Measurement Error without Exclusion: the Returns to College Selectivity and Characteristics

Karim Chalak*†
University of Virginia

Daniel Kim
University of Pennsylvania

May 22, 2017

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 the classical error-in-variables assumptions and relax the assumption that the proxy is excluded from the outcome equation. This enables the proxy to directly impact the outcome and allows for differential measurement error. Without the exclusion restriction, we first show that the coefficients on the latent variable, the proxy, and the covariates are not identified. Then, we derive the sharp identification regions for these coefficients under either or both of two auxiliary restrictions. The first restriction weakens the assumption of "no measurement error" by imposing an upper bound on the net of the covariates "noise to signal" ratio, i.e. the ratio of the variance of the measurement error to the variance of the latent variable given the covariates. The second restriction 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 data, we employ this framework to study the financial returns to college selectivity and characteristics. Here, college selectivity, defined as the average SAT score of a student cohort, serves as a proxy for the latent average scholastic ability and is included in the average earnings equation. We obtain an informative upper bound on the return to college selectivity which becomes smaller upon conditioning on the instructional expenditures per student and the completion rate. Further, we obtain tight bounds on the returns to the college characteristics and find that conditioning on the composition of majors reduces the magnitude of the bounds on the effect of some of these characteristics, such as the gender composition.

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 participants in the New York Camp Econometrics XII, the tenth annual West Coast Experiments Conference, and the seminars at Boston College and the University of Virginia as well as Jennifer Jones and John Pepper. Any errors are the responsibility of the authors.

1 Introduction

The decisions of whether and where to pursue a college degree are often life-altering and merit careful consideration. For instance, several news articles offer advice to students who are applying to, or choosing among, universities. A Wall Street Journal article (Prior, 2014) suggests to an aspiring entrepreneur, that "a big name doesn't always matter much," to be "mindful of debt," and that "you don't need to major in business." A New York Times article (Bruni, 2013) advises students to "favor schools with higher percentages of foreigners" and answers the question "does brand matter?" inconclusively. A National Public Radio commentary (O'Connell, 2007) suggests that "[college] size matters" and that "a name-brand college will not guarantee your success." An Atlantic article (Thompson, 2014), finds that "smaller schools with high concentrations of computer science and engineering students near large cities with thriving technology scenes [...] dominate the list" of schools with highest-earnings graduates. In addition to several important nonpecuniary outcomes, a student who is deliberating among colleges may consider the expected financial return in the labor market. What is the empirical evidence about the financial returns to college selectivity and characteristics in the United States?

One challenge in answering this question is the limited availability of comprehensive data on the education characteristics and earnings of individuals in the United States (see e.g. Kirkeboen, Leuven, and Mogstad (2016) who make use of rich administrative data in Norway to study the payoffs to different types of postsecondary education). As a result, the existing studies often rely on survey data on individuals attending a narrow subset of US colleges (e.g. Dale and Krueger, 2002, 2014) or are focused on data pertaining to students who applied to a particular public institution (e.g. Hoekstra, 2009). A second challenge in answering this question arises because students with higher unobserved scholastic ability and motivation may earn more and sort into colleges that are more selective or that have particular characteristics. This sorting makes it difficult to identify the financial returns to college selectivity and characteristics. To address this endogeneity problem, the literature relies on modelling how students choose a college (e.g. Brewer, Eide, and Ehrenberg, 1999), matching students based on observed characteristics such as their SAT scores and college application records (e.g. Dale and Krueger, 2002, 2014; Black and Smith, 2004), or contrasting the earnings of students who are just below a university's admission cutoff with the earnings of students who are just above the cutoff (see e.g. Hoekstra, 2009). The literature's findings are mixed; some studies report evidence for a positive return to college selectivity (e.g. Brewer, Eide, and Ehrenberg, 1999; Black and Smith, 2004; Hoekstra, 2009) while others do not (e.g. Dale and Krueger, 2002, 2014).

To help inform students and remedy the lack of information on college characteristics and student outcomes, the Obama Administration released the College Scorecard (CS) data which reports data on postsecondary institutions in the US. The CS data is aggregated at the institution level and includes information on the institution, student body, affordability, admission and academic attributes, and earnings outcomes. While the CS data has some limitations, as we discuss below, 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 develop a parsimonious econometric framework that is suitable for analyzing the CS data and employ this framework and data to measure the returns to college selectivity and characteristics.

Specifically, following the literature (e.g. Dale and Krueger, 2002, 2014; Hoxby, 2009), we define college selectivity as the average SAT equivalent score of the cohort of enrolled students. We then model the logarithm of the average earnings of the cohort six years after enrolling in the college as a function of the college's selectivity, a rich set of aggregate characteristics of the college, and the average unobserved ability of the cohort. We allow the average unobserved ability of the cohort to freely statistically depend on the aggregate college characteristics. To account for the possibility that students select into colleges based on unobserved ability, we let the average SAT score (i.e. the college selectivity) serve a second role as an error-laden proxy for the average unobserved ability. We then characterize the identification regions of the earnings equation coefficients and study the sensitivity of these regions to two types of restrictions. The first restricts the extent of the measurement error in how the average SAT score proxies the average ability. The second restricts the direction of the effects of the college selectivity and the average ability on the average earnings.

Thus, this paper's econometric method studies the identification of the effects of the latent variable U (average ability), the error-laden proxy W (average SAT score), and the covariates X (e.g. gender and major composition) on the outcome Y (average wage) when the proxy is included in the linear outcome equation. In this case, the measurement error is "differential" since the proxy may help predict the outcome even after conditioning on the latent variable. Specifically, the paper puts forward new 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). We apply our results to study the returns to college selectivity and characteristics when college selectivity can directly affect the average earnings, beyond its role as a proxy for the average unobserved ability. More generally, this paper's econometric analysis applies to many other contexts in which a researcher suspects that an error-laden proxy for a latent variable may have a direct impact on the outcome[1].

Compared to the literature, our econometric framework imposes minimal assumptions. In particular, we maintain the classical error-in-variables assumptions and relax the common assumption that the proxy is excluded from the outcome equation. Recall that when the proxy exclusion restriction is imposed, substituting a proxy for the latent variable is not harmless: a regression of the outcome on the proxy and covariates does not identify the effects of the latent variable and covariates. In particular, the regression coefficient associated with the proxy suffers from "attenuation bias." Nevertheless, one can obtain sharp bounds for these coefficients (see e.g. Klepper and Leamer, 1984; Bollinger, 2003). How do these sharp bounds change when we allow the proxy to be included in the outcome equation? We advance the econometrics literature on measurement error by removing the proxy exclusion restriction. First, we characterize the joint sharp identification region for the coefficients on the latent variable, the proxy, and the covariates as well as for the net of the covariates "signal to total variance ratio," i.e. the ratio of the variance of the latent variable given the covariates to the variance of the proxy given the covariates. When projecting this joint identification region onto the supports of the coefficients associated with the latent variable, the proxy, and the covariates, we show that these coefficients are not separately identified. This demonstrates the crucial role that the proxy exclusion restriction plays in ensuring

[1]For example, the GPA of an employee may serve as a proxy for her latent ability and may have a direct impact on her wage when it is known to the employer. Similarly, a patient may alter her behavior after learning the result of a medical test that serves as a proxy for a latent health factor.

the validity of the standard bounds discussed above. To proceed, we derive the joint and projected sharp identification regions under either or both of two auxiliary restrictions. The first restriction weakens the benchmark assumption of "no measurement error" by imposing an upper bound on the net of the covariates "noise to signal" ratio, i.e. the ratio of the variance of the measurement error to the variance of the latent variable given the covariates. For example, restricting this ratio to be less than 1 assumes that, given the covariates, at most (at least) half of the variance in the average SAT score, i.e. college selectivity, is due to the measurement error (the average ability). By varying this upper bound, a researcher can conduct a sensitivity analysis of how the measurement error in the proxy affects the sharp identification regions. The second restriction 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. For instance, in the empirical analysis, we sometimes assume that the effects of the average ability and the college selectivity on the average earnings have the same sign (e.g. non-negative). Nevertheless, we do not require a particular set of assumptions; rather, we establish the mapping from each one or combination of these auxiliary assumptions to the sharp identification set.

After discussing estimation and inference, we employ our framework to analyze the CS data. Under our two auxiliary assumptions, we obtain an informative upper bound on the returns to college selectivity and average ability. In particular, given the college characteristics and the major composition, a 100 points increase in the average SAT score (roughly the difference between Stanford and Boston College) leads to at most a 4.8% increase in the average earnings 6 years after enrollment. Further, these upper bounds become smaller when conditioning on the instructional expenditures per student and the completion rate. Specifically, the upper bound on the return to selectivity drops to 2.8% per 100 average SAT points. In addition, we obtain tight bounds on the financial returns to the characteristics of the institution, student body, and affordability as well as the major composition, instructional expenditures, and the completion rate and contrast these bounds with the regression estimates. Last, we show how conditioning on the major composition reduces the magnitude of the bounds on the effects of some university characteristics, such as the gender composition or whether a university offers a graduate degree.

This paper is organized as follows. Section 2 specifies the data generating process and assumptions and introduces the notation. Section 3 studies the identification of the outcome equation coefficients and the net of the covariates signal to total variance ratio in the cases of classical measurement error (under the exclusion restriction) and differential measurement error (without the exclusion restriction). It characterizes the joint and projected sharp identification regions when none, either, or both of the auxiliary restrictions are imposed. Section 4 illustrates these identification results using a numerical example. Section 5 discusses estimation and inference. Section 6 applies our framework to analyzes the CS data and to study the returns to college selectivity and characteristics. Section 7 concludes. Mathematical proofs are gathered in the Appendix.

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 generate the random vector* X *and variables* η, ε, U, W, *and* Y *such that*

$$Y = X'\beta + W\phi + U\delta + \eta \tag{S}$$

$$W = U + \varepsilon \tag{P}$$

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

We are interested in identifying the slope coefficients δ, ϕ, and β in equation (S). These are the ceteris paribus causal effects of the latent variable U, the proxy W, and the covariates X on the outcome Y respectively. Note that A_1 allows, but does not require, W to directly affect Y. Further, A_1 decomposes the proxy W (e.g. the average SAT score) into the "signal" component[2] U (e.g. the average ability) and the "noise" or error ε.

One difficulty for identification is due to U being unobserved and correlated with W and possibly X. Nevertheless, we maintain two standard assumptions on the other unobservables

[2]Here, the structure $Y = X'\beta + W\phi + V\gamma + \eta$ and $W = V\psi + \varepsilon$, with V unobserved, is observationally equivalent to (S, P). Provided the scale $\psi \neq 0$, only the ratio $\delta \equiv \frac{\gamma}{\psi}$ of the coefficients on V may be (partially) identified. To ease the notation, we use the simpler representation that defines $U \equiv V\psi$.

η and ε. In particular, A_2 assumes that the "disturbance" η is uncorrelated[3] with[4] $(X', U)'$.

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

Further, the measurement error ε is uncorrelated with (X', U, η).

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

When $\phi = 0$, A_1-A_3 are the classical error-in-variables assumptions (see e.g. Wooldridge, 2002, p. 80). In this case, additional restrictions on higher order moments of η, ε, U, and X can point identify $(\beta', \delta)'$ (see e.g. Lewbel, 1997; Erickson and Whited, 2000). We do not impose these stronger assumptions. Specifically, we study partial identification under the weak uncorrelation[5] assumptions A_2-A_3 and the linear specification in A_1.

We relax these benchmark assumptions by not requiring the exclusion restriction[6] that sets $\phi = 0$. This leads to a second difficulty for identification. 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. We allow ϕ to be nonzero and

[3]The analysis can be generalized to be conditional on additional covariates; we forgo this for simplicity.

[4]Thus, a regression of Y on $(1, X', W)'$ would point identify $(\beta', \phi + \delta)'$ had U been observed without measurement error. In the case of classical measurement error with $\phi = 0$, DiTraglia and Garcia-Jimeno (2017) study a setting in which U, and the instruments for U, may be correlated with η. We leave studying systems in which (X', U) are endogenous (i.e. correlated with η) and ϕ may be nonzero to other work.

[5]For example, if (η, ε) is mean independent of X then one may point identify β and δ by generating a sufficient number of instruments as functions of X. In this case, one may also relax the parametric specification in A_1. We leave studying the identification of average effects under such stronger mean independence assumptions to other work.

[6]Some papers study relaxing A_2-A_3 while maintaining the exclusion restriction $\phi = 0$. For example, Erickson (1993) assumes that $Cov[(\varepsilon, \eta)', (X', U)'] = 0$ and relaxes the assumption $Cov(\varepsilon, \eta) = 0$ by imposing a lower and upper bounds on $Corr(\varepsilon, \eta)$. Also, Hyslop and Imbens (2001) assumes that W is an optimal prediction of U and thus that ε is uncorrelated with W and correlated with U.

study the consequences of deviating from the exclusion restriction, imposed in the classical error-in-variables assumptions, on the identification[7] of ϕ, δ, and β. For instance, this enables us to study the return to college selectivity when the average SAT score W serves as a proxy for the average scholastic ability U and may directly affect the average earnings Y.

We briefly note that if U, W, and ε are non-scalar vectors of the same dimension in A_1-A_3 then additional sign restrictions are needed to guarantee that non-trivial bounds exist for δ and the elements of β even if one assumes that[8] $\phi = 0$ (see e.g. Klepper and Leamer, 1984; Bollinger, 2003). To keep the scope of the analysis focused, we study relaxing the proxy exclusion restriction in a linear equation with a scalar U and W and leave considering non-linear or higher-dimensional generalizations, which may be technically and computationally nontrivial, to other work.

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, when B and C are of equal dimension with $\sigma_{C,B}$ nonsingular, we use the following succinct notation for the linear instrumental variable (IV) regression estimand and residual

$$R_{A.B|C} \equiv \sigma_{C,B}^{-1}\sigma_{C,A} \quad \text{and} \quad \epsilon'_{A.B|C} \equiv [A - E(A)]' - [B - E(B)]' R_{A.B|C}$$

so that by construction $E(\epsilon_{A.B|C}) = 0$ and $Cov(C, \epsilon_{A.B|C}) = 0$. If $B = C$ we obtain the linear regression estimand $R_{A.B} \equiv R_{A.B|B}$ and residual $\epsilon_{A.B} \equiv \epsilon_{A.B|B}$. For example, $R_{Y.X}$ 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 $\mathfrak{R}^2_{A.B} \equiv \sigma_A^{-2}(\sigma_{A,B}\sigma_B^{-2}\sigma_{B,A})$ denote the population coefficient of determination (R-squared) from a regression of A on B.

Under A_1-A_3, we have that $Cov[(\eta, \varepsilon)', X] = 0$. Throughout, we also assume that $Var(X)$ is nonsingular unless otherwise specified. Thus, by projecting W and

$$Y = X'\beta + W(\phi + \delta) - \varepsilon\delta + \eta \tag{1}$$

[7] Our results complement the results in Imai and Yamamoto (2010) who study bounding the average effect of a binary misclassified treatment on a binary outcome under alternative assumptions on the differential measurement error.

[8] Here, W can be viewed as a proxy for both W and U. Then the assumption that the latent variables (W, U) and the measurement error $(0, \varepsilon)$ are uncorrelated, in e.g. Klepper and Leamer (1984) and Bollinger (2003), fails.

onto X, we obtain that $R_{W.X} = R_{U.X}$ and

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

Using the following shorthand notation for the residuals from a regression on $(1, X')'$:

$$Y^* \equiv \epsilon_{Y.X}, \qquad W^* \equiv \epsilon_{W.X}, \quad \text{and} \;\; U^* \equiv \epsilon_{U.X},$$

we employ the convenient system of projected linear equations:

$$Y^* = W^*\phi + U^*\delta + \eta, \tag{S^*}$$

$$W^* = U^* + \varepsilon, \tag{P^*}$$

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 Restrictions

In addition to A_1-A_3, we consider two additional plausible restrictions R_1 and R_2 that weaken two common benchmark assumptions. Unlike A_1-A_3, which we impose throughout the paper, we do not require the auxiliary assumptions R_1 and R_2. Instead, we study the identification gain that may result when the researcher imposes either one or both restrictions.

The first auxiliary restriction 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, i.e. the ratio of the variances of ε and U^* (see e.g. Klepper and Leamer (1984) who consider a similar restriction in the case of classical measurement error with $\phi = 0$).

Assumption R_1 *Bounded Net of X Noise to Signal Ratio: $\sigma^2_\varepsilon \leq \kappa\sigma^2_{U^*}$ 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, after projecting on X, the variance of the measurement error is at most as large as the variance of U, $\sigma^2_\varepsilon \leq \sigma^2_{U^*}$. From A_1-A_3 and equation (P^*), we have $\sigma^2_{W^*} = \sigma^2_{U^*} + \sigma^2_\varepsilon$. Thus, R_1 sets an upper $\frac{\kappa}{1+\kappa}$ on the share of $\sigma^2_{W^*}$ that is due to σ^2_ε. Equivalently, R_1 sets a lower bound $\frac{1}{1+\kappa}$ on ρ, the share of $\sigma^2_{W^*}$ (e.g. the variance of the average SAT score, net of X) that is due to $\sigma^2_{U^*}$ (e.g. the variance of the average ability, net of X):

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

We refer to ρ as the net of X "signal to total variance ratio." Applying equation (20) in DiTraglia and Garcia-Jimeno (2017) (see also Dale and Krueger, 2002, p. 1514) shows that

$$\rho \equiv \frac{\sigma^2_{U^*}}{\sigma^2_{W^*}} = \frac{\sigma^2_U - \sigma_{U,X}\sigma_X^{-2}\sigma_{X,U}}{\sigma^2_W - \sigma_{W,X}\sigma_X^{-2}\sigma_{X,W}} = \frac{\mathfrak{R}^2_{W.U} - \mathfrak{R}^2_{W.X}}{1 - \mathfrak{R}^2_{W.X}},$$

where we use $\sigma_{W,X} = \sigma_{U,X}$ and $\mathfrak{R}^2_{W.U} = \frac{\sigma^2_U}{\sigma^2_W}$. Specifically, the "reliability ratio" $\mathfrak{R}^2_{W.U}$ is a weighted average of $\mathfrak{R}^2_{W.X}$ and 1 with weight ρ:

$$\mathfrak{R}^2_{W.U} = (1-\rho)\mathfrak{R}^2_{W.X} + \rho. \tag{3}$$

Since $0 \leq \rho \leq 1$, we have $\mathfrak{R}^2_{W.X} \leq \mathfrak{R}^2_{W.U} \leq 1$. If $\sigma^2_\varepsilon = 0$ (no measurement error) then $\rho = 1$ and $\mathfrak{R}^2_{W.U} = 1$. Further, if $\sigma^2_{U^*} = 0$ (U and X are perfectly collinear) then $\rho = 0$ and $\mathfrak{R}^2_{W.U} = \mathfrak{R}^2_{W.X}$. It follows that R_1 is equivalent to $\sigma^2_\varepsilon \leq \tau\sigma^2_U$ or $\frac{1}{1+\tau} \leq \mathfrak{R}^2_{W.U}$ where $\tau = \frac{\kappa(1-\mathfrak{R}^2_{W.X})}{1+\kappa\mathfrak{R}^2_{W.X}}$. The closer one sets $\frac{1}{1+\tau}$ to $\mathfrak{R}^2_{W.X}$ in $\mathfrak{R}^2_{W.X} \leq \frac{1}{1+\tau} \leq \mathfrak{R}^2_{W.U}$, the worse U is allowed to predict W (i.e. the larger σ^2_ε and the smaller $\mathfrak{R}^2_{W.U}$ can be) and the better X is allowed to predict U (i.e. the smaller $\sigma^2_{U^*}$ and thus ρ in equation (3) can be).

A researcher can impose a lower bound on either $\mathfrak{R}^2_{W.U}$ or ρ. While fixing τ in $\sigma^2_\varepsilon \leq \tau\sigma^2_U$ restricts the variance of the measurement error ε in W irrespective of X, fixing κ in $\sigma^2_\varepsilon \leq \kappa\sigma^2_{U^*}$ restricts the variance of ε relative to that of U^* and thus to how well X predicts U. In particular, τ determines that $\frac{1}{1+\tau} \leq \mathfrak{R}^2_{W.U}$ whereas κ determines the minimum improvement in $\mathfrak{R}^2_{W.U}$ over $\mathfrak{R}^2_{W.X}$ (e.g. for a given κ, the better the covariates predict the average SAT score, the better one requires the average ability to predict the average SAT score). For example, setting $\kappa = 1$ assumes that $\mathfrak{R}^2_{W.U}$ is at least half as close to 1 than $\mathfrak{R}^2_{W.X}$ is. We employ R_1 to conduct a sensitivity analysis by letting κ (or τ) range over a domain that deviates from the no measurement error assumption $\kappa = 0$ (or $\tau = 0$). Conversely, one may study what value of κ is required for the identification region to admit an ex-ante plausible value or range for δ or β. To keep the exposition concise, we impose R_1 throughout the analysis and understand the results under A_1-A_3 alone as a special case in which $\kappa \to +\infty$ and R_1 is not binding.

The second auxiliary assumption that we study weakens the classical exclusion restriction $\phi = 0$ by specifying whether ϕ and δ have the same sign or the opposite sign (or are zero).

Assumption $\mathbf{R_2}$ *Coefficient Sign Restriction: $\phi\delta \geq 0$ (R_2^+) or $\phi\delta \leq 0$ (R_2^-).*

R_2^+ (R_2^-) implies that U and W affect Y in the same (opposite) direction. For instance, R_2^+ allows for the possibility that both the college selectivity (i.e. the average SAT score) W and the average ability U affect the average earnings Y positively.

3 Identification

We characterize the sharp identification regions for ϕ and δ as well as $(\phi+\delta)$, and consequently $\beta = R_{Y.X} - R_{W.X}(\phi+\delta)$, under A_1-A_3 alone, first under the exclusion restriction and then without imposing it. Then, we additionally impose R_1 or R_2 (when ϕ may be nonzero) or both. First, we digress briefly and consider the assumption that one element X_1 of X is excluded from the Y equation and may serve as an instrument for W. In this case, an IV regression point-identifies the remaining coefficients on X as well as $(\phi+\delta)$.

Proposition 3.1 *Assume A_1 and let $X = (X_1, X_2')'$ and $\beta = (\beta_1, \beta_2')'$. Suppose that the scalar $\beta_1 = 0$, $Cov[(\eta, \varepsilon)', X] = 0$, and $Cov(X, (W, X_2')')$ is nonsingular. Then*

$$(\phi+\delta, \beta_2')' = R_{Y.(W,X_2')'|X}.$$

Without exclusion restrictions on X or W, consider the moments $Var[(Y^*, W^*)']$. Under A_1-A_3, using the expressions for $\sigma^2_{W^*} \neq 0$ and $\sigma_{Y^*.W^*}$ from the proof of Theorem 3.2 below, we have that $R_{Y^*.W^*}$ is a weighted average of ϕ and $(\phi+\delta)$ with weight ρ:

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

If there is no measurement error ($\sigma^2_\varepsilon = 0$) then $\rho = 1$ and $(\phi+\delta)$ and β are point identified[9], $(\phi+\delta, \beta')' = R_{Y.(W,X')'}$. Further, if U^* is degenerate ($\sigma^2_{U^*} = 0$ and U and X are perfectly collinear) then $\rho = 0$ and ϕ is point identified, $\phi = R_{Y^*.W^*}$. In addition, using the expression for $\sigma^2_{Y^*}$ from the proof of Theorem 3.2, we have that

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

where, by definition, we have the inequality

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

[9] Recall the regression representation $R_{Y.(W,X')'} = (R_{Y^*.W^*}, R'_{Y.X} - R'_{W.X}R_{Y^*.W^*})'$.

When U and X are not perfectly collinear, i.e. $\rho \neq 0$, Theorem 3.2 employs equations (4, 5) to express δ, β, and ξ^2 as functions D, B, and C^2 of (ρ, ϕ). This mapping permits characterizing the joint sharp identification region for $(\rho, \phi, \delta, \beta)$ in terms of restrictions on (ρ, ϕ) only. It facilitates studying the consequences of deviating from the benchmark assumptions of no measurement error ($\rho = 1$) or the proxy exclusion restriction ($\phi = 0$).

Theorem 3.2 *Assume* A_1-A_3 *and let* $Var[(X', W)']$ *be nonsingular and* $0 < \rho \leq 1$. *Then*

$$\delta = D(\rho, \phi) \equiv \frac{1}{\rho}(R_{Y^*.W^*} - \phi),$$

$$\beta = B(\rho, \phi) \equiv R_{Y.X} - R_{W.X}\frac{1}{\rho}[R_{Y^*.W^*} - \phi(1-\rho)], \text{ and}$$

$$\xi^2 = C^2(\rho, \phi) \equiv \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - \frac{(1-\rho)}{\rho}(\phi - R_{Y^*.W^*})^2 - R^2_{Y^*.W^*}.$$

3.1 Excluded Proxy: Classical Measurement Error

First, consider the special case in which $\phi = 0$ and the measurement error is classical. Then equation (4) implies that $R_{Y^*.W^*} = \delta\rho$ and we obtain the classical "attenuation bias" whereby $R_{Y^*.W^*}$ understates the magnitude of δ and has its sign. Further, when $\rho\delta \neq 0$, and thus $R_{Y^*.W^*} \neq 0$, equation (5) implies that $\frac{1}{R_{W^*.Y^*}} = \delta + \frac{\xi^2}{\delta\rho}$. It follows that δ can be bounded in the interval with endpoints $R_{Y^*.W^*}$ and $\frac{1}{R_{W^*.Y^*}}$ corresponding to $\rho = 1$ and $\xi^2 = 0$ respectively (see e.g. Gini, 1921; Frisch, 1934. For the case of a vector U, see e.g. Klepper and Leamer, 1984; Bollinger, 2003). Corollary 3.3 collects these results and allows imposing the auxiliary restriction[10] R_1 in addition to the classical assumptions A_1-A_3 with $\phi = 0$.

Corollary 3.3 *Assume the conditions of Theorem 3.2 and suppose that* R_1 *holds. If* $\phi = 0$ *then* $(\rho, \phi, \delta, \beta)$ *are partially identified in the sharp set*

$$\mathcal{S}^{1,c}_{\rho,\phi,\delta,\beta} \equiv \left\{ (r, 0, D(r,0), B(r,0)) : \frac{1}{1+\kappa} \leq r \leq 1 \quad \text{and} \quad 0 \leq C^2(r,0) \right\}.$$

[10] In the case of classical measurement error with multiple latent variables and excluded proxies, Klepper and Leamer (1984) study imposing a common $\frac{1}{1+\tau}$ bound on all the signal to total variance ratios. They report non-sharp bounds on the vector of slope coefficients on the latent variables. See also Bekker, Kapteyn, and Wansbeek (1987) who report tighter bounds.

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

$$\mathcal{S}_{\rho}^{1,c} = \{\lambda\overline{\rho} + (1-\lambda) : 0 \le \lambda \le 1\} \quad \textit{where } \overline{\rho} \equiv \max\{R_{Y^*.W^*}R_{W^*.Y^*}, \frac{1}{1+\kappa}\}$$
$$\mathcal{S}_{\delta}^{1,c} = \{R_{Y^*.W^*}[\lambda + (1-\lambda)\frac{1}{\overline{\rho}}] : 0 \le \lambda \le 1\}, \textit{ and}$$
$$\mathcal{S}_{\beta}^{1,c} = \{R_{Y.X} - R_{W.X}R_{Y^*.W^*}[\lambda + (1-\lambda)\frac{1}{\overline{\rho}}] : 0 \le \lambda \le 1\}.$$

Corollary 3.3 derives the sharp[11] joint identification region for $(\rho, \phi, \delta, \beta)$ as well as the sharp[12] identification regions (projections) for ρ, δ, and β separately. First, consider the cases in which δ or β is point identified. In particular, if $R_{Y^*.W^*} = 0$ then $\delta = 0$ and $\beta = R_{Y.X}$. Also, if $R_{W.X} = 0$ then $\beta = R_{Y.X}$. Further, if $\kappa = 0$ then there is no measurement error and we have that $\rho = 1$, $\delta = R_{Y^*.W^*}$, and $\beta = R_{Y.X} - R_{W.X}R_{Y^*.W^*}$. Otherwise, if $\kappa \to +\infty$ so that R_1 is not imposed then the bounds in Corollary 3.3 reduce to the standard sharp bounds $\mathcal{S}_{\rho}^{c}$, $\mathcal{S}_{\delta}^{c}$, and $\mathcal{S}_{\beta}^{c}$ which set $\overline{\rho} = R_{Y^*.W^*}R_{W^*.Y^*} \ge \frac{1}{1+\kappa} \to 0$. Corollary 3.3 demonstrates how R_1 can lead to tighter sharp bounds.

3.2 Included Proxy: Differential Measurement Error

The bounds in Corollary 3.3 fail when $\phi \neq 0$. Corollary 3.4 employs (in)equalities (4-6) to study the partial identification region for $(\rho, \phi, \delta, \beta)$ when the measurement error may be differential.

Corollary 3.4 *Under the conditions of Theorem 3.2 and R.1, $(\rho, \phi, \delta, \beta)$ are partially identified in the sharp sets*

$$\mathcal{S}_{\rho,\phi,\delta,\beta}^{1} \equiv \left\{(r, f, D(r,f), B(r,f)) : \frac{1}{1+\kappa} \le r \le 1 \;\; \textit{and} \;\; 0 \le C^2(r,f)\right\}.$$

Further, ϕ and δ are not identified, $\mathcal{S}_{\phi}^{1} = \mathcal{S}_{\delta}^{1} = \mathbb{R}$, and ρ and β are partially identified in the sharp sets

$$\mathcal{S}_{\rho}^{1} = [\frac{1}{1+\kappa}, 1] \textit{ and}$$
$$\mathcal{S}_{\beta}^{1} = \{R_{Y.X} - R_{W.X}\{R_{Y^*.W^*} + \lambda[\kappa(\frac{\sigma_{Y^*}^2}{\sigma_{W^*}^2} - R_{Y^*.W^*}^2)]^{\frac{1}{2}}\} : -1 \le \lambda \le 1\}.$$

[11] A joint identification region such as $\mathcal{S}_{\rho,\phi,\delta,\beta}^{c,1}$ is sharp if for every $(r, f, d, b) \in \mathcal{S}_{\rho,\phi,\delta,\beta}^{c,1}$ there exists $(\tilde{U}, \tilde{\eta}, \tilde{\varepsilon})$ that satisfy A.2-A.3 such that $Y = X'b + Wf + \tilde{U}d + \tilde{\eta}$, $W = \tilde{U} + \tilde{\varepsilon}$, and $\frac{\sigma_{\tilde{U}^*}^2}{\sigma_{W^*}^2} = r$.

[12] A projected identification region such as $\mathcal{S}_{\beta}^{c,1}$ is sharp if for every $b \in \mathcal{S}_{\beta}^{c,1}$ there exists $(r, f, d, b) \in \mathcal{S}_{\rho,\phi,\delta,\beta}^{c,1}$.

Under R_1, the joint sharp identification region $\mathcal{S}^1_{\rho,\phi,\delta,\beta}$ is informative about $(\rho, \phi, \delta, \beta)$ since it rules out[13] certain elements (r, f, d, b) of $(0, 1] \times \mathbb{R}^{k+2}$. Corollary 3.4 shows that if $R_{W.X} = 0$ then $R_{Y.X}$ point identifies β. Further, if either $\kappa = 0$ (i.e. $\rho = 1$ and there is no measurement error) or $\frac{1}{\sigma^2_{W^*}} Var(\epsilon_{Y^*.W^*}) = \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - R^2_{Y^*.W^*} = 0$ (i.e. $\epsilon_{Y^*.W^*} = 0$), we have that $\beta = R_{Y.X} - R_{W.X} R_{Y^*.W^*}$. Using the expressions for δ and $0 \leq \xi^2$ from Theorem 3.2, note that $\epsilon_{Y^*.W^*} = 0$ implies that either $\rho = 1$ or $\delta = 0$. When these point identification conditions do not hold and R_1 is not imposed (i.e. when $\kappa \to +\infty$), projecting $\mathcal{S}^1_{\rho,\phi,\delta,\beta}$ onto the support $(0, 1]$ of ρ and the support $\mathbb{R}$ of ϕ, δ, and β_j for $j = 1, ..., k$ yields uninformative bounds (the full support). On the other hand, imposing R_1 yields two-sided sharp bounds $\mathcal{S}^1_\rho$ and $\mathcal{S}^1_\beta$ on ρ and β but ϕ and δ remain unidentified. Thus, R_1 permits studying the sensitivity of β to deviations from the "no measurement error" assumption $\kappa = 0$. Last, note that when $R_{Y^*.W^*} \neq 0$, $\mathcal{S}^1_\beta$ can be conveniently expressed in terms of regression coefficients

$$\mathcal{S}^1_\beta = \{R_{Y.X} - R_{W.X} R_{Y^*.W^*} \{1 + \lambda[\kappa(\frac{1}{R_{Y^*.W^*} R_{W^*.Y^*}} - 1)]^{\frac{1}{2}}\} : -1 \leq \lambda \leq 1\}.$$

Next, we consider the second useful auxiliary assumption R_2 which weakens the standard exclusion restriction $\phi = 0$. We begin by examining R_2^+, $\phi\delta \geq 0$. Recall that $\overline{\rho} \equiv \max\{R_{Y^*.W^*} R_{W^*.Y^*}, \frac{1}{1+\kappa}\}$. Further, let $E(r, f) \equiv f D(r, f) = \frac{1}{r} f (R_{Y^*.W^*} - f)$.

Corollary 3.5 *Under the conditions of Theorem 3.2, R_1, and R_2^+, $(\rho, \phi, \delta, \beta)$ is partially identified in the sharp set*

$$\mathcal{S}^{1,2^+}_{\rho,\phi,\delta,\beta} \equiv \left\{(r, f, D(r, f), B(r, f)) : \frac{1}{1+\kappa} \leq r \leq 1,\ 0 \leq C^2(r, f),\ \text{and } 0 \leq E(r, f)\right\}.$$

Further, let $F_\kappa \equiv \frac{1}{1+\kappa}(1 - \frac{1}{1+\kappa}) - R_{Y^*.W^*} R_{W^*.Y^*}(1 - R_{Y^*.W^*} R_{W^*.Y^*})$. Then ρ, ϕ, δ, and β are partially identified in the sharp sets

$$\mathcal{S}^{1,2^+}_\rho = [\frac{1}{1+\kappa}, 1],$$

$$\mathcal{S}^{1,2^+}_\phi = \{\lambda R_{Y^*.W^*} : 0 \leq \lambda \leq 1\},$$

$$\mathcal{S}^{1,2^+}_\delta = \begin{cases} \{\lambda R_{Y^*.W^*} : 0 \leq \lambda \leq 1\} & \text{if } \kappa = 0 \\ \{\lambda (1+\kappa) R_{Y^*.W^*} [\frac{1}{\kappa}(\frac{1}{\overline{\rho}} - 1)]^{\frac{1}{2}} : 0 \leq \lambda \leq 1\} & \text{if } F_\kappa \leq 0 \text{ and } 0 < \kappa \\ \{\lambda R_{Y^*.W^*} \frac{1}{\overline{\rho}} : 0 \leq \lambda \leq 1\} & \text{if } 0 \leq F_\kappa \end{cases}, \text{ and}$$

$$\mathcal{S}^{1,2^+}_\beta = \{R_{Y.X} - R_{W.X} R_{Y^*.W^*} \{1 + \lambda[\kappa(\frac{1}{\overline{\rho}} - 1)]^{\frac{1}{2}}\} : 0 \leq \lambda \leq 1\}.$$

[13] For example, let $(r, f) = (\frac{1}{2}, (M+1) R_{Y^*.W^*})$. Then one can choose the constant M such that $C^2(\frac{1}{2}, (M+1) R_{Y^*.W^*}) = \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - (M^2 + 1) R^2_{Y^*.W^*} \leq 0$.

Corollary 3.5 shows that, under R_1 and R_2^+, if $R_{Y^*.W^*} = 0$ then ϕ, δ, and β are point identified, $\phi = \delta = 0$ and $\beta = R_{Y.X}$. Further, if $R_{W.X} = 0$ then $\beta = R_{Y.X}$ and if $\kappa = 0$ or $\epsilon_{Y^*.W^*} = 0$ then $\beta = R_{Y.X} - R_{W.X}R_{Y^*.W^*}$. Otherwise, imposing R_2^+ alone (which corresponds to $\kappa \to +\infty$, $\overline{\rho} = R_{Y^*.W^*}R_{W^*.Y^*} \geq \frac{1}{1+\kappa} \to 0$, and $F_\kappa \to -R_{Y^*.W^*}R_{W^*.Y^*}(1 - R_{Y^*.W^*}R_{W^*.Y^*}) \leq 0$) yields the two-sided sharp bounds $\mathcal{S}_\phi^{2^+} = \mathcal{S}_\phi^{1,2^+}$ for ϕ and the one-sided sharp bounds $\mathcal{S}_\delta^{2^+} = \{\lambda R_{Y^*.W^*} : 0 \leq \lambda\}$ for δ and $\mathcal{S}_\beta^{2^+} = \{R_{Y.X} - R_{W.X}R_{Y^*.W^*}\lambda : 1 \leq \lambda\}$ for β. Note that $\mathcal{S}_\phi^{2^+}$ and $\mathcal{S}_\delta^{2^+}$ identify the common sign of ϕ and δ. Also, $\mathcal{S}_\beta^{2^+}$ may rule out that a component β_j of β is zero, in which case the corresponding X_j variable cannot serve as an instrument in Proposition 3.1. Moreover, Corollary 3.5 shows how imposing R_1 and R_2^+ yields bounded identification regions for ρ, ϕ, δ, and β that are tighter than those obtained under either R_1 or R_2^+ alone. Last, observe that if $\overline{\rho} = \frac{1}{1+\kappa} < 1$ then $\mathcal{S}_\beta^{1,2^+} = \mathcal{S}_\beta^{1,c}$ and incorrectly assuming that $\phi = 0$ (instead of $0 \leq \phi\delta$) has no bearing on the sharp bounds for β. However, this incorrectly leads to the region $\mathcal{S}_\delta^{1,c} \subseteq \mathcal{S}_\delta^{1,2^+}$ which need not contain δ.

The last corollary studies the identifying power of R_2^-, $\phi\delta \leq 0$.

Corollary 3.6 *Under the conditions of Theorem 3.2, R_1, and R_2^-, $(\rho, \phi, \delta, \beta)$ is partially identified in the sharp set*

$$\mathcal{S}_{\rho,\phi,\delta,\beta}^{1,2^-} \equiv \left\{(r, f, D(r,f), B(r,f)) : \frac{1}{1+\kappa} \leq r \leq 1,\ 0 \leq C^2(r,f),\ \textit{and}\ E(r,f) \leq 0\right\}.$$

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

$$\mathcal{S}_\rho^{1,2^-} = [\frac{1}{1+\kappa}, 1], \qquad \mathcal{S}_\phi^{1,2^-} = \mathcal{S}_\delta^{1,2^-} = \left\{ \begin{array}{cc} \{\lambda R_{Y^*.W^*} : \lambda \notin (0,1)\} & \textit{if } R_{Y^*.W^*} \neq 0 \\ \mathbb{R} & \textit{if } R_{Y^*.W^*} = 0 \end{array} \right. ,$$

and if $\overline{\rho} = \frac{1}{1+\kappa}$ then $\mathcal{S}_\beta^{1,2^-} = \mathcal{S}_\beta^1$ whereas if $\overline{\rho} = R_{Y^*.W^*}R_{W^*.Y^*}$ then

$$\mathcal{S}_\beta^{1,2^-} = \{R_{Y.X} - R_{W.X}R_{Y^*.W^*}\{\lambda\frac{1}{\rho} + (1-\lambda)[1 - (\kappa(\frac{1}{\rho} - 1))^{\frac{1}{2}}]\} : 0 \leq \lambda \leq 1\}.$$

Under R_1 and R_2^-, if $R_{W.X} = 0$ or $R_{Y^*.W^*} = 0$ then $\beta = R_{Y.X}$ and if $\kappa = 0$ or $\epsilon_{Y^*.W^*} = 0$ then $\beta = R_{Y.X} - R_{W.X}R_{Y^*.W^*}$. Further, the sharp identification regions for ϕ and δ obtained under R_2^- alone (i.e. when $\kappa \to +\infty$) coincide with those obtained under R_1 and R_2^-, $\mathcal{S}_\phi^{2^-} = \mathcal{S}_\delta^{2^-} = \mathcal{S}_\phi^{1,2^-} = \mathcal{S}_\delta^{1,2^-}$. This is a disconnected identification region which rules out that ϕ or δ is in the open interval with end points[14] 0 and $R_{Y^*.W^*}$. Moreover, under R_2^- only, we

[14] Although we do not pursue this here, we note that, unlike in Corollary 3.5, one may tighten the bounds in Corollary 3.6 by assigning a specific sign to ϕ and the opposite sign to δ.

obtain a one-sided sharp bound on β, $\mathcal{S}^{2^-}_{\beta} = \{R_{Y.X} - R_{W.X}\frac{1}{R_{W^*.Y^*}}\lambda : \lambda \leq 1\}$. Last, the sharp identification region $\mathcal{S}^{1,2^-}_{\beta}$ for β, obtained under R_1 and R_2^-, is tighter that the identification region $\mathcal{S}^{1}_{\beta}$, obtained under R_1, only if $\frac{1}{1+\kappa} \leq R_{Y^*.W^*}R_{W^*.Y^*}$.

4 Illustrative 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

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

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. We allow the components of $\eta_X = (\eta_{X_1}, \eta_{X_2})'$ to be correlated. It follows that $(X', Y, W)'$ is normally distributed and we can analytically express the identification regions for ρ, ϕ, δ, and β in Section 3 in terms of 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 $0 < \phi\delta$, R_2^+ holds. Also, we set $\sigma^2_U = 3$, $\sigma^2_\eta = 0.24$, $\sigma^2_\varepsilon = \sigma^2_{\eta_{X_1}} = \sigma^2_{\eta_{X_2}} = 1$, and $\sigma_{\eta_{X_1},\eta_{X_2}} = 0.2$. We obtain that $\rho = 0.685$.

Figure 1 illustrates the joint identification regions $\mathcal{S}^{1,c}_{\rho,\phi,\delta,\beta}$, $\mathcal{S}^{1}_{\rho,\phi,\delta,\beta}$, $\mathcal{S}^{1,2^+}_{\rho,\phi,\delta,\beta}$, and $\mathcal{S}^{1,2^-}_{\rho,\phi,\delta,\beta}$ obtained under this parametrization. Specifically, we illustrate these five-dimensional regions by plotting their projections onto the (ϕ, ρ), (ϕ, δ), and (β_1, β_2) spaces[15]. Each graph in Figure 1 superimposes the 4 projected identification regions that correspond to $\kappa = +\infty, 2, 1, 0.5$ (here the net of X noise to signal ratio $\frac{\sigma^2_\varepsilon}{\sigma^2_{U^*}} = 0.461$). The darker intersections correspond to the smaller κ values and are nested within the lighter regions. Sometimes the identification regions displayed in Figure 1 are unbounded. For example, $\mathcal{S}^{1}_{\beta_1,\beta_2}$ is an unbounded line when $\kappa = +\infty$ whereas $\mathcal{S}^{1,2^+}_{\rho,\phi}$ is a bounded set when $\kappa = 1$. Figure 1 illustrates how the vector of true population coefficients $(\rho, \phi, \delta, \beta)$ (that we label using a cross) is an element of the joint sharp identification regions $\mathcal{S}^{1}_{\rho,\phi,\delta,\beta}$ and $\mathcal{S}^{1,2^+}_{\rho,\phi,\delta,\beta}$. On the other hand, neither $\phi = 0$ nor $\phi\delta \leq 0$ holds and the sharp regions $\mathcal{S}^{1,c}_{\rho,\phi,\delta,\beta}$ and $\mathcal{S}^{1,2^-}_{\rho,\phi,\delta,\beta}$ do not contain $(\rho, \phi, \delta, \beta)$. Also, Figure 1

[15]Recall, from Theorem 3.2, that the joint identification regions $\mathcal{S}^{1,c}_{\rho,\phi,\delta,\beta}$, $\mathcal{S}^{1}_{\rho,\phi,\delta,\beta}$, $\mathcal{S}^{1,2^-}_{\rho,\phi,\delta,\beta}$, and $\mathcal{S}^{1,2^+}_{\rho,\phi,\delta,\beta}$ depend only on the two unknowns (ρ, ϕ). To draw the graphs in Figure 1, we first compute the population (in)equalities (as a function of (ρ, ϕ)) that determine the joint identification regions. We then use a grid search over the (ρ, ϕ) space to approximate these joint regions. Last, we project the grids for these joint regions onto each of the (ϕ, ρ), (ϕ, δ), and (β_1, β_2) two-dimensional spaces.

demonstrates how $\mathcal{S}^{1,c}_{\rho,\phi,\delta,\beta}$ is a subset of both $\mathcal{S}^{1,2^-}_{\rho,\phi,\delta,\beta}$ and $\mathcal{S}^{1,2^+}_{\rho,\phi,\delta,\beta}$, which in turn are subsets of $\mathcal{S}^{1}_{\rho,\phi,\delta,\beta}$. Last, it illustrates how $\mathcal{S}^{1,2^-}_{\rho,\phi,\delta,\beta}$ is disconnected and $\mathcal{S}^{1,2^-}_{\rho,\phi,\delta,\beta} \cup \mathcal{S}^{1,2^+}_{\rho,\phi,\delta,\beta} = \mathcal{S}^{1}_{\rho,\phi,\delta,\beta}$.

Table 1 illustrates the regression estimands and the projected identification regions from Corollaries 3.3, 3.4, 3.5, and 3.6 for $\kappa = +\infty, 2, 1, 0.5$. Thus, the bounds in Table 1 correspond to the regions from Figure 1, projected onto the ρ, ϕ, δ, β_1, and β_2 spaces. Column 1 reports the slope estimands from a regression of Y on $(1, W, X')'$. These estimands would identify (ϕ, β) had δ been zero. Columns 2, 3, 4, and 5 report the sharp bounds for ρ, ϕ, δ, and β under R_1 and the incorrect assumption $\phi = 0$ (column 2), under R_1 only (column 3), under R_1 and R_2^+ (column 4), and under R_1 and the incorrect assumption R_2^- (column 5). Note that, in this example, $\mathcal{S}^{1,c}_{\delta}$ and $\mathcal{S}^{1,c}_{\beta}$ as well as $\mathcal{S}^{1,2^-}_{\delta}$ and $\mathcal{S}^{1,2^-}_{\beta}$ do not contain δ and β. Also, since $\overline{\rho} = R_{Y^*.W^*} R_{W^*.Y^*} = 0.833$ for $\kappa = +\infty, 2, 1, 0.5$, $\mathcal{S}^{1,c}_{\rho,\phi,\delta,\beta}$ remains the same at these κ values. Since $0 \leq F_\kappa$ for $\kappa = 2, 1, 0.5$, $\mathcal{S}^{1,2^+}_{\delta}$ is also identical at these κ values[16]. Last, $\mathcal{S}^{1,2^+}_{\rho,\phi,\delta,\beta}$ improves substantially over $\mathcal{S}^{1}_{\rho,\phi,\delta,\beta}$ and both regions become tighter as κ decreases.

5 Estimation and Inference

Each of the identification sets[17] for ρ, ϕ, δ, and β_j for $j = 1, ..., k$ in Corollaries[18] 3.3, 3.4, 3.5, and 3.6 is of the form $\mathcal{S}_\theta$=$\{\theta(R; \lambda) : \lambda \in \Lambda\}$ where $\theta(\cdot; \lambda)$ is a function of the estimands[19]

$$R \equiv (R'_{Y.(W,X')'}, R'_{W.(Y,X')'}, R'_{Y.X}, R'_{W.X}, \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}})'$$

and λ is a nuisance parameter that is partially identified in a known set Λ. For example, the identification set $\mathcal{S}^{1,2^+}_{\beta}$ from Corollary 3.5 can be written as

$$\mathcal{S}^{1,2^+}_{\beta} = \{\beta^{1,2+}(R; \lambda) : \lambda \in \Lambda\} \equiv \{R_{Y.X} - R_{W.X} R_{Y^*.W^*}\{1 + \lambda[\kappa(\frac{1}{\overline{\rho}} - 1)]^{\frac{1}{2}}\} : \lambda \in [0, 1]\}.$$

[16]Consider slightly changing this parametrization to set $\sigma^2_\eta = 2$ and let $\kappa = +\infty, 2, 1, 0.5$. Then $\overline{\rho} = \frac{1}{1+\kappa}$ when $\kappa = 0.5$, leading $\mathcal{S}^{1,c}_{\rho,\phi,\delta,\beta}$ to become tighter at this κ value. Further, $F_\kappa \leq 0$ at $\kappa = +\infty, 2, 0.5$ and $\mathcal{S}^{1,2^+}_{\delta}$ shrinks as κ decreases. In this case, while $\mathcal{S}^{1,c}_{\rho,\phi,\delta,\beta}$ and $\mathcal{S}^{1,2^-}_{\rho,\phi,\delta,\beta}$ do not contain the true population coefficients vector $(\rho, \phi, \delta, \beta)$, $\mathcal{S}^{1,c}_{\beta}$ and $\mathcal{S}^{1,2^-}_{\beta}$ happen to include $\beta = (1, 0.7)'$ for all these κ values.

[17]Here, we focus on the estimation of the (projected) identification regions for ρ, ϕ, δ, and β_j for $j = 1, ..., k$. To estimate the joint identification regions for $(\rho, \phi, \delta, \beta)$, one can consider using the procedures in e.g. Kline and Tamer (2015) or Shi and Shum (2015).

[18]We omit the standard asymptotic distribution of the IV plug-in estimator for Proposition 3.1.

[19]If one supposes that $R_{Y^*.W^*} \neq 0$ then $\mathcal{S}^{1}_{\beta}$ can expressed in terms of regression coefficients only and it suffices to set $R \equiv (R'_{Y.(W,X')'}, R'_{W.(Y,X')'}, R'_{Y.X}, R'_{W.X})'$.

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

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

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:

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

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

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

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*

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

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

See e.g. White (2001) for standard 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 plug-in estimator $\hat{\Gamma}^* \equiv \hat{Q}^{-1}\hat{\Xi}\hat{Q}'^{-1}$ where $\hat{\Xi}$ is a heteroskedasticity-robust estimator for Ξ (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 $\theta(R;\lambda)$ for an identification region $\mathcal{S}_\theta$ sometimes depends on the signs of $G_\kappa \equiv \frac{1}{1+\kappa} - R_{Y^*.W^*}R_{W^*.Y^*}$ (i.e. the value of $\overline{\rho}$) and $F_\kappa \equiv \frac{1}{1+\kappa}(1-\frac{1}{1+\kappa}) - R_{Y^*.W^*}R_{W^*.Y^*}(1-R_{Y^*.W^*}R_{W^*.Y^*})$. Let $r_{Y^*.W^*} \equiv \frac{\sigma_{Y^*.W^*}}{\sigma_{Y^*}\sigma_{W^*}}$ denote the partial correlation between Y and W given X. Then $0 \leq G_\kappa$ if and only if $r_{Y^*.W^*} \in [-(\frac{1}{(1+\kappa)})^{\frac{1}{2}}, (\frac{1}{(1+\kappa)})^{\frac{1}{2}}]$. Further, $F_\kappa < 0$ if and only if $|r_{Y^*.W^*}|$ belongs to the open line segment with end points $(\frac{1}{(1+\kappa)})^{\frac{1}{2}}$ and $(\frac{\kappa}{1+\kappa})^{\frac{1}{2}}$. Suppose that the signs of G_κ and F_κ are known then one can construct a $1-\alpha$ (e.g.

95%) confidence interval $C_{1-\alpha}(\lambda)$ for $\theta(R;\lambda)$ for each $\lambda \in \Lambda$ using the delta method. Further, a confidence region $CR^{\theta}_{1-\alpha}$ for a partially identified parameter $\theta \in \mathcal{S}_{\theta}$ 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, the signs of G_{κ} and F_{κ}, or alternatively the value of $r_{Y^*.W^*}$, must be estimated and $CR^{\theta}_{1-\alpha}$ must be adjusted to account for this estimation. In particular, the projected identification regions from Section 3 can be rewritten in the form $\mathcal{S}_{\theta}$=$\{\tilde{\theta}(R;\pi) : \pi \in \Pi\}$ where $\pi = (\lambda, \tilde{r}) \in \Lambda \times \{r_{Y^*.W^*}\}$ determines the signs of $G_{\kappa}(\tilde{r})$ and $F_{\kappa}(\tilde{r})$, with $\tilde{\theta}(\cdot;\pi)$ continuously differentiable in R. For example, we have

$$\begin{aligned}
&\mathcal{S}^{1,2^+}_{\beta} = \{\tilde{\beta}^{1,2+}(R;\pi) : \pi \in \Pi\} \\
&\equiv \{R_{Y.X} - R_{W.X} R_{Y^*.W^*}\{1 + \lambda\{\kappa[\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}}\mathbf{1}\{G_{\kappa}(\tilde{r}) < 0\} + (1+\kappa)\mathbf{1}\{0 \leq G_{\kappa}(\tilde{r})\} - 1]\}^{\frac{1}{2}}\} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad : (\lambda, \tilde{r}) \in [0,1] \times \{r_{Y^*.W^*}\}\}.
\end{aligned}$$

Using the delta method, the plug-in estimator $\tilde{\theta}(\hat{R};\pi)$ for an element $\tilde{\theta}(R;\pi)$ of $\mathcal{S}_{\theta}$ is consistent and has a $\sqrt{n}$ asymptotic normal distribution[20]

$$\sqrt{n}(\tilde{\theta}(\hat{R};\pi) - \tilde{\theta}(R;\pi)) \xrightarrow{d} N(0, \nabla_R \tilde{\theta}(R;\pi)\Gamma \nabla_R \tilde{\theta}(R;\pi)').$$

This permits constructing a $1-\alpha_1$ confidence interval $C_{1-\alpha_1}(\pi)$ for $\tilde{\theta}(R;\pi)$ for each $\pi \in \Pi$. To obtain a $1-\alpha_1-\alpha_2$ (e.g. 95%) confidence region $CR^{\theta}_{1-\alpha_1-\alpha_2}$ for $\theta \in \mathcal{S}_{\theta}$, one can construct the confidence intervals $CR^{\tilde{r}}_{1-\alpha_2}$ for $r_{Y^*.W^*}$ and apply Proposition 3 of Chernozhukov, Rigobon, and Stoker (2010) to form the union[21]:

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

[20] The expressions for the gradients $\nabla_R \tilde{\theta}(R;\pi)$ for Corollaries 3.3-3.6 are omitted for brevity and are available from the authors upon request.

[21] We construct the confidence intervals for $\phi \in \mathcal{S}^{1,2^-}_{\phi}$ and $\delta \in \mathcal{S}^{1,2^-}_{\delta}$ similarly, by writing these sets in the form $\mathcal{S}_{\theta}$=$\{\tilde{\theta}(R;\pi) : \pi \in \Pi\}$ with $\pi = (\lambda, \tilde{r}) \in \Lambda \times \{R_{Y^*.W^*}\}$.

To construct $CR^{\tilde{r}}_{1-\alpha_2}$ we use the "Fisher z" variance stabilizing transformation[22] (see e.g. van der Vaart, 2000, p. 30-31). In the empirical analysis in Section 6, $r_{Y^*.W^*}$ is precisely estimated and G_κ and F_κ do not change signs for all elements $\tilde{r} \in CR^{\tilde{r}}_{1-\alpha_2}$ under a range of values for α_2 and κ. In this case, α_2 can be made very small and the adjustment in $CR^{\theta}_{1-\alpha_1-\alpha_2}$ is negligible. In particular, we set $\alpha_1 = 0.4999$ and $\alpha_2 = 0.0001$.

6 The Returns to College Selectivity and Characteristics

Several papers study the returns to college selectivity and characteristics. On the one hand, some studies provide evidence for a positive return to college quality. For example, Brewer, Eide, and Ehrenberg (1999) model how students choose a college and find a significant wage premium (e.g. six years after high school graduation) to attending an elite private college. Also, using propensity score matching, Black and Smith (2004) find that college quality has a positive, albeit sometimes imprecisely estimated, effect on wage. Further, when comparing the earnings of individuals around the admission cutoff point, Hoekstra (2009) finds that attending a flagship state university leads to 20% 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 after being admitted to equally selective colleges earn comparably. Dale and Krueger (2002, 2014) report similar findings using a "self-revelation" model in which a student's college application behavior (e.g. the average SAT score of the colleges to which the student applied) reveals his or her unobserved 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.

We contribute to this literature by applying this paper's framework to analyze the recently

[22] Let $\hat{r}_{Y^*.W^*} \equiv \frac{\sum_{i=1}^n \hat{\epsilon}_{Y.X,i}\hat{\epsilon}_{W.X,i}}{(\sum_{i=1}^n \hat{\epsilon}^2_{Y.X,i})^{\frac{1}{2}}(\sum_{i=1}^n \hat{\epsilon}^2_{W.X,i})^{\frac{1}{2}}}$ and $Z(\tilde{r}) \equiv \frac{1}{2}\log(\frac{1+\tilde{r}}{1-\tilde{r}})$. Under normality of (Y^*, W^*), we have $\sqrt{n}(Z(\hat{r}_{Y^*.W^*}) - Z(r_{Y^*.W^*})) \xrightarrow{d} N(0,1)$. To construct $CR^{\tilde{r}}_{1-\alpha_2}$, we use this approximation and apply the inverse transformation $\tilde{r} = \frac{e^{2z}-1}{e^{2z}+1}$ to the end points of the $1-\alpha_2$ confidence interval for $Z(r_{Y^*.W^*})$. An alternative uses the delta method to derive the asymptotic distribution of $\sqrt{n}\hat{R}_{Y^*.W^*}\hat{R}_{Y^*.W^*}$ and construct a confidence interval for $r^2_{Y^*.W^*} = R_{Y^*.W^*}R_{Y^*.W^*}$, in order to infer the sign of G_κ and F_κ.

released College Scorecard (CS) data. We specify a parsimonious model for the logarithm of the average earnings of a cohort as a function of the college selectivity (i.e. the cohort's average SAT score), a rich set of college characteristics, and the cohort's average unobserved ability. We allow the average unobserved ability to freely statistically depend on the college characteristics. Further, we let the average SAT score serve a second role as an error-laden proxy for the average unobserved ability. Thus, this model decomposes the average SAT score into the average ability, as the signal component, and a classical measurement error. In what follows, we report tight bounds on the returns to the college selectivity and characteristics as well as the average ability. Moreover, we study the sensitivity of these bounds to two types of restrictions. The first restricts the extent of the measurement error in how the average SAT score proxies the average ability. The second restricts the effects of the college selectivity and the average ability on the average earnings to have the same sign.

6.1 College Scorecard Data Description

CS reports data on several dimensions of the quality of higher education in the US. The data on these variables 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. In particular, CS reports data on the institution characteristics, student demographic and socioeconomic characteristics, admission and academic attributes, affordability, as well as earnings outcomes.

While CS is detailed and nationally comprehensive, it has some limitations that arise due to data unavailability or 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. Nevertheless, 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. Although "about two-thirds of institutions, collectively enrolling 83 percent of students, have only one main campus identifier" (CEA, p. 29), complex institutions with multiple branches may differ in how they aggregate and report data[23]. 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 non-enrolled Title IV students who are working 6 years after estimated college entry and this average 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 first 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 whether the aggregate features of these variables remains stable in the short run[24]. Here, we do not directly address these data imperfections and focus instead on the potential measurement error in how the average SAT score[25] measures the average scholastic ability. We refer the reader to the CS data documentation webpage[26] and to the CEA report for a detailed account of the CS data.

6.2 Definition of Variables and Sample Selection

We focus on the most recent cohort of students on which data is available. This is the cohort of students who enrolled in an academic institution in the fall[27] of 2007, graduated with a bachelor's degree in the spring of 2011, and were non-enrolled and working[28] in[29] 2013.

[23]For example, the data on the mean earnings are aggregated at the level of each of the 13 branches of the University of Wisconsin whereas they are aggregated across the 23 branches of the Pennsylvania State University (see CEA, p. 29).

[24]For example, Black and Smith (2004, p. 105) state that variables related to college quality "change only very slowly, so utilizing values from a single point in time adds little measurement error."

[25]We use data on SAT_AVG, drawn from IPEDS, for the average SAT score. The CS data dictionary defines SAT_AVG as the "average SAT equivalent score of students admitted" but does not provide a formula for how this is calculated. CEA, p. 42 states that "data from IPEDS are used to form average SAT score equivalents using reported ranges for ACT and SAT scores." We believe that this refers to the IPEDS data on the first and third quartiles of the SAT and ACT scores of the enrolled students.

[26]https://collegescorecard.ed.gov/data/documentation/

[27]CS covers the time period from 1996 to 2015. The CS 2012-13 data file contains the most recent data on earnings for the pooled cohorts that enrolled in Fall 2006 or Fall 2007. We use this earnings data together with data on other variables that pertain to the 2007 cohort. Also, we obtain generally similar results when we use the 2006-2007 pooled earnings data with the variables that pertain to the 2006 cohort instead.

[28]We set the average earnings as the main outcome variable of interest. We also examine setting, as the outcome variable, the share of individuals, including those with 0 earnings, who are non-enrolled and earning more than $\$25,000$ per year, six years after enrolling in an institution. Although less sizeable and informative for some coefficients, these bounds share similar features and directionality with the mean earnings results. For brevity, we forgo analyzing these results in detail.

[29]We study the short run financial returns to college quality. This focuses on non-enrolled individuals who are working 6 years after enrollment. It is also of interest to distinguish these short run effects from the

CS reports yearly data files which contain institution-level aggregate data that need not correspond to a specific student cohort. We proceed by drawing, from several CS data files, the data that we think is the most representative of the 2007 student cohort[30]. To ease the exposition, we sometimes omit referencing the details of the CS data construction. Instead, we refer the reader to Table 9 which 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 9 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. This yields a sample of 1710 institutions. After dropping observations with missing data[31], we obtain a sample of 1165 institutions. Table 2 reports summary statistics for the key variables that we employ in our analysis. For example, the average SAT score 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 5 most selective public institutions in our sample are College of William and Mary, Georgia Institute of Technology, SUNY College at Geneseo, University of Virginia, and University of Michigan-Ann Arbor. The standard deviation of SATAvg is 119.93, which corresponds roughly to the difference between Stanford and the University of Virginia. Although our analysis incorporates in X all the institutional characteristics that fall below the dividing line in Table 2, for brevity, we do not report estimates of the identification regions for the coefficients on these variable (the confidence regions for these coefficients often contain zero).

---

long run effects that may be channeled via attending graduate school or accumulating work experience. For instance, for the cohort of students who enrolled in the fall of 2002 and graduated with a bachelor's degree in the spring of 2006, we obtain generally similar results when setting the outcome to be the mean earnings of those who were non-enrolled and working in either 2008 (6 years after enrollment) or in 2013 (10 years after enrollment). We leave a detailed study of the dynamic aspects of the returns to college selectivity and characteristics to other work.

[30] We merge data on variables that are relevant to the 2007 cohort from 6 CS data files for the years 2007-08, 2010-11, 2011-12, 2012-13, 2013-14, and 2014-15.

[31] We drop 27 institutions that were missing from at least one of the CS data files that we consider. Among the remaining 1683 institutions that appear in all CS data files, 378 institutions are missing data on SATAvg. Among the remaining 1305 institutions, 21 are missing MnEarnWnEP6/Gt25KP6, 28 are missing Female, 45 are missing Dependent, 50 are missing ParEdPctPS, 1 is missing CostT4, 1 is missing NPT4, 31 are missing GDebtMdn, and 3 are missing RelAffilInd. Last, we drop 3 institutions with negative NPT4.

6.3 Estimates for the Returns to College Selectivity and Characteristics

Using the sample of 1165 institutions, we report regression estimates as well as bounds that account for classical or differential measurement error in how the average SAT score proxies the average ability. Throughout, Y denotes the logarithm of the average earnings 6 years after enrollment and W denotes the average SAT score (i.e. the college selectivity) which serves as an “included” proxy for the average unobserved ability U. We consider three nested settings for X, described next. These include important college characteristics that have been discussed in the literature.

In the first specification, X consists of characteristics that pertain to the (1) institution, (2) student body, and (3) affordability. In particular, X includes the following institution characteristics: 8 region indicators and 11 locale indicators[32] for the institution's location, indicators for whether the institution is minority-serving, a women-only college[33], has a religious affiliation, or awards a graduate degree, and an indicator for the institution's control (public versus private nonprofit). Further, X includes the following student body characteristics: the logarithm of the student population, 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, and we omit from X the share of students who are white as the reference group), the shares of students how are female, dependent, and with at least one post-secondary educated parent, as well as the logarithm of the average family income. Last, X includes these affordability characteristics: the logarithms of the average cost of attendance and average net price, the shares of students with a federal student loan and with a Pell grant, and the logarithm of the median student debt.

The first specification does not account for the composition of majors at each institution. However, the choice of major plays an important role in understanding the labor market outcomes (see e.g. Altonji, Arcidiacono, and Maurel, 2016; Kirkeboen, Leuven, and Mogstad, 2016). As such, the apparent effects in the first specification, may partly reflect that institutions with a particular selectivity and characteristics profile may be more specialized in majors that yield high (or low) labor market returns. To account for this possibility, the

[32] The sample does not contain US service schools (Region0) and we leave out, as reference groups, the indicator for the Plains region (Region4) and the the indicator for the rural remote locale (Locale43).

[33] The sample does not contain men-only colleges.

second specification augments X to include, in addition to the covariates in the first specification, the percentage of degrees awarded in each field of study in our sample[34] according to the Classification of Instructional Programs (CIP).

Last, we also examine some of the mechanisms that may explain the return to college selectivity. In particular, in addition to the variables from the second specification, the third specification includes in X the instructional expenditures per student and the completion rate within 150% of expected graduation time.

In the empirical analysis, we sometime assume R_1 ($\sigma_\varepsilon^2 \leq \kappa\sigma_{U^*}^2$) to impose an upper bound on the net of X noise to signal ratio. In this case, we set $\kappa = 1$ as our default restriction. This assumes that at most (at least) half of the variance in W^*, i.e. the average SAT score net of the college characteristics X, is due to measurement error ε (U^*, i.e. the average ability net of X). Specifically, in the above three specifications, $\mathfrak{R}^2_{W.X}$ is estimated to be 0.8183, 0.8612, and 0.8893 respectively. Setting $\kappa = 1$ assumes that $\mathfrak{R}^2_{W.U}$ is at least half as close to 1 than $\mathfrak{R}^2_{W.X}$ is. We also obtain qualitatively similar results when we use the estimate for $\mathfrak{R}^2_{W.X}$ to set κ such that[35] $0.9 = \frac{1}{1+\tau} \leq \mathfrak{R}^2_{W.U}$. More generally, Section 6.4 studies the consequences of varying the value of κ on the bounds estimates. In particular, we let κ range from 0 to 30, allowing the variance of U^* to account for either all ($\kappa = 0$) or almost none ($\kappa = 30$) of the variance of W^*.

6.3.1 Estimates given Institution, Student, and Cost Characteristics

Table 3 presents the results under the first specification. This accounts for the institution, student, and affordability characteristics and serves as a basis of comparison with the results in Section 6.3.2 which further account for the composition of majors. Column 1 reports the regression estimates $\hat{R}_{Y.(W,X')'}$ along with 95% confidence intervals in parentheses. Under the first specification, $\hat{R}_{Y.(W,X')'}$ consistently estimates the returns to college selectivity and characteristics ϕ and β if the average ability does not directly affect the average earnings, i.e. $\delta = 0$. Column 2 reports the bounds under R_1 in the classical measurement error case. This allows the average ability U, but not selectivity W, to directly affect the average earnings

[34] There is a total of 38 CIP fields of study. Among these, our sample includes 37 fields listed in Table 6 (PCIP29, the percentage of degrees awarded in Military Technologies and Applied Sciences, is always zero). Further, we choose PCIP45 (Social Sciences) as the reference field of study and omit it from X.

[35] This sets κ equal to 1.22, 2.58, and 9.35 in the first, second, and third specifications respectively.

(i.e. δ may be nonzero whereas $\phi = 0$). Columns 3 and 4 allow for differential measurement error by removing the exclusion restriction $\phi = 0$ so that both the college selectivity and the average ability can affect the mean earnings. Column 3 reports the results under R_1 whereas column 4 reports the results under R_1 and R_2^+, when the college selectivity and the average ability can affect the average earnings in the same direction, $\phi\delta \geq 0$. For brevity, we omit the results under R_2^+ (recall that $\mathcal{S}_{\phi}^{2^+} = \mathcal{S}_{\phi}^{1,2^+}$ and that $\mathcal{S}_{\delta}^{2^+}$ and $\mathcal{S}_{\beta_j}^{2^+}$ are half open intervals which identify the sign of δ and possibly that of β_j). Further, we do not report the results under R_1 and R_2^- ($\phi\delta \leq 0$) since we deem it plausible that selectivity and average ability affect mean earnings positively, $\phi \geq 0$ and $\delta \geq 0$.

As shown[36] in Table 3, under the first specification, if $\delta = 0$ then the regression coefficient $\hat{R}_{Y.(W,X')'}$ in column 1 estimates that a 100 point increase in SATAvg leads to a 5.1% approximate increase in the average earnings 6 years after enrollment, with a 95% confidence region $CR_{0.95}$ $(3.6\%, 6.6\%)$. Under R_1 and $\phi = 0$, selectivity is assumed to not affect the average earnings and the estimated bounds in column 2 on the return to a 100 point increase in the average ability U are $[5.1\%, 10.2\%]$ with $CR_{0.95}$ $(3.6\%, 13.15\%)$. When selectivity may affect the mean earnings, the returns to selectivity and the average ability are not identified under R_1 only, as shown in column 3. However, together R_1 and R_2^+ yield informative upper bounds on ϕ and δ. In particular, the estimated identification regions in column 4 for the ceteris paribus return to a 100 point increase in SATAvg or U are $[0, 5.1\%]$, with $CR_{0.95}$ $(0, 6.6\%)$, and $[0, 10.2\%]$, with $CR_{0.95}$ $(0, 13.2\%)$, respectively.

Next, consider the returns to the college characteristics X. First, note that $\hat{r}_{Y^*.W^*} = 0.2070$ with 99.99% confidence region (CR) $(0.0958, 0.3131)$. Thus, G_κ is precisely estimated to be positive. As discussed in Section 3 and shown in Table 3, in this case, replacing the exclusion restriction $\phi = 0$ with R_2^+ has no bearing on the identification region for β under R_1, i.e. $\mathcal{S}_{\beta}^{1,2^+} = \mathcal{S}_{\beta}^{1,c}$. Also, note that for certain coefficients, $\widehat{\mathcal{S}}_{\beta}^{1,2^+}$ is substantially tighter than $\widehat{\mathcal{S}}_{\beta}^{1}$. In general, the $\widehat{\mathcal{S}}_{\beta}^{1,2^+}$ bounds are tight around the regression estimate.

The bounds estimates in Table 3 for the returns to the institution characteristics suggest that the ceteris paribus difference in mean earnings between public and private institutions

[36] The entries in Tables 3, 4, and 5 display the ceteris paribus (approximate) percentage change in the average earnings due to: a 100 point increase in SATAvg or the average ability U, being a private institution (ControlInd), offering a graduate degree (HDeg), a percentage increase in any of the variables that appear in logarithmic form, or a percentage point increase in any of the variables that are reported in percentile form.

is statistically insignificant. Also, Table 3 reports a positive and statistically significant percentage difference in the mean earnings of undergraduate students who attended an institution that offers a graduate degree as opposed to one that does not. Table 3's bounds on the returns to the student body characteristics suggest that an increase in the shares of minority students (Black, Hispanic, or Asian) or a percentage increase in the average family income leads to a significant percentage increase in the average earnings and that an increase in the share of students who are female, dependent or have a postsecondary educated parent leads to a significant percentage decrease in the average earnings. For example, under this specification, the regression coefficient $\hat{R}_{Y.(W,X')'}$ in column 1 estimates that a percentage point increase in the share of females leads to a 0.4% decrease in the average earnings, with $CR_{0.95}$ $(-0.54\%, -0.27\%)$. Accounting for the effects of the unobserved ability under R_1 and R_2^+ in column 4 barely alters this gender earnings gap and yields the bounds $[-0.4\%, -0.39\%]$ for the coefficient on Female, with $CR_{0.95}$ $(-0.54\%, -0.26\%)$. Also, whereas a regression estimates that a percentage increase in the enrollment size has a statistically significant positive effect on the mean earnings, one cannot rule out under R_1 and R_2^+ that this effect is zero. Nor can one rule out under R_1 and R_2^+ that the effect of an increase in the share of non-resident aliens is zero. Last, the estimates for the returns to the affordability characteristics in Table 3 show that an increase in the shares of students with a federal student loan leads to a significant percentage increase in the average earnings and that a percentage point increase in the share of students who have a Pell grant or a percentage increase in the median debt leads to a significant percentage decrease in the average earnings. Further, we cannot rule out under R_1 and R_2^+ that the effect of a percentage increase in the average cost or the net price is zero, whereas a regression estimates the latter effect to be statistically significantly negative.

6.3.2 Estimates given the Composition of Majors

The second specification accounts for the composition of majors. This yields upper bounds for the returns to college selectivity and average ability that are slightly smaller than the bounds from the first specification. Further, the extent of the returns to the institution, student body, and affordability characteristics under R_1 and R_2^+ is often more modest than the first specification estimates suggest. For example, under R_1 and R_2^+, the bounds on the return to a percentage point increase in the share of females[37] become smaller in magnitude: $[-0.23\%, -0.20\%]$ with $CR_{0.95}$ $(-0.34\%, -0.1\%)$. Also, after conditioning on the major composition, we can no longer rule out at comfortable significance levels under R_1 and R_2^+ that the coefficient associated with offering a graduate degree, the share of students with a postsecondary-educated parent, or the share of students with a federal student loan is zero. Note that, unlike in the first specification, conditioning on the major composition renders the regression coefficient on the logarithm of the average cost statistically significantly positive. Nevertheless, we cannot rule out that this effect is zero under R_1 and R_2^+.

Last, this specification allows us to study the consequences of changing the major composition. To illustrate this, Table 4 includes the regression and bounds estimates for the coefficients associated with PICP13 (Education) and PCIP14 (Engineering). For instance, under R_1 and R_2^+, the estimated bounds on the return to shifting a percentage point of students away from the excluded field of study (PCIP45, Social Sciences) and toward Education are $[-0.23\%, -0.16\%]$ with $CR_{0.95}$ $(-0.38\%, 0.01\%)$ whereas the corresponding bounds for Engineering are $[0.27\%, 0.28\%]$ with $CR_{0.95}$ $(0.09\%, 0.47\%)$. Table 7 reports the regression estimates and bounds under R_1 and R_2^+ for all the CIP fields of study listed in Table 6.

6.3.3 Estimates given the Instructional Expenditures and the Completion Rate

The last specification studies whether attending a highly selective college leads to large average earnings in part due to large instructional expenditures per student and a high completion rate. In particular, after including these two variables in X, the regression estimate $\hat{R}_{Y.(W,X')'}$ in column 1 of Table 5 implies that if $\delta = 0$ then a 100 point increase in SATAvg leads to a 2.8% approximate increase in average earnings, with $CR_{0.95}$ $(1.3\%, 4.2\%)$. Under R_1 and classical measurement error with $\phi = 0$, the estimates for the bounds on the return to a 100 point increase in the average ability U become $[2.8\%, 5.5\%]$ with $CR_{0.95}$ $(1.3\%, 8.4\%)$. Under R_1 and R_2^+, with $\phi\delta \geq 0$, the bounds estimates in column 4 for the ceteris paribus return to a 100 point increase in SATAvg or U are $[0, 2.8\%]$, with $CR_{0.95}$ $(0, 4.2\%)$, and $[0, 5.5\%]$, with $CR_{0.95}$ $(0, 8.4\%)$, respectively. Thus, in comparison to the estimates in column 4 of Tables 3 and 4, accounting for the expenditures per student and the completion rate leads to a substantially smaller upper bound estimates for the returns

[37] 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.

to the college selectivity and the average ability[38]. In contrast, the bounds estimates for the return to the college characteristics X in Table 5 and the CIP fields of study in Table 8 are tight around $\hat{R}_{Y.(W,X')'}$ and generally similar to the estimates in Tables 4 and 7 respectively. Last, the bounds estimates under R_1 and R_2^+ for the return to a percent increase in the expenditures per student are $[0.07\%, 0.08\%]$, with $CR_{0.95}$ $(0.05\%, 0.1\%)$, and the bounds for the return to a percentage point increase in the completion rate are $[0.11\%, 0.19\%]$, with $CR_{0.95}$ $(0.01\%, 0.27\%)$.

6.4 Sensitivity to κ and Discussion

The bounds estimates for the returns to college selectivity and characteristics in Tables 3, 4, 5, 7, and 8 impose R_1 and employ the default setting $\kappa = 1$. Nevertheless, this paper's framework does not require a particular choice for κ and a researcher may conduct a sensitivity analysis that examines how the estimates change as κ varies. Figure 2 illustrates this by plotting the estimated bounds $\widehat{\mathcal{S}}_{\phi}^{1,2^+}$, $\widehat{\mathcal{S}}_{\delta}^{1,2^+}$, $\widehat{\mathcal{S}}_{\beta}^{1,2^+}$ (using the darker shade) and the 95% confidence regions $CR_{0.95}$ (using the lighter shade) for ϕ, δ, and β under R_1 and R_2^+ when κ ranges from[39] 0 to 30. It also plots $\widehat{\mathcal{S}}_{\beta}^{1}$ and $CR_{0.95}$ obtained under R_1 only. To ease the presentation for β, we focus on the coefficient associated with Female. Recall that setting $\kappa = 0$ assumes that there is no measurement error ($\sigma_{\varepsilon}^2 = 0$) whereas setting $\kappa = 30$ allows the variance of the measurement error to be 30 times as large as the variance of U net of X, $\sigma_{\varepsilon}^2 \leq 30\sigma_{U^*}^2$. Thus, setting $\kappa = 30$ permits X to explain the variance of U considerably better than the variance of W, so that $\sigma_{W^*}^2 = \sigma_{U^*}^2 + \sigma_{\varepsilon}^2$ may substantially exceed $\sigma_{U^*}^2$. For a given κ, Figure 2 shows how the bounds estimates and confidence regions for ϕ, δ, and β change across the three specifications that we consider. Further, it illustrates how, unlike ϕ, the bounds estimates and confidence regions for δ and β vary with κ. In particular, the kinks in the bounds estimates and $CR_{0.95}$ correspond to κ values when the F_{κ} or G_{κ} estimate changes sign and the significance of the sign of F_{κ} or G_{κ} at the 99.99% level changes, respectively.

[38]When augmenting X with $\log(InExpFTE)$ only, the bounds on the ceteris paribus return to a 100 point increase in SATAvg or U under R_1 and R_2^+ are $[0, 4\%]$, with $CR_{0.95}$ $(0, 5.3\%)$, and $[0, 8\%]$, with $CR_{0.95}$ $(0, 10.6\%)$, respectively. Further, augmenting X with $C150\_4$ only yields the bounds $[0, 3.3\%]$, with $CR_{0.95}$ $(0, 4.7\%)$, and $[0, 6.6\%]$, with $CR_{0.95}$ $(0, 9.4\%)$, respectively.

[39]Since $\mathfrak{R}_{W.X}^2$ is estimated to be 0.8183, 0.8612, and 0.8893 in the above three nested specifications, letting κ range from 0 to 30 corresponds to letting τ range from 0 to $\bar{\tau} = 0.2134$, 0.1552, and 0.12 respectively, so that $\mathfrak{R}_{W.X}^2 \leq \frac{1}{1+\bar{\tau}} \leq \frac{1}{1+\tau} \leq \mathfrak{R}_{W.U}^2$ and $\frac{1}{1+\bar{\tau}}$ is 0.8241, 0.8657, and 0.8929.

Note that $\widehat{\mathcal{S}}_{\beta}^{1}$ and especially $\widehat{\mathcal{S}}_{\beta}^{1,2^{+}}$ remain relatively tight around the regression estimate as the value of κ varies over a sizeable range. For example, under R_1 and R_2^{+}, the smallest κ value for which the $CR_{0.95}$ for the Female coefficient contains 0 is 9.2, 29.6, and 15.2 under the first, second, and third specification respectively. This suggests as possibilities that either the measurement error in how the average SAT score proxies the average ability is modest (i.e. ε in equation (1) is is close to being degenerate and $R_{Y.(W,X')'}$ approaches $(\phi+\delta,\beta')'$) or the average ability U has a small effect on the mean earnings (i.e. δ in equation (S) is close to 0 and $R_{Y.(W,X')'}$ approaches $(\phi,\beta')'$).

To conclude Section 6, we note that this empirical analysis may inherit some of limitations of the CS data that arise due to data unavailability or aggregation. Further, the analysis imposes several assumptions which may fail, including linearity, homogeneity of the slope coefficients, or that endogeneity arises due to one variable U. For example, this paper defines college selectivity as the average SAT score of a student cohort, as is often assumed in the literature, and a more general model would allow the average SAT score to serve as a proxy for both the college selectivity[40] and the average ability. However, this specification falls outside of the scope of this paper's econometrics framework which is concerned with a scalar latent variable. As such, the paper's estimates should be carefully interpreted if one suspects that the imposed assumptions do not hold or that there are unobserved confounders other than U. Nevertheless, this analysis contributes to the literature in two ways. First, it analyzes the rich CS data which comprehensively represents the post-secondary institutions in the US. Second, it does not impose several commonly employed assumptions. In particular, the analysis allows college selectivity W to serve as a error-laden proxy for the average ability U and to directly affect the average earnings. This dispenses with the exclusion restriction $\phi = 0$ that is commonly imposed in the measurement error literature. Also, the analysis allows for selection on unobserved ability and requires neither the availability of exogenous instruments[41] nor of conditioning covariates that ensure the exogeneity of $(W, X')'$.

[40] Black and Smith (2006) study a setting in which multiple proxies for the unobserved college quality are available but there is no selection on unobserved ability.

[41] For instance, in the second and third (but not the first) specifications, the bounds on the coefficient on ParEdPctPS (the share of students with at least one post-secondary educated parent) contain zero. Suppose that one assumes that ParEdPctPS is excluded from the average earnings equation in these specifications. Then one can apply Proposition 3.1 by using $X_1 = ParEdPctPS$ as an instrument. Specifically, the IV estimate for $\phi+\delta$ is 3.87% with 95% CI $(1.1\%, 6.63\%)$ in the second specification and is small and insignificant in the third specification, -0.67% with 95% CI $(-4.41\%, 3.07\%)$. In both specifications, the IV estimates

7 Conclusion

This paper studies the identification of 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. It extends the measurement error literature by removing the standard exclusion restriction that assumes that the coefficient on the proxy W in the outcome equation is set to zero. This accommodates a leading setting for differential measurement error that occurs when a variable W, which serves as a proxy for the latent variable U, may directly affect the outcome. First, the paper demonstrates the crucial role that the proxy exclusion restriction plays in ensuring the validity of the standard classical measurement error bounds in the literature since removing it renders the coefficients on the proxy W, the latent variable U, and the covariates X not separately identified. We then characterize the identification regions for these coefficients under two auxiliary restrictions. The first restriction imposes an upper bound κ on the net of X "noise to signal" ratio, i.e. the ratio of the variance of the measurement error ε to the variance of the latent variable U given X. This permits conducting a sensitivity analysis that examines the consequences of deviating from the "no measurement error" assumption $\kappa = 0$. The second auxiliary restriction specifies whether the effects ϕ and δ of the proxy W and the latent variable U on the outcome Y are of the same or the opposite direction ($\phi\delta \leq 0$ or $\phi\delta \geq 0$). This relaxes the exclusion restriction $\phi = 0$.

After discussing estimation and inference, we employ the paper's framework to study the returns to college selectivity and characteristics. We analyze the recently released College Scorecard (CS) data which reports comprehensive data, aggregated at the institution level, on postsecondary institutions in the US. Following the literature, we define college selectivity as the average SAT equivalent score for the enrolled student cohort. We then use a parsimonious model for the logarithm of the average earnings of the cohort as a function of the college selectivity, the characteristics of the college, and the average unobserved ability of the cohort. For the college characteristics, we consider characteristics of the institution, student body, and affordability as well as the major composition, the instructional expenditures, and the completion rate. We allow the average scholastic ability to statistically depend on these

for the coefficients on the remaining covariates are generally similar to the regression and bounds estimates.

college characteristics and let the average SAT score serve as an error-laden proxy for the average unobserved ability. Thus, selectivity (i.e. the average SAT score) serves as an "included" proxy for the average ability and may directly impact the average earnings.

Under our auxiliary restrictions, we report an informative upper bound on the returns to college selectivity and average ability 6 years after enrollment. In particular, given the college characteristics and the major composition, a 100 points increase in the average SAT score leads to at most a 4.8% increase in the average earnings, with $CR_{0.95}$ $(0, 6.1\%)$. We also report tight bounds on the returns to the college characteristics and contrast these with the regression estimates. In particular, for the institution characteristics, we cannot rule out that undergraduates who attend a public as opposed to a private institution, or an institution that offers a graduate degree as opposed to one that does not, have similar mean earnings ceteris paribus. For the student body characteristics, our estimates suggest that an increase in the share of minority students or in the average family income has a positive effect on the average earnings and that an increase in the share of female or dependent students has a negative effect on the average earnings. However, the evidence on the directions of the effects of the enrollment size, the share of non-resident aliens, or the share of students with a post-secondary educated parent is inconclusive. For the affordability characteristics, we find that an increase in the share of students with Pell grants or in the median student debt has a negative effect on the average earnings and we cannot rule out that the effect of an increase in the average cost, the net price, or the share of students with a federal student loan is zero. We also obtain tight bounds on the effects of a change in the major composition. For example, a shift away from the social sciences and toward engineering (English Language and Literature/Letters) increases (decreases) mean earnings. Further, we demonstrate how conditioning on the major composition reduces the magnitude of the bounds on the effects of certain college characteristics, such as the gender composition or offering a graduate degree. Last, we report bounds on the effects of the instructional expenditures per student and the completion rate on mean earnings and show how conditioning on these variables narrows the upper bounds on the returns to the college selectivity and the average ability.

A Mathematical Proofs

Proof of Proposition 3.1 Since $Cov[(\eta,\varepsilon)',X]=0$ and $Cov(X,(W,X_2')')$ is nonsingular, the result obtain from $Y=X_2'\beta_2+W(\phi+\delta)-\varepsilon\delta+\eta$.

Proof of Theorem 3.2: Since $Cov[(\eta,\varepsilon)',X]=0$ and $Var(X)$ is nonsingular, we obtain

$$\beta = R_{Y.X} - R_{W.X}(\phi+\delta).$$

A_2 and A_3 give $\sigma_{U^*,\varepsilon}=\sigma_{U^*,\eta}=\sigma_{\varepsilon,\eta}=0$. Using $R_{W.X}=R_{U.X}$ and $Y^*=U^*(\phi+\delta)+\varepsilon\phi+\eta$, we have

$$\begin{aligned}\sigma^2_{W^*} &= \sigma^2_{U^*}+\sigma^2_\varepsilon,\\ \sigma^2_{Y^*} &= (\phi+\delta)^2\sigma^2_{U^*}+\phi^2\sigma^2_\varepsilon+\sigma^2_\eta, \text{ and}\\ \sigma_{W^*,Y^*} &= (\phi+\delta)\sigma^2_{U^*}+\phi\sigma^2_\varepsilon.\end{aligned}$$

Since $Var[(X',W)']$ is nonsingular, we have that $\sigma^2_{W^*}\neq 0$ and we obtain

$$\begin{aligned}R_{Y^*.W^*} &= \frac{\sigma_{W^*,Y^*}}{\sigma^2_{W^*}} = \phi(1-\rho)+(\phi+\delta)\rho = \phi+\delta\rho, \text{ and}\\ \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} &= \phi^2(1-\rho)+(\phi+\delta)^2\rho+\xi^2.\end{aligned}$$

Since $\rho\neq 0$, we obtain $\delta = D(\rho,\phi)\equiv \frac{1}{\rho}(R_{Y^*.W^*}-\phi)$,

$$\beta = B(\rho,\phi)\equiv R_{Y.X}-R_{W.X}(\phi+D(\rho,\phi)) = R_{Y.X}-R_{W.X}\frac{1}{\rho}[R_{Y^*.W^*}-\phi(1-\rho)],$$

and

$$\begin{aligned}\xi^2 = C^2(\rho,\phi) &\equiv \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}}-\phi^2(1-\rho)-(\phi+D(\rho,\phi))^2\rho\\ &= \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}}-\phi^2(1-\rho)-\frac{1}{\rho}(R_{Y^*.W^*}-\phi(1-\rho))^2\\ &= \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}}-\frac{(1-\rho)}{\rho}(\phi-R_{Y^*.W^*})^2-R^2_{Y^*.W^*}.\end{aligned}$$

Proof of Corollary 3.3: Setting $\phi=0$ in $\mathcal{S}^1_{\rho,\phi,\delta,\beta}$ from Corollary 3.4, we obtain

$$0\leq C^2(\rho,0)\equiv \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}}-\frac{(1-\rho)}{\rho}R^2_{Y^*.W^*}-R^2_{Y^*.W^*} = \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}}-\frac{1}{\rho}R^2_{Y^*.W^*}.$$

Since $0 < \rho \leq 1$, it follows that when $0 < \sigma^2_{Y^*}$ we have

$$R_{Y^*.W^*}R_{W^*.Y^*} = \frac{R^2_{Y^*.W^*}\sigma^2_{W^*}}{\sigma^2_{Y^*}} \leq \rho \leq 1.$$

By R_1, we obtain $\max\{R_{Y^*.W^*}R_{W^*.Y^*}, \frac{1}{1+\kappa}\} \leq \rho \leq 1$. Setting $f = 0$ in the proof of Corollary 3.4 proves that $\mathcal{S}^{1,c}_{\rho,\phi,\delta,\beta}$, and thus $\mathcal{S}^{1,c}_{\rho}$, are sharp. The bounds $\mathcal{S}^{1,c}_{\delta}$ and $\mathcal{S}^{1,c}_{\beta}$ obtain by setting $\phi = 0$ in the expressions for $\delta = D(\rho, 0) = \frac{1}{\rho}R_{Y^*.W^*}$ and $\beta = B(\rho, 0) = R_{Y.X} - R_{W.X}\frac{1}{\rho}R_{Y^*.W^*}$ from Theorem 3.2. Since $\mathcal{S}^{1,c}_{\rho}$ is sharp, it follows from the mappings $D(\rho, 0)$ and $B(\rho, 0)$ that $\mathcal{S}^{1,c}_{\delta}$ and $\mathcal{S}^{1,c}_{\beta}$ are sharp.

Proof of Corollary 3.4: The identification set $\mathcal{S}^{1}_{\rho,\phi,\delta,\beta}$ obtains from R.1 and the moments $Var[(Y^*, W^*)']$, given by (in)equalities (4-6), using the expressions in Theorem 3.2. To show that $\mathcal{S}^{1}_{\rho,\phi,\delta,\beta}$ is sharp, let $d = D(r, f)$ and $b = B(r, f)$. We show that for each $(r, f, d, b) \in \mathcal{S}^{1}_{\rho,\phi,\delta,\beta}$ there exist random variables $(\tilde{U}, \tilde{\eta}, \tilde{\varepsilon})$ that satisfy A_2 and A_3 such that $Y = X'b + Wf + \tilde{U}d + \tilde{\eta}$ and $W = \tilde{U} + \tilde{\varepsilon}$ with $\frac{\sigma^2_{\tilde{U}^*}}{\sigma^2_{W^*}} = r$. For this, let V be any random variable such that $V^* \equiv \epsilon_{V.X}$ is nondegenerate and satisfies

$$\sigma_{W^*.V^*} = \sqrt{r}\sigma_{V^*}\sigma_{W^*} \quad \text{and} \quad \sigma_{Y^*.V^*} = \frac{1}{\sqrt{r}}\sigma_{W^*}\sigma_{V^*}[\frac{\sigma_{Y^*.W^*}}{\sigma^2_{W^*}} - f(1-r)].$$

Note that these covariance matrix restrictions are coherent. Specifically,

$$Var(V^*, W^*, Y^*) = \begin{bmatrix} \sigma^2_{V^*} & \sqrt{r}\sigma_{V^*}\sigma_{W^*} & \frac{\sigma_{W^*}\sigma_{V^*}}{\sqrt{r}}[\frac{\sigma_{Y^*.W^*}}{\sigma^2_{W^*}} - f(1-r)] \\ \sqrt{r}\sigma_{V^*}\sigma_{W^*} & \sigma^2_{W^*} & \sigma_{Y^*,W^*} \\ \frac{\sigma_{W^*}\sigma_{V^*}}{\sqrt{r}}[\frac{\sigma_{Y^*.W^*}}{\sigma^2_{W^*}} - f(1-r)] & \sigma_{Y^*,W^*} & \sigma^2_{Y^*} \end{bmatrix}$$

is positive semi-definite, since a basic calculation and $0 \leq C^2(r, f)$ yield the determinant

$$\det(\sigma^2_{(V^*,W^*,Y^*)'}) = (1-r)\sigma^2_{V^*}\sigma^4_{W^*}[\frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - \frac{1-r}{r}(R_{Y^*.W^*} - f)^2 - R^2_{Y^*.W^*}] = (1-r)\sigma^2_{V^*}\sigma^4_{W^*}C^2(r, f) \geq 0.$$

For instance, to construct V, set σ_{V^*} to some value (e.g. $\sigma_{V^*} = 1$) and let ϑ be any random variable that is uncorrelated with (Y^*, W^*) (e.g. a residual from a regression on (Y^*, W^*)). Then one can use the above restrictions on $\sigma_{W^*.V^*}$ and $\sigma_{Y^*.V^*}$ to construct $R_{V^*.(W^*,Y^*)'}$ and the scalar

$$\tau = \{\frac{1}{\sigma^2_{\vartheta}}[\sigma^2_{V^*} - R'_{V^*.(W^*,Y^*)'}\sigma^2_{(W^*,Y^*)}R_{V^*.(W^*,Y^*)'}]\}^{\frac{1}{2}}$$

(τ ensures that the variance of the generated V^* is $\sigma^2_{V^*}$) to construct

$$V^* = (W^*, Y^*)R_{V^*.(W^*,Y^*)'} + \tau\vartheta.$$

Last, $V = X'R_{V.X} + V^*$ obtains by setting $R_{V.X}$ to some value (e.g. 0).

Then it suffices to set $(\tilde{U}, \tilde{\varepsilon}, \tilde{\eta}) = (V^* R_{W^*.V^*} + X' R_{W.X}, \epsilon_{W.(X',V)'}, \epsilon_{Y.(X',V,\tilde{\varepsilon})'})$ so that $R_{\tilde{U}.X} = R_{W.X}$ and $\tilde{U}^* = V^* R_{W^*.V^*}$. In particular, by construction, $(X, \tilde{U}, \tilde{\varepsilon}, \tilde{\eta})$ satisfy A_2 and A_3 since $Cov[\tilde{\eta}, (X', \tilde{U})'] = 0$ and $Cov[\tilde{\varepsilon}, (\tilde{\eta}, X', \tilde{U})'] = 0$. Further, projecting the Y and W equations onto X gives

$$R_{Y.X} = b + R_{\tilde{U}.X}(f + d) = b + R_{W.X}(f + d)$$

and the following Y^* and W^* equations which are consistent with the construction of $(\tilde{U}, \tilde{\varepsilon}, \tilde{\eta})$:

$$\begin{aligned} Y^* &= W^* f + \tilde{U}^* d + \tilde{\eta} = V^* R_{W^*.V^*}(d + f) + \tilde{\varepsilon} f + \tilde{\eta} \\ W^* &= \tilde{U}^* + \tilde{\varepsilon} = V^* R_{W^*.V^*} + \tilde{\varepsilon}. \end{aligned}$$

In particular, we have $\tilde{\varepsilon} = \epsilon_{W.(X',V)'}$ and $\frac{\sigma^2_{\tilde{U}^*}}{\sigma^2_{W^*}} = \frac{\sigma^2_{W^*.V^*}}{\sigma^2_{V^*}\sigma^2_{W^*}} = r$. Last, we verify that $R_{Y^*.V^*} = R_{W^*.V^*}(d + f)$ and $R_{Y^*.\tilde{\varepsilon}} = f$. Since $Cov(V^*, \tilde{\varepsilon}) = 0$, this implies that $R_{Y^*.(V^*,\tilde{\varepsilon})'} = (R_{W^*.V^*}(d + f), f)'$ which is consistent with $\tilde{\eta} = \epsilon_{Y.(X',V,\tilde{\varepsilon})'}$. We have:

$$\begin{aligned} \frac{R_{Y^*.V^*}}{R_{W^*.V^*}} &= \frac{\sigma_{Y^*.V^*}}{\sigma_{W^*.V^*}} = \frac{\frac{1}{\sqrt{r}}\sigma_{W^*}\sigma_{V^*}[\frac{\sigma_{Y^*.W^*}}{\sigma^2_{W^*}} - f(1-r)]}{\sqrt{r}\sigma_{V^*}\sigma_{W^*}} = \frac{1}{r}[R_{Y^*.W^*} - f(1-r)] = d + f, \text{ and} \\ R_{Y^*.\tilde{\varepsilon}} &= \frac{\sigma_{Y^*.\tilde{\varepsilon}}}{\sigma^2_{\tilde{\varepsilon}}} = \frac{\sigma_{Y^*.W^*} - \frac{\sigma_{Y^*.V^*}\sigma_{W^*.V^*}}{\sigma^2_{V^*}}}{(1-r)\sigma^2_{W^*}} \\ &= \frac{1}{(1-r)\sigma^2_{W^*}}[\sigma_{Y^*.W^*} - \frac{\frac{1}{\sqrt{r}}\sigma_{W^*}\sigma_{V^*}[\frac{\sigma_{Y^*.W^*}}{\sigma^2_{W^*}} - f(1-r)]\sqrt{r}\sigma_{V^*}\sigma_{W^*}}{\sigma^2_{V^*}}] = f. \end{aligned}$$

Next, we derive the projected identification regions. $\mathcal{S}^1_\rho$ obtains by R_1. Further, recall that $B(r, f) = R_{Y.X} - R_{W.X}s$ where

$$s = S(r, f) \equiv f + D(r, f) = \frac{1}{r}[R_{Y^*.W^*} - f(1-r)].$$

For all $(1, f, d, b) \in \mathcal{S}^1_{\rho,\phi,\delta,\beta}$, we have that $s = R_{Y^*.W^*}$. Further, for all $(r, f, d, b) \in \mathcal{S}^1_{\rho,\phi,\delta,\beta}$ with $r \neq 1$, and corresponding $S(r, f)$, we have that

$$\begin{aligned} 0 \leq C^2(r, f) &= C^2(r, \frac{1}{(1-r)}(R_{Y^*.W^*} - rs)) = \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - \frac{(1-r)}{r}(f - R_{Y^*.W^*})^2 - R^2_{Y^*.W^*} \\ &= \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - \frac{r}{(1-r)}(s - R_{Y^*.W^*})^2 - R^2_{Y^*.W^*}. \end{aligned}$$

Since $\frac{1}{\kappa} \leq \frac{r}{1-r}$ by R_1, we obtain

$$\frac{1}{\kappa}(s - R_{Y^*.W^*})^2 \leq \frac{r}{(1-r)}(s - R_{Y^*.W^*})^2 \leq \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - R^2_{Y^*.W^*},$$

and therefore $s \in \mathcal{S}^1_{\phi+\delta} = \{R_{Y^*.W^*} + \lambda[\kappa(\frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - R^2_{Y^*.W^*})]^{\frac{1}{2}}\} : -1 \leq \lambda \leq 1\}$. The expression for $\mathcal{S}^1_\beta$ follows from $B(r,f) = R_{Y.X} - R_{W.X}s$.

To show that the projected regions are sharp, we show that every point in a projected region could have been generated by a point $(r,f,d,b) \in \mathcal{S}^1_{\rho,\phi,\delta,\beta}$. In particular, $\mathcal{S}^1_\rho$ is sharp since for each $r \in \mathcal{S}^1_\rho$, setting $f = R_{Y^*.W^*}$ gives $0 \leq C^2(r, R_{Y^*.W^*}) = \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - R^2_{Y^*.W^*}$ where we make use of the Cauchy Schwarz inequality. Further, $\mathcal{S}^1_\phi$ is not identified since for each $f \in \mathbb{R}$, setting $r = 1$ gives $0 \leq C^2(1,f) = \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - R^2_{Y^*.W^*}$. Similarly, $\mathcal{S}^1_\delta$ is not identified since for each $d \in \mathbb{R}$, setting $r = 1$ and $f = R_{Y^*.W^*} - rd$ gives $D(r,f) = d$ and $0 \leq C^2(1, R_{Y^*.W^*} - rd) = \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - R^2_{Y^*.W^*}$. Last, we show that $\mathcal{S}^1_{\phi+\delta}$, and thus $\mathcal{S}^1_\beta$, is sharp. For $\kappa = 0$, setting $r = 1$ gives $S(1,f) = R_{Y^*.W^*}$ and $0 \leq C^2(r,f) = \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - R^2_{Y^*.W^*}$. Otherwise, for $\kappa \neq 0$ and each $s \in \mathcal{S}^1_{\phi+\delta}$ corresponding to $\lambda_s \in [-1,1]$, setting $r = \frac{1}{1+\kappa}$ and $f = \frac{1}{(1-r)}(R_{Y^*.W^*} - rs)$ gives $S(r,f) = s$ and

$$\begin{aligned}
C^2(\frac{1}{1+\kappa}, \frac{1}{(1-r)}(R_{Y^*.W^*} - rs)) &= \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - \frac{r}{(1-r)}(s - R_{Y^*.W^*})^2 - R^2_{Y^*.W^*} \\
&= \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - \frac{1}{\kappa}\lambda_s^2\kappa(\frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - R^2_{Y^*.W^*}) - R^2_{Y^*.W^*} \\
&= (1-\lambda_s^2)(\frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - R^2_{Y^*.W^*}) \geq 0.
\end{aligned}$$

Proof of Corollary 3.5: The identification set $\mathcal{S}^{1,2^+}_{\rho,\phi,\delta,\beta}$ obtains from R_1, R_2^+, and the moments $Var[(Y^*, W^*)']$, given by (in)equalities (4-6), using the expressions in Theorem 3.2. Since $\mathcal{S}^{1,2^+}_{\rho,\phi,\delta,\beta} \subseteq \mathcal{S}^1_{\rho,\phi,\delta,\beta}$, the sharpness proof in Corollary 3.4 implies that $\mathcal{S}^{1,2^+}_{\rho,\phi,\delta,\beta}$ is sharp.

Next, we derive the projected identification regions. By R_1, $\rho \in \mathcal{S}^{1,2^+}_\rho$. By R_2^+, $0 \leq E(r,f) \equiv \frac{1}{r}f(R_{Y^*.W^*} - f)$ for all $(r,f,d,b) \in \mathcal{S}^{1,2^+}_{\rho,\phi,\delta,\beta}$ and thus $\phi \in \mathcal{S}^{1,2^+}_\phi$. $\mathcal{S}^{1,2^+}_\delta$ and $\mathcal{S}^{1,2^+}_\beta$ obtain by studying the behavior of $D(r,f)$ and $B(r,f)$ subject to the constraints defining $\mathcal{S}^{1,2^+}_{\rho,\phi,\delta,\beta}$. For brevity, we give a proof by contradiction for $\mathcal{S}^{1,2^+}_\delta$. Let $(r,f,d,b) \in \mathcal{S}^{1,2^+}_{\rho,\phi,\delta,\beta}$ and suppose that $d = D(r,f) \notin \mathcal{S}^{1,2^+}_\delta$. Then $f = R_{Y^*.W^*} - rd$ and we have that

$$\begin{aligned}
0 &\leq C^2(r, R_{Y^*.W^*} - rd) = \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - \frac{(1-r)}{r}(f - R_{Y^*.W^*})^2 - R^2_{Y^*.W^*} \\
&= \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - (1-r)rd^2 - R^2_{Y^*.W^*}, \text{ and} \\
0 &\leq E(r, R_{Y^*.W^*} - rd) = \frac{1}{r}f(R_{Y^*.W^*} - f) = (R_{Y^*.W^*} - rd)d.
\end{aligned}$$

Given R_1, it follows that

$$r(1-r) \leq \frac{1}{d^2}(\frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - R^2_{Y^*.W^*}) \quad \text{and} \quad \frac{1}{1+\kappa} \leq r \leq R_{Y^*.W^*}\frac{1}{d}.$$

If $R_{Y^*.W^*}\frac{1}{d} \leq 0$ then $r \leq R_{Y^*.W^*}\frac{1}{d} \leq 0$, a contradiction. Let $dR_{Y^*.W^*} > 0$ with $d \notin \mathcal{S}^{1,2^+}_{\delta}$. When $\kappa = 0$, this leads to the contradiction $1 \leq r \leq R_{Y^*.W^*}\frac{1}{d} < 1$. When $F_\kappa \leq 0$ and $0 < \kappa$, we obtain

$$r(1-r) < \frac{(\frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - R^2_{Y^*.W^*})}{(1+\kappa)^2 R^2_{Y^*.W^*}\frac{1}{\kappa}(\frac{1}{\rho}-1)} = \frac{1}{1+\kappa}(1-\frac{1}{1+\kappa})\frac{\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}} - 1}{(\frac{1}{\rho}-1)}, \text{ and}$$

$$\frac{1}{1+\kappa} \leq r < \frac{1}{(1+\kappa)[\frac{1}{\kappa}(\frac{1}{\rho}-1)]^{\frac{1}{2}}}$$

and when $0 \leq F_\kappa$, we obtain

$$r(1-r) < \frac{(\frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - R^2_{Y^*.W^*})}{R^2_{Y^*.W^*}\frac{1}{\overline{\rho}^2}} = \overline{\rho}^2(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}} - 1), \text{ and}$$

$$\frac{1}{1+\kappa} \leq r < R_{Y^*.W^*}\frac{1}{R_{Y^*.W^*}\frac{1}{\overline{\rho}}} = \overline{\rho}.$$

When $\overline{\rho} = \frac{1}{1+\kappa}$ this leads to the contradiction $\frac{1}{1+\kappa} \leq r < \frac{1}{1+\kappa}$. Further, when $\overline{\rho} = R_{W^*.Y^*}R_{Y^*.W^*}$ this leads to the contradictory inequalities $\frac{1}{1+\kappa} \leq r < R_{Y^*.W^*}R_{W^*.Y^*}$ and $r(1-r) < \min\{\frac{1}{1+\kappa}(1-\frac{1}{1+\kappa}), R_{Y^*.W^*}R_{W^*.Y^*}(1-R_{W^*.Y^*}R_{Y^*.W^*})\}$, where we make use of the fact that, when $\kappa \neq 0$, $F_\kappa \leq 0$ if and only if

$$\frac{1}{(1+\kappa)[\frac{1}{\kappa}(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}} - 1)]^{\frac{1}{2}}} \leq R_{Y^*.W^*}R_{W^*.Y^*}.$$

Last, $\mathcal{S}^{1,2^+}_{\beta}$ obtains from $B(r,f) = R_{Y.X} - R_{W.X}s$ and the bounds $\mathcal{S}^{1,2^+}_{\phi+\delta}$ for $s = S(r,f) \equiv \frac{1}{r}[R_{Y^*.W^*} - f(1-r)]$. Let $(r,f,d,b) \in \mathcal{S}^{1,2^+}_{\rho,\phi,\delta,\beta}$. If $r = 1$ then $s = R_{Y^*.W^*}$ whereas if $0 < r < 1$ then $f = \frac{1}{(1-r)}(R_{Y^*.W^*} - rs)$ and

$$0 \leq E(r, \frac{1}{(1-r)}(R_{Y^*.W^*} - rs)) = \frac{1}{r}f(R_{Y^*.W^*} - f) = \frac{1}{(1-r)^2}(R_{Y^*.W^*} - rs)(s - R_{Y^*.W^*}).$$

Given R_1, it follows that

$$|R_{Y^*.W^*}| \leq |s| \leq \frac{1}{r}|R_{Y^*.W^*}| \leq (1+\kappa)|R_{Y^*.W^*}| \quad \text{and} \quad 0 \leq sR_{Y^*.W^*}.$$

Thus, if $R_{Y^*.W^*} = 0$ then $\phi + \delta = 0$ and $\beta = R_{Y.X}$. Also, from Corollary 3.4, we have that $\phi + \delta \in \mathcal{S}^1_{\phi+\delta}$. And, when $R_{Y^*.W^*} \neq 0$, $1 + [\kappa(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}} - 1)]^{\frac{1}{2}} \leq 1 + \kappa$ if and only if $\frac{1}{1+\kappa} \leq R_{Y^*.W^*}R_{W^*.Y^*}$. It follows that $\phi + \delta \in \{R_{Y^*.W^*}(1 + \lambda[\kappa(\frac{1}{\rho} - 1)]^{\frac{1}{2}}) : 0 \leq \lambda \leq 1\}$.

Next, we show that the projected regions are sharp. $\mathcal{S}_\rho^{1,2^+}$ is sharp because for each $r \in \mathcal{S}_\rho^{1,2^+}$, setting $f = R_{Y^*.W^*}$ gives $0 \leq C^2(r, R_{Y^*.W^*})$ and $E(r, R_{Y^*.W^*}) = 0$. $\mathcal{S}_\phi^{1,2^+}$ is sharp because for each $f \in \mathcal{S}_\phi^{1,2^+}$, setting $r = 1$ gives $0 \leq C^2(1, f)$ and $0 \leq E(1, f)$. To show that $\mathcal{S}_\delta^{1,2^+}$ is sharp, for each $d \in \mathcal{S}_\delta^{1,2^+}$, corresponding to $\lambda_d \in [0, 1]$, set $f = R_{Y^*.W^*} - rd$ so that $d = D(r, f)$ and choose r as follows. If $R_{Y^*.W^*} = 0$ or $\kappa = 0$ then set $r = 1$ so that $0 \leq C^2(1, R_{Y^*.W^*} - rd) = \frac{\sigma_{Y^*}^2}{\sigma_{W^*}^2} - R_{Y^*.W^*}^2$, and

$$0 \leq E(1, R_{Y^*.W^*} - rd) = (R_{Y^*.W^*} - \lambda_d R_{Y^*.W^*})\lambda_d R_{Y^*.W^*} = \lambda_d(1 - \lambda_d)R_{Y^*.W^*}^2.$$

Now suppose that $R_{Y^*.W^*} \neq 0$. If $F_\kappa \leq 0$ and $0 < \kappa$ then set $r = \frac{1}{1+\kappa}$ so that

$$\begin{aligned} C^2(\frac{1}{1+\kappa}, R_{Y^*.W^*} - rd) &= \frac{\sigma_{Y^*}^2}{\sigma_{W^*}^2} - (1-r)rd^2 - R_{Y^*.W^*}^2 \\ &= \frac{\sigma_{Y^*}^2}{\sigma_{W^*}^2} - \frac{\kappa}{(1+\kappa)^2}\lambda_d^2(1+\kappa)^2 R_{Y^*.W^*}^2 \frac{1}{\kappa}(\frac{1}{\overline{\rho}} - 1) - R_{Y^*.W^*}^2 \\ &\geq \frac{\sigma_{Y^*}^2}{\sigma_{W^*}^2} - \lambda_d^2 R_{Y^*.W^*}^2(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}} - 1) - R_{Y^*.W^*}^2 \\ &= (1 - \lambda_d^2)(\frac{\sigma_{Y^*}^2}{\sigma_{W^*}^2} - R_{Y^*.W^*}^2) \geq 0. \end{aligned}$$

Further, we have that

$$\begin{aligned} E(\frac{1}{1+\kappa}, R_{Y^*.W^*} - rd) &= (R_{Y^*.W^*} - rd)d \\ &= \lambda_d(1+\kappa)R_{Y^*.W^*}^2[\frac{1}{\kappa}(\frac{1}{\overline{\rho}} - 1)]^{\frac{1}{2}} - \lambda_d^2\frac{1+\kappa}{\kappa}R_{Y^*.W^*}^2(\frac{1}{\overline{\rho}} - 1). \end{aligned}$$

If $\overline{\rho} = \frac{1}{1+\kappa}$ then

$$E(\frac{1}{1+\kappa}, R_{Y^*.W^*} - rd) = \lambda_d(1+\kappa)R_{Y^*.W^*}^2 - \lambda_d^2(1+\kappa)R_{Y^*.W^*}^2 = \lambda_d(1-\lambda_d)(1+\kappa)R_{Y^*.W^*}^2 \geq 0.$$

If $\overline{\rho} = R_{Y^*.W^*}R_{W^*.Y^*}$ then $\frac{1}{\kappa}(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}} - 1) \leq 1$ and

$$\begin{aligned} &E(\frac{1}{1+\kappa}, R_{Y^*.W^*} - rd) \\ &= \lambda_d R_{Y^*.W^*}^2(1+\kappa)[\frac{1}{\kappa}(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}} - 1)]^{\frac{1}{2}} - \lambda_d^2\frac{1+\kappa}{\kappa}R_{Y^*.W^*}^2(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}} - 1) \\ &\geq \lambda_d R_{Y^*.W^*}^2(1+\kappa)\frac{1}{\kappa}(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}} - 1) - \lambda_d^2\frac{1+\kappa}{\kappa}R_{Y^*.W^*}^2(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}} - 1) \\ &= \lambda_d(1-\lambda_d)\frac{1+\kappa}{\kappa}R_{Y^*.W^*}^2(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}} - 1) \geq 0. \end{aligned}$$

Otherwise, if $0 \leq F_\kappa$ then set $r = \overline{\rho}$ so that

$$C^2(\overline{\rho}, R_{Y^*.W^*} - rd) = \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - (1-r)rd^2 - R^2_{Y^*.W^*} = \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - \frac{1-\overline{\rho}}{\overline{\rho}}\lambda_d^2 R^2_{Y^*.W^*} - R^2_{Y^*.W^*}.$$

If $\overline{\rho} = \frac{1}{1+\kappa}$ then $\kappa \leq \frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}} - 1$ and

$$\begin{aligned}
C^2(\overline{\rho}, R_{Y^*.W^*} - rd) &= \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - \kappa\lambda_d^2 R^2_{Y^*.W^*} - R^2_{Y^*.W^*} \\
&\geq \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - (\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}} - 1)\lambda_d^2 R^2_{Y^*.W^*} - R^2_{Y^*.W^*} \\
&= (1-\lambda_d^2)(\frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - R^2_{Y^*.W^*}) \geq 0.
\end{aligned}$$

If $\overline{\rho} = R_{Y^*.W^*}R_{W^*.Y^*}$ then

$$\begin{aligned}
C^2(\overline{\rho}, R_{Y^*.W^*} - rd) &= \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - \frac{1 - R_{Y^*.W^*}R_{W^*.Y^*}}{R_{Y^*.W^*}R_{W^*.Y^*}}\lambda_d^2 R^2_{Y^*.W^*} - R^2_{Y^*.W^*} \\
&= (1-\lambda_d^2)(\frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - R^2_{Y^*.W^*}) \geq 0.
\end{aligned}$$

Further, we have

$$\begin{aligned}
E(\overline{\rho}, R_{Y^*.W^*} - rd) &= (R_{Y^*.W^*} - rd)d = (R_{Y^*.W^*} - \overline{\rho}\lambda_d R_{Y^*.W^*}\frac{1}{\overline{\rho}})\lambda_d R_{Y^*.W^*}\frac{1}{\overline{\rho}} \\
&= (1-\lambda_d)\lambda_d R^2_{Y^*.W^*}\frac{1}{\overline{\rho}} \geq 0.
\end{aligned}$$

Last, since $B(r, f) = R_{Y.X} - R_{W.X}s$, it suffices to show that $\mathcal{S}^{1,2^+}_{\phi+\delta}$, and thus $\mathcal{S}^{1,2^+}_{\beta}$, is sharp. If $R_{Y^*.W^*} = 0$ or $\kappa = 0$ then setting $r = 1$ and $f = R_{Y^*.W^*}$, so that $s = S(r, f) = R_{Y^*.W^*}$, gives $0 \leq C^2(1, R_{Y^*.W^*}) = \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - R^2_{Y^*.W^*}$ and $E(1, R_{Y^*.W^*}) = 0$. Otherwise, for $R_{Y^*.W^*} \neq 0$, $\kappa \neq 0$, and each $s \in \mathcal{S}^{1,2^+}_{\phi+\delta}$ corresponding to $\lambda_s \in [0, 1]$, setting $r = \frac{1}{1+\kappa}$ and $f = \frac{1}{(1-r)}(R_{Y^*.W^*} - rs)$, so that $S(r, f) = s$, yields

$$\begin{aligned}
&C^2(\frac{1}{1+\kappa}, \frac{1}{(1-r)}(R_{Y^*.W^*} - rs)) \\
&= \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - \frac{r}{(1-r)}(s - R_{Y^*.W^*})^2 - R^2_{Y^*.W^*} \\
&= \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - \frac{1}{\kappa}(R_{Y^*.W^*}\{1 + \lambda_s[\kappa(\frac{1}{\overline{\rho}} - 1)]^{\frac{1}{2}}\} - R_{Y^*.W^*})^2 - R^2_{Y^*.W^*} \\
&= \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - R^2_{Y^*.W^*}\lambda_s^2(\frac{1}{\overline{\rho}} - 1) - R^2_{Y^*.W^*} \geq 0
\end{aligned}$$

where the last inequality is shown above. Since $\kappa \geq \frac{1}{\rho} - 1$, we also have

$$
\begin{aligned}
& E(\frac{1}{1+\kappa}, \frac{1}{(1-r)}(R_{Y^*.W^*} - rs)) \\
&= \frac{1}{(1-r)^2}(R_{Y^*.W^*} - rs)(s - R_{Y^*.W^*}) \\
&= \frac{(1+\kappa)^2}{\kappa^2}(R_{Y^*.W^*} - \frac{1}{1+\kappa}R_{Y^*.W^*}\{1 + \lambda_s[\kappa(\frac{1}{\rho} - 1)]^{\frac{1}{2}}\})(R_{Y^*.W^*}\lambda_s[\kappa(\frac{1}{\rho} - 1)]^{\frac{1}{2}}) \\
&= \frac{1+\kappa}{\kappa^2}R^2_{Y^*.W^*}(\kappa - \lambda_s[\kappa(\frac{1}{\rho} - 1)]^{\frac{1}{2}})(\lambda_s[\kappa(\frac{1}{\rho} - 1)]^{\frac{1}{2}}) \\
&= \lambda_s \frac{1+\kappa}{\kappa}R^2_{Y^*.W^*}\{[\kappa(\frac{1}{\rho} - 1)]^{\frac{1}{2}} - \lambda_s(\frac{1}{\rho} - 1)\} \\
&\geq \lambda_s \frac{1+\kappa}{\kappa}R^2_{Y^*.W^*}[(\frac{1}{\rho} - 1)^2]^{\frac{1}{2}} - \lambda_s(\frac{1}{\rho} - 1)] = \lambda_s(1-\lambda_s)\frac{1+\kappa}{\kappa}R^2_{Y^*.W^*}(\frac{1}{\rho} - 1) \geq 0.
\end{aligned}
$$

Proof of Corollary 3.6: The identification region $\mathcal{S}^{1,2^-}_{\rho,\phi,\delta,\beta}$ obtains from R_1 , R_2^-, and the moments $Var[(Y^*, W^*)']$ given by (in)equalities (4-6), using the expressions in Theorem 3.2. Since $\mathcal{S}^{1,2^-}_{\rho,\phi,\delta,\beta} \subseteq \mathcal{S}^{1}_{\rho,\phi,\delta,\beta}$, the sharpness proof in Corollary 3.4 applies to $\mathcal{S}^{1,2^-}_{\rho,\phi,\delta,\beta}$.

Next, we derive the projected identification regions. By R_1, $\rho \in \mathcal{S}^{1,2^-}_{\rho}$. If $R_{Y^*.W^*} = 0$ then $\mathcal{S}^{1,2^-}_{\phi} = \mathcal{S}^{1,2^-}_{\delta} = \mathbb{R}$. If $R_{Y^*.W^*} \neq 0$, since $E(r,f) = \frac{1}{r}f(R_{Y^*.W^*} - f) \leq 0$ for all $(r,f,d,b) \in \mathcal{S}^{1,2^-}_{\rho,\phi,\delta,\beta}$, we have that $f \in \mathcal{S}^{1,2^-}_{\phi}$. Similarly, if $R_{Y^*.W^*} \neq 0$ let $(r,f,d,b) \in \mathcal{S}^{1,2^-}_{\rho,\phi,\delta,\beta}$ and suppose that $d = D(r,f) \notin \mathcal{S}^{1,2^-}_{\delta}$ (i.e. $d \in \{\lambda R_{Y^*.W^*} : 0 < \lambda < 1\}$) then $E(r, R_{Y^*.W^*} - rd) = (R_{Y^*.W^*} - rd)d \leq 0$ implies that $1 < \frac{R_{Y^*.W^*}}{d} \leq r$, a contradiction. Last, we derive $\mathcal{S}^{1,2^-}_{\beta}$. If $R_{Y^*.W^*}R_{W^*.Y^*} \leq \frac{1}{1+\kappa}$ then $\mathcal{S}^{1,2^-}_{\beta} = \mathcal{S}^{1}_{\beta}$, the bounds obtained from Corollary 3.4. To derive $\mathcal{S}^{1,2^-}_{\beta}$ when $\frac{1}{1+\kappa} \leq R_{Y^*.W^*}R_{W^*.Y^*}$, it suffices to derive the bound $\mathcal{S}^{1,2^-}_{\phi+\delta}$ for $\phi + \delta$ since $\beta = R_{Y.X} - R_{W.X}(\phi + \delta)$. Let $(r,f,d,b) \in \mathcal{S}^{1,2^-}_{\rho,\phi,\delta,\beta}$ and suppose that $s = S(r,f) \notin \mathcal{S}^{1,2^-}_{\phi+\delta}$. Then, from Corollary 3.4, we have that

$$
s \in \mathcal{S}^{1}_{\phi+\delta} \backslash \mathcal{S}^{1,2^-}_{\phi+\delta} = \{\lambda \frac{1}{R_{W^*.Y^*}} + (1-\lambda)R_{Y^*.W^*}(1 + [\kappa(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}} - 1)]^{\frac{1}{2}}) : 0 \leq \lambda < 1\}.
$$

Note that $R_{Y^*.W^*}R_{W^*.Y^*} < 1$ since if $R_{Y^*.W^*}R_{W^*.Y^*} = 1$ then $\mathcal{S}^{1}_{\phi+\delta} \backslash \mathcal{S}^{1,2^-}_{\phi+\delta}$ is empty. Further, note that s, $\frac{1}{R_{W^*.Y^*}}$, and $R_{Y^*.W^*}$ have the same sign. Last, since $\frac{1}{1+\kappa} \leq R_{Y^*.W^*}R_{W^*.Y^*}$, we obtain that

$$
|R_{Y^*.W^*}| < \frac{1}{|R_{W^*.Y^*}|} < |s| \leq |R_{Y^*.W^*}| (1 + [\kappa(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}} - 1)]^{\frac{1}{2}}).
$$

If $r = 1$ then $s = R_{Y^*.W^*} \in \mathcal{S}^{1,2^-}_{\delta}$. Thus, $r \neq 1$ and we have

$$
E(r, \frac{1}{1-r}(R_{Y^*.W^*} - rs)) = \frac{1}{(1-r)^2}(R_{Y^*.W^*} - rs)(s - R_{Y^*.W^*}) \leq 0.
$$

Further, $E(r,f)=0$ if and only if either $d=0$, and thus $s=R_{Y^*.W^*}$, or $f=0$, and thus $s\in\mathcal{S}_\delta^{1,c}$, contradicting $s\in\mathcal{S}^1_{\phi+\delta}\backslash\mathcal{S}^{1,2^-}_{\phi+\delta}$. Therefore, we must have $E(r,\frac{1}{1-r}(R_{Y^*.W^*}-rs))<0$, so that either $|s|<|R_{Y^*.W^*}|<\frac{1}{r}|R_{Y^*.W^*}|$ (which is ruled out by $|R_{Y^*.W^*}|<|s|$ above) or $|R_{Y^*.W^*}|<\frac{1}{r}|R_{Y^*.W^*}|<|s|$. In particular, we obtain that $\left|\frac{R_{Y^*.W^*}}{s}\right|<r$. Since

$$0\le C^2(r,\frac{1}{1-r}(R_{Y^*.W^*}-rs))=\frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}}-\frac{r}{1-r}(s-R_{Y^*.W^*})^2-R^2_{Y^*.W^*},$$

using $\left|\frac{R_{Y^*.W^*}}{s}\right|<r$ and $0<|s|-|R_{Y^*.W^*}|\le|s-R_{Y^*.W^*}|$ we obtain

$$|R_{Y^*.W^*}|\,|s-R_{Y^*.W^*}|\le\frac{|R_{Y^*.W^*}|}{|s|-|R_{Y^*.W^*}|}(s-R_{Y^*.W^*})^2<\frac{r}{1-r}(s-R_{Y^*.W^*})^2\le\frac{R_{Y^*.W^*}}{R_{W^*.Y^*}}-R^2_{Y^*.W^*}.$$

It follows that

$$|s-R_{Y^*.W^*}|<\left|\frac{1}{R_{W^*.Y^*}}-R_{Y^*.W^*}\right|.$$

But since s, $\frac{1}{R_{W^*.Y^*}}$, and $R_{Y^*.W^*}$ have the same sign this contradicts $|R_{Y^*.W^*}|<\frac{1}{|R_{W^*.Y^*}|}<|s|$.

$\mathcal{S}^{1,2^-}_\rho$ is sharp since for each $r\in\mathcal{S}^{1,2^-}_\rho$ setting $f=R_{Y^*.W^*}$ gives $0\le C^2(r,R_{Y^*.W^*})$ and $E(r,R_{Y^*.W^*})=0$. $\mathcal{S}^{1,2^-}_\phi$ is sharp since for each $f\in\{\lambda R_{Y^*.W^*}:\lambda\notin(0,1)\}$ and corresponding $\lambda_f\notin(0,1)$ when $R_{Y^*.W^*}\ne 0$ (or $f\in\mathbb{R}$ when $R_{Y^*.W^*}=0$) setting $r=1$ gives $0\le C^2(1,f)=\frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}}-R^2_{Y^*.W^*}$ and $E(1,f)=(1-\lambda_f)\lambda_f R^2_{Y^*.W^*}\le 0$ (or $E(1,f)=-f^2\le 0$ when $R_{Y^*.W^*}=0$). Similarly, $\mathcal{S}^{1,2^-}_\delta$ is sharp since for each $d\in\{\lambda R_{Y^*.W^*}:\lambda\notin(0,1)\}$ and corresponding $\lambda_d\notin(0,1)$ when $R_{Y^*.W^*}\ne 0$ (or $d\in\mathbb{R}$ when $R_{Y^*.W^*}=0$) setting $r=1$ and $f=R_{Y^*.W^*}-rd$ gives $D(r,f)=d$, $0\le C^2(1,R_{Y^*.W^*}-rd)=\frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}}-R^2_{Y^*.W^*}$, and $E(1,R_{Y^*.W^*}-rd)=(1-\lambda_d)\lambda_d R^2_{Y^*.W^*}\le 0$ (or $E(1,R_{Y^*.W^*}-rd)=-d^2\le 0$ when $R_{Y^*.W^*}=0$).

Given that $B(r,f)=R_{Y.X}-R_{W.X}S(r,f)$, in order to show that $\mathcal{S}^{1,2^-}_\beta$ is sharp it suffices to show that $\mathcal{S}^{1,2^-}_{\phi+\delta}$ is sharp. If $R_{Y^*.W^*}=0$ or $\kappa=0$ then setting $r=1$ and $f=R_{Y^*.W^*}$, so that $s=S(r,f)=R_{Y^*.W^*}$, gives $0\le C^2(1,R_{Y^*.W^*})=\frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}}-R^2_{Y^*.W^*}$, and $E(1,R_{Y^*.W^*})=0$. Otherwise, for $R_{Y^*.W^*}\ne 0$, $\kappa\ne 0$, and each $s\in\mathcal{S}^{1,2-}_{\phi+\delta}$ corresponding to λ_s, let $f(1-r)=(R_{Y^*.W^*}-rs)$ so that $S(r,f)=s$. Partition $\mathcal{S}^{1,2-}_{\phi+\delta}$ and choose either r or f as follows. For each $s\in\{R_{Y^*.W^*}(1-\lambda[\kappa(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}}-1)]^{\frac{1}{2}}):0\le\lambda\le 1\}$, set $r=\frac{1}{1+\kappa}$ so that

$$\begin{aligned}C^2(\frac{1}{1+\kappa},\frac{1}{1-r}(R_{Y^*.W^*}-rs))&=\frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}}-\frac{r}{(1-r)}(s-R_{Y^*.W^*})^2-R^2_{Y^*.W^*}\\&=(1-\lambda_s^2)(\frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}}-R^2_{Y^*.W^*})\ge 0,\end{aligned}$$

and

$$
\begin{aligned}
&E(\frac{1}{1+\kappa},\frac{1}{1-r}(R_{Y^*.W^*}-rs)) = \frac{1}{(1-r)^2}(R_{Y^*.W^*}-rs)(s-R_{Y^*.W^*})\\
&= \frac{(1+\kappa)}{\kappa^2}\{(1+\kappa)R_{Y^*.W^*} - [R_{Y^*.W^*} - \lambda_s R_{Y^*.W^*}[\kappa(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}}-1)]^{\frac{1}{2}}]\}\\
&\qquad\qquad \times [-\lambda_s R_{Y^*.W^*}[\kappa(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}}-1)]^{\frac{1}{2}}]\\
&= \frac{(1+\kappa)}{\kappa}[-\lambda_s R^2_{Y^*.W^*}[\kappa(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}}-1)]^{\frac{1}{2}} - \lambda_s^2 R^2_{Y^*.W^*}(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}}-1)] \le 0.
\end{aligned}
$$

Further, for each $s \in \{\lambda\frac{1}{\rho}R_{Y^*.W^*} + (1-\lambda)R_{Y^*.W^*} : 0 < \lambda \le 1\}$ set $f = 0$ so that

$$
\begin{aligned}
C^2(\frac{1}{s}R_{Y^*.W^*},0) &= \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - \frac{(1-r)}{r}(f-R_{Y^*.W^*})^2 - R^2_{Y^*.W^*}\\
&= \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - \frac{s}{R_{Y^*.W^*}}R^2_{Y^*.W^*}\\
&= \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - (\lambda_s\frac{1}{\rho}R_{Y^*.W^*} + (1-\lambda_s)R_{Y^*.W^*})R_{Y^*.W^*}\\
&\ge \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - (\lambda_s\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}}R_{Y^*.W^*} + (1-\lambda_s)R_{Y^*.W^*})R_{Y^*.W^*}\\
&= (1-\lambda_s)(\frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - R^2_{Y^*.W^*}) \ge 0
\end{aligned}
$$

and

$$
E(\frac{1}{s}R_{Y^*.W^*},0) = \frac{s}{R_{Y^*.W^*}}f(R_{Y^*.W^*}-f) = 0.
$$

Last, if $R_{Y^*.W^*}R_{W^*.Y^*} \le \frac{1}{1+\kappa}$ then for each $s \in \{\lambda(1+\kappa)R_{Y^*.W^*} + (1-\lambda)R_{Y^*.W^*}(1 + [\kappa(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}}-1)]^{\frac{1}{2}}) : 0 \le \lambda \le 1\}$ set $r = \frac{1}{1+\kappa}$ so that

$$
\begin{aligned}
&C^2(\frac{1}{1+\kappa},\frac{1}{1-r}(R_{Y^*.W^*}-rs))\\
&= \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - \frac{r}{(1-r)}(s-R_{Y^*.W^*})^2 - R^2_{Y^*.W^*}\\
&= \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - \frac{1}{\kappa}(\lambda_s(1+\kappa)R_{Y^*.W^*} + (1-\lambda_s)R_{Y^*.W^*}(1+[\kappa(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}}-1)]^{\frac{1}{2}} - R_{Y^*.W^*})^2 - R^2_{Y^*.W^*}\\
&= \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - R^2_{Y^*.W^*}(\lambda_s\kappa^{\frac{1}{2}} + (1-\lambda_s)(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}}-1)^{\frac{1}{2}})^2 - R^2_{Y^*.W^*}\\
&\ge \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}} - R^2_{Y^*.W^*}(\lambda_s(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}}-1)^{\frac{1}{2}} + (1-\lambda_s)(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}}-1)^{\frac{1}{2}})^2 - R^2_{Y^*.W^*} = 0
\end{aligned}
$$

and

$$E(\frac{1}{1+\kappa}, \frac{1}{1-r}(R_{Y^*.W^*} - rs)) = \frac{1}{(1-r)^2}(R_{Y^*.W^*} - rs)(s - R_{Y^*.W^*})$$
$$= \frac{(1+\kappa)}{\kappa^2}[(1+\kappa)R_{Y^*.W^*} - \lambda_s(1+\kappa)R_{Y^*.W^*} - (1-\lambda_s)R_{Y^*.W^*}(1+[\kappa(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}} - 1)]^{\frac{1}{2}})]$$
$$\times (\lambda_s(1+\kappa)R_{Y^*.W^*} + (1-\lambda_s)R_{Y^*.W^*}(1+[\kappa(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}} - 1)]^{\frac{1}{2}}) - R_{Y^*.W^*})$$
$$= \frac{(1+\kappa)}{\kappa^2}(1-\lambda_s)R^2_{Y^*.W^*}(\kappa - [\kappa(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}} - 1)]^{\frac{1}{2}})(\lambda_s\kappa + (1-\lambda_s)[\kappa(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}} - 1)]^{\frac{1}{2}}) \leq 0$$

since $R_{Y^*.W^*}R_{W^*.Y^*} \leq \frac{1}{1+\kappa}$ implies that $\kappa - [\kappa(\frac{1}{R_{Y^*.W^*}R_{W^*.Y^*}} - 1)]^{\frac{1}{2}} \leq 0$.

Proof of Theorem 5.1 Recall that, for random column vectors A and B, we have

$$A' = [E(A)' - E(B)'R_{A.B}] + B'R_{A.B} + \epsilon'_{A.B} \equiv (1, B')R^*_{A.B} + \epsilon'_{A.B}.$$

Given observations $\{A_i, B_i\}_{i=1}^n$, denote the linear regression intercept ($\hat{R}^0_{A.B}$) and slope ($\hat{R}_{A.B}$) estimators and the sample residual ($\hat{\epsilon}_{A.B,i}$) by:

$$\tilde{R}_{A.B} = (\hat{R}^0_{A.B}, \hat{R}'_{A.B})' \equiv (\frac{1}{n}\sum_{i=1}^n (1, B'_i)'(1, B'_i))^{-1}(\frac{1}{n}\sum_{i=1}^n (1, B'_i)'A'_i) \text{ and } \hat{\epsilon}'_{A.B,i} \equiv A'_i - (1, B'_i)\tilde{R}_{A.B}.$$

Further, collect into R^* and $\tilde{R}$ the following estimands and estimators

$$R^* \equiv (R^{*\prime}_{Y.(W,X')'}, R^{*\prime}_{W.(Y,X')'}, R^{*\prime}_{Y.X}, R^{*\prime}_{W.X}, \frac{\sigma^2_{Y^*}}{\sigma^2_{W^*}})' \text{ and } \tilde{R} \equiv (\tilde{R}'_{Y.(W,X')'}, \tilde{R}'_{W.(Y,X')'}, \tilde{R}'_{Y.X}, \tilde{R}'_{W.X}, \frac{\sum_{i=1}^n \hat{\epsilon}^2_{Y.X,i}}{\sum_{i=1}^n \hat{\epsilon}^2_{W.X,i}})'.$$

Last, let $\hat{\mu}^2_A = \frac{1}{n}\sum_{i=1}^n A_i A'_i$,

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

and

$$M = \frac{1}{n}\sum_{i=1}^n [(1, W_i, X'_i)\epsilon_{Y.(W,X')',i},\ (1, Y_i, X'_i)\epsilon_{Y.(W,X')',i},\ (1, X'_i)'\epsilon_{Y.X,i},\ (1, X'_i)'\epsilon_{W.X,i}, \epsilon^2_{Y.X,i} - \sigma^2_{Y^*}]'.$$

Since $Var(Y, W, X')$ is finite, we have that Q is finite and nonsingular. For a symmetric matrix C and a vector D, let C_1 denote the submatrix that removes the last row and column of C and let D_1 be the subvector that removes the last row of D. Then

$$\sqrt{n}(\tilde{R}_1 - R^*_1) = \hat{Q}_1^{-1}\sqrt{n}M_1 = (\hat{Q}_1^{-1} - Q_1^{-1})\sqrt{n}M_1 + Q_1^{-1}\sqrt{n}M_1,$$

exists in probability for all n sufficiently large. Since (i) gives $\hat{Q}_1^{-1} - Q_1^{-1} = o_p(1)$ and (ii) gives $\sqrt{n}M_1 \xrightarrow{d} N(0, \Xi_1)$, we obtain that $\sqrt{n}(\tilde{R}_1 - R_1^*) = Q_1^{-1}\sqrt{n}M_1 + o_p(1) \xrightarrow{d} N(0, \Gamma_1^*)$. In particular, it follows from $\sqrt{n}(\tilde{R}_{Y.X} - R_{Y.X}^*) = O_p(1)$, $\hat{\mu}^2_{(1,X')'} \xrightarrow{p} \mu^2_{(1,X')'}$, and $\frac{1}{n}\sum_{i=1}^n \epsilon_{Y.X,i}(1, X_i')' = E[\epsilon_{Y.X}(1, X')'] + o_p(1) = o_p(1)$ that

$$
\begin{aligned}
n^{-\frac{1}{2}}\sum_{i=1}^n \hat{\epsilon}^2_{Y.X,i} &= n^{-\frac{1}{2}}\sum_{i=1}^n (\epsilon_{Y.X,i} - (1, X_i')(\tilde{R}_{Y.X} - R_{Y.X}^*))^2 \\
&= n^{-\frac{1}{2}}\sum_{i=1}^n \epsilon^2_{Y.X,i} + (\tilde{R}_{Y.X} - R_{Y.X}^*)'\hat{\mu}^2_{(1,X')'}\sqrt{n}(\tilde{R}_{Y.X} - R_{Y.X}^*) \\
&- 2[\frac{1}{n}\sum_{i=1}^n \epsilon_{Y.X,i}(1, X_i')]\sqrt{n}(\tilde{R}_{Y.X} - R_{Y.X}^*) \\
&= n^{-\frac{1}{2}}\sum_{i=1}^n \epsilon^2_{Y.X,i} + o_p(1).
\end{aligned}
$$

Similarly, by (i), we have that

$$
\frac{1}{n}\sum_{i=1}^n \hat{\epsilon}^2_{Y.X,i} = E(\epsilon^2_{Y.X}) + o_p(1) = \sigma^2_{Y^*} + o_p(1) \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n \hat{\epsilon}^2_{W.X,i} = \sigma^2_{W^*} + o_p(1).
$$

Thus, since $n^{-1/2}\sum_{i=1}^n \epsilon^2_{Y.X,i}$ is $O_p(1)$ by (ii), we have

$$
\sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^n \hat{\epsilon}^2_{Y.X,i}}{\frac{1}{n}\sum_{i=1}^n \hat{\epsilon}^2_{W.X,i}} = (\sigma^2_{W^*})^{-1} n^{-\frac{1}{2}}\sum_{i=1}^n \epsilon^2_{Y.X,i} + o_p(1).
$$

Together with $\sqrt{n}(\tilde{R}_1 - R_1^*) = Q_1^{-1}\sqrt{n}M_1 + o_p(1)$, we obtain by (i) and (ii) that

$$
\sqrt{n}(\tilde{R} - R^*) = Q^{-1}\sqrt{n}M + o_p(1) \xrightarrow{d} N(0, \Gamma^*)
$$

and thus that the subvector $\sqrt{n}(\hat{R} - R) \xrightarrow{d} N(0, \Gamma)$.

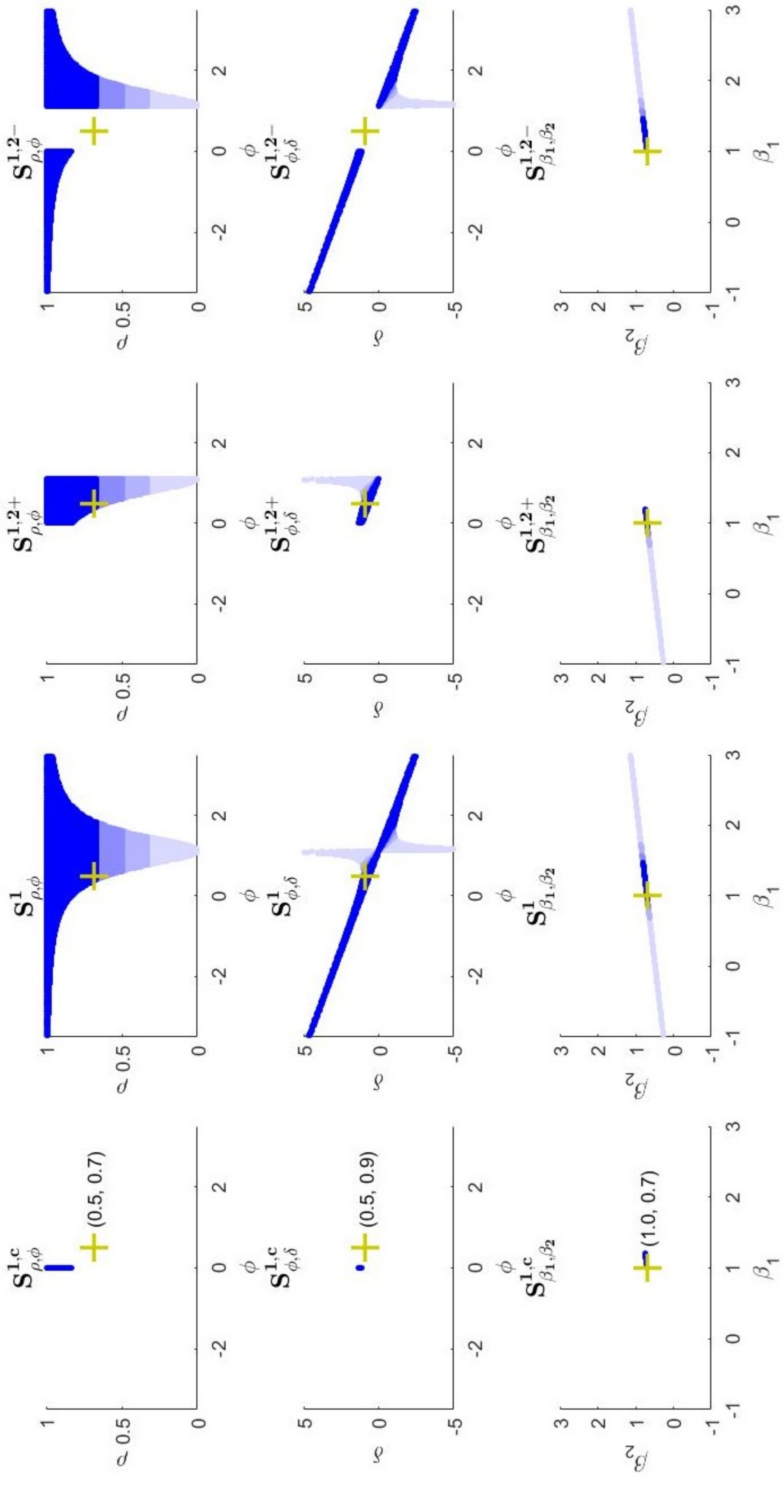

Figure 1: Sharp Identification Regions for $\kappa = +\infty(\text{light}), 2, 1, 0.5(\text{dark})$.

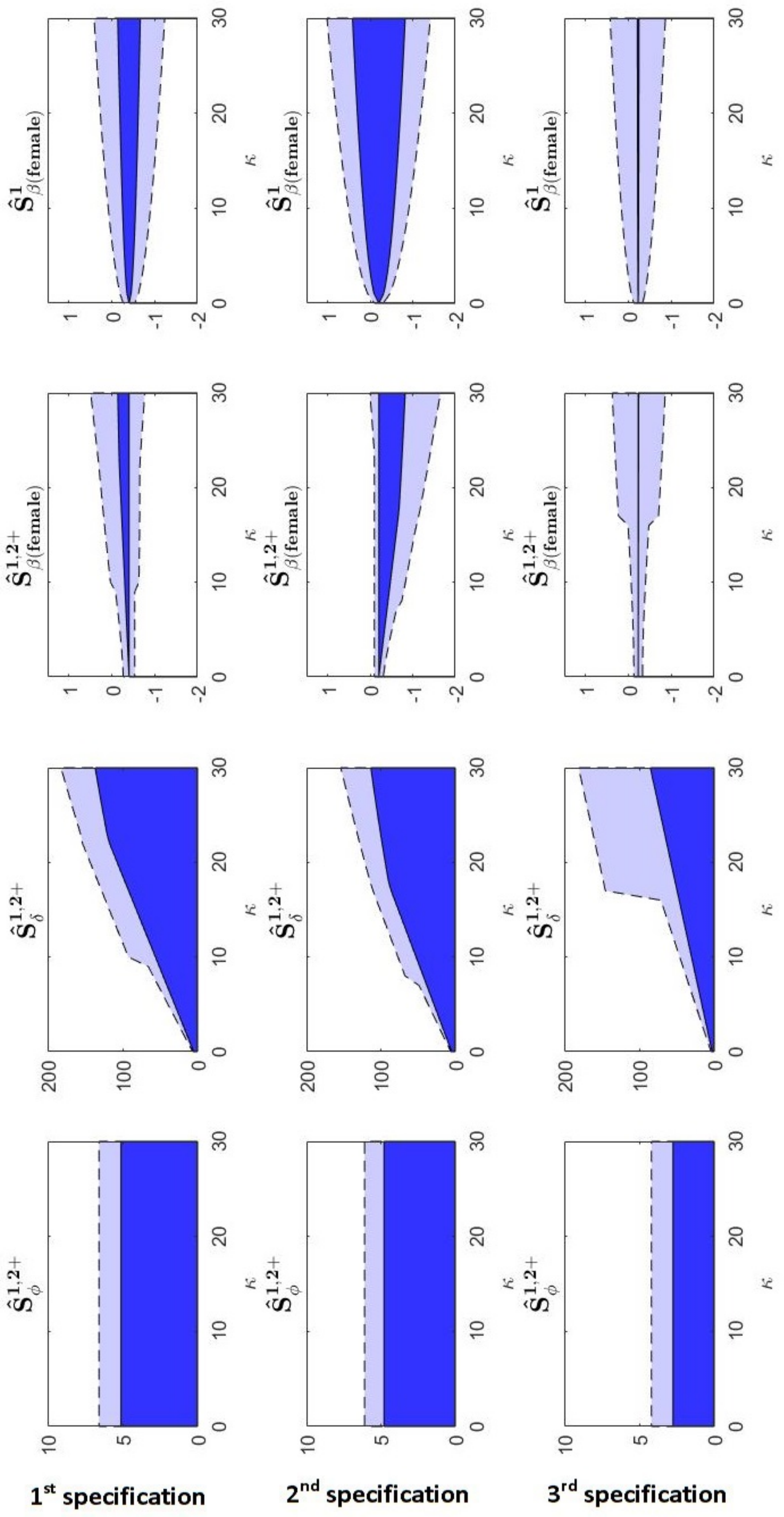

Figure 2: Returns Estimates and 95% Confidence Regions for $\kappa \in [0, 30]$.

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

	$R_{Y.(W,X')'}$	$\mathcal{S}_{\theta}^{1,c}$	$\mathcal{S}_{\theta}^{1}$	$\mathcal{S}_{\theta}^{1,2^{+}}$	$\mathcal{S}_{\theta}^{1,2\text{-}}$
			$\kappa \to \infty$		
ρ		$[0.832, 1]$	$(0, 1]$	$(0, 1]$	$(0, 1]$
ϕ	1.116	0	$[-\infty, \infty]$	$[0, 1.116]$	$\mathbb{R}\backslash(0, 1.116)$
δ		$[1.116, 1.341]$	$[-\infty, \infty]$	$[0, \infty]$	$\mathbb{R}\backslash(0, 1.116)$
β_1	1.207	$[1.043, 1.207]$	$[-\infty, \infty]$	$[-\infty, 1.207]$	$[1.043, \infty]$
β_2	0.745	$[0.709, 0.745]$	$[-\infty, \infty]$	$[-\infty, 0.745]$	$[0.709, \infty]$
			$\kappa = 2$		
ρ		$[0.832, 1]$	$[0.333, 1]$	$[0.333, 1]$	$[0.333, 1]$
ϕ		0	$[-\infty, \infty]$	$[0, 1.116]$	$\mathbb{R}\backslash(0, 1.116)$
δ		$[1.116, 1.341]$	$[-\infty, \infty]$	$[0, 1.341]$	$\mathbb{R}\backslash(0, 1.116)$
β_1		$[1.043, 1.207]$	$[0.691, 1.722]$	$[0.691, 1.207]$	$[1.043, 1.722]$
β_2		$[0.709, 0.745]$	$[0.633, 0.857]$	$[0.633, 0.745]$	$[0.709, 0.857]$
			$\kappa = 1$		
ρ		$[0.832, 1]$	$[0.5, 1]$	$[0.5, 1]$	$[0.5, 1]$
ϕ		0	$[-\infty, \infty]$	$[0, 1.116]$	$\mathbb{R}\backslash(0, 1.116)$
δ		$[1.116, 1.341]$	$[-\infty, \infty]$	$[0, 1.341]$	$\mathbb{R}\backslash(0, 1.116)$
β_1		$[1.043, 1.207]$	$[0.842, 1.571]$	$[0.842, 1.207]$	$[1.043, 1.571]$
β_2		$[0.709, 0.745]$	$[0.666, 0.824]$	$[0.666, 0.745]$	$[0.709, 0.824]$
			$\kappa = 0.5$		
ρ		$[0.832, 1]$	$[0.667, 1]$	$[0.667, 1]$	$[0.667, 1]$
ϕ		0	$[-\infty, \infty]$	$[0, 1.116]$	$\mathbb{R}\backslash(0, 1.116)$
δ		$[1.116, 1.341]$	$[-\infty, \infty]$	$[0, 1.341]$	$\mathbb{R}\backslash(0, 1.116)$
β_1		$[1.043, 1.207]$	$[0.949, 1.464]$	$[0.949, 1.207]$	$[1.043, 1.464]$
β_2		$[0.709, 0.745]$	$[0.689, 0.801]$	$[0.689, 0.745]$	$[0.709, 0.801]$

Table 1 reports the population regression estimands and sharp projected identification regions from Section 3 for $\kappa = +\infty, 2, 1, 0.5$. $r_{Y^*,W^*} = 0.9124$.

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

Variable	Name	Mean	Std. Dev.	Min	Max
Mean earnings	MnEarnWnEP6	36985.24	9513.16	15700	102700
Share earning over $25K/year	Gt25KP6	0.6528	0.1077	0.162	0.918
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
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 nonresident aliens	UGDSnRA	0.0291	0.0342	0	0.3617
Share of females	Female	0.5869	0.1076	0.0773	0.986
Share of dependents	Dependent	0.7504	0.1645	0.1238	0.9886
% with tertiary-educated parent	ParEdPctPS	0.6556	0.1025	0.4108	0.9381
Average family income	FamInc	70495.26	21358.66	17501.84	143865.7
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
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

Table 3: The Returns to College Selectivity and Characteristics ($\kappa = 1$)

	$\hat{R}_{Y.(W,X')'}$	$\hat{\mathcal{S}}_{\theta}^{1,c}$	$\hat{\mathcal{S}}_{\theta}^{1}$	$\hat{\mathcal{S}}_{\theta}^{1,2^+}$
$10^{-4}\times$SATAvg	5.105	0	$[-\infty, \infty]$	$[0, 5.105]$
	$(3.635, 6.576)$	$(0, 0)$	$(-\infty, \infty)$	$(0, 6.576)$
$10^{-4}\times$U (ability)		$[5.105, 10.21]$	$[-\infty, \infty]$	$[0, 10.21]$
		$(3.635, 13.15)$	$(-\infty, \infty)$	$(0, 13.15)$
$10^{-2}\times$ControlInd	2.263	$[2.263, 3.054]$	$[-1.474, 6.001]$	$[2.263, 3.054]$
	$(-2.849, 7.376)$	$(-2.851, 8.328)$	$(-8.163, 13)$	$(-2.851, 8.328)$
$10^{-2}\times$HDeg	4.228	$[4.228, 4.697]$	$[2.012, 6.444]$	$[4.228, 4.697]$
	$(1.747, 6.708)$	$(1.746, 7.263)$	$(-1.496, 10.08)$	$(1.746, 7.263)$
log(UGDS)	0.022	$[0.014, 0.022]$	$[-0.013, 0.056]$	$[0.014, 0.022]$
	$(0.006, 0.037)$	$(-0.002, 0.037)$	$(-0.034, 0.075)$	$(-0.002, 0.037)$
UGDSBlack	0.307	$[0.307, 0.353]$	$[0.087, 0.526]$	$[0.307, 0.353]$
	$(0.229, 0.385)$	$(0.229, 0.437)$	$(-0.035, 0.648)$	$(0.229, 0.437)$
UGDSHisp	0.255	$[0.255, 0.337]$	$[-0.129, 0.640]$	$[0.255, 0.337]$
	$(0.133, 0.377)$	$(0.133, 0.470)$	$(-0.328, 0.852)$	$(0.133, 0.470)$
UGDSAsian	1.035	$[0.893, 1.035]$	$[0.365, 1.704]$	$[0.893, 1.035]$
	$(0.727, 1.343)$	$(0.576, 1.343)$	$(0.022, 2.121)$	$(0.576, 1.343)$
UGDSnRA	0.187	$[0.187, 0.191]$	$[0.171, 0.203]$	$[0.187, 0.191]$
	$(-0.107, 0.481)$	$(-0.107, 0.485)$	$(-0.257, 0.599)$	$(-0.107, 0.485)$
Female	-0.401	$[-0.401, -0.391]$	$[-0.450, -0.353]$	$[-0.401, -0.391]$
	$(-0.536, -0.267)$	$(-0.536, -0.257)$	$(-0.622, -0.192)$	$(-0.536, -0.257)$
Dependent	-0.570	$[-0.570, -0.519]$	$[-0.811, -0.329]$	$[-0.570, -0.519]$
	$(-0.655, -0.485)$	$(-0.655, -0.428)$	$(-0.927, -0.213)$	$(-0.655, -0.428)$
ParEdPctPS	-0.442	$[-0.749, -0.442]$	$[-1.895, 1.011]$	$[-0.749, -0.442]$
	$(-0.601, -0.283)$	$(-0.972, -0.283)$	$(-2.134, 1.246)$	$(-0.972, -0.283)$
log(FamInc)	0.307	$[0.287, 0.307]$	$[0.212, 0.401]$	$[0.287, 0.307]$
	$(0.226, 0.387)$	$(0.203, 0.387)$	$(0.091, 0.518)$	$(0.203, 0.387)$
log(CostT4)	0.055	$[-0.014, 0.055]$	$[-0.272, 0.383]$	$[-0.014, 0.055]$
	$(-0.018, 0.129)$	$(-0.095, 0.129)$	$(-0.376, 0.482)$	$(-0.095, 0.129)$
log(NPT4)	-0.048	$[-0.048, -0.018]$	$[-0.189, 0.093]$	$[-0.048, -0.018]$
	$(-0.094, -0.003)$	$(-0.094, 0.028)$	$(-0.266, 0.155)$	$(-0.094, 0.028)$
PctFLoan	0.235	$[0.235, 0.266]$	$[0.088, 0.382]$	$[0.235, 0.266]$
	$(0.120, 0.350)$	$(0.120, 0.383)$	$(-0.067, 0.538)$	$(0.120, 0.383)$
PctPell	-0.699	$[-0.699, -0.680]$	$[-0.792, -0.606]$	$[-0.699, -0.680]$
	$(-0.870, -0.529)$	$(-0.870, -0.502)$	$(-1.029, -0.353)$	$(-0.870, -0.502)$
log(GDebtMdn)	-0.107	$[-0.107, -0.078]$	$[-0.244, 0.031]$	$[-0.107, -0.078]$
	$(-0.168, -0.046)$	$(-0.168, -0.017)$	$(-0.343, 0.109)$	$(-0.168, -0.017)$

Y is log(MnEarnWnEp6), W is SATAvg, and U is scholastic ability. In addition to the variables listed above, X contains 8 region indicators and 11 locale indicators as well as the shares of the remaining available race categories and indicators for whether the university has a special mission, is a women-only college, or has a religious affiliation. 95% confidence regions are reported in parentheses. $\hat{r}_{Y^*,W^*} = 0.2070$ with 99.99% CR $(0.0958, 0.3131)$.

Table 4: The Returns to College Selectivity and Characteristics Given Major Choice ($\kappa = 1$)

	$\hat{R}_{Y.(W,X')'}$	$\hat{\mathcal{S}}_\theta^{1,c}$	$\hat{\mathcal{S}}_\theta^{1}$	$\hat{\mathcal{S}}_\theta^{1,2^+}$
$10^{-4}\times$SATAvg	4.790	0	$[-\infty, \infty]$	$[0, 4.790]$
	$(3.487, 6.094)$	$(0, 0)$	$(-\infty, \infty)$	$(0, 6.094)$
$10^{-4}\times$U (ability)		$[4.790, 9.580]$	$[-\infty, \infty]$	$[0, 9.580]$
		$(3.486, 12.19)$	$(-\infty, \infty)$	$(0, 12.19)$
$10^{-2}\times$ControlInd	-2.085	$[-2.737, -2.085]$	$[-4.816, 0.646]$	$[-2.737, -2.085]$
	$(-5.978, 1.808)$	$(-6.817, 1.810)$	$(-10.41, 5.920)$	$(-6.817, 1.810)$
$10^{-2}\times$HDeg	0.145	$[0.145, 0.251]$	$[-0.299, 0.589]$	$[0.145, 0.251]$
	$(-1.721, 2.011)$	$(-1.722, 2.185)$	$(-2.990, 3.342)$	$(-1.722, 2.185)$
log(UGDS)	0.008	$[-0.001, 0.008]$	$[-0.031, 0.048]$	$[-0.001, 0.008]$
	$(-0.003, 0.020)$	$(-0.014, 0.020)$	$(-0.047, 0.064)$	$(-0.014, 0.020)$
UGDSBlack	0.170	$[0.170, 0.226]$	$[-0.062, 0.402]$	$[0.170, 0.226]$
	$(0.103, 0.237)$	$(0.103, 0.3)$	$(-0.157, 0.5)$	$(0.103, 0.3)$
UGDSHisp	0.216	$[0.216, 0.297]$	$[-0.123, 0.556]$	$[0.216, 0.297]$
	$(0.094, 0.338)$	$(0.094, 0.434)$	$(-0.273, 0.742)$	$(0.094, 0.434)$
UGDSAsian	0.765	$[0.687, 0.765]$	$[0.438, 1.092]$	$[0.687, 0.765]$
	$(0.540, 0.990)$	$(0.462, 0.990)$	$(0.191, 1.391)$	$(0.462, 0.990)$
UGDSnRA	0.137	$[0.137, 0.140]$	$[0.126, 0.148]$	$[0.137, 0.140]$
	$(-0.083, 0.357)$	$(-0.083, 0.357)$	$(-0.205, 0.457)$	$(-0.083, 0.357)$
Female	-0.202	$[-0.229, -0.202]$	$[-0.315, -0.089]$	$[-0.229, -0.202]$
	$(-0.306, -0.098)$	$(-0.337, -0.098)$	$(-0.462, 0.064)$	$(-0.337, -0.098)$
Dependent	-0.453	$[-0.453, -0.401]$	$[-0.670, -0.235]$	$[-0.453, -0.401]$
	$(-0.528, -0.378)$	$(-0.528, -0.322)$	$(-0.781, -0.134)$	$(-0.528, -0.322)$
ParEdPctPS	-0.044	$[-0.272, -0.044]$	$[-1, 0.912]$	$[-0.272, -0.044]$
	$(-0.177, 0.089)$	$(-0.437, 0.089)$	$(-1.197, 1.092)$	$(-0.437, 0.089)$
log(FamInc)	0.293	$[0.268, 0.293]$	$[0.186, 0.401]$	$[0.268, 0.293]$
	$(0.225, 0.362)$	$(0.194, 0.362)$	$(0.088, 0.490)$	$(0.194, 0.362)$
log(CostT4)	0.093	$[0.054, 0.093]$	$[-0.070, 0.255]$	$[0.054, 0.093]$
	$(0.037, 0.149)$	$(-0.005, 0.149)$	$(-0.153, 0.332)$	$(-0.005, 0.149)$
log(NPT4)	-0.026	$[-0.026, -0.008]$	$[-0.098, 0.046]$	$[-0.026, -0.008]$
	$(-0.062, 0.011)$	$(-0.062, 0.028)$	$(-0.155, 0.097)$	$(-0.062, 0.028)$
PctFLoan	0.072	$[0.072, 0.098]$	$[-0.034, 0.178]$	$[0.072, 0.098]$
	$(-0.009, 0.154)$	$(-0.009, 0.180)$	$(-0.150, 0.286)$	$(-0.009, 0.180)$
PctPell	-0.306	$[-0.307, -0.306]$	$[-0.309, -0.303]$	$[-0.307, -0.306]$
	$(-0.438, -0.174)$	$(-0.445, -0.168)$	$(-0.502, -0.117)$	$(-0.445, -0.168)$
log(GDebtMdn)	-0.144	$[-0.144, -0.123]$	$[-0.233, -0.056]$	$[-0.144, -0.123]$
	$(-0.198, -0.091)$	$(-0.198, -0.070)$	$(-0.307, 0.009)$	$(-0.198, -0.070)$
PCIP13 (Educ)	-0.226	$[-0.226, -0.159]$	$[-0.508, 0.055]$	$[-0.226, -0.159]$
	$(-0.383, -0.069)$	$(-0.383, 0.005)$	$(-0.743, 0.288)$	$(-0.383, 0.005)$
PCIP14 (Eng)	0.284	$[0.272, 0.284]$	$[0.231, 0.337]$	$[0.272, 0.284]$
	$(0.102, 0.467)$	$(0.085, 0.467)$	$(-0.009, 0.578)$	$(0.085, 0.467)$

The results use the specification in Table 3 and augment X with the (nonzero in the sample) percentage of degrees awarded in each field of study according to the Classification of Instructional Programs (we leave out Social Science, PCIP45). 95% confidence regions are reported in parentheses. Table 7 in the Appendix reports the estimates for the CIP fields of study coefficients. $\hat{r}_{Y^*,W^*} = 0.2323$ with 99.99% CR $(0.1220, 0.3369)$.

Table 5: The Return to Selectivity Given Major, Expenditures, and Completion ($\kappa = 1$)

	$\hat{R}_{Y.(W,X')'}$	$\hat{\mathcal{S}}_{\theta}^{1,c}$	$\hat{\mathcal{S}}_{\theta}^{1}$	$\hat{\mathcal{S}}_{\theta}^{1,2^+}$
$10^{-4}\times$SATAvg	2.750	0	$[-\infty,\infty]$	$[0, 2.750]$
	$(1.317, 4.183)$	$(0, 0)$	$(-\infty,\infty)$	$(0, 4.183)$
$10^{-4}\times$U (ability)		$[2.750, 5.5]$	$[-\infty,\infty]$	$[0, 5.5]$
		$(1.316, 8.367)$	$(-\infty,\infty)$	$(0, 8.367)$
$10^{-2}\times$ControlInd	-0.231	$[-0.505, -0.231]$	$[-2.395, 1.934]$	$[-0.505, -0.231]$
	$(-4.004, 3.543)$	$(-4.348, 3.544)$	$(-7.848, 7.256)$	$(-4.348, 3.544)$
$10^{-2}\times$HDeg	0.264	$[0.264, 0.365]$	$[-0.532, 1.061]$	$[0.264, 0.365]$
	$(-1.507, 2.036)$	$(-1.508, 2.167)$	$(-3.095, 3.768)$	$(-1.508, 2.167)$
log(UGDS)	0.003	$[0.001, 0.003]$	$[-0.013, 0.018]$	$[0.001, 0.003]$
	$(-0.009, 0.014)$	$(-0.011, 0.014)$	$(-0.028, 0.035)$	$(-0.011, 0.014)$
UGDSBlack	0.140	$[0.140, 0.169]$	$[-0.084, 0.364]$	$[0.140, 0.169]$
	$(0.077, 0.204)$	$(0.077, 0.238)$	$(-0.177, 0.458)$	$(0.077, 0.238)$
UGDSHisp	0.212	$[0.212, 0.249]$	$[-0.074, 0.499]$	$[0.212, 0.249]$
	$(0.105, 0.320)$	$(0.105, 0.367)$	$(-0.206, 0.677)$	$(0.105, 0.367)$
UGDSAsian	0.682	$[0.664, 0.682]$	$[0.541, 0.824]$	$[0.664, 0.682]$
	$(0.457, 0.908)$	$(0.441, 0.908)$	$(0.304, 1.118)$	$(0.441, 0.908)$
UGDSnRA	0.070	$[0.070, 0.071]$	$[0.059, 0.081]$	$[0.070, 0.071]$
	$(-0.133, 0.273)$	$(-0.133, 0.273)$	$(-0.225, 0.342)$	$(-0.133, 0.273)$
Female	-0.226	$[-0.226, -0.226]$	$[-0.229, -0.222]$	$[-0.226, -0.226]$
	$(-0.329, -0.123)$	$(-0.329, -0.123)$	$(-0.382, -0.062)$	$(-0.329, -0.123)$
Dependent	-0.457	$[-0.457, -0.426]$	$[-0.698, -0.215]$	$[-0.457, -0.426]$
	$(-0.530, -0.384)$	$(-0.530, -0.348)$	$(-0.804, -0.118)$	$(-0.530, -0.348)$
ParEdPctPS	-0.115	$[-0.207, -0.115]$	$[-0.842, 0.612]$	$[-0.207, -0.115]$
	$(-0.241, 0.012)$	$(-0.354, 0.012)$	$(-1.038, 0.783)$	$(-0.354, 0.012)$
log(FamInc)	0.273	$[0.269, 0.273]$	$[0.248, 0.297]$	$[0.269, 0.273]$
	$(0.207, 0.338)$	$(0.202, 0.338)$	$(0.157, 0.382)$	$(0.202, 0.338)$
log(CostT4)	0.038	$[0.027, 0.038]$	$[-0.046, 0.123]$	$[0.027, 0.038]$
	$(-0.016, 0.093)$	$(-0.027, 0.093)$	$(-0.127, 0.2)$	$(-0.027, 0.093)$
log(NPT4)	-0.007	$[-0.007, -0.003]$	$[-0.036, 0.022]$	$[-0.007, -0.003]$
	$(-0.042, 0.029)$	$(-0.042, 0.032)$	$(-0.092, 0.070)$	$(-0.042, 0.032)$
PctFLoan	0.064	$[0.064, 0.079]$	$[-0.059, 0.187]$	$[0.064, 0.079]$
	$(-0.015, 0.143)$	$(-0.015, 0.160)$	$(-0.168, 0.292)$	$(-0.015, 0.160)$
PctPell	-0.272	$[-0.278, -0.272]$	$[-0.321, -0.223]$	$[-0.278, -0.272]$
	$(-0.397, -0.147)$	$(-0.406, -0.147)$	$(-0.503, -0.052)$	$(-0.406, -0.147)$
log(GDebtMdn)	-0.129	$[-0.129, -0.123]$	$[-0.180, -0.078]$	$[-0.129, -0.123]$
	$(-0.178, -0.080)$	$(-0.178, -0.073)$	$(-0.244, -0.014)$	$(-0.178, -0.073)$
log(InExpFTE)	0.075	$[0.070, 0.075]$	$[0.035, 0.116]$	$[0.070, 0.075]$
	$(0.053, 0.098)$	$(0.047, 0.098)$	$(0.002, 0.151)$	$(0.047, 0.098)$
C150_4	0.185	$[0.112, 0.185]$	$[-0.390, 0.760]$	$[0.112, 0.185]$
	$(0.103, 0.268)$	$(0.008, 0.268)$	$(-0.506, 0.878)$	$(0.008, 0.268)$

The results use the specification in Table 4 and augment X with log(InExpFTE) and C1504. 95% confidence regions are reported in parentheses. Table 8 in the Appendix reports the estimates for the CIP fields of study coefficients. $\hat{r}_{Y^*,W^*} = 0.1257$ with 99.99% CR $(0.0124, 0.2358)$.

Table 6: Field of Study According to the Classification of Instructional Programs

PCIPxx	CIP field of study
PCIP01	Agriculture, Agriculture Operations, and Related Sciences
PCIP03	Natural Resources and Conservation
PCIP04	Architecture and Related Services
PCIP05	Area, Ethnic, Cultural, Gender, and Group Studies
PCIP09	Communication, Journalism, and Related Programs
PCIP10	Communications Technologies/Technicians and Support Services
PCIP11	Computer and Information Sciences and Support Services
PCIP12	Personal and Culinary Services
PCIP13	Education
PCIP14	Engineering
PCIP15	Engineering Technologies and Engineering-Related Fields
PCIP16	Foreign Languages, Literatures, and Linguistics
PCIP19	Family and Consumer Sciences/Human Sciences
PCIP22	Legal Professions and Studies
PCIP23	English Language and Literature/Letters
PCIP24	Liberal Arts and Sciences, General Studies and Humanities
PCIP25	Library Science
PCIP26	Biological and Biomedical Sciences
PCIP27	Mathematics and Statistics
PCIP30	Multi/Interdisciplinary Studies
PCIP31	Parks, Recreation, Leisure, and Fitness Studies
PCIP38	Philosophy and Religious Studies
PCIP39	Theology and Religious Vocations
PCIP40	Physical Sciences
PCIP41	Science Technologies/Technicians
PCIP42	Psychology
PCIP43	Homeland Security, Law Enforcement, Firefighting and Related Protective Services
PCIP44	Public Administration and Social Service Professions
PCIP45	Social Sciences
PCIP46	Construction Trades
PCIP47	Mechanic and Repair Technologies/Technicians
PCIP48	Precision Production
PCIP49	Transportation and Materials Moving
PCIP50	Visual and Performing Arts
PCIP51	Health Professions and Related Programs
PCIP52	Business, Management, Marketing, and Related Support Services
PCIP54	History

Table 7: The Returns to College Selectivity and Characteristics Given Major Choice ($\kappa = 1$)

	$\hat{R}_{Y.(W,X')'}$	$CR_{0.95}(R_{Y.(W,X')'})$	$\hat{\mathcal{S}}_{\theta}^{1,2^+}$	$CR_{0.95}^{\theta}(\mathcal{S}_{\theta}^{1,2^+})$
PCIP01	0.002	$(-0.292, 0.296)$	$[0.002, 0.099]$	$(-0.292, 0.403)$
PCIP03	-0.835	$(-1.226, -0.443)$	$[-0.835, -0.758]$	$(-1.226, -0.360)$
PCIP04	-0.334	$(-0.727, 0.059)$	$[-0.334, -0.264]$	$(-0.727, 0.1)$
PCIP05	-0.933	$(-1.897, 0.031)$	$[-1.147, -0.933]$	$(-2.127, 0.031)$
PCIP09	-0.132	$(-0.357, 0.092)$	$[-0.132, -0.037]$	$(-0.357, 0.201)$
PCIP10	0.031	$(-0.589, 0.650)$	$[0.031, 0.147]$	$(-0.589, 0.817)$
PCIP11	-0.217	$(-0.511, 0.077)$	$[-0.217, -0.214]$	$(-0.518, 0.091)$
PCIP12	-0.137	$(-1.719, 1.444)$	$[-0.137, 0.052]$	$(-1.720, 1.445)$
PCIP13	-0.226	$(-0.383, -0.069)$	$[-0.226, -0.159]$	$(-0.383, 0.005)$
PCIP14	0.284	$(0.102, 0.467)$	$[0.272, 0.284]$	$(0.085, 0.467)$
PCIP15	0.138	$(-0.128, 0.403)$	$[0.138, 0.241]$	$(-0.128, 0.516)$
PCIP16	-0.612	$(-1.063, -0.162)$	$[-0.612, -0.471]$	$(-1.063, 0.016)$
PCIP19	-0.153	$(-0.382, 0.076)$	$[-0.153, -0.050]$	$(-0.382, 0.194)$
PCIP22	-0.252	$(-0.942, 0.438)$	$[-0.309, -0.252]$	$(-1.036, 0.439)$
PCIP23	-0.794	$(-1.152, -0.435)$	$[-0.794, -0.763]$	$(-1.152, -0.380)$
PCIP24	-0.260	$(-0.425, -0.095)$	$[-0.260, -0.196]$	$(-0.425, -0.027)$
PCIP25	2.530	$(-1.063, 6.124)$	$[2.530, 3.743]$	$(-1.065, 7.630)$
PCIP26	-0.007	$(-0.251, 0.237)$	$[-0.007, 0.031]$	$(-0.251, 0.278)$
PCIP27	0.073	$(-0.615, 0.762)$	$[-0.232, 0.073]$	$(-1, 0.762)$
PCIP30	-0.149	$(-0.359, 0.061)$	$[-0.149, -0.058]$	$(-0.359, 0.160)$
PCIP31	-0.283	$(-0.508, -0.058)$	$[-0.283, -0.160]$	$(-0.508, 0.073)$
PCIP38	-0.410	$(-0.669, -0.152)$	$[-0.410, -0.372]$	$(-0.669, -0.105)$
PCIP39	-0.257	$(-0.419, -0.096)$	$[-0.257, -0.203]$	$(-0.419, -0.039)$
PCIP40	0.018	$(-0.886, 0.923)$	$[-0.148, 0.018]$	$(-1.111, 0.923)$
PCIP41	-1.952	$(-4.769, 0.864)$	$[-1.952, -1.322]$	$(-4.770, 1.660)$
PCIP42	-0.343	$(-0.561, -0.125)$	$[-0.343, -0.265]$	$(-0.561, -0.050)$
PCIP43	-0.117	$(-0.292, 0.057)$	$[-0.117, -0.007]$	$(-0.292, 0.180)$
PCIP44	0.039	$(-0.187, 0.265)$	$[0.039, 0.212]$	$(-0.187, 0.457)$
PCIP46	0.239	$(-2.019, 2.497)$	$[0.007, 0.239]$	$(-2.175, 2.498)$
PCIP47	-1.569	$(-3.633, 0.495)$	$[-1.579, -1.569]$	$(-3.634, 0.496)$
PCIP48	0.302	$(-1.697, 2.301)$	$[0.252, 0.302]$	$(-1.748, 2.302)$
PCIP49	0.727	$(0.257, 1.197)$	$[0.727, 0.836]$	$(0.257, 1.292)$
PCIP50	-0.436	$(-0.599, -0.273)$	$[-0.436, -0.368]$	$(-0.599, -0.202)$
PCIP51	0.297	$(0.134, 0.459)$	$[0.297, 0.376]$	$(0.134, 0.543)$
PCIP52	0.046	$(-0.111, 0.204)$	$[0.046, 0.122]$	$(-0.111, 0.282)$
PCIP54	0.055	$(-0.371, 0.481)$	$[0.055, 0.122]$	$(-0.371, 0.555)$

This table reports the estimates for the coefficients on the CIP fields of study under the specification in Table 4.

Table 8: The Return to Selectivity Given Major, Expenditures, and Completion ($\kappa = 1$)

	$\hat{R}_{Y.(W,X')'}$	$CR_{0.95}(R_{Y.(W,X')'})$	$\hat{\mathcal{S}}_{\theta}^{1,2^+}$	$CR_{0.95}^{\theta}(\mathcal{S}_{\theta}^{1,2^+})$
PCIP01	0.006	$(-0.275, 0.287)$	$[0.006, 0.051]$	$(-0.275, 0.337)$
PCIP03	-0.806	$(-1.190, -0.421)$	$[-0.806, -0.764]$	$(-1.190, -0.373)$
PCIP04	-0.310	$(-0.637, 0.018)$	$[-0.310, -0.286]$	$(-0.638, 0.035)$
PCIP05	-1.123	$(-2.080, -0.165)$	$[-1.203, -1.123]$	$(-2.156, -0.165)$
PCIP09	-0.132	$(-0.350, 0.086)$	$[-0.132, -0.094]$	$(-0.350, 0.125)$
PCIP10	0.019	$(-0.578, 0.615)$	$[0.019, 0.072]$	$(-0.578, 0.690)$
PCIP11	-0.145	$(-0.420, 0.129)$	$[-0.151, -0.145]$	$(-0.429, 0.129)$
PCIP12	-0.019	$(-1.554, 1.516)$	$[-0.019, 0.030]$	$(-1.555, 1.517)$
PCIP13	-0.185	$(-0.333, -0.037)$	$[-0.185, -0.161]$	$(-0.333, -0.012)$
PCIP14	0.312	$(0.139, 0.485)$	$[0.303, 0.312]$	$(0.129, 0.485)$
PCIP15	0.167	$(-0.092, 0.426)$	$[0.167, 0.206]$	$(-0.092, 0.469)$
PCIP16	-0.709	$(-1.114, -0.305)$	$[-0.709, -0.634]$	$(-1.114, -0.210)$
PCIP19	-0.119	$(-0.351, 0.114)$	$[-0.119, -0.078]$	$(-0.351, 0.155)$
PCIP22	-0.027	$(-0.737, 0.683)$	$[-0.080, -0.027]$	$(-0.798, 0.684)$
PCIP23	-0.670	$(-1.009, -0.330)$	$[-0.670, -0.670]$	$(-1.018, -0.322)$
PCIP24	-0.170	$(-0.336, -0.004)$	$[-0.170, -0.159]$	$(-0.337, 0.004)$
PCIP25	1.030	$(-2.747, 4.806)$	$[1.030, 1.637]$	$(-2.749, 5.350)$
PCIP26	0.047	$(-0.190, 0.283)$	$[0.047, 0.054]$	$(-0.190, 0.293)$
PCIP27	-0.153	$(-0.791, 0.485)$	$[-0.279, -0.153]$	$(-0.949, 0.486)$
PCIP30	-0.098	$(-0.301, 0.106)$	$[-0.098, -0.064]$	$(-0.302, 0.142)$
PCIP31	-0.225	$(-0.443, -0.007)$	$[-0.225, -0.174]$	$(-0.443, 0.045)$
PCIP38	-0.359	$(-0.580, -0.138)$	$[-0.359, -0.339]$	$(-0.580, -0.109)$
PCIP39	-0.218	$(-0.373, -0.062)$	$[-0.218, -0.194]$	$(-0.373, -0.040)$
PCIP40	0.049	$(-0.829, 0.928)$	$[-0.041, 0.049]$	$(-0.961, 0.929)$
PCIP41	-1.254	$(-3.751, 1.243)$	$[-1.254, -1.181]$	$(-3.756, 1.395)$
PCIP42	-0.270	$(-0.480, -0.060)$	$[-0.270, -0.245]$	$(-0.480, -0.039)$
PCIP43	-0.059	$(-0.228, 0.110)$	$[-0.059, -0.019]$	$(-0.228, 0.153)$
PCIP44	0.036	$(-0.187, 0.259)$	$[0.036, 0.114]$	$(-0.187, 0.348)$
PCIP46	-0.369	$(-2.384, 1.645)$	$[-0.376, -0.369]$	$(-2.416, 1.664)$
PCIP47	-1.374	$(-3.363, 0.614)$	$[-1.428, -1.374]$	$(-3.415, 0.615)$
PCIP48	0.407	$(-1.741, 2.554)$	$[0.361, 0.407]$	$(-1.744, 2.555)$
PCIP49	0.652	$(0.204, 1.099)$	$[0.652, 0.721]$	$(0.204, 1.167)$
PCIP50	-0.405	$(-0.559, -0.251)$	$[-0.405, -0.378]$	$(-0.559, -0.225)$
PCIP51	0.319	$(0.162, 0.477)$	$[0.319, 0.351]$	$(0.161, 0.509)$
PCIP52	0.122	$(-0.027, 0.271)$	$[0.122, 0.149]$	$(-0.027, 0.297)$
PCIP54	0.117	$(-0.295, 0.529)$	$[0.117, 0.131]$	$(-0.295, 0.543)$

This table reports the estimates for the coefficients on the CIP fields of study under the specification in Table 5.

Table 9: Variables Definition, Corresponding College Scorecard Variables, and Data Files

Variable Definition	CS Variable	CS Variable Definition	CS Datafile
Data using the pooled cohorts that enrolled in award years 2006-2007 and 2007-2008			
MnEarnWnEP6	mn_earn_wne_p6	Mean earnings (in 2015 USD) of students working and not enrolled 6 years after entry	2012-13
Gt25KP6	GT_25K_P6	Share of non-enrolled students earning over $25,000/year (in 2015 USD) 6 years after entry	2012-13
Female	FEMALE	Share of female students	2007-08
Dependent	DEPENDENT	Share of dependent students	2007-08
FamInc	FAMINC	Average family income (in 2015 USD)	2007-08
ParEdPctPS	PAR_ED_PCT_PS	Percent of students whose parents' highest educational level is some form of postsecondary education	2007-08
Data based on the cohort that enrolled in fall 2007 or academic year 2007-2008			
SATAvg	SAT_AVG	Average SAT equivalent score of students admitted	2007-08
C150_4	C150_4	Completion rate within 150% of expected time to completion for full-time, first-time, degree/certificate-seeking students at 4 year institutions	2013-14
Data using the pooled cohorts that completed in award years 2009-2010 and 2010-2011			
GDebtMdn	GRAD_DEBT_MDN	Median debt for students who have completed	2010-11
Data based on Fall 2010, award or academic year 2010-11, or fiscal year 2011			
Main	MAIN	Indicator for main campus	2010-11
PredDeg	PREDDEG	Predominant degree awarded (not classified: 0, certificate: 1, associate: 2, bachelor's: 3, : entirely graduate: 4)	2011-12
ControlInd: Indicator that is 1 if CONTROL is 2	CONTROL	Control of Institution (1: public, 2: private nonprofit, 3: private for profit)	2010-11
HDeg: Indicator is 1 if HIGHDEG is 4	HIGHDEG	Highest degree awarded (non-degree: 0, certificate: 1, associate: 2, bachelor's: 3, graduate: 4)	2011-12
PctFLoan	PCTFLOAN	Percent of all federal undergraduate students receiving a federal student loan	2011-12
PctPell	PCTPELL	Percentage of undergraduates who receive a Pell Grant	2011-12
UGDS	UGDS	Enrollment of undergraduate certificate/degree-seeking students	2010-11

UGDSBlack, UGDSHisp, UGDSAsian, UGDSnRA, UGDSWhite, UGDSAIAN, UGDSNHPI, UGDS2More, UGDSUnkn	UGDS_BLACK, UGDS_HISP, UGDS_ASIAN, UGDS_NRA, UGDS_WHITE, UGDS_AIAN, UGDS_NHPI, UGDS_2MOR, UGDS_UNKN	Total share of enrollment of undergraduate degree-seeking students who are Black, Hispanic, Asian, Non-Resident Aliens, White, American Indian/Alaska Native, Native Hawaiian/Pacific Islander, 2 or more races, or whose race is unknown	2010-11
PCIP## (38 variables)	PCIP##	Percentage of degrees awarded in one of 38 fields of study according to the Classification of Instructional Programs (CIP)	2011-12
InExpFTE	INEXPFTE	Instructional expenditures per full-time equivalent student (includes gradudate students)	2011-12
Data based on the cohort that enrolled in academic year 2010-11			
CostT4 =COSTT4_A +COSTT4_P	COSTT4_A/ COSTT4_P	Average cost of attendance for academic/program-year institutions	2011-12
NPT4 =NPT4_PUB +NPT4_PRIV	NPT4_PUB/ NPT4_PRIV	Average net price for Title IV institutions for public/private institutions	2011-12
Data based on academic year 2015-2016 or 2009-2010			
Region# (10 region indicators)	REGION	Region (0: U.S. Service Schools, 1: New England, 2: Mid East, 3: Great Lakes, 4: Plains, 5: Southeast, 6: Southwest, 7: Rocky Mountains, 8: Far West, 9: Outlying Areas)	2010-11
Locale## (12 locale indicators)	LOCALE	Locale of the institution (11-13: city large/midsize/small, 21-23: suburb large/midsize/small, 31-33: town fringe/distant/remote, 41-43: rural fringe/distant/remote)	2014-15
RelAffilInd (Indicator that is -1 if not reported, 0 if not applicable, and 1 otherwise)	RelAffil	Religous affiliation of the institution (65 values: either not reported, not applicable, or one of 63 listed religious affiliations)	2014-15
Data based on fiscal or calender year 2016			
SpecMis: Indicator that is 1 if any of the source variables is 1	HBCU, PBI, ANNHI, TRIBAL, AANAPII, HSI, NANTI	Flags for institutions that are/serve historically Black, predominantly Black, Alaska/Hawaiian native, Tribal, Asian American/Native American/Pacific Islander, Hispanic, or Native American non-Tribal	2014-15
WomenOnly	WOMENONLY	Flag for women-only college	2014-15

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.

Bruni, F. (2013), "How to Choose a College," *The New York Times*, January 5, 2013.

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

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

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.

DiTraglia, F. and C. Garcia-Jimeno (2017), "A Framework for Eliciting, Incorporating, and Disciplining Identification Beliefs in Linear Models," University of Pennsylvania Working Paper.

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.

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.

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

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.

Kline, B. and E. Tamer (2015), "Bayesian Inference in a Class of Partially Identified Models," *Quantitative Economics*, forthcoming.

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.

O'Connell, M. (2007). "How to Choose a College that's Right for You," *National Public Radio*, February 21, 2007.

Prior, A. (2014), "What College Can Teach the Aspiring Entrepreneur: What to Study, Where to Study - and Other Ways to Use Your Education to Be Ready to Start a Business," *The Wall Street Journal*, November 3, 2014.

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

Thompson, D. (2014), "Which College – and Which Major – Will Make you Richest," *The Atlantic*, March 26, 2014.

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

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.