Identification of Average Random Coefficients under Magnitude and Sign Restrictions on Confounding

Karim Chalak*†
Boston College

December 4, 2012

Abstract

This paper studies measuring the average effects $\bar{\beta}$ of X on Y in a structural system with random coefficients and confounding. We do not require (conditionally) exogenous regressors or instruments. Using proxies W for the confounders U, we ask how do the average direct effects of U on Y compare in magnitude and sign to those of U on W. Exogeneity and equi- or proportional confounding are limit cases yielding full identification. Alternatively, the elements of $\bar{\beta}$ are partially identified in a sharp bounded interval if W is sufficiently sensitive to U, and sharp upper or lower bounds may obtain otherwise. We extend this analysis to accommodate conditioning on covariates and a semiparametric separable specification as well as a panel structure and proxies included in the Y equation. After studying estimation and inference, we apply this method to study the financial return to education and the black-white wage gap.

Keywords: *causality, confounding, endogeneity, omitted variable, partial identification, proxy.*

1 Introduction

This paper studies identifying and estimating average causal effects in a structural system with random coefficients under restrictions on the magnitude and sign of confounding. To

*Karim Chalak is Assistant Professor, Department of Economics, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA 02467, USA. Email: chalak@bc.edu.

†Acknowledgments: The author thanks the participants in the Northwestern Junior Festival of Recent Developments in Microeconometrics, the Harvard Causal Inference Seminar, the Research in Econometrics Workshop at Boston College, and the California Econometrics Conference 2012 as well as Andrew Beauchamp, Stéphane Bonhomme, Donald Cox, Stefan Hoderlein, and Arthur Lewbel for helpful comments. I thank Rossella Calvi, Daniel Kim, and Tao Yang for excellent research assistance. Any errors are the author's responsibility.

illustrate the paper's main ideas, consider a Mincer (1974) earning structural equation, frequently employed in empirical work (see e.g. discussion in Card, 1999), given by

$$Y = \alpha_y + X'\bar{\beta} + U\bar{\delta}_y,$$

where Y denotes the logarithm of hourly wage, X denotes observed determinants of wage including years of education, and the scalar U, commonly referred to as "ability," denotes unobserved skill. To introduce the main ideas in their simplest form, we let U be scalar and consider constant slope coefficients $\bar{\beta}$ and $\bar{\delta}_y$ here but allow for a random intercept α_y which may be correlated with X. Because U is freely associated with X and may cause Y, we say that U is a "confounder" and X is "endogenous." Our object of interest here is $\bar{\beta}$, the vector of (average) direct effects of the elements of X on Y. Let W be a proxy for U that is possibly error-laden and given by

$$W = \alpha_w + U\bar{\delta}_w,$$

where, for now, we consider a constant slope coefficient $\bar{\delta}_w$ and random intercept α_w which may be correlated with U. For example, W may denote the logarithm of a test score commonly used as a proxy for "ability," such IQ (Intelligence Quotient) or KWW (Knowledge of the World of Work). This parsimonious specification facilitates comparing the slope coefficients on U in the Y and W equations while maintaining the commonly used log-level specification for the wage equation. In particular, $100\bar{\delta}_y$ and $100\bar{\delta}_w$ denote respectively the (average) approximate percentage changes in wage and test score due directly to a one unit increase in U. Alternatively, one could consider standardized variables. Last, let Z be a vector of potential instruments that are uncorrelated with α_y and α_w but freely correlated with U and therefore invalid. Let Z and X have the same dimension; for example, Z may equal X. Also, let $\tilde{Z} = Z - E(Z)$. We then have

$$E(\tilde{Z}Y) = E(\tilde{Z}X')\bar{\beta} + E(\tilde{Z}U)\bar{\delta}_y \quad \text{and} \quad E(\tilde{Z}W) = E(\tilde{Z}U)\bar{\delta}_w.$$

Provided $E(\tilde{Z}X')$ is nonsingular, we obtain

$$\bar{\beta} = E(\tilde{Z}X')^{-1}E(\tilde{Z}Y) - E(\tilde{Z}X')^{-1}E(\tilde{Z}W)\frac{\bar{\delta}_y}{\bar{\delta}_w}.$$

This expression for $\bar{\beta}$ involves two instrumental variables (IV) regression estimands $E(\tilde{Z}X')^{-1}$ $E(\tilde{Z}Y)$ and $E(\tilde{Z}X')^{-1}E(\tilde{Z}W)$. It also involves the unknown $\frac{\bar{\delta}_y}{\bar{\delta}_w}$ denoting the ratio of the (average) direct effect of U on Y to that of U on W. Importantly, the IV regression omitted variable bias (or inconsistency) $E(\tilde{Z}X')^{-1}E(\tilde{Z}W)\frac{\bar{\delta}_y}{\bar{\delta}_w}$ in measuring $\bar{\beta}$ is known up to this ratio. As we show, a similar expression for the average effects $\bar{\beta}$ of X on Y obtains in the case of random slope coefficients under suitable assumptions discussed below.

We ask the following questions:

1. How does the average direct effect of U on Y compare in magnitude to that of U on W?

2. How does the average direct effect of U on Y compare in sign to that of U on W?

Sometimes, economic theory and evidence can provide guidance to answering these questions. For example, Cawley, Heckman, and Vytlacil (2001) find that the fraction of wage variance explained by measures of cognitive ability is modest and that personality traits are correlated with earnings primarily through schooling attainment. Also, when ability is not revealed to employers, they may statistically discriminate based on observables such as education (see e.g. Altonji and Pierret (2001) and Arcidiacono, Bayer, and Hizmo (2010)). As such, in the context of the earning equation illustrative example, it may be reasonable to assume at least for certain subpopulations that, given the observables, wage is on average less sensitive or elastic to unobserved ability than the test score is. In particular, a one unit change in U may, on average, directly cause a higher percentage change in the test score than in wage. Second, ability may, on average, directly affect wage and the test score in the same direction.

The answers to these questions impose restrictions on the magnitude and sign of confounding which fully or partially identify the average effects $\bar{\beta}$ of X on Y. In particular, exogeneity is a limiting case, which obtains if the average direct effect $\bar{\delta}_y$ of U on Y is zero, yielding full (point) identification. Equiconfounding (or proportional confounding) is another limiting case, in which the average direct effect of U on Y equals (a known proportion of) that of U on W, also yielding full identification (see Chalak, 2012). Alternatively, restrictions on how the average direct effect of U on Y compares in magnitude and/or sign to that of U on W partially identify elements of $\bar{\beta}$, yielding sharp bounded intervals when the proxy W is sufficiently sensitive to the confounder U, and sharp lower or upper bounds otherwise.

This paper develops the analysis illustrated in this example and extends it to allow for random slope coefficients, multiple confounders, and a semiparametric separable specification as well as a panel structure and proxies included in the Y equation. It derives the sharp identification regions for the direct effects of X on Y in these cases under magnitude and sign restrictions on confounding. After discussing estimation and inference, the paper applies its results to study the return to education and the black-white wage gap. Using the data in Card (1995), we employ restrictions on confounding to partially identify in a sharp bounded interval the average financial return to education as well as the average black-white wage gap for given levels of unobserved ability and observables including education. We find that regression estimates provide an upper bound on the average return to education and black-white wage gap. We also find nonlinearity in the return to education, with the 12^{th}, 16^{th}, and 18^{th} years, corresponding to obtaining a high school, college, and possibly a graduate degree, yielding a high average return.

This paper's method provides a simple alternative to the common practice which informally assumes that conditioning on proxies ensures conditional exogeneity. In particular, even when α_w is independent of U, conditioning on W does not ensure that the coefficient on X from a regression of Y on $(1, X', W')'$ identifies $\bar{\beta}$. Conditioning on W may nevertheless attenuate the regression bias (see e.g. Ogburna and VanderWeele, 2012). This paper's method also provides a practical alternative to IV methods when there are fewer (conditionally) exogenous instruments than elements of X, as may be the case when allowing for year-specific incremental return to education.

The assumption that Y is, on average, less directly responsive to U than W is, underlying partial identification in a bounded interval, is a weakening of exogeneity[1]. Several recent papers employ alternative assumptions to partially identify constant coefficients associated with endogenous variables in a linear equation. For example, Altonji, Conley, Elder, and

[1] We let U denote the unobservables thought to be freely correlated with Z (conditional on the covariates) with the unobserved drivers of Y that are (conditionally) uncorrelated with Z absorbed into α_y. Recall that Z may equal X.

Taber (2011) obtain partial identification by assuming that the selection on unobservables occurs similarly to that on observables. Also, Reinhold and Woutersen (2009) and Nevo and Rosen (2012) partially identify constant coefficients in a linear equation under the assumptions that the instruments are less correlated with U than the endogenous variables are, and that these correlations are in the same direction. Bontemps, Magnac, and Maurin (2012) provide additional examples and a general treatment of set identified linear models. In a nonparametric setting, Manski and Pepper (2000) bound the average treatment effect under the assumption that the mean potential outcome varies monotonically with the instrument. This paper employs proxies to (partially) identify average effects under magnitude and sign restrictions on confounding. In particular, it does not impose assumptions on selection on observables or on how instruments and endogenous variables relate to the confounders. Which identifying assumption is more appropriate depends on the context.

This paper is organized as follows. Section 2 defines the notation and data generation assumption. Section 3 studies identification of average random coefficients under magnitude and sign restrictions on confounding. Section 4 extends the analysis to condition on covariates and Section 5 relaxes the Y equation to admit a semiparametric separable specification. Section 6 studies estimation and constructing confidence intervals. Section 7 applies the methods to study the return to education and the black-white wage gap. Section 8 concludes. Appendix A contains extensions to a panel structure and to cases where proxies are included in the Y equation. Mathematical proofs are gathered in Appendix B.

2 Data Generation and Notation

The next assumption defines the data generating process.

Assumption 1 (S.1) *(i) Let* $V \equiv (\underset{q\times 1}{S}', \underset{\ell\times 1}{Z}', \underset{k\times 1}{X}', \underset{m\times 1}{W}', \underset{1\times 1}{Y})'$ *be a random vector with unknown distribution* $P \in \mathcal{P}$. *(ii) Let a structural system* $\mathcal{S}_1$ *generate the random coefficients* $\theta \equiv (\underset{1\times 1}{\alpha_y}, \underset{m\times 1}{\alpha_w}', \underset{p\times 1}{\delta_y}', vec(\underset{p\times m}{\delta_w})', \underset{k\times 1}{\beta}')'$, *confounders* $\underset{p\times 1}{U}$, *covariates* S, *potential instruments* Z, *proxies* W, *causes* X, *and response* Y *such that*

$$Y = \alpha_y + X'\beta + U'\delta_y \quad \text{and} \quad W' = \alpha_w' + U'\delta_w,$$

with $Cov[Z, (Y, W')'] < \infty$. *Realizations of* V *are observed; those of* θ *and* U *are not.*

S.1(i) defines the notation for the observables. S.1(ii) imposes structure on the data generating process. In particular, S.1(ii) allows for random intercept and slope coefficients. Thus, for each individual i in a sample, we have

$$Y_i = \alpha_{y,i} + X_i'\beta_i + U_i'\delta_{y,i} \quad \text{and} \quad W_i' = \alpha_{w,i}' + U_i'\delta_{w,i}.$$

We suppress the index i when referring to population variables. The vector $U = (U_1, ..., U_p)'$ denotes unobserved confounders of X and Y. We observe realizations of a vector W of proxies for U. We generally allow for W and X to have common elements. We may also observe realizations of a vector of covariates S; otherwise we set $S = 1$. Last, we also observe realizations of a vector of potential instruments Z possibly equal to, or containing elements of, X. Importantly, we do not restrict the statistical dependence between Z or X and U, either unconditionally or conditional on S. Thus, elements of Z need not be valid instruments since these may be included in the Y equation and are freely (conditionally) correlated with U. Implicit in S.1(ii) is that X does not cause (structurally drive) the corresponding random slope coefficients β. Similarly, U does not cause δ_y and δ_w. As such, β is the vector of random direct effects of the elements of X on Y (i.e. holding fixed all random coefficients and variables other than β and the element of X intervened on). Similarly, δ_y is the vector of random direct effects of the elements of U on Y and δ_w are the random (direct) effects of U on W. Here, we're interested in measuring the average effects $\bar{\beta} \equiv E(\beta)$ of X on Y.

2.1 IV Regression Notation

For a generic $d \times 1$ random vector A we write:

$$\bar{A} \equiv E(A) \quad \text{and} \quad \tilde{A} \equiv A - \bar{A}.$$

For example, we write $\bar{\beta} \equiv E(\beta)$ and $\bar{\delta}_y \equiv E(\delta_y)$. Further, for generic random vectors B and C of equal dimension, we let

$$R_{a.b|c} \equiv E(CB')^{-1}E(CA') \quad \text{and} \quad \epsilon_{a.b|c}' \equiv A' - B'R_{a.b|c}$$

denote the IV regression estimand and residual respectively, so that by construction $E(C\epsilon_{a.b|c}') = 0$. For example, for $k = \ell$, we write $R_{\tilde{y}.\tilde{x}|\tilde{z}} = E(\tilde{Z}\tilde{X}')^{-1}E(\tilde{Z}\tilde{Y})$. Thus, $R_{\tilde{y}.\tilde{x}|\tilde{z}}$ is the slope coefficient associated with X in an IV regression of Y on $(1, X')'$ using instruments $(1, Z')'$. In the special case where $B = C$, we obtain the regression coefficients and residuals:

$$R_{a.b} \equiv R_{a.b|b} = E(BB')^{-1}E(BA') \quad \text{and} \quad \epsilon'_{a.b} \equiv \epsilon'_{a.b|b} = A' - B'R_{a.b}.$$

3 Identification Using Proxies for Confounders

We begin by characterizing $\bar{\beta}$ and studying conditions for full identification. Section 3.2 studies partial identification of elements of $\bar{\beta}$ under sign and magnitude restrictions on confounding.

3.1 Characterization and Full Identification

For illustration, consider the example from the Introduction with scalar proxy W and confounder U and constant slope coefficients:

$$Y = \alpha_y + X'\bar{\beta} + U\bar{\delta}_y \quad \text{and} \quad W = \alpha_w + U\bar{\delta}_w.$$

Using the IV regression succinct notation, recall from the Introduction that, provided $Cov(Z, (\alpha_y, \alpha_w)') = 0$ and $E(\tilde{Z}\tilde{X}')$ is nonsingular, we obtain:

$$\bar{\beta} = R_{\tilde{y}.\tilde{x}|\tilde{z}} - R_{\tilde{w}.\tilde{x}|\tilde{z}}\frac{\bar{\delta}_y}{\bar{\delta}_w}.$$

In particular, the IV regression (omitted variable) bias $R_{\tilde{w}.\tilde{x}|\tilde{z}}\frac{\bar{\delta}_y}{\bar{\delta}_w}$ depends on the ratio $\frac{\bar{\delta}_y}{\bar{\delta}_w}$ of the direct effect of U on the response Y to that of U on the proxy W. The next Theorem extends this result to allow for vectors U and W and for random slope coefficients. We set $S = 1$ in this Section to simplify the exposition; Section 4 explicitly conditions on covariates.

Theorem 3.1 *Assume S.1 with $S = 1$, $\ell = k$, $m = p$, and that*
(i) $E(\tilde{Z}\tilde{X}')$ and $\bar{\delta}_w$ are nonsingular,
(ii) $Cov(\alpha_y, Z) = 0$, $E(\tilde{\beta}|X, Z) = 0$, and $E(\tilde{\delta}_y|U, Z) = 0$,
(iii) $Cov(\alpha_w, Z) = 0$ and $E(\tilde{\delta}_w|U, Z) = 0$.

Let $\bar{\delta} \equiv \bar{\delta}_w^{-1}\bar{\delta}_y$ then

$$\bar{\beta} = R_{\tilde{y}.\tilde{x}|\tilde{z}} - R_{\tilde{w}.\tilde{x}|\tilde{z}}\bar{\delta}.$$

$B \equiv R_{\tilde{y}.\tilde{x}|\tilde{z}} - \bar{\beta} = R_{\tilde{w}.\tilde{x}|\tilde{z}}\bar{\delta}$ denotes the IV regression bias (or inconsistency) in measuring $\bar{\beta}$. When $Z = X$, Theorem 3.1 gives

$$\bar{\beta} = R_{\tilde{y}.\tilde{x}} - R_{\tilde{w}.\tilde{x}}\bar{\delta}.$$

Next, we discuss the conditions in Theorem 3.1. Condition (i) requires $\ell = k$ and $m = p$, with $E(\tilde{Z}\tilde{X}')$ and $\bar{\delta}_w$ nonsingular. More generally, if $\ell \geq k$ and $m \geq p$, and provided $E(\tilde{Z}\tilde{X}')$ and $\bar{\delta}_w$ are full rank, we can use $Q_{\tilde{z},\tilde{x}} \equiv E(\tilde{X}'\tilde{Z})P_z E(\tilde{Z}\tilde{X}')$, $Q_{\tilde{z},\tilde{y}} \equiv E(\tilde{X}'\tilde{Z})P_z E(\tilde{Z}\tilde{Y}')$, and $Q_{\tilde{z},\tilde{w}} \equiv E(\tilde{X}'\tilde{Z})P_z E(\tilde{Z}\tilde{W}')$ as well as $\bar{\delta}_w P_w \bar{\delta}_w'$ for some positive definite weighting matrices $\underset{\ell\times\ell}{P_z}$ and $\underset{m\times m}{P_w}$ to obtain

$$\bar{\beta} = Q_{\tilde{z},\tilde{x}}^{-1}Q_{\tilde{z},\tilde{y}} - Q_{\tilde{z},\tilde{x}}^{-1}Q_{\tilde{z},\tilde{w}}(P_w\bar{\delta}_w'(\bar{\delta}_w P_w \bar{\delta}_w')^{-1}\bar{\delta}_y).$$

We forgo this added generality in what follows to focus on restrictions on confounding with straightforward economic interpretation.

Conditions (ii) and (iii) are implied by the respectively stronger assumptions that the coefficients θ are mean independent of $(U, Z', X')'$ or constant. In particular, condition (ii) imposes assumptions on the random coefficients in the Y equation. It requires that Z is uncorrelated with α_y, β is mean independent[2] of (X, Z), and δ_y is mean independent of (U, Z). Roughly speaking, (ii) isolates U as the source of the difficulty in identifying $\bar{\beta}$. Had U been observed with $\ell = k+p$ and $E(\tilde{Z}(\tilde{X}', \tilde{U}'))$ nonsingular, (ii) would permit identifying the average slope coefficients via IV regression. While condition (ii) does not directly restrict the joint distribution of (β, U), S.1(ii) and the requirement that $E(\tilde{\beta}|X, Z) = 0$ in (ii) can restrict how β relates to U, e.g. in models where the return to education can depend on ability[3] (see e.g. Card 1999). In these cases, one can consider IV methods for the correlated random coefficient model, e.g. Wooldridge (1997, 2003) and Heckman and Vytlacil (1998). Similarly, condition (ii) can restrict how $\tilde{\delta}_y$ relates to Z (and X), e.g. in learning models where the return to ability can vary with experience and depend on

[2] We employ the unnecessary mean independence assumptions in conditions (ii) and (iii) of Theorem 3.1 because of their simple interpretation. However, zero covariances $E[\tilde{Z}X'\tilde{\beta}] = 0$, $E[\tilde{Z}U'\tilde{\delta}_y] = 0$, and $E[\tilde{Z}U'\tilde{\delta}_w] = 0$ suffice.

[3] One can consider extending this analysis to study specifications which allow for interaction terms involving observables and unobserved confounders. To keep a manageable scope of the paper, we leave studying (sign and magnitude) restrictions on confounding in nonseparable structural equations to other work.

educational attainment (e.g. Altonji and Pierret (2001) and Arcidiacono, Bayer, and Hizmo (2010)). Importantly, however, the conditions in Theorem 3.1 do not restrict the joint distribution of $(U, Z', X')'$ other than requiring that $E(\tilde{Z}\tilde{X}')$ is nonsingular. In particular, Z and X can be freely correlated with U and thus endogenous.

Condition (iii) restricts the random coefficients in the W equation. It requires Z to be uncorrelated with α_w, and the elements of δ_w to be mean independent of (U, Z). Thus, condition (iii) relates $Cov(Z, W)$ to $Cov(Z, U)$ without restricting the dependence between α_w and U, allowing W to be an error-laden proxy for U. Note that, even when δ_w is constant and α_w is independent of U, conditioning on W does not ensure that the coefficient on X from a regression of Y on $(1, X', W')'$ identifies $\bar{\beta}$.

To illustrate the consequences of Theorem 3.1, consider the example from the Introduction with scalar U and W. Observe that $R_{\tilde{y}.\tilde{x}|\tilde{z}}$ fully (point) identifies $\bar{\beta}$ under exogeneity. In this case, the IV regression bias disappears either because U does not determine Y, and in particular $\bar{\delta}_y = 0$, or because Z and U are uncorrelated, and thus $R_{\tilde{w}.\tilde{x}|\tilde{z}} = 0$. Alternatively, shape restrictions on the effects of U on Y and W can fully identify $\bar{\beta}$. This occurs for instance under signed proportional confounding where the sign of the ratio $\frac{\bar{\delta}_y}{\bar{\delta}_w}$ of the average direct effect of U on Y to that of U on W is known and its magnitude equals a known constant c. Under equiconfounding, $c = 1$, and U directly affects Y and W equally on average. In particular, $\bar{\beta}$ is fully identified under positive ($\bar{\delta}_y = \bar{\delta}_w$) or negative ($\bar{\delta}_y = -\bar{\delta}_w$) equiconfounding by $R_{\tilde{y}-\tilde{w}.\tilde{x}|\tilde{z}} = \bar{\beta}$ or $R_{\tilde{y}+\tilde{w}.\tilde{x}|\tilde{z}} = \bar{\beta}$ respectively.

More generally, U may be a vector of potential confounders. Often, to each U_h corresponds one proxy $W_h = \alpha_{w_h} + U_h\delta_{w_h}$ so that $W' = \alpha_w' + U'\delta_w$ with $\delta_w = diag(\delta_{w_1}, ..., \delta_{w_m})$. In this case,

$$\bar{\beta} = R_{\tilde{y}.\tilde{x}|\tilde{z}} - R_{\tilde{w}.\tilde{x}|\tilde{z}}\bar{\delta} = R_{\tilde{y}.\tilde{x}|\tilde{z}} - \textstyle\sum_{h=1}^{m} \frac{\bar{\delta}_{y_h}}{\bar{\delta}_{w_h}} R_{\tilde{w}_h.\tilde{x}|\tilde{z}}.$$

As before, under exogeneity $\bar{\beta} = R_{\tilde{y}.\tilde{x}|\tilde{z}}$ and, for example, under positive equiconfounding $\frac{\bar{\delta}_{y_h}}{\bar{\delta}_{w_h}} = 1$ for $h = 1, ..., m$ and $\bar{\beta} = R_{\tilde{y}.\tilde{x}|\tilde{z}} - \sum_{h=1}^{m} R_{\tilde{w}_h.\tilde{x}|\tilde{z}}$.

The next corollary extends these full identification results to the general case, with δ_w unrestricted and $\left|\bar{\delta}\right|$ equal to a known vector c of constants. However, it is useful throughout to keep in mind the leading one-to-one case where δ_w is a diagonal matrix with straightforward interpretation. We use subscripts to denote vector elements. For example,

$\bar{\beta}_j$ and $R_{\tilde{y}.\tilde{x}|\tilde{z},j}$ denote the j^{th} element of $\bar{\beta}$ and $R_{\tilde{y}.\tilde{x}|\tilde{z}}$ respectively, and $\bar{\delta}_h$ and c_h the h^{th} elements of $\bar{\delta}$ and c respectively.

Corollary 3.2 *Assume the conditions of Theorem 3.1 and let $j = 1, ..., k$. (i) If $B_j = 0$ (exogeneity) then $\bar{\beta}_j = R_{\tilde{y}.\tilde{x}|\tilde{z},j}$. (ii) If $\left|\bar{\delta}\right| = c$ (proportional confounding) then $\bar{\beta}_j = R_{\tilde{y}.\tilde{x}|\tilde{z},j} - \sum_{h=1}^{m} sign(\bar{\delta}_h) c_h R_{\tilde{w}_h.\tilde{x}|\tilde{z},j}$.*

Thus, it suffices for exogeneity that $\bar{\delta}_y = 0$ or $R_{\tilde{w}.\tilde{x}|\tilde{z}} = 0$. Further, signed proportional confounding with known c_h and $sign(\bar{\delta}_h), h = 1, ..., m$, point identifies $\bar{\beta}$.

3.2 Partial Identification

In the absence of conditions leading to full identification, magnitude and sign restrictions on the average direct effects of U on Y and W partially identify the elements of $\bar{\beta}$. To illustrate, consider the example from the Introduction with scalar U and W. We ask how does the average direct effect $\bar{\delta}_y$ of U on Y compares in magnitude and sign to the average effect $\bar{\delta}_w$ of U on W. As discussed above, exogeneity ($\bar{\delta}_y = 0$) and signed equi- or proportional confounding ($\bar{\delta}_y = c\bar{\delta}_w$ or $\bar{\delta}_y = -c\bar{\delta}_w$) are limit cases securing full identification. Next, we derive sharp identification regions for the elements of $\bar{\beta}$ under weaker sign and magnitude restrictions.

In particular, suppose that $R_{\tilde{w}.\tilde{x}|\tilde{z},j} \neq 0$ and $\left|\bar{\delta}\right| = \left|\frac{\bar{\delta}_y}{\bar{\delta}_w}\right| \leq 1$ so that the magnitude of the average direct effect of U on Y is not larger than that of U on W. Here, W is, on average, at least as directly responsive to U than Y is. Then, using the expression for $\bar{\beta}$, we obtain the following identification regions for $\bar{\beta}_j$ for $j = 1, ..., k$, depending on the sign of $\bar{\delta} R_{\tilde{w}.\tilde{x}|\tilde{z},j}$:

$$\bar{\beta}_j \in \mathcal{B}_j(\left|\bar{\delta}\right| \leq 1;\ \bar{\delta} R_{\tilde{w}.\tilde{x}|\tilde{z},j} \leq 0) = [R_{\tilde{y}.\tilde{x}|\tilde{z},j},\ R_{\tilde{y}.\tilde{x}|\tilde{z},j} + \left|R_{\tilde{w}.\tilde{x}|\tilde{z},j}\right|],$$
$$\bar{\beta}_j \in \mathcal{B}_j(\left|\bar{\delta}\right| \leq 1;\ 0 \leq \bar{\delta} R_{\tilde{w}.\tilde{x}|\tilde{z},j}) = [R_{\tilde{y}.\tilde{x}|\tilde{z},j} - \left|R_{\tilde{w}.\tilde{x}|\tilde{z},j}\right|,\ R_{\tilde{y}.\tilde{x}|\tilde{z},j}].$$

For instance, if we assume $\bar{\delta} \geq 0$ so that, on average, U directly affects Y and W in the same direction, we obtain:

$$\bar{\beta}_j \in \mathcal{B}_j(\left|\bar{\delta}\right| \leq 1;\ 0 \leq \bar{\delta};\ R_{\tilde{w}.\tilde{x}|\tilde{z},j} < 0) = [R_{\tilde{y}.\tilde{x}|\tilde{z},j},\ R_{\tilde{y}-\tilde{w}.\tilde{x}|\tilde{z},j}],$$
$$\bar{\beta}_j \in \mathcal{B}_j(\left|\bar{\delta}\right| \leq 1;\ 0 \leq \bar{\delta};\ 0 < R_{\tilde{w}.\tilde{x}|\tilde{z},j}) = [R_{\tilde{y}-\tilde{w}.\tilde{x}|\tilde{z},j},\ R_{\tilde{y}.\tilde{x}|\tilde{z},j}].$$

Similar bounds involving $R_{\tilde{y}+\tilde{w}.\tilde{x}|\tilde{z},j}$ instead of $R_{\tilde{y}-\tilde{w}.\tilde{x}|\tilde{z},j}$ obtain if $\bar{\delta} \leq 0$.

Instead, if $\left|\bar{\delta}\right| = \left|\frac{\bar{\delta}_y}{\bar{\delta}_w}\right| > 1$ so that W is, on average less directly responsive to U than Y is, then we obtain the following identification regions for $\bar{\beta}_j$ for $j = 1, ..., k$, depending on the sign of $\bar{\delta} R_{\tilde{w}.\tilde{x}|\tilde{z},j}$:

$$\bar{\beta}_j \in \mathcal{B}_j(1 < \left|\bar{\delta}\right|;\ \bar{\delta} R_{\tilde{w}.\tilde{x}|\tilde{z},j} < 0) = (R_{\tilde{y}.\tilde{x}|\tilde{z},j} + \left|R_{\tilde{w}.\tilde{x}|\tilde{z},j}\right|, +\infty),$$
$$\bar{\beta}_j \in \mathcal{B}_j(1 < \left|\bar{\delta}\right|;\ 0 < \bar{\delta} R_{\tilde{w}.\tilde{x}|\tilde{z},j}) = (-\infty,\ R_{\tilde{y}.\tilde{x}|\tilde{z},j} - \left|R_{\tilde{w}.\tilde{x}|\tilde{z},j}\right|).$$

Note that these identification regions exclude the IV estimand $R_{\tilde{y}.\tilde{x}|\tilde{z},j}$.

Wider intervals obtain under either magnitude or sign (but not both) restrictions on the average direct effects $\bar{\delta}_y$ and $\bar{\delta}_w$. In particular, if $\left|\bar{\delta}_y\right| \leq \left|\bar{\delta}_w\right|$, the proxy W is on average at least as directly responsive as Y is to U, and $\bar{\beta}_j$ is partially identified as follows:

$$\bar{\beta}_j \in \mathcal{B}_j(\left|\bar{\delta}\right| \leq 1) = [R_{\tilde{y}.\tilde{x}|\tilde{z},j} - \left|R_{\tilde{w}.\tilde{x}|\tilde{z},j}\right|,\ R_{\tilde{y}.\tilde{x}|\tilde{z},j} + \left|R_{\tilde{w}.\tilde{x}|\tilde{z},j}\right|].$$

Note that $\mathcal{B}_j(\left|\bar{\delta}\right| \leq 1)$ is twice as large as $\mathcal{B}_j(\left|\bar{\delta}\right| \leq 1;\ sign(\bar{\delta} R_{\tilde{w}.\tilde{x}|\tilde{z},j}))$. Also, the "closer" Z is to exogeneity, the smaller $\left|R_{\tilde{w}.\tilde{x}|\tilde{z},j}\right|$ is, and the tighter $\mathcal{B}_j(\left|\bar{\delta}\right| \leq 1;\ sign(\bar{\delta} R_{\tilde{w}.\tilde{x}|\tilde{z},j}))$ and $\mathcal{B}_j(\left|\bar{\delta}\right| \leq 1)$ are. Alternatively, if $\left|\bar{\delta}_w\right| < \left|\bar{\delta}_y\right|$, W is, on average, less directly responsive to U than Y is, and

$$\bar{\beta}_j \in \mathcal{B}_j(1 < \left|\bar{\delta}\right|) = (-\infty,\ R_{\tilde{y}.\tilde{x}|\tilde{z},j} - \left|R_{\tilde{w}.\tilde{x}|\tilde{z},j}\right|) \bigcup (R_{\tilde{y}.\tilde{x}|\tilde{z},j} + \left|R_{\tilde{w}.\tilde{x}|\tilde{z},j}\right|,\ +\infty).$$

In this case, the "farther" Z is from exogeneity, the larger $\left|R_{\tilde{w}.\tilde{x}|\tilde{z},j}\right|$ is, and the more informative $\mathcal{B}_j(1 < \left|\bar{\delta}\right|;\ sign(\bar{\delta} R_{\tilde{w}.\tilde{x}|\tilde{z},j}))$ and $\mathcal{B}_j(1 < \left|\bar{\delta}\right|)$ are.

Alone, sign restrictions determine the direction of the IV regression omitted variable bias. In particular, if $\bar{\delta} R_{\tilde{w}.\tilde{x}|\tilde{z},j} \leq 0$ we have:

$$\bar{\beta}_j \in \mathcal{B}_j(\bar{\delta} R_{\tilde{w}.\tilde{x}|\tilde{z},j} \leq 0) = [R_{\tilde{y}.\tilde{x}|\tilde{z},j},\ +\infty).$$

If instead, $\bar{\delta} R_{\tilde{w}.\tilde{x}|\tilde{z},j} \geq 0$ then we have:

$$\bar{\beta}_j \in \mathcal{B}_j(0 \leq \bar{\delta} R_{\tilde{w}.\tilde{x}|\tilde{z},j}) = (-\infty,\ R_{\tilde{y}.\tilde{x}|\tilde{z},j}].$$

The identification regions derived above under magnitude and/or sign restrictions for scalars U and W are sharp. Thus, any point in these regions is feasible under the maintained assumptions. In particular, for each element b of $\mathcal{B}_j$ there exists $\delta_y(b)$ and $\delta_w(b)$ such that, when $\delta_y = \delta_y(b)$ and $\delta_w = \delta_w(b)$, the joint distribution (θ, U, V) satisfies the conditions of Theorem 3.1 and the restrictions on $sign(\left|\bar{\delta}\right| - 1)$ and/or $sign(\frac{\bar{\delta}_y}{\bar{\delta}_w} R_{\tilde{w}.\tilde{x}|\tilde{z},j})$ underlying $\mathcal{B}_j$ hold. For this, it suffices to let $\delta_y(b)$ and $\delta_w(b)$ be degenerate and to set $\frac{\bar{\delta}_y(b)}{\bar{\delta}_w(b)} = \frac{1}{R_{\tilde{w}.\tilde{x}|\tilde{z},j}}(R_{\tilde{y}.\tilde{x}|\tilde{z},j} - b)$ according to $\mathcal{S}_1$.

These identification regions obtain in part by asking how the average direct effects $\bar{\delta}_y$ and $\bar{\delta}_w$ compare in magnitude. If this comparison is ambiguous, a researcher may be more confident imposing an upper or lower bound c on $\left|\frac{\bar{\delta}_y}{\bar{\delta}_w}\right|$ and similar sharp identification regions derive with $c\left|R_{\tilde{w}.\tilde{x}|\tilde{z},j}\right|$ replacing $\left|R_{\tilde{w}.\tilde{x}|\tilde{z},j}\right|$, and with exogeneity and signed proportional confounding as limit cases.

More generally, suppose that U is a vector and that there is a proxy $W_h = \alpha_{w_h} + U_h \delta_{w_h}$ for each confounder U_h, $h = 1, ..., m$, so that $\delta_w = diag(\delta_{w_1}, ..., \delta_{w_m})$. Then $\bar{\beta} = R_{\tilde{y}.\tilde{x}|\tilde{z}} - \sum_{h=1}^{m} \frac{\bar{\delta}_{y_h}}{\bar{\delta}_{w_h}} R_{\tilde{w}_h.\tilde{x}|\tilde{z}}$ and magnitude and/or sign restrictions on $\bar{\delta}_h = \frac{\bar{\delta}_{y_h}}{\bar{\delta}_{w_h}}$ yield the sharp identification regions $\mathcal{B}_j(\underset{h=1,...,m}{sign}(\left|\bar{\delta}_h\right| - 1); \underset{h=1,...,m}{sign}(\bar{\delta}_h R_{\tilde{w}_h.\tilde{x}|\tilde{z},j}))$ for $\bar{\beta}_j$, $j = 1, ..., k$. The next Corollary derives sharp identification regions for a general matrix δ_w and vector c of known constants. For simplicity, we assume that $R_{\tilde{w}_h.\tilde{x}|\tilde{z},j} \neq 0$ for $h = 1, ..., m$; if $R_{\tilde{w}_h.\tilde{x}|\tilde{z},j} = 0$ it can be dropped from the expression $\bar{\beta}_j = R_{\tilde{y}.\tilde{x}|\tilde{z},j} - \sum_{h=1}^{m} R_{\tilde{w}_h.\tilde{x}|\tilde{z},j} \bar{\delta}_h$ for $\bar{\beta}_j$ and it won't impact the bounds. To facilitate the exposition in subsequent sections, we let $A \equiv R_{\tilde{w}.\tilde{x}|\tilde{z}}$ so that $A_{jh} \equiv R_{\tilde{w}_h.\tilde{x}|\tilde{z},j}$. In what follows, when $g = 0$, we omit the inequalities and sums indexed by $h = 1, ..., g$. When $\left|\bar{\delta}_h\right| \leq c_h$, $sign(\bar{\delta}_h A_{jh})$ denotes either $\bar{\delta}_h A_{jh} \leq 0$ or $\bar{\delta}_h A_{jh} \geq 0$ for $h = 1, ..., m$.

Corollary 3.3 *Let $A_{jh} \equiv R_{\tilde{w}_h.\tilde{x}|\tilde{z},j} \neq 0$ and $c_h > 0$ for $j = 1, ..., k$ and $h = 1, ..., m$. Under the conditions of Theorem 3.1, $\bar{\beta}_j \in \mathcal{B}_j(\underset{h=1,...,m}{sign}(\left|\bar{\delta}_h\right| - c_h); \underset{h=1,...,m}{sign}(\bar{\delta}_h A_{jh}))$, for $j = 1, ..., k$, defined as follows, and these bounds are sharp:*

(a) $\mathcal{B}_j(\left|\bar{\delta}_h\right| \underset{h=1,...,m}{\leq} c_h; \underset{h=1,...,m}{sign}(\bar{\delta}_h A_{jh})) =$

$$[R_{\tilde{y}.\tilde{x}|\tilde{z},j} + \textstyle\sum_{h=1}^{m} \min\{sign(-\bar{\delta}_h A_{jh}) c_h \left|A_{jh}\right|, 0\},$$

$$R_{\tilde{y}.\tilde{x}|\tilde{z},j} + \textstyle\sum_{h=1}^{m} \max\{0, sign(-\bar{\delta}_h A_{jh}) c_h \left|A_{jh}\right|\}].$$

(b) For $g \in \{0, ..., m-1\}$*:*

(b.i) $\mathcal{B}_j(\left|\bar{\delta}_h\right| \underset{h=1,...,g}{\leq} c_h;\ c_h \underset{h=g+1,...,m}{<} \left|\bar{\delta}_h\right|;\ \underset{h=1,...,g}{sign}\ (\bar{\delta}_h A_{jh});\ \bar{\delta}_h A_{jh} \underset{h=g+1,...,m}{<} 0) =$

$$(R_{\tilde{y}.\tilde{x}|\tilde{z},j} + \textstyle\sum_{h=1}^{g} \min\{sign(-\bar{\delta}_h A_{jh}) c_h \left|A_{jh}\right|, 0\} + \sum_{h=g+1}^{m} c_h \left|A_{jh}\right|,\ +\infty).$$

(b.ii) $\mathcal{B}_j(\left|\bar{\delta}_h\right| \underset{h=1,...,g}{\leq} c_h;\ c_h \underset{h=g+1,...,m}{<} \left|\bar{\delta}_h\right|;\ \underset{h=1,...,g}{sign}\ (\bar{\delta}_h A_{jh});\ 0 \underset{h=g+1,...,m}{<} \bar{\delta}_h A_{jh}) =$

$$(-\infty,\ R_{\tilde{y}.\tilde{x}|\tilde{z},j} + \textstyle\sum_{h=1}^{g} \max\{0, sign(-\bar{\delta}_h A_{jh}) c_h \left|A_{jh}\right|\} - \sum_{h=g+1}^{m} c_h \left|A_{jh}\right|).$$

(c) For $g \in \{0, ..., m-2\}$ *and* $g' \in \{g+1, ..., m-1\}$*:*

$$\mathcal{B}_j(\left|\bar{\delta}_h\right| \underset{h=1,...,g}{\leq} c_h;\ c_h \underset{h=g+1,...,m}{<} \left|\bar{\delta}_h\right|;\ \underset{h=1,...,g}{sign}\ (\bar{\delta}_h A_{jh});$$
$$\bar{\delta}_h A_{jh} \underset{h=g+1,...,g'}{<} 0;\ 0 \underset{h=g'+1,...,m}{<} \bar{\delta}_h A_{jh}) = (-\infty,\ +\infty).$$

Note that these sharp identification regions contain $R_{\tilde{y}.\tilde{x}|\tilde{z},j}$ in (a), may but need not contain $R_{\tilde{y}.\tilde{x}|\tilde{z},j}$ in (b) when $g > 0$, and do not contain $R_{\tilde{y}.\tilde{x}|\tilde{z},j}$ in (b) when $g = 0$. Sharp identification regions under either magnitude or sign restrictions (but not both) (e.g. $\mathcal{B}_j(\underset{h=1,...,m}{sign}\ (\left|\bar{\delta}_h\right| - c_h)))$ can be derived by taking the appropriate unions of the identification regions under both types of restrictions. Further, note that different instruments or proxies may lead to different identification regions for $\bar{\beta}_j$, in which case $\bar{\beta}_j$ is identified in the intersection of these regions, provided it's nonempty.

Appendix A contains extensions building on this section's results obtaining sharp bounds on average coefficients under magnitude and sign restrictions on confounding. In particular, Section A.1 studies a panel structure with individual and time varying random coefficients without requiring "fixed effects." Section A.2 studies cases where the proxies W are a component of X, included in the Y equation. While the conditions in Theorem 3.1 do not rule out that W and X have common elements, they entail restrictions on Z in this case. Section A.2.1 studies the "under-"identification case where there are fewer (possibly one) valid instruments than the dimension of X. Section A.2.2 studies the case of multiple proxies for U that are included in the Y equation and allowed to be elements of Z.

4 Conditioning on Covariates

We extend the results in Section 3 to weaken the mean independence assumptions on the random coefficients in Theorem 3.1 to their conditional counterparts given a vector of covariates S. For this, for a generic $d \times 1$ random vector A we write:

$$\bar{A}(S) \equiv E(A|S) \quad \text{and} \quad \tilde{A}(S) \equiv A - \bar{A}(S).$$

For example, $\bar{\beta}(S) \equiv E(\beta|S)$ denotes the average direct effects of X on Y for the subpopulation with covariates S. Further, for generic random vectors B and C of equal dimension, we write

$$R_{a.b|c}(S) \equiv E(CB'|S)^{-1}\ E(CA'|S) \quad \text{and} \quad \epsilon'_{a.b|c}(S) \equiv A' - B'R_{a.b|c}(S),$$

so that by construction $E(C\epsilon'_{a.b|c}|S) = 0$. For example, for $k = \ell$ we write $R_{\tilde{y}(s).\tilde{x}(s)|\tilde{z}(s)}(S) \equiv E(\tilde{Z}(S)\tilde{X}'(S)|S)^{-1}E(\tilde{Z}(S)\tilde{Y}(S)|S)$. When $B = C$, we obtain $R_{a.b}(S) \equiv R_{a.b|b}(S)$ and $\epsilon'_{a.b}(S) \equiv \epsilon'_{a.b|b}(S)$. This notation reduces to that defined in Section 2 when S is degenerate, $S = 1$; we leave S implicit in this case.

Theorem 4.1 *Assume S.1 with $\ell = k$ and $m = p$, and that*
(i) $E(\tilde{Z}(S)\tilde{X}'(S)|S)$ and $\bar{\delta}_w(S)$ are nonsingular,
(ii) $Cov(\alpha_y, Z|S) = 0$, $E(\tilde{\beta}(S)|X, Z, S) = 0$, and $E(\tilde{\delta}_y(S)|U, Z, S) = 0$,
(iii) $Cov(\alpha_w, Z|S) = 0$ and $E(\tilde{\delta}_w(S)|U, Z, S) = 0$.
Let $\bar{\delta}(S) \equiv \bar{\delta}_w^{-1}(S)\bar{\delta}_y(S)$ then

$$\bar{\beta}(S) = R_{\tilde{y}(s).\tilde{x}(s)|\tilde{z}(s)}(S) - R_{\tilde{w}(s).\tilde{x}(s)|\tilde{z}(s)}(S)\bar{\delta}(S).$$

Here, $B(S) \equiv R_{\tilde{y}(s).\tilde{x}(s)|\tilde{z}(s)}(S) - \bar{\beta}(S) = R_{\tilde{w}(s).\tilde{x}(s)|\tilde{z}(s)}(S)\bar{\delta}(S)$ denotes the conditional IV regression bias in measuring $\bar{\beta}(S)$. When $Z = X$, Theorem 4.1 gives $\bar{\beta}(S) = R_{\tilde{y}(s).\tilde{x}(s)}(S) - R_{\tilde{w}(s).\tilde{x}(s)}(S)\bar{\delta}(S)$.

The conditional on covariates S conditions (ii) and (iii) in Theorem 4.1 weaken their unconditional uncorrelation and mean independence analogs in Theorem 3.1. Further, conditioning on S may render Z "closer" to exogeneity and the conditional IV regression bias $B(S)$ smaller. Theorem 4.1 reduces to Theorem 3.1 when S is degenerate, $S = 1$.

If the conditional average effects $\bar{\beta}(S)$, $\bar{\delta}_y(S)$, and $\bar{\delta}_w(S)$ are constant, so that the mean independence assumptions for the slope coefficients hold unconditionally, the law of iterated expectations gives

$$\bar{\beta} = E(\tilde{Z}(S)\tilde{X}'(S))^{-1}E(\tilde{Z}(S)\tilde{Y}(S)) - E(\tilde{Z}(S)\tilde{X}'(S))^{-1}E(\tilde{Z}(S)\tilde{W}(S))\bar{\delta}.$$

This expression involves two estimands from IV regressions of $\tilde{Y}(S)$ and $\tilde{W}(S)$ respectively on $\tilde{X}(S)$ using instrument $\tilde{Z}(S)$. Further, if the conditional expectations $\bar{Z}(S)$, $\bar{X}(S)$, $\bar{W}(S)$, and $\bar{Y}(S)$ are affine functions of S, we obtain

$$\bar{\beta} = E(\epsilon_{z.(1,s')'}\epsilon'_{x.(1,s')'})^{-1}E(\epsilon_{z.(1,s')'}\epsilon'_{y.(1,s')'}) - E(\epsilon_{z.(1,s')'}\epsilon'_{x.(1,s')'})^{-1}E(\epsilon_{z.(1,s')'}\epsilon'_{w.(1,s')'})\bar{\delta}.$$

Using partitioned regressions (Frisch and Waugh, 1933), the two residual-based IV estimands in the $\bar{\beta}$ expression can be recovered from $R_{\tilde{y}.(\tilde{x}',\tilde{s}')'|(\tilde{z}',\tilde{s}')'}$ and $R_{\tilde{w}.(\tilde{x}',\tilde{s}')'|(\tilde{z}',\tilde{s}')'}$ as the coefficients associated with $\tilde{X}$ (or as the coefficients on X in IV regressions of Y and W respectively on $(1, X', S')'$ using instruments $(1, Z', S')'$). Aside from these special cases, an expression for $\bar{\beta}$ derives by integrating the expression for $\bar{\beta}(S)$ in Theorem 4.1 over the distribution of S.

Thus, an element $\bar{\beta}_j(S)$ of $\bar{\beta}(S)$ is fully identified under conditional exogeneity ($B_j(S) = 0$) or conditional signed equi- or proportional confounding ($\left|\bar{\delta}(S)\right| = c(S)$, a vector of functions of S, with $sign(\bar{\delta}_h(S))$ and $c_h(S)$, $h = 1, ..., m$, known or estimable). Otherwise, the expression for $\bar{\beta}(S)$ can be used, along with sign and magnitude restrictions on elements of $\bar{\delta}(S) \equiv \bar{\delta}_w^{-1}(S)\bar{\delta}_y(S)$ involving the conditional average direct effects of U on Y and W given S, to partially identify $\bar{\beta}_j(S)$. This yields the sharp identification regions $\mathcal{B}_j(\underset{h=1,...,m}{sign}(\left|\bar{\delta}_h(S)\right| - c_h(S)); \underset{h=1,...,m}{sign}(\bar{\delta}_h(S)A_{jh}(S)))$ for $\bar{\beta}_j(S)$ defined as in Corollary 3.3 with $R_{\tilde{y}(s).\tilde{x}(s)|\tilde{z}(s)}(S)$ and $A(S) \equiv R_{\tilde{w}(s).\tilde{x}(s)|\tilde{z}(s)}(S)$ replacing $R_{\tilde{y}.\tilde{x}|\tilde{z}}$ and A respectively.

5 Semiparametric Separable Specification

The next assumption extends S.1 to allow for a semiparametric separable specification of the Y equation.

Assumption 2 (S.2) *(i) Let* $V \equiv (\underset{q\times 1}{S}', \underset{k\times 1}{X}', \underset{m\times 1}{W}', \underset{1\times 1}{Y})'$ *be a random vector with unknown distribution* $P \in \mathcal{P}$. *(ii) Let a structural system* $\mathcal{S}_2$ *generate the random coefficients* $\theta \equiv$

$(\underset{1\times 1}{\alpha_y}, \underset{m\times 1}{\alpha_w}', vec(\underset{p\times m}{\delta_w})')'$, *vector* U_y *of countable dimension, confounders* $\underset{p\times 1}{U}$, *covariates* S, *proxies* W, *causes* X, *and response* Y *such that*

$$Y = r(X, U_y) + U'\delta_y \quad \text{and} \quad W' = \alpha_w' + U'\delta_w,$$

with r *an unknown real-valued measurable function and* $E(Y, W')' < \infty$. *Realizations of* V *are observed whereas those of* θ, U_y, *and* U *are not.*

Similar to S.1(ii), S.2(ii) implicitly assumes that X does not cause U_y and that U does not cause δ_y and δ_w. Thus, for $x, x^* \in S_X$, the support of X, the conditional average direct effect of X on Y at (x, x^*) given S is $\bar{\beta}(x, x^*|S) \equiv E[r(x^*, U_y) - r(x, U_y)|S]$. The next Theorem gives conditions under which $\bar{\beta}(x, x^*|S)$ depends on the unknown $\bar{\delta}(S) \equiv \bar{\delta}_w^{-1}(S)\bar{\delta}_y(S)$ involving the conditional average direct effects of U on Y and W. We use $\perp$ to denote independence as in Dawid (1979).

Theorem 5.1 *Assume S.2 with* $m = p$ *and that*
(i) $\bar{\delta}_w(S)$ *is nonsingular,*
(ii) $U_y \perp X|S$ *and* $E(\tilde{\delta}_y(S)|U, X, S) = 0$,
(iii) $E(\tilde{\alpha}_w(S)|X, S) = 0$ *and* $E(\tilde{\delta}_w(S)|U, X, S) = 0$.
Let $\bar{\delta}(S) \equiv \bar{\delta}_w^{-1}(S)\bar{\delta}_y(S)$. *Then for* $x, x^* \in S_X$, *the support of* X,

$$\bar{\beta}(x, x^*|S) = E(Y|X = x^*, S) - E(Y|X = x, S) - [E(W'|X = x^*, S) - E(W'|X = x, S)]\bar{\delta}(S).$$

Letting $A(x, x^*|S) \equiv E(W'|X = x^*, S) - E(W'|X = x, S)$, the conditional nonparametric regression bias for $\bar{\beta}(x, x^*|S)$ is given by

$$B(x, x^*|S) = A(x, x^*|S)\bar{\delta}(S).$$

As in the linear case, $\bar{\beta}(x, x^*|S)$ is fully identified under conditional exogeneity or signed proportional confounding. In particular, if $B(x, x^*|S) = 0$ (conditional exogeneity) then $\bar{\beta}(x, x^*|S) = E(Y|X = x^*, S) - E(Y|X = x, S)$. Alternatively, if $\left|\bar{\delta}(S)\right| = c(S)$ with $c_h(S)$ and $sign(\bar{\delta}_h(S))$, $h = 1, ..., m$, known or estimable (conditional signed proportional confounding) then

$$\bar{\beta}(x, x^*|S) = E(Y|X = x^*, S) - E(Y|X = x, S) - \textstyle\sum_{h=1}^{m} sign(\bar{\delta}_h(S))c_h(S)A_h(x, x^*|S).$$

Otherwise, restrictions on the magnitude and sign of confounding partially identify $\bar{\beta}(x, x^*|S)$. In particular, $\bar{\beta}(x, x^*|S) \in \mathcal{B}(\underset{h=1,\ldots,m}{sign}(|\bar{\delta}_h(S)| - c_h(S)); \underset{h=1,\ldots,m}{sign}(\bar{\delta}_h(S)A_h(S)))$, where these sharp identification regions are defined analogously to Corollary 3.3 with $E(Y|X = x^*, S) - E(Y|X = x, S)$ and $A(x, x^*|S)$ replacing $R_{\tilde{y}.\tilde{x}|\tilde{z}}$ and A respectively.

6 Estimation and Inference

6.1 Asymptotic Normality

We obtain a consistent set estimator $\hat{\mathcal{B}}_j$ for an identification region $\mathcal{B}_j$ by using consistent estimators for the bounds. We focus on the results in Section 3 with scalar U and W, $m = p = 1$, as considered in the empirical application. The case of a vector of proxies can be derived analogously. In the scalar confounder case with $S = 1$, the bounds in $\mathcal{B}_j(sign(|\bar{\delta}| - 1);\ sign(\bar{\delta}R_{\tilde{w}.\tilde{x}|\tilde{z},j}))$ and $\mathcal{B}_j(sign(|\bar{\delta}| - 1))$ are $R_{\tilde{y}.\tilde{x}|\tilde{z}}$, $R_{\tilde{y}-\tilde{w}.\tilde{x}|\tilde{z}}$, or $R_{\tilde{y}+\tilde{w}.\tilde{x}|\tilde{z}}$. Since these are linear transformations of $(R'_{\tilde{y}.\tilde{x}|\tilde{z}}, R'_{\tilde{w}.\tilde{x}|\tilde{z}})'$ we derive the joint distribution of plug-in estimators $(\hat{R}'_{\tilde{y}.\tilde{x}|\tilde{z}}, \hat{R}'_{\tilde{w}.\tilde{x}|\tilde{z}})'$ for $(R'_{\tilde{y}.\tilde{x}|\tilde{z}}, R'_{\tilde{w}.\tilde{x}|\tilde{z}})'$. We note that these results also encompass conditioning on S under the conditions in Theorem 4.1 when $\bar{\beta}(S)$, $\bar{\delta}_y(S)$, and $\bar{\delta}_w(S)$ are constant and $\bar{Z}(S)$, $\bar{X}(S)$, $\bar{Y}(S)$, and $\bar{W}(S)$ are affine functions of S since the bounds in this case are linear transformations of the coefficients associated with $\tilde{X}$ in $R_{\tilde{y}.(\tilde{x}',\tilde{s}')'|(\tilde{z}',\tilde{s}')'}$ and $R_{\tilde{w}.(\tilde{x}',\tilde{s}')'|(\tilde{z}',\tilde{s}')'}$ as discussed in Section 4.1. For notational simplicity, we leave S implicit in what follows.

We stack the observations $\{A_i\}_{i=1}^n$ of a generic $d \times 1$ vector A into the $n \times d$ matrix $\mathbf{A}$. Also, we let $\tilde{A}_i \equiv A_i - \frac{1}{n}\sum_{i=1}^n A_i$. Further, for generic observations $\{A_i, B_i, C_i\}_{i=1}^n$ corresponding to A and the random vectors B and C of equal dimension, we let

$$\hat{R}_{a.b|c} \equiv (\mathbf{C}'\mathbf{B})^{-1}(\mathbf{C}'\mathbf{A}) = (\frac{1}{n}\sum_{i=1}^n C_i B_i')^{-1}(\frac{1}{n}\sum_{i=1}^n C_i A_i')$$

denote the IV regression estimator. We continue to denote the IV residuals by $\epsilon_{a.b|c}$ and we denote the sample residuals by $\hat{\epsilon}'_{a.b|c,i} \equiv A_i' - B_i'\hat{R}_{a.b|c}$.

The next theorem derives the asymptotic distribution of $\sqrt{n}((\hat{R}'_{\tilde{y}.\tilde{x}|\tilde{z}}, \hat{R}'_{\tilde{w}.\tilde{x}|\tilde{z}})' - (R'_{\tilde{y}.\tilde{x}|\tilde{z}}, R'_{\tilde{w}.\tilde{x}|\tilde{z}})')$, using standard arguments. For this, we let $Q \equiv diag(E(\tilde{Z}\tilde{X}'), E(\tilde{Z}\tilde{X}'))$.

Theorem 6.1 *Assume S.1(i) with* $m = 1$, $\ell = k$, $E[\tilde{Z}(\tilde{W}', \tilde{X}', \tilde{Y})]$ *finite and* $E(\tilde{Z}\tilde{X}')$ *non-singular uniformly in* $P \in \mathcal{P}$. *Suppose further that*

(i) $\frac{1}{n}\sum_{i=1}^{n} \tilde{Z}_i \tilde{X}_i' \xrightarrow{p} E(\tilde{Z}\tilde{X}')$ *uniformly in* $P \in \mathcal{P}$; *and*

(ii) $n^{-1/2}\sum_{i=1}^{n}(\tilde{Z}_i'\epsilon_{\tilde{y}.\tilde{x}|\tilde{z},i}, \tilde{Z}_i'\epsilon_{\tilde{w}.\tilde{x}|\tilde{z},i})' \xrightarrow{d} N(0, \Lambda)$ *uniformly in* $P \in \mathcal{P}$, *where*

$$\Lambda = \begin{bmatrix} E(\tilde{Z}\epsilon^2_{\tilde{y}.\tilde{x}|\tilde{z}}\tilde{Z}') & E(\tilde{Z}\epsilon_{\tilde{y}.\tilde{x}|\tilde{z}}\epsilon_{\tilde{w}.\tilde{x}|\tilde{z}}\tilde{Z}') \\ E(\tilde{Z}\epsilon_{\tilde{w}.\tilde{x}|\tilde{z}}\epsilon_{\tilde{y}.\tilde{x}|\tilde{z}}\tilde{Z}') & E(\tilde{Z}\epsilon^2_{\tilde{w}.\tilde{x}|\tilde{z}}\tilde{Z}') \end{bmatrix}$$

is finite and positive definite uniformly in $P \in \mathcal{P}$.

Then, uniformly in $P \in \mathcal{P}$,

$$\sqrt{n}((\hat{R}'_{\tilde{y}.\tilde{x}|\tilde{z}}, \hat{R}'_{\tilde{w}.\tilde{x}|\tilde{z}})' - (R'_{\tilde{y}.\tilde{x}|\tilde{z}}, R'_{\tilde{w}.\tilde{x}|\tilde{z}})') \xrightarrow{d} N(0, Q^{-1}\Lambda Q'^{-1}).$$

We refer the reader to e.g. Shorack (2000) and Imbens and Manski (2004, lemma 5) for primitive conditions ensuring the uniform law of large numbers and central limit theorem in assumptions (i, ii) of Theorem 6.1. Collect the IV estimands in $R \equiv (R'_{\tilde{y}.\tilde{x}|\tilde{z}}, R'_{\tilde{w}.\tilde{x}|\tilde{z}}, R'_{\tilde{y}-\tilde{w}.\tilde{x}|\tilde{z}}, R'_{\tilde{y}+\tilde{w}.\tilde{x}|\tilde{z}})'$ and their corresponding plug-in estimators in $\hat{R}$. Last, it is convenient to let $(\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4) \equiv (\epsilon_{\tilde{y}.\tilde{x}|\tilde{z}}, \epsilon_{\tilde{w}.\tilde{x}|\tilde{z}}, \epsilon_{\tilde{y}-\tilde{w}.\tilde{x}|\tilde{z}}, \epsilon_{\tilde{y}+\tilde{w}.\tilde{x}|\tilde{z}})$. Since $\hat{R}$ is a linear transformation of $(\hat{R}'_{\tilde{y}.\tilde{x}|\tilde{z}}, \hat{R}'_{\tilde{w}.\tilde{x}|\tilde{z}})'$, it follows from Theorem 6.1 that, uniformly in $P \in \mathcal{P}$,

$$\sqrt{n}(\hat{R} - R) \xrightarrow{d} N(0, \Sigma),$$

where $\underset{4k\times 4k}{\Sigma}$ is finite and positive definite uniformly in $P \in \mathcal{P}$, with 16 $k \times k$ blocks,

$$\sigma^2_{gh} \equiv E(\tilde{Z}\tilde{X}')^{-1}E(\tilde{Z}\epsilon_g\epsilon_h\tilde{Z}')E(\tilde{X}\tilde{Z}')^{-1} \text{ for } g, h = 1, 2, 3, 4.$$

We obtain uniformly in $P \in \mathcal{P}$ consistent estimators $\hat{\mathcal{B}}_j(sign(|\bar{\delta}| - 1);\ sign(\bar{\delta}\hat{R}_{\tilde{w}.\tilde{x}|\tilde{z}}))$ and $\hat{\mathcal{B}}_j(sign(|\bar{\delta}| - 1))$ for $\mathcal{B}_j(sign(|\bar{\delta}| - 1);\ sign(\bar{\delta}R_{\tilde{w}.\tilde{x}|\tilde{z},j}))$ and $\mathcal{B}_j(sign(|\bar{\delta}| - 1))$ using the appropriate estimators $\hat{R}_{\tilde{y}.\tilde{x}|\tilde{z}}$, $\hat{R}_{\tilde{y}-\tilde{w}.\tilde{x}|\tilde{z}}$, or $\hat{R}_{\tilde{y}+\tilde{w}.\tilde{x}|\tilde{z}}$ for the bounds. Under regularity conditions (e.g. White, 1980, 2001), a uniformly in $P \in \mathcal{P}$ consistent heteroskedasticity robust estimator for the block σ^2_{gh} of the asymptotic covariance is given by

$$\hat{\sigma}^2_{gh} \equiv (\frac{1}{n}\tilde{\mathbf{Z}}'\tilde{\mathbf{X}})^{-1}(\frac{1}{n}\sum_{i=1}^{n}\tilde{Z}_i\hat{\epsilon}_{g,i}\hat{\epsilon}_{h,i}\tilde{Z}_i')(\frac{1}{n}\tilde{\mathbf{X}}'\tilde{\mathbf{Z}})^{-1}.$$

For example, we estimate $\sigma^2_{\tilde{y}.\tilde{x}|\tilde{z}} \equiv Avar(\sqrt{n}(\hat{R}_{\tilde{y}.\tilde{x}|\tilde{z}} - R_{\tilde{y}.\tilde{x}|\tilde{z}}))$ using $\hat{\sigma}^2_{\tilde{y}.\tilde{x}|\tilde{z}} \equiv (\frac{1}{n}\tilde{\mathbf{Z}}'\tilde{\mathbf{X}})^{-1}$ $(\frac{1}{n}\sum_{i=1}^{n}\tilde{Z}_i\hat{\epsilon}^2_{\tilde{y}.\tilde{x}|\tilde{z},i}\tilde{Z}_i')(\frac{1}{n}\tilde{\mathbf{X}}'\tilde{\mathbf{Z}})^{-1}$.

6.2 Confidence Intervals

This subsection discusses constructing a $1-\alpha$ confidence interval (CI) for $\bar{\beta}_j$ that is partially identified in $\mathcal{B}_j(|\bar{\delta}| \leq 1)$ or $\mathcal{B}_j(|\bar{\delta}| \leq 1;\ 0 \leq \bar{\delta};\ sign(R_{\tilde{w}.\tilde{x}|\tilde{z},j}))$ with scalar confounder and proxy as considered in the empirical application. These sharp regions are of the form $[\beta_{l,j}, \beta_{u,j}]$, a bounded interval of finite width. Let $\hat{\beta}_{l,j}$ and $\hat{\beta}_{u,j}$ denote estimators for $\beta_{l,j}$ and $\beta_{u,j}$. By Theorem 6.1, $(\hat{\beta}_{l,j}, \hat{\beta}_{u,j})'$ is asymptotically normally distributed uniformly in $P \in \mathcal{P}$. Denote by $\hat{\sigma}^2_{l,j}$ the uniformly in $P \in \mathcal{P}$ consistent estimator for $\sigma^2_{l,j} \equiv Avar(\sqrt{n}(\hat{\beta}_{l,j} - \beta_{l,j}))$ and define $\hat{\sigma}^2_{u,j}$ similarly. We construct[4] a $1-\alpha$ confidence interval for $\bar{\beta}_j$ as:

$$[\hat{\beta}_{l,j} - c_\alpha \frac{\hat{\sigma}_{l,j}}{\sqrt{n}}, \hat{\beta}_{u,j} + c_\alpha \frac{\hat{\sigma}_{u,j}}{\sqrt{n}}],$$

with the critical value c_α appropriately chosen as we discuss next.

Consider a $1-\alpha$ CI for $\bar{\beta}_j \in \mathcal{B}_j(|\bar{\delta}| \leq 1)$. Picking $c_\alpha = c_{1,\alpha}$ with $\Phi(c_{1,\alpha}) - \Phi(-c_{1,\alpha}) = 1-\alpha$, where Φ denotes the standard normal cumulative density function, (e.g. $c_{1,0.05} = 1.96$) yields a $1-\alpha$ confidence interval $CI_{\mathcal{B}_j,1-\alpha}$ for the identification region $\mathcal{B}_j(|\bar{\delta}| \leq 1)$. However, $CI_{\mathcal{B}_j,1-\alpha}$ is a conservative confidence interval for $\bar{\beta}_j \in \mathcal{B}_j$ since when $\mathcal{B}_j$ has positive width, $\bar{\beta}_j$ can be close to at most $\beta_{l,j}$ or $\beta_{u,j}$. Further, as discussed in Imbens and Manski (2004), picking $c_\alpha = c_{2,\alpha}$ with $\Phi(c_{2,\alpha}) - \Phi(-c_{2,\alpha}) = 1 - 2\alpha$ (e.g. $c_{2,0.05} = 1.645$) yields a confidence interval whose coverage probabilities do not converge to $1-\alpha$ uniformly across different widths of $\mathcal{B}_j(|\bar{\delta}| \leq 1)$, e.g. for $R_{\tilde{w}.\tilde{x}|\tilde{z},j} = 0$ with point identification. Instead, we construct the uniformly valid confidence interval $CI_{\bar{\beta}_j,1-\alpha}$ for $\bar{\beta}_j \in \mathcal{B}_j$ by setting $c_\alpha = c_{3,\alpha}$ with

$$\Phi(c_{3,\alpha} + \frac{\sqrt{n}(\hat{\beta}_{u,j} - \hat{\beta}_{l,j})}{\max\{\hat{\sigma}_{l,j}, \hat{\sigma}_{u,j}\}}) - \Phi(-c_{3,\alpha}) = 1 - \alpha.$$

Here, by construction, $\hat{\beta}_{u,j} - \hat{\beta}_{l,j} = 2\left|\hat{R}_{\tilde{w}.\tilde{x}|\tilde{z},j}\right| \geq 0$ and it follows from lemma 4 in Imbens and Manski (2004) and lemma 3 and proposition 1 in Stoye (2009) that the confidence interval $CI_{\bar{\beta}_j,1-\alpha}$ is uniformly valid for $\bar{\beta}_j$ in $\mathcal{B}_j(|\bar{\delta}| \leq 1)$.

In the empirical application, in addition to $\hat{\mathcal{B}}_j(|\bar{\delta}| \leq 1)$, we also report estimates for the half as large sharp identification region $\mathcal{B}_j(|\bar{\delta}| \leq 1;\ 0 \leq \bar{\delta};\ sign(R_{\tilde{w}.\tilde{x}|\tilde{z},j}))$ and confidence intervals for $\bar{\beta}_j$ that is partially identified in this set. Note that, unlike for

[4] An alternative method considers the union over confidence intervals for $\bar{\beta}_j(\bar{\delta})$ generated for each $\bar{\delta} \in [-1, 1]$ or $\bar{\delta} \in [0, 1]$ as in Chernozhukov, Rigobon, and Stoker (2010).

$\mathcal{B}_j(|\bar{\delta}| \leq 1)$, this identification region depends on the sign of $R_{\tilde{w}.\tilde{x}|\tilde{z},j}$ which can be estimated. We leave studying the consequences of estimating $R_{\tilde{w}.\tilde{x}|\tilde{z},j}$ to other work to keep a manageable and sharp scope of this paper. Here, we follow the literature (e.g. Reinhold and Woutersen, 2009; Nevo and Rosen, 2012) and report the estimated identification interval $\hat{\mathcal{B}}_j(|\bar{\delta}| \leq 1;\ 0 \leq \bar{\delta} \mid sign(R_{\tilde{w}.\tilde{x}|\tilde{z},j}) = sign(\hat{R}_{\tilde{w}.\tilde{x}|\tilde{z},j}))$ for $\bar{\beta}_j$ and the confidence interval $CI_{\bar{\beta}_j,1-\alpha}(sign(R_{\tilde{w}.\tilde{x}|\tilde{z},j}) = sign(\hat{R}_{\tilde{w}.\tilde{x}|\tilde{z},j}))$ under the assumption that $sign(R_{\tilde{w}.\tilde{x}|\tilde{z},j}) = sign(\hat{R}_{\tilde{w}.\tilde{x}|\tilde{z},j})$. In addition, we report on the p-value for a t-test for the null hypothesis $R_{\tilde{w}.\tilde{x}|\tilde{z},j} = 0$ against the alternative hypothesis $sign(R_{\tilde{w}.\tilde{x}|\tilde{z},j}) = sign(\hat{R}_{\tilde{w}.\tilde{x}|\tilde{z},j})$. When the p-value for this one-sided test is larger than $\frac{1}{2}\alpha$, one can not reject the null $R_{\tilde{w}.\tilde{x}|\tilde{z},j} = 0$ against the alternative $R_{\tilde{w}.\tilde{x}|\tilde{z},j} \neq 0$ at the α significance level, or that $\hat{R}_{\tilde{y}.\tilde{x}|\tilde{z},j}$ consistently estimates $\bar{\beta}_j$.

7 Return to Education and the Black-White Wage Gap

We apply this paper's method to study the financial return to education and the black-white wage gap. Card (1999) surveys several studies measuring the causal effect of education on earning. Among these, studies using institutional features as instruments for education report estimates for the return to a year of education ranging from 6% to 15.3%. Although these IV estimates are higher than the surveyed regression estimates (which range from 5.2% to 8.5%), they are less precise with standard errors sometimes as large as nearly half the IV point estimates. On the other hand, within-family differenced regression estimates in the surveyed twins studies report smaller estimates for the return to education, ranging from 2.2% to 7.8%. See Card (1999, section 4) for a detailed account. Many studies document a black-white wage gap and try to understand its causes. For example, Neal and Johnson (1996) employ a test score to control for unobserved skill and argue that the black-white wage gap primarily reflects a skill gap rather than labor market discrimination. Lang and Manove (2011) provide a model which suggests that one should control for the test score as well as education when comparing the earnings of blacks and whites and document a substantial black-white wage gap in this case. See also Carneiro, Heckman, and Masterov (2005) and Fryer (2011) for studies of the black-white wage gap and its causes.

We consider the wage equation specified in Card (1995, table 2, column 1), and allow for random intercept and slope coefficients as well as a proxy W for U. In particular, the wage and proxy equations are

$$Y = \alpha_y + X'\beta + U\delta_y \quad \text{and} \quad W = \alpha_w + U\delta_w,$$

where Y denotes the logarithm of hourly wage and the vector X contains completed years of education, years of experience as well as its squared value, and three binary variables taking value 1 if a person is black, lives in the South, and lives in a metropolitan area (SMSA) respectively. The confounder U denotes unobserved skill or "ability" and is potentially correlated with elements of X. The proxy W for U denotes the logarithm of the Knowledge of the World of Work (KWW) test score, a test of occupational information. This parsimonious specification facilitates comparing the slope coefficients on U in the Y and W equations while maintaining the commonly used (e.g. Card, 1995) log-level specification of the wage equation. Thus, $100\delta_y$ and $100\delta_w$ denote respectively the approximate percentage changes in wage and KWW directly due to a one unit change in U. Alternatively, one could consider standardized variables. We are interested in $\bar{\beta}$ and especially the components $\bar{\beta}_1$ and $\bar{\beta}_4$ corresponding to education and the black indicator. Here, the average financial return to education is $100\bar{\beta}_1\%$ and the average black-white wage gap for given levels of unobserved ability U and observables X including education is $100\bar{\beta}_4\%$.

We use data drawn from the 1976 subset of the National Longitudinal Survey of Young Men (NLSYM), described in Card (1995). The sample[5] used in Card (1995) contains 3010 observations on individuals who reported valid wage and education. In addition to Y, W, and X, the sample contains data on covariates S and potential instruments Z discussed below. We drop 47 observations (1.56% of the total observations) with missing KWW score[6], as in some results in Card (1995), leading to a sample size of 2963.

We assume Theorem 4.1's conditions on the random coefficients, that $\bar{\beta}(S)$, $\bar{\delta}_y(S)$, and

[5] This sample is reported at http://davidcard.berkeley.edu/data_sets.html as well as in Wooldridge (2008).

[6] The sample also contains IQ score. However, we do not employ IQ as a proxy here since 949 observations (31.5% of the total observations) report missing IQ score. Using the available observations, the sample correlation between IQ and KWW is 0.43 and is strongly significant. Further, using the available observations, employing $\log(IQ)$ instead of $\log(KWW)$ as proxy often leads to tighter bounds and confidence intervals. This, however, could be partly due to sample selection.

$\bar{\delta}_w(S)$ are constants, and that $\bar{Z}(S)$, $\bar{X}(S)$, $\bar{W}(S)$, and $\bar{Y}(S)$ are affine functions of S. In addition, we maintain two assumptions. First, we assume that KWW is, on average, at least as directly sensitive or elastic to ability as wage is, $\left|\bar{\delta}_y\right| \leq \left|\bar{\delta}_w\right|$. Thus, a unit change in U leads, on average, to a direct percentage change in KWW that is at least as large as that in wage. This is a weakening of exogeneity which would require $\bar{\delta}_y = 0$ when X (or Z) and U are freely correlated given S. This assumption may hold if wage is primarily determined by observables, such as education. For example, Cawley, Heckman, and Vytlacil (2001) find that the fraction of wage variance explained by measures of cognitive ability is modest and that personality traits are correlated with earnings primarily through schooling attainment. Also, when ability is not revealed to employers, they may statistically discriminate based on observables such as education (see e.g. Altonji and Pierret (2001) and Arcidiacono, Bayer, and Hizmo (2010)). One may further weaken this assumption by assuming $\left|\bar{\delta}_y\right| \leq c\left|\bar{\delta}_w\right|$ for known $c > 1$, leading to qualitatively similar but larger identification regions. Second, we assume that, on average, ability directly affects KWW and wage in the same direction, $\frac{\bar{\delta}_y}{\bar{\delta}_w} \geq 0$. Alone, this sign restriction determines the direction of the (IV) regression bias. For example, it implies that a regression estimand gives an upper bound on the average return to education when the conditional correlation between $\log(KWW)$ and education is positive.

Table 1 reports results applying the methods discussed in Section 3, using $Z = X$ and $S = 1$. In particular, as in Card (1995, table 2, column 1), column 1 reports regression estimates using $\hat{R}_{\tilde{y}.\tilde{x},j}$ (which consistently estimates $\bar{\beta}_j$ under exogeneity) along with heteroskedasticity-robust standard errors (s.e.) and 95% confidence intervals (denoted by $CI_{0.95}$). The regression estimates for the return to education and the black-white wage gap, with robust s.e. in parentheses, are 7.3%, (0.4%), and −18.8% , (1.7%), respectively. Column 2 reports estimates $\hat{\mathcal{B}}_j(\left|\bar{\delta}\right| \leq 1;\ 0 \leq \bar{\delta} \mid sign(R_{\tilde{w}.\tilde{x}|\tilde{z},j}) = sign(\hat{R}_{\tilde{w}.\tilde{x}|\tilde{z},j}))$ of the sharp identification region obtained under sign and magnitude restrictions on confounding, along with the uniformly valid 95% confidence interval $CI_{\bar{\beta}_j,0.95}(sign(R_{\tilde{w}.\tilde{x}|\tilde{z},j}) = sign(\hat{R}_{\tilde{w}.\tilde{x}|\tilde{z},j}))$ for $\bar{\beta}_j$. Here, we maintain that $sign(R_{\tilde{w}.\tilde{x}|\tilde{z},j}) = sign(\hat{R}_{\tilde{w}.\tilde{x}|\tilde{z},j})$. The estimated identification region for the return to education is $[-0.02\%, 7.3\%]$ with $CI_{\bar{\beta}_j,0.95}$ $[-0.7\%, 8\%]$ and that for the black-white wage gap is $[-18.8\%, 3.7\%]$ with $CI_{\bar{\beta}_j,0.95}$ $[-21.6\%, 6.9\%]$. We also report

$\hat{R}_{\tilde{w}.\tilde{x},j}$, whose magnitude is the estimated width of the identification region, along with its robust standard error and indicate whether a t-test rejects the null hypothesis $R_{\tilde{w}.\tilde{x}|\tilde{z},j} = 0$ against the alternative hypothesis $sign(R_{\tilde{w}.\tilde{x}|\tilde{z},j}) = sign(\hat{R}_{\tilde{w}.\tilde{x}|\tilde{z},j})$ at the 10%, 5%, or 1% level. Last, column 4 reports estimates $\hat{\mathcal{B}}_j(|\bar{\delta}| \leq 1)$ of the twice as large identification region obtained under magnitude restrictions only, along with the uniformly valid 95% confidence intervals $CI_{\bar{\beta}_j,0.95}$. In sum, we find that regression estimates provide an upper bound for the average (here assumed linear) return to education as well as for the average black-white wage gap for given levels of unobserved ability and observables including education.

As in Card (1995, table 2, column 5), we condition on covariates S composed of 8 indicators for region of residence in 1966 and 1 for residence in SMSA in 1966, imputed[7] father and mother education plus 2 indicators for missing father or mother education, 8 binary indicators for interacted mother and father high school, college, or post graduate education, 1 indicator for the presence of the father and mother at age 14 and another indicator for having a single mother at age 14. Table 2 reports sharp bounds estimates, using the estimators $\hat{R}_{\tilde{y}.(\tilde{x}',\tilde{s}')'}$ and $\hat{R}_{\tilde{w}.(\tilde{x}',\tilde{s}')'}$ which condition linearly on covariates S. The results in Table 2 for the average return to education are similar to those in Table 1 with slightly tighter bounds for the average black-white wage gap under sign and magnitude restrictions, given by $[-18.7\%, 1.5\%]$ with $CI_{\bar{\beta}_j,0.95}$ $[-21.9\%, 5\%]$.

As discussed above, conditioning on the proxy W, as is commonly done, does not generally ensure recovering $\bar{\beta}$ from a regression of Y on $(1, X', W', S')'$ (except in special cases such as when α_w and δ_w are constants). Nevertheless, this may attenuate the regression bias (see e.g. Ogburna and VanderWeele, 2012). Table 3 reports estimates from a regression of Y on $(1, X', S^{*\prime})'$ with covariates $S^* = (S', KWW)'$. Conditioning on $W = \log(KWW)$ instead of KWW yields very similar estimates. These estimates and their confidence intervals lie respectively within the identification regions and $CI_{\bar{\beta}_j,0.95}$ reported in Table 2. In particular, this regression's estimates of the average return to education and black-white wage gap, with robust s.e. in parentheses, are 5.5%, (0.5%), and −14.3%,

[7] From the 2963 observations, 12% report missing mother's education and 23% report missing father's education. We follow Card (1995) and impute these missing values using the averages of the reported observations.

(2.1%), respectively. The estimate of the coefficient on KWW is small and significant, 0.8% with robust s.e. 0.1%.

We also augment X to include an interaction term $(Education - 12) \times Black$, multiplying the black binary indicator with years of education minus 12. Table 4 reports[8] the conditional on S results. Under sign and magnitude restrictions, the estimates for the sharp identification region for the average return to education for non-blacks is $[0.4\%, 6.8\%]$ with $CI_{\bar{\beta}_j,0.95}$ $[-0.4\%, 7.5\%]$, that for the average black-white return to education differential is $[-1.2\%, 1.7\%]$ with $CI_{\bar{\beta}_j,0.95}$ $[-2.4\%, 2.8\%]$, and that for the average black-white wage gap, corresponding to individuals with 12 years of education, is $[-19.3\%, 1.8\%]$ with $CI_{\bar{\beta}_j,0.95}$ $[-22.5\%, 5.5\%]$. Thus, the average return to education for the black subpopulation may differ slightly from the nonblack subpopulation, if at all. We follow Card (1995) and maintain that these average returns are equal.

Further, as in Card (1995), we employ an indicator for the presence of a four year college in the local labor market, age, and age squared as instruments for education, experience, and experience squared in the specification from Table 2 with covariates S. Note that this paper's method does not require Z to be exogenous (for example, Carneiro and Heckman (2002) provide evidence suggesting that distance to school may be endogenous). As reported in Table 5, the IV results yield wider identification regions for the average return to education with larger confidence intervals[9]. In particular, under sign and magnitude restrictions, the conditional on S IV-based identification region for the average return to education is estimated to be $[2.9\%,\ 13.4\%]$ with $CI_{\bar{\beta}_j,0.95}$ $[-6\%,\ 22\%]$ and that for the for the average black-white wage gap is $[-16.2\%,\ 2.6\%]$, which is slightly tighter than the regression-based estimate albeit with comparable $CI_{\bar{\beta}_j,0.95}$ $[-20.9\%,\ 7.5\%]$.

Last, the results in Table 6 relax the linear return to education assumption in the specification in Table 2 by including in X binary indicators for exceeding t years of education, where $t = 2, ..., 18$ as in the sample, instead of total years of education. Because we do not require exogenous instruments, our method can accommodate this less restrictive specifica-

[8] Similar results obtain when we do not condition on S.

[9] We also consider not conditioning on S and employing different instruments, such as an interaction of low parental education with college proximity as in Card (1995). However, the regression-based estimates of the identification regions are often narrower and have especially tighter confidence intervals.

tion. As before, regression estimates generally give an upper bound on the average return to education and the average black-white wage gap[10]. We find nonlinearity in the return to education, with the 12^{th}, 16^{th}, and 18^{th} year, corresponding to obtaining a high school, college, and possibly a graduate degree, yielding a high average return. For example, under sign and magnitude restrictions on confounding, the estimate of the identification region for the average return to the 12^{th} year is $[1.6\%,\ 14.6\%]$ with $CI_{\bar{\beta}_j,0.95}$ $[-4.2\%,\ 20\%]$ and that for the 16^{th} year is $[13.33\%,\ 19.5\%]$ with $CI_{\bar{\beta}_j,0.95}$ $[7.5\%,\ 25.1\%]$. Under a magnitude restriction on confounding only, the estimate of the identification region for the average return to the 18^{th} year is $[13.9\%,\ 15.9\%]$ with $CI_{\bar{\beta}_j,0.95}$ $[6.5\%,\ 24.1\%]$ and we cannot reject at comfortable significance levels the null that the width of this region is zero or that regression consistently estimates this return (the regression estimates are 14.9% with robust s.e. 4.5%). Graph 1 plots the estimates of the sharp identification regions and $CI_{\bar{\beta}_j,0.95}$ for the incremental average returns to the 8^{th} up to the 18^{th} year of education under sign and magnitude restrictions as well as magnitude restrictions only. Further, the estimate of the sharp identification region for the black-white wage gap under sign and magnitude restrictions is similar to that in Table 2 and given by $[-17.8\%,\ 1.9\%]$ with $CI_{\bar{\beta}_j,0.95}$ $[-21\%,\ 5.4\%]$.

This empirical analysis imposes assumptions including linearity or separability among observables and the confounder, restrictions on the random coefficients, the presence of one confounder U denoting "ability" which we proxy using $\log(KWW)$, and the assumptions $\left|\bar{\delta}_y\right| \leq \left|\bar{\delta}_w\right|$ and $\frac{\bar{\delta}_y}{\bar{\delta}_w} \geq 0$. Of course, one should interpret the results carefully if these assumptions are suspected to fail. For example, this analysis does not generally allow the return to education to depend on ability or explicitly study employers' learning of workers' abilities (see e.g. Altonji and Pierret (2001) and Arcidiacono, Bayer, and Hizmo (2010)). Also, if other confounders are present and valid instruments or proxies for these are not available then additional assumptions are needed to (partially) identify elements of $\bar{\beta}$. Nevertheless, this analysis does not require several commonly employed assumptions. In particular, it does not require regressor or instrument exogeneity or restrict the dependence between X or Z and U (given S). Also, it does not require a linear return to education. Further, it permits test scores to be error-laden proxies for unobserved ability.

[10]Similar results obtain when we do not condition on S.

8 Conclusion

This paper studies the identification of the average effects $\bar{\beta}$ of X on Y under magnitude and sign restrictions on confounding. We do not require (conditional) exogeneity of instruments or regressors. Using proxies W for the confounders U in a random coefficient structure, we ask how do the average direct effects of U on Y compare in magnitude and sign to those of U on W. Exogeneity (zero average direct effect) and equi- or proportional confounding (equal or equal to a known proportion direct effects) are limiting cases yielding full identification of $\bar{\beta}$. Alternatively, we partially identify elements of $\bar{\beta}$ in a sharp bounded interval when W is sufficiently sensitive to U, and may obtain sharp upper or lower bounds otherwise. The paper extends this analysis to accommodate conditioning on covariates and a semiparametric specification for the Y equation. Appendix A contains extensions to a panel structure with individual and time varying random coefficients and to cases with proxies included in the Y equation. After studying estimation and confidence intervals, the paper applies its methods to study the return to education and the black-white wage gap using data from the 1976 subset of NLSYM used in Card (1995). Under restrictions on confounding, we partially identify in a sharp bounded interval the average financial return to education as well as the average black-white wage gap for given levels of unobserved ability and observables including education. We find that regression estimates provide an upper bound on the average return to education and the black-white wage gap. We also find nonlinearity in the return to education with the 12^{th}, 16^{th}, and 18^{th} years, corresponding to obtaining a high school, college, and possibly a graduate degree, yielding a high average return. Important extensions for future work include identifying (other aspects of) the distribution of the effects β as well as measuring causal effects in nonseparable systems under restrictions on confounding.

A Appendix A: Extensions

Appendix A contains extensions to a panel structure and to cases with proxies included in the Y equation. Throughout, we do not explicitly condition on covariates to simplify the exposition.

A.1 Panel with Individual and Time Varying Random Coefficients

We consider a panel structure whereby we index the variables and coefficients $(\theta_t', V_t')'$ in S.1 by $t = 1, 2$. Here, U may denote time-invariant unobserved individual characteristics. We allow the proxy W_t for U to be an element $\underset{k_1 \times 1}{X_{1,t}}$ of X_t. Thus, for $t = 1, 2$:

$$Y_t = \alpha_{y,t} + X_t'\beta_t + U'\delta_{y,t} \quad \text{and} \quad X_{1,t}' = \alpha_{x_1,t}' + U'\delta_{x_1,t}.$$

This is a panel structure with individual and time varying random coefficients where we do not require a "fixed effect" and thus $\delta_{y,t}$ need not equal $\delta_{y,t'}$.

For $t, t' = 1, 2$, $t \neq t'$, we apply Theorem 3.1 using $X_{1,t'}$ as the proxy to derive an expression for $\bar{\beta}_t$. In particular, the conditions in Theorem 3.1 require that (i) $E(\tilde{Z}_t\tilde{X}_t')$ and $\bar{\delta}_{x_1,t}$ are nonsingular, (ii) $Cov(\alpha_{y,t}, Z_t) = 0$, $E(\tilde{\beta}_t|X_t, Z_t) = 0$, $E(\tilde{\delta}_{y,t}|U, Z_t) = 0$, and (iii) $Cov(\tilde{\alpha}_{x_1,t'}, Z_t) = 0$, $E(\tilde{\delta}_{x_1,t'}|U, Z_t) = 0$. Condition (iii) restricts the dependence of the random coefficient in the $X_{1,t'}$ equation with U and non-contemporaneous Z_t. Then, with $\bar{\delta}_t \equiv \bar{\delta}_{x_1,t'}^{-1}\bar{\delta}_{y,t}$, Theorem 3.1 gives for $t, t' = 1, 2$, $t \neq t'$:

$$\bar{\beta}_t = R_{\tilde{y}_t.\tilde{x}_t|\tilde{z}_t} - \bar{\delta}_t R_{\tilde{x}_{1,t'}.\tilde{x}_t|\tilde{z}_t}.$$

The IV regression bias is $B_t \equiv R_{\tilde{y}_t.\tilde{x}_t|\tilde{z}_t} - \bar{\beta}_t = \bar{\delta}_t R_{\tilde{x}_{1,t'}.\tilde{x}_t|\tilde{z}_t}$. Thus, $\bar{\beta}_t$ is fully identified under exogeneity ($B_t = 0$) or signed proportional confounding ($sign(\bar{\delta}_{h,t})$ and $|\bar{\delta}_{h,t}| = c_{h,t}$, $h = 1, ..., k_1$, known). Applying Corollary 3.3 in this setup with $A_t \equiv R_{\tilde{x}_{1,t'}.\tilde{x}_t|\tilde{z}_t}$ sharply partially identifies the elements of $\bar{\beta}_t$ for $t = 1, 2$ so that $\bar{\beta}_{j,t} \in \mathcal{B}_{j,t}(\underset{h=1,...,k_1}{sign}(|\bar{\delta}_{h,t}| - c_{h,t}); \underset{h=1,...,k_1}{sign}(\bar{\delta}_{h,t}A_{jh,t}))$. The restrictions on the magnitude and sign of the average direct effects of U on Y_t and $X_{1,t'}$ may be plausible, for example, if one suspects that Y at time t is less directly responsive to U than X_1 is in both times t and t'.

A.2 Included Proxies

Sometimes, a researcher may want to allow proxies W to directly impact the response Y. In this case, W is a component X_1 of X. While Theorem 3.1 does not rule out that $W = X_1$, the conditions of Theorem 3.1 entail restrictions on Z in this case. First, when $W = X_1$, conditions (i) and (iii) of Theorem 3.1 imply that all elements of Z must be correlated with U since $E(\tilde{Z}\tilde{X}')$ is singular otherwise. Second, when $W = X_1$, the requirement that $Cov(\alpha_w, Z) = 0$ in condition (iii) generally rules out that Z contains elements of X_1. The following two subsections accommodate these two cases respectively by providing alternative conditions enabling (partial) identification of elements of $\bar{\beta}$, under sign and magnitude restrictions on confounding.

A.2.1 "Under"-Identification Using Valid Instruments

When $W = X_1$, Theorem 3.1 requires that all the elements of Z are correlated with U. Sometimes a vector Z_1 of one or a few valid (e.g. randomized) instruments may be available, albeit the dimension of X may exceed that of Z_1. Nevertheless, a researcher may wish to employ the exogenous instrument Z_1. The next Theorem allows for this possibility and provides an expression for $\bar{\beta}$ which depends on the average direct effects of U on Y and X_1.

Theorem A.1 *Assume S.1 with* $\underset{\ell\times 1}{Z} \equiv (\underset{\ell_1\times 1}{Z_1}{}', \underset{\ell_2\times 1}{Z_2}{}')'$, $\underset{k\times 1}{X} \equiv (\underset{k_1\times 1}{X_1}{}', \underset{k_2\times 1}{X_2}{}')'$, $W = X_1$, *with* $\ell_1, \ell_2 \geq 0$, $\ell = k$, $k_1 = p$, *and*

(i) $E(\tilde{Z}\tilde{X}')$ *and* $\bar{\delta}_{x_1}$ *are nonsingular,*

(ii) $Cov(U, Z_1) = 0$,

(iii) $Cov(\alpha_y, Z) = 0$, $E(\tilde{\beta}|X, Z) = 0$, *and* $E(\tilde{\delta}_y|U, Z) = 0$,

(iv) $Cov(\alpha_{x_1}, Z_2) = 0$ *and* $E(\tilde{\delta}_{x_1}|U, Z_2) = 0$.

Let $A \equiv E(\tilde{Z}\tilde{X}')^{-1}\left[\ 0',\ \ E(\tilde{Z}_2\tilde{X}_1')'\ \right]'$ *and* $\bar{\delta} \equiv \bar{\delta}_{x_1}^{-1}\bar{\delta}_y$ *then*

$$\bar{\beta} = R_{\tilde{y}.\tilde{x}|\tilde{z}} - A\bar{\delta}.$$

The IV regression bias is $B \equiv R_{\tilde{y}.\tilde{x}|\tilde{z}} - \bar{\beta} = A\bar{\delta}$. The conditions in Theorem A.1 are analogous to those in Theorem 3.1, except that they assume that Z_1 is uncorrelated with U and let Z_1 freely depend on the coefficients in the proxy X_1 equation. Thus, if $Z = Z_2$

Theorem A.1 reduces to Theorem 3.1 with $W = X_1$, and if $Z = Z_1$ exogeneity holds. Here, $\bar{\beta}_j$ is fully identified under exogeneity ($B_j = 0$) or signed proportional confounding ($sign(\bar{\delta}_h)$ and $|\bar{\delta}_h| = c_h$, $h = 1, ..., k_1$, known). Otherwise, $\bar{\beta}_j$ is sharply partially identified in $\mathcal{B}_j(\underset{h=1,...,k_1}{sign}(|\bar{\delta}_h| - c_h); \underset{h=1,...,k_1}{sign}(\bar{\delta}_h A_{jh}))$ under assumptions on how the average direct effects of U on X_1 compare in magnitude and sign to those of U on Y.

A.2.2 Multiple Included Proxies

When $W = X_1$, the assumption $Cov(\alpha_w, Z) = 0$ in condition (iii) of Theorem 3.1 generally rules out that X_1 is a component of Z and therefore that $Z = X$. We relax this requirement and let $W = (X_1', X_2')'$ with X_1 and X_2 two vectors of proxies included in the equation for Y and where X_1, and possibly X_2, is a component of Z.

The next Theorem derives an expression for $\bar{\beta}$ which depends on the unknowns $\bar{\delta}_{x_1}^{-1}\bar{\delta}_y$ and $\bar{\delta}_{x_2}^{-1}\bar{\delta}_y$ involving the average direct effects of U on Y and those of U on X_1 and X_2. Here, we let $Z_1 = X_1$, with Z potentially equal to X.

Theorem A.2 *Assume S.1 and let* $W = (\underset{k_1 \times 1}{X_1}', \underset{k_2 \times 1}{X_2}')'$ *with* $X_g' = \alpha_{x_g}' + U'\delta_{x_g}$, *for* $g = 1, 2$, $\underset{k \times 1}{X} = (W', \underset{k_3 \times 1}{X_3}')'$, $Z_1 = X_1$, $\underset{\ell \times 1}{Z} \equiv (\underset{\ell_1 \times 1}{Z_1}', \underset{\ell_2 \times 1}{Z_2}')'$, $k_1 = k_2 = p$, $k_3 \geq 0$, $\ell = k$, *and that*
(i) $E(\tilde{Z}\tilde{X}')$, $\bar{\delta}_{x_1}$, $\bar{\delta}_{x_2}$ *are nonsingular,*
(ii) $Cov(\alpha_y, Z) = 0, E(\tilde{\beta}|X, Z) = 0,$ *and* $E(\tilde{\delta}_y|U, Z) = 0$,
(iii) $Cov(\alpha_{x_1}, (U', X_2', Z_2')') = 0$ *and* $E(\tilde{\delta}_{x_1}|U, X_2, Z_2) = 0$,
(iv) $Cov(\alpha_{x_2}, U) = 0$ *and* $E(\tilde{\delta}_{x_2}|U) = 0$.

Let $\bar{\delta}_1 \equiv \bar{\delta}_{x_1}^{-1}\bar{\delta}_y$ *and* $\bar{\delta}_2 \equiv \bar{\delta}_{x_2}^{-1}\bar{\delta}_y$ *then*

$$\bar{\beta} = R_{\tilde{y}.\tilde{x}|\tilde{z}} - E(\tilde{Z}\tilde{X}')^{-1} \begin{bmatrix} E(\tilde{Z}_1\tilde{X}_2')\bar{\delta}_2 \\ E(\tilde{Z}_2\tilde{X}_1')\bar{\delta}_1 \end{bmatrix}.$$

The IV regression bias is

$$B \equiv R_{\tilde{y}.\tilde{x}|\tilde{z}} - \bar{\beta} = E(\tilde{Z}\tilde{X}')^{-1} \begin{bmatrix} E(\tilde{Z}_1\tilde{X}_2')\bar{\delta}_2 \\ E(\tilde{Z}_2\tilde{X}_1')\bar{\delta}_1 \end{bmatrix}.$$

The conditions in Theorem A.2 extend those in Theorem 3.1 to allow the proxy X_1 to equal Z_1 except that they also restrict the dependence between the proxy X_2 and the coefficients α_{x_1} and δ_{x_1} in the equation for the proxy X_1 as well as the dependence between U and $(\alpha_{x_1}', \alpha_{x_2}', \delta_{x_2}')'$.

The expression for $\bar{\beta}$ in Theorem A.2 can be used to fully or sharply partially identify elements of $\bar{\beta}$.

Corollary A.3 *Assume the conditions of Theorem A.2 and let* $X_{2,3} \equiv (X_2', X_3')'$ *and* $j = 1, ..., k$. *(i) If* $B_j = 0$ *(exogeneity) then* $\bar{\beta}_j = R_{\tilde{y}.\tilde{x}|\tilde{z},j}$. *(ii) If* $\left|\bar{\delta}_1\right| = \kappa_1$ *and* $\left|\bar{\delta}_2\right| = \kappa_2$ *(proportional confounding) then*

$$\bar{\beta} = R_{\tilde{y}.\tilde{x}|\tilde{z}} - E(\tilde{Z}\tilde{X}')^{-1} \begin{bmatrix} \sum_{h=1}^{p} sign(\bar{\delta}_{2h}) E(\tilde{Z}_1 \tilde{X}_{2h}') \kappa_{2h} \\ \sum_{h=1}^{p} sign(\bar{\delta}_{1h}) E(\tilde{Z}_2 \tilde{X}_{1h}') \kappa_{1h} \end{bmatrix}.$$

In particular, let $c = (\kappa_1', \kappa_2')'$, $\bar{\delta} = (\bar{\delta}_1', \bar{\delta}_2')'$, $\underset{k_1 \times k_1}{P_1} \equiv E(\epsilon_{\tilde{z}_1.\tilde{z}_2|\tilde{x}_{2,3}} \tilde{X}_1')$, $\underset{(k_2+k_3)\times(k_2+k_3)}{P_2} \equiv E(\epsilon_{\tilde{z}_2.\tilde{z}_1|\tilde{x}_1} \tilde{X}_{2,3}')$, *and*

$$\underset{k\times(k_1+k_2)}{A} \equiv \begin{bmatrix} -R_{\tilde{x}_{2,3}.\tilde{x}_1|\tilde{z}_1} P_2^{-1} E(\tilde{Z}_2 \tilde{X}_1'), & P_1^{-1} E(\tilde{Z}_1 \tilde{X}_2') \\ P_2^{-1} E(\tilde{Z}_2 \tilde{X}_1'), & -R_{\tilde{x}_1.\tilde{x}_{2,3}|\tilde{z}_2} P_1^{-1} E(\tilde{Z}_1 \tilde{X}_2') \end{bmatrix}.$$

Then

$$\bar{\beta} = R_{\tilde{y}.\tilde{x}|\tilde{z}} - B = R_{\tilde{y}.\tilde{x}|\tilde{z}} - A\bar{\delta},$$

and

$$\bar{\beta}_j = R_{\tilde{y}.\tilde{x}|\tilde{z},j} - \textstyle\sum_{h=1}^{2p} sign(\bar{\delta}_h) A_{jh} c_h \quad \text{for } j = 1, ..., k.$$

Thus, $\bar{\beta}_j$ is fully identified under exogeneity ($B_j = 0$) or signed proportional confounding ($\left|\bar{\delta}_h\right| = c_h$ and $sign(\bar{\delta}_h)$, $h = 1, ..., 2p$, known). Otherwise, $\bar{\beta}_j$ is sharply partially identified in $\mathcal{B}_j(\underset{h=1,...,2p}{sign}(\left|\bar{\delta}_h\right| - c_h); \underset{h=1,...,2p}{sign}(\bar{\delta}_h A_{jh}))$, defined analogously to Corollary 3.3, under assumptions on how the average direct effects of U on X_1 and X_2 compare in magnitude and sign to those of U on Y.

B Appendix B: Mathematical Proofs

Proof of Theorem 3.1 Apply Theorem 4.1 with $S = 1$.

Proof of Corollary 3.2 The proof is immediate.

Proof of Corollary 3.3 We have the following bounds for $h = 1, ..., m$:
Given $sign(-\bar{\delta}_h R_{\tilde{w}_h.\tilde{x}|\tilde{z},j})$, if $\left|\bar{\delta}_h\right| \leq c_h$ then

$$\min\{sign(-\bar{\delta}_h R_{\tilde{w}_h.\tilde{x}|\tilde{z},j}) c_h \left|R_{\tilde{w}_h.\tilde{x}|\tilde{z},j}\right|, 0\} \leq -\bar{\delta}_h R_{\tilde{w}_h.\tilde{x}|\tilde{z},j} \leq \max\{0, sign(-\bar{\delta}_h R_{\tilde{w}_h.\tilde{x}|\tilde{z},j}) c_h \left|R_{\tilde{w}_h.\tilde{x}|\tilde{z},j}\right|\};$$

if $c_h < \left|\bar{\delta}_h\right|$ then

$$-\infty < -\bar{\delta}_h R_{\tilde{w}_h.\tilde{x}|\tilde{z},j} < -c_h \left|R_{\tilde{w}_h.\tilde{x}|\tilde{z},j}\right| \quad \text{if } -\bar{\delta}_h R_{\tilde{w}_h.\tilde{x}|\tilde{z},j} < 0,$$
$$c_h \left|R_{\tilde{w}_h.\tilde{x}|\tilde{z},j}\right| < -\bar{\delta}_h R_{\tilde{w}_h.\tilde{x}|\tilde{z},j} < +\infty \quad \text{if } 0 < -\bar{\delta}_h R_{\tilde{w}_h.\tilde{x}|\tilde{z},j}.$$

The identification regions then follow from $\bar{\beta}_j = R_{\tilde{y}.\tilde{x}|\tilde{z},j} - \sum_{h=1}^{m} R_{\tilde{w}_h.\tilde{x}|\tilde{z},j}\bar{\delta}_h$.

To prove sharpness, we show that for each element b of $\mathcal{B}_j$ there exists $\delta_y(b)$ and $\delta_w(b)$ such that, when $\delta_y = \delta_y(b)$ and $\delta_w = \delta_w(b)$, the joint distribution (θ, U, V) satisfies the conditions of Theorem 3.1 and the restrictions on $\underset{h=1,\ldots,m}{sign}\ (\left|\bar{\delta}_h\right| - c_h)$ and $\underset{h=1,\ldots,m}{sign}\ (\bar{\delta}_h A_{jh})$ underlying $\mathcal{B}_j$ hold. For each b, let $\delta_y(b) = \bar{\delta}_y(b)$ and $\delta_w(b) = \bar{\delta}_y(b)$, so that $E(\tilde{\delta}_y(b)|U, Z) = 0$ and $E(\tilde{\delta}_w(b)|U, Z) = 0$, and set e.g. $\bar{\delta}_w(b) = I$ so that $\bar{\delta}(b) \equiv \bar{\delta}_w^{-1}(b)\bar{\delta}_y(b) = \bar{\delta}_y(b)$. We now construct $\bar{\delta}(b)$ such that $b = R_{\tilde{y}.\tilde{x}|\tilde{z},j} - \sum_{h=1}^{m} A_{hj}\bar{\delta}_h(b)$. First, we partition $h = 1, ..., m$ in the restrictions underlying $\mathcal{B}_j$ such that

$$\begin{array}{ll} \left|\bar{\delta}_h\right| \leq c_h \text{ and } \bar{\delta}_h A_{jh} \leq 0 \ \text{ for } h = 1, ..., g''; & \left|\bar{\delta}_h\right| \leq c_h \text{ and } \bar{\delta}_h A_{jh} \geq 0 \ \text{ for } h = g''+1, ..., g; \\ c_h < \left|\bar{\delta}_h\right| \text{ and } \bar{\delta}_h A_{jh} < 0 \ \text{ for } h = g, ..., g'; & c_h < \left|\bar{\delta}_h\right| \text{ and } \bar{\delta}_h A_{jh} > 0 \ \text{ for } h = g'+1, ..., m, \end{array}$$

with some of these categories possibly empty. Then any element of the identification regions in (a), (b), or (c) can be expressed as

$$b = R_{\tilde{y}.\tilde{x}|\tilde{z},j} + a_1(b)\textstyle\sum_{h=1}^{g''} c_h \left|A_{jh}\right| - a_2(b)\sum_{h=g''+1}^{g} c_h \left|A_{jh}\right| + a_3(b)\sum_{h=g+1}^{g'} c_h \left|A_{jh}\right| - a_4(b)\sum_{h=g'+1}^{m} c_h \left|A_{jh}\right|,$$

for $0 \leq a_1(b), a_2(b) \leq 1$ and $1 < a_3(b), a_4(b)$, where we omit the sums that correspond to empty groups. It suffices then to put

$$\bar{\delta}_h(b) = -sign(A_{jh})a_1(b)c_h \ \text{ for } h = 1, ..., g''; \quad \bar{\delta}_h(b) = sign(A_{jh})a_2(b)c_h \ \text{ for } h = g''+1, ..., g;$$
$$\bar{\delta}_h(b) = -sign(A_{jh})a_3(b)c_h \ \text{ for } h = g+1, ..., g'; \quad \bar{\delta}_h(b) = sign(A_{jh})a_4(b)c_h \ \text{ for } h = g'+1, ..., m.$$

Proof of Theorem 4.1 S.1 ensures finiteness of moments. By (ii) we have

$$\begin{aligned} E(\tilde{Z}(S)\tilde{Y}(S)|S) &= E(\tilde{Z}(S)Y|S) = E(\tilde{Z}(S)(\alpha_y + X'\beta + U'\delta_y)|S) \\ &= E(\tilde{Z}(S)\tilde{X}'(S)|S)\bar{\beta}(S) + E(\tilde{Z}(S)U'|S)\bar{\delta}_y(S) \end{aligned}$$

and by (iii) we have

$$E(\tilde{Z}(S)\tilde{W}'(S)|S) = E(\tilde{Z}(S)W|S) = E(\tilde{Z}(S)(\alpha_w' + U'\delta_w)|S) = E(\tilde{Z}(S)U'|S)\bar{\delta}_w(S).$$

Since $E(\tilde{Z}(S)\tilde{X}'(S)|S)$ and $\bar{\delta}_w(S)$ are nonsingular, we have

$$R_{\tilde{y}(s).\tilde{x}(s)|\tilde{z}(s)}(S) = \bar{\beta}(S) + R_{\tilde{w}(s).\tilde{x}(s)|\tilde{z}(s)}(S)\bar{\delta}(S).$$

Proof of Theorem 5.1 S.2 ensures finiteness of moments. For $x \in S_X$, (ii) gives

$$E(Y|X=x,S) = E[r(x,U_y)|X=x,S] + E(U'\delta_y|X=x,S) = E[r(x,U_y)|S] + E(U'|X=x,S)\bar{\delta}_y(S)$$

and by (iii) we have

$$E(W'|X=x,S) = E(\alpha_w'|S) + E(U'|X=x,S)\bar{\delta}_w(S).$$

Since $\bar{\delta}_w(S)$ is nonsingular, we have

$$E(Y|X=x,S) = E[r(x,U_y)|S] + [E(W'|X=x,S) - E(\alpha_w'|S)]\bar{\delta}(S).$$

It follows that

$$\bar{\beta}(x,x^*|S) = E(Y|X=x^*,S) - E(Y|X=x,S) - [E(W'|X=x^*,S) - E(W'|X=x,S)]\bar{\delta}(S).$$

Proof of Theorem 6.1 Let $\hat{Q} \equiv \frac{1}{n}\sum_{i=1}^{n} diag(\tilde{Z}_i\tilde{X}_i', \tilde{Z}_i\tilde{X}_i')$ and $\hat{M} = \frac{1}{n}\sum_{i=1}^{n}(\tilde{Z}_i'\epsilon_{\tilde{y}.\tilde{x}|\tilde{z},i}, \tilde{Z}_i'\epsilon_{\tilde{w}.\tilde{x}|\tilde{z},i})'$. We have that

$$\sqrt{n}((\hat{R}_{\tilde{y}.\tilde{x}|\tilde{z}}', \hat{R}_{\tilde{w}.\tilde{x}|\tilde{z}}')' - (R_{\tilde{y}.\tilde{x}|\tilde{z}}', R_{\tilde{w}.\tilde{x}|\tilde{z}}')') = \hat{Q}^{-1}\sqrt{n}\hat{M} = (\hat{Q}^{-1} - Q^{-1})\sqrt{n}\hat{M} + Q^{-1}\sqrt{n}\hat{M},$$

exists in probability for all n sufficiently large uniformly in $P \in \mathcal{P}$ by (i) and since $E(\tilde{Z}\tilde{X}')$, and thus Q, is nonsingular uniformly in $P \in \mathcal{P}$. The result obtains since $\hat{Q}^{-1} - Q^{-1} = o_p(1)$ uniformly in $P \in \mathcal{P}$ by (i), and $\sqrt{n}\hat{M} \xrightarrow{d} N(0,\Lambda)$ uniformly in $P \in \mathcal{P}$ by (ii).

Proof of Theorem A.1 S.1 ensures finiteness of moments. By (iii), we have

$$E(\tilde{Z}\tilde{Y}') = E[\tilde{Z}(\alpha_y + X'\beta + U'\delta_y)] = E(\tilde{Z}\tilde{X}')\bar{\beta} + E(\tilde{Z}U')\bar{\delta}_y.$$

By (ii) and since $E(\tilde{Z}\tilde{X}')$ is nonsingular, we have

$$\bar{\beta} = R_{\tilde{y}.\tilde{x}|\tilde{z}} - E(\tilde{Z}\tilde{X}')^{-1}E(\tilde{Z}U')\bar{\delta}_y = R_{\tilde{y}.\tilde{x}|\tilde{z}} - E(\tilde{Z}\tilde{X}')^{-1}\left[\begin{array}{cc} 0', & E(\tilde{Z}_2U')' \end{array}\right]'\bar{\delta}_y.$$

By (iv) we have

$$E(\tilde{Z}_2\tilde{X}_1') = E(\tilde{Z}_2X_1') = E[\tilde{Z}_2(\alpha_{x_1}' + U'\delta_{x_1})] = E(\tilde{Z}_2U')\bar{\delta}_{x_1},$$

so that by (i)

$$E(\tilde{Z}_2U') = E(\tilde{Z}_2\tilde{X}_1')\bar{\delta}_{x_1}^{-1}.$$

It follows that

$$\bar{\beta} = R_{\tilde{y}.\tilde{x}|\tilde{z}} - E(\tilde{Z}\tilde{X}')^{-1}\left[\begin{array}{cc} 0', & E(\tilde{Z}_2\tilde{X}_1')' \end{array}\right]'\bar{\delta}_{x_1}^{-1}\bar{\delta}_y.$$

Proof of Theorem A.2 S.1 ensures finiteness of moments. By (ii), we have

$$E(\tilde{Z}\tilde{Y}) = E(\tilde{Z}Y) = E(\tilde{Z}\tilde{X}')\bar{\beta} + E(\tilde{Z}U')\bar{\delta}_y$$

Further, (iii) gives

$$E(\tilde{Z}_2\tilde{X}_1') = E[\tilde{Z}_2(\alpha_{x_1}' + U'\delta_{x_1})] = E(\tilde{Z}_2U')\bar{\delta}_{x_1},$$

and $Z_1 = X_1$, (iii), and (iv) give

$$\begin{aligned} E(\tilde{Z}_1\tilde{X}_2') &= \bar{\delta}_{x_1}'E(U\tilde{X}_2') = \bar{\delta}_{x_1}'E[\tilde{U}(\alpha_{x_2}' + U'\delta_{x_2})] = \bar{\delta}_{x_1}'E(\tilde{U}\tilde{U}')\bar{\delta}_{x_2}, \text{ and} \\ E(\tilde{Z}_1U') &= E[(\alpha_{x_1} + \delta_{x_1}'U)\tilde{U}'] = \bar{\delta}_{x_1}'E(\tilde{U}\tilde{U}'). \end{aligned}$$

Since $E(\tilde{Z}\tilde{X}')$, $\bar{\delta}_{x_1}$, and $\bar{\delta}_{x_2}$ are nonsingular, we have

$$\bar{\beta} = R_{\tilde{y}.\tilde{x}|\tilde{z}} - E(\tilde{Z}\tilde{X}')^{-1}\left[\begin{array}{c} E(\tilde{Z}_1\tilde{X}_2')\bar{\delta}_{x_2}^{-1}\bar{\delta}_y \\ E(\tilde{Z}_2\tilde{X}_1')\bar{\delta}_{x_1}^{-1}\bar{\delta}_y \end{array}\right].$$

Proof of Corollary A.3 The result follows from the expression for $\bar{\beta}$ in Theorem A.2. For the expression for $\bar{\beta}_j$, recall that $E(\tilde{Z}\tilde{X}')^{-1}$ is given by (e.g. Baltagi, 1999, p. 185):

$$E(\tilde{Z}\tilde{X}')^{-1} = \left[\begin{array}{cc} E(\tilde{Z}_1\tilde{X}_1'), & E(\tilde{Z}_1\tilde{X}_{2,3}') \\ E(\tilde{Z}_2\tilde{X}_1'), & E(\tilde{Z}_2\tilde{X}_{2,3}') \end{array}\right]^{-1} = \left[\begin{array}{cc} P_1^{-1}, & -R_{\tilde{x}_{2,3}.\tilde{x}_1|\tilde{z}_1}P_2^{-1} \\ -R_{\tilde{x}_1.\tilde{x}_{2,3}|\tilde{z}_2}P_1^{-1}, & P_2^{-1} \end{array}\right],$$

where

$$\begin{aligned} P_1 &\equiv E(\tilde{Z}_1\tilde{X}_1') - E(\tilde{Z}_1\tilde{X}_{2,3}')E(\tilde{Z}_2\tilde{X}_{2,3}')^{-1}E(\tilde{Z}_2\tilde{X}_1') = E(\epsilon_{\tilde{z}_1.\tilde{z}_2|\tilde{x}_{2,3}}\tilde{X}_1') \\ P_2 &\equiv E(\tilde{Z}_2\tilde{X}_{2,3}') - E(\tilde{Z}_2\tilde{X}_1')E(\tilde{Z}_1\tilde{X}_1')^{-1}E(\tilde{Z}_1\tilde{X}_{2,3}') = E(\epsilon_{\tilde{z}_2.\tilde{z}_1|\tilde{x}_1}\tilde{X}_{2,3}'). \end{aligned}$$

The result then follows from

$$\bar{\beta} = R_{\tilde{y}.\tilde{x}|\tilde{z}} - \left[\begin{array}{c} P_1^{-1}E(\tilde{Z}_1\tilde{X}_2')\bar{\delta}_2 - R_{\tilde{x}_{2,3}.\tilde{x}_1|\tilde{z}_1}P_2^{-1}E(\tilde{Z}_2\tilde{X}_1')\bar{\delta}_1 \\ -R_{\tilde{x}_1.\tilde{x}_{2,3}|\tilde{z}_2}P_1^{-1}E(\tilde{Z}_1\tilde{X}_2')\bar{\delta}_2 + P_2^{-1}E(\tilde{Z}_2\tilde{X}_1')\bar{\delta}_1 \end{array}\right].$$

Table 1: Regression-Based Estimates of Log Wage Equation under Restrictions on Confounding

		$\hat{R}_{\tilde{y}.\tilde{x},j}$	$\hat{\mathcal{B}}_j(\lvert\bar{\delta}\rvert \leq 1; 0 \leq \bar{\delta})$	$\hat{R}_{\tilde{w}.\tilde{x},j}$	$\hat{\mathcal{B}}_j(\lvert\bar{\delta}\rvert \leq 1)$
1	Education	0.073	[-0.0002,0.073]	0.074^{***}	[-0.0002,0.147]
	Robust s.e.	(0.004)	-	(0.002)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[0.066,0.081]	[-0.007,0.080]	-	[-0.007,0.155]
2	Experience	0.082	[0.033,0.082]	0.049^{***}	[0.033,0.130]
	Robust s.e.	(0.007)	-	(0.004)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[0.068,0.095]	[0.020,0.093]	-	[0.020,0.144]
3	$\frac{1}{100}$Experience^2	-0.213	[-0.213,-0.121]	-0.093^{***}	[-0.306,-0.121]
	Robust s.e.	(0.032)	-	(0.023)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[-0.276, -0.151]	[-0.266,-0.058]	-	[-0.373,-0.058]
4	Black indicator	-0.188	[-0.188,0.037]	-0.224^{***}	[-0.412,0.037]
	Robust s.e.	(0.017)	-	(0.012)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[-0.222,-0.154]	[-0.216,0.069]	-	[-0.449,0.069]

Notes: Y denotes the logarithm of hourly wage and X is composed of education, experience, experience squared, and binary indicators taking value 1 if a person is black, lives in the South, and lives in a metropolitan area (SMSA) respectively. Log(KWW) is used as predictive proxy W. The sample size is 2963. It's a subset of the 3010 observations used in Card (1995) and drawn from the 1976 subset of NLSYM. The estimates $\hat{\mathcal{B}}_j(\lvert\bar{\delta}\rvert \leq 1; 0 \leq \bar{\delta})$ and the corresponding $CI_{\bar{\beta}_j,.95}$ obtain under the assumption $sign(R_{\tilde{w}.\tilde{x},j}) = sign(\hat{R}_{\tilde{w}.\tilde{x},j})$. The *, **, or *** next to $\hat{R}_{\tilde{w}.\tilde{x},j}$ indicate that the p-value associated with a t-test for the null $R_{\tilde{w}.\tilde{x},j} = 0$ against the alternative hypothesis $sign(R_{\tilde{w}.\tilde{x},j}) = sign(\hat{R}_{\tilde{w}.\tilde{x},j})$ is less than 0.1, 0.05, or 0.01 respectively.

Table 2: Regression-Based Estimates of Log Wage Equation Conditioning on Covariates under Restrictions on Confounding

		$\hat{R}_{\tilde{y}.(\tilde{x}',\tilde{s}')',j}$	$\hat{\mathcal{B}}_j(\lvert\bar{\delta}\rvert \leq 1; 0 \leq \bar{\delta})$	$\hat{R}_{\tilde{w}.(\tilde{x}',\tilde{s}')',j}$	$\hat{\mathcal{B}}_j(\lvert\bar{\delta}\rvert \leq 1)$
1	Education	0.072	[0.001,0.072]	0.070***	[0.001,0.142]
	Robust s.e.	(0.004)	-	(0.002)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[0.064,0.079]	[-0.006,0.078]	-	[-0.006,0.150]
2	Experience	0.083	[0.035,0.083]	0.048***	[0.035,0.131]
	Robust s.e.	(0.007)	-	(0.004)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[0.070,0.096]	[0.022,0.094]	-	[0.022,0.145]
3	$\frac{1}{100}$Experience^2	-0.220	[-0.220,-0.133]	-0.087***	[-0.307,-0.133]
	Robust s.e.	(0.032)	-	(0.023)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[-0.283,-0.157]	[-0.273,-0.070]	-	[-0.373,-0.070]
4	Black indicator	-0.187	[-0.187,0.015]	-0.201***	[-0.388,0.015]
	Robust s.e.	(0.020)	-	(0.013)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[-0.225,-0.148]	[-0.219,0.050]	-	[-0.429,0.050]

Notes: The results extend the specification in Table 1 by conditioning on covariates S corresponding to 8 indicators for region of residence in 1966, an indicator for residence in SMSA in 1966, imputed father and mother education plus 2 indicators for missing father or mother education, 8 binary indicators for interacted parental high school, college, or post graduate education, an indicator for father and mother present at age 14, and an indicator for single mother at age 14. The remaining notes in Table 1 apply analogously here.

Table 3: Regression Estimates of Log Wage Equation Conditioning on Covariates and KWW

	Education	Experience	$\frac{1}{100}$Experience^2	Black indicator	KWW
$\hat{R}_{\tilde{y}.(\tilde{x}',\tilde{s^*}')',j}$	0.055	0.071	-0.198	-0.143	0.008
Robust s.e.	(0.005)	(0.007)	(0.032)	(0.021)	(0.001)
$CI_{.95}$	[0.046, 0.064]	[0.057,0.085]	[-0.261,-0.135]	[-0.184,-0.102]	[0.006,0.010]

Notes: This table reports estimates $\hat{R}_{\tilde{y}.(\tilde{x}',\tilde{s^*}')',j}$ from a regression of $\tilde{Y}$ on $\tilde{X}$ conditioning on covariates $\tilde{S^*}$ with $S^* = (S', KWW)'$ and Y, X, and S defined as in Tables 1 and 2.

Table 4: Regression-Based Estimates of Log Wage Equation with an Education and Race Interaction Term Conditioning on Covariates under Restrictions on Confounding

		$\hat{R}_{\tilde{y}.(\tilde{x}',\tilde{s}')',j}$	$\hat{\mathcal{B}}_j(\lvert\bar{\delta}\rvert \leq 1; 0 \leq \bar{\delta})$	$\hat{R}_{\tilde{w}.(\tilde{x}',\tilde{s}')',j}$	$\hat{\mathcal{B}}_j(\lvert\bar{\delta}\rvert \leq 1)$
1	Education	0.068	[0.004,0.068]	0.064***	[0.004,0.131]
	Robust s.e.	(0.004)	-	(0.003)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[0.060,0.076]	[-0.004,0.075]	-	[-0.004,0.140]
2	(Education-12)×Black	0.017	[-0.012,0.017]	0.029***	[-0.012,0.046]
	Robust s.e.	(0.006)	-	(0.005)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[0.005,0.030]	[-0.024,0.028]	-	[-0.024,0.060]
3	Experience	0.081	[0.036,0.081]	0.045***	[0.036,0.127]
	Robust s.e.	(0.007)	-	(0.004)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[0.068,0.095]	[0.023,0.092]	-	[0.023,0.140]
4	$\frac{1}{100}$Experience^2	-0.210	[-0.210,-0.139]	-0.070***	[-0.280,-0.139]
	Robust s.e.	(0.032)	-	(0.023)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[-0.273,-0.146]	[-0.263,-0.075]	-	[-0.347,-0.075]
5	Black indicator	-0.193	[-0.193,0.018]	-0.211***	[-0.403,0.018]
	Robust s.e.	(0.020)	-	(0.013)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[-0.231,-0.154]	[-0.225,0.055]	-	[-0.445,0.055]

Notes: The results augment the specification in Table 2 to include in X an interaction term multiplying years of education minus 12 with a black binay indicator. The remaining notes in Tables 1 and 2 apply analogously here.

Table 5: IV Regression-Based Estimates of Log Wage Equation Conditioning on Covariates under Restrictions on Confounding

		$\hat{R}_{\tilde{y}.(\tilde{x}',\tilde{s}')'\|(\tilde{z}',\tilde{s}')',j}$	$\hat{\mathcal{B}}_j(\|\bar{\delta}\| \leq 1; 0 \leq \bar{\delta})$	$\hat{R}_{\tilde{w}.(\tilde{x}',\tilde{s}')'\|(\tilde{z}',\tilde{s}')',j}$	$\hat{\mathcal{B}}_j(\|\bar{\delta}\| \leq 1)$
1	Education	0.134	[0.029,0.134]	0.106^{***}	[0.029,0.240]
	Robust s.e.	(0.052)	-	(0.032)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[0.033,0.236]	[-0.060,0.220]	-	[-0.060,0.350]
2	Experience	0.061	[0.006,0.061]	0.054^{***}	[0.006,0.115]
	Robust s.e.	(0.025)	-	(0.015)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[0.011,0.110]	[-0.036,0.102]	-	[-0.036,0.168]
3	$\frac{1}{100}$Experience^2	-0.113	[-0.113,0.009]	-0.122^{**}	[-0.235,0.009]
	Robust s.e.	(0.122)	-	(0.072)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[-0.352,0.127]	[-0.314,0.214]	-	[-0.493,0.214]
4	Black indicator	-0.162	[-0.162,0.026]	-0.189^{***}	[-0.351,0.026]
	Robust s.e.	(0.028)	-	(0.018)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[-0.218,-0.107]	[-0.209,0.075]	-	[-0.412,0.075]

Notes: The results employ the specification in Table 2 but use an indicator for whether there is a four year college in the local labor market, age, and age squared as instruments for education, experience, and experience squared. The remaining notes in Tables 1 and 2 apply analogously for the IV-based results here.

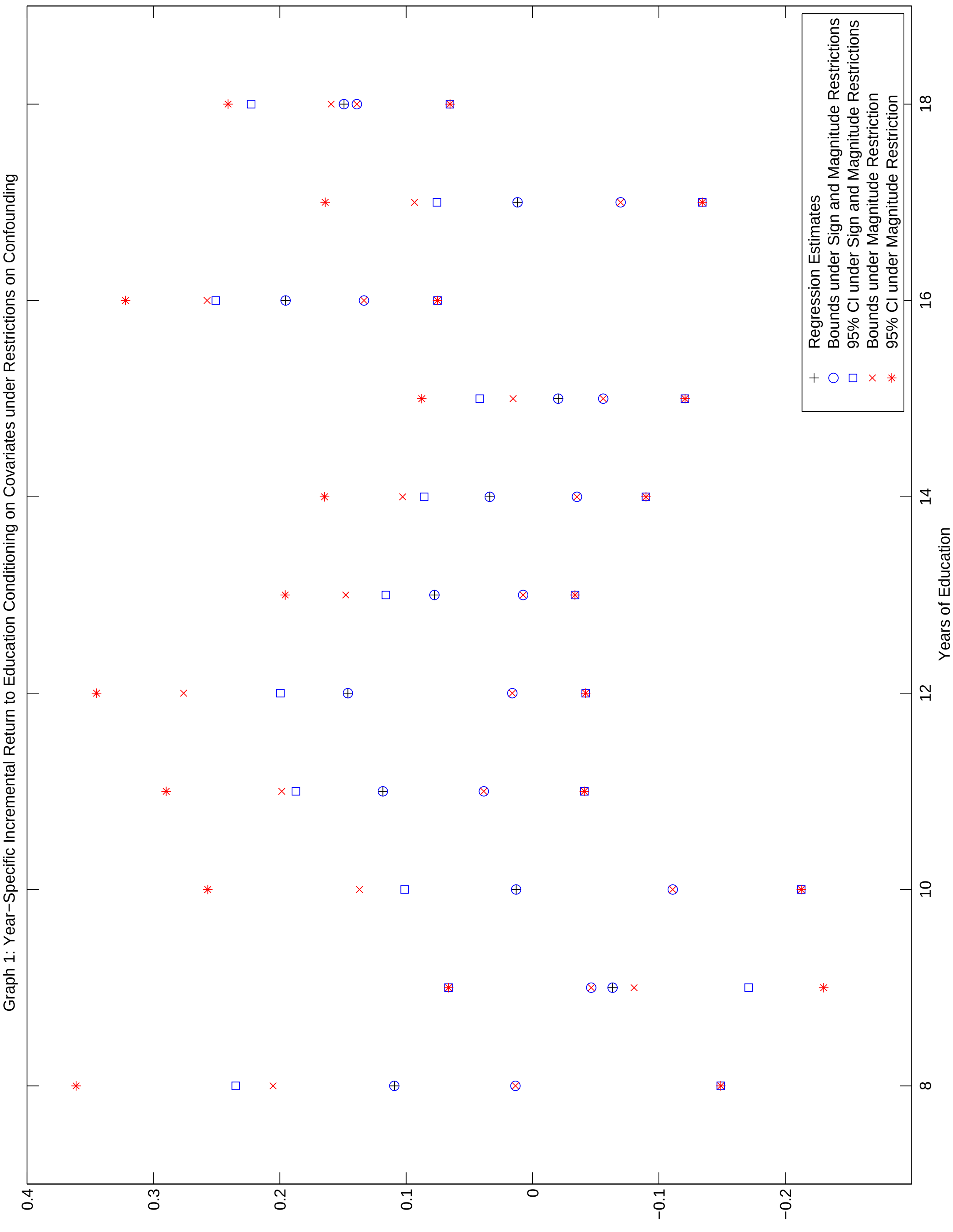
Table 6: Regression-Based Estimates of Log Wage Equation with Year-Specific Education Indicators Conditioning on Covariates under Restrictions on Confounding

		$\hat{R}_{\tilde{y}.(\tilde{x}',\tilde{s}')',j}$	$\hat{\mathcal{B}}_j(\lvert\bar{\delta}\rvert \leq 1; 0 \leq \bar{\delta})$	$\hat{R}_{\tilde{w}.(\tilde{x}',\tilde{s}')',j}$	$\hat{\mathcal{B}}_j(\lvert\bar{\delta}\rvert \leq 1)$
1	Educ $\geq$ 2 years	-0.444	[-1.438,-0.444]	0.994***	[-1.438,0.550]
	Robust s.e.	(0.059)	-	(0.175)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[-0.559,-0.329]	[-1.685,-0.347]	-	[-1.685,0.901]
2	Educ $\geq$ 3 years	-0.253	[-0.253,0.616]	-0.869***	[-1.123,0.616]
	Robust s.e.	(0.070)	-	(0.225)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[-0.390,-0.117]	[-0.368,0.970]	-	[-1.541,0.970]
3	Educ $\geq$ 4 years	0.446	[0.128,0.446]	0.318**	[0.128,0.764]
	Robust s.e.	(0.055)	-	(0.144)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[0.339,0.554]	[-0.137,0.536]	-	[-0.137,1.004]
4	Educ $\geq$ 5 years	-0.061	[-0.612,-0.061]	0.551***	[-0.612,0.490]
	Robust s.e.	(0.063)	-	(0.158)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[-0.186,0.063]	[-0.889,0.043]	-	[-0.889,0.773]
5	Educ $\geq$ 6 years	0.094	[0.078,0.094]	0.015	[0.078,0.109]
	Robust s.e.	(0.114)	-	(0.202)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[-0.130,0.317]	[-0.278,0.281]	-	[-0.278,0.514]
6	Educ $\geq$ 7 years	0.034	[-0.242,0.034]	0.276**	[-0.242,0.310]
	Robust s.e.	(0.118)	-	(0.144)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[-0.197,0.265]	[-0.527,0.228]	-	[-0.527,0.635]
7	Educ $\geq$ 8 years	0.109	[0.014,0.109]	0.096	[0.014,0.205]
	Robust s.e.	(0.076)	-	(0.060)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[-0.040,0.259]	[-0.149,0.235]	-	[-0.149,0.361]
8	Educ $\geq$ 9 years	-0.063	[-0.063,-0.046]	-0.017	[-0.080,-0.046]
	Robust s.e.	(0.065)	-	(0.047)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[-0.192,0.065]	[-0.171,0.066]	-	[-0.230,0.066]
9	Educ $\geq$ 10 years	0.013	[-0.111,0.013]	0.124***	[-0.111,0.137]
	Robust s.e.	(0.054)	-	(0.041)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[-0.092,0.118]	[-0.213,0.101]	-	[-0.213,0.257]
10	Educ $\geq$ 11 years	0.118	[0.039,0.118]	0.080***	[0.039,0.198]
	Robust s.e.	(0.042)	-	(0.031)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[0.036,0.200]	[-0.041,0.187]	-	[-0.041,0.290]
11	Educ $\geq$ 12 years	0.146	[0.016,0.146]	0.130***	[0.016,0.276]
	Robust s.e.	(0.032)	-	(0.021)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[0.082,0.210]	[-0.042,0.200]	-	[-0.042,0.345]
12	Educ $\geq$ 13 years	0.078	[0.007,0.078]	0.070***	[0.007,0.148]
	Robust s.e.	(0.023)	-	(0.014)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[0.032,0.123]	[-0.034,0.116]	-	[-0.034,0.196]
13	Educ $\geq$ 14 years	0.034	[-0.035,0.034]	0.069***	[-0.035,0.103]

	Robust s.e.	(0.032)	-	(0.016)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[-0.028,0.096]	[-0.090,0.086]	-	[-0.090,0.165]
14	Educ $\geq$ 15 years	-0.020	[-0.056,-0.020]	0.036**	[-0.056,0.015]
	Robust s.e.	(0.038)	-	(0.018)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[-0.094,0.054]	[-0.121,0.042]	-	[-0.121,0.088]
15	Educ $\geq$ 16 years	0.195	[0.133,0.195]	0.062***	[0.133,0.258]
	Robust s.e.	(0.034)	-	(0.016)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[0.130,0.261]	[0.075,0.251]	-	[0.075,0.322]
16	Educ $\geq$ 17 years	0.012	[-0.070,0.012]	0.082***	[-0.070,0.093]
	Robust s.e.	(0.039)	-	(0.014)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[-0.064,0.088]	[-0.134,0.076]	-	[-0.134,0.164]
17	Educ $\geq$ 18 years	0.149	[0.139,0.149]	0.010	[0.139,0.159]
	Robust s.e.	(0.045)	-	(0.015)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[0.062,0.237]	[0.065,0.223]	-	[0.065,0.241]
18	Experience	0.087	[0.052,0.087]	0.035***	[0.052,0.122]
	Robust s.e.	(0.008)	-	(0.005)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[0.072,0.102]	[0.038,0.099]	-	[0.038,0.137]
19	$\frac{1}{100}$Experience^2	-0.241	[-0.241,-0.227]	-0.014	[-0.256,-0.227]
	Robust s.e.	(0.037)	-	(0.024)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[-0.313,-0.169]	[-0.302,-0.158]	-	[-0.332,-0.158]
20	Black indicator	-0.178	[-0.178,0.019]	-0.196***	[-0.374,0.019]
	Robust s.e.	(0.020)	-	(0.012)	-
	$CI_{.95}$ and $CI_{\bar{\beta}_j,.95}$	[-0.216,-0.139]	[-0.210,0.054]	-	[-0.415,0.054]

Notes: The results extend the specification in Table 2 to include in X indicators for exceeding each year of education in the sample. The remaining notes in Tables 1 and 2 apply analogously here.

Graph 1: Year−Specific Incremental Return to Education Conditioning on Covariates under Restrictions on Confounding



References

Altonji, J. and C. Pierret (2001), "Employer Learning and Statistical Discrimination," *The Quarterly Journal of Economics*, 116, 313-350.

Altonji, J., T. Conley, T. Elder, and C. Taber (2011), "Methods for Using Selection on Observed Variables to Address Selection on Unobserved Variables," Yale University Department of Economics Working Paper.

Arcidiacono, P., P. Bayer, and A. Hizmo (2010), "Beyond Signaling and Human Capital: Education and the Revelation of Ability," *American Economic Journal: Applied Economics*, 2, 76–104.

Baltagi, B. (1999). *Econometrics, 2nd Edition*. Springer-Verlag, Berlin.

Bontemps, C., T. Magnac, and E. Maurin (2012), "Set Identified Linear Models," *Econometrica*, 80, 1129-1155.

Card, D. (1995), "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," In L.N. Christofides, E.K. Grant, and R. Swidinsky, editors, *Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp*. Toronto: University of Toronto Press.

Card, D. (1999), "The Causal Effect of Education on Earnings," in Ashenfelter, O. and Card, D. eds., *Handbook of Labor Economics*, vol. 3, Part A, Elsevier.

Carneiro, P. and J. Heckman (2002), "The Evidence on Credit Constraints in Post Secondary Schooling," *The Economic Journal*, 112, 705-734.

Carneiro, P., J. Heckman, and D. Masterov (2005), "Understanding the Sources of Ethnic and Racial Wage Gaps and Their Implications for Policy." In: Nelson, R and Nielsen, L, (eds.) *Handbook of Employment Discrimination Research: Rights and Realities.* Springer: Amsterdam, pp. 99 – 136.

Cawley J., J. Heckman, and E. Vytlacil (2001), "Three Observations on Wages and Measured Cognitive Ability," *Labour Economics*, 8, 419–442.

Chalak, K. (2012), "Identification without Exogeneity under Equiconfounding in Linear Recursive Structural Systems," in X. Chen and N. Swanson (eds.), *Causality, Prediction, and Specification Analysis: Recent Advances and Future Directions - Essays in Honor of Halbert L. White, Jr.*, Springer, pp. 27-55.

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

Dawid, A.P. (1979), "Conditional Independence in Statistical Theory" (with Discussion), *Journal of the Royal Statistical Society*, Series B, 41, 1-31.

Frisch, R. and F. Waugh (1933), "Partial Regressions as Compared with Individual Trends," *Econometrica*, 1, 939-953.

Fryer, R. (2011), "Racial Inequality in the 21st Century: The Declining Significance of Discrimination." In O. Ashenfelter and D. Card (eds.). *Handbook of Labor Economics.* Elsevier, 4B, pp. 855–971.

Heckman, J. and E. Vytlacil (1998), "Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling When the Return is Correlated with Schooling," *The Journal of Human Resources*, 33, 974-987.

Imbens, G. W. and C. F. Manski (2004), "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857.

Lang, K. and M. Manove (2011), "Education and Labor Market Discrimination," *American Economic Review*, 101, 1467–1496.

Manski, C. and J. Pepper (2000), "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68, 997-1010.

Mincer, J., (1974). *Schooling, Experience, and Earning.* New York: National Bureau of Economic Research.

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

Nevo, A. and A. M. Rosen (2012), "Identification With Imperfect Instruments," *Review of Economics and Statistics*, 94, 659–671.

Ogburna, E. L. and T. J. VanderWeele (2012), "On the Nondifferential Misclassification of a Binary Confounder," *Epidemiology*, 23, 433–439.

Reinhold, S. and T. Woutersen, (2009), "Endogeneity and Imperfect Instruments: Estimating Bounds for the Effect of Early Childbearing on High School Completion," University of Arizona Department of Economics Working Paper.

Shorack, G. (2000). *Probability for Statisticians.* New York: Springer-Verlag.

Stoye, J. (2009), "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 77, 1299–1315.

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. M. (1997), "On Two Stage Least Squares Estimation of the Average Treatment Effect in a Random Coefficient Model," *Economics Letters*, 56, 129–133.

Wooldridge, J. M. (2003), "Further Results on Instrumental Variables Estimation of Average Treatment Effects in the Correlated Random Coefficient Model," *Economics Letters*, 79, 185–191.

Wooldridge, J. M. (2008). *Introductory Econometrics: A Modern Approach.* South-Western College Publishing, 4th edition.