Assessing Omitted Variable Bias when the Controls are Endogenous*

Paul Diegert† Matthew A. Masten‡ Alexandre Poirier§

July 6, 2022

Abstract

Omitted variables are one of the most important threats to the identification of causal effects. Several widely used methods, including Oster (2019), have been developed to assess the impact of omitted variables on empirical conclusions. These methods all require an exogenous controls assumption: the omitted variables must be uncorrelated with the included controls. This is often considered a strong and implausible assumption. We provide a new approach to sensitivity analysis that allows for endogenous controls, while still letting researchers calibrate sensitivity parameters by comparing the magnitude of selection on observables with the magnitude of selection on unobservables. We illustrate our results in an empirical study of the effect of historical American frontier life on modern cultural beliefs. Finally, we implement these methods in the companion Stata module `regsensitivity` for easy use in practice.

JEL classification: C18; C21; C51

Keywords: Treatment Effects, Sensitivity Analysis, Unconfoundedness

*This paper was presented at Northwestern, Duke, Oxford, Brown, Texas A&M, the joint Bonn-Mannheim seminar, and the SEA 2021 conference. We thank audiences at those seminars and conferences for helpful conversations and comments. We thank Hongchang Guo for excellent research assistance. Masten thanks the National Science Foundation for research support under Grant 1943138.

†Department of Economics, Duke University, `paul.diegert@duke.edu`

‡Department of Economics, Duke University, `matt.masten@duke.edu`

§Department of Economics, Georgetown University, `alexandre.poirier@georgetown.edu`

1 Introduction

Angrist and Pischke (2015, page 74) argue that "careful reasoning about OVB [omitted variables bias] is an essential part of the 'metrics game." Largely for this reason, researchers have eagerly adopted new tools that let them quantitatively assess the impact of omitted variables on their results. In particular, researchers now widely use the sensitivity analysis methods developed in Altonji, Elder, and Taber (2005) and Oster (2019). These methods have been extremely influential, with about 3660 and 2560 Google Scholar citations as of July 2022, respectively. Looking beyond citations, researchers are actively using these methods. Every top 5 journal in economics is now regularly publishing papers which use the methods in Oster (2019). Aggregating across these five journals, from the three year period starting when Oster (2019) was published, 2019–2021, 33 published papers have used these methods, and often quite prominently.

These methods, however, rely on an assumption called *exogenous controls*. This assumption states that the omitted variables of concern are uncorrelated with all included covariates. For example, consider a classic regression of wages on education and controls like parental education. Typically we are worried that the coefficient on education in this regression is a biased measure of the returns to schooling because unobserved ability is omitted. To apply Oster's (2019) method for assessing the importance of this unobserved variable, we must assume that unobserved ability is uncorrelated with parent's education, along with all other included controls.[1]

Such exogeneity assumptions on the control variables are usually thought to be very strong and implausible. For example, Angrist and Pischke (2017) discuss how

> "The modern distinction between causal and control variables on the right-hand side of a regression equation requires more nuanced assumptions than the blanket statement of regressor-error orthogonality that's emblematic of the traditional econometric presentation of regression." (page 129)

Put differently: We usually do not expect the included control variables to be uncorrelated with the omitted variables; instead we merely hope that the treatment variable is uncorrelated with the omitted variables after adjusting for the included controls. These control variables are therefore usually thought to be endogenous.

In this paper we provide a new approach to sensitivity analysis that allows the included control variables to be endogenous, unlike Altonji et al. (2005) and Oster (2019). Like these previous papers, however, we measure the importance of unobserved variables by comparing them to the included covariates. Thus researchers can still measure how strong selection on unobservables must be relative to selection on observables in order to overturn their baseline findings.

Overview of Our Approach

In section 2 we describe the baseline model. The parameter of interest is β_{long}, the coefficient on a treatment variable X in a long OLS regression of an outcome variable Y on a constant, treatment X, the observed covariates W_1, and some unobserved covariates W_2. In section 4 we discuss three causal models which allow us to interpret this parameter causally, based on three different identification strategies: unconfoundedness, difference-in-differences, and instrumental variables. Since W_2 is unobserved, we cannot compute the long regression of Y on $(1, X, W_1, W_2)$ in the data. Instead, we can only compute the medium regression of Y

[1] Appendix A.1 in Oster (2019) briefly describes one approach to relaxing the exogenous controls assumption in her setting. We show that this approach changes the interpretation of her sensitivity parameter in a way that may change researchers' conclusions about the robustness of their results. We discuss this in detail in our appendix A.

on $(1, X, W_1)$. We begin by considering a baseline model with a no selection on unobservables assumption, which implies that the coefficients on X in the long and medium regressions are the same. Importantly, this baseline model does *not* assume the controls W_1 are exogenous. They are included solely to aid in identification of the coefficient on X in the long regression.

We then move to assess the importance of the no selection on unobservables assumption. In section 3 we develop a sensitivity analysis that does not rely on the exogenous controls assumption, while still allowing researchers to compare the magnitude of selection on observables with the magnitude of selection on unobservables. Our results use either one, two, or three sensitivity parameters; not all of our results require all three parameters. The first sensitivity parameter compares the relative magnitude of the coefficients on the observed and unobserved covariates in a treatment selection equation. This parameter thus measures the magnitude of selection on unobservables by comparing it with the magnitude of selection on observables. The second sensitivity parameter compares the relative magnitude of the coefficients on the observed and unobserved covariates in the outcome equation. The third sensitivity parameter controls the relationship between the observed and the unobserved covariates; this parameter thus measures the magnitude of control endogeneity.

We provide three main identification results. Our first result (Theorem 2) characterizes the identified set for β_{long}, the coefficient on treatment in the long regression of the outcome on the treatment and the observed and unobserved covariates. This theorem only requires that researchers make an assumption about a single sensitivity parameter—the relative magnitudes of selection on observables and unobservables. In contrast, Oster (2019) requires that researchers reason about two different sensitivity parameters. Moreover, our result allows for arbitrarily endogenous controls, unlike existing results in the literature. We provide a closed form, analytical expression for the identified set, which makes this result easy to use in practice. Using this result, we show how to do breakdown analysis: To find the largest magnitude of selection on unobservables relative to observables needed to overturn a specific baseline finding. This value is called a breakdown point, and can be used to measure the robustness of one's baseline results. We provide a simple expression for the breakdown point and recommend that researchers report estimates of it along with their baseline estimates. This estimated breakdown point provides a scalar summary of a study's robustness to selection on unobservables while allowing for arbitrarily endogenous controls.

If researchers are willing to partially restrict the magnitude of control endogeneity, then their results will be more robust to selection on unobservables. Our second result (Theorem 3) therefore characterizes the identified set for β_{long} when researchers make an assumption about two sensitivity parameters: the relative magnitude of selection on unobservables and the magnitude of control endogeneity. We again provide a simple closed form expression for the identified set, and then show how to use this result to do breakdown analysis. Finally, if researchers are willing to restrict the impact of unobservables on outcomes, then they can again obtain results that are more robust to selection on unobservables. In this case, the identified set is more difficult to characterize analytically (see Theorem 5 in the appendix). However, our third main result (Theorem 4) shows that we can nonetheless easily numerically compute objects like breakdown points. We also note that the identified set can be computed easily as well.

In section 5 we show how to use our results in empirical practice. We use data from Bazzi, Fiszbein, and Gebresilasse (2020, *Econometrica*) who studied the effect of historical American frontier life on modern cultural beliefs. Specifically, they test a well known conjecture that living on the American frontier cultivated individualism and antipathy to government intervention. They heavily rely on Oster's (2019) method to argue against the importance of omitted variables. Using our results, we obtain more nuanced conclusions about robustness. In particular, when allowing for endogenous controls, we find that effects obtained from questionnaire based outcomes are no longer robust, but effects from election and property tax outcomes remain robust. This analysis highlights that previous empirical findings of robustness based on Oster (2019), for example, may no longer be robust once the controls are allowed to be endogenous.

Related Literature

We conclude this section with a brief review of the literature. We focus on two literatures: The literature on endogenous controls and the literature on sensitivity analysis in linear or parametric models.

The idea that the treatment variable and the control variables should be treated asymmetrically in the assumptions goes back to at least Barnow, Cain, and Goldberger (1980). They developed the "linear control function estimator", which is based on an early parametric version of the now standard unconfoundedness assumption. Heckman and Robb (1985, page 190), Heckman and Hotz (1989), and Heckman and Vytlacil (2007, page 5035) all provide detailed discussions of this estimator. It was also one of the main estimators used in LaLonde (1986). Stock and Watson (2011) provide a textbook description of it on pages 230–233 and pages 250–251. Angrist and Pischke (2009) also discuss it at the end of their section 3.2.1. Also see Frölich (2008). Note that this earlier analysis was based on mean independence assumptions, while the analysis in our paper only uses linear projections. We do this so that our baseline model is not falsifiable, which allows us to avoid complications that arise in falsifiable models (e.g., see Masten and Poirier 2021). More recently, Hünermund and Louw (2020) remind researchers that most control variables are likely endogenous and hence their coefficients should not be interpreted as causal.

Although control variables are often thought to be endogenous, the literature on sensitivity analysis generally assumes the controls are exogenous. As mentioned earlier, this includes Altonji, Elder, and Taber (2005, 2008) and Oster (2019). However, Appendix A.1 of Oster (2019) describes one approach for relaxing the exogenous controls assumption by redefining her sensitivity parameter δ. We discuss this approach in detail in appendix A. There we show that such a redefinition implies that $\delta = 1$ is no longer a natural reference point. In particular, we show that this redefinition can change researchers' conclusions about the robustness of their results. This follows because, under endogenous controls, δ does not solely measure selection on unobservables when the controls are endogenous. Krauth (2016) explicitly allows for endogenous controls, but he relies on a similar redefinition approach as Oster (2019), which has similar drawbacks. See Appendix A.3 for more discussion. Cinelli and Hazlett (2020) develop an alternative to Oster (2019) that allows researchers to compare the relative strength of the observed and unobserved covariates on outcomes and on treatment selection. Their approach to calibration also imposes the exogenous controls assumption (see the last paragraph of their page 53, in their section 4.4). Like Oster (2019) and Krauth (2016), they also briefly mention a redefinition approach to allow for endogenous controls; in Appendix A.3 we show that it too has similar drawbacks as Oster's (2019) redefinition approach. Imbens (2003) starts from the standard unconfoundedness assumption which allows endogenous controls, but in his parametric assumptions (see his likelihood equation on page 128) he assumes that the unobserved omitted variable is independent of the observed covariates. Altonji, Conley, Elder, and Taber (2019) propose an approach to allow for endogenous controls based on imposing a factor model on all covariates, observable and unobservable. Their approach and ours are not nested; in particular, their results require the number of covariates to go to infinity, while we suppose the number of covariates is fixed. This difference arises because they explicitly model the covariate selection process. We instead take the covariates as given and impose assumptions directly on these covariates, rather than deriving such assumptions from a model of covariate selection. Our results also allow researchers to be fully agnostic about the relationship between the observed and unobserved covariates.

There are several other related papers on sensitivity analysis. The sensitivity parameters we use are defined based on the relative magnitude of different coefficients. That is similar to previous work by Chalak (2019), who shows how to use relative magnitude constraints to assess sensitivity to omitted variables when a proxy for the omitted variable is observed. Zhang, Cinelli, Chen, and Pearl (2021, section 7.3) discuss a sensitivity analysis that uses constraints on the relative magnitude of coefficients in a setting with exogenous controls. Finally, note that the standard unconfoundedness assumption (for example, chapter 12 in Imbens and Rubin 2015) allows for endogenous controls. For this reason, several papers that assess sensitivity to unconfoundedness also allow for endogenous controls. This includes Rosenbaum (1995, 2002), Masten and Poirier (2018), and Masten, Poirier, and Zhang (2021). These methods, however, do not provide formal results for calibrating their sensitivity parameters based on comparing selection on observables with selection on unobservables. These methods also focus on binary or discrete treatments, whereas the analysis in our paper can be used for continuous treatments as well.

Notation Remark

For random vectors A and B, let $\operatorname{cov}(A,B)$ be the $\dim(A) \times \dim(B)$ matrix whose (i,j)th element is $\operatorname{cov}(A_i, B_j)$. Define $A^{\perp B} = A - [\operatorname{var}(B)^{-1} \operatorname{cov}(B,A)]'B$. This is the sum of the residual from a linear projection of A onto $(1,B)$ and the intercept in that projection. Many of our equations therefore do not include intercepts because they are absorbed into $A^{\perp B}$ by definition. Note also that $A^{\perp B}$ is uncorrelated with each component of B, by definition. Let $R^2_{A \sim B \bullet C}$ denote the R-squared from a regression of $A^{\perp C}$ on $(1, B^{\perp C})$. This is sometimes called the partial R-squared.

2 The Baseline Model

Let Y and X be observed scalar variables. Let W_1 be a vector of observed covariates of dimension d_1. Let W_2 be an unobserved scalar covariate; we discuss vector W_2 in appendix B. Let $W = (W_1, W_2)$. Consider the OLS estimand of Y on $(1, X, W_1, W_2)$. Let $(\beta_{\text{long}}, \gamma_1, \gamma_2)$ denote the coefficients on (X, W_1, W_2). The following assumption ensures these coefficients and other OLS estimands we consider are well defined.

Assumption A1. The variance matrix of (Y, X, W_1, W_2) is finite and positive definite.

We can write

$$Y = \beta_{\text{long}} X + \gamma_1' W_1 + \gamma_2 W_2 + Y^{\perp X,W} \tag{1}$$

where $Y^{\perp X,W}$ is defined to be the OLS residual plus the intercept term, and hence is uncorrelated with each component of (X, W) by construction. Suppose our parameter of interest is β_{long}. In section 4 we discuss three causal models that lead to this specific OLS estimand as the parameter of interest, using either unconfoundedness, difference-in-differences, or instrumental variables as an identification strategy. Alternatively, it may be that we are simply interested in β_{long} as a descriptive statistic. The specific motivation for interest in β_{long} does not affect our technical analysis.

Next consider the OLS estimand of X on $(1, W_1, W_2)$. Let (π_1, π_2) denote the coefficients on (W_1, W_2). Then we can write

$$X = \pi_1' W_1 + \pi_2 W_2 + X^{\perp W} \tag{2}$$

where $X^{\perp W}$ is defined to be the OLS residual plus the intercept term, and hence is uncorrelated with each component of W by construction. Although equation (2) is not necessarily causal, we can think of the value of π_1 as representing "selection on observables" while π_2 represents "selection on unobservables." The following is thus a natural baseline assumption.

Assumption A2 (No selection on unobservables). $\pi_2 = 0$.

Let β_{med} denote the coefficient on X in the OLS estimand of Y on $(1, X, W_1)$. With no selection on unobservables, we have the following result.

Theorem 1. Suppose the joint distribution of (Y, X, W_1) is known. Suppose A1 and A2 hold. Then the following hold.

1. $\beta_{\text{long}} = \beta_{\text{med}}$. Consequently, β_{long} is point identified.
2. The identified set for γ_1 is $\mathbb{R}^{d_1}$.

This result allows for endogenous controls, in the sense that the observed covariates W_1 can be arbitrarily correlated with the unobserved covariate W_2. But it restricts the relationship between (X, W_1, W_2) in such a way that we can still point identify β_{long} even though W_1 and W_2 are arbitrarily correlated. The coefficient γ_1 on the observed controls, however, is completely unidentified. The difference between the roles of X and W_1 in Theorem 1 reflects the sentiment of Angrist and Pischke (2017) that we discussed in the introduction.

In practice, we are often worried that the no selection on unobservables assumption A2 does not hold. In section 3 we develop a new approach to assess the importance of this assumption.

3 Sensitivity Analysis

We have argued that, in practice, most controls are endogenous in the sense that they are potentially correlated with the omitted variables of concern. Consequently, methods for assessing the sensitivity of identifying assumptions for the treatment variable of interest should allow for the controls to be endogenous to some extent. In this section, we develop such a method. In section 3.1 we first describe the three sensitivity parameters that we use; note that we do not use all of these parameters in all of our results. In section 3.2 we then state our main identification results. In section 3.3 we make several remarks regarding interpretation of the sensitivity parameters. Finally, in section 3.4 we briefly discuss estimation and inference.

3.1 The Sensitivity Parameters

Recall from section 2 that our parameter of interest is β_{long}, the OLS coefficient on X in the long regression of Y on $(1, X, W_1, W_2)$. Since W_2 is not observed, we cannot compute this regression from the data. Instead, we can compute β_{med}, the OLS coefficient on X in the medium regression of $(1, X, W_1)$. The difference between these two regression coefficients is given by the classic omitted variable bias formula. We can write this formula as a function of the coefficient γ_2 on W_2 in the long regression outcome equation (1) and the coefficient π_2 on W_2 in the selection equation (2) as follows:

$$\beta_{\text{med}} - \beta_{\text{long}} = \frac{\gamma_2 \pi_2 (1 - R^2_{W_2 \sim W_1})}{\text{var}(X^{\perp W_1})} \tag{3}$$

where $R^2_{W_2 \sim W_1}$ denotes the population R^2 in a linear regression of the unobservable W_2 on the observables $(1, W_1)$. This bias is a function of the coefficient π_2. Hence π_2 is a natural sensitivity parameter. Using π_2 as a sensitivity parameter, however, would require researchers to make judgment calls about the absolute magnitude of this coefficient. This may be difficult. So, similar to the definition of Oster's δ (which we review in appendix A.2), we instead define a *relative* sensitivity parameter. Specifically, let $\| \cdot \|_{\Sigma_{\text{obs}}}$ denote the weighted Euclidean norm on $\mathbb{R}^{d_1}$: $\|a\|_{\Sigma_{\text{obs}}} = (a'\Sigma_{\text{obs}} a)^{1/2}$ where $\Sigma_{\text{obs}} \equiv \text{var}(W_1)$.

We then consider the following assumption.

Assumption A3. $|\pi_2| \leq \bar{r}_X \|\pi_1\|_{\Sigma_{\text{obs}}}$ for a known value of $\bar{r}_X \geq 0$.

To interpret assumption A3, we first normalize the variance of the unobserved W_2 to 1.

Assumption A4. $\text{var}(W_2) = 1$.

Using this normalization A4, assumption A3 can be written as

$$\sqrt{\text{var}(\pi_2 W_2)} \leq \bar{r}_X \cdot \sqrt{\text{var}(\pi_1' W_1)}.$$

So assumption A3 says that the association between treatment X and a one standard deviation increase in the index of unobservables is at most $\bar{r}_X$ times the association between treatment and a one standard deviation increase in the index of observables. Note that $\|\pi_1\|_{\Sigma_{\text{obs}}}$ is invariant to invertible linear transformations of W_1, including rescalings, since the index $\pi_1' W_1$ is invariant with respect to invertible linear transformations. This invariance ensures that $\bar{r}_X$ is a unit-free sensitivity parameter. We also explain how $\bar{r}_X$ is related to Oster's δ in Proposition 4 in appendix A.2.

The baseline model of section 2 corresponds to the case $\bar{r}_X = 0$, since it implies $\pi_2 = 0$. We relax the baseline model by considering values $\bar{r}_X > 0$. Our first main result (Theorem 2) describes the identified set using only A3. Researchers may also be willing to make additional restrictions so we consider two additional sensitivity parameters as well. These parameters are also motivated by the omitted variables bias formula in equation (3). The bias is a function of γ_2 so it is natural to also consider assumptions that restrict this parameter. Like our assumption on π_2, we consider a restriction on the relative magnitudes of γ_1 and γ_2, the coefficients of W_1 and W_2 in the outcome equation (1).

Assumption A5. $|\gamma_2| \leq \bar{r}_Y \|\gamma_1\|_{\Sigma_{\text{obs}}}$ for a known value of $\bar{r}_Y \geq 0$.

Maintaining the normalization A4, A5 has a similar interpretation as A3: It says that the association between the outcome and a one standard deviation increase in the index of unobservables is at most $\bar{r}_Y$ times the association between the outcome and a one standard deviation increase in the index of observables. $\|\gamma_1\|_{\Sigma_{\text{obs}}}$ is also invariant to invertible linear transformations of W_1 and hence $\bar{r}_Y$ is also a unit-free sensitivity parameter.

Finally, the omitted variable bias in equation (3) is a function of $R^2_{W_2 \sim W_1}$. So we also consider restrictions directly on the relationship between the observed and unobserved covariates.

Assumption A6. $R_{W_2 \sim W_1} \leq \bar{c}$ for a known value of $\bar{c} \in [0, 1]$.

Assumption A6 allows researchers to constrain the magnitude of control endogeneity. In particular, the exogenous controls assumption is equivalent to $R_{W_2 \sim W_1} = 0$ and hence can be obtained by setting $\bar{c} = 0$. Values $\bar{c} > 0$ allow for partially endogenous controls. Note that $R_{W_2 \sim W_1}$ is invariant to invertible linear transformations of W_1 as well as to the normalization on W_2. Finally, it will sometimes be useful to note that $R^2_{W_2 \sim W_1} = \| \operatorname{cov}(W_1, W_2)\|^2_{\Sigma^{-1}_{\text{obs}}}$. For interested readers, in remark 2 on page 76, we discuss the more general assumption $R_{W_2 \sim W_1} \in [\underline{c}, \overline{c}]$ for known $\underline{c}$ and $\overline{c}$ satisfying $0 \leq \underline{c} \leq \overline{c} \leq 1$ and show how to generalize our identification results to accommodate this assumption.

3.2 Identification

In this section we state our main results. For simplicity, we first normalize the treatment variable so that $\operatorname{var}(X) = 1$ and the covariates so that $\operatorname{var}(W_1) = I$. All of the results below can be rewritten without these normalizations, at the cost of additional notation. With these normalizations, $\| \cdot \|_{\Sigma^{-1}_{\text{obs}}} = \| \cdot \|_{\Sigma_{\text{obs}}}$. We use $\| \cdot \|$ to refer to this norm throughout.

Identification Using the $\bar{r}_X$ Restriction Only

Let $\mathcal{B}_I(\bar{r}_X)$ denote the identified set for β_{long} under the positive definite variance assumption A1, the normalization assumption A4, and the restriction A3 on π_2. In particular, this identified set does not impose the restriction A5 on γ_2 or the restriction A6 on $R^2_{W_2 \sim W_1}$. Let

$$\underline{B}(\bar{r}_X) = \inf \mathcal{B}_I(\bar{r}_X) \qquad \text{and} \qquad \overline{B}(\bar{r}_X) = \sup \mathcal{B}_I(\bar{r}_X)$$

denote its greatest lower bound and least upper bound. Our first main result, Theorem 2 below, provides simple, closed form expressions for these sharp bounds. Similarly, let $\mathcal{B}_I(\bar{r}_X, \bar{c})$ denote the identified set for β_{long} if we also impose A6. Let

$$\underline{B}(\bar{r}_X, \bar{c}) = \inf \mathcal{B}_I(\bar{r}_X, \bar{c}) \qquad \text{and} \qquad \overline{B}(\bar{r}_X, \bar{c}) = \sup \mathcal{B}_I(\bar{r}_X, \bar{c})$$

denote its greatest lower bound and least upper bound. Our second main result, Theorem 3 below, similarly provides simple, closed form expressions for these sharp bounds.

Let

$$k_0 = \operatorname{var}(X^{\perp W_1}) > 0, \qquad k_1 = \operatorname{cov}(Y, X^{\perp W_1}), \qquad \text{and} \qquad k_2 = \operatorname{var}(Y^{\perp W_1}) > 0.$$

The inequalities here follow from A1, positive definiteness of $\operatorname{var}(Y, X, W_1)$. (k_0, k_1, k_2) are the elements of the covariance matrix $\operatorname{var}(Y^{\perp W_1}, X^{\perp W_1})$. Moreover, note that the coefficient on X in the medium OLS regression of Y on $(1, X, W_1)$ can be written as $\beta_{\text{med}} = k_1/k_0$ by the FWL theorem. Below we will see that, after normalizing treatment and the covariates, the bounds only depend on the medium regression coefficient and the variance in outcomes and treatment after projecting out the observable covariates.

Before formally stating our first main result, we give a sketch derivation. We start by observing that having a known bound $\bar{r}_X$ on selection on unobservables relative to observables in A3 is equivalent to the existence of an $r_X \in \mathbb{R}^{d_1}$ such that $\pi_2 = \pi_1' r_X$ with $\|r_X\| \leq \bar{r}_X$ (Lemma 1 in appendix D). If we knew this r_X along with $c = \operatorname{cov}(W_1, W_2)$, then the observed first stage relationship $\operatorname{cov}(X, W_1)$ would be enough to point identify π_2 as

$$\pi_2 = \frac{r_X' \operatorname{cov}(W_1, X)}{1 + r_X' c}.$$

Define

$$z_X(r_X, c) = \frac{r_X' \operatorname{cov}(W_1, X)}{1 + r_X' c} \cdot \sqrt{1 - \|c\|^2}$$

as a scaled version of this point identified value of π_2. Then we can write the omitted variables bias equation

(3) as

$$(\beta_{\text{med}} - \beta_{\text{long}})^2 = \frac{z_X(r_X, c)^2 \gamma_2^2 (1 - \|c\|^2)}{k_0^2}$$

where we squared both sides to simplify derivations. We then note that the residual variance of outcomes constrains the second two terms in the numerator:

$$\begin{aligned} k_2 &\equiv \text{var}(Y^{\perp W_1}) \\ &= \gamma_2^2(1 - \|c\|^2) + 2\beta_{\text{long}} k_1 - \beta_{\text{long}}^2 k_0 + \text{var}(Y^{\perp X, W}) \end{aligned}$$

and hence

$$\gamma_2^2(1 - \|c\|^2) \leq k_2 - 2\beta_{\text{long}} k_1 + \beta_{\text{long}}^2 k_0$$

since $\text{var}(Y^{\perp X,W}) \geq 0$. Substituting this expression into the omitted variable bias formula gives

$$(\beta_{\text{med}} - \beta_{\text{long}})^2 \leq \frac{z_X(r_X, c)^2 (k_2 - 2\beta_{\text{long}} k_1 + \beta_{\text{long}}^2 k_0)}{k_0^2}.$$

Using $\beta_{\text{med}} = k_1/k_0$ we can then rearrange this inequality to obtain

$$(\beta_{\text{med}} - \beta_{\text{long}})^2 \leq \frac{z_X(r_X, c)^2 \left(\frac{k_2}{k_0} - \beta_{\text{med}}^2\right)}{k_0 - z_X(r_X, c)^2}.$$

The right hand side depends on the unknown values of r_X and c, but it is monotonically increasing in $z_X(r_X, c)$, so we can obtain bounds by maximizing this function over all (r_X, c) with $\|r_X\| \leq \bar{r}_X$ and $\|c\| \leq \bar{c}$. This maximum value is

$$\bar{z}_X(\bar{r}_X) = \begin{cases} \frac{\bar{r}_X}{\sqrt{1 - \bar{r}_X^2}} \sqrt{1 - k_0} & \text{if } \bar{r}_X < 1 \\ +\infty & \text{if } \bar{r}_X \geq 1. \end{cases}$$

Considering the first case, $\bar{r}_X < 1$, we obtain

$$(\beta_{\text{med}} - \beta_{\text{long}})^2 \leq \text{dev}(\bar{r}_X)^2$$

where we defined

$$\text{dev}(\bar{r}_X) = \sqrt{\frac{\bar{z}_X(\bar{r}_X)^2 (\frac{k_2}{k_0} - \beta_{\text{med}}^2)}{k_0 - \bar{z}_X(\bar{r}_X)^2}}.$$

And note that we need $\bar{z}_X(\bar{r}_X)^2 < k_0$ to avoid division by zero. Theorem 2 below formalizes these derivations. Our full proof is in appendix E, including our sharpness proof. For this theorem, we also use the following assumption.

Assumption A7. $\text{cov}(W_1, Y) \neq \text{cov}(W_1, X) \, \text{cov}(X, Y)$ and $\text{cov}(W_1, X) \neq 0$.

This assumption is not necessary, but it simplifies the proofs. A sufficient condition for A7 is $\beta_{\text{short}} \neq \beta_{\text{med}}$, where β_{short} is the coefficient on X in the short OLS regression of Y on $(1, X)$. This follows from $\beta_{\text{med}} - \beta_{\text{short}} = \text{cov}(W_1, X)'(\text{cov}(W_1, Y) - \text{cov}(W_1, X) \, \text{cov}(X, Y))$. We can now state our first main result.

Theorem 2. Suppose the joint distribution of (Y, X, W_1) is known. Suppose A1, A3, A4, and A7 hold.

Normalize $\text{var}(X) = 1$ and $\text{var}(W_1) = I$. If $\bar{z}_X(\bar{r}_X)^2 < k_0$, then

$$\underline{B}(\bar{r}_X) = \beta_{\text{med}} - \text{dev}(\bar{r}_X) \qquad \text{and} \qquad \overline{B}(\bar{r}_X) = \beta_{\text{med}} + \text{dev}(\bar{r}_X).$$

Otherwise, $\underline{B}(\bar{r}_X) = -\infty$ and $\overline{B}(\bar{r}_X) = +\infty$.

Theorem 2 characterizes the largest and smallest possible values of β_{long} when some selection on unobservables is allowed, the observed covariates are allowed to be arbitrarily correlated with the unobserved covariate, and we make no restrictions on the coefficients in the outcome equation. In fact, we prove that, with the exception of at most three singletons, the interval $[\underline{B}(\bar{r}_X), \overline{B}(\bar{r}_X)]$ is the identified set for β_{long} under these assumptions. Here we focus on the smallest and largest elements to avoid technical digressions that are unimportant for applications.

There are two important features of Theorem 2: First, it only requires researchers to reason about *one* sensitivity parameter, unlike some existing approaches, including Oster (2019). Second, and also unlike those results, it allows for arbitrarily endogenous controls. So this result allows researchers to examine the impact of selection on unobservables on their baseline results without also having to reason about the magnitude of endogenous controls.

Since Theorem 2 provides explicit expressions for the bounds, we can immediately derive a few of their properties. First, when $\bar{r}_X = 0$, the bounds collapse to β_{med}, the point estimand from the baseline model with no selection on unobservables. So we recover the baseline model as a special case. For small values of $\bar{r}_X > 0$, the bounds are no longer a singleton, but their length increases continuously as $\bar{r}_X$ increases away from zero. After normalizing treatment and the observed covariates, the rate at which they increase depends on just a few features of the data: The three elements of $\text{var}(Y^{\perp W_1}, X^{\perp W_1})$. We also see that the bounds are symmetric around β_{med}. Finally, the bounds can only be finite if $\bar{r}_X < 1$. We discuss interpretation of the magnitude of $\bar{r}_X$ in detail in section 3.3.

In practice, researchers often ask:

> How strong does selection on unobservables have to be relative to selection on observables in order to overturn our baseline findings?

We can use Theorem 2 to answer this question. Suppose in the baseline model we find $\beta_{\text{med}} \geq 0$. We are concerned, however, that $\beta_{\text{long}} \leq 0$, in which case our positive finding is driven solely by selection on unobservables. Define

$$\bar{r}_X^{\text{bp}} = \sup\{\bar{r}_X \geq 0 : b \geq 0 \text{ for all } b \in \mathcal{B}_I(\bar{r}_X)\}.$$

This value is called a *breakdown point.* It is the largest amount of selection on unobservables we can allow for while still concluding that β_{long} is nonnegative. Note that the breakdown point when $\beta_{\text{med}} \leq 0$ can be defined analogously.

Corollary 1. Suppose the assumptions of Theorem 2 hold. Then

$$\bar{r}_X^{\text{bp}} = \left(\frac{R^2_{Y \sim X \bullet W_1}}{\frac{R^2_{X \sim W_1}}{1 - R^2_{X \sim W_1}} + R^2_{Y \sim X \bullet W_1}} \right)^{1/2}.$$

The breakdown point described in Corollary 1 characterizes the magnitude of selection on unobservables relative to selection on observables needed to overturn one's baseline findings. One of our main recommendations is that researchers present estimates of this point as a scalar measure of the robustness of their results. We illustrate this recommendation in our empirical application in section 5.

Corollary 1 explicitly shows that this breakdown point depends on just two features of the observed data: The relationship between treatment and the outcome, after adjusting for the observed covariates, and the first stage relationship between treatment and the observed covariates. In particular, as the covariate adjusted relationship between outcomes and treatment strengthens, the breakdown point increases too. In contrast, as the relationship between treatment and covariates strengthens, the breakdown point decreases. This follows since we are using effects of the observed covariates to calibrate the magnitude of selection on unobservables. So when these observed covariates are strongly related to treatment, we need relatively less selection on unobservables to overturn our baseline findings. We discuss this point more in section 3.3.

Identification Using the $\bar{r}_X$ and $\bar{c}$ Restrictions

In some applications, the bounds in Theorem 2 may be quite large, even for small values of $\bar{r}_X$. In this case, researchers may be willing to restrict the relationship between the observed covariates and the omitted variable. So next we present a similar result, but now imposing A6. Let

$$\bar{z}_X(\bar{r}_X, \bar{c}) = \begin{cases} \dfrac{\bar{r}_X \sqrt{1 - \min\{\bar{c}, \bar{r}_X\}^2}}{1 - \bar{r}_X \min\{\bar{c}, \bar{r}_X\}} \sqrt{1 - k_0} & \text{if } \bar{r}_X \bar{c} < 1 \\ +\infty & \text{if } \bar{r}_X \bar{c} \geq 1. \end{cases}$$

Note that for $\bar{c} = 1$, $\bar{z}_X(\bar{r}_X, 1) = \bar{z}_X(\bar{r}_X)$ for all $\bar{r}_X \geq 0$. Also, $\bar{z}_X(\bar{r}_X, \bar{c}) = \bar{z}_X(\bar{r}_X)$ when $\bar{r}_X \leq \bar{c}$. As before, the sensitivity parameters will only affect the bounds via this function.

Theorem 3. Suppose the joint distribution of (Y, X, W_1) is known. Suppose A1, A3, A4, A6, and A7 hold. Normalize $\text{var}(X) = 1$ and $\text{var}(W_1) = I$. If $\bar{z}_X(\bar{r}_X, \bar{c})^2 < k_0$, then

$$\underline{B}(\bar{r}_X, \bar{c}) = \beta_{\text{med}} - \text{dev}(\bar{r}_X, \bar{c}) \qquad \text{and} \qquad \overline{B}(\bar{r}_X, \bar{c}) = \beta_{\text{med}} + \text{dev}(\bar{r}_X, \bar{c})$$

where

$$\text{dev}(\bar{r}_X, \bar{c}) = \sqrt{\frac{\bar{z}_X(\bar{r}_X, \bar{c})^2 \left(\frac{k_2}{k_0} - \beta_{\text{med}}^2\right)}{k_0 - \bar{z}_X(\bar{r}_X, \bar{c})^2}}.$$

Otherwise, $\underline{B}(\bar{r}_X, \bar{c}) = -\infty$ and $\overline{B}(\bar{r}_X, \bar{c}) = +\infty$.

The interpretation of Theorem 3 is similar to our earlier result Theorem 2. It characterizes the largest and smallest possible values of β_{long} when some selection on unobservables is allowed and the controls are allowed to be partially but not arbitrarily endogenous. We also make no restrictions on the coefficients in the outcome equation. As before, with the exception of at most three singletons, the interval $[\underline{B}(\bar{r}_X, \bar{c}), \overline{B}(\bar{r}_X, \bar{c})]$ is the identified set for β_{long} under these assumptions.

Earlier we saw that $\bar{r}_X < 1$ is necessary for the bounds of Theorem 2 to be finite. Theorem 3 shows that, if we are willing to restrict the value of $\bar{c}$, then we can allow for $\bar{r}_X > 1$ while still obtaining finite bounds. Thus there is a trade-off between (i) the magnitude of selection on unobservables we can allow for and (ii) the magnitude of control endogeneity. One way to summarize this trade-off is to use *breakdown frontiers* (Masten and Poirier 2020). Specifically, when $\beta_{\text{med}} \geq 0$, define

$$\bar{r}_X^{\text{bf}}(\bar{c}) = \sup\{\bar{r}_X \geq 0 : b \geq 0 \text{ for all } b \in \mathcal{B}_I(\bar{r}_X, \bar{c})\}.$$

For any fixed $\bar{c}$, $\bar{r}_X^{\text{bf}}(\bar{c})$ is a breakdown point: It is the largest magnitude of selection on unobservables relative to selection on observables that we can allow for while still concluding that our parameter of interest is nonnegative. As we vary $\bar{c}$, this breakdown point changes: It increases as $\bar{c}$ gets smaller, because we can allow for more selection on unobservables if we impose stronger restrictions on exogeneity of the observed covariates. Conversely, it decreases as $\bar{c}$ gets larger, because we can allow for less selection on unobservables if we allow for more endogeneity of the observed covariates. In particular, $\bar{r}_X^{\text{bf}}(1) = \bar{r}_X^{\text{bp}}$, the breakdown point of Corollary 1. Like that corollary, we can derive an analytical characterization of the function $\bar{r}_X^{\text{bf}}(\cdot)$, but we omit this for brevity.

Identification Using the $\bar{r}_X$, $\bar{c}$, and $\bar{r}_Y$ Restrictions

Finally, in some empirical settings the results may not be robust even if we impose exogenous controls ($\bar{c} = 0$). In these cases, we might be willing to restrict the impact of unobservables on outcomes; that is, we may be willing to impose A5. Let $\mathcal{B}_I(\bar{r}_X, \bar{r}_Y, \bar{c})$ denote the identified set for β_{long} under A1 and A3–A6. Unlike the two identified sets we considered above, this set is less analytically tractable. We provide a precise characterization in appendix D. Here we instead use our characterization to show how to do breakdown analysis. We also briefly explain how to easily compute $\mathcal{B}_I(\bar{r}_X, \bar{r}_Y, \bar{c})$ numerically.

Suppose we are interested in the robustness of the conclusion that $\beta_{\text{long}} \geq \underline{b}$ for some known scalar $\underline{b}$. For example, $\underline{b} = 0$. Define the function

$$\bar{r}_Y^{\text{bf}}(\bar{r}_X, \bar{c}, \underline{b}) = \sup\{\bar{r}_Y \geq 0 : b \geq \underline{b} \text{ for all } b \in \mathcal{B}_I(\bar{r}_X, \bar{c}, \bar{r}_Y)\}.$$

This is a three-dimensional breakdown frontier. In particular, we can use it to define the set

$$\text{RR} = \{(\bar{r}_X, \bar{r}_Y, \bar{c}) \in \mathbb{R}^2_{\geq 0} \times [0,1] : \bar{r}_Y \leq \bar{r}_Y^{\text{bf}}(\bar{r}_X, \bar{c}, \underline{b})\}.$$

Masten and Poirier (2020) call this the *robust region* because the conclusion of interest, $\beta_{\text{long}} \geq \underline{b}$, holds for any combination of sensitivity parameters in this region. The size of this region is therefore a measure of the robustness of our baseline conclusion.

Although we do not have a closed form expression for the smallest and largest elements of $\mathcal{B}_I(\bar{r}_X, \bar{c}, \bar{r}_Y)$, our next main result shows that we can still easily compute the breakdown frontier numerically. To state the result, we first define some additional notation. For any random vectors A and B, let $\sigma_{A,B} = \text{cov}(A, B)$. Define the sets

$$\begin{aligned}\mathcal{D} &= \mathbb{R} \times \{c \in \mathbb{R}^{d_1} : \|c\| < 1\} \times \mathbb{R} \\ \mathcal{D}^0 &= \{(z, c, b) \in \mathcal{D} : z\sqrt{1 - \|c\|^2}(\sigma_{W_1,Y} - b\sigma_{W_1,X}) - (k_1 - k_0 b)c \neq 0\}\end{aligned}$$

and define the functions

$$
\begin{aligned}
\text{devsq}(z) &= \frac{z^2(k_2/k_0 - \beta_{\text{med}}^2)}{k_0 - z^2} \\
\underline{r}_Y(z,c,b) &= \begin{cases} 0 & \text{if } b = \beta_{\text{med}} \\ \dfrac{|k_1 - k_0 b|}{\|z\sqrt{1-\|c\|^2}(\sigma_{W_1,Y} - b\sigma_{W_1,X}) - (k_1 - k_0 b)c\|} & \text{if } (z,c,b) \in \mathcal{D}^0 \text{ and } b \neq \beta_{\text{med}} \\ +\infty & \text{otherwise} \end{cases} \\
p(z,c;\bar{r}_X) &= \bar{r}_X^2 \|\sigma_{W_1,X}\sqrt{1-\|c\|^2} - cz\|^2 - z^2.
\end{aligned}
$$

We can now state our last main result.

Theorem 4. Suppose the joint distribution of (Y, X, W_1) is known. Suppose A1, A4, and A7 hold. Normalize $\text{var}(X) = 1$ and $\text{var}(W_1) = I$. Suppose $\sigma_{W_1,Y}$ and $\sigma_{W_1,X}$ are linearly independent. Suppose $d_1 \geq 2$. Let $\underline{b} \in \mathbb{R}$, $\bar{c} \in [0,1)$, and $\bar{r}_X, \bar{r}_Y \geq 0$.

1. If $\underline{b} \geq \beta_{\text{med}}$ then $\bar{r}_Y^{\text{bf}}(\bar{r}_X, \bar{c}, \underline{b}) = 0$.
2. If $\underline{B}(\bar{r}_X, \bar{c}) > \underline{b}$, then $\bar{r}_Y^{\text{bf}}(\bar{r}_X, \bar{c}, \underline{b}) = +\infty$.
3. If $\underline{B}(\bar{r}_X, \bar{c}) \leq \underline{b} < \beta_{\text{med}}$, then

$$
\begin{aligned}
\bar{r}_Y^{\text{bf}}(\bar{r}_X, \bar{c}, \underline{b}) = \min_{(z,c_1,c_2,b) \in (-\sqrt{k_0},\sqrt{k_0}) \times \mathbb{R} \times \mathbb{R} \times (-\infty, \underline{b}]} \quad & \underline{r}_Y(z, c_1\sigma_{W_1,Y} + c_2\sigma_{W_1,X}, b) \\
\text{subject to} \quad & p(z, c_1\sigma_{W_1,Y} + c_2\sigma_{W_1,X}; \bar{r}_X) \geq 0 \\
& (b - \beta_{\text{med}})^2 < \text{devsq}(z) \\
& \|c_1\sigma_{W_1,Y} + c_2\sigma_{W_1,X}\| \leq \bar{c}.
\end{aligned}
$$

Theorem 4 shows that the three dimensional breakdown frontier can be computed as the solution to a smooth optimization problem. Importantly, this problem only requires searching over a 4-dimensional space. In particular, this dimension does not depend on the dimension of the covariates W_1. Consequently, it remains computationally feasible even with a large number of observed covariates, as is often the case in empirical practice. For example, the results for our empirical application take about 15 seconds to compute.

Theorem 4 makes several minor technical assumptions. In particular, it assumes $\sigma_{W_1,Y}$ and $\sigma_{W_1,X}$ are linearly independent for simplicity. Moreover, the theorem assumes $d_1 \geq 2$: there are at least two observed covariates in W_1. This is not restrictive since the purpose of this result is primarily to show that the optimization problem does depend on the dimension of W_1. If $d_1 = 1$ then the breakdown frontier can instead be easily computed using equation (34) in the appendix, which only requires searching over a 3-dimensional space.

We conclude this subsection by noting that the identified set $\mathcal{B}_I(\bar{r}_X, \bar{r}_Y, \bar{c})$ can also be easily computed. Specifically, using derivations similar to the proof of Theorem 4, the sharp upper bound on this identified set is

$$\max_{(z,c_1,c_2,b)\in(-\sqrt{k_0},\sqrt{k_0})\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}} b \quad \text{subject to} \quad \begin{aligned} &\underline{r}_Y(z, c_1\sigma_{W_1,Y} + c_2\sigma_{W_1,X}, b) \leq \bar{r}_Y \\ &p(z, c_1\sigma_{W_1,Y} + c_2\sigma_{W_1,X}; \bar{r}_X) \geq 0 \\ &(b - \beta_{\text{med}})^2 < \text{devsq}(z) \\ &\|c_1\sigma_{W_1,Y} + c_2\sigma_{W_1,X}\| \leq \bar{c}. \end{aligned}$$

The sharp lower bound is obtained by replacing max with min.

3.3 Interpreting the Sensitivity Parameters

Thus far we have introduced the sensitivity parameters (section 3.1) and described their implications for identification (section 3.2). Next we make several remarks regarding how to interpret the magnitudes of these parameters.

Which Covariates to Calibrate Against?

As we discuss below, one of the main benefits of using *relative* sensitivity parameters like $\bar{r}_X$ is that $\bar{r}_X = 1$ is a natural reference point of "equal selection." However, the interpretation of this reference point depends on the choice of covariates that we calibrate against. Put differently, when we say that we compare "selection on unobservables to selection on observables," *which* observables do we mean?

To answer this, we split the observed covariates into two groups: (1) The control covariates, which we label W_0, and (2) The calibration covariates, which we continue to label W_1. Write equation (1) as

$$Y = \beta_{\text{long}} X + \gamma_0' W_0 + \gamma_1' W_1 + \gamma_2 W_2 + Y^{\perp X,W} \tag{1'}$$

where $W = (W_0, W_1, W_2)$. Likewise, write equation (2) as

$$X = \pi_0' W_0 + \pi_1' W_1 + \pi_2 W_2 + X^{\perp W}. \tag{2'}$$

The key difference from our earlier analysis is that, like in assumption A3, we will continue to only compare π_1 with π_2. That is, we only compare the omitted variable to the observed variables in W_1; we do not use W_0 for calibration. A similar remark applies to A5.

This distinction between control and calibration covariates is useful because in many applications we do not necessarily think the omitted variables have similar explanatory power as *all* of the observed covariates included in the model. For example, in our empirical application in section 5, we include state fixed effects as control covariates, but we do not use them for calibration.

We next briefly describe how to generalize our results in section 3.2 to account for this distinction. By the FWL theorem, a linear projection of Y onto $(1, X^{\perp W_0}, W_1^{\perp W_0}, W_2^{\perp W_0})$ has the same coefficients as equation (1). Likewise for a linear projection of X onto $(1, W_1^{\perp W_0}, W_2^{\perp W_0})$. Hence we can write

$$\begin{aligned} Y &= \beta_{\text{long}} X^{\perp W_0} + \gamma_1' W_1^{\perp W_0} + \gamma_2 W_2^{\perp W_0} + \widetilde{U} \\ X &= \pi_1' W_1^{\perp W_0} + \pi_2 W_2^{\perp W_0} + \widetilde{V} \end{aligned}$$

where

$$\widetilde{U} = Y^{\perp X^{\perp W_0}, W_1^{\perp W_0}, W_2^{\perp W_0}} \qquad \text{and} \qquad \widetilde{V} = X^{\perp W_1^{\perp W_0}, W_2^{\perp W_0}}.$$

By construction, $\widetilde{U}$ has zero covariance with $(X^{\perp W_0}, W_1^{\perp W_0}, W_2^{\perp W_0})$ and $\widetilde{V}$ has zero covariance with $(W_1^{\perp W_0}, W_2^{\perp W_0})$. Therefore our earlier results continue to hold when (X, W_1, W_2) are replaced with $(X^{\perp W_0}, W_1^{\perp W_0}, W_2^{\perp W_0})$. Finally, this change implies that $\bar{c}$ should be interpreted as an upper bound on $R_{W_2 \sim W_1 \bullet W_0}$, the square root of the R-squared from the regression of W_2 on W_1 after partialling out W_0.

What is a Robust Result?

How should researchers determine what values of $\bar{r}_X$ and $\bar{r}_Y$ are large and what values are small? Like in Altonji et al. (2005) and Oster (2019), these are *relative* sensitivity parameters. Consequently, the values $\bar{r}_X = 1$ and $\bar{r}_Y = 1$ are natural reference points. Specifically, when $\bar{r}_X < 1$, the magnitude of the coefficient on the unobservable W_2 in equation (2) is smaller than the magnitude of the coefficient on the observable controls W_1 in the outcome equation. This is one way to formalize the idea that "selection on unobservables is smaller than selection on observables." Likewise, when $\bar{r}_X > 1$, we can think of this as formalizing the idea that "selection on unobservables is larger than selection on observables." A similar interpretation applies to $\bar{r}_Y$. These interpretations do *not*, however, imply that the value 1 should be thought of as a universal, context independent cutoff between "small" and "large" values of these two sensitivity parameters.

Why? As we described above, researchers must choose which of their observed covariates should be used to calibrate against. Consequently, the choice of W_0 (and hence W_1) affects the interpretation of the magnitude of $\bar{r}_X$. One way in which this choice manifests itself is via its impact on the breakdown point: Including more relevant variables in W_1 will tend to will tend to decrease the breakdown point $\bar{r}_X^{\text{bp}}$, because the explanatory power of the observables we're calibrating against increases when we move variables from W_0 to W_1—see corollary 1. This observation does *not* necessarily imply that the result is becoming less robust, but rather that the standard by which we are measuring sensitivity is changing. If the calibration variables W_1 have a large amount of explanatory power, then even an apparently small value of $\bar{r}_X^{\text{bp}}$ like 0.3 could be considered to be robust. Conversely, when the calibration variables W_1 do not have much explanatory power, then even an apparently large value of $\bar{r}_X^{\text{bp}}$ like 3 could be considered sensitive.

This discussion can be summarized by the following relationship:

$$\text{Selection on Unobservables} = r \cdot (\text{Selection on Observables}). \tag{4}$$

The left hand side is the absolute magnitude of selection on unobservables, while the right hand side is the proportion r of the absolute magnitude of selection on observables. $\bar{r}_X^{\text{bp}}$ is a bound on r. Our discussion above merely states that even if the bound $\bar{r}_X^{\text{bp}}$ on r seems small, the magnitude of selection on unobservables allowed for can be very large if the magnitude of selection on observables is large. And conversely, even if $\bar{r}_X^{\text{bp}}$ seems large, the amount of unobserved selection allowed for must be small if the magnitude of selection on observables is also small.

Overall, the value of using relative sensitivity parameters like $\bar{r}_X$ is not that they allow us to obtain a universal threshold for what is or is not a robust result. Instead, it gives researchers a *unit free* measurement of sensitivity that is *interpretable* in terms of the effects of observed variables. Finally, note that the issues we've raised in this discussion equally apply to the existing methods in the literature as well, including Oster (2019); they are not unique to our analysis.

Assessing Exogenous Controls

Thus far we have discussed the interpretation of $\bar{r}_X$ and $\bar{r}_Y$. Next consider $\bar{c}$. This is a constraint on the covariance between the observed calibration covariates and the unobserved covariate. In particular, recall that it is a bound on $R_{W_2 \sim W_1 \bullet W_0}$. So what values of this parameter should be considered large, and what values should be considered small? One way to calibrate this parameter is to compute

$$c_k = R_{W_{1k} \sim W_{1,-k} \bullet W_0}$$

for each covariate k in W_1. That is, compute the square root of the population R-squared from the regression of W_{1k} on the rest of the calibration covariates $W_{1,-k}$, after partialling out the control covariates W_0. These numbers tell us two things. First, if many of these values are nonzero and large, we may worry that the exogenous controls assumption fails. That is, if W_2 is in some way similar to the observed covariates W_1, then we might expect that $R_{W_2 \sim W_1 \bullet W_0}$ is similar to some of the c_k's. So this gives us one method for assessing the plausibility of exogenous controls. Second, we can use the magnitudes of these values to calibrate our choice of $\bar{c}$, in analysis based on Theorems 3 or 4. For example, you could choose the largest value of c_k. A less conservative approach would be to select the median value.

3.4 Estimation and Inference

Thus far we have described population level identification results. In practice, we only observe finite sample data. Our identification results depend on the observables (Y, X, W_1) solely through their covariance matrix. In our empirical analysis in section 5, we apply our identification results by using sample analog estimators that replace $\text{var}(Y, X, W_1)$ with a consistent estimator $\widehat{\text{var}}(Y, X, W_1)$. For example, we let $\widehat{\beta}_{\text{med}}$ denote the OLS estimator of β_{med}, the coefficient on X in the medium regression of Y on $(1, X, W_1)$. We expect the corresponding asymptotic theory for estimation and inference on the bound functions to be straightforward, but for brevity we do not develop it in this paper. Inference on the breakdown points and frontiers could also be done as in Masten and Poirier (2020).

4 Causal Models

In this section we describe three different causal models in which the parameter β_{long} in equation (1) has a causal interpretation. These models are based on three different kinds of identification strategies: Unconfoundedness, difference-in-differences, and instrumental variables. Here we focus on simple models, but our analysis can be used anytime the causal parameter of interest can be written as the coefficient on a treatment variable in a long regression of the form in equation (1).

4.1 Unconfoundedness

Recall that Y denotes the realized outcome, X denotes treatment, W_1 denotes the observed covariates, and W_2 denotes the unobserved variables of concern. Let $Y(x)$ denote potential outcomes, where x is any logically possible value of treatment. Assume this potential outcome has the following form:

$$Y(x) = \beta_c x + \gamma_1' W_1 + \gamma_2 W_2 + U \tag{5}$$

where $(\beta_c, \gamma_1, \gamma_2)$ are unknown constants. The parameter of interest is β_c, the causal effect of treatment on the outcome. U is an unobserved random variable. Suppose the realized outcome satisfies $Y = Y(X)$. Consider the following assumption.

Linear Latent Unconfoundedness: $\text{corr}(X^{\perp W_1, W_2}, U^{\perp W_1, W_2}) = 0$.

This assumption says that, after partialling out the observed covariates W_1 and the unobserved variables W_2, treatment is uncorrelated with the unobserved variable U. This model has two unobservables, which are treated differently via this assumption. We call W_2 the *confounders* and U the *non-confounders*. W_2 are the unobserved variables which, when omitted, may cause bias. In contrast, as long as we adjust for (W_1, W_2), omitting U does not cause bias. Note that, given equation (5), linear latent unconfoundedness can be equivalently written as $\text{corr}(X^{\perp W_1, W_2}, Y(x)^{\perp W_1, W_2}) = 0$ for all logically possible values of treatment x.

Linear latent unconfoundedness is a linear parametric version of the nonparametric latent unconfoundedness assumption

$$Y(x) \perp\!\!\!\perp X \mid (W_1, W_2) \tag{6}$$

for all logically possible values of x. In particular, with the linear potential outcomes assumption of equation (5), nonparametric latent unconfoundedness (equation (6)) implies linear latent unconfoundedness. We use the linear parametric version to avoid overidentifying restrictions that can arise from the combination of linearity and statistical independence.

The following result shows that, in this model, the causal effect of X on Y can be obtained from β_{long}, the coefficient on X in the long regression described in equation (1).

Proposition 1. Consider the linear potential outcomes model (5). Suppose linear latent unconfoundedness holds. Suppose A1 holds. Then $\beta_c = \beta_{\text{long}}$.

Since W_2 is unobserved, however, this result cannot be used to identify β_c. Instead, suppose we believe the no selection on unobservables assumption A2. Recall that this assumption says that $\pi_2 = 0$, where π_2 is the coefficient on W_2 in the OLS estimand of X on $(1, W_1, W_2)$. Under this assumption, we obtain the following result. Recall that β_{med} denotes the coefficient on X in the medium regression of Y on $(1, X, W_1)$.

Corollary 2. Suppose the assumptions of Proposition 1 hold. Suppose A2 holds ($\pi_2 = 0$). Then $\beta_c = \beta_{\text{med}}$.

The selection on observables assumption A2 is usually thought to be quite strong, however. Nonetheless, since $\beta_c = \beta_{\text{long}}$, our results in section 3 can be used to assess sensitivity to selection on unobservables.

4.2 Difference-in-differences

Let $Y_t(x_t)$ denote potential outcomes at time t, where x_t is a logically possible value of treatment. For simplicity we do not consider models with dynamic effects of treatment or of covariates. Also suppose W_{2t} is a scalar for simplicity. Suppose

$$Y_t(x_t) = \beta_c x_t + \gamma_1' W_{1t} + \gamma_2 W_{2t} + V_t \tag{7}$$

where V_t is an unobserved random variable and $(\beta_c, \gamma_1, \gamma_2)$ are unknown parameters that are constant across units. The classical two way fixed effects model is a special case where

$$V_t = A + \delta_t + U_t. \tag{8}$$

where A is an unobserved random variable that is constant over time, δ_t is an unobserved constant, and U_t is an unobserved random variable.

Suppose there are two time periods, $t \in \{1, 2\}$. Let $Y_t = Y_t(X_t)$ denote the observed outcome at time t. For any time varying random variable like Y_t, let $\Delta Y = Y_2 - Y_1$. Then taking first differences of the observed outcomes yields

$$\Delta Y = \beta_c \Delta X + \gamma_1' \Delta W_1 + \gamma_2 \Delta W_2 + \Delta V.$$

Let β_{long} denote the OLS coefficient on ΔX from the long regression of ΔY on $(1, \Delta X, \Delta W_1, \Delta W_2)$.

Proposition 2. Consider the linear potential outcome model (7). Suppose the following exogeneity assumption holds:

- $\text{cov}(\Delta X, \Delta V) = 0$, $\text{cov}(\Delta W_2, \Delta V) = 0$, and $\text{cov}(\Delta W_1, \Delta V) = 0$.

Then $\beta_c = \beta_{\text{long}}$.

The exogeneity assumption in Proposition 2 says that ΔV is uncorrelated with all components of $(\Delta X, \Delta W_1, \Delta W_2)$. A sufficient condition for this is the two way fixed effects assumption (8) combined with the assumption that the U_t are uncorrelated with (X_s, W_{1s}, W_{2s}) for all t and s. Given this exogeneity assumption, the only possible identification problem is that ΔW_2 is unobserved. Hence we cannot adjust for this trend variable. If we assume, however, that treatment trends ΔX are not related to the unobserved trend ΔW_2, then we can point identify β_c. Specifically, consider the linear projection of ΔX onto $(1, \Delta W_1, \Delta W_2)$:

$$\Delta X = \pi_1'(\Delta W_1) + \pi_2(\Delta W_2) + (\Delta X)^{\perp \Delta W_1, \Delta W_2}.$$

Using this equation to define π_2, we now have the following result. Here we let β_{med} denote the coefficient on ΔX in the medium regression of ΔY on $(1, \Delta X, \Delta W_1)$.

Corollary 3. Suppose the assumptions of Proposition 2 hold. Suppose A2 holds ($\pi_2 = 0$). Then $\beta_c = \beta_{\text{med}}$.

This result implies that β_c is point identified when $\pi_2 = 0$. This assumption is a version of common trends, because it says that the unobserved trend ΔW_2 is not related to the trend in treatments, ΔX. Our results in section 3 allow us to analyze the impacts of failure of this common trends assumption on conclusions about the causal effect of X on Y, β_c. In particular, our results allow researchers to assess the failure of common trends by comparing the impact of observed time varying covariates with the impact of unobserved time varying confounders. In this context, allowing for endogenous controls means allowing for the trend in observed covariates to correlate with the trend in the unobserved covariates. Finally, note that this approach allows researchers to assess sensitivity to common trends even when it is not possible to examine pre-trends; that is, even when there are not multiple time periods where all units are untreated.

4.3 Instrumental variables

Let Z be an observed variable that we want to use as an instrument. Let $Y(z)$ denote potential outcomes, where z is any logical value of the instrument. Assume

$$Y(z) = \beta_c z + \gamma_1' W_1 + \gamma_2 W_2 + U$$

where U is an unobserved scalar random variable and $(\beta_c, \gamma_1, \gamma_2)$ are unknown constants. Thus β_c is the causal effect of Z on Y. In an instrumental variables analysis, this is typically called the reduced form causal effect, and $Y(z)$ are reduced form potential outcomes. Suppose $\text{cov}(Z,U) = 0$, $\text{cov}(W_2,U) = 0$, and $\text{cov}(W_1,U) = 0$. Then β_c equals the OLS coefficient on Z from the long regression of Y on $(1, Z, W_1, W_2)$. In this model, Theorem 1 implies that β_c is also obtained as the coefficient on Z in the medium regression of Y on $(1, Z, W_1)$, and thus is point identified. In this case, assumption A2 is an instrument exogeneity assumption. Our results in section 3 thus allow us to analyze the impacts of instrument exogeneity failure on conclusions about the reduced form causal effect of Z on Y, β_c.

In a typical instrumental variable analysis, the reduced form causal effect of the instrument on outcomes is not the main effect of interest. Instead, we usually care about the causal effect of a treatment variable on outcomes. The reduced form is often just an intermediate tool for learning about that causal effect. Our analysis in this paper can be used to assess the sensitivity of conclusions about this causal effect to failures of instrument exclusion or exogeneity too. This analysis is somewhat more complicated, however, and so we leave it for a separate paper. Nonetheless, empirical researchers do sometimes examine the reduced form directly to study the impact of instrument exogeneity failure. For example, see section D7 and table D15 of Tabellini (2020).

5 Empirical Application: The Frontier Experience and Culture

Where does culture come from? Bazzi et al. (2020) study the origins of people's preferences for or against government redistribution, intervention, and regulation. They provide the first systematic empirical analysis of a famous conjecture that living on the American frontier cultivated individualism and antipathy to government intervention. The idea is that life on the frontier was hard and dangerous, had little to no infrastructure, and required independence and self-reliance to survive. It was far from the federal government. And it was an opportunity for upward mobility through effort, rather than luck. These features then create cultural change, in particular, leading to "more pervasive individualism and opposition to redistribution". Overall, Bazzi et al. (2020) find evidence supporting this frontier life conjecture.

The main results in Bazzi et al. (2020) are based on an unconfoundedness identification strategy and use linear models. They note that "the main threat to causal identification of β lies in omitted variables" and hence they strongly rely on Oster's (2019) method to "show that unobservables are unlikely to drive our results" (page 2344). As we have discussed, however, this approach is based on the exogenous controls assumption. In this section, we apply our methods to examine the impact of allowing for endogenous controls on Bazzi et al.'s empirical conclusions. Overall, we come to a more nuanced conclusion about robustness: While they found that all of their analyses were robust to the presence of omitted variables, we find that their analysis using questionnaire based outcomes is quite sensitive, but their analysis using property tax levels and voting patterns is robust. We also find suggestive evidence that the controls are endogenous, which highlights the value of sensitivity analysis methods that allow for endogenous controls. We discuss all of these findings in more detail below.

5.1 Data

We first describe the variables and data sources. The main units of analysis are counties in the U.S., although we will also use some individual level data. The treatment X is the "total frontier experience" (TFE). This is defined as the number of years between 1790 and 1890 a country spent "on the frontier", divided by 10. A county is "on the frontier" if it had a population density less than 6 people per square mile and was within 100 km of the "frontier line". The frontier line is a line that divides sparse counties (less than or equal to

2 people per square mile) from less sparse counties. By definition, the frontier line changed over time in response to population patterns, but it did so unevenly, resulting in some counties being "on the frontier" for longer than others. Figure 3 in Bazzi et al. (2020) shows the spatial distribution of treatment.

The outcome variable Y is a measure of modern culture. They consider 8 different outcome variables. Since data is not publicly available for all of them, we only look at 5 of these. They can be classified into two groups. The first are questionnaire based outcomes:

1. *Cut spending on the poor.* This variable comes from the 1992 and 1996 waves of the American National Election Study (ANES), a nationally representative survey. In those waves, it asked

   > "Should federal spending be increased, decreased, or kept about the same on poor people?"

   Let $Y_{1i} = 1$ if individual i answered "decreased" and 0 otherwise.

2. *Cut welfare spending.* This variable comes from the Cooperative Congressional Election Study (CCES), waves 2014 and 2016. In those waves, it asked

   > "State legislatures must make choices when making spending decisions on important state programs. Would you like your legislature to increase or decrease spending on Welfare? 1. Greatly Increase 2. Slightly Increase 3. Maintain 4. Slightly Decrease 5. Greatly Decrease."

   Let $Y_{2i} = 1$ if individual i answered "slightly decrease" or "greatly decrease" and 0 otherwise.

3. *Reduce debt by cutting spending.* This variable also comes from the CCES, waves 2000–2014 (biannual). It asked

   > "The federal budget deficit is approximately [$ year specific amount] this year. If the Congress were to balance the budget it would have to consider cutting defense spending, cutting domestic spending (such as Medicare and Social Security), or raising taxes to cover the deficit. Please rank the options below from what would you most prefer that Congress do to what you would least prefer they do: Cut Defense Spending; Cut Domestic Spending; Raise Taxes."

   Let $Y_{3i} = 1$ if individual i chooses "cut domestic spending" as a first priority, and 0 otherwise.

These surveys also collected data on individual demographics, specifically age, gender, and race. The second group of outcome variables are based on behavior rather than questionnaire responses:

4. Y_{4i} is the average effective *property tax rate* in county i, based on data from 2010 to 2014 from the National Association of Home Builders (NAHB) data, which itself uses data from the American Community Survey (ACS) waves 2010–2014.

5. Y_{5i} is the average *Republican vote share* over the five presidential elections from 2000 to 2016 in county i, using data from Leip's Atlas of U.S. Presidential Elections.

Next we describe the observed covariates. We partition these covariates into W_1 and W_0 by following the implementation of Oster's (2019) approach in Bazzi et al. (2020). W_1, the calibration covariates which *are* used to calibrate selection on unobservables, is a set of geographic and climate controls: Centroid Latitude, Centroid Longitude, Land area, Average rainfall, Average temperature, Elevation, Average potential agricultural yield, and Distance from the centroid to rivers, lakes, and the coast. W_0, the control covariates which are *not* used to calibrate selection on unobservables, includes state fixed effects. The questionnaire

Table 1: The Effect of Frontier Life on Opposition to Government Intervention and Redistribution.

	Prefers Cut Public Spending on Poor	Prefers Cut Public Spending on Welfare	Prefers Reduce Debt by Spending Cuts	County Property Tax Rate	Republican Presidential Vote Share
	(1)	(2)	(3)	(4)	(5)
Panel A. Baseline Results					
Total Frontier Exp.	0.010	0.007	0.014	-0.034	2.055
	(0.004)	(0.003)	(0.002)	(0.007)	(0.349)
Mean of Dep Variable	0.09	0.40	0.41	1.02	60.04
Number of Individuals	2322	53,472	111,853	-	-
Number of Counties	95	1863	1963	2029	2036
Controls:					
Survey Wave FEs	X	X	X	-	-
Ind. Demographics	X	X	X	-	-
State Fixed Effects	X	X	X	X	X
Geographic/Climate	X	X	X	X	X
Panel B. Sensitivity Analysis (Exogenous Controls; Oster 2019)					
δ^{bp} (wrong)	16.01	3.1	5.9	-27.4	-8.5
δ^{bp} (correct)	2.28	3.05	2.58	90.7	-23.3
Panel C. Sensitivity Analysis (Endogenous Controls)					
$\bar{r}_X^{\text{bp}}$ ($\times 100$)	2.83	3.05	5.85	72.0	80.4

Note: Panel A and the first row of Panel B replicate columns 1, 2, 4, 6, and 7 of table 3 in Bazzi, Fiszbein, and Gebresilasse (2020), while the second row of Panel B and Panel C are new. As in Bazzi et al. (2020), Panel B uses Oster's rule of thumb choice $R^2_{\text{long}} = 1.3R^2_{\text{med}}$.

based outcomes use individual level data. For those analyses, we also include age, age-squared, gender, race, and survey wave fixed effects in W_0. In Bazzi et al. (2020), they were included in W_1. We instead include them in W_0 to keep the set of calibration covariates W_1 constant across the five main specifications. This allows us to directly compare the robustness of our baseline results across different specifications.

5.2 Baseline Model Results

Bazzi et al. (2020) has a variety of analyses. We focus on the subset of their main results for which replication data is publicly available. These are columns 1, 2, 4, 6, and 7 of their table 3. Panel A in our table 1 replicates those results. From columns (1)–(3) we see that individuals who live in counties with more exposure to the frontier prefer cutting spending on the poor, on welfare, and to reduce debt by spending cuts. Moreover, these point estimates are statistically significant at conventional levels. From columns (4) and (5), we see that counties with more exposure to the frontier have lower property taxes and are more likely to vote for Republicans. As Bazzi et al. (2020) argue, these baseline results therefore support the conjecture that frontier life led to opposition to government intervention and redistribution.

5.3 Assessing Selection on Observables

The baseline results in table 1 rely on a selection on observables assumption, that treatment X is exogenous after adjusting for the observed covariates (W_0, W_1). How plausible is this assumption? Bazzi et al. (2020)

say

> "The main threat to causal identification of β lies in omitted variables correlated with both contemporary culture and TFE. We address this concern in four ways. First, we rule out confounding effects of modern population density. Second, we augment [the covariates] to remove cultural variation highlighted in prior work. *Third, we show that unobservables are unlikely to drive our results.* Finally, we use an IV strategy that isolates exogenous variation in TFE due to changes in national immigration flows over time." (page 2344, emphasis added)

Their first two approaches continue to rely on selection on observables, and consist of including additional control variables. We focus on their third strategy: to use a formal econometric method to assess the importance of omitted variables.

Sensitivity Analysis Assuming Exogenous Controls

We start by summarizing the sensitivity analysis based on Oster (2019) (hereafter Oster), as used in Bazzi et al. (2020). Oster's analysis uses two sensitivity parameters: (i) δ, which we define in equation (9) in appendix A.2 and (ii) R^2_{long}, the R-squared from the long regression of Y on $(1, X, W_0, W_1, W_2)$, including the omitted variable of concern W_2. For any choice of $(\delta, R^2_{\text{long}})$, Oster's Proposition 2 derives the identified set for β_{long}. Oster's Proposition 3 derives the breakdown point for δ, as a function of R^2_{long}, for the conclusion that the identified set does not contain zero. Denote this point by $\delta^{\text{bp}}(R^2_{\text{long}})$. This is the smallest value of δ such that the identified set contains zero. Put differently: For any $\delta < \delta^{\text{bp}}(R^2_{\text{long}})$, the true value of β_{long} cannot be zero.

The second row of Panel B of table 1 shows sample analog estimates of this breakdown point, which is commonly referred to as *Oster's delta*. As in Bazzi et al. (2020), we use Oster's rule of thumb choice $R^2_{\text{long}} = 1.3R^2_{\text{med}}$. R^2_{med} is the R-squared from the medium regression of $Y^{\perp W_0}$ on $(1, X^{\perp W_0}, W_1^{\perp W_0})$, which can be estimated from the data. Thus the table shows estimates of $\delta^{\text{bp}}(1.3R^2_{\text{med}})$. The first row of Panel B shows the values of Oster's delta as reported in table 2 of Bazzi et al. (2020). These were incorrectly computed. It appears to us that, rather than using the correct expression in Proposition 3 of Oster (2019), they set the first displayed equation on page 193 of that paper equal to zero and solved for δ. That does not give the correct breakdown point. Note that we noticed this same mistake in several papers published in top 5 economics journals.

Bazzi et al. (2020) conclude:

> "Oster (2019) suggests $|\delta| > 1$ leaves limited scope for unobservables to explain the results" and therefore, based on their δ^{bp} estimates, "unobservables are unlikely to drive our results" (page 2344)

This conclusion remains unchanged if the same rule is applied to the correctly computed δ^{bp} estimates.

Assessing Exogenous Controls

As we have discussed, Oster's method combined with the $\delta = 1$ cutoff rule relies on the exogenous controls assumption. Is exogenous controls plausible in this application? The answer depends on which omitted variables W_2 we are concerned about. Bazzi et al. (2020) does not specifically describe the unmeasured omitted variables of concern, nor do they discuss the plausibility of exogenous controls. However, in their extra robustness checks they consider the variables listed in table 2.

Table 2: Additional Covariates Included by Bazzi et al. (2020) as Robustness Checks.

Contemporary population density	Sex ratio
Conflict with Native Americans	Rainfall risk
Employment share in manufacturing	Portage sites
Mineral resources	Prevalence of slavery
Immigrant share	Scotch-Irish settlement
Timing of railroad access	Birthplace diversity
Ruggedness	

The additional omitted variables of concern might therefore be similar to these variables. Thus the question is: Are *all* of the geographic/climate variables in W_1 uncorrelated with variables like these? This seems unlikely, especially since many of these additional variables are also geographic/climate type variables. Moreover, although this assumption is not falsifiable—since W_2 is unobserved—we can assess its plausibility by examining the correlation structure of the observed covariates. Specifically, we compute the parameters c_k defined in section 3.3. These are square roots of R-squareds from regressing each element of W_1 on the other elements, after partialling out W_0. Table 3 shows sample analog estimates of these c_k's.

The estimates in table 3 show a substantial range of correlation between the observed covariates in W_1. Recall that the exogenous controls assumption says that each element of W_1 is uncorrelated with W_2, after partialling out W_0. Thus if W_2 was included in this table it would have a value of zero. Therefore, if W_2 is a variable similar to the components of W_1 then we would expect exogenous controls to fail. This suggests that it is important to use sensitivity analysis methods that allow for endogenous controls.

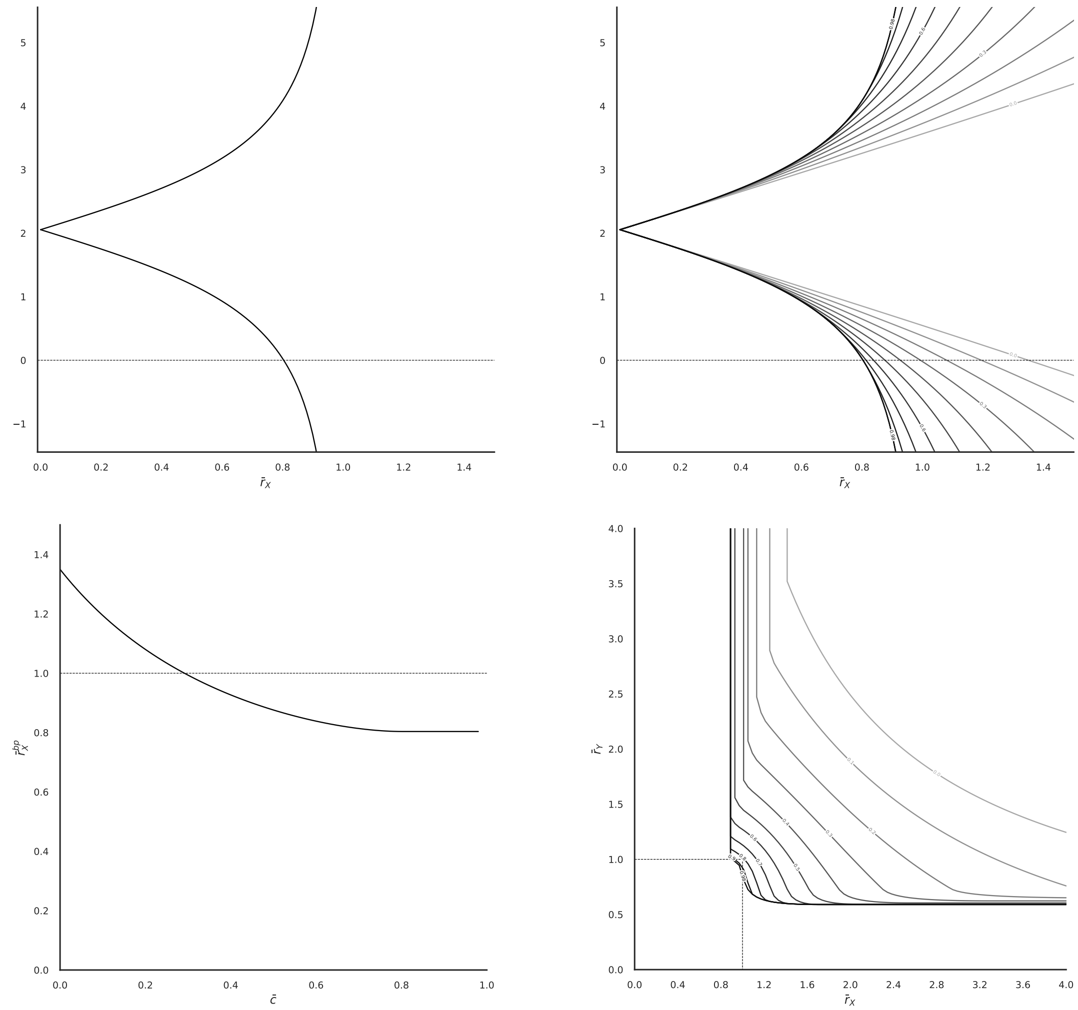
Table 3: Correlations Between Observed Covariates.

W_{1k}	$\widehat{R}_{W_{1k} \sim W_{1,-k} \bullet W_0}$
Average temperature	0.945
Centroid Latitude	0.936
Elevation	0.825
Average potential agricultural yield	0.805
Average rainfall	0.748
Distance from centroid to the coast	0.698
Centroid Longitude	0.659
Distance from centroid to rivers	0.367
Distance from centroid to lakes	0.316
Land area	0.313

Sensitivity Analysis Allowing For Endogenous Controls

Next we present the findings from the sensitivity analysis that we developed in section 3, which allows for endogenous controls. We begin with our simplest result, Theorem 2, that only uses a single sensitivity parameter $\bar{r}_X$. Panel C of table 1 reports sample analog estimates of the breakdown point $\bar{r}_X^{\text{bp}}$ described in Corollary 1. This is the largest amount of selection on unobservables, as a percentage of selection on observables, allowed for until we can no longer conclude that β_{long} is nonzero. Recall that, since this result allows for arbitrarily endogenous controls, Theorem 2 implies that $\bar{r}_X^{\text{bp}} < 1$. As discussed in section 3.3, however, this does not imply that these results should always be considered non-robust. Instead, when the calibration covariates W_1 are a set of variables that are important for treatment selection, researchers should

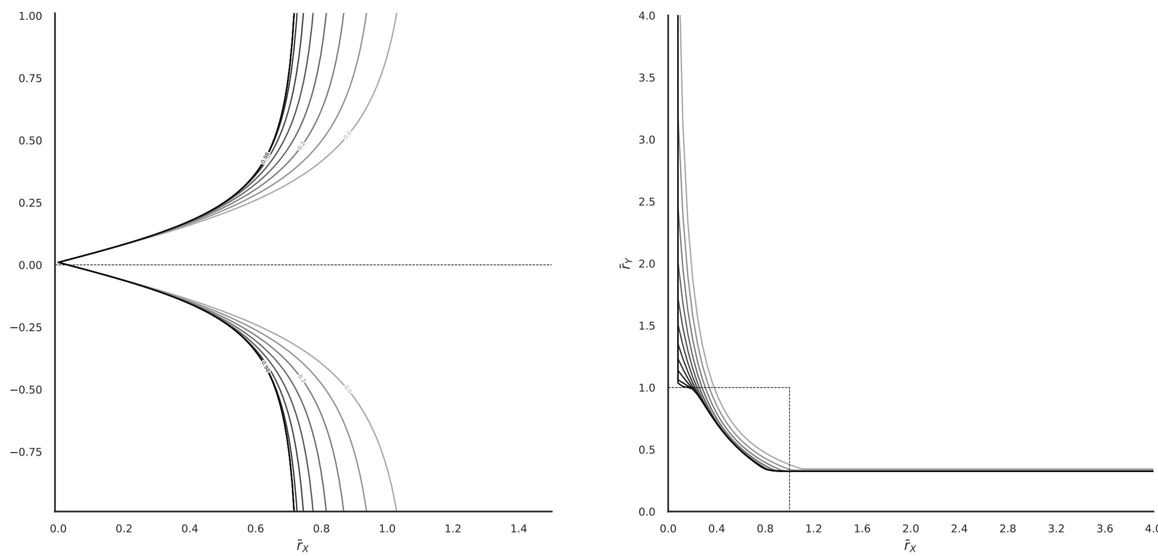
Figure 1: Sensitivity Analysis for Average Republican Vote Share. See body text for discussion.



consider large values of $\bar{r}_X^{\text{bp}}$ to indicate the robustness of their baseline results. For example, in columns (4) and (5) of Panel C we see that the breakdown point estimates for the two behavior based outcomes are 72% and 80.4%. For example, for the average Republication vote share outcome, we can conclude $\beta_{\text{long}} > 0$ as long as selection on unobservables is at most 80.4% as large as selection on observables. In contrast, the breakdown point estimates in columns (1)–(3) are substantially smaller: between about 3% and 6%. For these outcomes, we therefore only need selection on unobservables to be at least 3 to 6% as large as selection on observables to overturn our conclusion that $\beta_{\text{long}} > 0$. Thus, unlike the conclusions based on Oster's method, we find that the analysis using questionnaire based outcomes is highly sensitive to selection on unobservables. In contrast, the analysis using behavior based outcomes is quite robust to selection on unobservables. This contrast continues to hold after considering restrictions on the magnitude of endogenous controls and the impact of unobservables on outcomes too. We present these analyses next.

For brevity we discuss just one questionnaire based outcome, cut spending on the poor, and one behavior based outcome, average Republican vote share. Figure 1 shows the results for average Republican vote share. The top left plot shows the estimated identified set β_{long} as a function of $\bar{r}_X$, allowing for arbitrarily endogenous controls and no restrictions on the outcome equation. This is the set described by Theorem 2.

Figure 2: Sensitivity Analysis for Cut Spending on Poor. See body text for discussion.

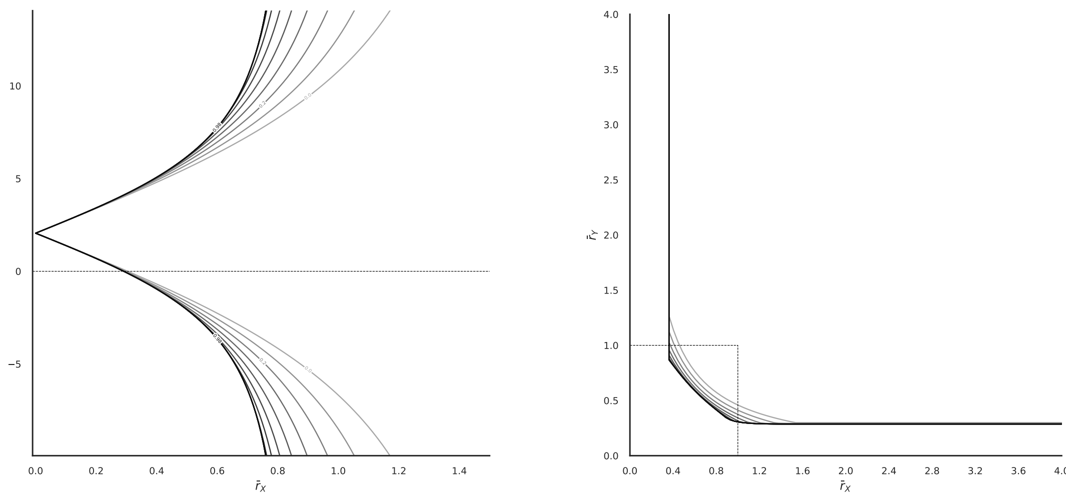


The horizontal intercept is the breakdown point $\bar{r}_X^{\text{bp}} = 80.4\%$, as reported in Panel C, column (5) of table 1. This result allows for arbitrarily endogenous controls.

If we are willing to somewhat restrict the magnitude of control endogeneity, we can allow for more selection on unobservables. The top right figure shows a sequence of estimated identified sets for β_{long} as a function of $\bar{r}_X$ on the horizontal axis, as described in Theorem 3. It starts at the darkest line, $\bar{c} = 1$ (arbitrarily endogenous controls), and then as the shading of the bound functions becomes lighter, we get closer to exogenous controls ($\bar{c} = 0$). Put differently: For any fixed value of $\bar{r}_X$, imposing stronger assumptions on exogeneity of the controls results in a smaller identified set. The bottom left picture shows the impact of assuming partially exogenous controls on the breakdown point for selection on unobservables. Specifically, this plot shows the estimated breakdown frontier $\bar{r}_X^{\text{bf}}(\bar{c})$. This function shows the horizontal intercept in the top right figure, as a function of $\bar{c}$. At $\bar{c} = 1$, we recover the breakdown point 80.4% that allows for arbitrarily endogenous controls. If we impose exogenous controls, however, and set $\bar{c} = 0$, we obtain a breakdown point around 135%. That is, under exogenous controls, we can allow for selection on unobservables of up to 135% as large as selection on observables before our baseline results break down. In fact, we only need $\bar{c}$ less than or equal to about 0.3 to obtain a breakdown point at or above 100%.

All of the analysis thus far has left the impact of unobservables on outcomes unrestricted. So in our final analysis we also consider the effect of restricting the impact of unobservables on outcomes. The bottom right plot in figure 1 shows the three-dimensional breakdown frontiers $\bar{r}_Y^{\text{bf}}(\bar{r}_X, \bar{c})$ described in Theorem 4. Any combination of sensitivity parameters $(\bar{r}_X, \bar{r}_Y, \bar{c})$ below this three-dimensional function lead to an identified set that allows us to conclude $\beta_{\text{long}} > 0$. This includes, for example, $\bar{r}_X = \bar{r}_Y = 110\%$ and $\bar{c} = 0.7$. Note that 0.7 is around the middle of the distribution of c_k values in table 3, and hence might be considered a moderate or slightly conservative value of the magnitude of control endogeneity. For this value, our baseline finding is robust to omitted variables that have up to 110% of the effect on treatment and outcomes as the observables. If we impose exogenous controls ($\bar{c} = 0$) then we can allow the impact of the omitted variable on outcomes to be 200% as large as the observables and the impact of the omitted variable on treatment to be up to about 240% as large as the observables, and yet still conclude that $\beta_{\text{long}} > 0$.

Figure 3: Effect of Calibration Covariates on Analysis For Republican Vote Share. See body text for discussion.



These findings suggest that the empirical conclusions for average Republican vote share are quite robust to failures of the selection on observables assumption. In contrast, we next consider the analysis for the cut spending on the poor outcome. Figure 2 shows the results. The left plot shows the estimated identified sets for β_{long} as a function of $\bar{r}_X$ on the horizontal axis, as described in Theorem 3. For $\bar{c} = 1$, the horizontal intercept gives an estimated value for $\bar{r}_X^{\text{bp}}$ of 2.83%, as reported in Panel C, column (1) of table 1. Moreover, as shown in the figure, even if we impose exogenous controls, $\bar{c} = 0$, the identified sets do not change much, and hence the breakdown point does not change much. The breakdown frontier $\bar{r}_X^{\text{bf}}(\bar{c})$ is essentially flat and hence we do not report it. These conclusions do not change substantially if we also impose restrictions on how omitted variables affect the outcomes. The right plot in figure 2 shows the estimated three-dimensional breakdown frontier. It shows that we can allow for larger amounts of selection on unobservables if we are willing to greatly restrict the impact of unobservables on outcomes. For example, if we allow for arbitrarily endogenous controls ($\bar{c} = 1$) then we can allow for the effect of omitted variables on the treatment and outcomes to be as much as 50% that of the effect of the observables while still concluding that $\beta_{\text{long}} > 0$. Alternatively, if we restrict the effect of omitted variables on outcomes to be at most 25% that of observables, then we can allow the omitted variables to affect treatment by more than 100% of the effect of the observables while still concluding that $\beta_{\text{long}} > 0$.

Overall, we see that there are some cases where the results for cutting spending on the poor could be considered robust. But there are also many cases where these results could be considered sensitive. In contrast, the results for average Republican vote share are robust across a wide range of relaxations of the baseline model. Similar findings hold for the other three outcome variables: The three results using the questionnaire based outcomes tend to be much more sensitive than the two results using behavior based outcomes.

The Effect of the Choice of Calibration Covariates

In section 3.3 we discussed the importance of choosing which variables to calibrate against (the variables in W_1) versus which variables to use as controls only (the variables in W_0). We next briefly illustrate this in our empirical application. The results in table 1 and figures 1 and 2 all include state fixed effects as controls, but do not use them for calibration; that is, these variables are in W_0. Next we consider the impact of instead putting them into W_1 and calibrating the magnitude of selection on unobservables against them, in addition to the geographic and climate controls already in W_1.

Figure 3 shows figures corresponding to the top right and bottom right plots in figure 1, but now also using state fixed effects for calibration. We first see that the identified sets for β_{long} (left plot) are larger, for any fixed $\bar{r}_X$. This makes sense because the *meaning* of $\bar{r}_X$ has changed with the change in calibration controls. In particular, the breakdown point $\bar{r}_X^{\text{bp}}$ is now about 30%, whereas previously it was about 80%. This change can be understood as a consequence of equation (4). By including state fixed effects—which have a large amount of explanatory power—in our calibration controls, we have increased the magnitude of selection on observables. Holding selection on unobservables fixed, this implies that r must decrease. This discussion reiterates the point that the magnitude of $\bar{r}_X$ must always be interpreted as dependent on the set of calibration controls. For example, our finding in figure 3 that the estimated $\bar{r}_X^{\text{bp}}$ is about 30% should not be interpreted as saying that the results are sensitive; in fact, an effect about 30% as large as these calibration covariates is substantially large, and so it may be that we do not expect the omitted variable to have such a large additional impact.

The right plot in figure 3 shows the estimated three-dimensional breakdown frontiers. The frontiers have all shifted inward, compared to the bottom right plot of figure 1 which did not use state fixed effects for calibration. Consequently, a superficial reading of this plot may suggest that the results for average Republican vote share are no longer robust. However, as we just emphasized in our discussion of the left plot, by including state fixed effects in the calibration covariates W_1, we are changing the meaning of all three sensitivity parameters. Since the expanded set of calibration covariates has substantial explanatory power, even a relaxation like $(\bar{r}_Y, \bar{r}_X, \bar{c}) = (50\%, 50\%, 1)$—which is below the breakdown frontier and hence allows us to conclude that β_{long} is positive—could be considered to be a large impact of omitted variables. So these figures do not change our overall conclusions about the robustness of the analysis for average Republican vote share.

Finally, as we emphasized in section 3.3, our discussion about the choice of calibration covariates are not unique to our analysis; they apply equally to all other methods that use covariates to calibrate the magnitudes of sensitivity parameters in some way.

5.4 Empirical Conclusions

Overall, a sensitivity analysis based on our new methods leads to a more nuanced empirical conclusion than originally obtained by Bazzi et al. (2020). We found that their analysis using questionnaire based outcomes is quite sensitive to the presence of omitted variables, while their analysis using property tax levels and voting patterns is robust. This has several empirical implications.

First, the questionnaire based outcomes are the most easily interpretable as measures of opposition to redistribution, regulation, and preferences for small government. In contrast, it is less clear that property taxes and Republican presidential vote share alone should be interpreted as direct measures of opposition to redistribution. So the fact that the questionnaire based outcomes are sensitive to the presence of omitted variables suggests that Bazzi et al.'s overall conclusion in support of the "frontier thesis" should be considered more tentative than previously stated. Second, it suggests that the impact of frontier life may occur primarily through broader behavior based channels like elections, rather than individuals' more specific policy preferences and behavior in their personal lives. It may be useful to explore this difference in future empirical work.

Finally, note that Bazzi et al. (2020) perform a wide variety of additional supporting analyses that we have not examined here. It would be interesting to apply our methods to these additional analyses, to see whether allowing for endogenous controls affects the sensitivity of these other analyses. In particular, their figure 5 considers another set of outcome variables: Republican vote share in each election from 1900 to 2016. In contrast, our analysis above looked only at one election outcome: the average Republican vote share over the five elections from 2000 to 2016. They use these additional baseline estimates along with a qualitative discussion of the evolution of Republican party policies over time to argue that the average Republican vote share outcome between 2000–2016 can be interpreted as a measure of opposition to redistribution. It would be interesting to see how these supporting results hold up to a sensitivity analysis that allows for endogenous controls.

6 Conclusion

As Angrist and Pischke (2017) emphasize, most researchers do not expect to identify causal effects for many variables at the same time. Instead, we target a single variable, called the treatment, while the other variables are called controls, and are included solely to aid identification of causal effects for the treatment variable. These control variables are therefore usually thought to be endogenous. And yet most of the available methods for doing sensitivity analysis rely on an assumption that these controls are exogenous. This raises the question of whether these methods for assessing sensitivity are themselves sensitive to allowing the controls to be endogenous. In this paper we provide a new approach to assessing the sensitivity of selection on observables assumptions in linear models. Our results have two key features that distinguish them from existing methods: First, they allow the controls to be endogenous. Second, our first main result only requires researchers to pick a single sensitivity parameter. In contrast, several existing methods rely on exogenous controls *and* require researchers to pick or reason about at least two different sensitivity parameters. Our results are also simple to implement in practice, via an accompanying Stata package `regsensitivity`. Finally, in our empirical application to Bazzi et al.'s (2020) study of the impact of frontier life on modern culture, we showed that allowing for endogenous controls does matter in practice, leading to more nuanced empirical conclusions than those obtained in Bazzi et al. (2020).

Internal and External Calibrations of Sensitivity Parameters

Our analysis raises several open questions for the broader literature on sensitivity analysis. A typical method specifies continuous relaxations or deviations from one's baseline assumptions and then asks: How much can we relax or deviate from our baseline assumptions until our conclusions breakdown? Answering this question requires *calibrating* the sensitivity parameters: How do we know when these sensitivity parameters are 'large' in some sense? A key insight of Altonji, Elder, and Taber (2005) was that we could answer this question by performing an *internal* calibration, by comparing the magnitude of the sensitivity parameters to the magnitude other parameters in the model. However, as we have discussed in this paper, the value of such internal calibrations is to provide (1) a unit free sensitivity parameter which (2) can be interpreted in terms of the effects of observed variables. It does *not* provide a single universal threshold for what is or is not a robust result. In particular, the choice of which observed variables to calibrate against will change the scale and interpretation of the sensitivity parameter. Consequently, the value 1 should be considered a unit free reference point, not a threshold for robustness.

This observation leads to several questions: How should researchers choose the covariates against which they calibrate? And for any given choice of covariates, if 1 is not the threshold for robustness, what is the threshold? The difficulty of answering these questions speaks to the difficulty of using a single dataset to assess sensitivity and to calibrate those sensitivity parameters. An alternative approach is *external* calibration, where a secondary dataset is used to calibrate the sensitivity parameters. This approach uses sensitivity parameters that are not defined relative to other parameters in the model, and does not require researchers to pick a set of covariates to calibrate against. Such absolute sensitivity parameters are common in the literature on nonparametric sensitivity analysis (e.g., Rosenbaum 1995, 2002 or Masten and Poirier 2018). This external calibration approach is also common in the literature on measurement error or missing data, where secondary datasets are used to assess the extent of measurement error or the strength of violations of missing at random assumptions. It is possible that some combination of both internal and external calibration approaches will lead to the most robust set of methods for assessing the role of selection on unobservables in empirical work.

References

Abadir, K. M. and J. R. Magnus (2005): *Matrix Algebra*, Cambridge University Press.

Altonji, J. G., T. Conley, T. E. Elder, and C. R. Taber (2019): "Methods for using selection on observed variables to address selection on unobserved variables," *Working paper*.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005): "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools," *Journal of Political Economy*, 113, 151–184.

——— (2008): "Using selection on observed variables to assess bias from unobservables when evaluating swan-ganz catheterization," *The American Economic Review P&P*, 98, 345–50.

Angrist, J. D. and J.-S. Pischke (2009): *Mostly Harmless Econometrics*, Princeton University Press.

——— (2015): *Mastering 'Metrics: The Path from Cause to Effect*, Princeton University Press.

——— (2017): "Undergraduate econometrics instruction: through our classes, darkly," *Journal of Economic Perspectives*, 31, 125–44.

Barnow, B. S., G. G. Cain, and A. S. Goldberger (1980): "Issues in the analysis of selectivity bias," in *Evaluation Studies 5*, Sage Publications, 42–59.

Bazzi, S., M. Fiszbein, and M. Gebresilasse (2020): "Frontier culture: The roots and persistence of "rugged individualism" in the United States," *Econometrica*, 88, 2329–2368.

Chalak, K. (2019): "Identification of average effects under magnitude and sign restrictions on confounding," *Quantitative Economics*, 10, 1619–1657.

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

Frölich, M. (2008): "Parametric and nonparametric regression in the presence of endogenous control variables," *International Statistical Review*, 76, 214–227.

Heckman, J. J. and V. J. Hotz (1989): "Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training," *Journal of the American Statistical Association*, 84, 862–874.

Heckman, J. J. and R. Robb (1985): "Alternative methods for evaluating the impact of interventions," *Longitudinal Analysis of Labor Market Data*, 156–245.

Heckman, J. J. and E. J. Vytlacil (2007): "Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments," *Handbook of Econometrics*, 6, 4875–5143.

Hünermund, P. and B. Louw (2020): "On the nuisance of control variables in regression analysis," *arXiv preprint arXiv:2005.10314v3*.

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

Imbens, G. W. and D. B. Rubin (2015): *Causal Inference for Statistics, Social, and Biomedical Sciences*, Cambridge University Press.

Krauth, B. (2016): "Bounding a linear causal effect using relative correlation restrictions," *Journal of Econometric Methods*, 5, 117–141.

LaLonde, R. J. (1986): "Evaluating the econometric evaluations of training programs with experimental data," *The American Economic Review*, 604–620.

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

——— (2020): "Inference on breakdown frontiers," *Quantitative Economics*, 11, 41–111.

——— (2021): "Salvaging falsified instrumental variable models," *Econometrica*, 89, 1449–1469.

Masten, M. A., A. Poirier, and L. Zhang (2021): "Assessing sensitivity to unconfoundedness: Estimation and inference," *arXiv preprint arXiv:2012.15716*.

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

Rosenbaum, P. R. (1995): *Observational Studies*, Springer.

——— (2002): *Observational Studies*, Springer, second ed.

Stock, J. H. and M. W. Watson (2011): *Introduction to Econometrics, 3rd Edition*, Pearson.

Tabellini, M. (2020): "Gifts of the immigrants, woes of the natives: Lessons from the age of mass migration," *The Review of Economic Studies*, 87, 454–486.

Zhang, C., C. Cinelli, B. Chen, and J. Pearl (2021): "Exploiting equality constraints in causal inference," in *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, ed. by A. Banerjee and K. Fukumizu, PMLR, vol. 130 of *Proceedings of Machine Learning Research*, 1630–1638.

A The Residualization Approach to Endogenous Controls

In this appendix we discuss an alternative approach to allowing for endogenous controls that has been suggested by several papers. This approach is based on redefining the sensitivity parameter so that it simultaneously measures the magnitude of endogenous controls as well as the magnitude of selection on unobservables. We argue that this approach substantially limits its usefulness, because it changes the scale of the sensitivity parameter. For Oster's (2019) analysis (hereafter Oster), we show this scale change implies that $\delta = 1$ is no longer a natural reference point. That is a concern because an important aspect of Oster's overall analysis centers on comparing certain values of δ with 1. For example, Oster's Proposition 1, Proposition 2, and Corollary 1 all assume that δ equals 1. Oster's Proposition 3 derives a breakdown point type value of δ, which is then compared to 1. Values larger than 1 are viewed as 'large' while values smaller than 1 are viewed as 'small'. Such comparisons are also how empirical researchers routinely use

Oster's results, like in the example we study in section 5. To correctly interpret these comparisons of δ with 1 as statements about whether selection on unobservables is larger or smaller than selection on observables, researchers must therefore make the exogenous controls assumption.

In section A.1 we first give a simple example where redefinition does not address an underlying identification problem. The purpose of this example is to show that redefinition is not always an appropriate method to relax an assumption, in a much simpler setting than the regression sensitivity analysis considered by Oster. In section A.2 we then discuss Oster's redefinition approach to allowing for endogenous controls. In section A.3 we briefly show that Krauth's (2016) and Cinelli and Hazlett's (2020) residualization approaches to endogenous controls have similar drawbacks as Oster's.

A.1 Projecting Unobservables, "Normalizations," and Interpreting Parameters

We say the controls W_1 are endogenous when they are correlated with the omitted variables of concern, W_2. These variables W_2, however, are not observed. Since they are unobserved, why is it not without loss of generality to simply assume that they are uncorrelated with W_1, $\text{cov}(W_1, W_2) = 0$? Is this not simply a "normalization"? We answer this question in section A.2. To build intuition, first consider a simple model where potential outcomes satisfy

$$Y(x) = \beta_c x + U,$$

where U is unobservable, X is a treatment variable, and $Y = Y(X)$. The parameter of interest is β_c. Similar to how $\delta = 1$ is a special value of Oster's parameter, $\beta_c = 0$ is often a special value of interest here. As we describe in section A.2 below, Oster (2019, Appendix A.1) suggests projecting W_2 onto $(1, W_1)$, obtaining the residual $W_2^{\perp W_1}$, and then replacing W_2 with $W_2^{\perp W_1}$ everywhere. Thus, even though $\text{cov}(W_1, W_2)$ is not necessarily zero, $\text{cov}(W_1, W_2^{\perp W_1}) = 0$. The analogous procedure here is to project U onto $(1, X)$, obtain the residual $U^{\perp X}$, and then replace U with $U^{\perp X}$ everywhere. So even though $\text{cov}(X, U)$ is not necessarily zero, $\text{cov}(X, U^{\perp X}) = 0$.

This procedure, however, is not without loss of generality. In particular, we can write

$$\begin{aligned} Y &= \beta_c X + (\rho X + U^{\perp X}) \\ &= (\beta_c + \rho) X + U^{\perp X} \end{aligned}$$

where $\rho = \text{cov}(U, X)/\text{var}(X)$ and $U^{\perp X} = U - \rho X$. This implies that OLS of Y on $(1, X)$ gives $\beta_c + \rho$ as the coefficient on X. Consequently, the interpretation of the OLS estimand depends substantially on whether we believe X is uncorrelated with the true unobservable U—in which case we learn the causal effect β_c from OLS—or whether we think the unobservable is really just a projection residual $U^{\perp X}$ that is only uncorrelated with X by construction—in which case OLS only gives us the non-causal projection coefficient. In the first case, 0 is a special value of interest, representing no causal effect. But in the second case, 0 is not necessarily a relevant reference point.

In this context, we see that "normalizing" $\text{cov}(X, U) = 0$ is not typically considered to be a solution to the problem that $\text{cov}(X, U)$ may be nonzero since the composite parameter $\beta_c + \rho$ does not have a causal interpretation when the true U is correlated with X. This observation motivates alternative identification strategies like instrumental variable methods, which explicitly allow for $\text{cov}(X, U) \neq 0$. In our discussion of redefinition and "normalizations" in the context of sensitivity analysis below, we will make a similar claim: Redefining the parameters to allow for nonzero correlations between W_1 and W_2 changes the interpretation of the sensitivity parameter substantially. In particular, 1 is no longer a relevant reference point after this redefinition. This observation therefore motivates the alternative sensitivity analysis that we develop in section 3.

A.2 Oster (2019)

Shortly after stating the exogenous controls assumption, Oster (2019) says

> "As in AET [Altonji, Elder, and Taber (2005)], the orthogonality of W_1 and W_2 is central to deriving the results and may be somewhat at odds with the intuition that the observables and the unobservables are "related." In practice, the weight of this assumption is in *how we think*

about the proportionality condition. In Appendix A.1, I show formally that if we begin with a case in which the elements of W_1 are correlated with W_2 we can always redefine W_2 such that the results hold under some value of δ." (page 192, emphasis added)

Here Oster explains that the main implication of her approach to removing the exogenous controls assumption is that it changes the interpretation of δ. In this section we argue that this reinterpretation substantially limits the usefulness of Oster's results. For example, we show that it can lead researchers to find that their conclusions are robust even when an analysis based on the original δ shows non-robustness, or vice versa.

A Brief Review of Oster's δ

First we briefly define and discuss Oster's sensitivity parameter. Here we continue to use the same notation as in our analysis in sections 2 and 3. We also continue to assume W_2 is a scalar, for simplicity. Following the analysis of Altonji et al. (2005), Oster (2019) recommends that we measure the magnitude of selection on unobservables via the parameter

$$\delta_{\text{orig}} = \frac{\operatorname{cov}(X, \gamma_2 W_2)}{\operatorname{var}(\gamma_2 W_2)} \Bigg/ \frac{\operatorname{cov}(X, \gamma_1' W_1)}{\operatorname{var}(\gamma_1' W_1)}, \tag{9}$$

which is commonly called *Oster's δ*. We denote it by δ_{orig} to distinguish it by the redefined version we discuss later. This parameter depends on two terms:

1. ("Selection on unobservables") In the numerator we regress X on $(1, \gamma_2 W_2)$ and get the coefficient on the index $\gamma_2 W_2$.
2. ("Selection on observables") In the denominator we regress X on $(1, \gamma_1' W_1)$ and get the coefficient on the index $\gamma_1' W_1$.

δ_{orig} is the ratio of these two regression coefficients. δ_{orig} is not known from the data since it depends on W_2, which is not observed. It also depends on (γ_1, γ_2) which are also not generally known. Finally, for reference, we formally define the *exogenous controls* assumption as follows.

Assumption A8 (Exogenous controls). $\operatorname{cov}(W_{1k}, W_2) = 0$ for all components W_{1k} of W_1.

Formal Analysis of Oster's Redefinition Approach

Oster derives all of her results using the sensitivity parameter δ_{orig} combined with assumption A8. To remove this assumption, she suggests replacing δ_{orig} with a different sensitivity parameter. We discuss this approach next.

For simplicity, suppose W_1 is scalar. W_2 in Oster's notation is equivalent to $\gamma_2 W_2$ in our notation. For the discussion below we'll assume $\gamma_2 = 1$. This is not required, but it implies that Oster's W_2 and ours are the same, which makes the derivations and comparisons clearer. With these simplifying assumptions, equation (9) becomes

$$\delta_{\text{orig}} = \frac{\operatorname{cov}(X, W_2)}{\operatorname{var}(W_2)} \Bigg/ \frac{\operatorname{cov}(X, \gamma_1 W_1)}{\operatorname{var}(\gamma_1 W_1)}.$$

In her Appendix A.1, Oster notes that if W_1 and W_2 are correlated, we can consider the linear projection of W_2 onto $(1, W_1)$:

$$W_2 = \rho W_1 + W_2^{\perp W_1}.$$

Here $\operatorname{cov}(W_1, W_2^{\perp W_1}) = 0$ by construction, and $\rho = \operatorname{cov}(W_2, W_1)/\operatorname{var}(W_1)$. Hence we could instead define δ based on the residuals $W_2^{\perp W_1}$ rather than the original covariate W_2:

$$\tilde{\delta} = \frac{\operatorname{cov}(X, W_2^{\perp W_1})}{\operatorname{var}(W_2^{\perp W_1})} \Bigg/ \frac{\operatorname{cov}(X, \gamma_1 W_1)}{\operatorname{var}(\gamma_1 W_1)}.$$

Oster defines this $\tilde{\delta}$ in her Appendix A.1 and then suggests that all of the previous analysis can proceed as before, using $\tilde{\delta}$ instead of δ_{orig}. This is slightly incorrect, however, because the coefficient on W_1 needs to be adjusted. For example, without this adjustment, which we describe below, Oster's Proposition 1 does not hold if we replace W_2 with $W_2^{\perp W_1}$ and δ_{orig} with $\tilde{\delta}$.

To see the issue, we can write the outcome equation as

$$\begin{aligned}
Y &= \beta_{\text{long}} X + \gamma_1 W_1 + W_2 + Y^{\perp X, W_1, W_2} \\
&= \beta_{\text{long}} X + \gamma_1 W_1 + (\rho W_1 + W_2^{\perp W_1}) + Y^{\perp X, W_1, W_2} \\
&= \beta_{\text{long}} X + (\gamma_1 + \rho) W_1 + W_2^{\perp W_1} + Y^{\perp X, W_1, W_2}.
\end{aligned}$$

This is the correct version of the outcome equation after residualizing the unobserved covariates. In particular, the ρW_1 term cannot be absorbed into the residual because then the residual would no longer be uncorrelated with W_1. Applying equation (9) to this equation yields

$$\delta_{\text{resid}} = \frac{\text{cov}(X, W_2^{\perp W_1})}{\text{var}(W_2^{\perp W_1})} \Bigg/ \frac{\text{cov}(X, (\gamma_1 + \rho) W_1)}{\text{var}((\gamma_1 + \rho) W_1)}. \tag{10}$$

This version of the sensitivity parameters matches the outcome equation, and thus all of Oster's results go through if we replace W_2 with $W_2^{\perp W_1}$ and δ_{orig} with δ_{resid}. Next we discuss the implications of this substitution. We provide two formal results: Proposition 3 relates δ_{resid} to the original sensitivity parameter δ_{orig}. Proposition 4 shows how to interpret δ_{resid} in terms of the three sensitivity parameters we use in our section 3 analysis.

First we compare Oster's redefined and original sensitivity parameters.

Proposition 3. Suppose $\text{var}(W_1) = 1$ and $\text{var}(W_2) = 1$. Suppose $\gamma_1 \neq 0$, $\gamma_2 = 1$, and $|\rho| \neq 1$. The following holds:

$$\delta_{\text{resid}} = \left(1 + \frac{\rho}{\gamma_1}\right) \frac{\delta_{\text{orig}} - \rho\gamma_1}{1 - \rho^2}.$$

Here we assume $\text{var}(W_1) = 1$, $\gamma_2 = 1$, and $\gamma_1 \neq 0$ for simplicity only, since the purpose of this proposition is solely to help clarify the difference between the two definitions of δ. Proposition 3 shows that the δ based on the residualized W_2 is a scaled version δ_{orig}, the δ based on the original unobserved covariate W_2. Moreover, the scale term depends on ρ, which measures of the endogeneity of the observed control W_1. We see that if $\rho = 0$, so that W_1 is an exogenous control, then the two versions of δ are the same. If $\rho \neq 0$, however, the two versions of δ can be very different.

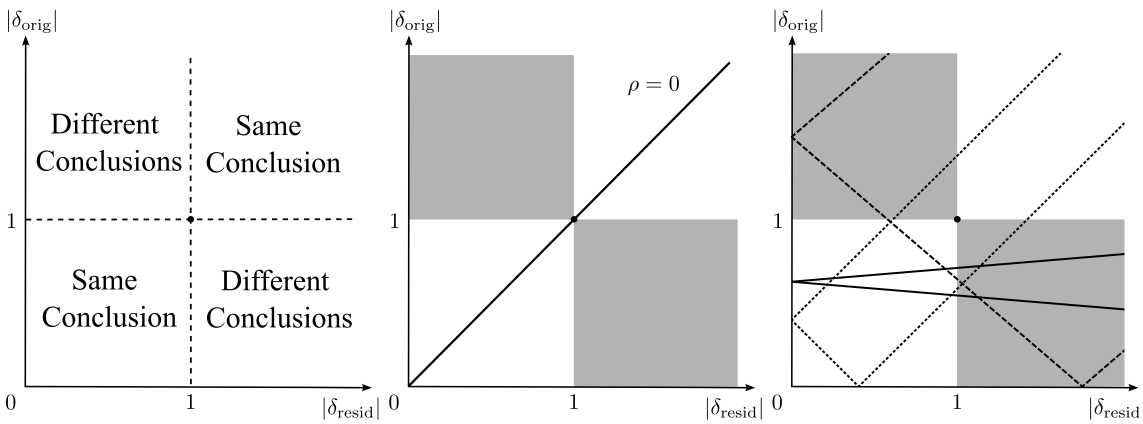
Proposition 3 has several implications. First, consider a researcher who wants to compute the identified set for β_{long} by applying Oster's Proposition 2. For example, this set could be used to present bias adjusted coefficient estimates. Suppose this researcher prefers to make statements about δ_{orig}. Without the exogenous controls assumption, Oster's Proposition 2 requires that they instead make statements about δ_{resid}. So to use that result, this researcher must first apply our Proposition 3 to translate a statement about δ_{orig} into a statement about δ_{resid}. In particular, to perform this translation, this researcher must also select specific values of the unknowns γ_1 and ρ (and, more generally, γ_2, although we have assumed it is 1 for simplicity here).

Second, consider a researcher who has used Oster's Proposition 3 to compute a breakdown point. This is the most common way that researchers use Oster's results. Without the exogenous controls assumption, Oster's Proposition 3 delivers a breakdown point on the δ_{resid} scale; denote it by $\delta_{\text{resid}}^{\text{bp}}$. Suppose the researcher wants to translate this breakdown point result into a result about δ_{orig}, which is a statement about selection on observables versus unobservables in terms of the original covariates. To perform this translation, the researcher can invert our result in Proposition 3 to get

$$\delta_{\text{orig}} = \delta_{\text{resid}} \frac{\gamma_1(1 - \rho^2)}{\gamma_1 + \rho} + \rho\gamma_1. \tag{11}$$

So once again, the translation from δ_{resid} depends on the specific unknowns γ_1 and ρ (and, more generally, γ_2). Moreover, equation (11) has several further implications about researchers' findings of robustness. As

Figure 4: The relationship between robustness conclusions drawn based on $|\delta_{\text{resid}}|$ and those drawn based on $|\delta_{\text{orig}}|$. Left: Regions where the two parameters yield either the same or different conclusions about robustness. Middle: Under exogenous controls, the two parameters are the same and always yield the same conclusions. Right: With endogenous controls, the two parameters can be very different and therefore yield different conclusions. This plot shows three values of (ρ, γ_1): $(0.9, 0.7)$–solid line, $(0.3, 5)$–dashed line, and $(0.05, 8)$–dotted line. See body text for further discussion.



mentioned earlier, an important aspect of Oster's overall analysis is that there is a very special value of δ: 1. Like in our empirical application in section 5, empirical researchers commonly compare the absolute value of Oster's breakdown point to 1 to assess whether their study is robust to selection on unobservables. If researchers are instead interested in using δ_{orig} to assess sensitivity, then the breakdown point $\delta^{\text{bp}}_{\text{resid}}$ must be translated using equation (11). Let

$$\delta^{\text{bp}}_{\text{orig}} = \delta^{\text{bp}}_{\text{resid}} \frac{\gamma_1(1-\rho^2)}{\gamma_1 + \rho} + \rho\gamma_1 \tag{12}$$

denote the translated value of the breakdown point. From this equation we immediately see that, when exogenous controls fails ($\rho \neq 0$), it is possible for $|\delta^{\text{bp}}_{\text{resid}}| > 1$ even though $|\delta^{\text{bp}}_{\text{orig}}| < 1$. That is, *researchers might conclude that their results* are *robust based on the residualized variables even though they would conclude their results are* not *robust based on the original variables.*

Figure 4 illustrates this property. All three figures plot the value of $|\delta_{\text{resid}}|$ on the horizontal axis and $|\delta_{\text{orig}}|$ on the vertical axis. These plots can be thought of as taking the plot of δ_{orig} versus δ_{resid} on the 2D plane and folding all four quadrants onto the positive quadrant. The left plot divides the area into two kinds of regions. The first region is where $\delta^{\text{bp}}_{\text{orig}}$ and $\delta^{\text{bp}}_{\text{resid}}$ lead to the same conclusion, based on Oster's recommendation of comparing them with the value 1. This happens when both are less than one in absolute value, or both are greater than one in absolute value. The second region is where they lead to different conclusions. The middle plot shows equation (11) when $\rho = 0$, so exogenous controls holds. In this case, $\delta_{\text{orig}} = \delta_{\text{resid}}$. Consequently, they always lead to the same conclusion about robustness. Here the shaded areas are where they lead to different conclusions. Since the two parameters are the same under exogenous controls, the line relating the two is the 45 degree line, which never passes through the shaded regions. The right plot shows the case with endogenous controls. Here we plot equation (11) for three different choices of (γ_1, ρ) with $\rho \neq 0$, showing the wide variety of possible relationships between δ_{orig} and δ_{resid}. With endogenous controls, we see that for many dgps, the lines cross the shaded region, and hence the two parameters can lead to different conclusions about robustness.

Thus far we have emphasized that researchers' conclusions about robustness depend greatly on whether they intend on measuring sensitivity using residualized variables or the original variables. Put differently: If researchers think that $\delta_{\text{orig}} = 1$ is a reasonable reference point, then by Proposition 3, $\delta_{\text{resid}} = 1$ is *not* the relevant reference point. Instead, researchers must translate between the two versions of the sensitivity parameter, and this translation requires thinking about both the coefficient on W_1 in the outcome equation and the magnitude of the relationship between W_1 and W_2.

More generally, as can be seen from equation (10), the statement "$\delta_{\text{resid}} = 1$" is an assumption on *three* different underlying relationships:

1. The relationship between W_1 and W_2, as measured by ρ.
2. The relationship between outcomes and (W_1, W_2), as measured by γ_1 and γ_2.
3. The relationship between X and (W_1, W_2), as measured by $\text{cov}(X, W_1)$ and $\text{cov}(X, W_2)$.

Only the third relationship is actually about treatment *selection* on covariates, however. Our next result provides another way to see that δ_{resid} measures three different underlying relationships simultaneously. Let

$$c = \text{corr}(W_1, W_2), \qquad r_Y = \frac{\gamma_2\sigma_2}{\gamma_1\sigma_1}, \qquad \text{and} \qquad r_X = \frac{\pi_2\sigma_2}{\pi_1\sigma_1}.$$

Here $(\gamma_1, \gamma_2, \pi_1, \pi_2)$ are defined as in section 2, $\sigma_1 = \sqrt{\text{var}(W_1)}$, and $\sigma_2 = \sqrt{\text{var}(W_2)}$. Note that $c = \rho$ if $\text{var}(W_1) = 1$ and $\text{var}(W_2) = 1$. Our sensitivity analysis in section 3 considers bounds $\bar{r}_X$ on $|r_X|$, $\bar{r}_Y$ on $|r_Y|$, and $\bar{c}$ on $|c|$. The following result therefore relates Oster's sensitivity parameter to these three underlying measures that our analysis makes assumptions about.

Proposition 4. Suppose γ_1 and π_1 are nonzero, so that δ_{resid} is well defined. Suppose $r_Y(1 + r_X c)$ is nonzero. Then

$$\delta_{\text{resid}} = \frac{r_X(1 + r_Y c)}{r_Y(1 + r_X c)}.$$

Proposition 4 generalizes the derivations in section 6.3 of Cinelli and Hazlett (2020), who looked at the special case where the controls are exogenous, $\text{cov}(W_1, W_2) = 0$, and showed

$$\delta_{\text{orig}} = \frac{r_X}{r_Y}.$$

This follows by setting $c = 0$ in Proposition 4 and then applying Proposition 3. They emphasized that, under the exogenous controls assumption, Oster's parameter is a *double ratio*: It compares the relative effect selection on unobservables to observables (r_X) *relative to* the relative effect of unobservables versus observables on outcomes (r_Y). Thus a large value of δ_{orig} is consistent with a *small* amount of selection on unobservables (small r_X) so long as the unobservables matter *less* for outcomes than the observables (small*er* r_Y). They write that Oster "asks users to reason about a quantity that is very difficult to understand" (page 63). They focused on the fact that γ_1 and γ_2 appear in the definition of δ_{orig}. They continued to assume exogenous controls, however. Here we point out that if you also allow for endogenous controls, then the interpretation of Oster's sensitivity parameter δ_{resid} becomes even more delicate.

Empirical Illustration

Thus far we have shown theoretically how the interpretation of δ_{resid} depends on more than just selection on unobservables versus observables. Next we discuss the implications of this interpretation in the context of a simple empirical example: Assessing the impact of omitted ability when measuring the returns to education.

Suppose the outcome Y is log wages, treatment X is education, the observable W_1 is parents' education, and the omitted variable of concern W_2 is ability. Suppose for simplicity that we normalize the variance of W_1 and W_2 to 1. First we discuss how to interpret the sensitivity parameters we use in section 3. $\bar{r}_X$ is the maximum effect of a one standard deviation increase in ability on the amount of education received relative to the effect of a one standard deviation increase of parent's education on the amount of education received. $\bar{r}_Y$ is defined similarly, but swapping education for the outcome, log wages. $\bar{c}$ is the maximum correlation between parents' education and the child's ability. In this setting, the exogenous controls assumption ($\bar{c} = 0$) is unlikely to hold, since we typically expect parents with higher education to have children with higher unobserved ability.

Oster's original sensitivity parameter δ_{orig} equals the ratio λ_2/λ_1 of two regression coefficients:

1. λ_2, the coefficient from OLS of education on a constant and $\gamma_2 \times$ ability.

2. λ_1, the coefficient from OLS of education on a constant and $\gamma_1 \times$ parents' education.

And recall that γ_2 is the coefficient on ability in OLS of log wages on a constant, child's education, parents' education, and ability. γ_1 is the coefficient on parent's education in that same regression. This ratio λ_2/λ_1 is arguably a straightforward sensitivity parameter to interpret. However, to use Oster's results along with this specific interpretation, we must assume exogenous controls, $\text{cov}(W_1, W_2) = 0$. That is, we must assume child's ability and parents' education are uncorrelated.

To avoid this exogenous controls assumption, we can instead use Oster's redefined sensitivity parameter, δ_{resid}. This parameter is also a ratio λ_2^*/λ_1^* of two regression coefficients:

1. λ_2^*, the coefficient from OLS of education on a constant and $\gamma_2 \times$residualized ability, where residualized ability is the residual from a projection of ability onto parents' education.

2. λ_1^*, the coefficient from OLS of education on a constant and

$$\gamma_1 \times \text{parents' education} + \text{corr}(\text{ability}, \text{parents' education}) \times \gamma_2 \times \text{parents' education}.$$

These two parameters λ_2^* and λ_1^* are arguably substantially more difficult to interpret than the original parameters λ_2 and λ_1. In particular, notice that ability now explicitly enters both parameters.

The motivation of Oster's original sensitivity parameter δ_{orig} is to compare the effect of $\gamma_2 \times$ ability on treatment X, relative to the effect of $\gamma_1 \times$ parents' education on treatment X. We could then interpret δ_{orig} substantively. For example, one could argue that a large δ_{orig} is unlikely since it would mean that ability is much more important than parents' education in determining the level of education that the child receives. This kind of discussion is not valid when using δ_{resid}, however. That is because the comparison is now between γ_2 times residualized ability and the complicated composite variable

$$\gamma_1 \times \text{parents' education} + \text{corr}(\text{ability}, \text{parents' education}) \times \gamma_2 \times \text{parents' education}.$$

In particular, if child's ability and parents' education are strongly correlated (ρ far from zero), then these two new variables, $\gamma_2(W_2 - \rho W_1)$ and $(\gamma_1 + \rho\gamma_2)W_1$, are very different from the original variables $(\gamma_2 W_2, \gamma_1 W_1)$. Consequently, statements like "we think parents' education is much more important than ability in determining the level of education that the child receives" are no longer helpful when attempting to interpret the magnitude of δ_{resid}. Put differently: The original goal was to interpret the separate impacts of observables from unobservables, but using δ_{resid} now requires us to make comparisons of variables that mix both observables and unobservables.

Summary

Overall, we have shown that relaxing exogenous controls by replacing δ_{orig} with δ_{resid} implies that statements about δ_{resid} are not statements that compare the magnitude of selection on unobservables with the magnitude of selection on observables. Instead, it also requires researchers to make an implicit judgment about endogeneity of the control variables. This undermines the argument that $\delta_{\text{resid}} = 1$ is a natural reference point. Moreover, it implies that researchers may conclude that their results are robust based on examining δ_{resid}, whereas a translation to δ_{orig} may instead show that their results are not robust.

In contrast, our analysis in section 3 allows researchers to reason about all three relevant relationships separately: the endogeneity of the controls via $\bar{c}$, the relative magnitudes of selection on unobservables versus on observables via $\bar{r}_X$, and the relative effects of the covariates on outcomes via $\bar{r}_Y$. Moreover, our first main result (Theorem 2) does not require any assumptions on $\bar{c}$ or $\bar{r}_Y$ at all, thus allowing researchers to isolate the impact of selection on unobservables alone on the sensitivity of their findings.

A.3 Other Papers

In this section we briefly show that Krauth's (2016) and Cinelli and Hazlett's (2020) analyses have similar drawbacks as Oster's (2019). To ease comparisons, here we translate their definitions into our notation. Suppose for simplicity that W_1 and W_2 are scalars and that there are no control covariates W_0. First consider Krauth (2016). As in section A.2, consider the linear projection of W_2 onto $(1, W_1)$:

$$W_2 = \rho W_1 + W_2^{\perp W_1}.$$

Also as in section A.2, the outcome equation can be written as

$$Y = \beta_{\text{long}} X + (\gamma_1 + \gamma_2 \rho) W_1 + \gamma_2 W_2^{\perp W_1} + Y^{\perp X, W_1, W_2}.$$

Krauth defines the sensitivity parameter

$$\lambda = \frac{\text{corr}(X, \gamma_2 W_2^{\perp W_1})}{\text{corr}(X, (\gamma_1 + \gamma_2 \rho) W_1)}. \tag{13}$$

Here we explicitly include the projection of W_2 onto W_1 in the definition of this sensitivity parameter. Krauth does this implicitly in his notation; see his footnote 1 on page 119. Compare equation (13) with equation (10) for Oster's δ_{resid}. Like Oster's parameter, Krauth's λ is a composite sensitivity parameter that depends on three different underlying relationships. Our discussion of the difficulties of interpreting Oster's δ_{resid} therefore apply similarly to Krauth's λ.

Next consider Cinelli and Hazlett (2020). In their section 4.4, they formally assume exogenous controls and work with two ratio type sensitivity parameters (their equation (21)),

$$k_{X,\text{orig}} = \frac{R^2_{X \sim W_2}}{R^2_{X \sim W_1}} \qquad \text{and} \qquad k_{Y,\text{orig}} = \frac{R^2_{Y \sim W_2 \bullet X}}{R^2_{Y \sim W_1 \bullet X}}.$$

Like Oster (2019) and Krauth (2016), they also suggest that endogenous controls can be allowed by redefining their sensitivity parameters. Specifically, define

$$k_{X,\text{resid}} = \frac{R^2_{X \sim W_2^{\perp W_1}}}{R^2_{X \sim W_1}} \qquad \text{and} \qquad k_{Y,\text{resid}} = \frac{R^2_{Y \sim W_2^{\perp W_1} \bullet X}}{R^2_{Y \sim W_1 \bullet X}}.$$

They suggest that their theoretical results of section 4.4 derived under the exogenous controls assumption continue to hold when the controls are endogenous, so long as we replace W_2 with $W_2^{\perp W_1}$ and $(k_{X,\text{orig}}, k_{Y,\text{orig}})$ with $(k_{X,\text{resid}}, k_{Y,\text{resid}})$. For brevity, we focus on the interpretation of $k_{X,\text{resid}}$, although similar remarks apply to $k_{Y,\text{resid}}$. The following result shows how the original and residualized sensitivity parameters of Cinelli and Hazlett (2020) are related and is analogous to our Proposition 3 above in our analysis of Oster (2019). Recall that $c = \text{corr}(W_1, W_2)$.

Proposition 5. Suppose $\text{var}(W_1) = 1$ and $\text{var}(W_2) = 1$. Suppose $\text{corr}(X, W_1) \neq 0$ and $\text{corr}(W_1, W_2) \neq 1$. Then

$$k_{X,\text{resid}} = \frac{1}{1 - c^2} \left(\frac{\text{corr}(X, W_2)}{\text{corr}(X, W_1)} - c \right)^2.$$

Proposition 5 shows how to write the residualized sensitivity parameter as a function of the ratio of the correlation between treatment and the original unobserved covariate W_2 and the correlation between treatment and the observed covariate W_1. Note that

$$k_{X,\text{orig}} = \left(\frac{\text{corr}(X, W_2)}{\text{corr}(X, W_1)} \right)^2$$

is just the square of this ratio, and hence Proposition 5 relates $k_{X,\text{resid}}$ to $k_{X,\text{orig}}$. An immediate corollary of Proposition 5 is that when exogenous controls holds ($c = 0$), $k_{X,\text{resid}} = k_{X,\text{orig}}$. However, under endogenous controls ($c \neq 0$), these two sensitivity parameters are generally not the same. Indeed, they can be very different. In particular, like Oster (2019), Cinelli and Hazlett (2020) discuss using the value 1 as a cutoff for determining robustness. Hence we could use Proposition 5 to draw a similar figure to 4 relating these two sensitivity parameters. This implies that, for many dgps with endogenous controls, these two parameters can lead to different conclusions about robustness.

Overall, in this subsection we have shown that the residualized versions of Krauth's (2016) and Cinelli and Hazlett's (2020) sensitivity parameters have similar drawbacks as Oster's (2019). Unlike these three papers, our analysis in section 3 uses three distinct sensitivity parameters to measure the three relevant relationships that affect the magnitude of omitted variable bias: the relationship between Y and (W_1, W_2), the relationship between treatment X and (W_1, W_2), and the magnitude of control endogeneity. Our results therefore allow users to both reason about and vary these relationships separately rather than simultaneously.

A.4 Proofs

Proof of Proposition 3. We have

$$\begin{aligned}
\delta_{\text{resid}} &= \frac{\text{cov}(X, W_2^{\perp W_1})}{\text{var}(W_2^{\perp W_1})} \Bigg/ \frac{\text{cov}(X, (\gamma_1 + \rho)W_1)}{\text{var}((\gamma_1 + \rho)W_1)} \\
&= (\gamma_1 + \rho)\frac{\text{cov}(X, W_2^{\perp W_1})}{\text{var}(W_2^{\perp W_1})} \Bigg/ \frac{\text{cov}(X, W_1)}{\text{var}(W_1)} \\
&= \frac{\gamma_1 + \rho}{\gamma_1}\frac{\text{cov}(X, W_2^{\perp W_1})}{\text{var}(W_2^{\perp W_1})} \Bigg/ \frac{\text{cov}(X, \gamma_1 W_1)}{\text{var}(\gamma_1 W_1)} \\
&= \left(1 + \frac{\rho}{\gamma_1}\right)\tilde{\delta}.
\end{aligned}$$

Next note that

$$\text{var}(W_2^{\perp W_1}) = 1 - \rho^2$$

since $\text{var}(W_1) = 1$ and $\text{var}(W_2) = 1$. Also,

$$\text{cov}(X, W_2^{\perp W_1}) = \text{cov}(X, W_2) - \frac{\rho}{\gamma_1}\text{cov}(X, \gamma_1 W_1).$$

So

$$\begin{aligned}
\tilde{\delta} &= \frac{\text{cov}(X, W_2) - \frac{\rho}{\gamma_1}\text{cov}(X, \gamma_1 W_1)}{\text{cov}(X, \gamma_1 W_1)}\frac{\text{var}(\gamma_1 W_1)}{\text{var}(W_2^{\perp W_1})} \\
&= \delta_{\text{orig}}\frac{\text{var}(W_2)}{\text{var}(W_2^{\perp W_1})} - \frac{\rho}{\gamma_1}\frac{\text{var}(\gamma_1 W_1)}{\text{var}(W_2^{\perp W_1})} \\
&= \frac{\delta_{\text{orig}} - \rho\gamma_1}{1 - \rho^2}.
\end{aligned}$$

□

Proof of Proposition 4. Let $\sigma_{12} = \text{cov}(W_1, W_2)$. Recall that $\rho = \sigma_{12}/\sigma_1^2$. We have

$$\begin{aligned}
\delta_{\text{resid}} &= \frac{\text{cov}(X, \gamma_2 W_2^{\perp W_1})}{\text{var}(\gamma_2 W_2^{\perp W_1})} \Bigg/ \frac{\text{cov}(X, (\gamma_1 + \gamma_2\rho)W_1)}{\text{var}((\gamma_1 + \gamma_2\rho)W_1)} \\
&= \frac{(\gamma_1 + \gamma_2\rho)}{\gamma_2}\frac{\text{cov}(X, W_2^{\perp W_1})}{\text{var}(W_2^{\perp W_1})} \Bigg/ \frac{\text{cov}(X, W_1)}{\text{var}(W_1)} \\
&= \frac{(\gamma_1 + \gamma_2\sigma_{12}/\sigma_1^2)}{\gamma_2}\frac{\text{cov}(X, W_2^{\perp W_1})}{\text{cov}(X, W_1)}\frac{\sigma_1^2}{\text{var}(W_2^{\perp W_1})} \\
&= \frac{\gamma_1\sigma_1^2 + \gamma_2\sigma_{12}}{\gamma_2}\frac{\text{cov}(X, W_2^{\perp W_1})}{\text{cov}(X, W_1)}\frac{1}{\text{var}(W_2^{\perp W_1})} \\
&= \sigma_1\sigma_2(r_Y^{-1} + c)\frac{\text{cov}(X, W_2^{\perp W_1})}{\text{cov}(X, W_1)}\frac{1}{\text{var}(W_2^{\perp W_1})}.
\end{aligned}$$

Next we have

$$\begin{aligned}
\operatorname{cov}(X, W_1) &= \operatorname{cov}(\pi_1 W_1 + \pi_2 W_2 + X^{\perp W_1, W_2}, W_1) \\
&= \pi_1 \operatorname{var}(W_1) + \pi_2 \operatorname{cov}(W_1, W_2) \\
&= \pi_1 \sigma_1^2 + \pi_2 \sigma_{12} \\
&= \pi_1 \sigma_1^2 + \pi_2 \sigma_2 \frac{\sigma_{12}}{\sigma_2} \\
&= \pi_1 \sigma_1^2 \left(1 + r_X c\right)
\end{aligned}$$

and

$$\begin{aligned}
\operatorname{cov}(X, W_2) &= \pi_1 \operatorname{cov}(W_1, W_2) + \pi_2 \operatorname{var}(W_2) \\
&= \pi_1 \sigma_{12} + \pi_2 \sigma_2^2 \\
&= \sigma_1 \sigma_2 \left(\pi_1 \frac{\sigma_{12}}{\sigma_1 \sigma_2} + \frac{\pi_2 \sigma_2}{\sigma_1}\right) \\
&= \pi_1 \sigma_1 \sigma_2 (c + r_X).
\end{aligned}$$

So

$$\begin{aligned}
\operatorname{cov}(X, W_2^{\perp W_1}) &= \operatorname{cov}(X, W_2) - \rho \operatorname{cov}(X, W_1) \\
&= \pi_1 \sigma_1 \sigma_2 (c + r_X) - \frac{\sigma_{12}}{\sigma_1^2} \pi_1 \sigma_1^2 \left(1 + r_X c\right) \\
&= \pi_1 \sigma_1 \sigma_2 (c + r_X - c - r_X c^2) \\
&= \pi_1 \sigma_1 \sigma_2 r_X (1 - c^2)
\end{aligned}$$

and

$$\begin{aligned}
\operatorname{var}(W_2^{\perp W_1}) &= \operatorname{var}(W_2 - \rho W_1) \\
&= \operatorname{var}(W_2) + \rho^2 \operatorname{var}(W_1) - 2 \operatorname{cov}(\rho W_1, W_2) \\
&= \sigma_2^2 (1 - c^2).
\end{aligned}$$

Hence

$$\begin{aligned}
\delta_{\text{resid}} &= \sigma_1 \sigma_2 (r_Y^{-1} + c) \frac{\operatorname{cov}(X, W_2^{\perp W_1})}{\operatorname{cov}(X, W_1)} \frac{1}{\operatorname{var}(W_2^{\perp W_1})} \\
&= \sigma_1 \sigma_2 (r_Y^{-1} + c) \frac{\pi_1 \sigma_1 \sigma_2 r_X (1 - c^2)}{\pi_1 \sigma_1^2 \left(1 + r_X c\right)} \frac{1}{\sigma_2^2 (1 - c^2)} \\
&= \frac{r_X (1 + r_Y c)}{r_Y (1 + r_X c)}.
\end{aligned}$$

$\square$

Proof of Proposition 5. We have

$$\begin{aligned}
&R_{X \sim W_2^{\perp W_1}} \\
&= |\operatorname{corr}(X, W_2^{\perp W_1})| \\
&= \left| \frac{\operatorname{cov}(X, W_2^{\perp W_1})}{\sqrt{\operatorname{var}(X)} \sqrt{\operatorname{var}(W_2^{\perp W_1})}} \right|
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{\operatorname{var}(X)}\sqrt{\operatorname{var}(W_2^{\perp W_1})}} \left|\operatorname{cov}(X, W_2) - c\operatorname{cov}(X, W_1)\right| \\
&= \frac{1}{\sqrt{\operatorname{var}(X)}\sqrt{\operatorname{var}(W_2^{\perp W_1})}} \left|\operatorname{corr}(X, W_2)\sqrt{\operatorname{var}(X)}\sqrt{\operatorname{var}(W_2)} - c\cdot\operatorname{corr}(X, W_1)\sqrt{\operatorname{var} X}\sqrt{\operatorname{var}(W_1)}\right| \\
&= \frac{1}{\sqrt{\operatorname{var}(W_2^{\perp W_1})}} \left|\operatorname{corr}(X, W_2) - c\cdot\operatorname{corr}(X, W_1)\right|.
\end{aligned}
$$

Note also that

$$
\begin{aligned}
\operatorname{var}(W_2^{\perp W_1}) &= \operatorname{var}(W_2 - cW_1) \\
&= 1 - c^2
\end{aligned}
$$

given the normalizations. Using this and dividing through by $R_{X\sim W_1} = |\operatorname{corr}(X, W_1)|$ gives

$$
\frac{R_{X\sim W_2^{\perp W_1}}}{R_{X\sim W_1}} = \frac{1}{\sqrt{1-c^2}} \left|\frac{\operatorname{corr}(X, W_2)}{\operatorname{corr}(X, W_1)} - c\right|.
$$

Squaring both sides gives the desired result. □

B Generalization to a Vector of Unobserved Covariates

In sections 2 and 3 we assumed W_2 was a scalar. In this section we show how to generalize our analysis to allow for a vector of unobservable covariates. Our approach in this section is similar to how Altonji et al. (2005) and Oster (2019) allow for vector unobservables.

Let $\widetilde{W}_2$ denote the vector of unobservables. Replace equation (2) with

$$
X = \pi_1' W_1 + \widetilde{\pi}_2' \widetilde{W}_2 + X^{\perp W_1, \widetilde{W}_2}.
$$

Write

$$
\begin{aligned}
\widetilde{\pi}_2' \widetilde{W}_2 &= \sqrt{\operatorname{var}(\widetilde{\pi}_2' \widetilde{W}_2)} \cdot \frac{\widetilde{\pi}_2' \widetilde{W}_2}{\sqrt{\operatorname{var}(\widetilde{\pi}_2' \widetilde{W}_2)}} \\
&\equiv \pi_2 \cdot W_2.
\end{aligned}
$$

Thus we define π_2 to be the standard deviation of the index $\widetilde{\pi}_2' \widetilde{W}_2$ and define W_2 to be the standardized version of this index. All of our technical analysis now applies using this definition of W_2. Importantly, researchers do not need to specify whether they think W_2 is a vector or not a priori; the same technical results hold in both cases. That said, there are a few important points to keep in mind.

1. By construction, $\operatorname{var}(W_2) = 1$. So the normalization assumption A4 continues to hold.
2. Next we'll show that this generalization is consistent with our analysis of the scalar W_2 case. Specifically, suppose $\widetilde{W}_2$ is scalar. Then

$$
\begin{aligned}
\pi_2 &= \sqrt{\operatorname{var}(\widetilde{\pi}_2 \widetilde{W}_2)} \\
&= |\widetilde{\pi}_2| \sqrt{\operatorname{var}(\widetilde{W}_2)}
\end{aligned}
$$

and

$$
W_2 = \frac{\widetilde{W}_2}{\sqrt{\operatorname{var}(\widetilde{W}_2)}}.
$$

Thus W_2 is the standardized version of the original unobserved covariate and π_2 is the standardized version of its coefficient. This is precisely the meaning behind our original normalization assumption A4.

3. Consider A3. Using the above definition of W_2, it is equivalent to

$$\sqrt{\operatorname{var}(\widetilde{\pi}_2'\widetilde{W}_2)} \leq \bar{r}_X \cdot \sqrt{\operatorname{var}(\pi_1' W_1)}.$$

Thus the interpretation of this assumption does not change when the unobservables are a vector.

4. Next consider A6, the assumption that $\|\operatorname{cov}(W_1, W_2)\|_{\Sigma_{\text{obs}}^{-1}} \leq \bar{c}$. In the vector unobservable case, W_2 is the standardized index variable. So A6 can be interpreted as a restriction on the relationship between each observed variable with this standardized index variable. In particular, we can still write

$$\|\operatorname{cov}(W_1, W_2)\|_{\Sigma_{\text{obs}}^{-1}} = \sqrt{R^2_{\widetilde{\pi}_2'\widetilde{W}_2 \sim W_1}},$$

the square root of the R-squared from the regression of the index of unobservables $\widetilde{\pi}_2'\widetilde{W}_2$ on the vector of observables W_1.

5. Consider the following two regressions:

   (a) OLS of Y on $(1, X, W_1, \widetilde{W}_2)$. Denote the coefficient on X by $\beta_{\text{long,vec}}$. Let $(\gamma_{1,\text{vec}}, \gamma_{2,\text{vec}})$ denote the coefficients on $(W_1, \widetilde{W}_2)$.

   (b) OLS of Y on $(1, X, W_1, \widetilde{\pi}_2'\widetilde{W}_2)$. Denote the coefficient on X by $\beta_{\text{long,index}}$. Let $(\gamma_{1,\text{index}}, \gamma_{2,\text{index}})$ denote the coefficients on $(W_1, \widetilde{\pi}_2'\widetilde{W}_2)$.

Then we have $\beta_{\text{long,vec}} = \beta_{\text{long,index}}$. To see this, by definition of $(\pi_1, \widetilde{\pi}_2)$ as linear projection coefficients, we have

$$X = \pi_1' W_1 + \widetilde{\pi}_2'\widetilde{W}_2 + X^{\perp W_1, \widetilde{W}_2}.$$

Consequently, the coefficients on $(W_1, \widetilde{\pi}_2'\widetilde{W}_2)$ from a linear projection of X on $(1, W_1, \widetilde{\pi}_2'\widetilde{W}_2)$ are $(\pi_1, 1)$. That is, we can write

$$X = \pi_1' W_1 + 1 \cdot (\widetilde{\pi}_2'\widetilde{W}_2) + X^{\perp W_1, \widetilde{\pi}_2'\widetilde{W}_2}.$$

This implies that $X^{\perp W_1, \widetilde{W}_2} = X^{\perp W_1, \widetilde{\pi}_2'\widetilde{W}_2}$. Hence, by the FWL theorem,

$$\beta_{\text{long,vec}} = \frac{\operatorname{cov}(Y, X^{\perp W_1, \widetilde{W}_2})}{\operatorname{var}(X^{\perp W_1, \widetilde{W}_2})} = \frac{\operatorname{cov}(Y, X^{\perp W_1, \widetilde{\pi}_2'\widetilde{W}_2})}{\operatorname{var}(X^{\perp W_1, \widetilde{\pi}_2'\widetilde{W}_2})} = \beta_{\text{long,index}}.$$

This shows that the definition of the parameter of interest, β_{long}, does not depend on whether W_2 is a vector or a scalar.

6. Finally, we discuss the interpretation of A5. Unlike the previous result, it is generally the case that

$$\sqrt{\operatorname{var}(\gamma_{2,\text{vec}}'\widetilde{W}_2)} \neq \sqrt{\operatorname{var}(\gamma_{2,\text{index}} W_2)} \quad \text{and} \quad \sqrt{\operatorname{var}(\gamma_{1,\text{vec}}' W_1)} \neq \sqrt{\operatorname{var}(\gamma_{1,\text{index}}' W_1)}.$$

Hence it is important to keep in mind that A5 is an assumption on the index version of the regression. That is, this assumption says

$$\sqrt{\operatorname{var}(\gamma_{2,\text{index}} W_2)} \leq \bar{r}_X \cdot \sqrt{\operatorname{var}(\gamma_{1,\text{index}}' W_1)}.$$

So it is a comparison of the impact of the index W_2 on outcomes relative to the vector of covariates W_1. Note that this point only applies when using the sensitivity parameter $\bar{r}_Y$. It is not relevant when applying our first two main results, Theorems 2 and 3.

Overall, our approach to allowing for a vector of unobservables does not change how researchers use our technical results. Instead, it only requires a slight adjustment to the interpretation of assumptions A5 and A6.

C Proofs for Section 2

In the following proofs, let $\mathbb{L}(A \mid B)$ denote the linear projection of the scalar random variable A on the random column vector B, defined as

$$\mathbb{L}(A \mid B) = \mathbb{E}(AB')\mathbb{E}(BB')^{-1}B.$$

Also note that

$$\mathbb{L}(A \mid 1, B) = \operatorname{cov}(A, B) \operatorname{var}(B)^{-1}B + k$$

where $k = \mathbb{E}(A) - \operatorname{cov}(A, B) \operatorname{var}(B)^{-1}\mathbb{E}(B)$ is a constant. Finally, recall that we define

$$A^{\perp B} = A - \operatorname{cov}(A, B) \operatorname{var}(B)^{-1}B$$

as the random component of this projection. This allows us to decompose $A = \rho' B + A^{\perp B}$ where $\rho' = \operatorname{cov}(A, B) \operatorname{var}(B)^{-1}$, and where $A^{\perp B}$ is uncorrelated with all components of B.

Proof of Theorem 1. ***Part 1***. We first show that $\beta_{\text{long}} = \beta_{\text{med}}$. By the FWL theorem,

$$\beta_{\text{med}} = \frac{\operatorname{cov}(Y, X^{\perp W_1})}{\operatorname{var}(X^{\perp W_1})}.$$

By the no selection on unobservables assumption A2 ($\pi_2 = 0$) and hence

$$X = \pi_1' W_1 + X^{\perp W}.$$

From this equation we have $\operatorname{cov}(X, W_1) = \pi_1' \operatorname{var}(W_1)$. Hence

$$\begin{aligned}
X^{\perp W_1} &= X - \operatorname{cov}(X, W_1) \operatorname{var}(W_1)^{-1}W_1 \\
&= X - \pi_1' W_1 \\
&= X^{\perp W}.
\end{aligned}$$

Next, write

$$\begin{aligned}
Y &= \beta_{\text{long}} X + \gamma_1' W_1 + \gamma_2 W_2 + Y^{\perp X,W} \\
&= \beta_{\text{long}}(\pi_1' W_1 + X^{\perp W}) + \gamma_1' W_1 + \gamma_2 W_2 + Y^{\perp X,W}.
\end{aligned}$$

So

$$\begin{aligned}
\operatorname{cov}(Y, X^{\perp W_1}) &= \operatorname{cov}(Y, X^{\perp W}) \\
&= \operatorname{cov}(\beta_{\text{long}}(\pi_1' W_1 + X^{\perp W}) + \gamma_1' W_1 + \gamma_2 W_2 + Y^{\perp X,W}, X^{\perp W}) \\
&= \beta_{\text{long}} \operatorname{var}(X^{\perp W}) + \operatorname{cov}(Y^{\perp X,W}, X^{\perp W}) \\
&= \beta_{\text{long}} \operatorname{var}(X^{\perp W_1}) + \operatorname{cov}(Y^{\perp X,W}, X - \pi_1' W_1) \\
&= \beta_{\text{long}} \operatorname{var}(X^{\perp W_1}).
\end{aligned}$$

Thus

$$\beta_{\text{long}} = \frac{\operatorname{cov}(Y, X^{\perp W_1})}{\operatorname{var}(X^{\perp W_1})}$$

as desired.

Part 2. Next we'll show that the identified set for γ_1 is $\mathbb{R}^{d_1}$. To show this, fix $g_1 \in \mathbb{R}^{d_1}$. We will construct a joint distribution of (W_2, W_1, X, Y) that is consistent with (a) the population distribution of observables (Y, X, W_1), (b) assumptions A1 and A2, and (c) equation (1) holding with $\gamma_1 = g_1$.

Part 2(i): Construction of (W_2, W_1, X, Y) **joint distribution**.

First we derive some useful expressions that we often use below. Under A1 and A2, we have

$$\pi_1 = \text{var}(W_1)^{-1} \text{cov}(W_1, X)$$

and

$$\begin{aligned}\beta_{\text{long}} &= \frac{\text{cov}(Y, X^{\perp W_1})}{\text{var}(X^{\perp W_1})} \\ &= \frac{\text{cov}(Y, X) - \text{cov}(Y, W_1)\pi_1}{\text{var}(X) - \text{cov}(X, W_1)\pi_1}.\end{aligned}$$

where the last line follows by substituting $X^{\perp W_1} = X - \pi_1' W_1$.

Next, since we want to ensure that equation (1) holds with $\gamma_1 = g_1$ for our constructed joint distribution of (W_2, W_1, X, Y), we start by finding values $g_2 \in \mathbb{R}$ and $c \in \mathbb{R}^{d_1}$ such that

$$\text{cov}(W_1, Y) = \beta_{\text{long}} \text{cov}(W_1, X) + \text{var}(W_1)g_1 + g_2 c. \tag{14}$$

This equation is the covariance of W_1 with equation (1) where the coefficients are $(\beta_{\text{long}}, g_1, g_2)$ and where $c = \text{cov}(W_1, W_2)$. For any fixed g_1, equation (14) has multiple solutions (g_2, c). Note that $g_2 \neq 0$ by assumption A1. We will select specific solutions (g_2, c) below.

For any choice of (g_2, c), define the $(1 + d_1 + 2) \times (1 + d_1 + 2)$ matrix

$$\Sigma(g_2, c) = \begin{pmatrix} 1 & c' & \pi_1' c & \beta_{\text{long}}\pi_1' c + g_1' c + g_2 \\ c & \text{var}(W_1) & \text{cov}(W_1, X) & \text{cov}(W_1, Y) \\ \pi_1' c & \text{cov}(X, W_1) & \text{var}(X) & \text{cov}(X, Y) \\ \beta_{\text{long}}\pi_1' c + g_1' c + g_2 & \text{cov}(Y, W_1) & \text{cov}(Y, X) & \text{var}(Y) \end{pmatrix}.$$

Note that the bottom right $(d_1+2) \times (d_1+2)$ block corresponds to $\text{var}(W_1, X, Y) \in \mathbb{R}^{(d_1+2)\times(d_1+2)}$, a known variance matrix. The top row of $\Sigma(g_2, c)$ corresponds to W_2. Overall, we will think of this matrix as a covariance matrix for (W_2, W_1, X, Y). Specifically, we will show that for a fixed $g_1 \in \mathbb{R}^{d_1}$, and for a pair of values (g_2^*, c^*) that satisfy equation (14), the matrix $\Sigma(g_2^*, c^*)$ is positive definite. Then we show how to use this matrix to construct W_2 in a way that is consistent with the above equations and assumptions.

To check whether a matrix is positive definite, we use Sylvester's criterion. It states that a matrix is positive definite if and only if all its leading principal minors have positive determinants. First, we show that the $(1 + d_1) \times (1 + d_1)$ matrix

$$\Sigma_2(c) = \begin{pmatrix} 1 & c' \\ c & \text{var}(W_1) \end{pmatrix}$$

is positive definite. By symmetry, $\Sigma_2(c)$ is positive definite if and only if

$$\begin{pmatrix} \text{var}(W_1) & c \\ c' & 1 \end{pmatrix}$$

is positive definite. By A1, $\text{var}(W_1)$ is positive definite. By another application of Sylvester's criterion we then know that

$$\begin{pmatrix} \text{var}(W_1) & c \\ c' & 1 \end{pmatrix}$$

is positive definite if and only if its determinant is positive. By the determinant formula for partitioned matrices, this determinant equals

$$\det \begin{pmatrix} \text{var}(W_1) & c \\ c' & 1 \end{pmatrix} = \det(\text{var}(W_1)) \cdot \left(1 - c' \text{var}(W_1)^{-1} c\right).$$

Notice that $1 = \text{var}(W_2)$ and $c' \text{var}(W_1)^{-1}$ will be the coefficient $\text{cov}(W_2, W_1) \text{var}(W_1)^{-1}$ in $\mathbb{L}(W_2 \mid 1, W_1)$ for the (W_2, W_1) distribution we are constructing. $\det(\text{var}(W_1)) > 0$ by A1. Next consider the term in parentheses. By equation (14), it is positive if

$$\begin{aligned} c' \text{var}(W_1)^{-1} c = \frac{1}{g_2^2} &(\text{cov}(Y, W_1) - \beta_{\text{long}} \text{cov}(X, W_1) - g_1' \text{var}(W_1)) \\ &\cdot \text{var}(W_1)^{-1} (\text{cov}(W_1, Y) - \beta_{\text{long}} \text{cov}(W_1, X) - \text{var}(W_1) g_1) \\ < 1 & \end{aligned}$$

or, equivalently, if

$$\begin{aligned} g_2^2 > &(\text{cov}(Y, W_1) - \beta_{\text{long}} \text{cov}(X, W_1) - g_1' \text{var}(W_1)) \\ &\text{var}(W_1)^{-1} (\text{cov}(W_1, Y) - \beta_{\text{long}} \text{cov}(W_1, X) - \text{var}(W_1) g_1) \\ \equiv &L(g_1). \end{aligned}$$

Thus the first $(1 + d_1)$ leading principal minors are positive definite when $g_2^2 > L(g_1)$.

Second, define the $(2 + d_1) \times (2 + d_1)$ matrix

$$\Sigma_3(c) = \begin{pmatrix} 1 & c' & \pi_1' c \\ c & \text{var}(W_1) & \text{cov}(W_1, X) \\ \pi_1' c & \text{cov}(X, W_1) & \text{var}(X) \end{pmatrix}.$$

This is the next leading principal minor of $\Sigma(g_2, c)$. We will show its determinant is also positive when $g_2^2 > L(g_1)$. By the partitioned matrix determinant formula,

$$\det(\Sigma_3(c)) = \det(\Sigma_2(c)) \cdot \left(\text{var}(X) - \begin{pmatrix} \pi_1' c & \text{cov}(X, W_1) \end{pmatrix} \Sigma_2(c)^{-1} \begin{pmatrix} \pi_1' c \\ \text{cov}(X, W_1) \end{pmatrix} \right).$$

Note that

$$\begin{pmatrix} \pi_1' c & \text{cov}(X, W_1) \end{pmatrix} \Sigma_2(c)^{-1}$$

will be the coefficient $\begin{pmatrix} \text{cov}(X, W_2) & \text{cov}(X, W_1) \end{pmatrix} \text{var}(W_2, W_1)^{-1}$ in $\mathbb{L}(X \mid 1, W_1, W_2)$ for the distribution of (X, W_1, W_2) we are constructing. We have already shown that $\det(\Sigma_2(c)) > 0$ when $g_2^2 > L(g_1)$. So consider the term in parentheses:

$$\begin{aligned} &\text{var}(X) - \begin{pmatrix} \pi_1' c & \text{cov}(X, W_1) \end{pmatrix} \Sigma_2(c)^{-1} \begin{pmatrix} \pi_1' c \\ \text{cov}(X, W_1) \end{pmatrix} \\ &= \text{var}(X) - \frac{1}{1 - c' \text{var}(W_1)^{-1} c} \begin{pmatrix} \pi_1' c & \text{cov}(X, W_1) \end{pmatrix} \\ &\quad \cdot \begin{pmatrix} 1 & -c' \text{var}(W_1)^{-1} \\ -\text{var}(W_1)^{-1} c & \text{var}(W_1)^{-1}(1 - c' \text{var}(W_1)^{-1} c) + \text{var}(W_1)^{-1} c c' \text{var}(W_1)^{-1} \end{pmatrix} \\ &\quad \cdot \begin{pmatrix} \pi_1' c \\ \text{cov}(W_1, X) \end{pmatrix} \\ &= \text{var}(X) - \frac{1}{1 - c' \text{var}(W_1)^{-1} c} \begin{pmatrix} \pi_1' c & \text{cov}(X, W_1) \end{pmatrix} \\ &\quad \cdot \begin{pmatrix} \pi_1' c - \pi_1' c \\ -(\pi_1' c) \text{var}(W_1)^{-1} c + (1 - c' \text{var}(W_1)^{-1} c) \pi_1 + \text{var}(W_1)^{-1} c c' \pi_1 \end{pmatrix} \end{aligned}$$

$$
\begin{aligned}
&= \text{var}(X) - \frac{1}{1 - c' \text{var}(W_1)^{-1} c} \begin{pmatrix} \pi_1' c & \text{cov}(X, W_1) \end{pmatrix} \begin{pmatrix} 0 \\ (1 - c' \text{var}(W_1)^{-1} c) \pi_1 \end{pmatrix} \\
&= \text{var}(X) - \text{cov}(X, W_1) \text{var}(W_1)^{-1} \text{cov}(W_1, X) \\
&= \text{var}(X^{\perp W_1}) \\
&> 0.
\end{aligned}
$$

The first line follows by the partitioned inverse formula and by $g_2^2 > L(g_1)$ (which we showed is equivalent to $c' \text{var}(W_1)^{-1} c < 1$). The second and fourth lines follow by the definition of π_1. The last line follows by A1.

Finally, we compute the determinant of $\Sigma(g_2, c)$ itself. By the partitioned matrix determinant formula, it equals

$$
\begin{aligned}
&\det(\Sigma(g_2, c)) \\
&= \det(\Sigma_3(c)) \\
&\cdot \left(\text{var}(Y) - \begin{pmatrix} \beta_{\text{long}} \pi_1' c + g_1' c + g_2 & \text{cov}(Y, W_1) & \text{cov}(Y, X) \end{pmatrix} \Sigma_3(c)^{-1} \begin{pmatrix} \beta_{\text{long}} \pi_1' c + g_1' c + g_2 \\ \text{cov}(W_1, Y) \\ \text{cov}(Y, X) \end{pmatrix} \right).
\end{aligned}
$$

Note that

$$
\begin{pmatrix} \beta_{\text{long}} \pi_1' c + g_1' c + g_2 & \text{cov}(Y, W_1) & \text{cov}(Y, X) \end{pmatrix} \Sigma_3(c)^{-1}
$$

will be the coefficient

$$
\begin{pmatrix} \text{cov}(Y, W_2) & \text{cov}(Y, W_1) & \text{cov}(Y, X) \end{pmatrix} \Sigma_3(c)^{-1}
$$

in $\mathbb{L}(Y \mid 1, W_2, W_1, X)$ for the distribution of (Y, W_2, W_1, X) we are constructing. We have already shown that $\det(\Sigma_3(c)) > 0$ when $g_2^2 > L(g_1)$. We finish by showing that the second term in the expression for $\det(\Sigma(g_2, c))$ is positive so long as (g_1, g_2) satisfies one additional constraint. Specifically, we compute

$$
\begin{aligned}
&\text{var}(Y) - \begin{pmatrix} \beta_{\text{long}} \pi_1' c + g_1' c + g_2 & \text{cov}(Y, X) & \text{cov}(Y, W_1) \end{pmatrix} \Sigma_3(c)^{-1} \begin{pmatrix} \beta_{\text{long}} \pi_1' c + g_1' c + g_2 \\ \text{cov}(Y, X) \\ \text{cov}(W_1, Y) \end{pmatrix} \\
&= \text{var}(Y) - \beta_{\text{long}}^2 \text{var}(X) + g_1' \text{var}(W_1) g_1 + 2\beta_{\text{long}} g_1' \text{cov}(W_1, X) \\
&\quad - 2\beta_{\text{long}} \pi_1' \text{cov}(W_1, Y) - 2 g_1' \text{cov}(W_1, Y) + 2\beta_{\text{long}}^2 \pi_1' \text{cov}(W_1, X) - g_2^2 \qquad (15) \\
&\equiv U(g_1) - g_2^2.
\end{aligned}
$$

For details on how to obtain equation (15), see Appendix G. Thus this determinant is positive if $g_2^2 < U(g_1)$.

Putting all of these results together, by Sylvester's criterion, we have shown that $\Sigma(g_2, c)$ is positive definite whenever g_2 satisfies $L(g_1) < g_2^2 < U(g_1)$. To show there exists such a $g_2 \in \mathbb{R}$, we show that $0 \leq L(g_1) < U(g_1)$ for all $g_1 \in \mathbb{R}$. To see this, first note that $L(g_1)$ is a quadratic and $\text{var}(W_1)$ is positive definite, so that $L(g_1) \geq 0$ for all $g_1 \in \mathbb{R}$. To see that $L(g_1) < U(g_1)$ for all $g_1 \in \mathbb{R}$, note that

$$
\begin{aligned}
&U(g_1) - L(g_1) \\
&= \text{var}(Y) - \beta_{\text{long}}^2 \text{var}(X) + g_1' \text{var}(W_1) g_1 + 2\beta_{\text{long}} g_1' \text{cov}(W_1, X) \\
&\quad - 2\beta_{\text{long}} \pi_1' \text{cov}(W_1, Y) - 2 g_1' \text{cov}(W_1, Y) + 2\beta_{\text{long}}^2 \pi_1' \text{cov}(W_1, X) \\
&\quad - (\text{cov}(Y, W_1) - \beta_{\text{long}} \text{cov}(X, W_1) - g_1' \text{var}(W_1)) \text{var}(W_1)^{-1} (\text{cov}(W_1, Y) - \beta_{\text{long}} \text{cov}(W_1, X) - \text{var}(W_1) g_1) \\
&= \text{var}(Y) - \beta_{\text{long}}^2 \text{var}(X) + g_1' \text{var}(W_1) g_1 + 2\beta_{\text{long}} g_1' \text{cov}(W_1, X) \\
&\quad - 2\beta_{\text{long}} \pi_1' \text{cov}(W_1, Y) - 2 g_1' \text{cov}(W_1, Y) + 2\beta_{\text{long}}^2 \pi_1' \text{cov}(W_1, X) \\
&\quad - \text{cov}(Y, W_1) \text{var}(W_1)^{-1} \text{cov}(W_1, Y) - \beta_{\text{long}}^2 \pi_1' \text{cov}(W_1, X) - g_1' \text{var}(W_1) g_1 \\
&\quad + 2\beta_{\text{long}} \text{cov}(Y, W_1) \pi_1 + 2 \text{cov}(Y, W_1) g_1 - 2\beta_{\text{long}} \text{cov}(X, W_1) g_1
\end{aligned}
$$

$$
\begin{aligned}
&= \operatorname{var}(Y) - \beta_{\text{long}}^2 \operatorname{var}(X) + \beta_{\text{long}}^2 \pi_1' \operatorname{cov}(W_1, X) - \operatorname{cov}(Y, W_1) \operatorname{var}(W_1)^{-1} \operatorname{cov}(W_1, Y) \\
&= \operatorname{var}(Y) + \left(\frac{\operatorname{cov}(Y, X) - \operatorname{cov}(Y, W_1)\pi_1}{\operatorname{var}(X) - \operatorname{cov}(X, W_1)\pi_1} \right)^2 (\pi_1' \operatorname{cov}(W_1, X) - \operatorname{var}(X)) - \operatorname{cov}(Y, W_1) \operatorname{var}(W_1)^{-1} \operatorname{cov}(W_1, Y) \\
&= \operatorname{var}(Y) - \frac{\operatorname{cov}(Y, X^{\perp W_1})^2}{\operatorname{var}(X^{\perp W_1})} - \operatorname{cov}(Y, W_1) \operatorname{var}(W_1)^{-1} \operatorname{cov}(W_1, Y) \\
&= \frac{1}{\operatorname{var}(X^{\perp W_1})} \left(\operatorname{var}(Y) \operatorname{var}(X^{\perp W_1}) - \operatorname{cov}(Y, X^{\perp W_1})^2 - \operatorname{cov}(Y, W_1) \operatorname{var}(W_1)^{-1} \operatorname{cov}(W_1, Y) \operatorname{var}(X^{\perp W_1}) \right) \\
&= \frac{1}{\operatorname{var}(X^{\perp W_1})} \left(\operatorname{var}(Y^{\perp W_1}) \operatorname{var}(X^{\perp W_1}) - \operatorname{cov}(Y, X^{\perp W_1})^2 \right) \\
&= \frac{1}{\operatorname{var}(X^{\perp W_1})} \left(\operatorname{var}(Y^{\perp W_1}) \operatorname{var}(X^{\perp W_1}) - \operatorname{cov}(Y^{\perp W_1}, X^{\perp W_1})^2 \right) \\
&> 0.
\end{aligned}
$$

The second line follows by distributing terms and using our equation for π_1. The last inequality follows from the Cauchy-Schwarz inequality and A1, which ensures that the inequality is strict. Also note that $\operatorname{var}(X^{\perp W_1}) > 0$ by A1.

Therefore, let g_2^* satisfy $g_2^{*2} \in (L(g_1), U(g_1))$ and

$$
c^* = \frac{\operatorname{cov}(W_1, Y) - \beta_{\text{long}} \operatorname{cov}(W_1, X) - \operatorname{var}(W_1) g_1}{g_2^*}.
$$

Then $\Sigma(g_2^*, c^*)$ is positive definite and hence a valid covariance matrix.

We next use this matrix to construct the unobservable W_2. Here and below we use the following notation: For any symmetric positive definite matrix A, let $A = \text{Chol}(A)\text{Chol}(A)'$ denote its unique Cholesky decomposition, where $\text{Chol}(A)$ is a lower triangular matrix. Returning to our problem, let

$$
\begin{pmatrix} \widetilde{W}_1 \\ \widetilde{X} \\ \widetilde{Y} \end{pmatrix} = \text{Chol}(\operatorname{var}(W_1, X, Y)^{-1})' \begin{pmatrix} W_1 \\ X \\ Y \end{pmatrix}
$$

denote the whitened vector of observables (W_1, X, Y). Note that $\operatorname{var}(\widetilde{W}_1, \widetilde{X}, \widetilde{Y}) = I$. Next, let $\widetilde{W}_2$ be a unit variance random variable that is uncorrelated with $(\widetilde{W}_1, \widetilde{X}, \widetilde{Y})$. Then we define

$$
\begin{pmatrix} W_2 \\ W_1 \\ X \\ Y \end{pmatrix} = \text{Chol}(\Sigma(g_2^*, c^*))' \begin{pmatrix} \widetilde{W}_2 \\ \widetilde{W}_1 \\ \widetilde{X} \\ \widetilde{Y} \end{pmatrix},
$$

which undoes the whitening. That is, $\operatorname{var}(W_2, W_1, X, Y) = \Sigma(g_2^*, c^*)$. Thus we have constructed a joint distribution of (W_2, W_1, X, Y). Finally, we can use this distribution to define the residuals

$$
Y^{\perp X, W} = Y - \beta_{\text{long}} X + g_1' W_1 + g_2^* W_2 \qquad \text{and} \qquad X^{\perp W} = X - \pi_1' W_1.
$$

Part 2(ii): This distribution is consistent with the distribution of (Y, X, W_1).

By definition, the marginal distribution of (W_1, X, Y) from the joint distribution of (W_2, W_1, X, Y) that we have constructed equals the marginal distribution of the observed (W_1, X, Y). This follows since our construction leaves (W_1, X, Y) unchanged due to the lower triangular structure of $\text{Chol}(\Sigma(g_2^*, c^*))$ and $\text{Chol}(\operatorname{var}(W_1, X, Y)^{-1})$.

Part 2(iii): This distribution is consistent with A1 and A2 and with equation (1) **where** $\gamma_1 = g_1$.

Checking A1. By definition, the covariance matrix of (W_2, W_1, X, Y) is $\Sigma(g_2^*, c^*)$, which is positive definite. Therefore A1 holds.

Checking that equation (1) *holds with* $\gamma_1 = g_1$. By construction, equation (14) holds at (g_1, g_2^*, c^*):

$$\operatorname{cov}(W_1, Y) = \beta_{\text{long}} \operatorname{cov}(W_1, X) + \operatorname{var}(W_1) g_1 + g_2^* c^*.$$

Using our expression for β_{long}, we have that that

$$\begin{aligned}
\operatorname{cov}(Y, X) &= \beta_{\text{long}} \operatorname{var}(X^{\perp W_1}) + \operatorname{cov}(Y, W_1)\pi_1 \\
&= \beta_{\text{long}} \operatorname{var}(X^{\perp W_1}) + \pi_1'(\beta_{\text{long}} \operatorname{cov}(W_1, X) + \operatorname{var}(W_1) g_1 + g_2^* c^*) \\
&= \beta_{\text{long}} \operatorname{var}(X) + g_1' \operatorname{var}(W_1)\pi_1 + g_2^* \pi_1' c^*.
\end{aligned}$$

Directly from $\Sigma(g_2^*, c^*)$, we have that $\operatorname{cov}(Y, W_2) = \beta_{\text{long}} \pi_1' c^* + g_1' c^* + g_2^*$. We use these to calculate

$$\begin{aligned}
&\mathbb{L}(Y \mid 1, X, W) \\
&= \begin{pmatrix} \operatorname{cov}(Y, W_2) & \operatorname{cov}(Y, X) & \operatorname{cov}(Y, W_1) \end{pmatrix} \begin{pmatrix} 1 & \pi_1' c^* & c^{*\prime} \\ \pi_1' c^* & \operatorname{var}(X) & \operatorname{cov}(X, W_1) \\ c^* & \operatorname{cov}(W_1, X) & \operatorname{var}(W_1) \end{pmatrix}^{-1} \begin{pmatrix} W_2 \\ X \\ W_1 \end{pmatrix} + \text{const.} \\
&= \begin{pmatrix} \beta_{\text{long}} \pi_1' c^* + g_1' c^* + g_2^* & \beta_{\text{long}} \operatorname{var}(X) + g_1' \operatorname{var}(W_1)\pi_1 + g_2^* \pi_1' c^* & \beta_{\text{long}} \pi_1' \operatorname{var}(W_1) + g_1' \operatorname{var}(W_1) + g_2^* c^{*\prime} \end{pmatrix} \\
&\quad \cdot \begin{pmatrix} 1 & \pi_1' c^* & c^{*\prime} \\ \pi_1' c^* & \operatorname{var}(X) & \operatorname{cov}(X, W_1) \\ c^* & \operatorname{cov}(W_1, X) & \operatorname{var}(W_1) \end{pmatrix}^{-1} \begin{pmatrix} W_2 \\ X \\ W_1 \end{pmatrix} + \text{const.} \\
&= \beta_{\text{long}} X + g_1' W_1 + g_2^* W_2 + \text{const.} \qquad (16)
\end{aligned}$$

We give a full derivation of equation (16) in Appendix G.

Checking A2. From the definition of $\Sigma(g_2^*, c^*)$, we have that $\operatorname{cov}(X, W_2) = \pi_1' c^*$. Using this, we we next show that $\pi_2 = 0$. We have

$$\begin{aligned}
&\mathbb{L}(X \mid 1, W) \\
&= \begin{pmatrix} \operatorname{cov}(X, W_2) & \operatorname{cov}(X, W_1) \end{pmatrix} \begin{pmatrix} 1 & c^{*\prime} \\ c^* & \operatorname{var}(W_1) \end{pmatrix}^{-1} \begin{pmatrix} W_2 \\ W_1 \end{pmatrix} + \text{const.} \\
&= \begin{pmatrix} \pi_1' c^* & \pi_1' \operatorname{var}(W_1) \end{pmatrix} \frac{1}{1 - c^{*\prime} \operatorname{var}(W_1)^{-1} c^*} \\
&\quad \cdot \begin{pmatrix} 1 & -c^{*\prime} \operatorname{var}(W_1)^{-1} \\ -\operatorname{var}(W_1)^{-1} c^* & \operatorname{var}(W_1)^{-1}(1 - c^{*\prime} \operatorname{var}(W_1)^{-1} c^*) + \operatorname{var}(W_1)^{-1} c^* c^{*\prime} \operatorname{var}(W_1)^{-1} \end{pmatrix} \begin{pmatrix} W_2 \\ W_1 \end{pmatrix} + \text{const.} \\
&= 0 \cdot W_2 + \pi_1' W_1 + \text{const.}
\end{aligned}$$

Thus we see that equation (2) holds with $\pi_2 = 0$. This concludes the proof. □

D The Identified Set For β_{long} With Fixed $(\bar{r}_X, \bar{r}_Y, \bar{c})$

In this appendix we characterize the identified set for β_{long}, the coefficient on X in the long regression of Y on $(1, X, W_1, W_2)$, under assumptions A3–A6 hold. That is, we use information from all three sensitivity parameters to learn about β_{long} (Theorem 5 below). This is an important preliminary step in deriving our main results in section 3.2.

D.1 Statement of the Result

Before stating the result, we define some notation. Let

$$\mathbf{A}(r_X, r_Y, c, \pi_1) = \begin{pmatrix} \operatorname{var}(X) & \operatorname{cov}(X, W_1) \\ (\operatorname{var}(W_1) + c r_X' + r_Y c' + r_Y r_X')\pi_1 & \operatorname{var}(W_1) + r_Y c' \end{pmatrix}$$

and

$$\Pi_1(r_X, c) = \{p_1 \in \mathbb{R}^{d_1} : \operatorname{cov}(X, W_1) = p_1' \left(\operatorname{var}(W_1) + r_X c'\right)\}.$$

Define

$$\mathcal{B}(r_X, r_Y, c) = \{b \in \mathbb{R} : \text{Equations (18)–(21) hold for some } p_1 \in \Pi_1(r_X, c) \text{ and some } g_1 \in \mathbb{R}^{d_1}\} \tag{17}$$

where these equations are

$$\operatorname{cov}\big(Y, (X, W_1)\big) = \begin{pmatrix} b & g_1' \end{pmatrix} \mathbf{A}(r_X, r_Y, c, p_1) \tag{18}$$

$$\begin{aligned} \operatorname{var}(Y) > b^2 \operatorname{var}(X) &+ g_1' \left(\operatorname{var}(W_1) + r_Y r_Y' + 2 r_Y c'\right) g_1 \\ &+ 2b p_1' (\operatorname{var}(W_1) + r_X c' + c r_Y' + r_X r_Y') g_1 \end{aligned} \tag{19}$$

$$\operatorname{var}(X) > p_1' (\operatorname{var}(W_1) + 2 r_X c' + r_X r_X') p_1 \tag{20}$$

$$1 > c' \operatorname{var}(W_1)^{-1} c. \tag{21}$$

We can now state the result.

Theorem 5. Suppose the joint distribution of (Y, X, W_1) is known. Suppose A1 holds. Suppose A3–A6 hold. Then the identified set for β_{long} is

$$\mathcal{B}_I(\bar{r}_X, \bar{r}_Y, \bar{c}) = \bigcup_{(r_X, r_Y, c) : \|r_X\|_{\Sigma_{\text{obs}}^{-1}} \leq \bar{r}_X, \|r_Y\|_{\Sigma_{\text{obs}}^{-1}} \leq \bar{r}_Y, \|c\|_{\Sigma_{\text{obs}}^{-1}} \leq \bar{c}} \mathcal{B}(r_X, r_Y, c). \tag{22}$$

Remark 1. The following alternative characterization of the set $\mathcal{B}(r_X, r_Y, c)$ will sometimes be useful:

$$\mathcal{B}(r_X, r_Y, c) = \{b \in \mathbb{R} : \text{ the below six equations hold for some } (p_1, g_1) \in \mathbb{R}^{2d_1}\}$$

where

$$\begin{aligned} \operatorname{cov}(Y, X) &= b \operatorname{var}(X) + g_1' (\operatorname{var}(W_1) + c r_X' + r_Y c' + r_Y r_X') p_1 \\ \operatorname{cov}(Y, W_1) &= b \operatorname{cov}(X, W_1) + g_1' (\operatorname{var}(W_1) + r_Y c') \\ \operatorname{cov}(X, W_1) &= p_1' (\operatorname{var}(W_1) + r_X c') \\ \operatorname{var}(Y) &> b^2 \operatorname{var}(X) + g_1' \left(\operatorname{var}(W_1) + r_Y r_Y' + 2 r_Y c'\right) g_1 + 2b(\operatorname{cov}(Y, X) - b \operatorname{var}(X)) \\ \operatorname{var}(X) &> p_1' (\operatorname{var}(W_1) + 2 r_X c' + r_X r_X') p_1 \\ 1 &> c' \operatorname{var}(W_1)^{-1} c. \end{aligned}$$

D.2 Proofs

We use the following lemma to prove Theorem 5.

Lemma 1.

1. A3 is equivalent to the following statement: There exists an $r_X \in \mathbb{R}^{d_1}$ such that $\pi_2 = r_X' \pi_1$ and $\|r_X\|_{\Sigma_{\text{obs}}^{-1}} \leq \bar{r}_X$.

2. A5 is equivalent to the following statement: There exists an $r_Y \in \mathbb{R}^{d_1}$ such that $\gamma_2 = r_Y'\gamma_1$ and $\|r_Y\|_{\Sigma_{\text{obs}}^{-1}} \leq \bar{r}_Y$.

Proof of Lemma 1. We start with part 1. First, suppose A3 holds. If $\pi_1 = 0$, then this trivially holds by setting $r_X = 0$. Suppose $\pi_1 \neq 0$. Then let

$$r_X = \frac{\pi_2}{\|\pi_1\|^2_{\Sigma_{\text{obs}}}} \Sigma_{\text{obs}} \pi_1.$$

We can see that

$$\begin{aligned} r_X' \pi_1 &= \frac{\pi_2}{\|\pi_1\|^2_{\Sigma_{\text{obs}}}} \pi_1' \Sigma_{\text{obs}} \pi_1 \\ &= \frac{\pi_2}{\|\pi_1\|^2_{\Sigma_{\text{obs}}}} \cdot \|\pi_1\|^2_{\Sigma_{\text{obs}}} \\ &= \pi_2. \end{aligned}$$

Also,

$$\begin{aligned} \|r_X\|_{\Sigma_{\text{obs}}^{-1}} &= \left| \frac{\pi_2}{\|\pi_1\|^2_{\Sigma_{\text{obs}}}} \right| \cdot \|\Sigma_{\text{obs}}\pi_1\|_{\Sigma_{\text{obs}}^{-1}} \\ &= \left| \frac{\pi_2}{\|\pi_1\|^2_{\Sigma_{\text{obs}}}} \right| \sqrt{\pi_1' \Sigma_{\text{obs}} \Sigma_{\text{obs}}^{-1} \Sigma_{\text{obs}} \pi_1} \\ &= \frac{|\pi_2|}{\|\pi_1\|_{\Sigma_{\text{obs}}}} \\ &\leq \bar{r}_X. \end{aligned}$$

The first line follows by the definition of r_X. The last line follows by A3. Hence A3 implies the first statement in the lemma.

Next assume the first statement in the lemma holds. Recall that for any symmetric positive definite matrix A, we let $A = \text{Chol}(A)\text{Chol}(A)'$ denote its unique Cholesky decomposition, where $\text{Chol}(A)$ is a lower triangular matrix. Then

$$\begin{aligned} |\pi_2| &= |r_X' \pi_1| \\ &= |r_X' \text{Chol}(\Sigma_{\text{obs}})^{-1} \text{Chol}(\Sigma_{\text{obs}}) \pi_1| \\ &\leq \|r_X' \text{Chol}(\Sigma_{\text{obs}})^{-1}\| \cdot \|\text{Chol}(\Sigma_{\text{obs}})\pi_1\| \\ &= \|r_X\|_{\Sigma_{\text{obs}}^{-1}} \|\pi_1\|_{\Sigma_{\text{obs}}} \\ &\leq \bar{r}_X \|\pi_1\|_{\Sigma_{\text{obs}}}. \end{aligned}$$

The first inequality follows from the Cauchy Schwarz ienquality. The second inequality is assumed in the first statement of the lemma. Therefore A3 holds.

The proof of part 2 is analogous to the proof of part 1. □

Proof of Theorem 5. We prove this result in two parts. First we show that the true value of β_{long} is in $\mathcal{B}_I(\bar{r}_X, \bar{r}_Y, \bar{c})$. Second we show sharpness: Any b in $\mathcal{B}_I(\bar{r}_X, \bar{r}_Y, \bar{c})$ is consistent with the distribution of the observables and the model assumptions.

Part 1 (Outer set). By A3 and A5, Lemma 1 implies that there exists $r_X, r_Y \in \mathbb{R}^{d_1}$ such that $\pi_2 = r_X'\pi_1$ and $\gamma_2 = r_Y'\gamma_1$ with $\|r_X\|_{\Sigma_{\text{obs}}^{-1}} \leq \bar{r}_X$ and $\|r_Y\|_{\Sigma_{\text{obs}}^{-1}} \leq \bar{r}_Y$. Let $c = \text{cov}(W_1, W_2)$ denote the true covariance between the observed covariates and the unobserved covariate W_2. Note that $\|c\|_{\Sigma_{\text{obs}}^{-1}} \leq \bar{c}$ holds by A6. By the definition of $\mathcal{B}_I(\bar{r}_X, \bar{r}_Y, \bar{c})$, it therefore suffices to show that $\beta_{\text{long}} \in \mathcal{B}(r_X, r_Y, c)$ for these values of (r_X, r_Y, c).

We have

$$\underset{(1\times d_1)}{\operatorname{cov}(X, W_1)} = \operatorname{cov}(\pi_1' W_1 + \pi_2 W_2 + X^{\perp W}, W_1)$$
$$= \pi_1' \operatorname{var}(W_1) + \pi_2 \operatorname{cov}(W_1, W_2)'$$
$$= \pi_1' \operatorname{var}(W_1) + (\pi_1' r_X) c'$$
$$= \pi_1'(\operatorname{var}(W_1) + r_X c').$$

Thus $\pi_1 \in \Pi_1(r_X, c)$. Similarly,

$$\begin{aligned}\operatorname{cov}(X, W_2) &= \operatorname{cov}(\pi_1' W_1 + \pi_2 W_2 + X^{\perp W}, W_2)\\ &= \pi_1' \operatorname{cov}(W_1, W_2) + \pi_2 \operatorname{var}(W_2)\\ &= \pi_1' c + \pi_1' r_X\\ &= \pi_1'(c + r_X).\end{aligned}$$

We used $\operatorname{var}(W_2) = 1$ in the third line. Next consider

$$\begin{aligned}\operatorname{cov}(Y, X) &= \operatorname{cov}(\beta_{\text{long}} X + \gamma_1' W_1 + \gamma_2 W_2 + Y^{\perp X,W}, X)\\ &= \beta_{\text{long}} \operatorname{var}(X) + \gamma_1' \operatorname{cov}(W_1, X) + \gamma_2 \operatorname{cov}(W_2, X)\\ &= \beta_{\text{long}} \operatorname{var}(X) + \gamma_1'(\operatorname{var}(W_1) + c r_X')\pi_1 + \gamma_1' r_Y (c + r_X)'\pi_1\\ &= \beta_{\text{long}} \operatorname{var}(X) + \gamma_1'(\operatorname{var}(W_1) + c r_X' + r_Y c' + r_Y r_X')\pi_1\end{aligned}$$

and

$$\underset{(1\times d_1)}{\operatorname{cov}(Y, W_1)} = \operatorname{cov}(\beta_{\text{long}} X + \gamma_1' W_1 + \gamma_2 W_2 + Y^{\perp X,W}, W_1)$$
$$= \beta_{\text{long}} \operatorname{cov}(X, W_1) + \gamma_1' \operatorname{var}(W_1) + \gamma_2 \operatorname{cov}(W_2, W_1)$$
$$= \beta_{\text{long}} \operatorname{cov}(X, W_1) + \gamma_1' \operatorname{var}(W_1) + (\gamma_1' r_Y) c'$$
$$= \beta_{\text{long}} \operatorname{cov}(X, W_1) + \gamma_1'(\operatorname{var}(W_1) + r_Y c').$$

These two results imply that

$$\operatorname{cov}(Y, (X, W_1)) = \begin{pmatrix}\beta_{\text{long}} & \gamma_1'\end{pmatrix} \mathbf{A}(r_X, r_Y, c, \pi_1).$$

So equation (18) holds at the true β_{long} with $(p_1, g_1) = (\pi_1, \gamma_1)$. Next we verify that the inequalities in equations (19), (20), and (21) also hold at the true β_{long} with $(p_1, g_1) = (\pi_1, \gamma_1)$.

1. Consider inequality (19). By A1, $\operatorname{var}(Y^{\perp X,W}) > 0$. By equation (1), $\mathbb{L}(Y \mid 1, X, W_1, W_2) = X\beta_{\text{long}} + \gamma_1' W_1 + \gamma_2 W_2 + \text{const.}$, since $Y^{\perp X,W}$ is uncorrelated with (X, W). Therefore,

$$\begin{aligned}\operatorname{var}(Y) &= \operatorname{var}(\mathbb{L}(Y \mid 1, X, W_1, W_2)) + \operatorname{var}(Y^{\perp X,W})\\ &> \operatorname{var}(\mathbb{L}(Y \mid 1, X, W_1, W_2))\\ &= \operatorname{var}(X\beta_{\text{long}} + \gamma_1' W_1 + \gamma_2 W_2)\\ &= \beta_{\text{long}}^2 \operatorname{var}(X) + \gamma_1' \operatorname{var}(W_1)\gamma_1 + \gamma_2^2\\ &\qquad + 2\beta_{\text{long}}\gamma_1' \operatorname{cov}(W_1, X) + 2\beta_{\text{long}}\gamma_2 \operatorname{cov}(X, W_2) + 2\gamma_1' \operatorname{cov}(W_1, W_2)\gamma_2\\ &= \beta_{\text{long}}^2 \operatorname{var}(X) + \gamma_1' \operatorname{var}(W_1)\gamma_1 + (r_Y'\gamma_1)^2\\ &\qquad + 2\beta_{\text{long}}\gamma_1'(\operatorname{var}(W_1) + c r_X')\pi_1 + 2\beta_{\text{long}}(r_Y'\gamma_1)\pi_1'(c + r_X) + 2\gamma_1' c r_Y' \gamma_1\\ &= \beta_{\text{long}}^2 \operatorname{var}(X) + \gamma_1' \left(\operatorname{var}(W_1) + r_Y r_Y' + 2 c r_Y'\right)\gamma_1 + 2\beta_{\text{long}}\pi_1'(\operatorname{var}(W_1) + r_X c' + (c + r_X) r_Y')\gamma_1\end{aligned}$$

   Hence inequality (19) holds.

2. Consider inequality (20). By A1, $\text{var}(X^{\perp W}) > 0$. By equation (2), $\mathbb{L}(X \mid 1, W_1, W_2) = \pi_1' W_1 + \pi_2 W_2 +$ const., since $X^{\perp W}$ is uncorrelated with W. Therefore,

$$\begin{aligned} \text{var}(X) &> \text{var}(\mathbb{L}(X \mid 1, W_1, W_2)) \\ &= \text{var}(\pi_1' W_1 + \pi_2 W_2) \\ &= \pi_1' \text{var}(W_1) + 2\pi_1' c \pi_2 + \pi_2^2 \\ &= \pi_1' \text{var}(W_1) + 2\pi_1' c r_X' \pi_1 + (\pi_1' r_X)(r_X' \pi_1) \\ &= \pi_1' (\text{var}(W_1) + 2 c r_X' + r_X r_X') \pi_1. \end{aligned}$$

   Hence inequality (20) holds.

3. Consider inequality (21). We have

$$\begin{aligned} c' \text{var}(W_1)^{-1} c &= \text{cov}(W_2, W_1) \text{var}(W_1)^{-1} \text{cov}(W_1, W_2) \\ &= R^2_{W_2 \sim W_1} \\ &< 1. \end{aligned}$$

   The last line follows by A1. Hence inequality (21) holds.

Putting all of these results together we have shown that $\beta_{\text{long}} \in \mathcal{B}(r_X, r_Y, c)$ and therefore $\beta_{\text{long}} \in \mathcal{B}_I(\bar{r}_X, \bar{r}_Y, \bar{c})$.

Part 2 (Sharpness). Let $b \in \mathcal{B}_I(\bar{r}_X, \bar{r}_Y, \bar{c})$. We will construct a joint distribution for (Y, X, W_1, W_2) for which $\beta_{\text{long}} = b$ and which is consistent with all of our assumptions and the observed distribution of (Y, X, W_1).

Part 1: Constructing a covariance matrix. Since $b \in \mathcal{B}_I(\bar{r}_X, \bar{r}_Y, \bar{c})$, let (r_X, r_Y, c) be such that $b \in \mathcal{B}(r_X, r_Y, c)$, $\|r_X\|_{\Sigma^{-1}_{\text{obs}}} \le \bar{r}_X$, $\|r_Y\|_{\Sigma^{-1}_{\text{obs}}} \le \bar{r}_Y$, and $\|c\|_{\Sigma^{-1}_{\text{obs}}} \le \bar{c}$. Using the definition of $\mathcal{B}(r_X, r_Y, c)$, let $(p_1, g_1) \in \Pi_1(r_X, c) \times \mathbb{R}^{d_1}$ be such that

$$\text{cov}(Y, (X, W_1)) = \begin{pmatrix} b & g_1' \end{pmatrix} \mathbf{A}(r_X, r_Y, c, p_1).$$

Define $\gamma_1 = g_1$, $\pi_1 = p_1$, $\gamma_2 = r_Y' \gamma_1$, and $\pi_2 = r_X' \pi_1$. By Lemma 1, this implies that A3 and A5 hold.

Next we construct a random variable W_2 together with (W_1, X, Y) such that A1, A4, and A6 hold, and such that equations (1) and (2) hold. To do so, define the matrix

$$\Sigma = \begin{pmatrix} 1 & c' & p_1'(c + r_X) & b p_1'(c + r_X) + g_1'(c + r_Y) \\ c & \text{var}(W_1) & \text{cov}(W_1, X) & \text{cov}(W_1, Y) \\ p_1'(c + r_X) & \text{cov}(X, W_1) & \text{var}(X) & \text{cov}(X, Y) \\ b p_1'(c + r_X) + g_1'(c + r_Y) & \text{cov}(Y, W_1) & \text{cov}(Y, X) & \text{var}(Y) \end{pmatrix}.$$

Note that the bottom right $(d_1 + 2) \times (d_1 + 2)$ block is $\text{var}(W_1, X, Y)$, the variance matrix of (W_1, X, Y). We next show that Σ is positive definite. To do this, we apply Sylvester's criterion as in the proof of Theorem 1. First, we show that $(1 + d_1) \times (1 + d_1)$ matrix

$$\Sigma_2 = \begin{pmatrix} 1 & c' \\ c & \text{var}(W_1) \end{pmatrix}$$

is positive definite. Repeating the arguments from part 2(i) of the proof of Theorem 1, Σ_2 is positive definite if and only if $c' \text{var}(W_1)^{-1} c < 1$. This inequality holds by equation (21).

Next consider the $(1 + d_1 + 1) \times (1 + d_1 + 1)$ leading principal minor of Σ. Denote it by Σ_3. By the formula for determinants of partitioned matrices, it has determinant equal to

$$\det(\Sigma_3) = \det \begin{pmatrix} 1 & c' & (c + r_X)' p_1 \\ c & \text{var}(W_1) & (\text{var}(W_1) + c r_X') p_1 \\ p_1'(c + r_X) & p_1'(\text{var}(W_1) + r_X c') & \text{var}(X) \end{pmatrix}$$

$$
\begin{aligned}
&= \det(\Sigma_2) \cdot \left(\operatorname{var}(X) - \begin{pmatrix} p_1'(c + r_X) & p_1'(\operatorname{var}(W_1) + r_X c') \end{pmatrix} \Sigma_2^{-1} \begin{pmatrix} (c + r_X)' p_1 \\ (\operatorname{var}(W_1) + c r_X') p_1 \end{pmatrix} \right) \\
&= \det(\Sigma_2) \cdot \left(\operatorname{var}(X) - \begin{pmatrix} p_1' r_X & p_1' \end{pmatrix} \begin{pmatrix} (c + r_X)' p_1 \\ (\operatorname{var}(W_1) + c r_X') p_1 \end{pmatrix} \right) \\
&= \det(\Sigma_2) \cdot \left(\operatorname{var}(X) - p_1' \left(r_X r_X' + r_X c' + c r_X' + \operatorname{var}(W_1) \right) p_1 \right) \\
&= \det(\Sigma_2) \cdot \left(\operatorname{var}(X) - p_1' \left(r_X r_X' + 2 r_X c' + \operatorname{var}(W_1) \right) p_1 \right) \\
&> 0.
\end{aligned}
$$

In the first line we used the formula for $\operatorname{cov}(X, W_1)$ that we derived earlier. In the third line we used

$$
\Sigma_2^{-1} = \frac{1}{1 - c' \operatorname{var}(W_1)^{-1} c} \begin{pmatrix} 1 & -c' \operatorname{var}(W_1)^{-1} \\ -\operatorname{var}(W_1)^{-1} c & \operatorname{var}(W_1)^{-1}(1 - c' \operatorname{var}(W_1)^{-1} c) + \operatorname{var}(W_1)^{-1} c c' \operatorname{var}(W_1)^{-1} \end{pmatrix}
$$

and

$$
\begin{aligned}
&\frac{1}{1 - c' \operatorname{var}(W_1)^{-1} c} \begin{pmatrix} p_1'(c + r_X) & p_1'(\operatorname{var}(W_1) + r_X c') \end{pmatrix} \\
&\qquad\qquad \cdot \begin{pmatrix} -c' \operatorname{var}(W_1)^{-1} \\ \operatorname{var}(W_1)^{-1}(1 - c' \operatorname{var}(W_1)^{-1} c) + \operatorname{var}(W_1)^{-1} c c' \operatorname{var}(W_1)^{-1} \end{pmatrix} = p_1'.
\end{aligned}
$$

The fifth line uses $p_1'(c r_X') p_1 = p_1'(r_X c') p_1$. The last line follows since we already showed that $\det(\Sigma_2) > 0$, and from inequality (20), assumed in the definition of $\mathcal{B}(r_X, r_Y, c)$.

Finally, the determinant of the entire matrix is

$$
\begin{aligned}
\det(\Sigma) &= \det(\Sigma_3) \cdot \left(\operatorname{var}(Y) - \begin{pmatrix} b p_1'(c + r_X) + g_1'(c + r_Y) \\ b \operatorname{cov}(X, W_1) + g_1'(\operatorname{var}(W_1) + r_Y c') \\ b \operatorname{var}(X) + g_1'(\operatorname{var}(W_1) + c r_X' + r_Y c' + r_Y r_X') p_1 \end{pmatrix}' \right. \\
&\qquad \begin{pmatrix} 1 & c' & (c + r_X)' p_1 \\ c & \operatorname{var}(W_1) & (\operatorname{var}(W_1) + c r_X') p_1 \\ p_1'(c + r_X) & p_1'(\operatorname{var}(W_1) + r_X c') & \operatorname{var}(X) \end{pmatrix}^{-1} \\
&\qquad \left. \begin{pmatrix} b (c + r_X)' p_1 + (c + r_Y)' g_1 \\ b \operatorname{cov}(W_1, X) + (\operatorname{var}(W_1) + c r_Y') g_1 \\ b \operatorname{var}(X) + g_1'(\operatorname{var}(W_1) + c r_X' + r_Y c' + r_Y r_X') p_1 \end{pmatrix} \right) \\
&= \det(\Sigma_3) \cdot \left(\operatorname{var}(Y) - \begin{pmatrix} g_1' r_Y & g_1' & b \end{pmatrix} \begin{pmatrix} b (c + r_X)' p_1 + (c + r_Y)' g_1 \\ b (\operatorname{var}(W_1) + c r_X') p_1 + (\operatorname{var}(W_1) + c r_Y') g_1 \\ b \operatorname{var}(X) + g_1'(\operatorname{var}(W_1) + c r_X' + r_Y c' + r_Y r_X') p_1 \end{pmatrix} \right) \\
&= \det(\Sigma_3) \cdot \Big(\operatorname{var}(Y) - (b^2 \operatorname{var}(X) + g_1' \operatorname{var}(W_1) g_1 \\
&\qquad + 2 b p_1'(\operatorname{var}(W_1) + r_X c') g_1 + (g_1' r_Y)^2 + 2 b g_1' r_Y p_1'(c + r_X) + 2 g_1' r_Y g_1' c) \Big) \\
&> 0.
\end{aligned}
$$

The first line follows since Σ_3 is invertible and by the determinant formula for partitioned matrices. The second line follows by similar calculations as earlier. The last line follows since we have already shown that $\det(\Sigma_3) > 0$, and from inequality (19) in the definition of $\mathcal{B}(r_X, r_Y, c)$. Thus the matrix Σ is positive definite by Sylvester's Criterion.

Part 2: Constructing the joint distribution of (Y, X, W_1, W_2). Let

$$\begin{pmatrix} \widetilde{W}_1 \\ \widetilde{X} \\ \widetilde{Y} \end{pmatrix} = \text{Chol}(\text{var}(W_1, X, Y)^{-1})' \begin{pmatrix} W_1 \\ X \\ Y \end{pmatrix}$$

denote the whitened vector of observables (W_1, X, Y). Let $\widetilde{W}_2$ be a random variable that has unit variance and is uncorrelated with $(\widetilde{W}_1, \widetilde{X}, \widetilde{Y})$. Define

$$\begin{pmatrix} W_2 \\ W_1 \\ X \\ Y \end{pmatrix} = \text{Chol}(\Sigma)' \begin{pmatrix} \widetilde{W}_2 \\ \widetilde{W}_1 \\ \widetilde{X} \\ \widetilde{Y} \end{pmatrix},$$

which undoes the whitening. This equation defines a joint vector of random variables (W_2, W_1, X, Y).

Part 3: Verifying consistency with the data. This random vector has variance matrix Σ. Moreover, it leaves (W_1, X, Y) unchanged due to the triangular structure of $\text{Chol}(\Sigma)$ and $\text{Chol}(\text{var}(W_1, X, Y)^{-1})$. Therefore the (W_1, X, Y) marginal distribution is the same as the observed data.

Part 4: Verifying consistency with the assumptions. We already showed that A3 and A5 hold. A4 holds by definition of W_2. Next note that the constructed (W_2, W_1, X, Y) are not perfectly multicollinear as their variance matrix is positive definite. Therefore A1 holds. A6 holds by $\text{cov}(W_1, W_2) = c$ and $\|c\|_{\Sigma_{\text{obs}}^{-1}} \leq \bar{c}$.

Finally, we show that this distribution of (W_2, W_1, X, Y) together with $(b, p_1, g_1, r_X, r_Y, c)$ imply a distribution for $(Y^{\perp X,W}, X^{\perp W})$ such that equations (1) and (2) hold with the correct coefficients. Specifically, let

$$Y^{\perp X,W} = Y - bX - g_1' W_1 - r_Y' g_1 W_2$$

and

$$X^{\perp W} = X - p_1' W_1 - r_X' p_1 W_2.$$

We next verify that $Y^{\perp X,W}$ and $X^{\perp W}$ are uncorrelated with (X, W_1, W_2) and (W_1, W_2) respectively. We start by calculating the covariances of $Y^{\perp X,W}$ and (X, W_1, W_2). First,

$$\begin{aligned}
\text{cov}(Y^{\perp X,W}, X) &= \text{cov}(Y - bX - g_1' W_1 - \gamma_2 W_2, X) \\
&= \text{cov}(Y, X) - b\,\text{var}(X) - \text{cov}(X, W_1) g_1 - r_Y' g_1 p_1' (r_X + c) \\
&= (b\,\text{var}(X) + g_1'(\text{var}(W_1) + c r_X' + r_Y c' + r_Y r_X') p_1) \\
&\quad\; - b\,\text{var}(X) - g_1'(\text{var}(W_1) + c r_X') p_1 - g_1'(r_Y r_X' + r_Y c') p_1 \\
&= 0.
\end{aligned}$$

The third equality follows from the equation $\text{cov}(Y, (X, W_1)) = \begin{pmatrix} b & g_1' \end{pmatrix} \mathbf{A}(r_X, r_Y, c, p_1)$.

Next,

$$\begin{aligned}
\text{cov}(Y^{\perp X,W}, W_1) &= \text{cov}(Y - bX - g_1' W_1 - \gamma_2 W_2, W_1) \\
&= \text{cov}(Y, W_1) - b\,\text{cov}(X, W_1) - g_1'\,\text{var}(W_1) - r_Y' g_1 c' \\
&= (b\,\text{cov}(X, W_1) + g_1'(\text{var}(W_1) + r_Y c')) - b\,\text{cov}(X, W_1) - g_1'\,\text{var}(W_1) - g_1' r_Y c' \\
&= 0.
\end{aligned}$$

The third equality also follows from the equation $\text{cov}(Y, (X, W_1)) = \begin{pmatrix} b & g_1' \end{pmatrix} \mathbf{A}(r_X, r_Y, c, p_1)$. Similarly,

$$\begin{aligned}
\text{cov}(Y^{\perp X,W}, W_2) &= \text{cov}(Y - bX - g_1 W_1 - \gamma_2 W_2, W_2) \\
&= \text{cov}(Y, W_2) - b\,\text{cov}(X, W_2) - g_1' c - r_Y' g_1 \\
&= (b p_1'(c + r_X) + g_1'(c + r_Y)) - b p_1'(c + r_X) - g_1' c - g_1' r_Y
\end{aligned}$$

$$= 0.$$

The second line uses $\text{var}(W_2) = 1$. The third line substitutes the values of $\text{cov}(Y, W_2)$ and $\text{cov}(X, W_2)$ from Σ, the variance matrix for our constructed distribution of (W_2, W_1, X, Y).

As for $X^{\perp W}$, we can see that

$$\begin{aligned}
\text{cov}(X^{\perp W}, W_1) &= \text{cov}(X, W_1) - p_1' \text{var}(W_1) - r_X' p_1 c' \\
&= p_1'(\text{var}(W_1) + r_X c') - p_1' \text{var}(W_1) - p_1' r_X c' \\
&= 0 \\
\text{cov}(X^{\perp W}, W_2) &= \text{cov}(X, W_2) - p_1' c - r_X' p_1 \\
&= p_1'(r_X + c) - p_1' c - p_1' r_X \\
&= 0.
\end{aligned}$$

Here we used $\text{cov}(X, W_1) = p_1'(\text{var}(W_1) + r_X c')$ (part of the definition of $\mathcal{B}(r_X, r_Y, c)$), and $\text{cov}(X, W_2) = p_1'(r_X + c)$, which follows from the construction of (W_2, W_1, X, Y) to have variance matrix Σ.

Overall, we have constructed a joint distribution of (W_2, W_1, X, Y) that is consistent with the assumptions, the observed data, and for which $\beta_{\text{long}} = b$. In particular, assumptions A1 and A3–A6 hold, and equations (1) and (2) hold for this distribution of (W_2, W_1, X, Y), the residuals $Y^{\perp X,W}$ and $X^{\perp W}$, and the corresponding coefficients $(b, p_1, \pi_2, g_1, \gamma_2, r_X, r_Y, c)$. It follows that the set $\mathcal{B}_I(\bar{r}_X, \bar{r}_Y, \bar{c})$ is sharp and therefore the proof is complete. □

E Proofs for Section 3

Proof of equation (3). By the FWL theorem,

$$\beta_{\text{med}} = \frac{\text{cov}(Y, X^{\perp W_1})}{\text{var}(X^{\perp W_1})}.$$

Next,

$$\begin{aligned}
\text{cov}(Y, X^{\perp W_1}) &= \text{cov}(X\beta_{\text{long}} + \gamma_1' W_1 + \gamma_2 W_2 + Y^{\perp X,W}, X^{\perp W_1}) \\
&= \beta_{\text{long}} \text{var}(X^{\perp W_1}) + \gamma_2 \text{cov}(W_2, X^{\perp W_1})
\end{aligned}$$

and

$$\begin{aligned}
\text{cov}(W_2, X^{\perp W_1}) &= \text{cov}(W_2, X - \text{cov}(X, W_1) \text{var}(W_1)^{-1} W_1) \\
&= \text{cov}(W_2, X) - \text{cov}(X, W_1) \text{var}(W_1)^{-1} \text{cov}(W_1, W_2) \\
&= \pi_1' \text{cov}(W_1, W_2) + \pi_2 - (\pi_1' \text{var}(W_1) + \pi_2 \text{cov}(W_2, W_1)) \text{var}(W_1)^{-1} \text{cov}(W_1, W_2) \\
&= \pi_2 - \pi_2 \text{cov}(W_2, W_1) \text{var}(W_1)^{-1} \text{cov}(W_1, W_2) \\
&= \pi_2(1 - R^2_{W_2 \sim W_1}).
\end{aligned}$$

Putting these together gives the desired result. □

E.1 Useful Definitions and Lemmas

Before proving the main results of section 3.2, we first define some useful notation and state and prove some useful lemmas. Recall from remark 1 that equation (18) can be written as the following two equalities:

$$\sigma_{X,Y} - b = p_1'(I + r_X c' + c r_Y' + r_X r_Y') g_1 \tag{23}$$

$$(I + c r_Y') g_1 = \sigma_{W_1,Y} - b\sigma_{W_1,X}. \tag{24}$$

Here we have imposed the normalizations $\text{var}(X) = 1$ and $\text{var}(W_1) = I$. We maintain these normalizations throughout this section. We also use the notation $\sigma_{A,B} = \text{cov}(A, B)$.

Lemma 2. Let $\text{var}(Y, X, W_1)$ be positive definite. Then $k_2 k_0 > k_1^2$.

Proof of Lemma 2. By the determinant formula for partitioned matrices,

$$\begin{aligned}
\det \begin{pmatrix} \sigma_Y^2 & \sigma_{X,Y} & \sigma_{W_1,Y}' \\ \sigma_{X,Y} & 1 & \sigma_{W_1,X}' \\ \sigma_{W_1,Y} & \sigma_{W_1,X} & I \end{pmatrix} &= \det(I) \cdot \det \left(\begin{pmatrix} \sigma_Y^2 & \sigma_{X,Y} \\ \sigma_{X,Y} & 1 \end{pmatrix} - \begin{pmatrix} \sigma_{W_1,Y}' \\ \sigma_{W_1,X}' \end{pmatrix} I^{-1} \begin{pmatrix} \sigma_{W_1,Y} & \sigma_{W_1,X} \end{pmatrix} \right) \\
&= \det \left(\begin{pmatrix} \sigma_Y^2 & \sigma_{X,Y} \\ \sigma_{X,Y} & 1 \end{pmatrix} - \begin{pmatrix} \|\sigma_{W_1,Y}\|^2 & \sigma_{W_1,Y}' \sigma_{W_1,X} \\ \sigma_{W_1,Y}' \sigma_{W_1,X} & \|\sigma_{W_1,X}\|^2 \end{pmatrix} \right) \\
&= \det \begin{pmatrix} \sigma_Y^2 - \|\sigma_{W_1,Y}\|^2 & \sigma_{X,Y} - \sigma_{W_1,Y}' \sigma_{W_1,X} \\ \sigma_{X,Y} - \sigma_{W_1,Y}' \sigma_{W_1,X} & 1 - \|\sigma_{W_1,X}\|^2 \end{pmatrix} \\
&= \det \begin{pmatrix} k_2 & k_1 \\ k_1 & k_0 \end{pmatrix} \\
&= k_2 k_0 - k_1^2.
\end{aligned}$$

The fourth line follows by definition of these constants. By assumption, the variance matrix of (Y, X, W_1) is positive definite. Therefore $k_2 k_0 - k_1^2 > 0$. $\square$

The next lemma provides a useful way of rewriting certain quadratic forms that arise in the characterization of $\mathcal{B}(r_X, r_Y, c)$.

Lemma 3. Let $(I + cr_1')w_1 = v_1$ and $(I + cr_2')w_2 = v_2$. Then

$$w_1'(I + r_1 c' + cr_2' + r_1 r_2')w_2 = v_1' v_2 + (w_1' r_1)(w_2' r_2)(1 - \|c\|^2).$$

Proof of Lemma 3. Note that $I + r_1 c' + cr_2' + r_1 r_2' = (I + r_1 c')(I + cr_2') + (1 - \|c\|^2) r_1 r_2'$. Therefore

$$\begin{aligned}
w_1'(I + r_1 c' + cr_2' + r_1 r_2')w_2 &= w_1'(I + r_1 c')(I + cr_2')w_2 + (1 - \|c\|^2)(w_1' r_1)(r_2' w_2) \\
&= v_1' v_2 + (1 - \|c\|^2)(w_1' r_1)(r_2' w_2).
\end{aligned}$$

$\square$

We next use the previous lemma to provide a simplified characterization of the set $\mathcal{B}(r_X, r_Y, c)$.

Lemma 4. Let $\text{var}(Y, X, W_1)$ be positive definite. Then $b \in \mathcal{B}(r_X, r_Y, c)$ if and only if the following six equations hold for some $(p_1, g_1) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_1}$:

$$\begin{aligned}
(I + cr_X')p_1 &= \sigma_{W_1,X} \\
(I + cr_Y')g_1 &= \sigma_{W_1,Y} - b\sigma_{W_1,X} \\
(p_1' r_X)(g_1' r_Y)(1 - \|c\|^2) &= k_1 - bk_0 && (25) \\
(p_1' r_X)^2(1 - \|c\|^2) &< k_0 && (26) \\
(g_1' r_Y)^2(1 - \|c\|^2) &< k_0(\beta_{\text{med}} - b)^2 + k_0 \left(\frac{k_2}{k_0} - \beta_{\text{med}}^2 \right) && (27) \\
\|c\|^2 &< 1. && (28)
\end{aligned}$$

Proof of Lemma 4. By the definition of $\mathcal{B}(r_X, r_Y, c)$, $b \in \mathcal{B}(r_X, r_Y, c)$ if and only if there exists $(p_1, g_1) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_1}$ such that $(I + cr_X')p_1 = \sigma_{W_1,X}$, $(I + cr_Y')g_1 = \sigma_{W_1,Y} - b\sigma_{W_1,X}$, and equations (19), (20), (21), and (23) hold. We will show that each of these four equations is equivalent to one of the four numbered equations in the lemma.

Part 1: (23) *is equivalent to* (25). Since $(I + cr_X')p_1 = \sigma_{W_1,X}$ and $(I + cr_Y')g_1 = \sigma_{W_1,Y} - b\sigma_{W_1,X}$, the conditions of Lemma 3 hold with $(r_1, r_2) = (r_X, r_Y)$, $(w_1, w_2) = (p_1, g_1)$, and $(v_1, v_2) = (\sigma_{W_1,X}, \sigma_{W_1,Y} -$

$b\sigma_{W_1,X}$). The lemma therefore gives

$$p_1'(I + r_X c' + c r_Y' + r_X r_Y') g_1 = \sigma_{W_1,Y}' \sigma_{W_1,X} - b\|\sigma_{W_1,X}\|^2 + (p_1' r_X)(g_1' r_Y)(1 - \|c\|^2).$$

Hence (23) can be written as

$$\begin{aligned} \sigma_{XY} - b &= p_1'(I + r_X c' + c r_Y' + r_X r_Y') g_1 \\ &= \sigma_{W_1,Y}' \sigma_{W_1,X} - b\|\sigma_{W_1,X}\|^2 + (p_1' r_X)(g_1' r_Y)(1 - \|c\|^2) \end{aligned}$$

or, equivalently, as

$$k_1 - b k_0 = (p_1' r_X)(g_1' r_Y)(1 - \|c\|^2).$$

Part 2: (20) *is equivalent to* (26). Apply Lemma 3 with $r_1 = r_2 = r_X$, $w_1 = w_2 = p_1$, $v_1 = v_2 = \sigma_{W_1,X}$ to obtain

$$\begin{aligned} p_1'(I + r_X r_X' + 2c r_X') p_1 &= p_1'(I + r_X r_X' + c r_X' + r_X c') p_1 \\ &= \|\sigma_{W_1,X}\|^2 + (p_1' r_X)^2 (1 - \|c\|^2). \end{aligned}$$

Therefore (20) is equivalent to

$$\begin{aligned} 1 &> p_1'(I + r_X r_X' + 2c r_X') p_1 \\ &= \|\sigma_{W_1,X}\|^2 + (p_1' r_X)^2 (1 - \|c\|^2) \end{aligned}$$

which is equivalent to

$$k_0 > (p_1' r_X)^2 (1 - \|c\|^2).$$

Part 3: (19) *is equivalent to* (27). Apply Lemma 3 with $r_1 = r_2 = r_Y$, $w_1 = w_2 = g_1$, $v_1 = v_2 = \sigma_{W_1,Y} - b\sigma_{W_1,X}$ to find that

$$g_1'(I + r_Y r_Y' + 2c r_Y') g_1 = \|\sigma_{W_1,Y} - b\sigma_{W_1,X}\|^2 + (g_1' r_Y)^2 (1 - \|c\|^2).$$

Therefore

$$\begin{aligned} \sigma_Y^2 &> b^2 + g_1'(I + r_Y r_Y + 2c r_Y') g_1 + 2b p_1'(I + r_X c' + c r_Y' + r_X r_Y') g_1 \\ &= b^2 + \|\sigma_{W_1,Y} - b\sigma_{W_1,X}\|^2 + (g_1' r_Y)^2 (1 - \|c\|^2) + 2b(\sigma_{X,Y} - b) \\ &= b^2 + \|\sigma_{W_1,Y}\|^2 + b^2 \|\sigma_{W_1,X}\|^2 - 2b \sigma_{W_1,X}' \sigma_{W_1,Y} + (g_1' r_Y)^2 (1 - \|c\|^2) + 2b(\sigma_{X,Y} - b). \end{aligned}$$

The second line uses our application of Lemma 3 as well as equation (23). This inequality is equivalent to

$$\begin{aligned} (g_1' r_Y)^2 (1 - \|c\|^2) &< b^2 k_0 - 2b k_1 + k_2 \\ &= k_0 (b^2 - 2b(k_1/k_0) + k_2/k_0) \\ &= k_0 (b^2 - 2b\beta_{\text{med}} + \beta_{\text{med}}^2 + k_2/k_0 - \beta_{\text{med}}^2) \\ &= k_0 (b - \beta_{\text{med}})^2 + k_0 (k_2/k_0 - \beta_{\text{med}}^2). \end{aligned}$$

The third line uses $k_1/k_0 = \beta_{\text{med}}$.

Part 4: (21) *is equivalent to* (28). This follows from the normalization $\text{var}(W_1) = I$. $\square$

Our previous lemma characterized $\mathcal{B}(r_X, r_Y, c)$. The next lemma characterizes every element of this set which is not equal to β_{med}. By treating the element β_{med} separately, we obtain an even simpler characterization of the elements in $\mathcal{B}(r_X, r_Y, c)$.

Lemma 5. Let $\text{var}(Y, X, W_1)$ be positive definite. Let $b \neq \beta_{\text{med}}$. Then $b \in \mathcal{B}(r_X, r_Y, c)$ if and only if there exists $z \in \mathbb{R} \setminus \{0\}$ such that the following hold:

$$k_1 - bk_0 = r_Y' \left(z\sqrt{1 - \|c\|^2}(\sigma_{W_1,Y} - b\sigma_{W_1,X}) - (k_1 - bk_0)c \right) \tag{29}$$

$$z = r_X' \left(\sqrt{1 - \|c\|^2}\sigma_{W_1,X} - cz \right) \tag{30}$$

$$z^2 < k_0 \tag{31}$$

$$(b - \beta_{\text{med}})^2 < \text{devsq}(z) \tag{32}$$

$$\|c\|^2 < 1. \tag{33}$$

Proof of Lemma 5. Throughout the proof, keep in mind that the lemma assumes $b \neq \beta_{\text{med}}$.

Step 1 ($\Rightarrow$). First we show that $b \in \mathcal{B}(r_X, r_Y, c)$ implies (29)–(33) hold for some $z \in \mathbb{R} \setminus \{0\}$.

By Lemma 4, $b \in \mathcal{B}(r_X, r_Y, c)$ is equivalent to the existence of $(p_1, g_1) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_1}$ such that $(I + cr_X')p_1 = \sigma_{W_1,X}$, $(I + cr_Y')g_1 = \sigma_{W_1,Y} - b\sigma_{W_1,X}$, and equations (25), (26), (27), and (28) hold. Fix such $(p_1, g_1) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_1}$ and let

$$z = p_1' r_X \sqrt{1 - \|c\|^2}.$$

First we show that $z \neq 0$. To see this, note that $z = 0$ implies $p_1' r_X = 0$ since $\|c\|^2 < 1$. Equation (25) then implies that $k_1 - bk_0 = 0$. This is a contradiction since $\beta_{\text{med}} = k_1/k_0$ and we assumed $b \neq \beta_{\text{med}}$. Therefore $z \neq 0$.

Second, we show that (25) implies (29). Multiplying both sides of (25) by $(1 + r_Y' c)$, we have that

$$\begin{aligned}
(k_1 - bk_0)(1 + r_Y' c) &= (p_1' r_X)(1 - \|c\|^2)(r_Y' g_1)(1 + r_Y' c) \\
&= (p_1' r_X)(1 - \|c\|^2) r_Y' (I + c r_Y') g_1 \\
&= (p_1' r_X)(1 - \|c\|^2) r_Y' (\sigma_{W_1,Y} - b\sigma_{W_1,X}) \\
&= r_Y' \left((p_1' r_X)\sqrt{1 - \|c\|^2}(\sigma_{W_1,Y} - b\sigma_{W_1,X})\sqrt{1 - \|c\|^2} \right) \\
&= r_Y' (\sigma_{W_1,Y} - b\sigma_{W_1,X}) z\sqrt{1 - \|c\|^2}.
\end{aligned}$$

The fifth line follows by the definition of z. Rearranging the last equality yields (29).

Third, we show that (30) holds for this choice of z. Note that $p_1 = \sigma_{W_1,X} - c(p_1' r_X)$. Thus

$$\begin{aligned}
z &= r_X' p_1 \sqrt{1 - \|c\|^2} \\
&= r_X' \left(\sigma_{W_1,X}\sqrt{1 - \|c\|^2} - c(p_1' r_X \sqrt{1 - \|c\|^2}) \right) \\
&= r_X' (\sigma_{W_1,X}\sqrt{1 - \|c\|^2} - cz).
\end{aligned}$$

Fourth, we show that (26) implies (31). Substituting $p_1' r_X \sqrt{1 - \|c\|^2} = z$ into (26) gives $z^2 < k_0$, which is equation (31).

Fifth, we show that equations (25) and (27) imply equation (32). Dividing both sides of (25) by $p_1' r_X \sqrt{1 - \|c\|^2}$ gives

$$g_1' r_Y \sqrt{1 - \|c\|^2} = \frac{k_1 - bk_0}{p_1' r_X \sqrt{1 - \|c\|^2}}.$$

Substitute this into (27) to get

$$\begin{aligned}
k_0(\beta_{\text{med}} - b)^2 + k_0 \left(\frac{k_2}{k_0} - \beta_{\text{med}}^2 \right) &> (g_1' r_Y)^2 (1 - \|c\|^2) \\
&= \frac{(k_1 - bk_0)^2}{(p_1' r_X)^2 (1 - \|c\|^2)} \\
&= \frac{k_0^2 (\beta_{\text{med}} - b)^2}{z^2}.
\end{aligned}$$

Rearranging this inequality and recalling that $k_0 - z^2 > 0$, we obtain

$$(\beta_{\text{med}} - b)^2 < \frac{z^2\left(\frac{k_2}{k_0} - \beta_{\text{med}}^2\right)}{k_0 - z^2}.$$

Therefore equation (32) holds. Finally, note that equations (28) and (33) are identical.

Step 2 ($\Leftarrow$). Next we show that if equations (29)–(33) hold for some $z \in \mathbb{R} \setminus \{0\}$ then $b \in \mathcal{B}(r_X, r_Y, c)$.

Let

$$\begin{aligned} p_1 &= \sigma_{W_1,X} - c\frac{z}{\sqrt{1-\|c\|^2}} \\ g_1 &= \sigma_{W_1,Y} - b\sigma_{W_1,X} - c\frac{k_1 - bk_0}{z\sqrt{1-\|c\|^2}}. \end{aligned}$$

First, we verify that $(I + cr_X')p_1 = \sigma_{W_1,X}$:

$$\begin{aligned} (I + cr_X')p_1 &= (I + cr_X')\left(\sigma_{W_1,X} - c\frac{z}{\sqrt{1-\|c\|^2}}\right) \\ &= \sigma_{W_1,X} + c(r_X'\sigma_{W_1,X}) - c\frac{(1 + r_X'c)z}{\sqrt{1-\|c\|^2}} \\ &= \sigma_{W_1,X} + c(r_X'\sigma_{W_1,X}) - c(r_X'\sigma_{W_1,X}) \\ &= \sigma_{W_1,X}. \end{aligned}$$

The third line follows by (30).

Second, we verify that $(I + cr_Y')g_1 = \sigma_{W_1,Y} - b\sigma_{W_1,X}$:

$$\begin{aligned} (I + cr_Y')g_1 &= (I + cr_Y')\left(\sigma_{W_1,Y} - b\sigma_{W_1,X} - c\frac{k_1 - bk_0}{z\sqrt{1-\|c\|^2}}\right) \\ &= \sigma_{W_1,Y} - b\sigma_{W_1,X} + c\left((\sigma_{W_1,Y} - b\sigma_{W_1,X})'r_Y\right) - c\frac{(1 + r_Y'c)(k_1 - bk_0)}{z\sqrt{1-\|c\|^2}} \\ &= \sigma_{W_1,Y} - b\sigma_{W_1,X} + c\left((\sigma_{W_1,Y} - b\sigma_{W_1,X})'r_Y\right) - c\big(r_Y'(\sigma_{W_1,Y} - b\sigma_{W_1,X})\big) \\ &= \sigma_{W_1,Y} - b\sigma_{W_1,X}. \end{aligned}$$

The third line follows by (29).

Third, note that by the definition of p_1 and equation (30),

$$p_1'r_X\sqrt{1-\|c\|^2} = z.$$

Similarly, by the definition of g_1 and equation (29),

$$g_1'r_Y\sqrt{1-\|c\|^2} = \frac{k_1 - bk_0}{z}.$$

Multiplying these two equations together gives

$$(p_1'r_X)(g_1'r_Y)(1 - \|c\|^2) = z \cdot \frac{k_1 - bk_0}{z} = k_1 - bk_0,$$

Hence equation (25) holds. Fourth, we see that

$$(p_1'r_X)^2(1 - \|c\|^2) = z^2$$

$$< k_0$$

by equation (31). Therefore inequality (26) holds. Finally, since $z^2 < k_0$ by inequality (31), inequality (32),

$$(\beta_{\text{med}} - b)^2 < \frac{z^2(k_2/k_0 - \beta_{\text{med}}^2)}{k_0 - z^2}$$

is equivalent to

$$(k_1 - bk_0)^2(k_0 - z^2) < z^2(k_2k_0 - k_1^2)$$

and to

$$(k_1 - bk_0)^2 k_0 < z^2(k_1 - bk_0)^2 + z^2(k_2k_0 - k_1^2).$$

Using this inequality we find that

$$\begin{aligned}
(g_1' r_Y)^2(1 - \|c\|^2) &= \frac{k_0(k_1 - bk_0)^2}{k_0 z^2} \\
&< \frac{1}{k_0 z^2} \cdot \left(z^2(k_1 - bk_0)^2 + z^2(k_2k_0 - k_1^2)\right) \\
&= \frac{(k_1 - bk_0)^2}{k_0} + \frac{k_2k_0 - k_1^2}{k_0} \\
&= k_0(b - \beta_{\text{med}})^2 + k_0(k_2/k_0 - \beta_{\text{med}}^2).
\end{aligned}$$

Therefore inequality (27) holds. Finally, note that equations (28) and (33) are identical.

We have shown that equations (25), (26), (27), and (28) hold, as well as $(I + cr_Y')g_1 = \sigma_{W_1,Y} - b\sigma_{W_1,X}$ and $(I + cr_X')p_1 = \sigma_{W_1,X}$. Hence Lemma 4 implies that $b \in \mathcal{B}(r_X, r_Y, c)$. $\square$

For the next several lemmas, we use the function

$$z_X(r_X, c) = \frac{r_X' \sigma_{W_1,X}\sqrt{1 - \|c\|^2}}{1 + r_X' c}.$$

This function is well defined for all (r_X, c) such that $r_X' c \neq -1$. When $r_X' c \neq -1$, equation (30) is equivalent to $z = z_X(r_X, c)$.

The following lemma examines the set $\mathcal{B}(r_X, r_Y, c)$ when $r_Y = 0$. It characterizes when this set is the singleton containing β_{med}. We use this result to handle the value β_{med} separately from the other elements of $\mathcal{B}(r_X, r_Y, c)$.

Lemma 6. Let $\text{var}(Y, X, W_1)$ be positive definite. Let $\|c\| < 1$. Then

$$\mathcal{B}(r_X, 0, c) = \{\beta_{\text{med}}\}$$

if and only if there exists a $p_1 \in \mathbb{R}^{d_1}$ such that $(I + cr_X')p_1 = \sigma_{W_1,X}$ and (26) holds. In particular, $\mathcal{B}(r_X, 0, c) = \{\beta_{\text{med}}\}$ for all (r_X, c) such that $r_X' c \neq -1$ and $z_X(r_X, c)^2 < k_0$.

Proof of Lemma 6. ($\Rightarrow$) Suppose $\mathcal{B}(r_X, 0, c) = \{\beta_{\text{med}}\}$. By Lemma 4, there exists a $p_1 \in \mathbb{R}^{d_1}$ such that $(I + cr_X')p_1 = \sigma_{W_1,X}$ and equation (26) holds.

($\Leftarrow$) Conversely, suppose there exists a $p_1 \in \mathbb{R}^{d_1}$ such that $(I + cr_X')p_1 = \sigma_{W_1,X}$ and equation (26) holds.

1. First we show that $\beta_{\text{med}} \in \mathcal{B}(r_X, 0, c)$. We do this by setting $g_1 = \sigma_{W_1,Y} - \beta_{\text{med}}\sigma_{W_1,X}$ and then using the characterization of $\mathcal{B}(r_X, r_Y, c)$ given in Lemma 4.

   By assumption, equations (26) and (28) hold, as well as the first displayed equation in Lemma 4, $(I + cr_X')p_1 = \sigma_{W_1,X}$. The second displayed equation also holds since our specific choice of g_1 combined with $r_Y = 0$ gives $(I + cr_Y')g_1 = g_1 = \sigma_{W_1,Y} - \beta_{\text{med}}\sigma_{W_1,X}$. Equation (25) holds immediately from setting $r_Y = 0$ and $b = \beta_{\text{med}}$.

All that remains is to verify inequality (27). Substitute $b = \beta_{\text{med}}$ and $r_Y = 0$ into this inequality to get

$$0 < k_0 \left(\frac{k_2}{k_0} - \beta_{\text{med}}^2 \right)$$

or, equivalently,

$$0 < \frac{k_2 k_0 - k_1^2}{k_0}.$$

This inequality holds by $k_0 > 0$ and by Lemma 2. Thus we have verified that all of the conditions in Lemma 4 hold. Hence the lemma shows that $\beta_{\text{med}} \in \mathcal{B}(r_X, 0, c)$.

2. Next we show that $\{\beta_{\text{med}}\} \supseteq \mathcal{B}(r_X, 0, c)$. Let $b \in \mathcal{B}(r_X, 0, c)$. Then equation (25) holds with $r_Y = 0$. That is, $0 = k_1 - bk_0$. Hence $b = \beta_{\text{med}}$.

Putting these two steps together gives $\{\beta_{\text{med}}\} = \mathcal{B}(r_X, 0, c)$.

We conclude by proving the final sentence of the lemma. Suppose (r_X, c) are such that $r_X' c \neq -1$ and $z_X(r_X, c)^2 < k_0$. Then, since $I + r_X c'$ is invertible when $r_X' c \neq -1$ (Abadir and Magnus 2005, exercise 4.28 with $A = I$), we can let $p_1 = (I + c r_X')^{-1} \sigma_{W_1,X}$. Then $r_X' \sigma_{W_1,X} = r_X'(I + c r_X') p_1 = r_X' p_1 (1 + c' r_X)$, and therefore $r_X' p_1 = \frac{r_X' \sigma_{W_1,X}}{1 + c' r_X}$. Thus

$$\begin{aligned}(r_X' p_1)^2 (1 - \|c\|^2) &= \frac{(r_X' \sigma_{W_1,X})^2 (1 - \|c\|^2)}{(1 + c' r_X)^2} \\ &= z_X(r_X, c)^2 \\ &< k_0.\end{aligned}$$

Hence equation (26) holds. By the first part of the proof we therefore have $\mathcal{B}(r_X, 0, c) = \{\beta_{\text{med}}\}$. □

Let

$$A(\bar{r}_X, \bar{c}) = \{(r_X, c, b, z) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_1} \times \mathbb{R} \times \mathbb{R} \setminus \{0\} : \|r_X\| \leq \bar{r}_X, \|c\| \leq \bar{c}, (30)\text{–}(33) \text{ hold}\}.$$

This is the set of (r_X, c, b, z) that satisfies four of the five equations in the characterization of $\mathcal{B}(r_X, r_Y, c)$ given in Lemma 5. Our next lemma provides an analytical characterization of the smallest value of $\|r_Y\|$ such that the fifth and final equation (numbered (29)) in the characterization of Lemma 5 holds for that r_Y. Recall that the function $\underline{r}_Y(z, c, b)$ is defined just prior to the statement of Theorem 4 in section 3.2.

Lemma 7. Let $\text{var}(Y, X, W_1)$ be positive definite. Fix $(r_X, c, b, z) \in A(\bar{r}_X, \bar{c})$. Then

$$\inf \left\{ \|r_Y\| : r_Y \in \mathbb{R}^{d_1}, k_1 - bk_0 = r_Y' \left(z\sqrt{1 - \|c\|^2} (\sigma_{W_1,Y} - b\sigma_{W_1,X}) - (k_1 - bk_0)c \right) \right\} = \underline{r}_Y(z, c, b).$$

Proof of Lemma 7. First we recall a few equations and definitions. The fifth equation in the characterization of Lemma 5, equation (29), is

$$k_1 - bk_0 = r_Y' \left(z\sqrt{1 - \|c\|^2} (\sigma_{W_1,Y} - b\sigma_{W_1,X}) - c(k_1 - bk_0) \right). \tag{29}$$

We then have the sets

$$\mathcal{D} = \mathbb{R} \times \{c \in \mathbb{R}^{d_1} : \|c\| < 1\} \times \mathbb{R}$$

and

$$\mathcal{D}^0 = \{(z, c, b) \in \mathcal{D} : z\sqrt{1 - \|c\|^2} (\sigma_{W_1,Y} - b\sigma_{W_1,X}) - (k_1 - k_0 b)c \neq 0\}.$$

The set $\mathcal{D}$ is the set of feasible values of (z, c, b), while $\mathcal{D}^0$ is the subset of (z, c, b) values such that the coefficient on r_Y in equation (29) is nonzero.

We compute the infimum in the lemma by considering a three different cases. The proof concludes by combining all three of these cases into the definition of $\underline{r}_Y(z, c, b)$.

Case 1. We show that $\inf\left\{\|r_Y\| : r_Y \in \mathbb{R}^{d_1}, (29) \text{ holds for } (r_Y, c, b, z)\right\} = 0$ if $b = \beta_{\text{med}}$.

When $b = \beta_{\text{med}}$, equation (29) is

$$0 = r_Y' \left(z\sqrt{1-\|c\|^2}(\sigma_{W_1,Y} - \beta_{\text{med}}\sigma_{W_1,X}) - c(k_1 - \beta_{\text{med}} k_0)\right).$$

For all (z, c), this equation holds for $r_Y = 0$. Hence

$$\inf\{\|r_Y\| : b \in \mathcal{B}(r_X, r_Y, c), r_Y \in \mathbb{R}^{d_1}\} = 0.$$

Case 2. We show that $\inf\left\{\|r_Y\| : r_Y \in \mathbb{R}^{d_1}, (29) \text{ holds for } (r_Y, c, b, z)\right\} = +\infty$ when $b \neq \beta_{\text{med}}$ and $(z, c, b) \in \mathcal{D} \setminus \mathcal{D}^0$.

Suppose (z, c, b) are such that $b \neq \beta_{\text{med}}$. Since $(z, c, b) \notin \mathcal{D}^0$, the coefficient on r_Y in equation (29) must be zero for these values:

$$z\sqrt{1-\|c\|^2}(\sigma_{W_1,Y} - b\sigma_{W_1,X}) - c(k_1 - bk_0) = 0.$$

But since $b \neq \beta_{\text{med}}$, the left hand side of equation (29) is nonzero. Hence there does not exist an r_Y such that (29) holds.

Case 3. We show that

$$\inf\left\{\|r_Y\| : r_Y \in \mathbb{R}^{d_1}, (29) \text{ holds for } (r_Y, c, b, z)\right\} = \frac{|k_1 - k_0 b|}{\|z\sqrt{1-\|c\|^2}(\sigma_{W_1,Y} - b\sigma_{W_1,X}) + (k_1 - k_0 b)c\|}$$

if $b \neq \beta_{\text{med}}$ and $(z, c, b) \in \mathcal{D}^0$.

Applying the Cauchy Schwarz inequality to equation (29), we find that for any r_Y such that $b \in \mathcal{B}(r_X, r_Y, c) \setminus \{\beta_{\text{med}}\}$,

$$\begin{aligned} |k_1 - bk_0| &= \left| r_Y' \left(z\sqrt{1-\|c\|^2}(\sigma_{W_1,Y} - b\sigma_{W_1,X}) - c(k_1 - bk_0)\right)\right| \\ &\leq \|r_Y\| \left\| z\sqrt{1-\|c\|^2}(\sigma_{W_1,Y} - b\sigma_{W_1,X}) - c(k_1 - bk_0)\right\|. \end{aligned}$$

Rearranging this inequality gives

$$\|r_Y\| \geq \frac{|k_1 - k_0 b|}{\|z\sqrt{1-\|c\|^2}(\sigma_{W_1,Y} - b\sigma_{W_1,X}) + (k_1 - k_0 b)c\|},$$

where the denominator is non-zero since $(z, c, b) \in \mathcal{D}^0$. This lower bound is attained by letting $r_Y = r_Y^*(z, c, b)$ where

$$r_Y^*(z, c, b) = (k_1 - bk_0)\frac{z\sqrt{1-\|c\|^2}(\sigma_{W_1,Y} - b\sigma_{W_1,X}) - (k_1 - bk_0)c}{\left\| z\sqrt{1-\|c\|^2}(\sigma_{W_1,Y} - b\sigma_{W_1,X}) - (k_1 - bk_0)c\right\|^2}.$$

By construction, equation (29) is satisfied with this choice of r_Y, since

$$r_Y^*(z, c, b)'(z\sqrt{1-\|c\|^2}(\sigma_{W_1,Y} - b\sigma_{W_1,X}) - (k_1 - k_0 b)c) = k_1 - bk_0.$$

Since it attains the lower bound,

$$\inf\left\{\|r_Y\| : r_Y \in \mathbb{R}^{d_1}, (29) \text{ holds for } (r_Y, z, c, b)\right\} = \|r_Y^*(z, c, b)\|$$

Finally, note that $\|r_Y^*(z, c, b)\| = \underline{r}_Y(z, c, b)$. $\square$

To the state the next lemma, define the following bounds for b:

$$\underline{b}(z) = \beta_{\text{med}} - \sqrt{\text{devsq}(z)} \qquad \text{and} \qquad \overline{b}(z) = \beta_{\text{med}} + \sqrt{\text{devsq}(z)}.$$

Our next lemma characterizes the union of $\mathcal{B}(r_X, r_Y, c)$ over all $r_Y \in \mathbb{R}^{d_1}$. In particular, it shows that this union is essentially an interval. This lemma is not quite our main result, however, because it is stated for r_X and c, rather than for $\bar{r}_X$ and $\bar{c}$. Nonetheless, our main results will show how to use this lemma to obtain a characterization of $\mathcal{B}_I(\bar{r}_X, \bar{c})$.

Lemma 8. Let $\text{var}(Y, X, W_1)$ be positive definite. Let $\sigma_{W_1,Y} \neq \sigma_{X,Y}\sigma_{W_1,X}$ and $\|c\| < 1$. For any (r_X, c) such that $r_X' c \neq -1$ and $0 < z_X(r_X, c)^2 < k_0$, we have

$$\bigcup_{r_Y \in \mathbb{R}^{d_1}} \mathcal{B}(r_X, r_Y, c) = \big(\underline{b}(z_X(r_X, c)), \overline{b}(z_X(r_X, c))\big) \setminus \mathcal{B}^0(z_X(r_X, c), c)$$

where $\mathcal{B}^0(z_X(r_X, c), c)$ is a set containing at most one point.

Proof of Lemma 8. Recall that by Lemma 5, for $b \neq \beta_{\text{med}}$, $b \in \mathcal{B}(r_X, r_Y, c)$ if and only if (29)–(33) hold for some $z \in \mathbb{R} \setminus \{0\}$. We'll show two characterizations:

1. First, given $r_X' c \neq -1$, $z_X(r_X, c)^2 < k_0$, and $\|c\| < 1$, (30) and (31) hold if and only if $z = z_X(r_X, c)$. To see this, when $r_X' c \neq -1$, (30) is equivalent to $z_X(r_X, c) = z$ because

   $$z = r_X'(\sigma_{W_1,X}\sqrt{1 - \|c\|^2} - zc) \qquad \text{holds if and only if} \qquad z = \frac{r_X' \sigma_{W_1,X}\sqrt{1 - \|c\|^2}}{1 + r_X' c}$$

   and the far right hand side expression is precisely the definition of $z_X(r_X, c)$. Furthermore, setting $z = z_X(r_X, c)$, $z_X(r_X, c)^2 < k_0$ if and only if (31) holds. Hence, under our maintained assumptions, (30) and (31) hold if and only if $z = z_X(r_X, c)$.

2. Next, assuming $z = z_X(r_X, c)$, (32) can be written as

   $$(b - \beta_{\text{med}})^2 < \text{devsq}(z_X(r_X, c)).$$

   This is equivalent to

   $$\beta_{\text{med}} - \sqrt{\text{devsq}(z_X(r_X, c))} < b < \beta_{\text{med}} + \sqrt{\text{devsq}(z_X(r_X, c))}$$

   or

   $$b \in \big(\underline{b}(z_X(r_X, c)), \overline{b}(z_X(r_X, c))\big).$$

These two equivalences show that for any $b \in \big(\underline{b}(z_X(r_X, c)), \overline{b}(z_X(r_X, c))\big) \setminus \{\beta_{\text{med}}\}$, $(r_X, c, b, z_X(r_X, c))$ satisfy (30), (31), and (32). Therefore,

$$\begin{aligned}
&\bigcup_{r_Y \in \mathbb{R}^{d_1}} \mathcal{B}(r_X, r_Y, c) \\
&= \mathcal{B}(r_X, 0, c) \cup \left(\bigcup_{r_Y \in \mathbb{R}^{d_1}} \mathcal{B}(r_X, r_Y, c) \right) \\
&= \{\beta_{\text{med}}\} \cup \left(\bigcup_{r_Y \in \mathbb{R}^{d_1}} \mathcal{B}(r_X, r_Y, c) \setminus \{\beta_{\text{med}}\} \right) \\
&= \{\beta_{\text{med}}\} \cup \left(\bigcup_{r_Y \in \mathbb{R}^{d_1}} \{b \in \mathbb{R} \setminus \{\beta_{\text{med}}\} : \text{(29)–(33) hold for some } z \in \mathbb{R} \setminus \{0\}\} \right)
\end{aligned}$$

$$
\begin{aligned}
&= \{\beta_{\text{med}}\} \cup \left(\bigcup_{r_Y \in \mathbb{R}^{d_1}} \left\{ b \in \mathbb{R} \setminus \{\beta_{\text{med}}\} : (29) \text{ holds for } z = z_X(r_X,c) \text{ and } b \in \big(\underline{b}(z_X(r_X,c)), \bar{b}(z_X(r_X,c))\big) \right\} \right) \\
&= \{\beta_{\text{med}}\} \cup \left(\bigcup_{r_Y \in \mathbb{R}^{d_1}} \left\{ b \in \big(\underline{b}(z_X(r_X,c)), \bar{b}(z_X(r_X,c))\big) \setminus \{\beta_{\text{med}}\} : \right.\right. \\
&\qquad\qquad \left.\left. k_1 - bk_0 = r_Y' \left(z_X(r_X,c)\sqrt{1-\|c\|^2}(\sigma_{W_1,Y} - b\sigma_{W_1,X}) - c(k_1 - bk_0) \right) \right\} \right) \\
&= \big(\underline{b}(z_X(r_X,c)), \bar{b}(z_X(r_X,c))\big) \setminus \\
&\quad \left\{ b \in \mathbb{R} \setminus \{\beta_{\text{med}}\} : k_1 - bk_0 \neq r_Y' \left(z_X(r_X,c)\sqrt{1-\|c\|^2}(\sigma_{W_1,Y} - b\sigma_{W_1,X}) - c(k_1 - bk_0) \right) \text{ for all } r_Y \in \mathbb{R}^{d_1} \right\} \\
&= \left\{ b \in \big(\underline{b}(z_X(r_X,c)), \bar{b}(z_X(r_X,c))\big) : (z_X(r_X,c), c, b) \in \mathcal{D}^0 \right\} \\
&= \big(\underline{b}(z_X(r_X,c)), \bar{b}(z_X(r_X,c))\big) \setminus \{ b \in \mathbb{R} : (z_X(r_X,c), c, b) \in \mathcal{D} \setminus \mathcal{D}^0 \} \\
&\equiv \big(\underline{b}(z_X(r_X,c)), \bar{b}(z_X(r_X,c))\big) \setminus \mathcal{B}^0(z_X(r_X,c), c).
\end{aligned}
$$

The second equality follows by Lemma 6, which gives $\{\beta_{\text{med}}\} = \mathcal{B}(r_X, 0, c)$ since $r_X' c \neq -1$ and $z_X(r_X,c)^2 < k_0$. The third equality follows by Lemma 5. The fourth equality follows by our derivations above. The fifth equality follows by substituting in the definition of equation (29). In the sixth equality we used the fact that $z_X(r_X,c) > 0$ and so $\big(\underline{b}(z_X(r_X,c)), \bar{b}(z_X(r_X,c))\big)$ is a nonempty interval, which contains β_{med} by definition. Moreover, $\beta_{\text{med}} \notin \mathcal{B}^0(z_X(r_X,c),c)$ for all (r_X,c) because there always exists an $r_Y \in \mathbb{R}^{d_1}$ such that equation (29) holds at $b = \beta_{\text{med}}$. Thus β_{med} is never removed when we subtract the set $\mathcal{B}^0(z_X(r_X,c),c)$. The seventh equality follows since, as we showed in the proof of Lemma 7, for any (r_X,c,b,z) such that $b \neq \beta_{\text{med}}$ and such that (30), (31), (32), and (33) hold,

$$\{r_Y \in \mathbb{R}^{d_1} : (29) \text{ holds for } (r_Y, c, b, z)\} \neq \emptyset$$

if and only if $(z,c,b) \in \mathcal{D}^0$. In the last line we defined

$$\mathcal{B}^0(z_X(r_X,c),c) = \{b \in \mathbb{R} : (z_X(r_X,c), c, b) \in \mathcal{D} \setminus \mathcal{D}^0\}.$$

We complete the proof by showing that this set is either a singleton or empty.

Recall that $(z,c,b) \in \mathcal{D} \setminus \mathcal{D}^0$ if and only if

$$
\begin{aligned}
0 &= z\sqrt{1-\|c\|^2}(\sigma_{W_1,Y} - b\sigma_{W_1,X}) - (k_1 - bk_0)c \\
&= \left(z\sqrt{1-\|c\|^2}\sigma_{W_1,Y} - k_1 c \right) - b\left(z\sqrt{1-\|c\|^2}\sigma_{W_1,X} - k_0 c \right) \\
&\equiv A_0(z,c) - bA_1(z,c).
\end{aligned}
$$

For a given (z,c) this is system of linear equations in b. So it has at most one solution unless the vectors $A_0(z,c)$ and $A_1(z,c)$ satisfy $A_0(z,c) = A_1(z,c) = 0$. This happens if and only if

$$
\begin{aligned}
z\sqrt{1-\|c\|^2}\sigma_{W_1,Y} &= k_1 c \\
z\sqrt{1-\|c\|^2}\sigma_{W_1,X} &= k_0 c.
\end{aligned}
$$

Since $z \neq 0$ in this set and $\|c\| < 1$, we have $z\sqrt{1-\|c\|^2} \neq 0$. So $A_0(z,c) = A_1(z,c) = 0$ if and only if $\sigma_{W_1,Y} = (k_1/k_0)\sigma_{W_1,X}$. This equation is equivalent to

$$
\begin{aligned}
0 &= k_0\sigma_{W_1,Y} - k_1\sigma_{W_1,X} \\
&= (1 - \|\sigma_{W_1,X}\|^2)\sigma_{W_1,Y} - (\sigma_{X,Y} - \sigma_{W_1,X}'\sigma_{W_1,Y})\sigma_{W_1,X} \\
&= \sigma_{W_1,Y} - \|\sigma_{W_1,X}\|^2\sigma_{W_1,Y} - \sigma_{X,Y}\sigma_{W_1,X} + (\sigma_{W_1,Y}'\sigma_{W_1,X})\sigma_{W_1,X} \\
&= \sigma_{W_1,Y} - \|\sigma_{W_1,X}\|^2(k_1/k_0)\sigma_{W_1,X} - \sigma_{X,Y}\sigma_{W_1,X} + \left(((k_1/k_0)\sigma_{W_1,X})'\sigma_{W_1,X}\right)\sigma_{W_1,X}
\end{aligned}
$$

$$= \sigma_{W_1,Y} - \sigma_{X,Y}\sigma_{W_1,X}.$$

But now we see that this equation does not hold by assumption. Consequently we know that $A_0(z,c)$ and $A_1(z,c)$ are not both zero. This shows that there is at most one value b such that $(z_X(r_X,c),c,b) \in \mathcal{D}\backslash\mathcal{D}^0$. $\square$

Our next lemma shows that $\bar{z}_X(\cdot)$ is a bound for $z_X(\cdot)$.

Lemma 9. Suppose $\sigma_{W_1,X} \neq 0$. Then

$$\sup\{z_X(r_X,c)^2 : \|r_X\| \leq \bar{r}_X, \|c\| \leq \bar{c}, \|c\| \neq 1\} = \bar{z}_X(\bar{r}_X,\bar{c})^2.$$

Proof of Lemma 9. Recall that, by definition,

$$|z_X(r_X,c)| = \frac{|r_X'\sigma_{W_1,X}|\sqrt{1-\|c\|^2}}{|1+r_X'c|}.$$

We prove the result by considering two cases.

Case 1: Suppose $\bar{c}\bar{r}_X \geq 1$. In this case, by definition, $\bar{z}_X(\bar{r}_X,\bar{c})^2 = +\infty$. It therefore suffices to show that there exists a sequence $\{(r_X^{(n)},c^{(n)}) : \|r_X^{(n)}\| \leq \bar{r}_X, \|c^{(n)}\| \leq \bar{c}, \|c^{(n)}\| \neq 1\}$ such that $z_X(r_X,c)^2$ diverges to $+\infty$ along this sequence. For $n \geq 1$, define the sequence

$$r_X^{(n)} = -\frac{n-1}{n}\frac{1}{\bar{c}\|\sigma_{W_1,X}\|}\sigma_{W_1,X} \qquad \text{and} \qquad c^{(n)} = \frac{n-1}{n}\frac{\bar{c}}{\|\sigma_{W_1,X}\|}\sigma_{W_1,X}.$$

First we check that this sequence satisfies the norm constraints:

$$\|r_X^{(n)}\| = \left\|-\frac{n-1}{n}\frac{1}{\bar{c}\|\sigma_{W_1,X}\|}\sigma_{W_1,X}\right\| = \frac{n-1}{n\bar{c}} \leq \frac{(n-1)\bar{r}_X}{n} < \bar{r}_X$$

since $\bar{c}\bar{r}_X \geq 1$, and

$$\|c^{(n)}\| = \left\|\frac{n-1}{n}\frac{\bar{c}}{\|\sigma_{W_1,X}\|}\sigma_{W_1,X}\right\| = \frac{n-1}{n}\cdot\bar{c} < \bar{c}.$$

In particular, note that $\|c^{(n)}\| \neq 1$. Next we evaluate the function $|z_X(\cdot,\cdot)|$ along this sequence:

$$\begin{aligned}
|z_X(r_X^{(n)},c^{(n)})| &= \frac{|\frac{n-1}{n}\frac{1}{\bar{c}\|\sigma_{W_1,X}\|}\sigma_{W_1,X}'\sigma_{W_1,X}|\sqrt{1-\|c^{(n)}\|^2}}{\left|1-\left(\frac{n-1}{n}\right)^2\frac{1}{\bar{c}\|\sigma_{W_1,X}\|}\sigma_{W_1,X}'\frac{\bar{c}}{\|\sigma_{W_1,X}\|}\sigma_{W_1,X}\right|} \\
&= \frac{\frac{n-1}{n}\frac{\|\sigma_{W_1,X}\|}{\bar{c}}\sqrt{1-\bar{c}^2\left(\frac{n-1}{n}\right)^2}}{1-\left(\frac{n-1}{n}\right)^2} \\
&= \frac{\|\sigma_{W_1,X}\|}{\bar{c}}\sqrt{\left(\frac{n(n-1)}{2n-1}\right)^2\left(1-\bar{c}^2\left(\frac{n-1}{n}\right)^2\right)} \\
&= \frac{\|\sigma_{W_1,X}\|}{\bar{c}}\sqrt{\frac{(n-1)^2(n^2-\bar{c}^2(n-1)^2)}{(2n-1)^2}}.
\end{aligned}$$

Since $\|\sigma_{W_1,X}\| \neq 0$ by $\sigma_{W_1,X} \neq 0$, this expression is nonzero for all n. The term inside the square root goes to $+\infty$ by L'Hospital's rule. Thus

$$\lim_{n\to\infty}|z_X(r_X^{(n)},c^{(n)})| = +\infty.$$

Case 2: Suppose $\bar{c}\bar{r}_X < 1$. Consider (r_X,c) such that $\|r_X\| \leq \bar{r}_X$ and $\|c\| \leq \bar{c}$. By the Cauchy Schwarz inequality, $|r_X' \sigma_{W_1,X}| \leq \|r_X\| \|\sigma_{W_1,X}\|$ and

$$r_X' c \in \Big[- \|r_X\|\|c\|, \|r_X\|\|c\| \Big] \subseteq (0,1).$$

The inclusion in $(0,1)$ follows since $\bar{c}\bar{r}_X < 1$. This implies

$$|1 + r_X' c| = 1 + r_X' c \geq 1 - \|r_X\|\|c\| > 0.$$

Therefore

$$\begin{aligned} |z_X(r_X, c)| &= \frac{|r_X' \sigma_{W_1,X}| \sqrt{1 - \|c\|^2}}{|1 + r_X' c|} \\ &\leq \frac{\|r_X\| \|\sigma_{W_1,X}\| \sqrt{1 - \|c\|^2}}{1 - \|r_X\|\|c\|} \\ &\equiv f(\|r_X\|, \|c\|). \end{aligned}$$

This implies that

$$\begin{aligned} \sup\{|z_X(r_X, c)| : \|r_X\| \leq \bar{r}_X, \|c\| \leq \bar{c}\} &\leq \sup\{f(\|r_X\|, \|c\|) : \|r_X\| \leq \bar{r}_X, \|c\| \leq \bar{c}\} \\ &= \sup\{f(a_X, a_c) : (a_X, a_c) \in [0, \bar{r}_X] \times [0, \bar{c}]\}. \end{aligned}$$

For now we omit the $\|c\| \neq 1$ constraint; below we show that it will not bind at the supremum. Next we show that

$$\sup\{f(a_X, a_c) : (a_X, a_c) \in [0, \bar{r}_X] \times [0, \bar{c}]\} = \bar{z}(\bar{r}_X, \bar{c}).$$

To see this, first note that $f(a_X, a_c)$ is continuous on $[0, \bar{r}_X] \times [0, \bar{c}]$, a compact set, by $\bar{r}_X \bar{c} < 1$. $f(a_X, a_c)$ is nondecreasing in a_X for any $a_c \in [0, \bar{c}]$. Hence $f(a_X, a_c) \leq f(\bar{r}_X, a_c)$ for any $(a_X, a_c) \in [0, \bar{r}_X] \times [0, \bar{c}]$.

To maximize $f(\bar{r}_X, a_c)$ over $a_c \in [0, \bar{c}]$, consider the derivative of $f(\bar{r}_X, a_c)$ with respect to a_c:

$$\begin{aligned} \frac{\partial f(\bar{r}_X, a_c)}{\partial a_c} &= \bar{r}_X \|\sigma_{W_1,X}\| \frac{-a_c(1 - \bar{r}_X a_c) + \bar{r}_X(1 - a_c^2)}{\sqrt{1 - a_c^2}(1 - \bar{r}_X a_c)^2} \\ &= \frac{\bar{r}_X \|\sigma_{W_1,X}\| (\bar{r}_X - a_c)}{\sqrt{1 - a_c^2}(1 - \bar{r}_X a_c)^2}. \end{aligned}$$

Suppose $\bar{r}_X > 0$. Then this derivative is positive whenever $a_c < \bar{r}_X$, zero when $a_c = \bar{r}_X$, and negative when $a_c > \bar{r}_X$. So if $\bar{r}_X \leq \bar{c}$, $a_c = \bar{r}_X$ maximizes this function. Otherwise, $f(\bar{r}_X, \bar{c})$ is maximized at the boundary value $a_c = \bar{c}$. Hence $\operatorname{argmax}\{f(\bar{r}_X, a_c) : a_c \in [0, \bar{c}]\} = \min\{\bar{r}_X, \bar{c}\}$. If $\bar{r}_X = 0$, $f(\bar{r}_X, a_c)$ is constant and is trivially maximized at $a_c = \min\{\bar{r}_X, \bar{c}\}$.

Putting the previous few steps together gives

$$\begin{aligned} \sup\{|z_X(r_X, c)| : \|r_X\| \leq \bar{r}_X, \|c\| \leq \bar{c}\} &\leq \sup\{f(a_X, a_c) : (a_X, a_c) \in [0, \bar{r}_X] \times [0, \bar{c}]\} \\ &= f(\bar{r}_X, \min\{\bar{r}_X, \bar{c}\}) \\ &= \frac{\bar{r}_X \|\sigma_{W_1,X}\| \sqrt{1 - \min\{\bar{r}_X, \bar{c}\}^2}}{1 - \bar{r}_X \min\{\bar{r}_X, \bar{c}\}} \\ &= \bar{z}_X(\bar{r}_X, \bar{c}). \end{aligned}$$

The last line follows by the definition of $\bar{z}_X$.

We conclude the proof by showing that the inequality in fact holds with equality:

$$\sup\{|z_X(r_X, c)| : \|r_X\| \leq \bar{r}_X, \|c\| \leq \bar{c}\} = \bar{z}_X(\bar{r}_X, \bar{c}).$$

We do this by finding (r_X^*, c^*) such that $\|r_X^*\| \leq \bar{r}_X$, $\|c^*\| \leq \bar{c}$, and such that $|z_X(r_X^*, c^*)| = \bar{z}_X(\bar{r}_X, \bar{c})$.

Specifically, let

$$r_X^* = \frac{\bar{r}_X}{\|\sigma_{W_1,X}\|}\sigma_{W_1,X} \qquad \text{and} \qquad c^* = -\frac{\min\{\bar{r}_X, \bar{c}\}}{\|\sigma_{W_1,X}\|}\sigma_{W_1,X}.$$

First note that $\|r_X^*\| = \bar{r}_X \leq \bar{r}_X$ and $\|c^*\| = \min\{\bar{r}_X, \bar{c}\} \leq \bar{c}$. Moreover,

$$\begin{aligned}
|z_X(r_X^*, c^*)| &= \frac{|r_X^{*\prime}\sigma_{W_1,X}|\sqrt{1 - \|c^*\|^2}}{|1 + r_X^{*\prime}c^*|} \\
&= \frac{\left|\frac{\bar{r}_X}{\|\sigma_{W_1,X}\|}\|\sigma_{W_1,X}\|^2\right|\sqrt{1 - \min\{\bar{r}_X, \bar{c}\}^2}}{\left|1 - \frac{\bar{r}_X \min\{\bar{r}_X, \bar{c}\}\|\sigma_{W_1,X}\|^2}{\|\sigma_{W_1,X}\|^2}\right|} \\
&= \frac{\bar{r}_X\|\sigma_{W_1,X}\|\sqrt{1 - \min\{\bar{r}_X, \bar{c}\}^2}}{1 - \bar{r}_X \min\{\bar{r}_X, \bar{c}\}} \\
&= \bar{z}_X(\bar{r}_X, \bar{c}).
\end{aligned}$$

Finally, note that $\min\{\bar{r}_X, \bar{c}\} < 1$ since $\bar{r}_X\bar{c} < 1$. Hence $\|c^*\| \neq 1$. This shows that adding the constraint $\|c\| \neq 1$ does not change the optimizer. $\square$

We use the next two lemmas to prove our third main result, Theorem 4. The first provides a simple criterion for showing that (30) holds for some $z \in \mathbb{R}$. Recall that this is one of the five equations in the characterization of when some $b \neq \beta_{\text{med}}$ is in $\mathcal{B}(r_X, r_Y, c)$ given in Lemma 5. Recall that the function $p(z, c; \bar{r}_X)$ is defined just prior to the statement of Theorem 4 in section 3.2.

Lemma 10. Assume $\sigma_{W_1,X} \neq 0$. For any $(z, c, \bar{r}_X)$ with $z^2 < k_0$, $\|c\| < 1$, $\bar{r}_X \geq 0$, there exists an $r_X \in \mathbb{R}^{d_1}$ with $\|r_X\| \leq \bar{r}_X$ such that

$$z = r_X'(\sigma_{W_1,X}\sqrt{1 - \|c\|^2} - cz) \tag{30}$$

if and only if $p(z, c; \bar{r}_X) \geq 0$.

Proof of Lemma 10. Let $(z, c, \bar{r}_X) \in \mathbb{R} \times \mathbb{R}^{d_1} \times \mathbb{R}_{\geq 0}$ be such that $\|c\| < 1$ and $z^2 < k_0$.

($\Rightarrow$) Suppose there exists an r_X such that $\|r_X\| \leq \bar{r}_X$ and $z = r_X'(\sigma_{W_1,X}\sqrt{1 - \|c\|^2} - cz)$. Then

$$\begin{aligned}
p(z, c; \bar{r}_X) &= \bar{r}_X^2\|\sigma_{W_1,X}\sqrt{1 - \|c\|^2} - zc\|^2 - z^2 \\
&\geq \|r_X\|^2\|\sigma_{W_1,X}\sqrt{1 - \|c\|^2} - zc\|^2 - z^2 \\
&\geq \left(r_X'(\sigma_{W_1,X}\sqrt{1 - \|c\|^2} - cz)\right)^2 - z^2 \\
&= z^2 - z^2 \\
&= 0.
\end{aligned}$$

The first line follows by definition of p. The second line follows from $\|r_X\| \leq \bar{r}_X$. The third line follows by the Cauchy Schwarz inequality. The fourth line follows by assumption.

($\Leftarrow$) Suppose $p(z, c; \bar{r}_X) \geq 0$. We will construct an $r_X \in \mathbb{R}^{d_1}$ with $\|r_X\| \leq \bar{r}_X$ and such that equation (30) holds. We consider two cases.

1. Suppose $z = 0$. Let $r_X = 0$ Then, $\|r_X\| = 0 \leq \bar{r}_X$ and

   $$z = r_X'(\sigma_{W_1,X}\sqrt{1 - \|c\|^2} - cz)$$

   holds since both sides equal zero.

2. Suppose $z \neq 0$. Let

   $$r_X = \frac{z}{\|\sigma_{W_1,X}\sqrt{1 - \|c\|^2} - zc\|^2}(\sigma_{W_1,X}\sqrt{1 - \|c\|^2} - zc).$$

r_X is well defined since

$$\|\sigma_{W_1,X}\sqrt{1-\|c\|^2} - zc\|^2 = \frac{p(z,c;\bar{r}_X)}{\bar{r}_X^2} + \frac{z^2}{\bar{r}_X^2} > 0$$

by $z \neq 0$ and $p(z,c;\bar{r}_X) \geq 0$. Also note that $p(z,c;\bar{r}_X) \geq 0$ and $z \neq 0$ together imply that $\bar{r}_X > 0$. Next, $\|r_X\| \leq \bar{r}_X$ follows from

$$\begin{aligned}
\|r_X\|^2 - \bar{r}_X^2 &= \frac{z^2}{\|\sigma_{W_1,X}\sqrt{1-\|c\|^2} - zc\|^2} - \bar{r}_X^2 \\
&= \|\sigma_{W_1,X}\sqrt{1-\|c\|^2} - zc\|^{-2}\left(z^2 - \bar{r}_X^2\|\sigma_{W_1,X}\sqrt{1-\|c\|^2} - zc\|^2\right) \\
&= -\|\sigma_{W_1,X}\sqrt{1-\|c\|^2} - zc\|^{-2} p(z,c;\bar{r}_X) \\
&\leq 0.
\end{aligned}$$

Finally, we can verify that

$$\begin{aligned}
&r_X'(\sigma_{W_1,X}\sqrt{1-\|c\|^2} - cz) \\
&= \frac{z}{\|\sigma_{W_1,X}\sqrt{1-\|c\|^2} - zc\|^2}(\sigma_{W_1,X}\sqrt{1-\|c\|^2} - zc)'(\sigma_{W_1,X}\sqrt{1-\|c\|^2} - zc) \\
&= z.
\end{aligned}$$

□

The next lemma allows us to search over $c \in \text{span}\{\sigma_{W_1,X}, \sigma_{W_1,Y}\}$ rather than $\mathbb{R}^{d_1}$, reducing the dimension of the minimization problem by $d_1 - 2$.

Lemma 11. Suppose $(\sigma_{W_1,X}, \sigma_{W_1,X})$ are linearly independent and $d_1 \geq 2$. Let $b \neq \beta_{\text{med}}$, $z \in \mathbb{R} \setminus \{0\}$, $\bar{r}_X \in [0,\infty)$, and $\bar{c} \in [0,1)$. Then

$$\begin{aligned}
&\inf\{\underline{r}_Y(z,c,b) : p(z,c;\bar{r}_X) \geq 0, c \in \mathbb{R}^{d_1}, \|c\| \leq \bar{c}\} \\
&= \inf\{\underline{r}_Y(z,c,b) : p(z,c;\bar{r}_X) \geq 0, c \in \text{span}\{\sigma_{W_1,X}, \sigma_{W_1,Y}\}, \|c\| \leq \bar{c}\}.
\end{aligned}$$

Proof of Lemma 11. If $d_1 = 2$ the lemma follows immediately because $\mathbb{R}^{d_1} = \text{span}\{\sigma_{W_1,X}, \sigma_{W_1,Y}\}$. So suppose $d_1 > 2$. We break this proof in two parts.

Part 1: We first show that for any $a \in [0, \bar{c}]$,

$$\text{argmin}\{\underline{r}_Y(z,c,b) : p(z,c;\bar{r}_X) \geq 0, c \in \mathbb{R}^{d_1}, \|c\| = a\} \subseteq \text{span}\{\sigma_{W_1,X}, \sigma_{W_1,Y}\}.$$

We prove this by contrapositive: We show that for any $c_{\text{outside}} \notin \text{span}\{\sigma_{W_1,X}, \sigma_{W_1,Y}\}$, $c_{\text{outside}} \notin \text{argmin}\{\underline{r}_Y(z,c,b) : p(z,c;\bar{r}_X) \geq 0, c \in \mathbb{R}^{d_1}, \|c\| = a\}$. So let $c_{\text{outside}} \in \mathbb{R}^{d_1} \setminus \text{span}\{\sigma_{W_1,X}, \sigma_{W_1,Y}\}$. Note that such a point exists since $d_1 > 2 = \dim(\text{span}\{\sigma_{W_1,X}, \sigma_{W_1,Y}\})$. Let

$$B(a,z,\bar{r}_X) = \{c \in \mathbb{R}^{d_1} : p(z,c;\bar{r}_X) \geq 0, \|c\| = a\}$$

denote the constraint set. The result holds immediately if $c_{\text{outside}} \notin B(a,z,\bar{r}_X)$, so suppose $c_{\text{outside}} \in B(a,z,\bar{r}_X)$. It therefore suffices to construct a $c_{\text{better}} \in \text{span}\{\sigma_{W_1,X}, \sigma_{W_1,Y}\}$ with $c_{\text{better}} \in B(a,z,\bar{r}_X)$ and $\underline{r}_Y(z, c_{\text{better}}, b) < \underline{r}_Y(z, c_{\text{outside}}, b)$.

By Gram-Schmidt orthogonalization, the vectors

$$v_1 = \frac{\sigma_{W_1,X}}{\|\sigma_{W_1,X}\|} \qquad \text{and} \qquad v_2 = \frac{\|\sigma_{W_1,X}\|^2\sigma_{W_1,Y} - (\sigma_{W_1,Y}'\sigma_{W_1,X})\sigma_{W_1,X}}{\|\sigma_{W_1,X}\|\sqrt{\|\sigma_{W_1,Y}\|^2 - (\sigma_{W_1,X}'\sigma_{W_1,Y})^2}}$$

form an orthonormal basis for the subspace $\text{span}\{\sigma_{W_1,X}, \sigma_{W_1,Y}\}$, since we assumed $\sigma_{W_1,X}$ and $\sigma_{W_1,Y}$ are linearly independent. Since we can write $\mathbb{R}^{d_1} = \text{span}\{v_1, v_2\} \oplus \text{span}\{v_1, v_2\}^{\perp}$, c_{outside} can be uniquely represented as

$$c_{\text{outside}} = a_1 v_1 + a_2 v_2 + a_3 v_3$$

where (a_1, a_2, a_3) are such that $a_1^2 + a_2^2 + a_3^2 = a^2$, $v_3 \in \text{span}\{v_1, v_2\}^{\perp}$, and $\|v_3\| = 1$. Since $c_{\text{outside}} \notin \text{span}\{\sigma_{W_1,X}, \sigma_{W_1,Y}\}$, $a_3 \neq 0$.

Recall that our goal is to find an element c_{better} inside $\text{span}\{v_1, v_2\} = \text{span}\{\sigma_{W_1,X}, \sigma_{W_1,Y}\}$ that has norm equal to a and which leads to a smaller value of the objective function than c_{outside}. To this end, define

$$c^+ = a_1 v_1 + \left(\sqrt{a_2^2 + a_3^2}\right) v_2 \qquad \text{and} \qquad c^- = a_1 v_1 - \left(\sqrt{a_2^2 + a_3^2}\right) v_2.$$

These points satisfy the span and norm requirements: $c^+, c^- \in \text{span}\{\sigma_{W_1,X}, \sigma_{W_1,Y}\}$ and $\|c^+\| = \|c^-\| = \|c_{\text{outside}}\| = a$. In fact, we will show that one of c^+ or c^- is the desired point c_{better}. To do this, we have to show that (i) they are in the constraint set and (ii) at least one of them leads to a strictly smaller value of the objective function.

Step (i). Here we show that c^+ and c^- are in the constraint set $B(a, z, \bar{r}_X)$. We already showed that their norm equals a. So we only need to show that $p(z, c; \bar{r}_X) \geq 0$ for c equal to c^+ or c^-. To do this, write

$$\begin{aligned} p(z, c; \bar{r}_X) &= \bar{r}_X^2 \|\sigma_{W_1,X} \sqrt{1 - \|c\|^2} - cz\|^2 - z^2 \\ &= z^2(\|c\|^2 \bar{r}_X^2 - 1) - 2z\bar{r}_X^2(\sigma_{W_1,X}' c)\sqrt{1 - \|c\|^2} + \bar{r}_X^2 \|\sigma_{W_1,X}\|^2(1 - \|c\|^2). \end{aligned}$$

The second line follows by expanding the norm. Since $\|c^+\| = \|c^-\| = \|c_{\text{outside}}\| = a$, $v_1' \sigma_{W_1,X} = \|\sigma_{W_1,X}\|$, and $v_2' \sigma_{W_1,X} = v_3' \sigma_{W_1,X} = 0$, evaluating this function at any of the three points c_{outside}, c^+, or c^- gives the same value:

$$\begin{aligned} p(z, c_{\text{outside}}; \bar{r}_X) &= z^2(a^2 \bar{r}_X^2 - 1) - 2a_1 z \bar{r}_X^2 \|\sigma_{W_1,X}\| \sqrt{1 - a^2} + \bar{r}_X^2 \|\sigma_{W_1,X}\|^2 (1 - a^2) \\ &= p(z, c^+; \bar{r}_X) \\ &= p(z, c^-; \bar{r}_X). \end{aligned}$$

Since $c_{\text{outside}} \in B(a, z, \bar{r}_X)$, $p(z, c_{\text{outside}}; \bar{r}_X) \geq 0$. We therefore have $p(z, c^-; \bar{r}_X) \geq 0$ and $p(z, c^+; \bar{r}_X) \geq 0$. Thus $c^+, c^- \in B(a, z, \bar{r}_X)$.

Step (ii). Next we show that either $\underline{r}_Y(z, c^+, b) < \underline{r}_Y(z, c_{\text{outside}}, b)$ or $\underline{r}_Y(z, c^-, b) < \underline{r}_Y(z, c_{\text{outside}}, b)$.

First, we have

$$z\sqrt{1 - \|c_{\text{outside}}\|^2}(\sigma_{W_1,X} - b\sigma_{W_1,Y}) - (k_1 - k_0 b)c_{\text{outside}} \neq 0.$$

To see this, note that this quantity is a linear combination of $(\sigma_{W_1,X}, \sigma_{W_1,Y}, c_{\text{outside}})$. Its coefficients are nonzero because $b \neq \beta_{\text{med}}$, $z \neq 0$, and $\|c_{\text{outside}}\| = a \leq \bar{c} < 1$. Moreover, $\sigma_{W_1,X}$ and $\sigma_{W_1,Y}$ are linearly independent, and c_{outside} is not in $\text{span}\{\sigma_{W_1,X}, \sigma_{W_1,Y}\}$. This linear independence therefore implies that the linear combination cannot be zero. Thus $(z, c_{\text{outside}}, b) \in \mathcal{D}^0$ by definition of $\mathcal{D}^0$. By definition of $\underline{r}_Y(z, c, b)$, we therefore have $0 < \underline{r}_Y(z, c_{\text{outside}}, b) < \infty$. Consequently, $\underline{r}_Y(z, c_{\text{outside}}, b)^{-1}$ is well defined.

For any c with $\|c\| = a$, expanding the norm in $\underline{r}_Y(z, c, b)$ gives

$$\underline{r}_Y(z, c, b)^{-2} = \frac{z^2(1 - a^2)\|\sigma_{W_1,Y} - b\sigma_{W_1,X}\|^2 - 2z\left((\sigma_{W_1,Y} - b\sigma_{W_1,X})'c\right)\sqrt{1 - a^2}(k_1 - bk_0) + a^2(k_1 - bk_0)^2}{(k_1 - bk_0)^2}.$$

Therefore

$$\begin{aligned} &(k_1 - bk_0)^2 \left(\underline{r}_Y(z, c^+, b)^{-2} - \underline{r}_Y(z, c_{\text{outside}}, b)^{-2}\right) \\ &= 2z\sqrt{1 - a^2}(k_1 - bk_0)(\sigma_{W_1,Y} - b\sigma_{W_1,X})'(c_{\text{outside}} - c^+) \\ &= 2z\sqrt{1 - a^2}(k_1 - bk_0)(\sigma_{W_1,Y} - b\sigma_{W_1,X})'(a_2 v_2 + a_3 v_3 - \sqrt{a_2^2 + a_3^2} v_2) \\ &= 2z\sqrt{1 - a^2}(k_1 - bk_0)\left((\sigma_{W_1,Y} - b\sigma_{W_1,X})' v_2\right)(a_2 - \sqrt{a_2^2 + a_3^2}). \end{aligned}$$

The last line follows since $v_3 \in \text{span}\{\sigma_{W_1,X}, \sigma_{W_1,Y}\}^\perp$. Likewise,

$$
\begin{aligned}
&(k_1 - bk_0)^2 \left(\underline{r}_Y(z, c^-, b)^{-2} - \underline{r}_Y(z, c_{\text{outside}}, b)^{-2}\right) \\
&= 2\sqrt{1-a^2}(k_1 - bk_0)z\left((\sigma_{W_1,Y} - b\sigma_{W_1,X})'v_2\right)(a_2 + \sqrt{a_2^2 + a_3^2}).
\end{aligned}
$$

Recall that $z \neq 0$ by assumption, $a^2 < 1$ by $\bar{c} < 1$, and $k_1 \neq bk_0$ by $b \neq \beta_{\text{med}}$. Also,

$$
(\sigma_{W_1,Y} - b\sigma_{W_1,X})'v_2 = \sigma'_{W_1,Y} v_2 > 0.
$$

The equality follows since v_2 is orthogonal to $\sigma_{W_1,X}$. The inequality follows from

$$
\begin{aligned}
\sigma'_{W_1,Y} v_2 &= \frac{\|\sigma_{W_1,X}\|^2 \|\sigma_{W_1,Y}\|^2 - (\sigma'_{W_1,Y}\sigma_{W_1,X})^2}{\|\sigma_{W_1,X}\|\sqrt{\|\sigma_{W_1,Y}\|^2 - (\sigma'_{W_1,X}\sigma_{W_1,Y})^2}} \\
&> 0,
\end{aligned}
$$

which itself follows from the Cauchy-Schwarz inequality and linear independence of $(\sigma_{W_1,X}, \sigma_{W_1,Y})$. Moreover, since $a_3 \neq 0$,

$$
a_2 - \sqrt{a_2^2 + a_3^2} < 0 < a_2 + \sqrt{a_2^2 + a_3^2}.
$$

Therefore either

$$
(k_1 - bk_0)^2 \left(\underline{r}_Y(z, c^-, b)^{-2} - \underline{r}_Y(z, c_{\text{outside}}, b)^{-2}\right) > 0
$$

or

$$
(k_1 - bk_0)^2 \left(\underline{r}_Y(z, c^+, b)^{-2} - \underline{r}_Y(z, c_{\text{outside}}, b)^{-2}\right) > 0.
$$

Since $k_1 \neq bk_0$ by $b \neq \beta_{\text{med}}$, this implies that either $\underline{r}_Y(z, c^+, b) < \underline{r}_Y(z, c_{\text{outside}}, b)$ or $\underline{r}_Y(z, c^-, b) < \underline{r}_Y(z, c_{\text{outside}}, b)$.

Part 2: We finish the proof by showing that the result of Part 1 implies the lemma's claim. We have

$$
\begin{aligned}
&\inf\{\underline{r}_Y(z, c, b) : p(z, c; \bar{r}_X) \geq 0, c \in \mathbb{R}^{d_1}, \|c\| \leq \bar{c}\} \\
&= \inf\{\inf\{\underline{r}_Y(z, c, b) : p(z, c; \bar{r}_X) \geq 0, c \in \mathbb{R}^{d_1}, \|c\| = a\} : a \in [0, \bar{c}]\} \\
&= \inf\{\inf\{\underline{r}_Y(z, c, b) : p(z, c; \bar{r}_X) \geq 0, c \in \text{span}\{\sigma_{W_1,X}, \sigma_{W_1,Y}\}, \|c\| = a\} : a \in [0, \bar{c}]\} \\
&= \inf\{\underline{r}_Y(z, c, b) : p(z, c; \bar{r}_X) \geq 0, c \in \text{span}\{\sigma_{W_1,X}, \sigma_{W_1,Y}\}, \|c\| \leq \bar{c}\}.
\end{aligned}
$$

The first and third equality follow from a sequential minimization. The second equality follows from follows from Part 1 of this proof. □

E.2 Proofs of Main Results

We are now ready to prove our main results. Note that we can write

$$
\underline{B}(\bar{r}_X) = \inf \bigcup_{(r_X, r_Y, c): \|r_X\| \leq \bar{r}_X} \mathcal{B}(r_X, r_Y, c)
$$

and

$$
\overline{B}(\bar{r}_X) = \sup \bigcup_{(r_X, r_Y, c): \|r_X\| \leq \bar{r}_X} \mathcal{B}(r_X, r_Y, c).
$$

By Theorem 5, these are the infimums and supremums of $\mathcal{B}_I(\bar{r}_X)$, the identified set for β_{long} when both r_Y and c are not constrained. We can similarly write

$$
\underline{B}(\bar{r}_X, \bar{c}) = \inf \bigcup_{(r_X, r_Y, c): \|r_X\| \leq \bar{r}_X, \|c\| \leq \bar{c}} \mathcal{B}(r_X, r_Y, c)
$$

$$\overline{B}(\bar{r}_X, \bar{c}) = \sup \bigcup_{(r_X, r_Y, c): \|r_X\| \leq \bar{r}_X, \|c\| \leq \bar{c}} \mathcal{B}(r_X, r_Y, c).$$

By Theorem 5, these are the infimums and supremums of $\mathcal{B}_I(\bar{r}_X, \bar{c})$, the identified set for β_{long} when r_Y is not constrained.

We prove Theorem 3 first since Theorem 2 will be shown as the special case where $\bar{c} = 1$.

Proof of Theorem 3. Let $\mathcal{R}(\bar{r}_X, \bar{c}) = \{(r_X, c) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_1} : \|r_X\| \leq \bar{r}_X, \|c\| \leq \bar{c}, \|c\| \neq 1\}$. First note that

$$\begin{aligned}
\mathcal{B}_I(\bar{r}_X, \bar{c}) &= \bigcup_{(r_X, r_Y, c): \|r_X\| \leq \bar{r}_X, \|c\| \leq \bar{c}} \mathcal{B}(r_X, r_Y, c) \\
&= \bigcup_{(r_X, r_Y, c): \|r_X\| \leq \bar{r}_X, \|c\| \leq \bar{c}, \|c\| \neq 1} \mathcal{B}(r_X, r_Y, c) \\
&= \bigcup_{(r_X, c) \in \mathcal{R}(\bar{r}_X, \bar{c}), r_Y \in \mathbb{R}^{d_1}} \mathcal{B}(r_X, r_Y, c).
\end{aligned}$$

The first line holds by Theorem 5. The second line follows since $\|c\| < 1$ holds whenever $b \in \mathcal{B}(r_X, r_Y, c)$, by definition of $\mathcal{B}(r_X, r_Y, c)$. The last line follows by definition of $\mathcal{R}(\bar{r}_X, \bar{c})$. Next we consider the two cases specified in the theorem statement separately.

Case 1. $\bar{z}_X(\bar{r}_X, \bar{c})^2 < k_0$.

By Lemma 9, for all $(r_X, c) \in \mathcal{R}(\bar{r}_X, \bar{c})$:

$$z_X(r_X, c)^2 \leq \bar{z}_X(\bar{r}_X, \bar{c})^2.$$

This result combined with $\bar{z}_X(\bar{r}_X, \bar{c})^2 < k_0 < \infty$ and the fact that the denominator of $z_X(r_X, c)^2$ is $(1 + r_X' c)^2$ implies that $r_X' c \neq -1$ for all $(r_X, c) \in \mathcal{R}(\bar{r}_X, \bar{c})$. Partition $\mathcal{R}(\bar{r}_X, \bar{c}) = \mathcal{R}^{=0}(\bar{r}_X, \bar{c}) \cup \mathcal{R}^{\neq 0}(\bar{r}_X, \bar{c})$ where

$$\begin{aligned}
\mathcal{R}^{=0}(\bar{r}_X, \bar{c}) &= \{(r_X, c) \in \mathcal{R}(\bar{r}_X, \bar{c}) : z_X(r_X, c) = 0\} \\
\mathcal{R}^{\neq 0}(\bar{r}_X, \bar{c}) &= \{(r_X, c) \in \mathcal{R}(\bar{r}_X, \bar{c}) : z_X(r_X, c) \neq 0\}.
\end{aligned}$$

This partition lets us write

$$\mathcal{B}_I(\bar{r}_X, \bar{c}) = \left\{ \bigcup_{(r_X, c) \in \mathcal{R}^{=0}(\bar{r}_X, \bar{c}), r_Y \in \mathbb{R}^{d_1}} \mathcal{B}(r_X, r_Y, c) \right\} \bigcup \left\{ \bigcup_{(r_X, c) \in \mathcal{R}^{\neq 0}(\bar{r}_X, \bar{c}), r_Y \in \mathbb{R}^{d_1}} \mathcal{B}(r_X, r_Y, c) \right\}.$$

Next we characterize each of these unions.

1. First let b be in the union on the left. Then there is a $(r_X, c) \in \mathcal{R}^{=0}(\bar{r}_X, \bar{c})$ and $r_Y \in \mathbb{R}^{d_1}$ such that $b \in \mathcal{B}(r_X, r_Y, c)$. Recall the definition of $z_X(r_X, c)$:

   $$z_X(r_X, c) = \frac{r_X' \sigma_{W_1, X} \sqrt{1 - \|c\|^2}}{1 + r_X' c}.$$

   By definition of $\mathcal{R}(\bar{r}_X, \bar{c})$, $\|c\| < 1$. Consequently, $z_X(r_X, c) = 0$ implies $r_X' \sigma_{W_1, X} = 0$. Since $b \in \mathcal{B}(r_X, r_Y, c)$, there is a $p_1 \in \mathbb{R}^{d_1}$ such that $(I + c r_X') p_1 = \sigma_{W_1, X}$ holds, by Lemma 4. Since $r_X' c \neq -1$, this equation implies that $r_X' p_1 = 0$. Also by Lemma 4, there is a $g_1 \in \mathbb{R}^{d_1}$ such that $(p_1' r_X)(g_1' r_Y)(1 - \|c\|^2) = k_1 - b k_0$ holds. But since $r_X' p_1 = 0$, we must have $k_1 - b k_0 = 0$. Thus $b = \beta_{\text{med}}$. Thus we have shown that any element in the union on the left must be β_{med}. All we have left here is to show that this union is not empty, so that there does in fact exist one element in it. To see that it is non-empty, note that $b = \beta_{\text{med}}$ combined with the choices $r_Y = 0$, $g_1 = \sigma_{W_1, Y} - \beta_{\text{med}} \sigma_{W_1, X}$, $p_1 = \sigma_{W_1, X}$, $r_X = 0$, and any c with $\|c\| < 1$ imply that $b \in \mathcal{B}(r_X, r_Y, p)$ by Lemma 4 and that $(r_X, c) \in \mathcal{R}^{=0}(r_X, c)$.

Thus we conclude that

$$\left\{ \bigcup_{(r_X,c)\in\mathcal{R}^{=0}(\bar{r}_X,\bar{c}),r_Y\in\mathbb{R}^{d_1}} \mathcal{B}(r_X,r_Y,c) \right\} = \{\beta_{\text{med}}\}.$$

2. Next consider $(r_X,c) \in \mathcal{R}^{\neq 0}(\bar{r}_X,\bar{c})$. By Lemma 9, $0 < z_X(r_X,c)^2 \leq \bar{z}_X(\bar{r}_X,\bar{c})^2 < k_0$ for all $(r_X,c) \in \mathcal{R}^{\neq 0}(\bar{r}_X,\bar{c})$. So Lemma 8 shows that, for all $(r_X,c) \in \mathcal{R}^{\neq 0}(\bar{r}_X,\bar{c})$,

$$\bigcup_{r_Y\in\mathbb{R}^{d_1}} \mathcal{B}(r_X,r_Y,c) = \big(\underline{b}(z_X(r_X,c)),\bar{b}(z_X(r_X,c))\big) \setminus \mathcal{B}^0(z_X(r_X,c),c).$$

Using these characterizations of the two partition sets gives

$$\begin{aligned}
\mathcal{B}_I(\bar{r}_X,\bar{c}) &= \{\beta_{\text{med}}\} \cup \bigcup_{(r_X,c)\in\mathcal{R}^{\neq 0}(\bar{r}_X,\bar{c})} \big(\underline{b}(z_X(r_X,c)),\bar{b}(z_X(r_X,c))\big) \setminus \mathcal{B}^0(z_X(r_X,c),c) \\
&= \{\beta_{\text{med}}\} \cup \bigcup_{(r_X,c)\in\mathcal{R}(\bar{r}_X,\bar{c})} \big(\underline{b}(z_X(r_X,c)),\bar{b}(z_X(r_X,c))\big) \setminus \mathcal{B}^0(z_X(r_X,c),c).
\end{aligned}$$

The second line follows since $\big(\underline{b}(z_X(r_X,c)),\bar{b}(z_X(r_X,c))\big) = \emptyset$ when $z_X(r_X,c) = 0$. We can now derive our desired expression for $\underline{B}(\bar{r}_X,\bar{c})$ as follows:

$$\begin{aligned}
\underline{B}(\bar{r}_X,\bar{c}) &= \inf \mathcal{B}_I(\bar{r}_X,\bar{c}) \\
&= \inf \left[\{\beta_{\text{med}}\} \cup \left(\bigcup_{(r_X,c)\in\mathcal{R}(\bar{r}_X,\bar{c})} \big(\underline{b}(z_X(r_X,c)),\bar{b}(z_X(r_X,c))\big) \setminus \mathcal{B}^0(z_X(r_X,c),c) \right) \right] \\
&= \inf\{\underline{b}(z_X(r_X,c)) : (r_X,c) \in \mathcal{R}(\bar{r}_X,\bar{c})\} \\
&= \beta_{\text{med}} - \sup\left\{ \sqrt{\text{devsq}(z_X(r_X,c))} : \|r_X\| \leq \bar{r}_X, \|c\| \leq \bar{c}, \|c\| \neq 1 \right\}.
\end{aligned}$$

The first line follows by definition. The second line follows by our above derivations. In the third line we used the fact that $\underline{b}(z) \leq \beta_{\text{med}}$ for all $z \geq 0$. The fourth line follows by the definition of $\underline{b}(\cdot)$. Next, recall that

$$\text{devsq}(z) = \frac{z^2\left(\frac{k_2}{k_0} - \beta_{\text{med}}^2\right)}{k_0 - z^2}.$$

By $k_2/k_0 > \beta_{\text{med}}^2$ (Lemma 2), devsq(z) is increasing in $|z|$ over all $z^2 < k_0$. Therefore, since $z_X(r_X,c)^2 < k_0$ for all $(r_X,c) \in \mathcal{R}(\bar{r}_X,\bar{c})$, Lemma 9 implies that the infimum of $\underline{b}(z_X(r_X,c))$ is attained at

$$\underline{B}(\bar{r}_X,\bar{c}) = \beta_{\text{med}} - \sqrt{\text{devsq}(\bar{z}_X(\bar{r}_X,\bar{c}))}.$$

The analysis for $\overline{B}(\bar{r}_X,\bar{c})$ is analogous.

Case 2. $\bar{z}_X(\bar{r}_X,\bar{c})^2 \geq k_0$.

Let $\{\bar{r}_X^{(n)}\}$ be a sequence such that $\bar{r}_X^{(n)} \leq \bar{r}_X$ and $\bar{z}_X(\bar{r}_X^{(n)},\bar{c})^2 < k_0$ for all $n \in \mathbb{N}$. Then for any $n \in \mathbb{N}$,

$$\begin{aligned}
\overline{B}(\bar{r}_X,\bar{c}) &= \sup \bigcup_{(r_X,r_Y,c):\|r_X\|\leq\bar{r}_X,\|c\|\leq\bar{c}} \mathcal{B}(r_X,r_Y,c) \\
&\geq \sup \bigcup_{(r_X,r_Y,c):\|r_X\|\leq\bar{r}_X^{(n)},\|c\|\leq\bar{c}} \mathcal{B}(r_X,r_Y,c) \\
&= \overline{B}(\bar{r}_X^{(n)},\bar{c})
\end{aligned}$$

$$= \beta_{\text{med}} + \sqrt{\frac{\bar{z}_X(\bar{r}_X^{(n)}, \bar{c})^2 (k_2/k_0 - \beta_{\text{med}}^2)}{k_0 - \bar{z}_X(\bar{r}_X^{(n)}, \bar{c})^2}}.$$

The first and third lines follow by definition. The second line follows from $\bar{r}_X^{(n)} \leq \bar{r}_X$. The fourth line follows by our analysis in Case 1, since $\bar{z}_X(\bar{r}_X^{(n)}, \bar{c})^2 < k_0$. Suppose $\bar{z}_X(\bar{r}_X^{(n)}, \bar{c})^2 \nearrow k_0$ as $n \to \infty$. Then

$$\lim_{n\to\infty} \left(\beta_{\text{med}} + \sqrt{\frac{\bar{z}_X(\bar{r}_X^{(n)}, \bar{c})^2 (k_2/k_0 - \beta_{\text{med}}^2)}{k_0 - \bar{z}_X(\bar{r}_X^{(n)}, \bar{c})^2}} \right) = +\infty.$$

This follows from

$$\lim_{z^2 \nearrow k_0} \frac{z^2}{k_0 - z^2} = +\infty,$$

since $k_0 > 0$. This would be sufficient to obtain $\overline{B}(\bar{r}_X, \bar{c}) = +\infty$.

So all that remains to be shown is that there exists such a sequence $\{\bar{r}_X^{(n)}\}$ with $\bar{r}_X^{(n)} \leq \bar{r}_X$ and $\bar{z}_X(\bar{r}_X^{(n)}, \bar{c})^2 \nearrow k_0$ as $n \to \infty$. We consider two cases.

1. Suppose $\bar{c} = 0$. Then $\bar{z}_X(\bar{r}_X^{(n)}, \bar{c})^2 = (\bar{r}_X^{(n)})^2 \|\sigma_{W_1,X}\|^2$. This function is continuous in $\bar{r}_X^{(n)}$. Moreover, since $\|\sigma_{W_1,X}\| \neq 0$ by A7, it has range $[0, \infty)$. Thus the desired sequence exists.

2. Suppose $\bar{c} > 0$. To ensure that $\bar{z}_X(\bar{r}_X^{(n)}, \bar{c})^2 < k_0$ for all $n \in \mathbb{N}$, we need at least $\bar{r}_X^{(n)}\bar{c} < 1$; otherwise we would have $\bar{z}_X(\bar{r}_X^{(n)}, \bar{c})^2 = +\infty$ for some n. Thus our sequence must satisfy $\bar{r}_X^{(n)} < 1/\bar{c}$. Next notice that $\bar{z}_X(\bar{r}_X^{(n)}, \bar{c})^2$ is continuous in $\bar{r}_X^{(n)}$ over the set $[0, 1/\bar{c})$. Moreover, its range on this set is $[0, \infty)$. This follows from continuity combined with the boundary values $\bar{z}_X(0, \bar{c}) = 0$ and

   $$\begin{aligned} \lim_{\bar{r}_X \nearrow 1/\bar{c}} \bar{z}_X(\bar{r}_X, \bar{c}) &= \lim_{\bar{r}_X \nearrow 1/\bar{c}} \frac{\bar{r}_X \|\sigma_{W_1,X}\| \sqrt{1 - \min\{\bar{c}, \bar{r}_X\}^2}}{1 - \bar{r}_X \min\{\bar{r}_X, \bar{c}\}} \\ &= +\infty. \end{aligned}$$

   To see the second equality, note that $1/\bar{c} \geq 1$ since $\bar{c} = 1$ is the logically largest possible value of this parameter. Hence $\min\{\bar{r}_X, \bar{c}\} \to \bar{c}$ as $\bar{r}_X \nearrow 1/\bar{c}$. So $1 - \bar{r}_X \bar{c} \to 0$ along this sequence. So the denominator term converges to zero. The numerator converges the constant $(1/\bar{c})\|\sigma_{W_1,X}\|\sqrt{1 - \bar{c}^2}$. For $\bar{c} < 1$, putting these results together shows that the desired sequence $\bar{r}_X^{(n)}$ exists. For $\bar{c} = 1$, any sequence $\bar{r}_X \nearrow 1$ has $\min\{\bar{r}_X, \bar{c}\} = \bar{r}_X$ and hence

   $$\bar{z}_X(\bar{r}_X, \bar{c}) = \frac{\bar{r}_X}{\sqrt{1 - \bar{r}_X^2}} \|\sigma_{W_1,X}\|,$$

   which converges to $+\infty$ as $\bar{r}_X \nearrow 1$.

The proof of $\underline{B}(\bar{r}_X, \bar{c}) = -\infty$ is analogous. □

Proof of Theorem 2. This result follows as a corollary of Theorem 3 by setting $\bar{c} = 1$. Moreover, note that $\bar{z}_X(\bar{r}_X, 1) = \bar{z}_X(\bar{r}_X)$ for all $\bar{r}_X$. □

Proof of Corollary 1. Note that we can equivalently write

$$\bar{r}_X^{\text{bp}} = \inf\{\bar{r}_X \geq 0 : b \in \mathcal{B}_I(\bar{r}_X) \text{ for some } b \leq 0\}.$$

If $\beta_{\text{med}} \geq 0$, then

$$\begin{aligned} \bar{r}_X^{\text{bp}} &= \inf\{\bar{r}_X \geq 0 : b \in [\underline{B}(\bar{r}_X), \overline{B}(\bar{r}_X)] \text{ for some } b \leq 0\} \\ &= \inf\{\bar{r}_X \geq 0 : \text{dev}(\bar{r}_X)^2 = \beta_{\text{med}}^2\}. \end{aligned}$$

The second equality follows by the monotonicity and continuity of $\text{dev}(\bar{r}_X)$. If $\beta_{\text{med}} \leq 0$, the same equality is obtained (and hence why we work with squares here).

We now show that $\bar{r}_X^{\text{bp}}$ is the unique non-negative solution to $\beta_{\text{med}}^2 = \text{dev}(\bar{r}_X)^2$. If $\bar{r}_X$ is such that $k_0 - \bar{z}_X(\bar{r}_X)^2 \leq 0$, then $\bar{r}_X^{\text{bp}} < \bar{r}_X$ since $(\underline{B}(\bar{r}_X), \overline{B}(\bar{r}_X)) = (-\infty, +\infty)$. Therefore, assume that $\bar{r}_X$ is such that $k_0 - \bar{z}_X(\bar{r}_X)^2 > 0$. Then this equality can be written as

$$\begin{aligned}
\frac{k_1^2}{k_0^2} &= \frac{\bar{z}_X(\bar{r}_X)^2\left(\frac{k_2}{k_0} - \frac{k_1^2}{k_0^2}\right)}{k_0 - \bar{z}_X(\bar{r}_X)^2} \\
&= \frac{\frac{\bar{r}_X^2(1-k_0)}{1-\bar{r}_X^2}\frac{k_2k_0-k_1^2}{k_0^2}}{k_0 - \frac{\bar{r}_X^2(1-k_0)}{1-\bar{r}_X^2}} \\
&= \frac{\bar{r}_X^2(1-k_0)(k_2k_0-k_1^2)}{k_0^3(1-\bar{r}_X^2) - k_0^2\bar{r}_X^2(1-k_0)} \\
&= \frac{\bar{r}_X^2(1-k_0)(k_2k_0-k_1^2)}{k_0^2(k_0-\bar{r}_X^2)}.
\end{aligned}$$

Rearranging, we obtain

$$k_1^2\left(k_0 - \bar{r}_X^2\right) = \bar{r}_X^2(1-k_0)(k_2k_0-k_1^2).$$

Solving for $\bar{r}_X^2$ we find that

$$\begin{aligned}
\left(\bar{r}_X^{\text{bp}}\right)^2 &= \frac{k_0k_1^2}{(1-k_0)(k_2k_0-k_1^2)+k_1^2} \\
&= \frac{k_1^2/k_0k_2}{(1-k_0)/k_0 + k_1^2/k_0k_2}.
\end{aligned}$$

Note that

$$\frac{k_1^2}{k_0k_2} = \frac{k_1^2}{k_0^2}\frac{k_0}{k_2} = \frac{\text{var}(\beta_{\text{med}}X^{\perp W_1})}{\text{var}(Y^{\perp W_1})} = R^2_{Y\sim X\bullet W_1}$$

and

$$1 - k_0 = 1 - \frac{\text{var}(X^{\perp W_1})}{\text{var}(X)} = R^2_{X\sim W_1}$$

by $\text{var}(X) = 1$. Therefore

$$\left(\bar{r}_X^{\text{bp}}\right)^2 = \frac{R^2_{Y\sim X\bullet W_1}}{\frac{R^2_{X\sim W_1}}{1-R^2_{X\sim W_1}} + R^2_{Y\sim X\bullet W_1}}.$$

Taking the positive square root gives

$$\bar{r}_X^{\text{bp}} = \left(\frac{R^2_{Y\sim X\bullet W_1}}{\frac{R^2_{X\sim W_1}}{1-R^2_{X\sim W_1}} + R^2_{Y\sim X\bullet W_1}}\right)^{1/2}.$$

□

Proof of Theorem 4. Note that we can equivalently write

$$\bar{r}_Y^{\text{bf}}(\bar{r}_X, \bar{c}, \underline{b}) = \inf\{\bar{r}_Y \geq 0 : b \in \mathcal{B}_I(\bar{r}_X, \bar{c}, \bar{r}_Y) \text{ for some } b \leq \underline{b}\}.$$

We use this version in our proof. We will consider each case stated in the theorem separately.

Case 1 If $\underline{b} \geq \beta_{\text{med}}$, then $\bar{r}_Y^{\text{bf}}(\bar{r}_X, \bar{c}; \underline{b}) = 0$.

By Lemma 6, $\beta_{\text{med}} \in \mathcal{B}(0,0,0)$. This implies

$$\beta_{\text{med}} \in \bigcup_{(r_X,c):\|r_X\|\le \bar{r}_X, \|c\|\le \bar{c}} \mathcal{B}(r_X,0,c) = \mathcal{B}_I(\bar{r}_X,0,\bar{c}).$$

We then have

$$\inf\{\bar{r}_Y \ge 0 : b \in \mathcal{B}_I(\bar{r}_X,\bar{r}_Y,\bar{c}) \text{ for some } b \le \underline{b}\} = 0$$

by letting $b = \beta_{\text{med}} \le \underline{b}$.

Case 2 If $\underline{B}(\bar{r}_X,\bar{c}) > \underline{b}$, then $\bar{r}_Y^{\text{bf}}(\bar{r}_X,\bar{c};\underline{b}) = +\infty$.

Note that

$$\begin{aligned}\mathcal{B}_I(\bar{r}_X,\bar{r}_Y,\bar{c}) &= \bigcup_{(r_X,r_Y,c):\|r_X\|\le \bar{r}_X, \|r_Y\|\le \bar{r}_Y, \|c\|\le \bar{c}} \mathcal{B}(r_X,r_Y,c) \\ &\subseteq \bigcup_{(r_X,r_Y,c):\|r_X\|\le \bar{r}_X, \|c\|\le \bar{c}} \mathcal{B}(r_X,r_Y,c).\end{aligned}$$

Therefore,

$$\begin{aligned}&\{\bar{r}_Y \ge 0 : b \in \mathcal{B}_I(\bar{r}_X,\bar{r}_Y,\bar{c}) \text{ for some } b \le \underline{b}\} \\ &\subseteq \left\{\bar{r}_Y \ge 0 : b \in \bigcup_{(r_X,r_Y,c):\|r_X\|\le \bar{r}_X, \|c\|\le \bar{c}} \mathcal{B}(r_X,r_Y,c) \text{ for some } b \le \underline{b}\right\} \\ &= \emptyset.\end{aligned}$$

The last line follows by

$$\underline{B}(\bar{r}_X,\bar{c}) = \inf \bigcup_{(r_X,r_Y,c):\|r_X\|\le \bar{r}_X, \|c\|\le \bar{c}} \mathcal{B}(r_X,r_Y,c) > \underline{b}.$$

Hence

$$\inf\{\bar{r}_Y \ge 0 : b \in \mathcal{B}_I(\bar{r}_X,\bar{r}_Y,\bar{c}) \text{ for some } b \le \underline{b}\} = +\infty.$$

Case 3 If $\underline{B}(\bar{r}_X,\bar{c}) \le \underline{b} < \beta_{\text{med}}$.

First, note that

$$\begin{aligned}\bar{r}_Y^{\text{bf}}(\bar{r}_X,\bar{c},\underline{b}) &= \inf\{\bar{r}_Y \ge 0 : b \in \mathcal{B}_I(\bar{r}_X,\bar{r}_Y,\bar{c}) \text{ for some } b \le \underline{b}\} \\ &= \inf\{\bar{r}_Y \ge 0 : \text{There is a } (b,r_X,r_Y,c) \text{ s.t. } b \in \mathcal{B}(r_X,r_Y,c), \|r_X\| \le \bar{r}_X, \|r_Y\| \le \bar{r}_Y, \|c\| \le \bar{c}, b \le \underline{b}\} \\ &= \inf\{\|r_Y\| : \text{There is a } (b,r_X,r_Y,c) \text{ s.t. } b \in \mathcal{B}(r_X,r_Y,c), \|r_X\| \le \bar{r}_X, r_Y \in \mathbb{R}^{d_1}, \|c\| \le \bar{c}, b \le \underline{b}\} \\ &= \inf\left\{\|r_Y\| : r_Y \in \bigcup_{\|r_X\|\le \bar{r}_X, \|c\|\le \bar{c}, b \le \underline{b}} \{r_Y \in \mathbb{R}^{d_1} : b \in \mathcal{B}(r_X,r_Y,c)\}\right\}.\end{aligned}$$

The second equality follows from the definition of $\mathcal{B}_I(\bar{r}_X,\bar{r}_Y,\bar{c})$. The third holds because $\bar{r}_Y \ge \|r_Y\|$, so it is minimized by setting $\bar{r}_Y = \|r_Y\|$.

Next we can write

$$\begin{aligned}&\bigcup_{\|r_X\|\le \bar{r}_X, \|c\|\le \bar{c}, b \le \underline{b}} \{r_Y \in \mathbb{R}^{d_1} : b \in \mathcal{B}(r_X,r_Y,c)\} \\ &= \bigcup_{\|r_X\|\le \bar{r}_X, \|c\|\le \bar{c}, b \le \underline{b}} \{r_Y \in \mathbb{R}^{d_1} : (29)\text{–}(33) \text{ hold for } (r_X,r_Y,c,b,z) \text{ and for some } z \in \mathbb{R}\setminus\{0\}\}\end{aligned}$$

$$= \bigcup_{\|r_X\| \le \bar{r}_X, \|c\| \le \bar{c}, b \le \underline{b}} \{r_Y \in \mathbb{R}^{d_1} : (29) \text{ holds for } (r_X, r_Y, c, b, z) \text{ where } (r_X, c, b, z) \in A(\bar{r}_X, \bar{c}), z \in \mathbb{R} \setminus \{0\}\}$$
$$= \bigcup_{(r_X, c, b, z) \in A(\bar{r}_X, \bar{c}), b \le \underline{b}} \{r_Y \in \mathbb{R}^{d_1} : (29) \text{ holds for } (r_X, r_Y, c, b, z)\}.$$

The first equality follows by Lemma 5, which we can apply since we consider b satisfying $b \le \underline{b} < \beta_{\text{med}}$. The second equality follows by definition of the set $A(\bar{r}_X, \bar{c})$. The third equality follows since the inequalities $\|r_X\| \le \bar{r}_X$ and $\|c\| \le \bar{c}$ are part of the definition of $A(\bar{r}_X, \bar{c})$.

Thus

$$\begin{aligned}
\bar{r}_Y^{\text{bf}}(\bar{r}_X, \bar{c}, \underline{b}) &= \inf \left\{ \|r_Y\| : r_Y \in \bigcup_{(r_X, c, b, z) \in A(\bar{r}_X, \bar{c}), b \le \underline{b}} \{r_Y \in \mathbb{R}^{d_1} : (29) \text{ holds for } (r_X, r_Y, c, b, z)\} \right\} \\
&= \inf \Big\{ \inf \big\{ \|r_Y\| : r_Y \in \mathbb{R}^{d_1}, (29) \text{ holds for } (r_X, r_Y, c, b, z) \big\} : (r_X, c, b, z) \in A(\bar{r}_X, \bar{c}), b \le \underline{b} \Big\} \\
&= \inf \left\{ \underline{r}_Y(z, c, b) : (r_X, c, b, z) \in A(\bar{r}_X, \bar{c}), b \le \underline{b} \right\}.
\end{aligned}$$

The last line follows by Lemma 7. We have just shown that $\bar{r}_Y^{\text{bf}}(\bar{r}_X, \bar{c}, \underline{b})$ can be computed as the solution to a constrained minimization problem with objective function $\underline{r}_Y(z, c, b)$. We complete the proof by showing that the constraints $(r_X, c, b, z) \in A(\bar{r}_X, \bar{c}), b \le \underline{b}$ here are equivalent to the constraints given in the statement of the theorem.

The constraint set is

$$\begin{aligned}
&\{(r_X, c, b, z) \in A(\bar{r}_X, \bar{c}) : b \le \underline{b}\} \\
&= \{(r_X, c, b, z) : (30)\text{–}(32) \text{ hold}, \|r_X\| \le \bar{r}_X, \|c\| \le \bar{c}, b \le \underline{b}, z \in \mathbb{R} \setminus \{0\}\} \\
&= \{(r_X, c, b, z) : (b - \beta_{\text{med}})^2 < \text{devsq}(z), z^2 < k_0, \\
&\qquad\qquad z = r_X'(\sqrt{1 - \|c\|^2}\sigma_{W_1,X} - cz), \|r_X\| \le \bar{r}_X, \|c\| \le \bar{c}, b \le \underline{b}, z \in \mathbb{R} \setminus \{0\}\} \\
&= \{(r_X, c, b, z) : (b - \beta_{\text{med}})^2 < \text{devsq}(z), 0 < z^2 < k_0, \\
&\qquad\qquad z = r_X'(\sqrt{1 - \|c\|^2}\sigma_{W_1,X} - cz), \|r_X\| \le \bar{r}_X, \|c\| \le \bar{c}, b \le \underline{b}\} \\
&= \{(r_X, c, b, z) : (b - \beta_{\text{med}})^2 < \text{devsq}(z), 0 < z^2 < k_0, p(z, c; \bar{r}_X) \ge 0, \|c\| \le \bar{c}, b \le \underline{b}, r_X \in \mathbb{R}^{d_1}\}.
\end{aligned}$$

The first equality follows by definition of $A(\bar{r}_X, \bar{c})$. The second equality follows by definition of equations (30)–(32). The third equality is just a slight simplification. The fourth equality follows from Lemma 10, which shows that for any $(z, c, \bar{r}_X)$ such that $z^2 < k_0$, $\|c\| < 1$, there exists a r_X with $\|r_X\| \le \bar{r}_X$ and $z = r_X'(\sigma_{W_1,X}\sqrt{1 - \|c\|^2} - cz)$ if and only if $p(z, c; \bar{r}_X) \ge 0$. Note that here we also use $\bar{c} < 1$.

Thus we have shown that

$$\bar{r}_Y^{\text{bf}}(\bar{r}_X, \bar{c}, \underline{b}) = \inf\{\underline{r}_Y(z, c, b) : (b - \beta_{\text{med}})^2 < \text{devsq}(z), 0 < z^2 < k_0, p(z, c; \bar{r}_X) \ge 0, \|c\| \le \bar{c}, b \le \underline{b}\}. \quad (34)$$

We can further write this as

$$\begin{aligned}
&\bar{r}_Y^{\text{bf}}(\bar{r}_X, \bar{c}, \underline{b}) \\
&= \inf\{\inf\{\underline{r}_Y(z, c, b) : p(z, c; \bar{r}_X) \ge 0, c \in \mathbb{R}^{d_1}, \|c\| \le \bar{c}\} : (b - \beta_{\text{med}})^2 < \text{devsq}(z), 0 < z^2 < k_0, b \le \underline{b}\} \\
&= \inf\{\inf\{\underline{r}_Y(z, c, b) : p(z, c; \bar{r}_X) \ge 0, c \in \text{span}\{\sigma_{W_1,X}, \sigma_{W_1,Y}\}, \|c\| \le \bar{c}\} : \\
&\qquad\qquad (b - \beta_{\text{med}})^2 < \text{devsq}(z), 0 < z^2 < k_0, b \le \underline{b}\} \\
&= \inf\{\underline{r}_Y(z, c_1\sigma_{W_1,Y} + c_2\sigma_{W_1,X}, b) : (b - \beta_{\text{med}})^2 < \text{devsq}(z), 0 < z^2 < k_0, \\
&\qquad\qquad p(z, c_1\sigma_{W_1,Y} + c_2\sigma_{W_1,X}; \bar{r}_X) \ge 0, \|c_1\sigma_{W_1,Y} + c_2\sigma_{W_1,X}\| \le \bar{c}, (c_1, c_2) \in \mathbb{R}^2, b \le \underline{b}\} \\
&= \inf\{\underline{r}_Y(z, c_1\sigma_{W_1,Y} + c_2\sigma_{W_1,X}, b) : (b - \beta_{\text{med}})^2 < \text{devsq}(z), z^2 < k_0,
\end{aligned}$$

$$p(z, c_1\sigma_{W_1,Y} + c_2\sigma_{W_1,X}; \bar{r}_X) \geq 0, \|c_1\sigma_{W_1,Y} + c_2\sigma_{W_1,X}\| \leq \bar{c}, (c_1,c_2) \in \mathbb{R}^2, b \leq \underline{b}\}.$$

The second equality follows by Lemma 11, which uses the assumptions $\bar{c} < 1$, $d_1 \geq 2$, and $(\sigma_{W_1,X}, \sigma_{W_1,X})$ are linearly independent. Moreover, note that the outer infimum constraint set implies that $z \neq 0$ and $b \neq \beta_{\text{med}}$ (since $\underline{b} < \beta_{\text{med}}$), which are also needed to apply Lemma 11. The third equality follows by combining the two infimums and using the definition of the span. The last equality follows from $\{b \in \mathbb{R} : (b - \beta_{\text{med}})^2 < \text{devsq}(0)\} = \emptyset$, since $\text{devsq}(0) = 0$, so that allowing $z = 0$ does not change this infimum. □

Remark 2 (Extension to $\underline{c} \leq R_{W_2 \sim W_1} \leq \bar{c}$)**.** In this remark we briefly discuss how to extend our identification results to the more general assumption $R_{W_2 \sim W_1} \in [\underline{c}, \bar{c}]$ for known $\underline{c}$ and $\bar{c}$ satisfying $0 \leq \underline{c} \leq \bar{c} \leq 1$. Recall that letting $c = \text{cov}(W_1, W_2)$, and given our normalization $\text{var}(W_1) = I$, this assumption is equivalent to $\underline{c} \leq \|c\| \leq \bar{c}$.

For brevity, we only consider Theorem 3, which is our identified set for β_{long} using the restrictions $\|r_X\| \leq \bar{r}_X$ and $\|c\| \leq \bar{c}$. From the proof of Theorem 3 notice that the constraint on $\|c\|$ is only used in the very last step of case 1. That step invokes Lemma 9, which shows that

$$\sup\{z_X(r_X, c)^2 : \|r_X\| \leq \bar{r}_X, \|c\| \leq \bar{c}\} = \bar{z}_X(\bar{r}_X, \bar{c})^2.$$

This lemma can easily be generalized to include the additional assumption $\underline{c} \leq \|c\|$. To see this, notice that in the proof of the lemma we use the function

$$f(\|r_X\|, \|c\|) = \frac{\|r_X\|\|\sigma_{W_1,X}\|\sqrt{1 - \|c\|^2}}{1 - \|r_X\|\|c\|}.$$

We first show this function is maximized at $f(\bar{r}_X, \|c\|)$ for any value of $\|c\|$. Then we show that

$$\frac{\partial f(\bar{r}_X, a)}{\partial a} \begin{cases} > 0 & \text{if } a < \bar{r}_X \\ = 0 & \text{if } a = \bar{r}_X \\ < 0 & \text{if } a > \bar{r}_X. \end{cases}$$

Note that a must be in the set $[\underline{c}, \bar{c}]$. Therefore this is a simple constrained optimization problem. The solution is

$$\begin{aligned} a^* &= \begin{cases} \underline{c} & \text{if } \bar{r}_X < \underline{c} \\ \bar{r}_X & \text{if } \bar{r}_X \in [\underline{c}, \bar{c}] \\ \bar{c} & \text{if } \bar{r}_X > \bar{c} \end{cases} \\ &= \min\{\max\{\bar{r}_X, \underline{c}\}, \bar{c}\}. \end{aligned}$$

Thus the maximized value of f is

$$\begin{aligned} f(\bar{r}_X, \min\{\max\{\bar{r}_X, \underline{c}\}, \bar{c}\}) &= \frac{\bar{r}_X\|\sigma_{W_1,X}\|\sqrt{1 - \min\{\max\{\bar{r}_X, \underline{c}\}, \bar{c}\}^2}}{1 - \bar{r}_X \min\{\max\{\bar{r}_X, \underline{c}\}, \bar{c}\}} \\ &\equiv \bar{z}_X(\bar{r}_X, \underline{c}, \bar{c}). \end{aligned}$$

Notice that $\bar{z}_X(\bar{r}_X, 0, \bar{c}) = \bar{z}_X(\bar{r}_X, \bar{c})$. To verify that this upper bound can be attained, modify c^* to be

$$c^* = -\frac{\min\{\max\{\bar{r}_X, \underline{c}\}, \bar{c}\}}{\|\sigma_{W_1,X}\|}\sigma_{W_1,X}.$$

The rest of the proof of Lemma 9 continues to hold. In particular, the proof of case 1 continues to hold without modification. Thus we have shown that

$$\sup\{z_X(r_X, c)^2 : \|r_X\| \leq \bar{r}_X, \underline{c} \leq \|c\| \leq \bar{c}\} = \bar{z}_X(\bar{r}_X, \underline{c}, \bar{c})^2.$$

With this extension of the lemma, the generalization of Theorem 3 then uses the bounds

$$\underline{B}(\bar{r}_X, \underline{c}, \bar{c}) = \beta_{\text{med}} - \sqrt{\text{devsq}(\bar{z}_X(\bar{r}_X, \underline{c}, \bar{c}))} \qquad \text{and} \qquad \overline{B}(\bar{r}_X, \underline{c}, \bar{c}) = \beta_{\text{med}} + \sqrt{\text{devsq}(\bar{z}_X(\bar{r}_X, \underline{c}, \bar{c}))}.$$

Finally, note that we assumed $\bar{c} < 1$ in this analysis, because our invertibility assumption A1 rules out $c = 1$, perfect multicollinearity between W_1 and W_2. All of our analysis could be generalized to relax this assumption and allow for $c = 1$, at the cost of additional technical derivations. In particular, note that when $\underline{c} = \bar{c} = 1$ then β_{long} is point identified and equal to β_{med}. That is intuitive since, in this case, there are no omitted variables since $\underline{c} = \bar{c} = 1$ implies that we know the unobservable W_2 is an affine function of the observables W_1. Our analysis is consistent with this case since, by examining the $\bar{z}_X(\bar{r}_X, \underline{c}, \bar{c})$ function, we see that for any fixed $\bar{r}_X$, $\bar{z}_X(\bar{r}_X, \underline{c}, \bar{c}) \to 0$ as $\underline{c}, \bar{c} \to 1$. Hence our bounds collapse to β_{med} as the correlation between the observables and the unobservables is known to become arbitrarily close to one.

F Proofs for Section 4

Proof of Proposition 1. We have

$$\begin{aligned}
\text{cov}(Y, X^{\perp W_1, W_2}) &= \text{cov}(\beta_c X + \gamma_1' W_1 + \gamma_2' W_2 + U, X^{\perp W_1, W_2}) \\
&= \beta_c \, \text{cov}(X, X^{\perp W_1, W_2}) + \text{cov}(U, X^{\perp W_1, W_2}) \\
&= \beta_c \, \text{cov}(X^{\perp W_1, W_2}, X^{\perp W_1, W_2}) + \text{cov}(U^{\perp W_1, W_2}, X^{\perp W_1, W_2}) \\
&= \beta_c \, \text{var}(X^{\perp W_1, W_2}).
\end{aligned}$$

The first line follows by the linear potential outcomes assumption. The last line follows by linear latent unconfoundedness. Thus

$$\beta_c = \frac{\text{cov}(Y, X^{\perp W_1, W_2})}{\text{var}(X^{\perp W_1, W_2})}.$$

The term on the right equals β_{long} by the FWL theorem. Note that the denominator here is nonzero by A1. ☐

Proof of Corollary 2. Theorem 1 implies that $\beta_{\text{long}} = \beta_{\text{med}}$. Proposition 1 implies that $\beta_c = \beta_{\text{long}}$. ☐

Proof of Proposition 2. We have

$$\begin{aligned}
\text{cov}(\Delta Y, (\Delta X)^{\perp \Delta W_1, \Delta W_2}) &= \beta_c \, \text{cov}(\Delta X, (\Delta X)^{\perp \Delta W_1, \Delta W_2})) + \text{cov}(\Delta V, (\Delta X)^{\perp \Delta W_1, \Delta W_2}) \\
&= \beta_c \, \text{var}\left((\Delta X)^{\perp \Delta W_1, \Delta W_2}\right).
\end{aligned}$$

The first line follows by linear potential outcomes. The second line follows by exogeneity. The result then follows by the FWL theorem. ☐

Proof of Corollary 3. Theorem 1 implies that $\beta_{\text{long}} = \beta_{\text{med}}$. Proposition 2 implies that $\beta_c = \beta_{\text{long}}$. ☐

G Additional Details on the Derivation of Equations (15) and (16)

In this section, we provide detailed calculations that lead to equations (15) and (16). We begin with equation (16) since we'll use some of these derivations to show equation (15). As a preliminary step, we first compute

$$\text{var}(W_2, X, W_1)^{-1} = \begin{pmatrix} 1 & \pi_1' c & c' \\ \pi_1' c & \text{var}(X) & \text{cov}(X, W_1) \\ c & \text{cov}(W_1, X) & \text{var}(W_1) \end{pmatrix}^{-1} \equiv \begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1}$$

where we defined

$$A = \begin{pmatrix} 1 & \pi_1' c \\ \pi_1' c & \text{var}(X) \end{pmatrix}, \qquad B = C' = \begin{pmatrix} c' \\ \text{cov}(X, W_1) \end{pmatrix}, \qquad \text{and} \qquad D = \text{var}(W_1).$$

Note that $D = \operatorname{var}(W_1)$ is invertible. $A - BD^{-1}C$ is also invertible because

$$\begin{aligned}
A - BD^{-1}C &= \begin{pmatrix} 1 & \pi_1' c \\ \pi_1' c & \operatorname{var}(X) \end{pmatrix} - \begin{pmatrix} c' \\ \operatorname{cov}(X, W_1) \end{pmatrix} \operatorname{var}(W_1)^{-1} \begin{pmatrix} c & \operatorname{cov}(W_1, X) \end{pmatrix} \\
&= \begin{pmatrix} 1 & \pi_1' c \\ \pi_1' c & \operatorname{var}(X) \end{pmatrix} - \begin{pmatrix} c' \operatorname{var}(W_1)^{-1} c & c' \pi_1 \\ \pi_1' c & \pi_1' \operatorname{cov}(W_1, X) \end{pmatrix} \\
&= \begin{pmatrix} 1 - c' \operatorname{var}(W_1)^{-1} c & 0 \\ 0 & \operatorname{var}(X^{\perp W_1}) \end{pmatrix}.
\end{aligned}$$

Recall that c in equation (16) is such that $c' \operatorname{var}(W_1)^{-1} c < 1$ and that $\operatorname{var}(X^{\perp W_1})$ is invertible by A1. Therefore we can use the following partitioned inverse formula to compute

$$\begin{aligned}
&\begin{pmatrix} 1 & \pi_1' c & c' \\ \pi_1' c & \operatorname{var}(X) & \operatorname{cov}(X, W_1) \\ c & \operatorname{cov}(W_1, X) & \operatorname{var}(W_1) \end{pmatrix}^{-1} \\
&= \begin{pmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1} BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{pmatrix} \\
&= \begin{pmatrix} \begin{pmatrix} (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} & 0 \\ 0 & \operatorname{var}(X^{\perp W_1})^{-1} \end{pmatrix} & -\begin{pmatrix} (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} & 0 \\ 0 & \operatorname{var}(X^{\perp W_1})^{-1} \end{pmatrix} \begin{pmatrix} c' \\ \operatorname{cov}(X, W_1) \end{pmatrix} \operatorname{var}(W_1)^{-1} \\ -\operatorname{var}(W_1)^{-1} \begin{pmatrix} c & \operatorname{cov}(W_1, X) \end{pmatrix} \begin{pmatrix} (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} & 0 \\ 0 & \operatorname{var}(X^{\perp W_1})^{-1} \end{pmatrix} & \operatorname{var}(W_1)^{-1} + \operatorname{var}(W_1)^{-1} \begin{pmatrix} c & \operatorname{cov}(W_1, X) \end{pmatrix} \begin{pmatrix} (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} & 0 \\ 0 & \operatorname{var}(X^{\perp W_1})^{-1} \end{pmatrix} \begin{pmatrix} c' \\ \operatorname{cov}(X, W_1) \end{pmatrix} \operatorname{var}(W_1)^{-1} \end{pmatrix} \\
&= \begin{pmatrix} \begin{pmatrix} (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} & 0 \\ 0 & \operatorname{var}(X^{\perp W_1})^{-1} \end{pmatrix} & -\begin{pmatrix} (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} & 0 \\ 0 & \operatorname{var}(X^{\perp W_1})^{-1} \end{pmatrix} \begin{pmatrix} c' \operatorname{var}(W_1)^{-1} \\ \pi_1' \end{pmatrix} \\ -\begin{pmatrix} \operatorname{var}(W_1)^{-1} c & \pi_1 \end{pmatrix} \begin{pmatrix} (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} & 0 \\ 0 & \operatorname{var}(X^{\perp W_1})^{-1} \end{pmatrix} & \operatorname{var}(W_1)^{-1} + \begin{pmatrix} \operatorname{var}(W_1)^{-1} c & \pi_1 \end{pmatrix} \begin{pmatrix} (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} & 0 \\ 0 & \operatorname{var}(X^{\perp W_1})^{-1} \end{pmatrix} \begin{pmatrix} c' \operatorname{var}(W_1)^{-1} \\ \pi_1' \end{pmatrix} \end{pmatrix} \\
&= \begin{pmatrix} \begin{pmatrix} (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} & 0 \\ 0 & \operatorname{var}(X^{\perp W_1})^{-1} \end{pmatrix} & -\begin{pmatrix} (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} c' \operatorname{var}(W_1)^{-1} \\ \operatorname{var}(X^{\perp W_1})^{-1} \pi_1' \end{pmatrix} \\ -\begin{pmatrix} \operatorname{var}(W_1)^{-1} c (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} & \pi_1 \operatorname{var}(X^{\perp W_1})^{-1} \end{pmatrix} & \operatorname{var}(W_1)^{-1} + \operatorname{var}(W_1)^{-1} c c' \operatorname{var}(W_1)^{-1} (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} + \pi_1 \pi_1' \operatorname{var}(X^{\perp W_1})^{-1} \end{pmatrix} \\
&= \begin{pmatrix} (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} & 0 & -(1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} c' \operatorname{var}(W_1)^{-1} \\ 0 & \operatorname{var}(X^{\perp W_1})^{-1} & -\operatorname{var}(X^{\perp W_1})^{-1} \pi_1' \\ -\operatorname{var}(W_1)^{-1} c (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} & -\pi_1 \operatorname{var}(X^{\perp W_1})^{-1} & \operatorname{var}(W_1)^{-1} + \operatorname{var}(W_1)^{-1} c c' \operatorname{var}(W_1)^{-1} (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} + \pi_1 \pi_1' \operatorname{var}(X^{\perp W_1})^{-1} \end{pmatrix}.
\end{aligned}$$

Recall that here $c = \operatorname{cov}(W_1, W_2)$. Given this expression for $\operatorname{var}(W_2, X, W_1)^{-1}$, we now compute the non-constant term in $\mathbb{L}(Y \mid 1, W_2, X, W_1)$:

$$\begin{aligned}
&\begin{pmatrix} \operatorname{cov}(Y, W_2) & \operatorname{cov}(Y, X) & \operatorname{cov}(Y, W_1) \end{pmatrix} \operatorname{var}(W_2, X, W_1)^{-1} \begin{pmatrix} W_2 \\ X \\ W_1 \end{pmatrix} \\
&= \begin{pmatrix} \beta_{\text{long}} \pi_1' c + g_1' c + g_2 & \beta_{\text{long}} \operatorname{var}(X) + g_1' \operatorname{var}(W_1) \pi_1 + g_2 \pi_1' c & \beta_{\text{long}} \pi_1' \operatorname{var}(W_1) + g_1' \operatorname{var}(W_1) + g_2 c' \end{pmatrix} \begin{pmatrix} 1 & \pi_1' c & c' \\ \pi_1' c & \operatorname{var}(X) & \operatorname{cov}(X, W_1) \\ c & \operatorname{cov}(W_1, X) & \operatorname{var}(W_1) \end{pmatrix}^{-1} \begin{pmatrix} W_2 \\ X \\ W_1 \end{pmatrix} \\
&= \begin{pmatrix} \beta_{\text{long}} \pi_1' c + g_1' c + g_2 & \beta_{\text{long}} \operatorname{var}(X) + g_1' \operatorname{var}(W_1) \pi_1 + g_2 \pi_1' c & \beta_{\text{long}} \pi_1' \operatorname{var}(W_1) + g_1' \operatorname{var}(W_1) + g_2 c' \end{pmatrix} \\
&\times \begin{pmatrix} (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} & 0 & -(1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} c' \operatorname{var}(W_1)^{-1} \\ 0 & \operatorname{var}(X^{\perp W_1})^{-1} & -\operatorname{var}(X^{\perp W_1})^{-1} \pi_1' \\ -\operatorname{var}(W_1)^{-1} c (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} & -\pi_1 \operatorname{var}(X^{\perp W_1})^{-1} & \operatorname{var}(W_1)^{-1} + \operatorname{var}(W_1)^{-1} c c' \operatorname{var}(W_1)^{-1} (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} + \pi_1 \pi_1' \operatorname{var}(X^{\perp W_1})^{-1} \end{pmatrix} \\
&\times \begin{pmatrix} W_2 \\ X \\ W_1 \end{pmatrix} \\
&\equiv T_1 W_2 + T_2 X + T_3' W_1,
\end{aligned}$$

where we defined

$$\begin{aligned}
T_1 &= (\beta_{\text{long}} \pi_1' c + g_1' c + g_2)(1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} \\
&\qquad + (\beta_{\text{long}} \pi_1' \operatorname{var}(W_1) + g_1' \operatorname{var}(W_1) + g_2 c')(-\operatorname{var}(W_1)^{-1} c (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1}) \\
&= (\beta_{\text{long}} \pi_1' c + g_1' c + g_2)(1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} - \beta_{\text{long}} \pi_1' c (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} \\
&\qquad - g_1' c (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} - g_2 c' \operatorname{var}(W_1)^{-1} c (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} \\
&= g_2 (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} - g_2 c' \operatorname{var}(W_1)^{-1} c (1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} \\
&= g_2 \frac{1 - c' \operatorname{var}(W_1)^{-1} c}{1 - c' \operatorname{var}(W_1)^{-1} c} \\
&= g_2
\end{aligned}$$

$$
\begin{aligned}
T_2 &= (\beta_{\text{long}} \operatorname{var}(X) + g_1' \operatorname{var}(W_1)\pi_1 + g_2\pi_1' c) \operatorname{var}(X^{\perp W_1})^{-1} \\
&\quad - (\beta_{\text{long}}\pi_1' \operatorname{var}(W_1) + g_1' \operatorname{var}(W_1) + g_2 c')\pi_1 \operatorname{var}(X^{\perp W_1})^{-1} \\
&= \beta_{\text{long}} \operatorname{var}(X^{\perp W_1})^{-1}(\operatorname{var}(X) - \pi_1' \operatorname{var}(W_1)\pi_1) \\
&\quad + g_1'(\operatorname{var}(W_1)\pi_1 - \operatorname{var}(W_1)\pi_1) \operatorname{var}(X^{\perp W_1})^{-1} + g_2(\pi_1' c - c'\pi_1) \operatorname{var}(X^{\perp W_1})^{-1} \\
&= \beta_{\text{long}}
\end{aligned}
$$

$$
\begin{aligned}
T_3' &= -(\beta_{\text{long}}\pi_1' c + g_1' c + g_2)(1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} c' \operatorname{var}(W_1)^{-1} \\
&\quad - (\beta_{\text{long}} \operatorname{var}(X) + g_1' \operatorname{var}(W_1)\pi_1 + g_2\pi_1' c) \operatorname{var}(X^{\perp W_1})^{-1}\pi_1' \\
&\quad + (\beta_{\text{long}}\pi_1' \operatorname{var}(W_1) + g_1' \operatorname{var}(W_1) + g_2 c')\times \\
&\qquad (\operatorname{var}(W_1)^{-1} + \operatorname{var}(W_1)^{-1} cc' \operatorname{var}(W_1)^{-1}(1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} + \pi_1\pi_1' \operatorname{var}(X^{\perp W_1})^{-1}) \\
&\equiv \beta_{\text{long}} T_{31} + g_1' T_{32} + g_2 T_{33}.
\end{aligned}
$$

In the expression for T_3' we defined

$$
\begin{aligned}
T_{31} &= -\pi_1' c(1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} c' \operatorname{var}(W_1)^{-1} - \operatorname{var}(X) \operatorname{var}(X^{\perp W_1})^{-1}\pi_1' \\
&\quad + \pi_1' \operatorname{var}(W_1)(\operatorname{var}(W_1)^{-1} + \operatorname{var}(W_1)^{-1} cc' \operatorname{var}(W_1)^{-1}(1 - c' \operatorname{var}(W_1)^{-1} c)^{-1} + \pi_1\pi_1' \operatorname{var}(X^{\perp W_1})^{-1}) \\
&= \frac{-\pi_1' cc' \operatorname{var}(W_1)^{-1}}{1 - c' \operatorname{var}(W_1)^{-1} c} - \operatorname{var}(X) \operatorname{var}(X^{\perp W_1})^{-1}\pi_1' + \pi_1' \\
&\quad + \frac{\pi_1' cc' \operatorname{var}(W_1)^{-1}}{1 - c' \operatorname{var}(W_1)^{-1} c} + \pi_1' \operatorname{var}(W_1)\pi_1\pi_1' \operatorname{var}(X^{\perp W_1})^{-1} \\
&= -\operatorname{var}(X) \operatorname{var}(X^{\perp W_1})^{-1}\pi_1' + \pi_1' + \pi_1' \operatorname{var}(W_1)\pi_1\pi_1' \operatorname{var}(X^{\perp W_1})^{-1} \\
&= \frac{1}{\operatorname{var}(X^{\perp W_1})}\big(- \operatorname{var}(X) + \operatorname{var}(X^{\perp W_1}) + \pi_1' \operatorname{var}(W_1)\pi_1\big)\pi_1' \\
&= \frac{1}{\operatorname{var}(X^{\perp W_1})}\big(\operatorname{var}(X^{\perp W_1}) - \operatorname{var}(X^{\perp W_1})\big)\pi_1' \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
T_{32} &= -\frac{cc' \operatorname{var}(W_1)^{-1}}{1 - c' \operatorname{var}(W_1)^{-1} c} - \operatorname{var}(W_1)\pi_1\pi_1' \operatorname{var}(X^{\perp W_1})^{-1} + I \\
&\quad + \frac{cc' \operatorname{var}(W_1)^{-1}}{1 - c' \operatorname{var}(W_1)^{-1} c} + \operatorname{var}(W_1)\pi_1\pi_1' \operatorname{var}(X^{\perp W_1})^{-1} \\
&= I
\end{aligned}
$$

$$
\begin{aligned}
T_{33} &= -\frac{c' \operatorname{var}(W_1)^{-1}}{1 - c' \operatorname{var}(W_1)^{-1} c} - \pi_1' c\pi_1' \operatorname{var}(X^{\perp W_1})^{-1} + c' \operatorname{var}(W_1)^{-1} \\
&\quad + \frac{c' \operatorname{var}(W_1)^{-1} cc' \operatorname{var}(W_1)^{-1}}{1 - c' \operatorname{var}(W_1)^{-1} c} + c'\pi_1\pi_1' \operatorname{var}(X^{\perp W_1})^{-1} \\
&= c' \operatorname{var}(W_1)^{-1} \left(\frac{-1}{1 - c' \operatorname{var}(W_1)^{-1} c} + 1 + \frac{c' \operatorname{var}(W_1)^{-1} c}{1 - c' \operatorname{var}(W_1)^{-1} c} \right) \\
&\quad + \pi_1' c\big(\pi_1' \operatorname{var}(X^{\perp W_1})^{-1} - \pi_1' \operatorname{var}(X^{\perp W_1})^{-1}\big) \\
&= c' \operatorname{var}(W_1)^{-1} \left(-1 + 1 - c' \operatorname{var}(W_1)^{-1} c + c' \operatorname{var}(W_1)^{-1} c\right) \frac{1}{1 - c' \operatorname{var}(W_1)^{-1} c} \\
&= 0.
\end{aligned}
$$

Therefore $T_3 = g_1'$. Putting everything together gives

$$\mathbb{L}(Y \mid 1, X, W_1, W_2) = \beta_{\text{long}} X + g_1' W_1 + g_2 W_2 + \text{const.}$$

as desired.

To show that equation (15) holds, we have

$$\begin{aligned}
&\text{var}(Y) - \begin{pmatrix} \beta_{\text{long}}\pi_1' c + g_1' c + g_2 & \text{cov}(Y,X) & \text{cov}(Y,W_1) \end{pmatrix} \times \\
&\qquad\qquad \begin{pmatrix} 1 & \pi_1' c & c' \\ \pi_1' c & \text{var}(X) & \text{cov}(X, W_1) \\ c & \text{cov}(W_1, X) & \text{var}(W_1) \end{pmatrix}^{-1} \begin{pmatrix} \beta_{\text{long}}\pi_1' c + g_1' c + g_2 \\ \text{cov}(Y,X) \\ \text{cov}(W_1, Y) \end{pmatrix} \\
&= \text{var}(Y) - \begin{pmatrix} g_2 & \beta_{\text{long}} & g_1' \end{pmatrix} \begin{pmatrix} \beta_{\text{long}}\pi_1' c + g_1' c + g_2 \\ \text{cov}(Y,X) \\ \text{cov}(W_1, Y) \end{pmatrix} \\
&= \text{var}(Y) - \begin{pmatrix} g_2 & \beta_{\text{long}} & g_1' \end{pmatrix} \begin{pmatrix} \beta_{\text{long}}\pi_1' c + g_1' c + g_2 \\ \beta_{\text{long}} \text{var}(X) + g_1' \text{cov}(W_1, X) + g_2 \pi_1' c \\ \beta_{\text{long}} \text{cov}(W_1, X) + \text{var}(W_1) g_1 + g_2 c \end{pmatrix} \\
&= \text{var}(Y) - \beta_{\text{long}}^2 \text{var}(X) - g_1' \text{var}(W_1) g_1 - g_2^2 \\
&\qquad - 2\beta_{\text{long}} \text{cov}(X, W_1) g_1 - 2 g_1' \text{cov}(W_1, W_2) g_2 - 2\beta_{\text{long}} g_2 \pi_1' \text{cov}(W_1, W_2) \\
&= \text{var}(Y) - \beta_{\text{long}}^2 \text{var}(X) - g_1' \text{var}(W_1) g_1 - g_2^2 - 2\beta_{\text{long}} \pi_1' \text{var}(W_1) g_1 - 2 g_1' c g_2 - 2\beta_{\text{long}} g_2 \pi_1' c \\
&= \text{var}(Y) - \beta_{\text{long}}^2 \text{var}(X) - g_1' \text{var}(W_1) g_1 - g_2^2 - 2\beta_{\text{long}} \pi_1' \text{var}(W_1) g_1 \\
&\qquad - 2 g_1' (\text{cov}(W_1, Y) - \beta_{\text{long}} \text{cov}(W_1, X) - \text{var}(W_1) g_1) \\
&\qquad - 2\beta_{\text{long}} \pi_1' (\text{cov}(W_1, Y) - \beta_{\text{long}} \text{cov}(W_1, X) - \text{var}(W_1) g_1) \\
&= \text{var}(Y) - \beta_{\text{long}}^2 \text{var}(X) - g_1' \text{var}(W_1) g_1 - 2\beta_{\text{long}} \text{cov}(X, W_1) g_1 \\
&\qquad - 2 g_1' \text{cov}(W_1, Y) + 2 g_1' \beta_{\text{long}} \text{cov}(W_1, X) + 2 g_1' \text{var}(W_1) g_1 \\
&\qquad - 2\beta_{\text{long}} \pi_1' \text{cov}(W_1, Y) + 2\beta_{\text{long}}^2 \pi_1' \text{cov}(W_1, X) + 2\beta_{\text{long}} \pi_1' \text{var}(W_1) g_1 - g_2^2 \\
&= \text{var}(Y) - \beta_{\text{long}}^2 \text{var}(X) + g_1' \text{var}(W_1) g_1 - 2 g_1' \text{cov}(W_1, Y) \\
&\qquad - 2\beta_{\text{long}} \pi_1' \text{cov}(W_1, Y) + 2\beta_{\text{long}} g_1' \text{cov}(W_1, X) + 2\beta_{\text{long}}^2 \pi_1' \text{cov}(W_1, X) - g_2^2 \\
&= U(g_1) - g_2^2.
\end{aligned}$$

The first equality follows by the same derivations as we used to show equation (16). The second equality follows from two applications of equation (14): We directly use equation (14) to get an expression for $\text{cov}(W_1, Y)$. We also use it indirectly to get an expression for $\text{cov}(Y, X)$. To see this, recall that equation (14) is

$$g_2 c = \text{cov}(W_1, Y) - \beta_{\text{long}} \text{cov}(W_1, X) - \text{var}(W_1) g_1. \tag{14}$$

Since

$$\beta_{\text{long}} = \frac{\text{cov}(Y, X) - \text{cov}(Y, W_1)\pi_1}{\text{var}(X) - \text{cov}(X, W_1)\pi_1}$$

we can rearrange this equation and substitute in equation (14) to get

$$\begin{aligned}
\text{cov}(Y, X) &= \beta_{\text{long}}(\text{var}(X) - \text{cov}(X, W_1)\pi_1) + \pi_1' \text{cov}(W_1, Y) \\
&= \beta_{\text{long}}(\text{var}(X) - \text{cov}(X, W_1)\pi_1) + \pi_1'(g_2 c + \beta_{\text{long}} \text{cov}(W_1, X) + \text{var}(W_1) g_1) \\
&= \beta_{\text{long}} \text{var}(X) + g_1' \text{cov}(W_1, X) + g_2 \pi_1' c.
\end{aligned}$$

Thus the second equality above holds. The final equality above follows from the definition of $U(g_1)$.