^2_{\tilde{Y}}}{\sigma^2_{\tilde{W}}} \leq C^2(r,f) \right\}.$$

Further, ϕ and δ are not identified, $\mathcal{F}_{\kappa,\tau} = \mathcal{D}_{\kappa,\tau} = \mathbb{R}$, and ρ, $\phi + \delta$, and β are partially identified in the sharp sets

$$\mathcal{R}_{\kappa,\tau} = [\frac{1}{1+\kappa}, 1] \;\; \textit{and}$$

$$\mathcal{G}_{\kappa,\tau} = \{b_{\tilde{Y}.\tilde{W}} + \lambda[\kappa(\tau\frac{\sigma^2_{\tilde{Y}}}{\sigma^2_{\tilde{W}}} - b^2_{\tilde{Y}.\tilde{W}})]^{\frac{1}{2}}\} : -1 \leq \lambda \leq 1\} \quad \textit{with } \mathcal{B}_{\kappa,\tau} = \{b_{Y.X} - b_{W.X} g : g \in \mathcal{G}_{\kappa,\tau}\}.$$

The proof of Corollary 3.2 shows that the joint identification region $\mathcal{S}_{\kappa,\tau}$ is sharp since for every $(r, f, d, g, b) \in \mathcal{S}_{\kappa,\tau}$ there exists $(U^*, \eta^*, \varepsilon^*)$, with $\frac{\sigma^2_{\tilde{U}^*}}{\sigma^2_{\tilde{W}}} = r$ and $\frac{\sigma^2_{\tilde{\eta}^*}}{\sigma^2_{\tilde{W}}} = C^2(r,f)$, that satisfy A_2-A_5 and that could have generated Y and W according to A_1. Corollary 3.2 also derives the identification regions for $\rho, \phi, \delta, \phi + \delta$, and β separately. Each of these projected regions is sharp - for example, for every $d \in \mathcal{D}_{\kappa,\tau}$ there exists $(r, f, d, g, b) \in \mathcal{S}_{\kappa,\tau}$.

When $\kappa \to +\infty$ and $R^2_{\tilde{W}.\tilde{Y}} \neq \tau$ (i.e. $\tau\frac{\sigma^2_{\tilde{Y}}}{\sigma^2_{\tilde{W}}} \neq b^2_{\tilde{Y}.\tilde{W}}$), projecting $\mathcal{S}_{\kappa,\tau}$ onto the support $(0, 1]$ of ρ or the support $\mathbb{R}$ of ϕ, δ, $\phi+\delta$, and β_l for $l = 1, ..., k$ yields the full support. (Here, we drop the superfluous κ, τ subscripts, $\mathcal{S} \equiv \mathcal{S}_{\infty,\tau}$.) Without the proxy exclusion restriction $\phi = 0$, none of these parameters is identified under A_1-A_3 and A_5. When $\kappa < \infty$, Corollary 3.2 yields two-sided sharp bounds for ρ, $\phi+\delta$, and β whereas ϕ and δ remain unidentified. Last, we note that the paper's bounds enable characterizing the bias of several key estimands. For example, using the regression representation $b_{Y.(W,X')'} = (b_{\tilde{Y}.\tilde{W}}, b'_{Y.X} - b'_{W.X} b_{\tilde{Y}.\tilde{W}})'$, $\mathcal{B}_{\kappa,\tau}$ reveals that the magnitude of the bias of the coefficient on X in $b_{Y.(W,X')'}$ is at most $|b_{W.X}|\, [\kappa(\tau\frac{\sigma^2_{\tilde{Y}}}{\sigma^2_{\tilde{W}}} - b^2_{\tilde{Y}.\tilde{W}})]^{\frac{1}{2}}$.

3.2 Identification Regions under A_1-A_6

Next, we impose A_6. We begin by examining A_6^+, $\phi\delta \geq 0$. For this, we let $E(r,f) \equiv fD(r,f) = \frac{1}{r}f(b_{\tilde{Y}.\tilde{W}} - f)$. Also, we define the maximum L and the indicator T

$$L \equiv \max\{\frac{1}{\tau}R^2_{\tilde{W}.\tilde{Y}}, \frac{1}{1+\kappa}\} \quad \text{and} \quad T \equiv \mathbf{1}\{R^2_{\tilde{W}.\tilde{Y}} \in \{(1-\lambda)\frac{\tau}{1+\kappa} + \lambda\frac{\tau\kappa}{1+\kappa} : 0 < \lambda < 1\}\}.$$

Corollary 3.3 *Under the conditions of Theorem 3.1, A_4, A_5, and A_6^+, $(\rho, \phi, \delta, \phi+\delta, \beta)$ is partially identified in the sharp set*

$$\mathcal{S}^+_{\kappa,\tau} \equiv \left\{(r, f, D(r,f), G(r,f), B(r,f)) : \frac{1}{1+\kappa} \leq r \leq 1, \quad (1-\tau)\frac{\sigma^2_{\tilde{Y}}}{\sigma^2_{\tilde{W}}} \leq C^2(r,f), \textit{ and } 0 \leq E(r,f)\right\}.$$

Further, ρ, ϕ, δ, $\phi+\delta$, and β are partially identified in the sharp sets

$$\begin{aligned}
&\mathcal{R}^+_{\kappa,\tau} = [\frac{1}{1+\kappa}, 1], \\
&\mathcal{F}^+_{\kappa,\tau} = \{\lambda b_{\tilde{Y}.\tilde{W}} : 0 \leq \lambda \leq 1\}, \\
&\mathcal{D}^+_{\kappa,\tau} = \begin{cases} \{\lambda(1+\kappa)b_{\tilde{Y}.\tilde{W}}[\frac{1}{\kappa}(\frac{1}{L}-1)]^{\frac{1}{2}} : 0 \leq \lambda \leq 1\} & \textit{if } T = 1 \textit{ and } \kappa > 0 \\ \{\lambda b_{\tilde{Y}.\tilde{W}}\frac{1}{L} : 0 \leq \lambda \leq 1\} & \textit{if } T = 0 \textit{ or } \kappa = 0 \end{cases}, \\
&\mathcal{G}^+_{\kappa,\tau} = \{b_{\tilde{Y}.\tilde{W}}\{1 + \lambda[\kappa(\frac{1}{L}-1)]^{\frac{1}{2}}\} : 0 \leq \lambda \leq 1\} \quad \textit{with } \mathcal{B}^+_{\kappa,\tau} = \{b_{Y.X} - b_{W.X}g : g \in \mathcal{G}^+_{\kappa,\tau}\}.
\end{aligned}$$

When $\kappa \to +\infty$, we obtain the two-sided sharp bounds $\mathcal{F}^+ = \mathcal{F}^+_{\kappa,\tau}$ and, except when $R^2_{\tilde{W}.\tilde{Y}} = \tau$ (and thus $L = 1$), the one-sided sharp bounds $\mathcal{D}^+ = \{\lambda b_{\tilde{Y}.\tilde{W}} : 0 \leq \lambda\}$, $\mathcal{G}^+ = \{b_{\tilde{Y}.\tilde{W}}\lambda : 1 \leq \lambda\}$, and $\mathcal{B}^+ = \{b_{Y.X} - b_{W.X}b_{\tilde{Y}.\tilde{W}}\lambda : 1 \leq \lambda\}$. Note that $\mathcal{F}^+$ and $\mathcal{D}^+$ identify the common sign of ϕ and δ. Imposing A_1-A_6^+ with $\kappa < \infty$ yields bounded identification regions for ρ, ϕ, δ, $\phi+\delta$, and β that can be tighter than those obtained when $\kappa \to +\infty$, $\tau = 1$, or without A_6^+.

Corollary 3.4 examines the identifying power of A_6^-, $\phi\delta \leq 0$.

Corollary 3.4 *Under the conditions of Theorem 3.1, A_4, A_5, and A_6^-, $(\rho, \phi, \delta, \phi+\delta, \beta)$ is partially identified in the sharp set*

$$\mathcal{S}^-_{\kappa,\tau} \equiv \left\{(r, f, D(r,f), G(r,f), B(r,f)) : \frac{1}{1+\kappa} \leq r \leq 1, \quad (1-\tau)\frac{\sigma^2_{\tilde{Y}}}{\sigma^2_{\tilde{W}}} \leq C^2(r,f), \textit{ and } E(r,f) \leq 0\right\}.$$

Further, ρ, ϕ, δ, $\phi+\delta$, and β are partially identified in the sharp sets

$$\mathcal{R}^-_{\kappa,\tau} = [\frac{1}{1+\kappa}, 1], \quad \mathcal{F}^-_{\kappa,\tau} = \mathcal{D}^-_{\kappa,\tau} = \begin{cases} \{\lambda b_{\tilde{Y}.\tilde{W}} : \lambda \notin (0,1)\} & \textit{if } b_{\tilde{Y}.\tilde{W}} \neq 0 \\ \mathbb{R} & \textit{if } b_{\tilde{Y}.\tilde{W}} = 0 \end{cases},$$

and if $L = \frac{1}{1+\kappa}$ *then* $\mathcal{G}^-_{\kappa,\tau} = \mathcal{G}_{\kappa,\tau}$ *with* $\mathcal{B}^-_{\kappa,\tau} = \mathcal{B}_{\kappa,\tau}$ *whereas if* $L = \frac{1}{\tau} R^2_{\tilde{W}.\tilde{Y}}$ *then*

$$\mathcal{G}^-_{\kappa,\tau} = \{b_{\tilde{Y}.\tilde{W}}\{\lambda\frac{1}{L}+(1-\lambda)[1-(\kappa(\frac{1}{L}-1))^{\frac{1}{2}}]\} : 0 \le \lambda \le 1\} \textit{ with } \mathcal{B}^-_{\kappa,\tau} = \{b_{Y.X}-b_{W.X}g : g \in \mathcal{G}^-_{\kappa,\tau}\}.$$

When $\kappa \to +\infty$, we obtain the same sharp identification regions for ϕ and δ as when $\kappa < \infty$, $\mathcal{F}^- = \mathcal{D}^- = \mathcal{F}^-_{\kappa,\tau} = \mathcal{D}^-_{\kappa,\tau}$. This is a disconnected region which rules out that ϕ or δ is in the open interval with end points 0 and $b_{\tilde{Y}.\tilde{W}}$. Further, except when $R^2_{\tilde{W}.\tilde{Y}} = 0$ or $R^2_{\tilde{W}.\tilde{Y}} = \tau$, we obtain the one-sided sharp bounds $\mathcal{G}^-_{\tau} = \{b_{\tilde{Y}.\tilde{W}} \frac{\tau}{R^2_{\tilde{W}.\tilde{Y}}}\lambda : \lambda \le 1\}$ and $\mathcal{B}^-_{\tau} = \{b_{Y.X} - b_{W.X}b_{\tilde{Y}.\tilde{W}} \frac{\tau}{R^2_{\tilde{W}.\tilde{Y}}}\lambda : \lambda \le 1\}$. When $\kappa < \infty$, the sharp identification regions $\mathcal{G}^-_{\kappa,\tau}$ and $\mathcal{B}^-_{\kappa,\tau}$ are tighter than $\mathcal{G}_{\kappa,\tau}$ and $\mathcal{B}_{\kappa,\tau}$ only if $\frac{\tau}{1+\kappa} < R^2_{\tilde{W}.\tilde{Y}}$. Last, unlike in Corollary 3.3, assigning specific signs to ϕ and δ may tighten the bounds in Corollary 3.4 - we do not pursue this here for brevity.

Last, Corollary 3.5 imposes A_1-A_6^0 so that the proxy exclusion restriction $\phi = 0$ holds. The resulting bounds are nested in the bounds obtained under A_1-A_6^+ and A_1-A_6^-.

Corollary 3.5 *Under the conditions of Theorem 3.1,* A_4*,* A_5*, and* A_6^0*,* (ρ, δ, β) *is partially identified in the sharp set*

$$\mathcal{S}^0_{\kappa,\tau} \equiv \left\{(r, D(r,0), B(r,0)) : \frac{1}{1+\kappa} \le r \le 1 \quad \textit{and} \quad (1-\tau)\frac{\sigma^2_{\tilde{Y}}}{\sigma^2_{\tilde{W}}} \le C^2(r,0)\right\}.$$

Further, ρ*,* δ*, and* β *are partially identified in the sharp sets*

$$\mathcal{R}^0_{\kappa,\tau} = \{\lambda L + (1-\lambda) : 0 \le \lambda \le 1\}$$
$$\mathcal{D}^0_{\kappa,\tau} = \{b_{\tilde{Y}.\tilde{W}}[\lambda + (1-\lambda)\frac{1}{L}] : 0 \le \lambda \le 1\} \textit{ with } \mathcal{B}^0_{\kappa,\tau} = \{b_{Y.X} - b_{W.X}d : d \in \mathcal{D}^0_{\kappa,\tau}\}.$$

When $\kappa \to +\infty$, $\tau = 1$, and $R^2_{\tilde{W}.\tilde{Y}} \neq 0$, the bounds in Corollary 3.5 reduce to the standard sharp bounds $\mathcal{R}^0$, $\mathcal{D}^0$, and $\mathcal{B}^0$ with $L = R^2_{\tilde{W}.\tilde{Y}}$ (see e.g. Gini, 1921; Frisch, 1934, Klepper and Leamer, 1984; Bollinger, 2003). (If $R^2_{\tilde{W}.\tilde{Y}} = 0$, $\mathcal{R}^0 = (0,1]$ and $\mathcal{D}^0 = \{b_{\tilde{Y}.\tilde{W}}\lambda : 0 \le \lambda\} = \{0\}$.) Setting $\kappa < \infty$ or $\tau < 1$ can yield tighter bounds.

To conclude, we briefly comment on the case when a researcher assumes that several variables serve as proxies for U. When suitable, one can apply this paper's framework using each proxy separately by including the other proxies in X. Examining the analytical expressions for the identification regions then reveals whether a bound obtained using an included proxy has a narrower width than that obtained using another included or excluded proxy. For instance, consider two proxies W_h, $h = 1, 2$ that satisfy A_1-A_5 and either A_6^+ ($\delta\phi_h \geq 0$) or A_6^0 ($\phi_h = 0$), with the same sufficiently small κ and large τ such that $L = \frac{1}{1+\kappa}$ and $T = 1$ for either proxy. For $h, h' = 1, 2$, $h \neq h'$, let b_h denote the coefficient on W_h in a regression of Y on $(X', W_h, W_{h'})'$. In this case, when letting W_h serve as the proxy and $W_{h'}$ as a covariate, the smaller $|b_h|$ is the narrower the widths of the bounds on δ and, under A_6^+, ϕ_h and $\phi_h + \delta$ are. Similarly, the smaller the l^{th} component of $\left|b_h b_{W_h.(X', W_{h'})'}\right|$ is the narrower the width of the bounds on the l^{th} component of $(\beta', \phi_{h'})'$ is. Moreover, tighter bounds obtain by taking the intersection of the bounds that use each of the multiple proxies separately. We leave studying the sharp identification regions in the presence of multiple included and/or excluded proxies to other work.

4 Numerical Example

It is instructive to consider an example that illustrates the shape of the identification regions in Section 3. Specifically, let X, Y, and W be generated, according to A_1, by

$$Y = X'\beta + W\phi + U\delta + \eta, \qquad X' = U\varphi + \eta_X', \quad \text{and} \quad W = U + \varepsilon,$$

where $\underset{2\times 1}{X} = (X_1, X_2)'$. Further, let U, η, ε, and η_X be jointly independent and normally distributed with mean zero so that A_2 and A_3 hold. It follows that $(X', Y, W)'$ is normally distributed and we can analytically express the identification regions for $\rho, \phi, \delta, \phi+\delta$, and β in Section 3 using the elements of $Var[(U, \eta, \varepsilon, \eta_X')']$. To illustrate these identification regions, we set $\beta = (1, 0.7)'$, $\phi = 0.5$, $\delta = 0.9$, and $\varphi = (0.35, 0.14)$. Since $0 < \phi\delta$, A_6^+ holds. Also, we set $\sigma_U^2 = 3$, $\sigma_\eta^2 = 0.4$, $\sigma_\varepsilon^2 = \sigma_{\eta_{X_1}}^2 = \sigma_{\eta_{X_2}}^2 = 1$, and $\sigma_{\eta_{X_1}, \eta_{X_2}} = 0.2$. We obtain $\rho = 0.685$ and $R^2_{\tilde{W}.\tilde{Y}} = 0.805$ and set (κ, τ) such that $\frac{\sigma_\varepsilon^2}{\sigma_{\tilde{U}}^2} = 0.461 \leq \kappa$ and $\tilde{R}_*^2 = 0.918 \leq \tau$.

Using a grid search, we approximate the joint identification regions $\mathcal{S}_{\kappa,\tau}$, $\mathcal{S}^+_{\kappa,\tau}$, $\mathcal{S}^-_{\kappa,\tau}$, and $\mathcal{S}^0_{\kappa,\tau}$ obtained under this parametrization. Figure 1 illustrates these regions by plotting their projections onto the (ϕ, ρ), (ϕ, δ), and (β_1, β_2) spaces. Each graph in Figure 1 superimposes the 4 projected identification regions that correspond to $(\kappa, \tau) = (+\infty, 1)$, $(2, 1)$, $(2, 0.95)$, and $(0.5, 0.92)$. The darker intersections correspond to smaller κ or τ values (or both) and are nested within the lighter regions. Sometimes the identification regions displayed in Figure 1 are unbounded. For example, $\mathcal{B}$ is an unbounded line whereas the projection of $\mathcal{S}^+_{\kappa,\tau}$ on the (ρ, ϕ) space is a bounded set when $\kappa < \infty$. Figure 1 illustrates how the vector of population coefficients (which we mark using a plus sign) is an element of the joint sharp identification regions $\mathcal{S}_{\kappa,\tau}$ and $\mathcal{S}^{+}_{\kappa,\tau}$. On the other hand, neither $\phi\delta \leq 0$ nor $\phi = 0$ holds and $\mathcal{S}^{-}_{\kappa,\tau}$ and $\mathcal{S}^{0}_{\kappa,\tau}$ do not contain $(\rho, \phi, \delta, \phi + \delta, \beta)$. Last, Figure 1 illustrates how $\mathcal{S}^{-}_{\kappa,\tau}$ is disconnected, $\mathcal{S}^{0}_{\kappa,\tau} \subseteq \mathcal{S}^{+}_{\kappa,\tau} \cap \mathcal{S}^{-}_{\kappa,\tau}$, and $\mathcal{S}^{+}_{\kappa,\tau} \cup \mathcal{S}^{-}_{\kappa,\tau} = \mathcal{S}_{\kappa,\tau}$.

Using the analytical expressions in Section 3, Table 1 reports the bounds for ρ, ϕ, δ, $\phi + \delta$, β_1, and β_2 that correspond to the regions in Figure 1. It reports the sharp bounds obtained under A_1-A_5 (column 1), A_1-A_6^{+} (column 2), and the incorrect assumptions A_1-A_6^{-} (column 3) and A_1-A_6^{0} (column 4). Column 5 reports the regression estimand $b_{Y.(W,X')'}$ that identifies $(\phi + \delta, \beta)$ if $\rho = 1$ or $\delta = 0$. As Table 1 shows, the projections for $\mathcal{S}^{-}_{\kappa,\tau}$ and $\mathcal{S}^{0}_{\kappa,\tau}$ do not contain ϕ, δ, $\phi+\delta$, and β. In contrast, $\mathcal{S}^{+}_{\kappa,\tau}$ improves over $\mathcal{S}_{\kappa,\tau}$ and both regions contain the true parameter values and become tighter as κ or/and τ decrease(s).

5 Estimation and Inference

We conduct inference on each of the partially identified parameters ρ, ϕ, δ, $\phi + \delta$, and β_l, $l = 1, ..., k$, in Corollaries 3.2 to 3.5 (see e.g. Shi and Shum (2015) or Kline and Tamer (2016) for inference procedures on the joint identification regions). Each of these identification regions is of the form $\theta \in \mathcal{H} = \{H(P; \lambda) : \lambda \in \Lambda\}$ where $H(\cdot; \lambda)$ is a function of the estimands

$$P \equiv (b'_{Y.(W,X')'}, b'_{W.(Y,X')'}, b'_{Y.X}, b'_{W.X}, \frac{\sigma^2_{\tilde{Y}}}{\sigma^2_{\tilde{W}}})'$$

and λ is a nuisance parameter that is partially identified in a known set Λ. (We use $b_{W.(Y,X')'}$ to form $R^2_{\tilde{W}.\tilde{Y}} = b_{\tilde{Y}.\tilde{W}} b_{\tilde{W}.\tilde{Y}}$ and can dispense with it from P and use $R^2_{\tilde{W}.\tilde{Y}} = b^2_{\tilde{Y}.\tilde{W}}(\frac{\sigma^2_{\tilde{Y}}}{\sigma^2_{\tilde{W}}})^{-1}$ instead.) For example,

$$\mathcal{B}^{+}_{\kappa,\tau} = \{B^{+}_{\kappa,\tau}(P; \lambda) : \lambda \in \Lambda\} \equiv \{b_{Y.X} - b_{W.X} b_{\tilde{Y}.\tilde{W}} \{1 + \lambda[\kappa(\frac{1}{L} - 1)]^{\frac{1}{2}}\} : \lambda \in [0, 1]\}.$$

We estimate an identification region $\mathcal{H}$ consistently using $\widehat{\mathcal{H}} = \{H(\hat{P}; \lambda) : \lambda \in \Lambda\}$ where $\hat{P}$ denotes the plug-in estimator for P:

$$\hat{P} \equiv (\hat{b}'_{Y.(W,X')'}, \hat{b}'_{W.(Y,X')'}, \hat{b}'_{Y.X}, \hat{b}'_{W.X}, \frac{\sum_{i=1}^{n} \hat{\epsilon}^2_{Y.X,i}}{\sum_{i=1}^{n} \hat{\epsilon}^2_{W.X,i}})'.$$

Specifically, given observations $\{A_i, B_i\}_{i=1}^{n}$ corresponding to random column vectors A and B, let $\bar{A} \equiv \frac{1}{n}\sum_{i=1}^{n} A_i$ and denote the linear regression estimator and sample residual by:

$$\hat{b}_{A.B} \equiv [\frac{1}{n}\sum_{i=1}^{n}(B_i - \bar{B})(B_i - \bar{B})']^{-1}[\frac{1}{n}\sum_{i=1}^{n}(B_i - \bar{B})(A_i - \bar{A})'] \text{ and } \hat{\epsilon}'_{A.B,i} \equiv (A_i - \bar{A})' - (B_i - \bar{B})'\hat{b}_{A.B}.$$

Standard arguments show that the estimator $\hat{P}$ for P is $\sqrt{n}$ consistent and asymptotically normally distributed. For this, let $\mu_A^2 = E(AA')$ and define the $7+4k$ square diagonal matrix

$$Q \equiv diag\{\mu^2_{(1,W,X')'}, \mu^2_{(1,Y,X')'}, \mu^2_{(1,X')'}, \mu^2_{(1,X')'}, \sigma^2_{\tilde{W}}\}.$$

Theorem 5.1 *Assume A_1(i) and that Q is nonsingular. Suppose further that:*
(i) $\frac{1}{n}\sum_{i=1}^{n}(1, Y_i, W_i, X_i')'(1, Y_i, W_i, X_i') \xrightarrow{p} \mu^2_{(1,Y,W,X')'}$ *and*

$$(ii) \quad n^{-1/2}\sum_{i=1}^{n}\begin{bmatrix}(1, W_i, X_i')'\epsilon_{Y.(W,X')',i} \\ (1, Y_i, X_i')'\epsilon_{W.(Y,X')',i} \\ (1, X_i')'\epsilon_{Y.X,i} \\ (1, X_i')'\epsilon_{W.X,i} \\ \epsilon^2_{Y.X,i} - \sigma^2_{\tilde{Y}}\end{bmatrix} \xrightarrow{d} N(0, \Xi) \quad \textit{where } \Xi \equiv Var\begin{bmatrix}(1, W, X')'\epsilon_{Y.(W,X')'} \\ (1, Y, X')'\epsilon_{W.(Y,X')'} \\ (1, X')'\epsilon_{Y.X} \\ (1, X')'\epsilon_{W.X} \\ \epsilon^2_{Y.X}\end{bmatrix}.$$

Then $\sqrt{n}(\hat{P} - P) \xrightarrow{d} N(0, \Gamma)$ where Γ obtains by removing the $1, 3+k, 5+2k,$ and $6+3k$ intercept rows and columns from $\Gamma^* \equiv Q^{-1}\Xi Q'^{-1}$.

See e.g. White (2001) for primitive conditions for the law of large numbers and central limit theorem in Theorem 5.1. We estimate Γ using the relevant submatrix of the heteroskedasticity-robust plug-in estimator $\hat{\Gamma}^* \equiv \hat{Q}^{-1}\hat{\Xi}\hat{Q}'^{-1}$ (see e.g. White, 1980). For example, we estimate $Var(X\epsilon_{Y.X})$ using $\frac{1}{n}\sum_{i=1}^{n} X_i\hat{\epsilon}_{Y.X,i}, \hat{\epsilon}_{Y.X,i}X_i'$.

In Section 3, the function $H(P;\lambda)$ for an identification region $\mathcal{H}$ sometimes depends on the value of $R^2_{\tilde{W}.\tilde{Y}}$ via L and T. If $R^2_{\tilde{W}.\tilde{Y}}$ is known then one can construct a $1-\alpha$ (e.g. 95%) confidence interval $C_{1-\alpha}(\lambda)$ for $H(P;\lambda)$ for each $\lambda \in \Lambda$ using the delta method. A confidence region $CR^{\theta}_{1-\alpha}$ for a partially identified parameter $\theta \in \mathcal{H}$ then obtains by applying Proposition 2 of Chernozhukov, Rigobon, and Stoker (2010) and forming the union:

$$CR^{\theta}_{1-\alpha} = \bigcup_{\lambda\in\Lambda} C_{1-\alpha}(\lambda).$$

In applications, $R^2_{\tilde{W}.\tilde{Y}}$ must be estimated and $CR^{\theta}_{1-\alpha}$ needs to be adjusted to account for this estimation. Let $r_{\tilde{Y}.\tilde{W}} \equiv \frac{\sigma_{\tilde{Y}.\tilde{W}}}{\sigma_{\tilde{Y}}\sigma_{\tilde{W}}}$ denote the partial correlation between Y and W given X and rewrite $\mathcal{H}$ in the form $\mathcal{H} = \{\ddot{H}(P;\pi) : \pi \in \Pi\}$ where $\pi = (\lambda, \ddot{r}) \in \Lambda \times \{r_{\tilde{Y}.\tilde{W}}\}$ determines $R^2_{\tilde{W}.\tilde{Y}}$, with $\ddot{H}(\cdot;\pi)$ continuously differentiable in P. For example, we have

$$\begin{aligned}\mathcal{B}^+_{\kappa,\tau} &= \{\ddot{B}^+_{\kappa,\tau}(P;\pi) : \pi \in \Pi\} \\ &\equiv \{b_{Y.X} - b_{W.X}b_{\tilde{Y}.\tilde{W}}\{1 + \lambda\{\kappa[\frac{\tau}{\ddot{r}^2}\mathbf{1}\{\frac{\tau}{1+\kappa} < \ddot{r}^2\} + (1+\kappa)\mathbf{1}\{\ddot{r}^2 \leq \frac{\tau}{1+\kappa}\} - 1]\}^{\frac{1}{2}}\} \\ &\qquad : (\lambda, \ddot{r}) \in [0,1] \times \{r_{\tilde{Y}.\tilde{W}}\}\}.\end{aligned}$$

By the delta method, the plug-in estimator $\ddot{H}(\hat{P};\pi)$ for an element $\ddot{H}(P;\pi)$ of $\mathcal{H}$ obeys

$$\sqrt{n}(\ddot{H}(\hat{P};\pi) - \ddot{H}(P;\pi)) \xrightarrow{d} N(0, \nabla_P \ddot{H}(P;\pi)\Gamma\nabla_P \ddot{H}(P;\pi)').$$

This permits constructing a $1-\alpha_1$ confidence interval $C_{1-\alpha_1}(\pi)$ for $\ddot{H}(P;\pi)$ with $\pi \in \Pi$. To obtain a $1-\alpha_1-\alpha_2$ (e.g. 95%) confidence region $CR^{\theta}_{1-\alpha_1-\alpha_2}$ for $\theta \in \mathcal{H}$, we construct a confidence interval $CR^{\ddot{r}}_{1-\alpha_2}$ for $r_{\tilde{Y}.\tilde{W}}$ and apply Proposition 3 of Chernozhukov, Rigobon, and Stoker (2010) to form the union:

$$CR^{\theta}_{1-\alpha_1-\alpha_2} = \bigcup_{\pi \in \Lambda \times CR^{\ddot{r}}_{1-\alpha_2}} C_{1-\alpha_1}(\pi).$$

To construct $CR^{\ddot{r}}_{1-\alpha_2}$, we use the "Fisher z" variance stabilizing transformation (see e.g. van der Vaart, 2000, p. 30-31). For brevity, we describe how we construct $CR^{\ddot{r}}_{1-\alpha_2}$ and report the expressions for the gradients $\nabla_P \ddot{H}(P;\pi)$ for Corollaries 3.2 to 3.5 in Section B of the Supplement. In the empirical analysis in Section 6, we set $\alpha_1 = 0.04$ and $\alpha_2 = 0.01$.

6 The Returns to College Selectivity and Characteristics

As discussed in Monks (2000, p. 283), together with a student's individual characteristics, the attributes of the college that a students attends may influence his or her earnings through accumulating human capital and by signaling the student's ability to employers. We illustrate this paper's results by studying the returns to college selectivity and other characteristics as well as the student characteristics using the recent College Scorecard (CS) dataset. CS is nationally comprehensive and reports data, aggregated at the institution level, on a wide array of the attributes of postsecondary institutions in the US.

Following Black and Smith (2006, p. 703), we consider an education production function determined by "various college-level inputs [...] such as the average SAT score of the entering class, expenditures per student, and so on" and by "other factors affecting earnings and college quality choice." Specifically, we let the earnings of student j at college i be given by

$$Y_{ij} = f(X^{c\prime}_{ij}, X^{s\prime}_{ij}, W_i, U_{ij}, \eta_{ij}) = X^{c\prime}_{ij}\beta^c + X^{s\prime}_{ij}\beta^s + W_i\phi + U_{ij}\delta + \eta_{ij}$$

where Y_{ij} denotes student $j's$ earnings, W_i is the average SAT (equivalent) score of the student's cohort at college i, $X_{ij} = (X^{c\prime}_{ij}, X^{s\prime}_{ij})'$ collects the other characteristics X^c_{ij} that may depend on the college (e.g. the college's control (public or private non-profit) or the student's field of study) and the student's demographic and socioeconomic characteristics X^s_{ij}, U_{ij} is student j's unobserved scholastic "ability," and η_{ij} is an equation disturbance such that $Cov[\eta_{ij}, (X'_{ij'}, W_i, U_{ij'})'] = 0$ for all i, j, j'. The empirical illustration treats the average SAT score W_i as an error-free measure of "college selectivity" (as in e.g. Dale and Krueger (2002, 2014) and Hoxby (2009)) which, along with the other college and student characteristics, can directly affect earnings. Black and Smith (2006, p. 704) demonstrate the identification difficulties that arise when imposing the "simplifying assumption of a 'one-factor' model, in which quality has a single dimension" measured by an error-laden "single college quality measure." They then study approaches that use multiple college characteristics as excluded error-laden proxies for the latent college quality. Here, we do not impose a one-factor model for college quality. Instead, we estimate the coefficients ϕ, δ, and $\beta = (\beta^{c\prime}, \beta^{s\prime})'$ of the education production function which admits the multiple college characteristics $(X^{c\prime}_{ij}, W_i)$ as inputs.

An important challenge in identifying ϕ, δ, and β arises because students with higher unobserved ability U_{ij} may earn more and enroll in colleges with a particular selectivity and characteristics profile. In this case, $Cov[U_{ij}, (X'_{ij}, W_i)'] \neq 0$ and a regression of Y_{ij} on $(X'_{ij}, W_i)'$ does not identify $(\beta', \phi)'$. To proceed, the literature sometimes assumes that students with similar observed characteristics do not systematically select into colleges based on their unobserved ability (e.g. Monks (2000) and Black and Smith (2006). See Black and Smith (2004) for a nonparametric analysis). Because accounting for only the student's demographic and socioeconomic characteristics X^s_{ij} is not very likely to ensure this condition, the literature conditions on $(X^{s\prime}_{ij}, W_{ij})'$, where W_{ij} denotes a test score that measures the ability of student j at college i. In the context of the above linear production function, this "selection on observables" assumption is helpful because if the linear projection of U_{ij} on $(X'_{ij}, W_i, W_{ij})'$ depends only on $(X^{s\prime}_{ij}, W_{ij})'$ then a regression of Y_{ij} on $(X'_{ij}, W_i, W_{ij})'$ identifies $(\beta^{c\prime}, \phi)$, albeit not β^s or δ. To gain confidence in the resulting estimates, it is useful to study their sensitivity to this assumption. In particular, a key sufficient condition for selection on observables occurs when the test score W_{ij} is a perfect measure of ability U_{ij}. However, if the test score measures ability with error $W_{ij} = U_{ij} + \varepsilon_{ij}$ then this condition is not guaranteed to hold (e.g. Bollinger, 2003). The empirical application studies the consequences of deviating from the selection on observables assumption by allowing the SAT (equivalent) score W_{ij} to measure ability U_{ij} with classical measurement error ε_{ij}, such that $Cov[\varepsilon_{ij}, (X'_{ij'}, U_{ij'}, \eta_{ij'})'] = 0$ for all i, j, j', and letting a student's unobserved ability U_{ij} freely depend on the college and student characteristics X_{ij}.

A second identification challenge arises because the CS data is aggregated at the institution level - the individual data is not observed. Let $A_i \equiv \frac{1}{N_i}\sum_{j=1}^{N_i} A_{ij}$ denote the average of A_{ij} across N_i students in college i. Averaging the Y_{ij} and W_{ij} equations across N_i yields

$$Y_i = X_i'\beta + W_i\phi + U_i\delta + \eta_i \quad \text{and} \quad W_i = U_i + \varepsilon_i.$$

Here, A_1-A_3 hold, $Cov[\eta_i, (X_i', U_i)'] = 0$ and $Cov[\varepsilon_i, (X_i', U_i, \eta_i)'] = 0$. However, because the average SAT score W_i may directly affect the average earnings, W_i violates the proxy exclusion restriction and the standard measurement error bounds are not valid. Instead, we use the aggregate equations to estimate the identification regions for ϕ, δ, $\phi + \delta$ (the total effect on average earnings of enrolling a cohort with a higher average ability), and β under A_1-A_3 and the auxiliary assumptions A_4-A_6. Here, A_4 restricts the extent of the measurement error in how the average SAT score W_i proxies the average ability U_i, A_5 places an upper bound on the fit of the aggregate equation that would obtain had W_i measured U_i without error, and A_6 restricts the effects ϕ and δ of college selectivity and the student's ability on his or her earnings to have the same sign.

6.1 College Scorecard Data

CS reports comprehensive data on several dimensions of the higher education institutions in the US over the last few decades. The data are aggregated at the institution level and drawn from various sources including the Integrated Postsecondary Education Data System (IPEDS), National Student Loan Data System (NSLDS), and administrative earnings data from tax records maintained by the Department of Treasury.

CS has several advantageous features. The literature often analyzes survey data on students who attend a small or moderate number of institutions that tend to be prestigious. In contrast, CS covers a large number of post-secondary institutions in the US. In addition to student demographic and socioeconomic characteristics, CS reports data on several characteristics of the institution, including its setting, selectivity, affordability, fields of study, expenditures per student, completion rate, and earnings outcomes. Moreover, CS contains data drawn from administrative records which may be less prone to reporting error.

While CS is detailed and nationally comprehensive, we summarize at the outset some of its limitations that are partly due to data aggregation. First, the data based on NSLDS and tax records cover only "Title IV" undergraduate students. This subpopulation of students who receive federal aid may differ from the general population. Yet, the Title IV subpopulation amounts to roughly "seventy percent of all graduating postsecondary students" and seems "reasonably similar to the overall population of a school in terms of student characteristics" (Council of Economic Advisors, 2015 (thereafter CEA), p. 26-27). Second, CS employs the IPEDS definition of an institution and, although "about two-thirds of institutions, collectively enrolling 83 percent of students, have only one main campus identifier" (CEA, p. 29), complex institutions may differ in how they aggregate and report data across multiple branches. Third, CS uses various student cohort definitions that "are imperfect and vary for different metrics" (CEA, p. 30). For example, the mean earnings variable is based on Title IV students who are non-enrolled and working e.g. 6 years after estimated college entry and is reported for a pooled cohort across two consecutive entry years (e.g. the 2006-2007 and 2007-2008 entry cohorts). On the other hand, the average annual total cost of attendance is based on all full-time, first-time, degree-seeking Title IV undergraduate students who enrolled in an institution during the academic year (e.g. 2010-2011). The extent to which this inconsistency in cohort definitions can impact our estimates depends, in part, on how stable the aggregate data is in the short run and across the cohorts used.

While addressing these data limitations is desirable, we do not pursue this here and we keep the empirical illustration focused on demonstrating the consequences of allowing the SAT score to measure scholastic ability with error. For a detailed account of CS, we refer the reader to its documentation webpage (https://collegescorecard.ed.gov/data/documentation/) and to the CEA report.

6.2 Sample Selection

We focus on the recent cohort of students who enrolled in an academic institution in the fall of 2007 and were non-enrolled and working in 2013. As discussed above, CS reports yearly data files which contain institution-level aggregate data that need not correspond to a uniform student cohort. To proceed, we draw from several CS files, the data that we deem most representative of the 2007 student cohort. Table S1 in the Supplement defines the variables that we employ in our analysis and specifies the CS variable(s) that we use in constructing each of our variables. Further, Table S1 specifies the level of aggregation used in reporting each CS variable and the CS data file from which it is drawn.

We restrict our sample to the main campus of bachelor's degree granting institutions that are either public or private non-profit. Although of policy interest, we exclude for-profit institutions to focus on institutions that operate in a similar context and for which more data is available (for-profit institutions differ from other institutions in several dimensions including the admission requirements, funding, and online education. See e.g. Deming, Goldin, and Katz (2012)). This yields a sample of 1710 institutions. After dropping 27 institutions that were missing from at least one of the relevant CS data files, we exclude 378 institutions that are missing data on SATAvg, the average SAT (or ACT equivalent) score (some of these institutions are specialized in particular fields such as arts and design, music, religion, or medical and health sciences). Last, we exclude institutions with missing data on the other variables, leading to the final sample of 1165 institutions. Table 2 reports summary statistics for the aggregate variables in the sample. For example, the average of SATAvg is 1052.76 and the minimum and maximum average scores are 726 and 1491. The 5 most selective institutions in our sample are Harvard, Princeton, Yale, MIT, and Dartmouth. The standard deviation of SATAvg is 119.93, which corresponds roughly to the difference between Stanford and the University of Virginia. For brevity, we do not report results for the coefficients on the variables that fall below the dividing line in Table 2.

6.3 Main Specification

We let Y denote the average earnings 6 years after enrollment (we do not observe wage or hours worked). W denotes the average SAT equivalent score which serves as an "included" proxy for the average unobserved ability U. $X = (X^{c\prime}, X^{s\prime})'$ consists of several aggregate college and student characteristics that have been discussed in the literature.

In particular, X^c includes 8 region indicators and 11 locale indicators for the institution's location, indicators for whether the institution is minority-serving, a women-only college (our sample does not contain men-only colleges), has a religious affiliation, awards a graduate degree, has a private non-profit (as opposed to public) control, and the undergraduate student population size. Further, X^c includes the average cost of attendance, average net price, the shares of students with a federal student loan and with a Pell grant, the median student debt (we use the median debt to approximate the average debt which CS does not report), the instructional expenditures per student, and the completion rate within 150% of expected graduation time. An advantageous feature of CS is that it reports data on the fields of study which play an important role in understanding the labor market outcomes (see e.g. Altonji, Arcidiacono, and Maurel, 2016; Kirkeboen, Leuven, and Mogstad, 2016). Omitting the fields of study may lead to apparent effects that may partly reflect that students or institutions with particular characteristics may specialize in fields that yield high (or low) labor market returns. As such, we include in X^c the shares of degrees awarded in each field of study in our sample according to the Classification of Instructional Programs (CIP) (out of the total 38 CIP fields of study, our sample includes 37 fields listed in Table S2 in the Supplement. We exclude PCIP45 (Social Sciences) as the reference field).

The student characteristics X^s consist of the following averages of demographic and socioeconomic indicators or variables: the shares of each available race category (Black, Hispanic, Asian, American Indian/Alaska Native, Native Hawaiian/Pacific Islander, two or more races, race is unknown, and non-resident alien - we omit White as the reference group), the shares of students how are female, dependent, have at least one post-secondary educated parent, and the average family income.

The empirical analysis sometimes imposes A_4 ($\sigma^2_{\varepsilon} \leq \kappa\sigma^2_{\tilde{U}}$ or $R^2_{W.X} \leq \kappa' \leq R^2_{W.U}$) and A_5 ($\tilde{R}^2_* \leq \tau$ or $R^2_{Y.(X',W)'} \leq R^2_* \leq \tau'$). Given the rich set of college and student characteristics in X, the estimates $\hat{R}^2_{W.X}$ and $\hat{R}^2_{Y.(X',W)'}$ are 0.891 and 0.814. To report bounds without requiring SAT to measure ability perfectly ($\kappa = 0$), we choose conservative κ and τ default values. We set $\kappa = 11.09$ so that the κ' estimate is $\hat{\kappa}' = 0.9 \geq 0.891$. Similarly, we set $\tau = 0.736$ so that $\hat{\tau}' = 0.95 \geq 0.814$. More generally, Section 6.5 conducts a sensitivity analysis that allows κ and τ to range over $[0, 30]$ and $[\bar{R}^2_{\tilde{Y}.\tilde{W}}, 1]$ where $\bar{R}^2_{\tilde{Y}.\tilde{W}} = \max\{r^2 : r \in CR^{\tilde{r}}_{1-\alpha_2}\}$.

6.4 Results

Table 3 reports bounds and point estimates under sequentially stronger assumptions. Column 3 reports the regression estimates $\hat{b}_{Y.(W,X')'}$ along with 95% confidence intervals in parentheses. This consistently estimates $(\phi + \delta, \beta')'$ if either $\kappa = 0$ (the average SAT score measures the average ability without error) or $\delta = 0$ (a student's ability does not directly affect his or her earnings). Next, we examine the consequences of deviating from $\kappa = 0$. Recall that if $\kappa \to \infty$ and ϕ may be nonzero then none of the coefficients are identified. Further, imposing A_6^+ ($\phi\delta \geq 0$) only yields the half-lines $\mathcal{F}^+$, $\mathcal{D}^+$, $\mathcal{G}^+$, and $\mathcal{B}_l^+$ with bounds that correspond to the regression estimand. (We focus on A_6^+ as opposed A_6^- ($\phi\delta \leq 0$) since we deem it plausible that the effects of college selectivity and ability on earnings are nonnegative). To improve on these bounds, columns 1 and 2 report the bounds under A_1-A_5 and A_1-A_6^+ respectively using the default setting $(\kappa, \tau) = (11.09, 0.736)$, i.e. $(\hat{\kappa}', \hat{\tau}') = (0.9, 0.95)$. We note that $\hat{R}^2_{\tilde{W}.\tilde{Y}}$ is small, 0.0164. Thus, the bounds under A_1-A_6^+ when $\kappa = 11.09$ and τ is either 0.736 or 1 coincide. Further, the standard bounds that set $\phi = 0$, $\kappa \to \infty$, and $\tau = 1$ are wide with especially wide confidence regions.

First, consider the returns to college selectivity and student ability. Under $\delta = 0$, the regression coefficient $\hat{b}_{Y.(W,X')'}$ in column 3 estimates that a 100 point increase in SATAvg (roughly the difference between Stanford and Boston College) increases a student's earnings 6 years after enrollment by $\$1,339$, with a 95% confidence region ($CR_{0.95}$) ($\$607, \$2,071$). As shown in column 1, ϕ and δ are not identified under A_1-A_5 and the bounds for $\phi + \delta$ are wide. However, A_1-A_6^+ yield considerably more informative bounds on ϕ, δ, and $\phi + \delta$ (and β below). In particular, the bounds in column 2 on the return to a 100 point increase in SATAvg or U are $[\$0, \$1,339]$, with $CR_{0.95}$ ($\$0, \$2,071$), and $[\$0, \$16,190]$, with $CR_{0.95}$ ($\$0, \$25,464$), respectively. Further, the bounds on $\phi + \delta$ are $[\$1,339, \$16,190]$ with $CR_{0.95}$ ($\$572, \$25,464$) (this bounds the total (direct and mediated by an increase in SATAvg) effect on mean earnings due to a 100 points increase in the average ability of the student cohort).

Next, we comment on the returns to the college and student characteristics. For certain characteristics, the sign of the effect is not recovered under any of the considered assumptions. This includes the college's control, whether the institution offers a graduate degree, and the net price. An intermediate case occurs when the $CR_{0.95}$ for an effect does not contain 0 under the assumption $\kappa = 0$ but includes 0 under A_1-A_5 and A_1-A_6^+ when $90\% \leq R^2_{U.W}$ and $R^2_* \leq 95\%$. Among the college characteristics X^c, this includes the effects of the enrollment size, cost, student debt, certain fields of study, and whether a student completes his or her degree within 6 years. Among the student characteristics X^s, this includes the student's family income. For example, the regression's estimate for the premium to majoring in Engineering relative to the Social Sciences is $\$12,224$ with $CR_{0.95}$ $(\$2,598, \$21,850)$ whereas the bounds for this premium under A_1-A_6^+ are $[\$5,512, \$12,224]$ with $CR_{0.95}$ $(-\$9,853, \$22,310)$ (Table S3 in the Supplement reports the estimates for all the CIP fields of study). Last, for certain effects, the sign is not very sensitive to deviations from $(\kappa, \tau) = (0, 1)$ but the magnitude may be. Among X^c, this includes whether a student has a federal student loan or a Pell grant and the instructional expenditures per student. For example, the effect of a $\$1,000$ increase in instructional expenditures on a student's earnings is bounded under A_1-A_6^+ by $[\$166, \$290]$, $CR_{0.95}$ $(\$1, \$440)$, with the upper bound corresponding to the regression estimate. Among X^s, this includes the Black, Hispanic, and Asian race shares and the Female share. First, we note that conditioning on SATAvg in a basic regression of the mean earnings on the race and gender shares (relative to White and Male) renders the otherwise negative and significant coefficients on the Black, Hispanic, and Female shares smaller and insignificant and the significant and positive coefficient on the Asian share smaller (see e.g. Neal and Johnson, 1996). As Table 3 shows, further accounting for the college and student characteristics and allowing for measurement error in how SAT measures ability (see e.g. Bollinger, 2003) bounds the coefficients on the Black, Hispanic, and Asian shares in the positive range (Monks (2000) reports similar patterns). Moreover, the regression estimates for the Female coefficient is $-\$12,700$ with $CR_{0.95}$ $(-\$18,509, -\$6.891)$ and the bounds under A_1-A_6^+ are slightly wider $[-\$13,094, -\$12,700]$ with $CR_{0.95}$ $(-\$22,507, -\$3,681)$.

Last, accounting for the fields of study reduces the magnitude of the bounds on the returns to college selectivity and certain college and student characteristics. This is shown in Table S4 in the Supplement when replicating Table 3 without conditioning on the fields of study. For instance, under A_1-A_6^+ and $(\hat{\kappa}', \hat{\tau}') = (0.9, 0.95)$, the return to attending an institution that offers a graduate degree is small and not significantly different from zero in Table 3 whereas, similar to Monks (2000), this return appears positive and larger $[\$1,916, \$2,220]$ with $CR_{0.95}$ $(\$898, \$3,352)$ in Table S4. Similarly, under A_1-A_6^+, the coefficient on Female becomes larger in magnitude, $[-\$19,697, -\$17,914]$ with $CR_{0.95}$ $(-\$26,163, -\$10,838)$, than is reported in Table 3 (see e.g. Turner and Bowen (1999), Zafar (2013), and Gemici and Wiswall (2014) who study the gender gap in major choices in the US.)

6.5 Sensitivity to κ and τ

Table 3 impose the default setting for $(\kappa, \tau) = (11.09, 0.736)$ which sets $(\hat{\kappa}', \hat{\tau}') = (0.9, 0.95)$. More generally, we conduct a sensitivity analysis that examines how the estimates change as κ and τ vary. Figure 2 illustrates this by plotting the bounds $\hat{\mathcal{F}}^+_{\kappa,\tau}$, $\hat{\mathcal{D}}^+_{\kappa,\tau}$, $\hat{\mathcal{B}}^+_{\kappa,\tau}$ (using the darker shade) and the 95% confidence regions $CR_{0.95}$ (using the lighter shade) for ϕ, δ, and β. To ease the presentation for β, we focus on the coefficients associated with $C150\_4$ (the completion within 150% of expected graduation time) and Female. The first panel in Figure 2 sets $\tau = 1$ and lets κ range over $[0, 30]$ (i.e. $\hat{R}^2_{W.X} = 0.891 \leq \kappa' \in [0.8945, 1]$). Since $\hat{\mathcal{F}}^+$, $\hat{\mathcal{D}}^+$, and $\hat{\mathcal{B}}^+$ do not depend on τ, the second panel sets κ to the default value $\kappa = 11.09$ and lets τ range from $[\bar{R}^2_{\tilde{Y}.\tilde{W}}, 1]$ (i.e. $\hat{R}^2_{Y.(W,X')'} = 0.814 \leq \tau' \in [0.8187, 1]$). Figure 2 illustrates how, unlike ϕ, the bounds and confidence regions for δ and β vary with (κ, τ). Further, it shows that the smallest integer κ (corresponding $\hat{\kappa}'$ value) such that the $CR_{0.95}$ for the coefficient on $C150\_4$ or Female contains 0 is 2 (0.927) or 18 (0.897) respectively. When $\kappa = 11.09$ and $\tau \in [\bar{R}^2_{\tilde{Y}.\tilde{W}}, 1]$, the $CR_{0.95}$ for the coefficient on $C150\_4$ (resp. Female) always (resp. never) contains 0. Thus, the bounds on the Female coefficient are wider and less sensitive to A_4 than the bounds on $C150\_4$ are.

6.6 Additional Analyses and Discussion

Section 6.6 reports additional analyses (in Tables S4-S8 of the Supplement) and extensions.

6.6.1 Earnings Outcomes

Section 6.4 considers the mean earnings of federally aided students who are non-enrolled and working 6 years after college entry. To include the unemployed in the analysis, we use the same specification and consider an alternative labor market outcome: the share of individuals, including those with 0 earnings, who are non-enrolled and earning more than \$25,000 per year, 6 years after college entry. This threshold "corresponds approximately to the median wage of workers ages 25 to 34 with only a high-school degree" (CEA, p. 25). The bounds on the coefficients on SATAvg and several college and student characteristics include zero or are small in magnitude. However, the effects of the fields of study and of having a loan or a Pell grant on the probability of earning more than \$25,000 per year are significant and of a similar direction than the mean earnings results (see Table S4).

Further, to gauge the consequences of excluding the enrolled students (e.g. those pursuing a graduate degree), we replicate the analysis using the cohort of students who enrolled in 2002 and contrast the results when the earnings outcomes are measured in 2008 compared to 2012 (6 or 10 years after enrollment). As described in Table S5, we construct the sample for the 2002 cohort as closely as possible to the 2007 cohort (CS reports a coarser race category definition for the 2002 cohort. Also, we use the earliest available data from academic years 2007-08 or 2008-09 for the average cost and net price and the loan and Pell grant shares). Tables S6 and S7 reports these results. The upper bound on the return to college selectivity is slightly larger over the longer horizon, $[\$0, \$2,377]$ with $CR_{0.95}$ $(\$0, \$3,191)$, and the returns to the other college and student characteristics are generally comparable over these two time horizons. We note that the earnings outcomes of students 6 years after enrollment capture the short run labor market returns of timely completion of an undergraduate degree. Whereas the earnings outcomes 10 years after enrollment embody the returns that may be channeled via attending graduate school or accumulating work experience.

6.6.2 Specification and Aggregation

The literature sometimes imposes a log-linear specification (constant percentage effect) for the individual earnings equation. CS reports aggregate data on several variables (e.g. $\frac{1}{N_i}\sum_{j=1}^{N_i} Y_{ij}$) but not on certain transformations of these variables (e.g. $\frac{1}{N_i}\sum_{j=1}^{N_i}\log(Y_{ij})$). Substituting $\log(\frac{1}{N_i}\sum_{j=1}^{N_i} Y_{ij})$ for $\frac{1}{N_i}\sum_{j=1}^{N_i}\log(Y_{ij})$ introduces a specification error in the sense that the bounds obtained using the aggregate equation are no longer guaranteed to correspond to the coefficients from the individual earnings equation. In the specification above, the variables enter in levels (similar to e.g. Kirkeboen, Leuven, and Mogstad, 2016). In this case, the coefficients from the individual and aggregate equations coincide. Nevertheless, Table S8 replicates the analysis for the aggregate earnings equation using a log-linear specification and reports qualitatively similar results (with a comparable $\hat{R}^2_{Y.(W,X')'}$, 0.851 versus 0.814), albeit these are less easily relatable to the individual earnings equation coefficients.

6.6.3 Connection to the Literature

The findings on the returns to college selectivity and characteristics are mixed. On the one hand, some studies document a positive return to certain college qualities. For example, Brewer, Eide, and Ehrenberg (1999) find a significant selection-adjusted wage and earnings premia (e.g. 6 or 10 years after high school graduation) to attending an elite (based on Barron's Profile of American Colleges) private college. Monks (2000, table 4) reports similar findings. For example, relative to competitive institutions (based on Barron's index), the wage premium to attending highly or most competitive institutions is 13.1%. Using a matching estimator, Black and Smith (2004, Table 7) find that college quality (measured by an index based on faculty salaries, freshman retention rate, and the average SAT score) has a positive effect on wage. For example, a move from the first to the fourth quartile of college quality leads to a 13.9% increase in wage for men (7.8% for women), albeit this is somewhat imprecisely estimated. Further, when comparing the earnings of individuals around the admission cutoff point, Hoekstra (2009) finds that attending a flagship state university leads to 22% higher earnings for white men 10 to 15 years after high school graduation. On the other hand, other studies do not find strong evidence for a large return to college selectivity. For example, Dale and Krueger (2002) find that students who attend varyingly selective colleges (measured by the average SAT score) after being admitted to equally selective colleges earn comparably. Dale and Krueger (2002, 2014) report similar findings using a "self-revelation" model in which the student's application record (e.g. the average SAT score at the colleges to which the student applied) reveals his or her ability. Further, Kirkeboen, Leuven, and Mogstad (2016) find that the effect of attending a more selective institution is small relative to the substantial effect that the field of study has on earnings.

The paper's empirical application and each of the above papers uses a different sample, specification, covariates, and identifying assumptions. Moreover, they report the effects of various interventions (e.g. a change in the average SAT score or in a quality index) on various outcomes (e.g. wage or earnings). As such, a formal comparison of the estimates is not straightforward. Nevertheless, to get a rough approximation, we focus on the return to college selectivity and contrast some of these estimates to the paper's upper bound $\$2,377$ for the 10 year return to 100 SATAvg points. For example, using the 2002 cohort, the difference in the average SATAvg across the lower and upper quartiles of institutions in our sample is $1,212 - 922 = 290$. Using the upper bound $\$2,377$, this difference increases earnings by at most $\$6,893$ which corresponds to 16.4% of the weighted (by college enrollment size) average earnings $\$41,849$ in the lower quartile of institutions. This upper bound admits some of the related estimates discussed above. Also, Hoekstra (2009, Section VI and Table 3) describes that it is likely that the applicants who were nearly accepted to the flagship state university attended a public college in-state and that the average SAT differential between the flagship university (where the average SAT score is between $1,000$ and $1,100$) and 7 alternative in-state public universities ranges from 65 to 147 points. Using the paper's upper bound $\$2,377$, a 147 points increase in SATAvg increases earnings by at most $\$3,494$ or 8.3% of the weighted average earnings $\$41,901$ among public institutions with an average SAT score between 853 and 953. This upper bound is smaller than the 22% local to the discontinuity estimate of the return to attending the flagship state university in Hoekstra (2009). Among other possibilities, this difference may be due to a nonlinearity in the return or to accounting for characteristics, such as the field of study or completing a degree, that may explain a part of the apparent return to college selectivity.

6.6.4 Future Work

The empirical illustration assumes that SAT measures ability with classical measurement error, $W_{ij} = U_{ij} + \varepsilon_{ij}$. The literature documents a high correlation between SAT scores and other tests, such as the Armed Services Vocational Aptitude Battery and the Raven's Advanced Progressive Matrices (see e.g. Frey and Detterman (2004) who argue in favor of the "appropriateness of the SAT as a measure of [general intelligence]"). An interesting extension would relax $W_{ij} = U_{ij}$ even further to allow ε_{ij} to be correlated with U_{ij} or X_{ij} or both. This would allow the error in how SAT proxies ability to statistically depend on the ability level or on covariates such as family income (for example, some studies document positive but small effects of coaching or repeating the test on SAT scores (DerSimonian and Laird, 1983; Vigdor and Clotfelter, 2003; Domingue and Briggs, 2009).

Further, the analysis assumes that the average SAT equivalent score measures college selectivity without error. It would be of interest to consider the more general model

$$Y_i = X_i'\beta + S_i\phi + U_i\delta + \eta_i \quad \text{with} \quad W_i = U_i + \varepsilon_i \quad \text{and} \quad W_i = \alpha S_i + \varsigma_i,$$

where W_i measures both the unobserved college selectivity S_i and the average ability U_i with error. Here, we let $\alpha = 1$ and $\varsigma_i = 0$ so that $W_i = S_i$ and focus on relaxing the selection on observables assumption by allowing the SAT score to measure ability with error. Section

C of the Supplement demonstrates how the extended model above introduces additional parameters to the system of (in)equalities encoded in $Var[(\tilde{Y}, \tilde{W})']$ which may therefore alter the paper's identification regions. We leave studying the sharp identification regions for the parameters in this extended model under restrictions analogous to A_4-A_6 to future work.

6.6.5 Discussion

The empirical illustration contributes to the literature by analyzing the CS aggregate data. While CS is rich and comprehensive, the analysis may inherit some of its limitations that are partly due to data aggregation. Further, the analysis imposes several assumptions which may fail, including a linear specification with homogenous slope coefficients or that endogeneity arises due to one variable U (ability) that is observed with classical measurement error. As such, the empirical results should be interpreted carefully if one suspects that these assumptions do not hold. Nevertheless, the analysis allows ability to freely depend on the college and student characteristics and does not require exogenous instruments. As a result, the coefficients are not point identified. Instead, the analysis studies the consequences of deviating from the selection on observables assumption by allowing the SAT score to proxy ability with error. The resulting bounds are wider and more sensitive to the extent of the measurement error in W for some variables, such as completing a degree, than others, such as gender or instructional expenditures per student.

7 Conclusion

This paper studies identifying the coefficients in a linear equation when data on the outcome Y, covariates X, and an error-laden proxy W for a latent variable U are available. We maintain that the error in the proxy is classical and relax the proxy exclusion restriction which sets the coefficient on W in the outcome equation to zero. This accommodates a leading setting for differential measurement error that occurs when the latent variable U and its proxy W may directly affect the outcome. First, we show that, without the proxy exclusion restriction, the effects of U, W, and X are not separately identified. This demonstrates the crucial role that the proxy exclusion restriction plays in ensuring the validity of the standard classical measurement error bounds. We then characterize the sharp identification regions for these effects under any configuration of three auxiliary assumptions. The first imposes an upper bound on the noise to signal ratio. The second places an upper bound on the coefficient of determination that would obtain in the outcome equation had W measured U without error. The third specifies whether the latent variable and its proxy affect the outcome in the same or the opposite direction, if at all. These auxiliary assumptions enable a sensitivity analysis that examines the consequences of deviating from the no measurement error assumption, controlling the fit of the model, and weakening the proxy exclusion restriction. Using the recent College Scorecard aggregate data, we illustrate our framework by studying the financial returns to college selectivity and characteristics as well as student characteristics when the average SAT score at an institution may directly affect earnings and serves as a proxy for the average ability of the student cohort. Useful extensions for future work would accommodate a nonlinear specification or multiple latent variables and (included) proxies.

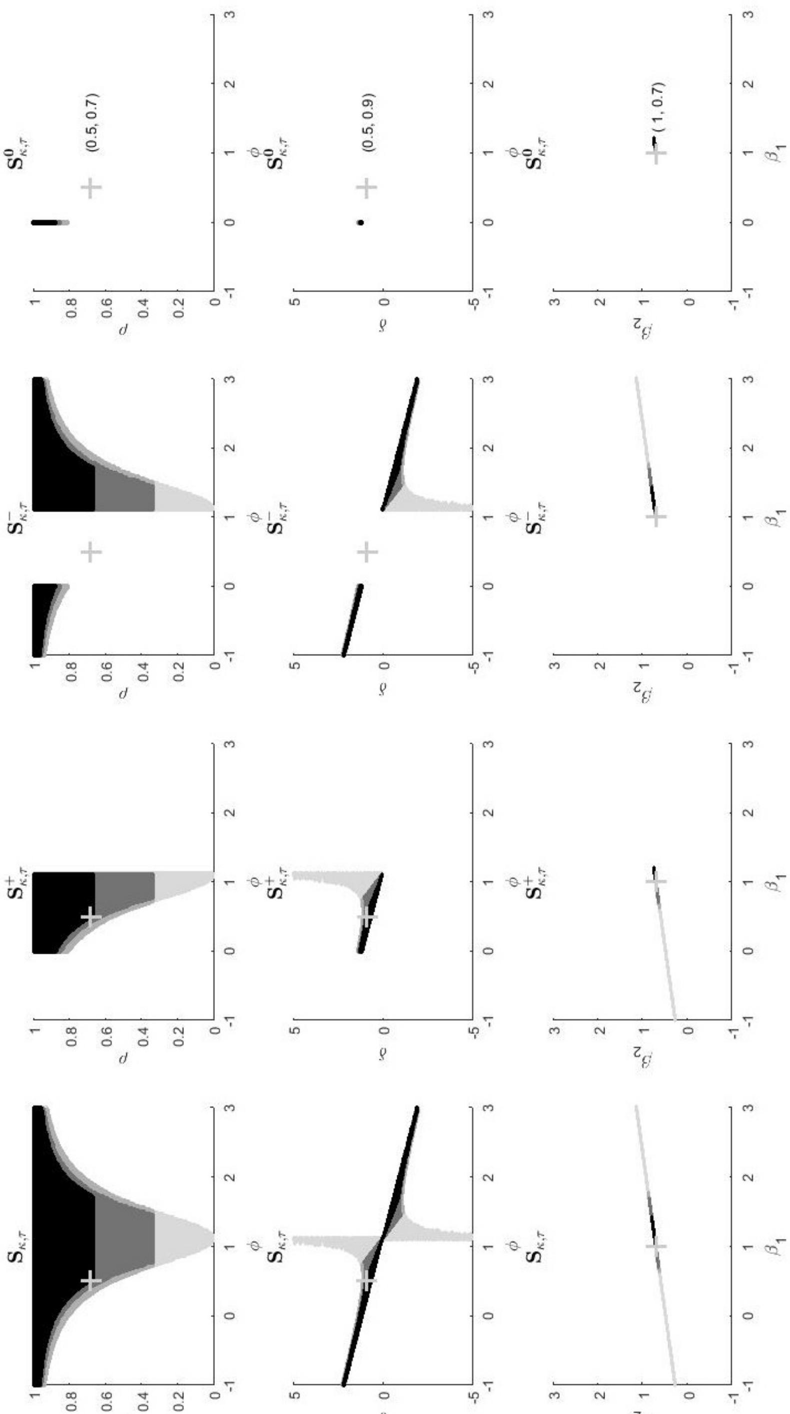

Figure 1: Identification regions for $(\kappa, \tau) = (+\infty, 1)$ (light), $(2, 1)$, $(2, 0.95)$, and $(0.5, 0.92)$ (dark).

Table 1: Numerical Example ($\rho = 0.685$, $\phi = 0.5$, $\delta = 0.9$, $\beta_1 = 1$, $\beta_2 = 0.7$)

	$\mathcal{S}_{\kappa,\tau}$	$\mathcal{S}^+_{\kappa,\tau}$	$\mathcal{S}^-_{\kappa,\tau}$	$\mathcal{S}^0_{\kappa,\tau}$	$b_{Y.(W,X')'}$
			$\kappa \to \infty, \tau = 1$		
ρ	$[0, 1]$	$[0, 1]$	$[0, 1]$	$[0.805, 1]$	
ϕ	$[-\infty, \infty]$	$[0, 1.116]$	$\mathbb{R}\backslash(0, 1.116)$	0	
δ	$[-\infty, \infty]$	$[0, \infty]$	$\mathbb{R}\backslash(0, 1.116)$	$[1.116, 1.386]$	
$\phi + \delta$	$[-\infty, \infty]$	$[1.116, \infty]$	$[-\infty, 1.386]$	$[1.116, 1.386]$	1.116
β_1	$[-\infty, \infty]$	$[-\infty, 1.207]$	$[1.010, \infty]$	$[1.010, 1.207]$	1.207
β_2	$[-\infty, \infty]$	$[-\infty, 0.745]$	$[0.702, \infty]$	$[0.702, 0.745]$	0.745
			$\kappa = 2, \tau = 1$		
ρ	$[0.333, 1]$	$[0.333, 1]$	$[0.333, 1]$	$[0.805, 1]$	
ϕ	$[-\infty, \infty]$	$[0, 1.116]$	$\mathbb{R}\backslash(0, 1.116)$	0	
δ	$[-\infty, \infty]$	$[0, 1.386]$	$\mathbb{R}\backslash(0, 1.116)$	$[1.116, 1.386]$	
$\phi + \delta$	$[0.340, 1.892]$	$[1.116, 1.892]$	$[0.340, 1.386]$	$[1.116, 1.386]$	
β_1	$[0.642, 1.771]$	$[0.642, 1.207]$	$[1.010, 1.771]$	$[1.010, 1.207]$	
β_2	$[0.622, 0.868]$	$[0.622, 0.745]$	$[0.702, 0.868]$	$[0.702, 0.745]$	
			$\kappa = 2, \tau = 0.95$		
ρ	$[0.333, 1]$	$[0.333, 1]$	$[0.333, 1]$	$[0.848, 1]$	
ϕ	$[-\infty, \infty]$	$[0, 1.116]$	$\mathbb{R}\backslash(0, 1.116)$	0	
δ	$[-\infty, \infty]$	$[0, 1.317]$	$\mathbb{R}\backslash(0, 1.116)$	$[1.116, 1.317]$	
$\phi + \delta$	$[0.447, 1.785]$	$[1.116, 1.785]$	$[0.447, 1.317]$	$[1.116, 1.317]$	
β_1	$[0.720, 1.693]$	$[0.720, 1.207]$	$[1.061, 1.693]$	$[1.061, 1.207]$	
β_2	$[0.639, 0.851]$	$[0.639, 0.745]$	$[0.713, 0.851]$	$[0.713, 0.745]$	
			$\kappa = 0.5, \tau = 0.92$		
ρ	$[0.667, 1]$	$[0.667, 1]$	$[0.667, 1]$	$[0.875, 1]$	
ϕ	$[-\infty, \infty]$	$[0, 1.116]$	$\mathbb{R}\backslash(0, 1.116)$	0	
δ	$[-\infty, \infty]$	$[0, 1.275]$	$\mathbb{R}\backslash(0, 1.116)$	$[1.116, 1.275]$	
$\phi + \delta$	$[0.818, 1.414]$	$[1.116, 1.414]$	$[0.818, 1.275]$	$[1.116, 1.275]$	
β_1	$[0.990, 1.423]$	$[0.990, 1.207]$	$[1.091, 1.423]$	$[1.091, 1.207]$	
β_2	$[0.698, 0.792]$	$[0.698, 0.745]$	$[0.720, 0.792]$	$[0.720, 0.745]$	

Population identification regions and regression estimands. $\frac{\sigma^2_{\varepsilon}}{\sigma^2_{\tilde{U}}} = 0.461$, $\tilde{R}^2_* = 0.918$, and $R^2_{\tilde{W}.\tilde{Y}} = 0.805$.

Table 2: Summary Statistics for the CS Sample of 1165 Institutions.

Variable	Name	Mean	Std. Dev.	Min	Max
Mean earnings among who work	MnEarnWnEP6	36985.24	9513.16	15700	102700
Share earning over $25K/year	Gt25KP6	0.6528	0.1077	0.162	0.918
College Characteristics					
Average SAT score	SATAvg	1052.76	119.93	726	1491
Private control indicator	ControlInd	0.6206	0.4854	0	1
Grad degree-awarding indicator	HDeg	0.8635	0.3434	0	1
Undergraduate enrollment	UGDS	5955.3	7143.88	178	56232
Average cost of attendance	CostT4	30090.46	11886.28	9917	57590
Average net price	NPT4	17638.82	6529.17	1081	39560
Percent with Federal student loan	PctFLoan	0.6032	0.1618	0.0334	1
Percent with Pell grant	PctPell	0.3679	0.1452	0.0738	0.9351
Median student debt	GDebtMdn	19598.26	3732.304	4500	35500
Percent of degrees in 38 fields	PCIP_##				
Percent of Education degrees	PCIP13	0.0804	0.0789	0	0.6452
Percent of Engineering degrees	PCIP14	0.0316	0.0903	0	0.9088
Expenditure per student	InExpFTE	9416.8	7578.22	1938	107380
Completion rate	C150_4	0.5489	0.1697	0.049	0.9779
Student Characteristics					
Share of Blacks	UGDSBlack	0.1281	0.1858	0	0.9955
Share of Hispanics	UGDSHisp	0.0774	0.1097	0	1
Share of Asians	UGDSAsian	0.0397	0.0593	0	0.5054
Share of females	Female	0.5869	0.1076	0.0773	0.986
Average family income	FamInc	70495.26	21358.66	17501.84	143865.7
Additional College and Student Characteristics in X					
10 region indicators	Region_##				
New England indicator	Region_1	0.0833	0.2764	0	1
Southeast indicator	Region_5	0.2592	0.4384	0	1
12 locale indicators	Locale_##				
City indicator	Locale_11	0.206	0.4046	0	1
Rural remote indicator	Locale_43	0.0069	0.0826	0	1
Minority-serving indicator	SpecMis	0.1554	0.3624	0	1
Women-only college indicator	WomenOnly	0.012	0.109	0	1
Religious affiliation indicator	RelAffilInd	0.4094	0.4919	0	1
Share of Whites	UGDSWhite	0.6409	0.2222	0	0.9666
% Nat. Hawaiian/Pacific Islander	UGDSNHPI	0.0023	0.0083	0	0.1448
% two or more races	UGSD2mor	0.015	0.0178	0	0.2617
% American Indian/Alaska Nat.	UGDSAian	0.0077	0.0209	0	0.325
% whose race is unknown	UGDSUnkn	0.0598	0.0721	0	0.675
Share of dependents	Dependent	0.7504	0.1645	0.1238	0.9886
Share of nonresident aliens	UGDSnRA	0.0291	0.0342	0	0.3617
% with tertiary-educated parent	ParEdPctPS	0.6556	0.1025	0.4108	0.9381

Table 3: The Returns to College Selectivity and Characteristics

$\hat{\kappa}' = 0.9$, $\hat{\tau}' = 0.95$	$\hat{\mathcal{S}}_{\kappa,\tau}$	$\hat{\mathcal{S}}^+_{\kappa,\tau}$	$\hat{b}_{Y(W,X')'}$
$10^{-2}\times$SATAvg	$[-\infty, \infty]$	$[0, 1.339]$	-
	$(-\infty, \infty)$	$(0, 2.071)$	
$10^{-2}\times$U (ability)	$[-\infty, \infty]$	$[0, 16.190]$	-
	$(-\infty, \infty)$	$(0, 25.464)$	
$10^{-2}\times$(SATAvg,U)	$[-28.155, 30.833]$	$[1.339, 16.190]$	1.339
	$(-31.208, 34.482)$	$(0.572, 25.464)$	$(0.607, 2.071)$
ControlInd	$[-2.684, 0.178]$	$[-1.253, -0.533]$	-1.253
	$(-7.945, 5.439)$	$(-3.666, 2.601)$	$(-2.739, 0.232)$
HDeg	$[-0.993, 0.566]$	$[-0.214, 0.179]$	-0.214
	$(-3.627, 3.279)$	$(-1.406, 1.763)$	$(-0.943, 0.516)$
$10^{-3}\times$UGDS	$[-0.254, 0.145]$	$[-0.155, -0.055]$	-0.055
	$(-0.439, 0.322)$	$(-0.285, 0.002)$	$(-0.108, -0.001)$
$10^{-3}\times$CostT4	$[-0.480, 0.735]$	$[-0.178, 0.127]$	0.127
	$(-0.761, 1.027)$	$(-0.421, 0.238)$	$(0.022, 0.233)$
$10^{-3}\times$NPT4	$[-0.448, 0.330]$	$[-0.059, 0.136]$	-0.059
	$(-0.761, 0.629)$	$(-0.203, 0.355)$	$(-0.196, 0.078)$
PctFLoan	$[-11.205, 18.775]$	$[3.785, 11.333]$	3.785
	$(-20.679, 28.552)$	$(0.508, 18.474)$	$(0.657, 6.913)$
PctPell	$[-15.704, -6.005]$	$[-13.296, -10.855]$	-10.855
	$(-33.301, 10.346)$	$(-24.043, -2.549)$	$(-16.127, -5.583)$
$10^{-3}\times$GDebtMdn	$[-0.639, 0.187]$	$[-0.226, -0.018]$	-0.226
	$(-0.995, 0.539)$	$(-0.357, 0.221)$	$(-0.351, -0.101)$
PCIP23 (English)	$[-33.702, -19.630]$	$[-30.209, -26.666]$	-26.666
	$(-82.597, 24.598)$	$(-61.141, 0.723)$	$(-43.311, -10.020)$
PCIP14 (Engineering)	$[-1.107, 25.554]$	$[5.512, 12.224]$	12.224
	$(-23.839, 48.679)$	$(-9.853, 22.310)$	$(2.598, 21.850)$
$10^{-3}\times$InExpFTE	$[0.044, 0.535]$	$[0.166, 0.290]$	0.290
	$(-0.117, 0.782)$	$(0.001, 0.440)$	$(0.147, 0.433)$
C150_4	$[-70.137, 89.279]$	$[-30.564, 9.571]$	9.571
	$(-84.730, 104.313)$	$(-56.614, 13.618)$	$(5.709, 13.433)$
UGDSBlack	$[-23.676, 35.776]$	$[6.050, 21.018]$	6.050
	$(-33.377, 46.374)$	$(2.817, 32.591)$	$(2.964, 9.136)$
UGDSHisp	$[-30.195, 47.237]$	$[8.521, 28.016]$	8.521
	$(-45.189, 64.714)$	$(3.950, 43.561)$	$(4.158, 12.884)$
UGDSAsian	$[12.793, 52.229]$	$[22.582, 32.511]$	32.511
	$(-8.444, 77.330)$	$(6.281, 45.563)$	$(20.054, 44.967)$
Female	$[-13.482, -11.918]$	$[-13.094, -12.700]$	-12.700
	$(-28.934, 5.022)$	$(-22.507, -3.681)$	$(-18.509, -6.891)$
$10^{-3}\times$FamInc	$[0.044, 0.140]$	$[0.068, 0.092]$	0.092
	$(-0.074, 0.257)$	$(-0.010, 0.146)$	$(0.044, 0.139)$

Y is $10^{-3}\times$MnEarnWnEp6, W is $10^{-2}\times$SATAvg, and U is scholastic ability scaled by 10^{-2}. In addition, X^c contains 8 region indicators, 11 locale indicators, the shares of the remaining CIP fields of study, indicators for whether the institution has a special mission, is a women only college, or has a religious affiliation and X^s contains the shares of the remaining race categories, non-resident aliens, and dependent students. 95% confidence regions are reported in parentheses.

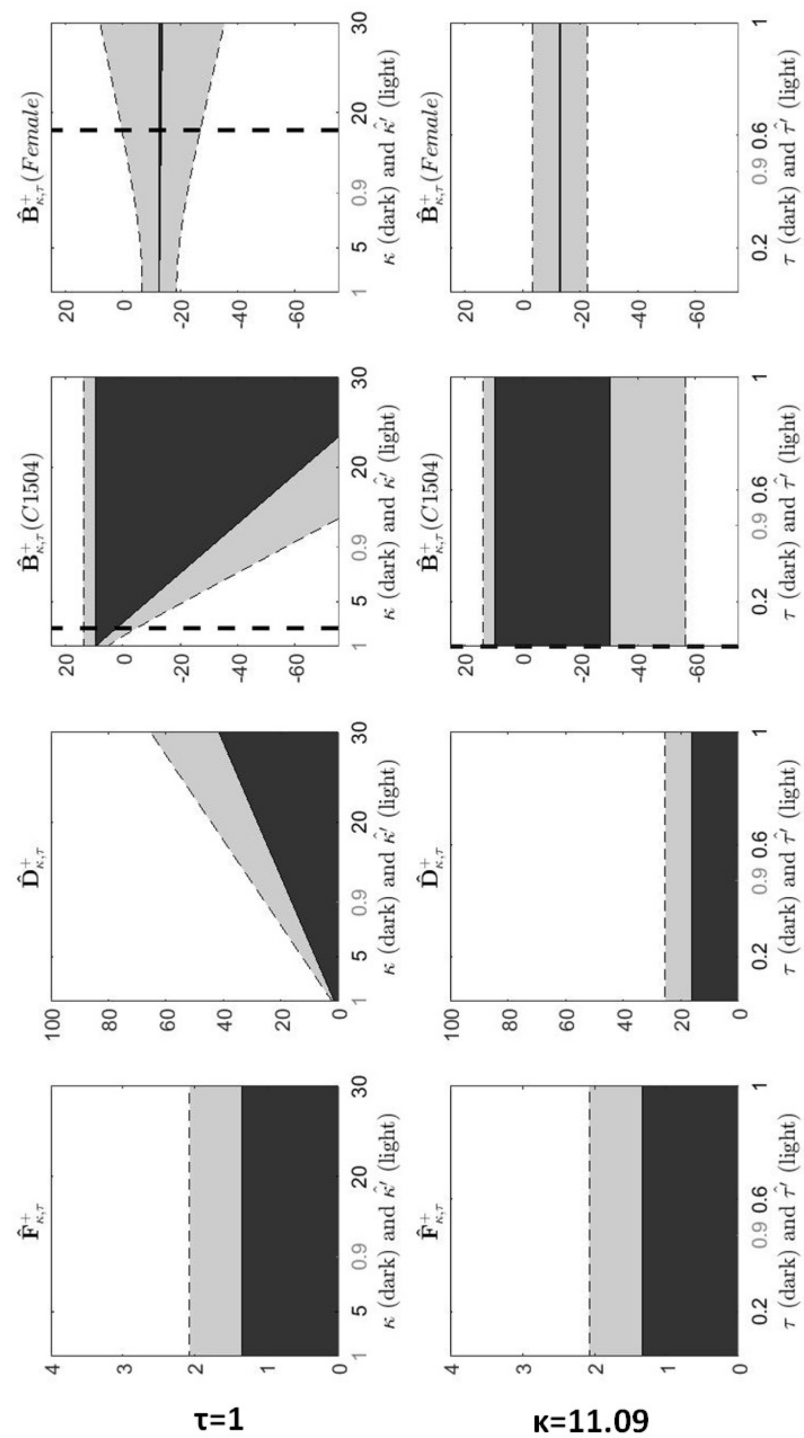

Figure 2: Bounds and 95% confidence regions when $\tau = 1$ and $\kappa \in [0, 30]$ (first panel) and $\kappa = 11.09$ and $\tau \in [\bar{R}^2_{\tilde{Y}.\tilde{W}}, 1]$ (second panel). The vertical thick dashed line indicates the smallest κ or τ value such that the confidence region contains zero.

References

Altonji, J., P. Arcidiacono, and A. Maurel (2016), "The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects." In E. Hanushek, S. Machin. and L. Woessmann (eds.). *Handbook of the Economics of Education.* Elsevier, 5, 305-396.

Bekker, P., A. Kapteyn, and T. Wansbeek (1987), "Consistent Sets of Estimates for Regressions with Correlated or Uncorrelated Measurement Errors in Arbitrary Subsets of all Variables," *Econometrica*, 55, 1223-1230.

Black, D. and J. Smith (2004), "How Robust is the Evidence on the Effects of College Quality? Evidence from Matching," *Journal of Econometrics*, 121, 99-124.

Black, D. and J. Smith (2006), "Estimating the Returns to College Quality with Multiple Proxies for Quality," *Journal of Labor Economics*, 24, 701-728.

Bollinger, C. (1996), "Bounding Mean Regressions When a Binary Regressor is Mismeasured," *Journal of Econometrics*, 73, 387-399.

Bollinger, C. (2003), "Measurement Error in Human Capital and the Black-White Wage Gap," *Review of Economics and Statistics*, 85, 578-585.

Bound, J., Brown, C., Mathiowetz, N. (2001), "Measurement error in survey data," in: Heckman, J. and Leamer, E. eds., *Handbook of Econometrics*, vol. 5.

Brewer, D., E. Eide, and R. Ehrenberg (1999), "Does It Pay to Attend an Elite Private College? Cross-Cohort Evidence on the Effects of College Type on Earnings," *The Journal of Human Resources*, 34, 104-123.

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

Chen, X., H. Hong, and D. Nekipelov (2011) "Nonlinear Models of Measurement Errors," *Journal of Economic Literature*, 49, 901–937.

Chesher, A. (1991), "The Effect of Measurement Error," *Biometrika*, 78, 451-462.

Chernozhukov, V., R. Rigobon, and T. Stoker (2010), "Set Identification and Sensitivity Analysis with Tobin Regressors," *Quantitative Economics*, 1, 255–277.

Conley, T., C. Hansen, and P. Rossi (2012), "Plausibly Exogenous," *Review of Economics and Statistics*, 94, 260–272.

Council of Economic Advisors (2015), "Using Federal Data To Measure and Improve the Performance of U.S. Institutions of Higher Education," Executive Office of the President of the Unites States.

Dale, S. and A. Krueger (2002), "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables," *Quarterly Journal of Economics*, 117, 1491-1527.

Dale, S. and A. Krueger (2014), "Estimating the Effects of College Characteristics over the Career Using Administrative Earnings Data," *Journal of Human Resources*, 49, 323-358.

Deming, D., C. Goldin, and L. Katz (2012), "The For-Profit Postsecondary School Sector: Nimble Critters or Agile Predators?" *Journal of Economic Perspectives*, 139-64.

DerSimonian R. and N. Laird (1983) "Evaluating the Effect of Coaching on SAT Scores:

A Meta-Analysis" *Harvard Educational Review*, 53, 1-15.

DiTraglia, F. and C. Garcia-Jimeno (2016), "A Framework for Eliciting, Incorporating, and Disciplining Identification Beliefs in Linear Models," NBER Working Paper Series.

Domingue, B. and D. Briggs (2009) "Using Linear Regression and Propensity Score Matching to Estimate the Effect of Coaching on the SAT," *Multiple Linear Regression Viewpoints*, 35, 12-29.

Erickson, T. (1993), "Restricting Regression Slopes in the Errors-in-Variables Model by Bounding the Error Correlation," *Econometrica*, 61, 959-969.

Erickson, T. and T. Whited (2002), "Two-Step GMM Estimation of the Errors-in-Variables Model Using High-Order Moments," *Econometric Theory*, 18, 776-799.

Frey M. and D. Detterman (2004), "Scholastic Assessment or g? The Relationship Between the Scholastic Assessment Test and General Cognitive Ability," *Psychological Science*, 15, 373-378.

Frisch, R. (1934). *Statistical Confluence Analysis by Means of Complete Regression Systems*. Oslo, Norway: University Institute of Economics.

Gemici, A. and M. Wiswall (2014), "Evolution of Gender Differences in Post-Secondary Human Capital Investments: College Majors," *International Economic Review*, 55, 23-56.

Geronimus, A., J. Bound, and L. Neidert (1996), "On the Validity of Using Census Geocode Characteristics to Proxy Individual Socioeconomic Characteristics," *Journal of the American Statistical Association*, 91, 529-537.

Gini, C. (1921), "Sull'Interpolazione di una Retta Quando i Valori della Variabile Indipendente Sono Affetti da Errori Accidentali," *Metroeconomica*, 1, 63-82.

Gundersen, C., B. Kreider, and J. Pepper (2012) "The Impact of the National School Lunch Program on child health: A Nonparametric Bounds Analysis," *Journal of Econometrics*, 166, 79–91.

Hoekstra, M. (2009), "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach," *Review of Economics and Statistics*, 91, 717–724.

Hoxby, C. (2009), "The Changing Selectivity of American Colleges," *Journal of Economic Perspectives*, 23, 95–118.

Hu, Y. (2008) "Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables: A General Solution," *Journal of Econometrics*, 144, 27–61.

Hyslop, D. and G. Imbens (2001), "Bias From Classical and Other Forms of Measurement Error," *Journal of Business and Economic Statistics*, 19, 475-481.

Imai, K. and T. Yamamoto (2010), "Causal Inference with Differential Measurement Error: Nonparametric Identification and Sensitivity Analysis," *American Journal of Political Science*, 54, 543-560.

Kirkeboen, L., E. Leuven, and M. Mogstad (2016) "Field of Study, Earnings, and Self-Selection," *Quarterly Journal of Economics*, 1057–1111.

Klepper, S. and E. Leamer (1984), "Consistent Sets of Estimates for Regressions with

Errors in All Variables," *Econometrica*, 52, 163-184.

Klepper, S. (1988), "Regressor Diagnostics for the Classical Errors-in-Variables Model," *Journal of Econometrics*, 37, 225-250.

Kline, B. and E. Tamer (2016), "Bayesian Inference in a Class of Partially Identified Models," *Quantitative Economics*, 7, 329–366.

Kreider , B., J. Pepper, C. Gundersen, and D. Jolliffe (2012) "Identifying the Effects of SNAP (Food Stamps) on Child Health Outcomes When Participation Is Endogenous and Misreported," *Journal of the American Statistical Association*, 107, 958-975.

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

Lewbel, A. (2007), "Estimation of Average Treatment Effects with Misclassification," *Econometrica*, 75, 537–551.

Mahajan, A. (2006), "Identification and Estimation of Regression Models with Misclassification," *Econometrica*, 74, 631-665.

Monks, J. (2000), "The returns to individual and college characteristics Evidence from the National Longitudinal Survey of Youth," *Economics of Education Review*, 19, 279–289.

Neal, D. and W. Johnson (1996), "The Role of Premarket Factors in Black-White Wage Differences," *Journal of Political Economy*, 104, 869-895.

Shi, X. and M. Shum (2015), "Simple Two Stage Inference for a Class of Partially Identified Models," *Econometric Theory*, 31, 493–520.

Turner, S. and W. Bowen (1999), "Choice of Major: The Changing (Unchanging) Gender Gap," *Industrial and Labor Relations Review*, 52, 289-313.

van der Vaart, A. (2000). *Asymptotic Statistics*. Cambridge University Press

Vigdor J. and C. Clotfelter (2003) "Retaking the SAT," *Journal of Human Resources*, 38, 1-33.

White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817-838.

White, H. (2001). *Asymptotic Theory for Econometricians*. New York: Academic Press.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press.

Zafar, B. (2013), "College Major Choice and the Gender Gap," *Journal of Human Resources*, 48, 545-595.