

NBER WORKING PAPER SERIES

UNOBSERVABLE SELECTION AND COEFFICIENT STABILITY:  
THEORY AND VALIDATION

Emily Oster

Working Paper 19054  
<http://www.nber.org/papers/w19054>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
May 2013

Ling Zhong, Unika Shrestha, Damian Kozbur, Guillaume Pouliot, David Birke and Angela Li provided excellent research assistance. I thank David Cesarini, Raj Chetty, Todd Elder, Amy Finkelstein, Larry Katz, Matt Gentzkow, Matt Notowidigdo, Chad Syverson, Manisha Shah, Azeem Shaikh, Jesse Shapiro, Bryce Steinberg, Matt Taddy, Heidi Williams and participants in seminar at Brown, University of Chicago Booth School of Business, Wharton and Yale for helpful comments. I am grateful to a number of authors for providing R-squared values from their work and to Amir Sufi for re-running analyses on request. I gratefully acknowledge financial support from the Neubauer Family. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2013 by Emily Oster. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Unobservable Selection and Coefficient Stability: Theory and Validation  
Emily Oster  
NBER Working Paper No. 19054  
May 2013, Revised January 2014  
JEL No. C01,I1,I12

### **ABSTRACT**

A common heuristic for evaluating robustness of results to omitted variable bias is to look at coefficient movements after inclusion of controls. This heuristic is informative only if selection on observables is proportional to selection on unobservables. I formalize this link, drawing on theory in Altonji, Elder and Taber (2005) and show how, with this assumption, coefficient movements, along with movements in R-squared values, can be used to calculate omitted variable bias. I discuss empirical implementation and describe a formal bounding argument to replace the coefficient movement heuristic. I show two validation exercises suggesting that this bounding argument would perform well empirically. I discuss application of this procedure to a large set of publications in economics, and use evidence from randomized studies to draw guidelines as to appropriate bounding values.

Emily Oster  
University of Chicago  
Booth School of Business  
5807 South Woodlawn Ave  
Chicago, IL 60637  
and NBER  
eoster@uchicago.edu

# 1 Introduction

Concerns about omitted variable bias are common to most or all non-experimental empirical work in economics, other social sciences and the natural sciences. And although randomized experiments are common in natural sciences and are becoming increasingly common within economics, the majority of empirical work in both settings is still not randomized.<sup>1</sup> Within economics, a common heuristic for evaluating the robustness of a result to omitted variable bias concerns is to look at the sensitivity of the treatment effect to added controls. In a review of non-structural, non-experimental empirical work published in three general interest economics journals<sup>2</sup> in 2012, 75% of papers explored the sensitivity of the results to varying control sets, and a number of these papers were quite explicit about the relationship between coefficient stability and omitted variable bias.<sup>3</sup>

The baseline assumptions underlying the linear model do not imply that coefficient movements are informative about bias. The heuristic is informative only with the added assumption that the selection on observables variables is informative about the selection on unobservables (Altonji, Elder and Taber (2005); Murphy and Topel (1990)). This connection is rarely made explicit in empirical work and the underlying assumption is generally untested.

The goal in this paper is to develop a formal statement of robustness related to the coefficient stability heuristic. I formalize the link between coefficient stability and omitted variable bias through the proportional selection on observed and unobserved variables. I develop implementation guidelines and suggest that robustness can be described by an identified set of the treatment effect. I perform two validation examples - one based on constructed data and one linking possibly biased observational relationships to external causal estimates - which suggest this procedure performs well. Finally, I apply this adjustment to a set of papers in economics and use insights from randomized data to suggest bounds for constructing the identified set.

A key result is that coefficient movements alone are not sufficient to learn about bias. It is also necessary to take into account movements in the R-squared values when controls are added. Small coefficient movements with small R-squared movements can imply more bias than large movements in both. In the analysis of the economics literature, I show significant overlap in the distribution of coefficient movements between result which do and do not survive the full robustness adjustment.

I begin in Section 2 with the formal theory, which adopts the setup and assumption of Altonji, Elder and Taber (2005) (hence, AET). I consider the following model:  $Y = \beta X + W_1 + W_2$ , where  $W_1$  is observed and  $W_2$  is unobserved and the coefficient of interest is  $\beta$ . Note that  $\beta$  cannot be recovered from regression

---

<sup>1</sup>For example: in 2012 *JAMA* published 133 major research papers, only 53 of which were randomized. The *American Journal of Public Health* published 128, only 14 of which were randomized. The combination of the *American Economic Review*, the *Quarterly Journal of Economics* and the *Journal of Political Economy* published 69 empirical, non-structural papers, only 11 of which were randomized.

<sup>2</sup>*American Economic Review*, *Journal of Political Economy* and *Quarterly Journal of Economics*.

<sup>3</sup>For example, Chiappori et al (2012) state: "It is reassuring that the estimates are very similar in the standard and the augmented specifications, indicating that our results are unlikely to be driven by omitted variables bias." Similarly, Lacetera et al (2012) state: "These controls do not change the coefficient estimates meaningfully, and the stability of the estimates from columns 4 through 7 suggests that controlling for the model and age of the car accounts for most of the relevant selection."

because of the unobserved elements in the model. I describe the *proportional selection assumption*, which formally links the relationship between  $X$  and the observed variables to the relationship between  $X$  and the unobserved variables. This link invokes a degree of proportionality, denoted  $\delta$ .

Under this assumption I show that  $\beta$  can be recovered from: (1) the coefficients on  $X$  with and without controls for observed variables; (2) the R-squared values from controlled and uncontrolled regressions; and (3) a value for  $\delta$ . The result shows that coefficient movements *do* relate to omitted variable bias, but they must be scaled by movements in the R-squared values.

Section 3 turns to implementation. Performing the adjustment described requires the coefficient and R-squared values from regressions, which are recoverable, but also an assumption about  $\delta$ . A key practical issue is that, because the model specified has no error term,  $\delta$  above captures both the relative importance of the unobserved variables related to  $X$  and the idiosyncratic variation in  $Y$ . This value has no clear analog in the data.

In the first part of Section 3, I redefine the model as  $Y = \beta X + W_1 + \tilde{W}_2 + \epsilon$ , where  $\epsilon$  is orthogonal to  $X$ ,  $W_1$  and  $\tilde{W}_2$ . This unpacks  $\delta$  into two components: (1)  $\tilde{\delta}$ , the proportional selection between  $W_1$  and the unobservables related to  $X$  ( $\tilde{W}_2$ ) and (2) the R-squared of the full regression with controls for  $X$ ,  $W_1$  and  $\tilde{W}_2$ , which I denote  $R_{max}$ .

The advantage to this unpacking is that it allows us to use knowledge about the measurement error or idiosyncratic variation in  $Y$  to constrain the problem. Consider estimating the wage returns to education in a setting where omitted family background is a concern, and imagine two possible wage measures: this week's wages, or the average weekly wage over the last year. As an overall measure of earning power, wages measured based on a single week have more measurement error. Without separating  $\delta$  into components we know only that it should be lower for the single-week wages. By separating out the components, we can use knowledge of the relative degree of measurement error to define the two  $R_{max}$  values, and then note that  $\tilde{\delta}$  should actually be the same in the two analyses.

Section 3 provides a formal way to describe robustness of results under this model. I show, using logic akin to the partial identification literature, that by assuming bounding values for  $R_{max}$  and  $\tilde{\delta}$  one can generate an identified set for the treatment effect. This set can then be subject to robustness questions: Does it contain zero? Does it contain values far from what we would conclude based on the controlled effect? A key consideration is bounding values for  $\tilde{\delta}$  and  $R_{max}$ . I argue that  $\tilde{\delta} \in [0, 1]$  is a plausible bound on the degree of selection, and note that  $R_{max}$  is necessarily bounded between the controlled R-squared and 1, although knowledge of the problem may suggest a bounding value less than 1.

Section 3 provides some simulation evidence on the performance of this adjustment, primarily illustrating the importance of the R-squared movements. What this simulation does not answer is the question of how this adjustment performs in non-simulated data or whether the identified set logic would improve

inference. I evaluate the performance of this adjustment in two ways in Section 4.

First, I use NLSY data to construct a dataset relating education and wages; the data is constructed such that we know the “true” treatment effect based on a full set of family background controls. I then evaluate the performance of this adjustment by excluding combinations of controls from the “observed” set. I estimate the value of  $\tilde{\delta}$  which would be produced by each excluded set and calculate the bias-adjusted treatment effect. I show that 83% of the  $\tilde{\delta}$  values are in the range from 0 to 1, the bounds suggested in Section 3, meaning that in 83% of cases the identified set would include the true effect. This may actually undervalue this performance as the control set is selected here at random rather than based on using the most important controls first, as would be common in practice. The data is constructed with a realistic amount of idiosyncratic variation in the outcome, and I demonstrate that ignoring this variation (i.e. assuming the R-squared in the full model is 1) would lead to a flawed adjustment.

In a second test I use NLSY data to estimate the relationship between maternal behavior and child outcomes. A major concern in this setting is the confound by socioeconomic status. I match the possibly biased observational estimates with external evidence of the causal effect from randomized data or comprehensive meta-analyses (this is close in spirit to Lalonde (1986)).<sup>4</sup> I then ask whether the proportional selection adjustment would separate true from false associations within these possibly biased observational correlations. I find that the adjustment performs well: The identified set under this adjustment excludes zero in the two cases where external evidence confirms a link, and overlaps with zero in the three cases where the external evidence rejects the link.

The evidence in this section suggests this adjustment would improve inference. Although these settings by no means capture all of the areas in which coefficient stability heuristics are used, they are helpful largely because the basic empirical issue – pro-social behavior linked with socioeconomic status – is shared by many settings within economics.

In the final section of the paper I turn to the application of this procedure to the economics literature. I focus on two questions. First: How do stability statements in published papers in economics hold up to a version of this adjustment? Second: Is it possible to generate tighter bounds on  $R_{max}$ ? This latter issue is crucial since the evidence in Section 4 suggests that the bound of  $R_{max} = 1$  will lead us to conclude too few results are robust.

I begin with a sample of papers in the *American Economic Review*, *Journal of Political Economy*, *Quarterly Journal of Economics* and *Econometrica*, published between 2008 and 2013 and satisfying citation cutoffs ( $>20$  for 2008-2010 and  $>10$  for 2011-2013). I extract all relationships for which a coefficient stability heuristic is reported (58 papers; 134 results). I consider the sensitivity of these results to the proportional selection adjustment, with  $\tilde{\delta} \in [0, 1]$  and varying bounds on  $R_{max}$ . My primary definition of robustness is

---

<sup>4</sup>I also look at sibling fixed effects estimates within this dataset as another way to fully control for family background, with identical results.

whether the identified set includes zero; this is estimated on the set of relationships for which inclusion of controls moves the effect toward zero. I also consider an auxiliary definition based on whether the identified set falls fully within some standard error bounds of the controlled effect.

Under either notion of robustness, about 40% of results are robust to a value of  $R_{max} = 1$ . I show other bounds on  $R_{max}$  which are a function of the fully controlled R-squared. These capture the idea that there is variation in how predictable outcomes are, and this variation can be roughly inferred from how much is predicted by the observables. Denoting the fully controlled R-squared as  $\tilde{R}$ , I explore robustness to  $R_{max} = \Pi\tilde{R}$ , with varying values of  $\Pi$ . About 50% of results are robust to a value of  $\Pi = 3$ , and 65% to a value of  $\Pi = 2$ .

This shows considerable variation in the robustness of these stability claims, but does not suggest a bounding value. For that, I turn to randomized results. The claim that the coefficient is unchanged by inclusion of controls implicitly suggests that the treatment is assigned as if randomly. If that is the case, then the coefficient movement should be within the bounds we would expect to occur if treatment were randomized. Fortunately for this exercise, it is common in randomized papers to show coefficients with and without controls, either as a balancing test or to increase precision.

I draw a sample of all randomized papers from the *American Economic Review*, *Journal of Political Economy*, *Quarterly Journal of Economics*, *Econometrica* and *American Economic Journal: Applied Economics* between 2008 and 2013 which report coefficients with and without controls (33 papers; 76 results). Randomized data is more robust to this adjustment than non-randomized, but I show that assuming  $R_{max} = 1$  would still lead to rejection of 30% of randomized results. I derive cutoffs based on values of  $\Pi$  which would allow 95% of randomized results to survive: for the primary robustness definition, this value is  $\Pi = 2.2$ .

This provides a full robustness reporting standard. Showing that the identified set with bounds  $\tilde{\delta} \in [0, 1]$  and  $R_{max} \in [\tilde{R}, \min\{2.2\tilde{R}, 1\}]$  excludes zero would suggest robustness in the range of what would be seen if the treatment were randomized. For results where inclusion of controls moves the coefficient away from zero, this identified set may be evaluated based on whether it leads to similar magnitude conclusions.

In the full sample of non-randomized results considered, about 65% would survive this bounding robustness argument. I show considerable overlap in coefficient movements across groups which do and do not survive this robustness standard, indicating that additional information is provided by this calculation over an approach that explicitly calculates coefficient movements. I conclude this section by discussing three example papers and showing how the conclusions would be altered (or not) by this adjustment.

## 2 Theory

Consider the regression model

$$Y = \beta X + W_1 + W_2 \quad (1)$$

$X$  represents the treatment and the coefficient of interest is  $\beta$ .  $W_1$  and  $W_2$  represent confounders. Specifically,  $W_1$  is a vector which is a linear combination of observed control variables  $w_j^o$  multiplied by their true coefficients:  $W_1 = \sum_{j=1}^{J_o} w_j^o \gamma_j^o$ .  $W_2$  is a vector which is a linear combination of *unobserved* control variables  $w_j^u$ , again multiplied by their true coefficients:  $W_2 = \sum_{j=1}^{J_u} w_j^u \gamma_j^u$ . Note that  $W_2$  may contain some components which are orthogonal to  $X$ , including any measurement error in  $Y$ .

I assume that  $Cov(W_1, W_2) = 0$  and, without loss of generality, that  $Var(X) = 1$ . The assumption of orthogonality between  $W_1$  and  $W_2$  is discussed in more detail below. The covariance matrix associated with the vector  $[X, W_1, W_2]'$  is positive definite. Note that without further assumptions on the relationship between  $X$ ,  $W_1$  and  $W_2$  there is no information provided about the bias associated with  $W_2$  by seeing the bias from  $W_1$ .

Define the proportional selection relationship as  $\delta \frac{\sigma_{1X}}{\sigma_{11}} = \frac{\sigma_{2X}}{\sigma_{22}}$ , where  $\sigma_{iX} = Cov(W_i, X)$ ,  $\sigma_{ii} = Var(W_i)$  and  $\delta$  is the coefficient of proportionality. I assume that  $\delta > 0$  and refer to this as the *proportional selection assumption*. This implies that the relationship between  $X$  and the vector containing the observables is informative about the relationship between  $X$  and the vector containing the unobservables.

Define the coefficient resulting from the short regression of  $Y$  on  $X$  as  $\mathring{\beta}$  and the R-squared from that regression as  $\mathring{R}$ . Define the coefficient from the intermediate regression of  $Y$  on  $X$  and  $W_1$  as  $\tilde{\beta}$  and the R-squared as  $\tilde{R}$ . Note these are in-sample values.

The omitted variable bias on  $\mathring{\beta}$  and  $\tilde{\beta}$  is controlled by the auxiliary regressions of (1)  $W_1$  on  $X$ ; (2)  $W_2$  on  $X$ ; and (3)  $W_2$  on  $X$  and  $W_1$ . Denote the in-sample coefficient on  $X$  from regressions of  $W_1$  and  $W_2$  on  $X$  as  $\hat{\lambda}_{W_1|X}$  and  $\hat{\lambda}_{W_2|X}$ , respectively and the coefficient on  $X$  from a regression of  $W_2$  on  $X$  and  $W_1$  as  $\hat{\lambda}_{W_2|X, W_1}$ . Denote the population analogs of these values  $\lambda_{W_1|X}$ ,  $\lambda_{W_2|X}$  and  $\lambda_{W_2|X, W_1}$ .

All estimates are implicitly indexed by  $n$ . Probability limits are taken as  $n$  approaches infinity. All observations are independent and identically distributed according to model (1). By standard omitted variable bias formulas, I can express the probability limits of the short and intermediate regression coefficients in terms of these values:

$$\begin{aligned} \mathring{\beta} &\xrightarrow{p} \beta + \lambda_{W_1|X} + \lambda_{W_2|X} \\ \tilde{\beta} &\xrightarrow{p} \beta + \lambda_{W_2|X, W_1} \end{aligned}$$

Lemma 1 defines the probability limit of the coefficient difference.

**Lemma 1.**  $(\mathring{\beta} - \tilde{\beta}) \xrightarrow{p} \sigma_{1X} \frac{\sigma_{11}^2 - \sigma_{1X}^2 (\delta \sigma_{22} + \sigma_{11})}{\sigma_{11} (\sigma_{11} - \sigma_{1X}^2)}$

*Proof.* This follows directly from the probability limits of the auxiliary regression coefficients under the proportional selection assumption. Proof details are in Appendix A.  $\square$

Denote the sample variance of  $Y$  as  $\hat{\sigma}_{yy}$  and note that  $\hat{\sigma}_{yy} \xrightarrow{p} \sigma_{yy}$ . Lemma 2 defines probability limits for functions of the R-squared values.

**Lemma 2.**  $(\tilde{R} - \hat{R})\hat{\sigma}_{yy} \xrightarrow{p} \frac{[\sigma_{11}^2 - \sigma_{1X}^2(\sigma_{11} + \delta\sigma_{22})]^2}{\sigma_{11}^2(\sigma_{11} - \sigma_{1X}^2)}$  and  $(1 - \tilde{R})\hat{\sigma}_{yy} \xrightarrow{p} \frac{\sigma_{22}[\sigma_{11}^2 - \sigma_{1X}^2(\sigma_{11} + \delta^2\sigma_{22})]}{\sigma_{11}(\sigma_{11} - \sigma_{1X}^2)}$ .

*Proof.* This follows directly from the auxiliary regression coefficients and Lemma 1. Proof details are in Appendix A.  $\square$

Define the following:

$$\beta^* = \begin{cases} \tilde{\beta} - \left[ \hat{\beta} - \tilde{\beta} \right] \frac{1 - \tilde{R}}{\tilde{R} - \hat{R}} & \text{if } \delta = 1 \\ \tilde{\beta} - \left[ \frac{\sqrt{[\hat{\beta} - \tilde{\beta}]^2 [\Theta^2 + \Theta(4\delta(1-\delta)[\hat{\beta} - \tilde{\beta}]^2 [1 - \tilde{R}])]} - \Theta[\hat{\beta} - \tilde{\beta}]}{2(1-\delta)[\hat{\beta} - \tilde{\beta}]^2 [\tilde{R} - \hat{R}]} \right] & \text{if } \delta \neq 1, \sigma_{1X} \geq 0 \\ \tilde{\beta} - \left[ \frac{-\sqrt{[\hat{\beta} - \tilde{\beta}]^2 [\Theta^2 + \Theta(4\delta(1-\delta)[\hat{\beta} - \tilde{\beta}]^2 [1 - \tilde{R}])]} - \Theta[\hat{\beta} - \tilde{\beta}]}{2(1-\delta)[\hat{\beta} - \tilde{\beta}]^2 [\tilde{R} - \hat{R}]} \right] & \text{if } \delta \neq 1, \sigma_{1X} < 0 \end{cases}$$

where  $\Theta = \left( [\tilde{R} - \hat{R}]^2 \hat{\sigma}_{yy} + [\hat{\beta} - \tilde{\beta}]^2 [\tilde{R} - \hat{R}] \right)$ .

**Proposition 1.**  $\beta^* \xrightarrow{p} \beta$ .

*Proof.* I outline the proof here, with details in Appendix A. Recall that the bias of interest to calculate is  $\hat{\lambda}_{W_2|X, W_1}$  which, under the proportional selection assumption and by Lemma 1, converges in probability to  $\frac{\delta\sigma_{22}\sigma_{1X}}{\sigma_{11} - \sigma_{1X}^2}$ .  $\delta$  is assumed to be known so the unknown variables are  $\sigma_{11}$ ,  $\sigma_{22}$  and  $\sigma_{1X}$ .

By Lemmas 1 and 2 we have:

$$\begin{aligned} \hat{\beta} - \tilde{\beta} &\xrightarrow{p} \sigma_{1X} \frac{\sigma_{11}^2 - \sigma_{1X}^2(\delta\sigma_{22} + \sigma_{11})}{\sigma_{11}(\sigma_{11} - \sigma_{1X}^2)} \\ [\tilde{R} - \hat{R}] \hat{\sigma}_{yy} &\xrightarrow{p} \frac{[\sigma_{11}^2 - \sigma_{1X}^2(\sigma_{11} + \delta\sigma_{22})]^2}{\sigma_{11}^2(\sigma_{11} - \sigma_{1X}^2)} \\ [1 - \tilde{R}] \hat{\sigma}_{yy} &\xrightarrow{p} \frac{\sigma_{22} [\sigma_{11}^2 - \sigma_{1X}^2(\sigma_{11} + \delta^2\sigma_{22})]}{\sigma_{11}(\sigma_{11} - \sigma_{1X}^2)} \end{aligned}$$

This defines a system of three equations in the three unknowns of interest. Solving this system completes the proof.  $\square$

For values of  $\delta$  close to 1, the simple expression  $\tilde{\beta} - \delta \left[ \hat{\beta} - \tilde{\beta} \right] \frac{1 - \tilde{R}}{\tilde{R} - \hat{R}}$  will be an approximation for  $\beta$ . The exact value diverges from this as  $\delta$  gets significantly larger than 1. A standard error for this estimator can be calculated with a bootstrap.



It is worth briefly discussing a central assumption in deriving these results – namely, that  $W_1$  and  $W_2$  are orthogonal. This is conceptually somewhat at odds with the idea that these are “related” observables and unobservables. In practice, this assumption can be generated mechanically, but it does influence how we should think about the proportionality condition.

Consider the example where  $W_1$  and  $W_2$  together capture a full picture of socioeconomic status, and  $W_1$  contains standard demographic controls – education categories, income categories, race. The assumption is, then, that  $W_2$  captures everything else that we do not see – details about education and income, IQ and achievement measures, etc. If we simply observed these variables they would be by definition correlated with the components of  $W_1$ . What  $W_2$  actually captures is the variables *after they are residualized* with respect to the elements in  $W_1$ .

This does not pose an issue for the implementation procedure I discuss below – in running the intermediate regression and controlling for the components of  $W_1$  the relevant omitted information is by definition orthogonal. However, when we consider the relative importance of the observed and unobserved variables in explaining variation in  $X$ , it is important to keep in mind that the unobservable set is residualized.

The setup and assumptions in this section are drawn directly from AET. The observation that the bias is proportional to  $\frac{\sigma_{22}\sigma_{1,X}}{\sigma_{11}-\sigma_{1,X}^2}$  is echoed in their work. The discussion here differs in two ways. First, I show explicitly the relationship between bias and coefficient movements. Second, the estimator they derive is consistent only if  $\beta = 0$  or  $\delta = 1$ . The more complicated formulas above arise from the bias on  $W_1$  in the intermediate regression. Further discussion is in Appendix A.2.

### 3 Implementation

This section develops an empirically tractable robustness concept based on the theory in Section 2.

#### 3.1 Components of $\delta$

In Section 2,  $W_2$  in the full model is defined to capture all of the residual variation in  $Y$  which is not explained by  $X$  and  $W_1$ . This includes some components which are related to  $X$  and  $W_1$  through proportional selection.  $W_2$  may also include components which are orthogonal to  $X$  or are related to  $X$  in a way about which  $W_1$  is not informative.

Focus on the simplest case where there is some idiosyncratic variation (say, measurement error) in  $Y$ . The proportionality value,  $\delta$ , defined as  $\delta \frac{\sigma_{1X}}{\sigma_{11}} = \frac{\sigma_{2X}}{\sigma_{22}}$  is sensitive to the degree of this measurement error<sup>5</sup> and the relationship between  $W_1$ ,  $W_2$  and  $X$ . It is possible to unpack  $\delta$  into these two elements.

---

<sup>5</sup>This is because  $\sigma_{22}$  captures measurement error in  $Y$ .

Define  $W_2 = \tilde{W}_2 + \epsilon$ , where  $Cov(X, \epsilon) = 0$ ,  $Cov(W_1, \epsilon) = 0$  and  $Cov(\tilde{W}_2, \epsilon) = 0$ . Define  $\sigma_{\tilde{2}X} = Cov(\tilde{W}_2, X)$ ,  $\sigma_{\tilde{2}\tilde{2}} = Var(\tilde{W}_2)$ . The full model can be rewritten as:

$$Y = \beta X + W_1 + \tilde{W}_2 + \epsilon \quad (2)$$

With this assumption, we can separate  $\delta$  into two components.

First, define  $\tilde{\delta}$  as the proportionality value relating  $W_1$  and  $\tilde{W}_2$ :  $\tilde{\delta} \frac{\sigma_{1X}}{\sigma_{11}} = \frac{\sigma_{\tilde{2}X}}{\sigma_{\tilde{2}\tilde{2}}}$ . This  $\tilde{\delta}$  captures how much of  $X$  is explained by the observables versus the unobservables, but in this case the unobservables contain only the omitted variables which are related to  $X$  and are proxied by  $W_1$ .

Second, define the overall R-squared of the model above - with controls for  $X$ ,  $W_1$  and  $\tilde{W}_2$  - as  $R_{max}$ . Note that  $R_{max} < 1$  as long as  $\epsilon \neq 0$ . The difference between  $R_{max}$  and 1 captures the degree of idiosyncratic variation in  $Y$ .

We can now define  $\beta^{*'}$ , a slight modification of  $\beta^*$ :

$$\beta^{*'} = \begin{cases} \tilde{\beta} - \left[ \hat{\beta} - \tilde{\beta} \right] \frac{R_{max} - \tilde{R}}{\tilde{R} - \tilde{R}} & \text{if } \tilde{\delta} = 1 \\ \tilde{\beta} - \left[ \frac{\sqrt{[\hat{\beta} - \tilde{\beta}]^2 [\Theta^2 + \Theta(4\tilde{\delta}(1-\tilde{\delta})[\hat{\beta} - \tilde{\beta}]^2 [R_{max} - \tilde{R}])]} - \Theta[\hat{\beta} - \tilde{\beta}]}{2(1-\tilde{\delta})[\hat{\beta} - \tilde{\beta}]^2 [\tilde{R} - \tilde{R}]} \right] & \text{if } \tilde{\delta} \neq 1, \sigma_{1X} \geq 0 \\ \tilde{\beta} - \left[ \frac{-\sqrt{[\hat{\beta} - \tilde{\beta}]^2 [\Theta^2 + \Theta(4\tilde{\delta}(1-\tilde{\delta})[\hat{\beta} - \tilde{\beta}]^2 [R_{max} - \tilde{R}])]} - \Theta[\hat{\beta} - \tilde{\beta}]}{2(1-\tilde{\delta})[\hat{\beta} - \tilde{\beta}]^2 [\tilde{R} - \tilde{R}]} \right] & \text{if } \tilde{\delta} \neq 1, \sigma_{1X} < 0 \end{cases}$$

with  $\Theta = \left( [\tilde{R} - \tilde{R}]^2 \hat{\sigma}_{yy} + [\hat{\beta} - \tilde{\beta}]^2 [\tilde{R} - \tilde{R}] \right)$  as in Section 2.

**Corollary 1.**  $\beta^{*'} \xrightarrow{P} \beta$ .

*Proof.* See Appendix A. □

This discussion focuses on the case where  $W_2$  contains only a portion related to  $W_1$  and an idiosyncratic component. This discussion can be extended, however, to a case where there is a second (possibly observed) factor which is orthogonal to  $W_1$  but correlated with  $X$  and  $Y$ . If we consider the education and wage example, this could capture something like sex: we would not think of sex as an important socioeconomic confound, but failing to control for it would bias  $X$ . In particular, consider the case where the full model is

$$Y = \beta X + W_1 + \tilde{W}_2 + m + \epsilon \quad (3)$$

where  $m$  is orthogonal to  $W_1$ ,  $\tilde{W}_2$  and  $\epsilon$  and the assumptions about orthogonality with  $\epsilon$  are as above.

**Corollary 2.** If  $m$  is included in the short and intermediate regression, then  $\beta^{*'} \xrightarrow{P} \beta$ . If  $m$  is not included in either regression then  $\beta^{*'} \xrightarrow{P} \bar{\beta}$  where  $\bar{\beta}$  is defined by  $Y = \bar{\beta}X + \Psi_1 W_1 + \Psi_2 \tilde{W}_2 + \bar{\epsilon}$ .

*Proof.* See Appendix A. □

This corollary simply notes that if there is a second confounding factor which is orthogonal to  $W_1$  and  $\tilde{W}_2$ , the effect  $\beta$  can be recovered in a similar way, as long as  $m$  can be observed and included in both regressions. In this case  $R_{max}$  is the R-squared we would observe if  $X$ ,  $W_1$ ,  $\tilde{W}_2$  and  $m$  were all included in the regression.

These corollaries are a simple extension of Proposition 1. If  $R_{max} = 1$  then  $\delta = \tilde{\delta}$  and  $\beta^* = \beta^{*'}$ .

Together,  $\tilde{\delta}$  and  $R_{max}$  define a single value of  $\delta$  (note that the converse is not true). In this sense, there is no additional information provided by unpacking  $\delta$  into these components. Practically, however, this separation is useful. Using this procedure to estimate  $\beta$ , or to make an argument about robustness, requires assuming a value for  $\delta$ . This value, which captures both idiosyncratic variation in the outcome and the selection relationship, does not have a natural interpretation in the data. In contrast, both  $\tilde{\delta}$  and  $R_{max}$  do. This is not to say that these values are recoverable from the data directly - they are not - but that because they have a natural interpretation it may be easier to describe and evaluate the assumptions made about them.

### 3.2 Implementation in Regression

Consider a researcher interested in estimating a treatment effect  $\beta$  of  $X$  on  $Y$  in a linear model. The researcher is concerned about confounding variable, and observes a vector of controls  $\mathbf{W}_1$ . There are a set of unobserved variables, which relate to  $Y$ ,  $X$  and an index based on  $\mathbf{W}_1$  as defined in the proportional selection assumption. Practically, the researcher is concerned that they observe only a subset of the possible omitted variables.

There are two regressions that the researcher can observe, shown in equations 4 and 5 below. The first controls only for  $X$ , the variable of interest. The second adds controls for the observed confounders. Each of these produces a coefficient on  $X$ .

$$Y = \hat{\alpha} + \hat{\beta}X + \hat{\epsilon} \tag{4}$$

$$Y = \tilde{\alpha} + \tilde{\beta}X + \Psi\mathbf{W}_1 + \tilde{\epsilon} \tag{5}$$

Referring back to the expressions for  $\beta^*$  and  $\beta^{*'}$  above, equation (4) here recovers  $\hat{\beta}$  and  $\hat{R}$ . Similarly, equation (5) recovers  $\tilde{\beta}$  and  $\tilde{R}$ .

The result in Proposition (1) shows how  $\beta$  can be recovered using these regression values and an assumption about  $\delta$ . Corollary (1) shows how  $\beta$  can be recovered using these regression values and assumptions about  $R_{max}$  and  $\tilde{\delta}$ . *Stata* code accompanying this paper performs this calculation.<sup>6</sup>

---

<sup>6</sup>The command is `psacalc` and it is available through `ssc`.

## Point Estimation

Given values for  $R_{max}$  and  $\tilde{\delta}$ , the object  $\beta$  is point identified. As I noted above, both of these objects have natural interpretations in the data and in some cases it may be reasonable to make assumption about their value.  $\beta$  will also be point identified with an assumption about  $\delta$ .

## Bounding and Robustness Calculations

In many cases the exact values of  $\tilde{\delta}$  and  $R_{max}$  will not be clear. In such cases, it may be more feasible to make robustness statements based on bounding values for these objects. In practice, discussion of coefficient movements are typically done as part of a robustness argument; the discussion here will suggest a more formal way to make similar contributions.

I conceptualize this in a partial identification framework (Tamer, 2010; Manski, 2003). Consider the estimator  $\beta^{*'}(R_{max}, \tilde{\delta})$  which is defined above and is an asymptotically consistent estimator of  $\beta$  under known values of  $R_{max}$  and  $\tilde{\delta}$ . Without any additional assumptions, I note that  $R_{max}$  is bounded between  $\tilde{R}$  (the controlled regression coefficient) and 1. Under the proportional selection assumption,  $\tilde{\delta}$  is bounded below at 0 and some arbitrary upper bound  $\bar{\delta}$ .

The estimator below delivers the identified set for  $\beta$ .

$$\Delta_S = \{\beta \in \mathbb{R} : \beta = \beta^{*'}(R_{max}, \tilde{\delta}), \text{ for some } R_{max} \in [\tilde{R}, 1] \text{ and } \delta \in [0, \bar{\delta}]\}$$

This set is bounded on one side by  $\tilde{\beta}$ , which is the value of  $\beta$  delivered when  $R_{max} = \tilde{R}$  or  $\delta = 0$  (or both). Without more assumptions, the other bound is either positive or negative infinity, since  $\bar{\delta}$  is unbounded. The insight of partial identification is that it may be possible to use additional intuition from the problem to further bound both  $R_{max}$  and  $\tilde{\delta}$  values.

Consider first the issue of bounding  $\tilde{\delta}$ . I argue that for many problems,  $\tilde{\delta} = 1$  may be a reasonable upper bound. Recall that  $\tilde{\delta}$  captures the relative importance of the index of observed and unobserved variables in explaining  $X$ . The bound of  $\tilde{\delta} = 1$  suggests the observables are at least as important as the unobservables. One reason to favor this is that researchers typically focus their data collection efforts (or their choice of regression controls) on the controls they believe *ex ante* are the most important. A second is that  $\tilde{W}_2$  is residualized with respect to  $W_1$  so, conceptually, we want to think of the omitted variables having been stripped of the portion related to the included ones. Ultimately, this is an empirical issue, and I will discuss evidence for this bound in Section 4.

In the case of  $R_{max}$  it may be possible to generate a bound smaller than 1 by, for example, considering measurement error in  $Y$  or evaluating variation in  $Y$  which cannot be related to  $X$  because it results from choices made after  $X$  is determined. Define an assumed upper bound on  $R_{max}$  as  $\overline{R_{max}}$ , with  $\overline{R_{max}} \leq 1$ .

With these two bounding assumptions the identified set is:  $\Delta_s = [\tilde{\beta}, \beta^{*'}(\overline{R_{max}}, 1)]$ .

Empirically, the question of interest in considering  $\Delta_s$  is whether the conclusions based on the full set are similar to what we would draw based on observing the controlled coefficient  $\tilde{\beta}$ . If inclusion of controls moves the coefficient toward zero, one natural question is whether the set includes zero. Regardless of the direction of movement one could ask whether the bounds of the set are outside the confidence interval on  $\tilde{\beta}$  – this effectively asks whether the magnitude conclusions based on the controlled coefficient are robust.

This suggested robustness leaves open the question of what is a reasonable  $\overline{R_{max}}$  to assume in describing the identified set. I discuss this in two specific empirical contexts in Section 4 and in more detail in the context of the economics literature in Section 5.

### 3.3 Numerical Example

Before moving to explore the performance of this adjustment in data, it is useful to illustrate the mechanisms here with numerical simulation. I will focus on illustrating the importance of the R-squared movements in addition to coefficient changes.

To begin, assume a setup in which it is known that  $\tilde{\delta} = 1$  and  $R_{max} = 1$ . Assume that a completely uncontrolled regression of  $Y$  on  $X$  yields a coefficient of 0.5 and an R-squared of 0.1. Consider three scenarios for the controlled coefficient: (1)  $\tilde{\beta} = .45$ , (2)  $\tilde{\beta} = .3$  and (3)  $\tilde{\beta} = .15$ . Relying on only the coefficient movements, the last result looks the least robust. That is, observation that the coefficient moves from 0.5 to 0.15 would likely lead to much more concern about the remaining unobservables than the move from 0.5 to 0.45. However, this information is incomplete without knowing  $\tilde{R}$ .

Figure 1 graphs, for the three coefficient movement scenarios, the distribution of bias-adjusted  $\beta$  values. I have truncated the distribution in all cases at -2.5, but used text to describe the coefficients under one extreme case where  $\tilde{R} = .11$ . This figure illustrates large overlap in the distributions. Clearly values of  $\beta > 0.15$  will be delivered only by the smaller coefficient movements (and values  $\beta > 0.3$  only by the smallest one) but below  $\beta = 0.15$  there is full overlap. At the value of  $\tilde{R} = .11$ , the larger coefficient movement implies a more negative effect, but as the R-squared movement gets arbitrarily small, all three cases could have infinitely negative adjusted  $\beta$  values. A perhaps simpler way to say this is that all three coefficient movements are fully consistent with true effect of less than zero. Such an effect would be delivered by  $\tilde{R} < .19$  with the small move,  $\tilde{R} < .46$  with the medium move, and  $\tilde{R} < .73$  with the big move. Clearly, the last is the most sensitive, but without knowing the movement in R-squared, we cannot draw conclusions.

A second example is useful to illustrate in simulated data how  $R_{max}$  and  $\tilde{\delta}$  may separately vary and what this will imply about the relationship between initial coefficient movements and the remaining bias. I simulate data on  $\{X, W_1, W_2\}$  based on assumed values of variances, covariances and means. I then construct

the outcome  $Y$  as a function of these variables:

$$Y = 0.3X + W_1 + \tilde{W}_2 + \epsilon$$

where  $\epsilon \sim N(0, 1)$  and  $Cov(X, \epsilon) = Cov(W_1, \epsilon) = Cov(\tilde{W}_2, \epsilon) = 0$ . I estimate the movements in coefficients and R-squared values when  $W_1$  is included in the regression of  $Y$  on  $X$ .

I construct several simulations which vary values of  $\tilde{\delta}$  and the variance of  $\tilde{W}_2$ . Throughout the example I assume that  $Var(W_1) = Var(X) = 1$ ,  $Cov(X, W_1) = 0.2$  and all variables ( $X, W_1$  and  $\tilde{W}_2$ ) have a mean of 1. Table 1 shows the simulated results. The first three rows use a value of  $\tilde{\delta} = 1$  and vary  $Var(\tilde{W}_2)$ ; the bottom three use a value of  $\tilde{\delta} = .5$  and vary  $Var(\tilde{W}_2)$ . The variation in  $\tilde{\delta}$  does not alter the intuition here, so I will focus on the case where  $\tilde{\delta} = 1$ .

This table demonstrates that the variance of  $\tilde{W}_2$  plays a huge role in how coefficient movements should be interpreted, precisely because this variation drives the difference between  $R_{max}$  and  $\tilde{R}$ , as well as how that difference compares to the difference between  $\tilde{R}$  and  $\hat{R}$ . When  $Var(\tilde{W}_2) = 1$  - the same variance as  $W_1$  - the R-squared moves about halfway toward  $R_{max}$  with inclusion of controls. In this case, if we assume  $\tilde{\delta} = 1$  we could come close to the true  $\beta$  by simply adjusting the coefficients again by the amount they move with controls.<sup>7</sup> When  $Var(\tilde{W}_2)$  is much smaller than 1 - we assume 0.1 here - the R-squared moves nearly all the way to  $R_{max}$  and the remaining bias is very small. Conversely, when  $Var(\tilde{W}_2)$  is large relative to  $Var(W_1)$  the R-squared moves only a tiny fraction of the way to  $R_{max}$  and the remaining bias is very large relative to the coefficient movements.

This last example illustrates the key example of where either ignoring R-squared movements *or* not assuming a correct value for  $R_{max}$  will cause serious errors. In this case the coefficient movement with inclusion of controls is very small, and one might be tempted to conclude the result was very robust. But because there is so much explanatory power in the unobservables, the remaining error is very large.

The final column of the table shows the calculated  $\beta^{*f}$  values, using the  $\tilde{\delta}$  and correct  $R_{max}$  and demonstrates that the mechanics of the adjustment are correct: all values are very tightly estimated around 0.3, the true coefficient used to construct the simulation.

## 4 Empirical Validation

The results above provide a way to recover an estimate of causal treatment effects under the assumption that selection on observables and unobservables is proportional. The discussion makes clear that the bias does

---

<sup>7</sup>This could actually be done without seeing the R-squared at all. This is the procedure that is suggested in Bellows and Miguel (2009) and followed in Nunn and Wantchekon (2011). It relies heavily, however, on this assumption that the variance of the observed and unobserved components are equal (this assumption is not stated in those papers, which do not discuss R-squared movements). The Nunn and Wantchekon (2011) case will be discussed in more detail later.

relate to coefficient movements in these cases, which is encouraging for common robustness calculations. The simulated data shows that this works mechanically, and illustrates the importance of these R-squared movements. However, this theoretical discussion does not provide any insight as to whether the proportional selection assumption is valid in empirical contexts and, by extension, whether the robustness calculation suggested would improve inference.

In this section I describe two exercises which explore how this adjustment performs in the data. I use two example settings which are familiar to economists. In the first example, I generate a dataset in which I know the true effect by construction and explore how this adjustment would perform when various confounds are “unobserved”. In the second example, I compute possibly biased estimates, perform the adjustment, and compare the resulting conclusions to external evidence on causal impacts. I ask whether the adjusted coefficients generate more accurate conclusions than the simple controlled estimates.

## 4.1 Constructed Data: Returns to Education

I begin with a canonical example in economics, the relationship between wages and years of education. One issue with estimates of standard Mincer regressions is the confound with family background: people whose mothers have more education, for example, are more likely to be educated but also have higher wages for other reasons.<sup>8</sup> Using data from the NLSY I construct a dataset in which I define the “true” return to education as the impact of education controlling for a full set of family background characteristics. I then consider the bias - both in simple controlled regressions and after this adjustment is performed - in hypothetical cases in which I do not observe the full set of controls. This exercise will allow me to see how the adjustment performs *and* to estimate values of  $\tilde{\delta}$  and ask how they compare to the bounds suggested in Section 3.

### 4.1.1 Data and Empirical Strategy

I use data from the NLSY-79 cohort. I am concerned with the impact of years of education on log wages, and I begin by considering the standard Mincer regression of log wages on educational attainment. I use the higher of the two educational levels recorded in 1981 and 1986 and the higher of the two wage values recorded in 1996 and 1998. Experience and experience-squared are calculated in the typical way (experience = age - education years - 6). I also control for individual sex.

My concern is with confounding by demographics and family background. I capture this with eight variables: region of residence, race, marital status, mother’s education, father’s education, mother’s occupation, father’s occupation and number of siblings. All variables are controlled for fully flexibly, with dummies. Summary statistics for these data appear in Appendix B.

---

<sup>8</sup>A second obvious issue is the confound with ability. It would be possible to do an exercise similar to this one with that confound. Since the exercise here is not about finding the causal effect of education on wages, but is simply about exploring this adjustment, there is no loss to ignoring the issue of ability.

I construct a dataset by regressing log wages on education, experience, sex and the full set of family background data. I generate fitted values, and then take these as the “true” effects in the model - that is, the effect on education we see in this regression is the unbiased treatment effect in the constructed data.

The regression of this fitted value on the full set of controls has an R-squared of 1 by construction. In practice, however, wages are not fully predicted by family background or individual characteristics. I therefore add an orthogonal error term to this fitted value. To generate a magnitude for this term I regress the log wage measure used here on log wages in 1992 or 1994 (again, I take the higher of the 2). This regression has an R-squared of about 0.45. I argue that family background, education, etc, should not explain more of the outcome than the previous year’s wages, since these variables all contribute to that wage. I therefore add an orthogonal error term to the fitted value such that the ultimate regression R-squared is about 0.45.

It is important to note that the addition of this error term is done largely for realism; it will be instructive to explore errors that may be introduced by incorrectly assuming that  $R_{max} = 1$ . However, the calculations of  $\tilde{\delta}$  is not sensitive to this addition.

Given this constructed dataset, the empirical exercise is straightforward. I iterate through excluding all sets of controls (up to 6 of the 8). In each case I: (1) calculate the  $\tilde{\delta}$  implied by the included and excluded control set; (2) calculate  $\beta^{*’}$  with this  $\tilde{\delta}$  and the true  $R_{max}$ ; (3) calculate whether the identified set bounded by  $\tilde{\beta}$  and  $\beta^{*’}(R_{max}, 1)$  contains the true effect. ; and (4) calculate  $\beta^{*’}$  with this  $\tilde{\delta}$  and the assuming  $R_{max} = 1$ .

#### 4.1.2 Results

Figure 2a shows the distributions of the true  $\beta$  and the estimated  $\tilde{\beta}$  and the values of  $\beta^{*’}$ . The true effect in the constructed data is 0.090, with a standard error of 0.003. The  $\beta^{*’}$  values cluster at the true effect value. This is a simple numerical check of the procedure in realistic data: if we know the true  $R_{max}$  and the true  $\tilde{\delta}$  the adjustment works as it should. Not surprisingly, the estimated  $\tilde{\beta}$  distribution is shifted substantially to the right from the true  $\beta$ . Controlled estimates are systematically biased to estimate excess returns to education.

Figure 2b shows the values of  $\tilde{\delta}$  calculated in this exercise. This value is not mechanical: nothing in the setup constrains any particular value of  $\tilde{\delta}$ . In the figure, I show the full distribution of  $\tilde{\delta}$  and the [0,1] bounds that I suggest would be appropriate in many settings.

The average  $\tilde{\delta}$  is 0.528 and 83% of values fall within the [0,1] range. Only 4 (of 211) values are negative. The cases with values of  $\tilde{\delta} > 1$  are instructive. These are combinations of controls where the index of the omitted variables are more important in explaining education than the included ones. Of the 32 cases with  $\tilde{\delta} > 1$ , 90% of them excluded either maternal or paternal education. This makes clear that these variables are among the most important confounds; this should not be surprising and, indeed, it seems likely that researchers would think to include these first, before considering data on parental occupation or number of siblings. Put differently, if we consider control set selection not at random as I do here but with the idea that the most



important controls are selected first, it is likely that the  $[0, 1]$  bound would be even more accurate. The fact that the average  $\tilde{\delta}$  is less than 1 supports the idea of 1 as a bound on  $\tilde{\delta}$ , rather than as an average value.

I can comment on the identified set logic described in Section 3. Given the  $\tilde{\delta}$  values, it is straightforward to observe that if we calculate the set  $[\tilde{\beta}, \beta^{*'}(R_{max}, 1)]$ , in 83% of cases the set will include the true value. In an additional 7% of cases the set will overlap with the confidence interval of the true  $\beta$ , suggesting that the error in the set is small.

In this setting, it seems that this adjustment works well. The confidence interval of controlled estimates captures the true value of  $\beta$  only 62% of the time, whereas 82% of the time the adjusted identified set using the true  $R_{max}$  and a  $\tilde{\delta} \in [0, 1]$  bound contains the true value.

As a final point, Figure 2c shows a graph similar to Figure 2a, but with values of  $\beta^{*'}$  which would be calculated with  $R_{max} = 1$  and the true  $\tilde{\delta}$ . These values are systematically biased downward (effectively, the adjustment is too large because we are mistakenly assuming the R-squared would go to 1 with the full control set) and the error can be extremely large. This suggests significant thought is necessary in considering  $R_{max}$ , an issue I take up in more detail in Section 5.

## 4.2 Observational Data: Maternal Behavior and Child Outcomes

A second approach to validation is to take a setting in which we have some possibly biased observational relationships and we think we have a sense of the true effect from external sources. Given this, the question is whether this approach can separate “true” from “false” associations.<sup>9</sup>

In this section I undertake this type of validation exercise in the context of the link between maternal behaviors, infant birth weight and child IQ. These relationships are of some interest in economics, and of wider interest in public health and public policy circles. A literature in economics demonstrates that health shocks while children are in the womb can influence early outcomes and later cognitive skills (e.g. Almond and Currie, 2011; Almond and Mazumder, 2011). A second literature, largely in epidemiology and public health, suggests that even much smaller variations in behavior – occasional drinking during pregnancy, not breastfeeding – could impact child IQ and birth weight. These latter studies, in particular, are subject to significant omitted variable concerns, largely associated with omitted socioeconomic status. I consider five relationships in all: the relationship between child IQ and breastfeeding, drinking during pregnancy, low birth weight/prematurity and the relationship between birth weight (as the outcome) and maternal drinking and smoking in pregnancy.

### 4.2.1 Data

I use NLSY data, this time from the Children and Young Adult sample, which has information on the children of NLSY participants. I measure IQ with PIAT test scores for children 4 to 8 and birth weight with birth

---

<sup>9</sup>Altonji, Elder and Taber (2008) do a version of this in a single context.

weight in grams as reported by the mother. In the latter analysis I include all children. In all cases I control for child sex and, with IQ, for their age. These are not considered as part of the confounding set.

The IQ treatments are: months of breastfeeding, any drinking of alcohol in pregnancy and an indicator for being low birth weight and premature (<2500 grams and <37 weeks of gestation). The birth weight treatments are maternal smoking and drinking intensity during pregnancy. I measure socioeconomic status, the confounding category, with child race, maternal age, maternal education, maternal income and maternal marital status. Summary statistics for these data appear in Appendix B.

### 4.2.2 Empirical Strategy

The empirical strategy here is straightforward. I run regressions with and without the socioeconomic controls to extract  $\hat{\beta}$ ,  $\hat{R}$ ,  $\tilde{\beta}$  and  $\tilde{R}$ . I adopt a bounding value for  $R_{max}$  drawn from within sibling correlations (Mazumder, 2011). In theory,  $R_{max}$  should reflect how much of the variation in child IQ and birth weight could be explained if we had full controls for family background; I argue this is the thought experiment approximated by the sibling fixed effect R-squared. The figures are 0.61 for IQ and 0.53 for birth weight.

Given this  $R_{max}$  bound, I first calculate the identified set  $[\tilde{\beta}, \beta^*(R_{max}, 1)]$ . For completeness, I do two other calculations. First, I calculate the value of  $R_{max}$  which would produce  $\beta = 0$  assuming that  $\tilde{\delta} = 1$  and compare this value to the  $R_{max}$  bound. Second, I find the value of  $\tilde{\delta}$  which would produce  $\beta = 0$  under the assumed  $R_{max}$  and compare this to  $\tilde{\delta} = 1$ . It is important to note that these latter two analyses contain the same information as the identified set.

The conclusions from these robustness calculations are compared to the conclusions we expect to get if we were able to estimate the true model. To ask whether the adjusted coefficient gets it right, we need to know what the correct answer is.

I use two types of evidence. First, I consider external evidence from randomized trials (where available) and meta-analyses. Randomized evidence suggests that breastfeeding is not linked with full-scale IQ (Kramer et al, 2008) and most evidence does not suggest an impact of occasional maternal drinking on child IQ (see, for example: Falgreen-Eriksen et al, 2012; O’Callaghan et al, 2007).<sup>10</sup> In contrast, low birth weight and prematurity do seem to be consistently linked to low IQ (Salt and Redshaw, 2006), a link which also has a biological underpinning (de Kieviet et al, 2012). Occasional maternal drinking is typically not thought to impact birth weight (Henderson, Gray and Brocklehurst, 2007), but there is better evidence that smoking does (e.g. from trials of smoking cessation programs as in Lumley et al, 2009).

Second, I consider the conclusions one would draw from sibling fixed effects regressions in the NLSY data described above, which provides a more “within sample” test of fully controlling for family background.

<sup>10</sup>Although the question of whether occasional maternal drinking lowers IQ is an issue with some controversy, as I show below the observational data here actually estimates *positive* impacts of maternal drinking on IQ, and the fact that those effects are not causal is not a subject of much debate.

Of course, sibling fixed effects estimates may be subject to their own concerns about causality, so it is perhaps comforting that the conclusions are the same from either source.

### 4.2.3 Results

Table 2 reports the results: Panel A shows results on IQ, Panel B on birth weight.

The first column shows treatment effects, standard errors and R-squared values with only sex (or age and sex in the case of IQ) as controls. Column 2 shows similar values with the full control set. More breastfeeding is associated with higher IQ in these regressions, and low birth weight is associated with lower child IQ. More maternal drinking appears in these data to be associated with *higher* child IQ later, an finding which has no biological support and is extremely likely to be due to selection. Both samples show smoking and drinking are associated with lower birth weight. All analyses reported here show significant effects with the full set of controls. Interpreting these results in a naive way, one would conclude that each has a significant link with child outcomes.

Column 3 reports whether external evidence, summarized above, suggests a causal impact. As noted, low birth weight does seem to be linked to IQ and smoking is linked to low birth weight, but the other relationships do not have broad support. Column 4 shows sibling fixed effects regressions, which show similar conclusions. The only difference is in the impact of low birth weight on child IQ, where the NLSY regression coefficient is significant only at the 11% level.

Column 5 shows the identified set, bounded by the  $R_{max}$  estimates in the top row of each panel and  $\tilde{\delta} = 1$ . This procedure performs well. The two cases in which the identified set does not include zero are those where the external evidence suggest significant results. Put differently, if one were to use the rule of accepting the effect as causal only if the identified set excluded zero, this would lead to the same conclusions as the external evidence. In all cases the identified set includes the sibling fixed effect estimates.

In Column 6 I show the value of  $R_{max}$  which would produce  $\beta = 0$ ; comparing these to the  $R_{max}$  values in the top of each panel yields the same conclusions as in Column 5. Finally, Column 7 shows that the effects confirmed in external data are those which have values of  $\tilde{\delta} > 1$  required to produce  $\beta = 0$ . Again, the conclusion are the same.

There are two final points to make about this analysis. First, similar to the wage analysis above, the value of  $\tilde{\delta}$  which matches the adjusted effects to the sibling fixed effect values is less than 1 - it is 0.42 - pointing to the value of 1 as a bound. Second, doing these calculations with a value of  $R_{max} = 1$  as the bound would lead us to reject all the associations - including the two which are confirmed in outside data.

The results in this section suggest the robustness framework suggested performs well. It also makes clear the importance of taking into account the movements in R-squared in addition to movements in coefficients. In this latter example, if we based our analysis only on the size (say, in percent terms) of the coefficient

movements we would conclude the link between drinking and low birth weight is much more robust than the link between low birth weight and IQ since the former moves only 10% and the latter 30%. In fact, the low birth weight and IQ link has more external support. This is confirmed by the identified set conclusions, and mechanically it is reflective of the much larger change in R-squared in the low birth weight - IQ relationship.

## 5 Application to Economics Literature

The evidence in the previous section suggests that a version of this approach would improve inference. Both exercises suggest that a bounding value of  $\tilde{\delta} \in [0, 1]$  performs well. Both cases also illustrate the importance of the choice of  $R_{max}$ , in particular noting that assuming  $R_{max} = 1$  when the truth is below this will lead to over-adjustment. I turn now to the application of this approach within the economics literature. I consider two questions. First: How do stability statements in published papers in economics hold up to a version of this adjustment? Second: Is it possible to say anything more about realistic bounds on  $R_{max}$ ?

I begin by analyzing stability comments made in non-randomized data within economics, and show the degree of robustness of the results to varying  $R_{max}$  cutoff. Second, I use evidence from *randomized* data within economics to develop a bound for  $R_{max}$ . The third subsection describes three examples.

### 5.1 Stability of Coefficients in Non-Randomized Data

The data for this section comes from the published literature in economics. I extract all papers in the *American Economic Review*, *Quarterly Journal of Economics*, *The Journal of Political Economy* and *Econometrica* from 2008-2010 with at least 20 citations in the ISI Web of Science, and those from 2011-2013 in the same journals with at least 10 citations. From these papers I extract all results where the researcher explores the sensitivity of the result to a control set. In cases where the R-squared is not reported (29%), I use replication files or ask the researcher. Two results were excluded due to inability to obtain R-squared values. The final sample includes 58 papers, with 134 total results.

The full set of citations used appears in Appendix D.

The empirical exercise here is straightforward. I extract coefficient and R-squared values from the paper. Note that in cases where controls are included sequentially, I compare the fewest-controls to the most-controls set. For each result, I calculate the identified set with  $\tilde{\delta} = 1$  and varying values of  $R_{max}$ .

I consider  $R_{max} = 1$  as one bound. I also consider a parametrization of  $R_{max}$  as a function of  $\tilde{R} : R_{max} = \min\{\Pi\tilde{R}, 1\}$  with varying values of  $\Pi$ . This function allows for that some outcomes have more measurement error or noise than others, and suggests that the degree of variation accounted for by the observables (including the treatment) may be informative as to the degree accounted for by the unobservables. An alternative would be to use  $R_{max} = \tilde{R} + \Pi(\tilde{R} - \hat{R})$ , which captures a similar assumption. I work through

this version in Appendix C, and show the conclusions are extremely similar.

Having calculated the identified set using these  $R_{max}$  values, I consider two standards for robustness. My primary analysis focuses on the subset of results for which the inclusion of controls moves the coefficient towards zero, and simply asks whether the set includes zero. I also consider whether the bounds of the set fall within  $\pm 2.8$  standard errors of the controlled estimate, an analysis which can be done including results where controls move the coefficient away from zero. This second standard captures a test of whether the magnitude conclusions from the controlled estimate are shared by the adjusted estimate.

The results appear in Figures 3a and 3b. Figure 3a shows the primary robustness with rejection of zero; Figure 3b uses all results and shows the magnitude test. These graphs show the share of relationships which would survive varying values of  $\Pi$ , with  $R_{max} = \min\{\Pi\tilde{R}, 1\}$ . In either case, we find about 40% of results would survive  $R_{max} = 1$ . Within the others, there is a wide distribution of robustness; some share of results would not survive even quite small differences between  $\tilde{R}$  and  $R_{max}$ .

To quantify this, Panel A of Table 3 shows the share of these results which would survive  $R_{max} = 1$  and three values of  $\Pi$ . Twenty-five to thirty percent of studies would not survive  $\Pi = 1.5$ . Considering the rejection-of-zero robustness, within this set that is not robust to  $\Pi = 1.5$ , the average study fails at a value of  $\Pi = 1.2$  or, in point estimate terms, a predicted increase in R-squared of 0.09 with inclusion of unobserved controls.

One issue in interpreting these results is that the author of these papers may not be intending these results as a test of omitted variable bias. To address this, I limit to the large subset of papers in which the authors either explicitly comment on the coefficient stability (since remaining omitted variable bias is the only reason that would matter) or explicitly comment on omitted variable bias. Within this subsample, consider the analog of Columns (1) and (3) of Table 3: 35% of these papers would survive  $R_{max} = 1$  and 63% would survive  $R_{max} = 2\tilde{R}$ . This is very similar to, if anything lower than, the overall sample, suggesting it is not the case that the papers which fail by this criteria do so because this is not the intended test.

## 5.2 Evidence on Stability Cutoffs from Randomized Data

The evidence above suggests that even within a sample of papers which argue for coefficient stability there is a lot of variation in the degree of stability. A natural following question is whether we can suggest any guidance about where one might draw the line - specifically, is there some value of  $\Pi$  above which we should consider a result robust?

I argue that one place to look for such guidance is in reports from randomized data. Randomized experiments are becoming increasingly common within economics and papers reporting results of these experiments often include regressions with and without controls. Sometimes these are explicitly used to test balance in the experiment, although it is also commonly done to increase precision. Assuming that the data is

correctly randomized, if the sample size were infinite, the effects would not be expected to move at all. In practice, with finite data, coefficients can move a bit simply due to very small differences across groups.

When non-randomized papers invoke a coefficient stability heuristic to argue the results they observe are causal, they are (perhaps implicitly) suggesting that the treatment is as good as random. Including controls doesn't change the coefficient because there is no confounding; this is exactly the argument we know holds in randomized cases. Given this, I argue we can use the stability of randomized data as a guide to how much stability we would expect in non-randomized data *if the treatment were assigned exogenously*: is the coefficient stability within the range the researcher would expect with a randomly assigned treatment?

The approach in this section is to assume effects estimated in randomized data are causal and to therefore assume that they should survive this adjustment procedure.<sup>11</sup> I then ask what value of  $\Pi$  in the  $R_{max}$  parametrization would make this true.

The baseline set of papers for this analysis is all randomized papers (lab or field) published in the *American Economic Review*, *Quarterly Journal of Economics*, *Journal of Political Economy*, *Econometrica* and the *American Economic Journal - Applied Economics* in the period 2008 through 2013.<sup>12</sup> I extract from this all papers which report sensitivity of a treatment effect to controls. In cases where there are multiple effects reported (i.e. multiple outcomes), I include all effects. In cases where R-squared values were not provided in the paper (22%), I used replication files if available and otherwise asked the authors of the papers. The one result with no R-squared values available was not included. The final sample includes 33 papers with 76 results.

The full set of references is in Appendix D.

I undertake the same analysis as in the non-randomized data: calculate the identified set under  $\tilde{\delta} = 1$  and varying  $R_{max}$  and compare the results to the two standards for robustness.

Figures 4a and 4b show the distributions of sensitivity for the randomized data. A first thing to note is that these results are more robust than the non-randomized results. I have graphed them on the same scale for comparability. Sixty to seventy percent of randomized results would survive a cutoff of  $R_{max} = 1$ . Nearly all would survive a cutoff of  $R_{max} = 2\tilde{R}$ , much greater than for the non-randomized results. Panel B of Table 3 shows the survival shares for this dataset explicitly under the varying  $R_{max}$  cutoffs.

It is not surprising that the randomized results are more robust. The fact that they do not all survive  $R_{max} = 1$  is due to the fact that even small changes in coefficient can be blown up with this assumption. To consider an example, Jensen (2012) estimates the impact of a job recruitment program on (among other outcomes) child BMI (Table 4 and Appendix B). Adding controls changes the coefficient on treatment from

<sup>11</sup>An obvious concern is that, perhaps, these papers are not correctly randomized. This would lead me to a standard which was too lax. I address this in two ways. First, I have focused on papers published in highly ranked journals, increasing the chance that the randomization was of high quality. Second, I will draw guidelines which fit 95% of papers, thus accepting that a small share of randomized papers may suffer from true lack of balance and should not be used to guide this approach.

<sup>12</sup>I include AEJ-Applied because it has published a large number of experimental papers. This journal begins in 2009.

0.24 to 0.20, a seemingly quite small difference. The R-squared moves from 0.007 to 0.017. Assuming that  $R_{max} = 1$  would produce a coefficient of -3.73. Given that this program was randomized and the randomization passes standard balancing tests, it seems implausible that this is the actual effect. Much more plausible is the possibility that even with all correlated unobservables included it would not be possible to explain close to all of the variation in child BMI.

I use these data to develop robustness cutoff values. I base these on the value of  $\Pi$  which would allow 95% of results to survive. In testing rejection of zero, this value is 2.2. This value suggests that a bound where the unobservables explain as much as the observables (where the latter includes the treatment). This is somewhat parallel to the bounding on  $\tilde{\delta}$ .

To argue for a level of stability which would be expected from a randomized treatment, non-randomized effects should show that the set  $[\tilde{\beta}, \beta^{*'}(\min\{2.2\tilde{R}, 1\}, 1)]$  excludes zero. Applying this to the non-randomized data above, I find that 64% of results would survive this standard. This set would be valuable to report even in cases where the controls cause the coefficient to move away from zero; in that case the question would be whether considering the full set would lead to very different conclusions than the controlled estimate.

A similar analysis applies to the second robustness standard, were I find the cutoff generated by the randomized data is  $\Pi = 1.5$ . This suggests an alternative way to test for robustness by showing that  $\beta^{*'}(\min\{1.5\tilde{R}, 1\}, 1)$  is within 2.8 standard errors of the controlled effect. This allows survival of 69% of non-randomized results.

A final question to ask is how the conclusions here differ from those we would draw if we relied only on coefficient movements. Figure 5 graphs the percent reduction (in absolute value) in coefficient for results which do and do not survive the  $\beta^{*'}(\min\{2.2\tilde{R}, 1\}, 1)$  cutoff (I consider only results where the coefficient moves towards zero). The smaller the coefficient movement the more robust the result (this almost has to be true given that coefficient movements are an input to the result) but the distributions are almost fully overlapping. Only at the very tails would coefficient movements alone differentiate these groups, suggesting significant additional information is provided by doing the full adjustment.

### 5.3 Examples

Before concluding, it is useful to consider the application of this cutoff value in several examples. I choose three papers from the literature which are very explicit about coefficient stability or (in one case) do calculations in the spirit of this adjustment and for which inclusion of controls reduces the coefficient. In each case I show the identified set  $[\tilde{\beta}, \beta^{*'}(\min\{2.2\tilde{R}, 1\}, 1)]$  as well as the  $R_{max}$  and  $\tilde{\delta}$  values which produce  $\beta = 0$ .

### 5.3.1 Mian and Sufi (2009)

Mian and Sufi (2009) deal with the impact of sub-prime borrowers on the mortgage crisis. In their Table V, they estimate the impact of fraction sub-prime borrowers by zip code on mortgage origination. They worry about other characteristics of the zip code, and in their most fully controlled specification control for fixed effect for three-digit zip code prefix. They comment explicitly, in the text, on the stability of the coefficients along with the increase in R-squared (this is among very few papers to comment on the latter).

Panel A of Table 4 shows the robustness calculations for this paper. To echo the data above, I consider what happens when moving from the least controlled to most controlled regression in Table (4): Column (1) to Column (4). The results in this paper is robust to this adjustment. The full identified set is quite small and does not come close to including zero. The magnitude conclusions in the paper would be quite similar if we drew any figure from the identified set.

### 5.3.2 Nunn and Wantchekon (2011)

Nunn and Wantchekon (2011) analyze the impact of slave trades on mistrust in Africa. They worry about unobserved differences across areas, and present a number of arguments to support the interpretation of their results as causal. In Section IV.B. they undertake direct calculations based on the theory in Altonji, Elder and Taber (2005). They use coefficient movements in their regressions to calculate the value of  $\tilde{\delta}$  which would be required to produce  $\beta = 0$ . They argue the results are robust because all the calculated values of  $\tilde{\delta}$  are greater than 1, a cutoff which they also favor.

However, although they do not make this explicit, the calculations they undertake - which rely only on coefficient movements - are correct only if  $R_{max} = \tilde{R} + (\tilde{R} - \hat{R})$ . In words, this assumes the unobservables explain as much of the outcome as the observables (ignoring that the treatment is part of the observables). The R-squared in their regressions do not move much; as an example, in the first row of Table 4, considering the “Trust Relatives” measure, adding controls increases R-squared from 0.115 to 0.133. Their adjustment assumes that the fully controlled R-squared would be 0.151. As I note in the discussion in Section 3.3, if the unobservable index has a large variance, this assumption of equal increase in R-squared will not hold.<sup>13</sup>

I consider how these results might change if we use the  $R_{max} = \min\{2.2\tilde{R}, 1\}$  value. Their paper has five outcomes and they show the results in several ways. I consider rows 1 and 2 of their Table 4 and recompute their analysis with the new  $R_{max}$  values. The results are shown in Table 4.

Some, but not all, of the results in their paper survive this version of the adjustment. Three of the ten  $\tilde{\delta}$  values are above 1, meaning that three of the ten identified sets continue to reject zero under this adjustment. The other seven admit zero. This suggests that the conclusion from this table, at least, would have been

---

<sup>13</sup>As noted above, Appendix C shows the analysis in the above sections using the formulation :  $R_{max} = \tilde{R} + \Pi(\tilde{R} - \hat{R})$ . The best fit value of  $\Pi$  there is 4.95, and the conclusions - in general and about this particular paper - are unchanged.



altered by taking the full adjustment - with the randomized-based cutoff for  $R_{max}$ - into account.

### 5.3.3 Spolaore and Wacziarg, 2009

Spolaore and Wacziarg (2009) estimate the impact of genetic distance on economic distance between countries. They worry about unobserved differences across country groups - geographic distance, shared continent, etc. Their Table 4, and the surrounding discussion, includes various control sets and looks at coefficient stability.

Panel C of Table 4 shows the robustness calculation for this paper. As above, I consider the comparison of the baseline results to the most-controlled column (in their case, this is with continent-pair fixed effects). The identified set here does include zero.

It is worth noting that this paper also shows three intermediate regressions with other control sets (Columns (2)-(4) of Table 4) where the coefficient is more stable (although the R-squared moves less). The adjusted sets based on these columns would exclude zero. In a case like this it may be useful to delve further into the appropriate  $R_{max}$  (perhaps using more information drawn from the setting) or think more about whether the unobservables are closer to the controls in Column (5) or to those in Columns (2) - (4). This latter point is akin to the discussion of this issue in Murphy and Topel (1990).

## 6 Conclusion

This paper develops a formal language for discussing robustness of treatment effects, related to the popular heuristic of exploring coefficient sensitivity to controls. I connect this heuristic to the assumption of proportional selection on observed and unobserved variables (Altonji, Elder and Taber (2005)). I describe an implementation strategy for generating bounds on treatment effects and show validation in two empirical contexts. Applying this to the economics literature, and drawing guidelines for expected coefficient sensitivity from randomized results, I develop a full bounding argument.

I suggest researchers report an identified set for the treatment effect  $[\tilde{\beta}, \beta^{*'}(\min\{2.2\tilde{R}, 1\}, 1)]$  where  $\beta^{*'} = \tilde{\beta} - \tilde{\delta} \frac{(\tilde{\beta} - \tilde{\beta})(R_{max} - \tilde{R})}{(\tilde{R} - \tilde{R})}$ . Showing that such a set excludes zero would indicate a level of robustness in the range of what would be seen if the treatment were randomized. For results where inclusion of controls moves the coefficient away from zero, a similar bound could be compared to the magnitude conclusions from the controlled coefficient.

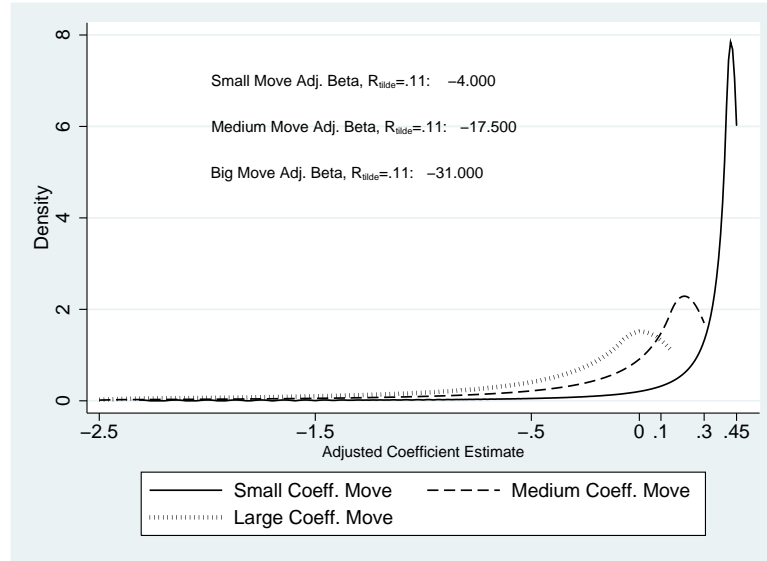
The particular  $R_{max} = \min\{2.2\tilde{R}, 1\}$  bound is a useful summary, but it is important to note that there may be many cases where one is able to be more precise about  $R_{max}$  using more knowledge of the problem. The examples in Section 4 illustrate two ways one might develop this intuition. The core insight here is that the movement in R-squared values relative to this  $R_{max}$  must be taken into account when developing any coefficient stability argument.

## References

- Almond, Douglas and Bhashkar Mazumder**, “Health Capital and the Prenatal Environment: The Effect of Ramadan Observance during Pregnancy,” *American Economic Journal: Applied Economics*, October 2011, 3 (4), 56–85.
- **and Janet Currie**, “Killing Me Softly: The Fetal Origins Hypothesis,” *Journal of Economic Perspectives*, Summer 2011, 25 (3), 153–72.
- Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber**, “Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools,” *Journal of Political Economy*, 2005, 113 (1), 151–184.
- **, Todd Elder, and Christopher R. Taber**, “Using Selection on Observed Variables to Assess Bias from Unobservables When Evaluating Swan-Ganz Catheterization,” *American Economic Review*, 2008, 98 (2), 345–50.
- Chiappori, Pierre-Andrei, Sonia Oreffice, and Climent Quintana-Domeque**, “Fatter Attraction: Anthropometric and Socioeconomic Matching on the Marriage Market,” *Journal of Political Economy*, 2012, 120 (4), 659 – 695.
- de Kieviet, J. F., L. Zoetebier, R. M. van Elburg, R. J. Vermeulen, and J. Oosterlaan**, “Brain development of very preterm and very low-birthweight children in childhood and adolescence: a meta-analysis,” *Dev Med Child Neurol*, Apr 2012, 54 (4), 313–323.
- Falgreen-Eriksen, H. L., E. L. Mortensen, T. Kilburn, M. Underbjerg, J. Bertrand, H. Stavring, T. Wimberley, J. Grove, and U. S. Kesmodel**, “The effects of low to moderate prenatal alcohol exposure in early pregnancy on IQ in 5-year-old children,” *BJOG*, Sep 2012, 119 (10), 1191–1200.
- Henderson, J., R. Gray, and P. Brocklehurst**, “Systematic review of effects of low-moderate prenatal alcohol exposure on pregnancy outcome,” *BJOG*, Mar 2007, 114 (3), 243–252.
- Jensen, Robert**, “Do Labor Market Opportunities Affect Young Women’s Work and Family Decisions? Experimental Evidence from India,” *The Quarterly Journal of Economics*, 2012, 127 (2), 753–792.
- Kramer, M. S., F. Aboud, E. Mironova et al.**, “Breastfeeding and child cognitive development: new evidence from a large randomized trial,” *Arch. Gen. Psychiatry*, May 2008, 65 (5), 578–584.
- Lacetera, Nicola, Devin G. Pope, and Justin R. Sydnor**, “Heuristic Thinking and Limited Attention in the Car Market,” *American Economic Review*, August 2012, 102 (5), 2206–36.
- LaLonde, Robert J**, “Evaluating the Econometric Evaluations of Training Programs with Experimental Data,” *American Economic Review*, September 1986, 76 (4), 604–20.
- Lumley, J., C. Chamberlain, T. Dowswell, S. Oliver, L. Oakley, and L. Watson**, “Interventions for promoting smoking cessation during pregnancy,” *Cochrane Database Syst Rev*, 2009, (3), CD001055.
- Manski, C.F.**, *Partial Identification of Probability Distributions* Springer Series in Statistics, Springer, 2003.
- Mazumder, Bhashkar**, “Family and Community Influences on Health and Socioeconomic Status: Sibling Correlations Over the Life Course,” *The B.E. Journal of Economic Analysis & Policy*, 2011, 11 (3), 1.
- Mian, Atif and Amir Sufi**, “The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis,” *The Quarterly Journal of Economics*, 2009, 124 (4), 1449–1496.
- Murphy, Kevin and Robert Topel**, “Efficiency Wages Reconsidered: Theory and Evidence,” in “Advances in the Theory and Measurement of Unemployment” 1990, pp. 204–240.

- Nunn, Nathan and Leonard Wantchekon**, “The Slave Trade and the Origins of Mistrust in Africa,” *American Economic Review*, 2011, *101* (7), 3221–3252.
- O’Callaghan, F. V., M. O’Callaghan, J. M. Najman, G. M. Williams, and W. Bor**, “Prenatal alcohol exposure and attention, learning and intellectual ability at 14 years: a prospective longitudinal study,” *Early Hum. Dev.*, Feb 2007, *83* (2), 115–123.
- Salt, A. and M. Redshaw**, “Neurodevelopmental follow-up after preterm birth: follow up after two years,” *Early Hum. Dev.*, Mar 2006, *82* (3), 185–197.
- Spolaore, Enrico and Romain Wacziarg**, “The Diffusion of Development,” *Quarterly Journal of Economics*, MAY 2009, *124* (2), 469–529.
- Tamer, Elie**, “Partial Identification in Econometrics,” *Annual Review of Economics*, 09 2010, *2* (1), 167–195.

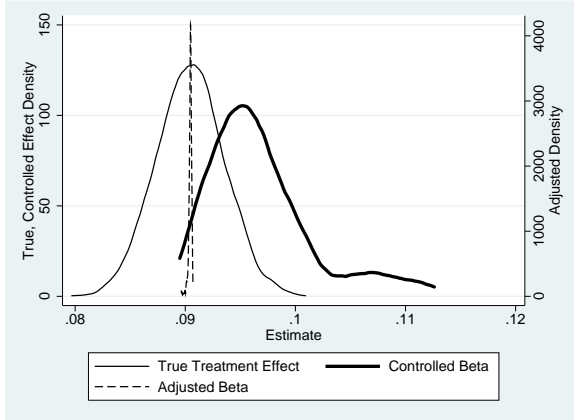
Figure 1: Simulated Effects by Controlled R-Squared



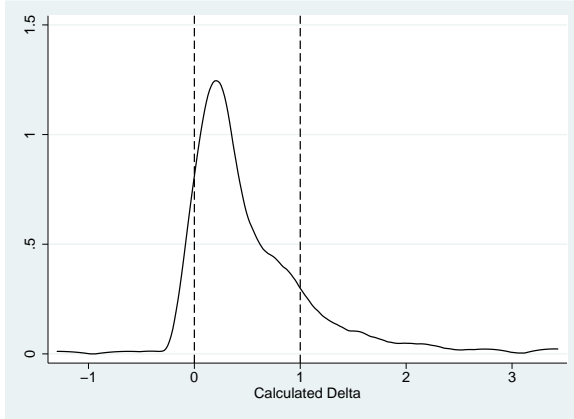
*Notes:* This figure shows calculated values of the bias-adjusted treatment effect in a constructed example with varying coefficient movements. In all cases I assume the coefficient with no controls is equal to 0.5, the uncontrolled R-squared is 0.1,  $\tilde{\delta} = 1$  and  $R_{max} = 1$ . The solid line shows the distribution of adjusted effects assuming the coefficient with controls is 0.45 and with varying controlled R-squared. The dashed line shows a similar distribution assuming the controlled coefficient is 0.30; the dotted line shows this with a controlled coefficient of 0.15. Discussion is in Section 3.3

Figure 2: NLSY Wage Data Simulation

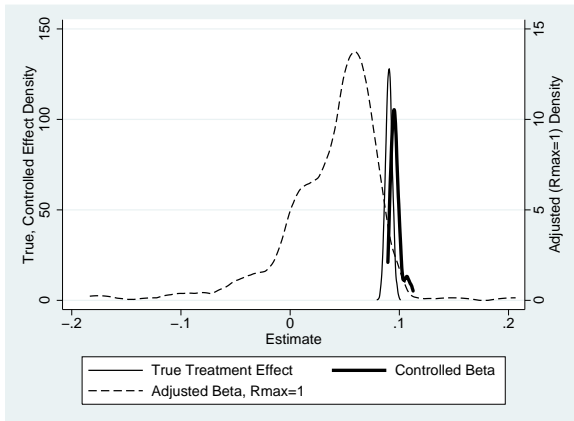
(a) “True”, Controlled and Adjusted Beta



(b) Distribution of Estimated Delta



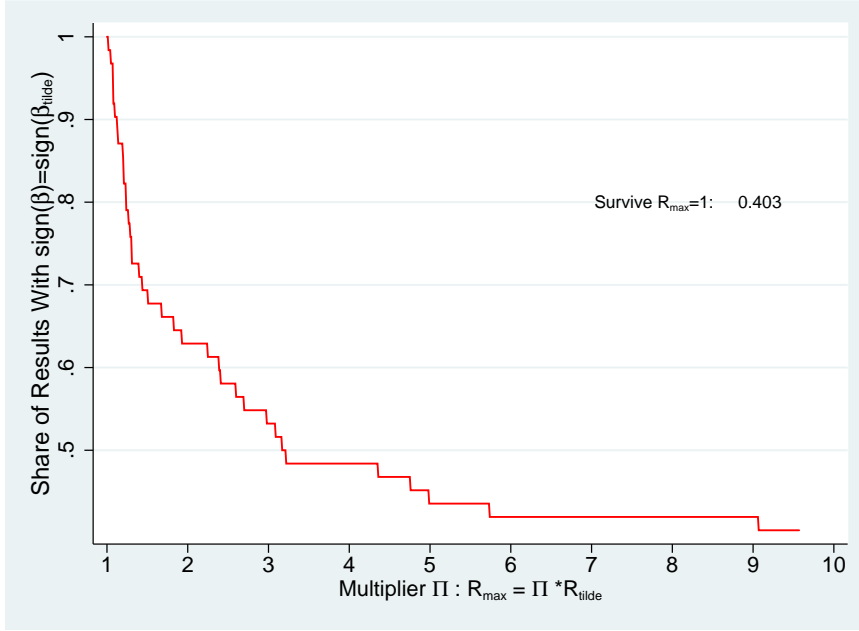
(c) “True”, Controlled and  $R_{max} = 1$  Adjusted Beta



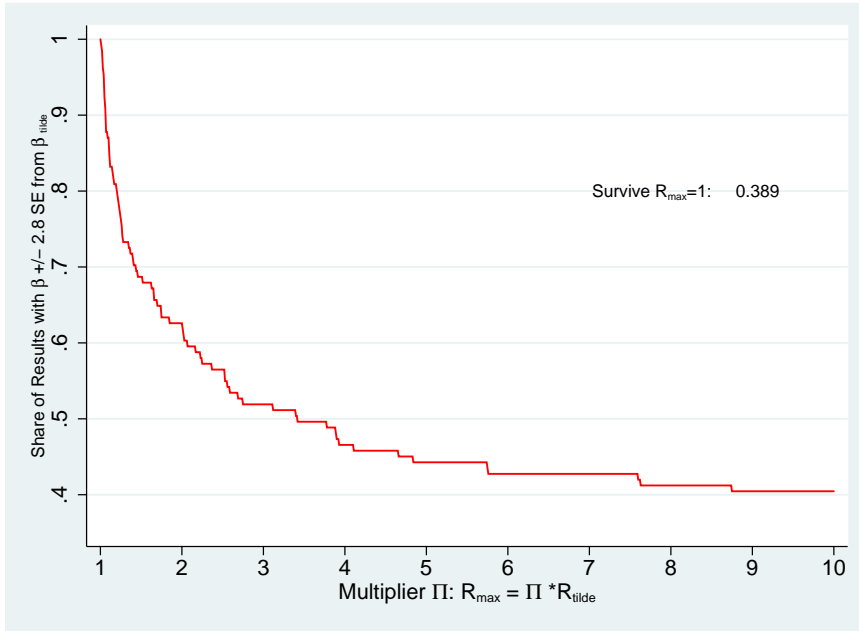
Notes: These figures show results from the validation using the constructed NLSY wage dataset. The analysis is described in Section 4.

Figure 3: **Robustness of Stability Results in Economics Literature**

(a) Rejection of Zero,  $R_{max} = \Pi \tilde{R}$ .



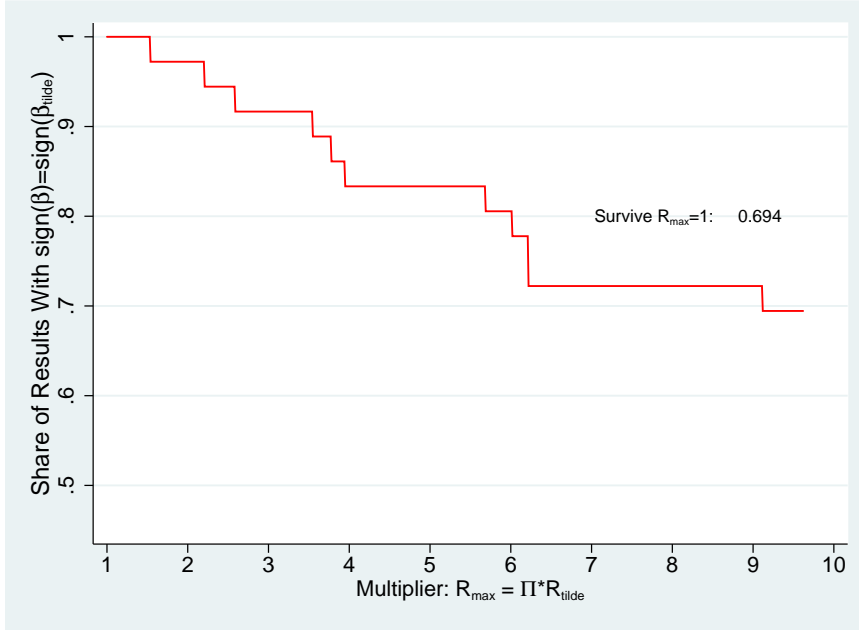
(b) Results within  $\pm 2.8$  SE,  $R_{max} = \Pi \tilde{R}$ .



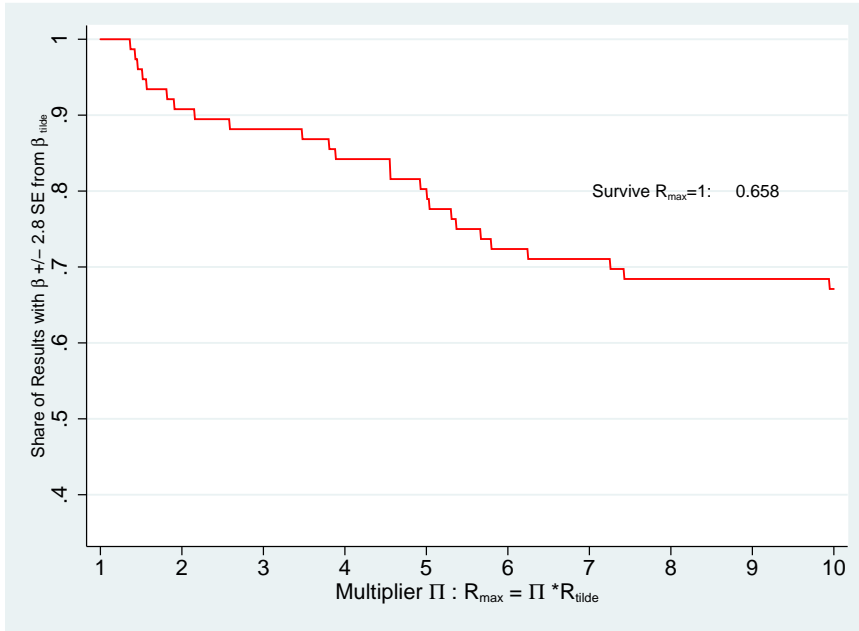
*Notes:* These graphs show the performance of non-randomized results under the proportional selection adjustment. Each figure graphs the share of results which would survive varying parametrizations of  $R_{max}$ , in all cases assuming  $\tilde{\delta} = 1$ . Sub-Figure a indicates the share of results which would survive  $R_{max} = \Pi \tilde{R}$  for varying values of  $\Pi$ , with the survival in this case meaning the identified set does not include zero. This figure contains only relationships where the effect is significant with controls and adding the controls moves the coefficient toward zero. Sub-Figure b indicates the share of results for which the full identified set would be within 2.8 standard errors of the controlled coefficient. This sub-figure includes all relationships.

Figure 4: Results from Randomized Data

(a) Rejection of Zero,  $R_{max} = \Pi \tilde{R}$ .

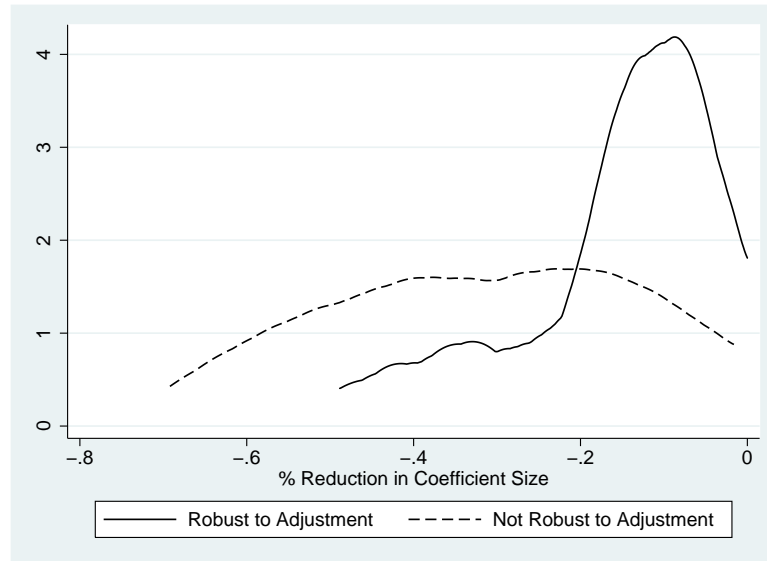


(b) Results within  $\pm 2.8$  SE,  $R_{max} = \Pi \tilde{R}$ .



*Notes:* These graphs show the performance of randomized results under the proportional selection adjustment. Each figure graphs the share of results which would survive varying parametrizations of  $R_{max}$ , in all cases assuming  $\tilde{\delta} = 1$ . Sub-Figure a indicates the share of results which would survive  $R_{max} = \Pi \tilde{R}$  for varying values of  $\Pi$ , with the survival in this case meaning the identified set does not include zero. This figure contains only relationships where the effect is significant with controls and adding the controls moves the coefficient toward zero. Sub-Figure b indicates the share of results for which the full identified set would be within 2.8 standard errors of the controlled coefficient. This sub-figure includes all relationships.

Figure 5: Relationship between Full Robustness and Coefficient Movement



*Notes:* This graph shows the range of coefficient movements in non-randomized studies divided into those which are robust to the proportional selection adjustment with  $R_{max} = 2.2\tilde{R}$  (solid line) and those which are not (dotted line). This includes only relationships in which the inclusion of controls moves the coefficient toward zero.



Table 1: **Example from Simulated Data**

Simulation Inputs		Simulation Outputs					
$\tilde{\delta}$	$Var(\tilde{W}_2)$	$\hat{\beta}[\hat{R}]$	$\tilde{\beta}[\tilde{R}]$	$R_{max}$	<i>Change in Coeff.</i>	<i>Remaining Bias</i>	<i>Adjusted <math>\beta</math></i>
1	0.1	0.52 [.12]	0.32 [.53]	0.57	0.198	0.017	0.308
1	1	0.69 [.14]	0.50 [.41]	0.69	0.189	0.198	0.299
1	10	2.50 [.46]	2.38 [.49]	0.93	0.116	2.08	0.309
0.5	0.1	0.51 [.11]	0.31 [.52]	0.56	0.199	0.009	0.307
0.5	1	0.60 [.11]	0.40 [.39]	0.70	0.196	0.102	0.295
0.5	10	1.49 [.17]	1.33 [.22]	0.92	0.159	1.03	0.291

*Notes:* This table shows calculations based on simulated data described in Section 3.3. The simulation constructs the outcome as follows:  $Y = .3X + W_1 + \tilde{W}_2 + \epsilon$  where  $\epsilon \sim N(0, 1)$ ,  $Cov(X, \epsilon) = Cov(W_1, \epsilon) = Cov(\tilde{W}_2, \epsilon) = 0$ ,  $Var(W_1) = Var(X) = 1$  and the means of  $X$ ,  $W_1$  and  $\tilde{W}_2$  are all equal to 1. The values  $\hat{\beta}$  and  $\hat{R}$  come from regressions of  $Y$  on  $X$  with no other controls;  $\tilde{\beta}$  and  $\tilde{R}$  come from regression which control for  $W_1$ .

Table 2: Maternal Behavior, Child IQ and Birth Weight

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Child IQ, Standardized (NLSY) ( $R_{max} = .61$ )						
<i>Treatment Variable</i>	<i>Baseline Effect</i> ( <i>Std. Error</i> ), [ $R^2$ ]	<i>Controlled Effect</i> ( <i>Std. Error</i> ), [ $R^2$ ]	<i>Null Reject?</i> ( <i>extrnl. evid.</i> )	<i>Sibling FE</i> <i>Estimate</i>	<i>Identified Set</i>	$R_{max}$ for $\beta = 0$ ( $\tilde{\delta} = 1$ ) $\tilde{\delta}$ for $\beta = 0$ <i>Given <math>R_{max}</math></i>
Breastfeed (Months)	0.045*** (.003) [.045]	0.017*** (.002) [.256]	No	-0.007 (.005)	[-0.028, 0.017]	0.391 0.382
Drink in Preg. (Any)	0.176*** (.026) [.008]	0.050*** (.023) [.249]	No	0.026 (.036)	[-0.138, 0.050]	0.345 0.265
LBW + Preterm	-0.188*** (.057) [.004]	-0.125*** (.050) [.251]	Yes	-0.111 (.070)	[-0.124, -0.033] <sup>†</sup>	0.743 1.37
Panel B: Birth Weight in Grams (NLSY) ( $R_{max} = .53$ )						
<i>Treatment Variable</i>	<i>Baseline Effect</i>	<i>Effect with Full Controls</i>	<i>Null Reject?</i> ( <i>extrnl. evid.</i> )	<i>Sibling FE</i> <i>Estimate</i>	<i>Identified Set</i>	$R_{max}$ for $\beta = 0$ ( $\tilde{\delta} = 1$ ) $\tilde{\delta}$ for $\beta = 0$ <i>Given <math>R_{max}</math></i>
Smoking in Preg	-206.7*** (16.0) [.030]	-196.5*** (16.5) [.068]	Yes	-90.1*** (31.8)	[-196.5, -73.6] <sup>†</sup>	0.806 1.59
Drink in Preg. (Amt)	-25.1*** (6.06) [.011]	-21.8*** (5.99) [.056]	No	-6.31 (8.64)	[-21.8, 12.9]	0.353 0.63

*Notes:* This table shows the validation results for the analysis of the impact of maternal behavior on child birth weight and IQ. Baseline effects include only controls for child sex and age dummies in the case of IQ. Full controls: race, age, education, income, marital status. Sibling fixed effects estimates come from NLSY in all panels. The identified set in Column (5) is bounded below by  $\tilde{\beta}$  and above by  $\beta^*$  calculated based on  $R_{max}$  given in the top row of each panel and  $\tilde{\delta} = 1$ .<sup>†</sup> Identified set excludes zero. The  $R_{max}$  calculation in Column (6) is done under the assumption that  $\tilde{\delta} = 1$ . \* significant at 10% level, \*\* significant at 5% level, \*\*\* significant at 1% level.

Table 3: Robustness of Stability Results

Panel A: Non-Randomized Data, Share of Results which Survive $\tilde{\delta} = 1$ , varying $R_{max}$				
	$R_{max} = 1$	$R_{max} = \min(3\tilde{R}; 1)$	$R_{max} = \min(2\tilde{R}; 1)$	$R_{max} = \min(1.5\tilde{R}; 1)$
Share With Adjusted $\beta$ Same Sign as $\tilde{\beta}$				
<i>Sample: Add Controls, Moves toward Zero</i>	41%	54%	64%	70%
Share with Adjusted $\beta$ +/- 2.8 SE of $\tilde{\beta}$				
<i>Sample: All</i>	38%	52%	63%	69%
Panel B: Randomized Data, Share of Results which Survive $\tilde{\delta} = 1$ , varying $R_{max}$				
	$R_{max} = 1$	$R_{max} = \min(3\tilde{R}; 1)$	$R_{max} = \min(2\tilde{R}; 1)$	$R_{max} = \min(1.5\tilde{R}; 1)$
Share With Adjusted $\beta$ Same Sign as $\tilde{\beta}$				
<i>Sample: Add Controls, Moves toward Zero</i>	69%	91%	97%	100%
Share with Adjusted $\beta$ +/- 2.8 SE of $\tilde{\beta}$				
<i>Sample: All</i>	63%	87%	91%	95%

*Notes:* This table describes the survival of non-randomized (Panel A) and randomized (Panel B) results under the proportional selection adjustment. Both panels show the share of results which would survive  $\tilde{\delta}$  with varying  $R_{max}$  values. I consider two definitions of survival: (1) the identified set does not include zero and (2) the outer bound of the set is within 2.8 standard errors of  $\tilde{\beta}$ . The first of these is considered only for results which move toward zero when controls are added.

Table 4: Examples from Economics Literature

Panel A: Mian and Sufi (2009)					
<i>Result Description</i>	<i>Baseline Effect</i> ( <i>Std. Error</i> )[ $R^2$ ]	<i>Controlled Effect</i> ( <i>Std. Error</i> )[ $R^2$ ]	$R_{max}$ for $\beta = 0$ ( $\tilde{\delta} = 1$ )	$\tilde{\delta}$ for $\beta = 0$ ( $R_{max} = 2.2\tilde{R}$ )	<i>Identified Set</i> $[\tilde{\beta}, \beta^*(2.2\tilde{R}, 1)]$
Table V, Column (1) to (4)	0.49 (.033) [.44]	0.431 (.079) [.94]	$> 1$	54.19	[0.431, 0.423]
Panel B: Nunn and Wantchekon (2011)					
<i>Result Description</i>	<i>Baseline Effect</i> ( <i>Std. Error</i> )[ $R^2$ ]	<i>Controlled Effect</i> ( <i>Std. Error</i> )[ $R^2$ ]	$R_{max}$ for $\beta = 0$ ( $\tilde{\delta} = 1$ )	$\tilde{\delta}$ for $\beta = 0$ ( $R_{max} = 2.2\tilde{R}$ )	<i>Identified Set</i> $[\tilde{\beta}, \beta^*(2.2\tilde{R}, 1)]$
Row 1, Trust Relatives	-0.163 (.043) [.115]	-0.133 (.036) [.133]	0.21	0.48	[-0.133, 0.141]
Row 1, Trust Neighbors	-0.196 (.045) [.117]	-0.159 (.034) [.156]	0.32	0.89	[-0.159, 0.020]
Row 1, Trust Local Council	-0.147 (.031) [.163]	-0.111 (.022) [.196]	0.30	0.43	[-0.111, 0.149]
Row 1, Intragroup Trust	-0.179 (.041) [.115]	-0.144 (.031) [.144]	0.26	0.69	[-0.144, 0.064]
Row 1, Intergroup Trust	-0.126 (.032) [.091]	-0.097 (.027) [.112]	0.18	0.52	[-0.097, 0.088]
Row 2, Trust Relatives	-0.193 (.043) [.106]	-0.178 (.031) [.130]	0.41	1.81	[-0.178, -0.080]
Row 2, Trust Neighbors	-0.238 (.044) [.115]	-0.202 (.029) [.159]	0.47	1.63	[-0.202, -0.078]
Row 2, Trust Local Council	-0.177 (.027) [.175]	-0.128 (.021) [.205]	0.28	0.32	[-0.128, 0.274]
Row 2, Intragroup Trust	-0.208 (.041) [.121]	-0.187 (.032) [.155]	0.47	1.70	[-0.187, -0.078]
Row 2, Intergroup Trust	-0.145 (.031) [.093]	-0.115 (.030) [.119]	0.21	0.68	[-0.115, 0.054]
Panel C: Spolaore and Wacziarg (2009)					
<i>Result Description</i>	<i>Baseline Effect</i> ( <i>Std. Error</i> )[ $R^2$ ]	<i>Controlled Effect</i> ( <i>Std. Error</i> )[ $R^2$ ]	$R_{max}$ for $\beta = 0$ ( $\tilde{\delta} = 1$ )	$\tilde{\delta}$ for $\beta = 0$ ( $R_{max} = 2.2\tilde{R}$ )	<i>Identified Set</i> $[\tilde{\beta}, \beta^*(2.2\tilde{R}, 1)]$
Table 4, Column (1) to (5)	6.357 (.996) [.11]	4.134 (1.046) [.22]	0.42	0.75	[-1.37, 4.13]

*Notes:* This table summarizes the sensitivity of results in three papers to the proportional selection adjustment selected. Mian and Sufi (2009) estimate the relationship between sub-prime borrowers and mortgage origination by zip code. The uncontrolled results for Mian and Sufi (2009) are slightly different from the paper because these limit to the same sample for controlled and uncontrolled results. Nunn and Wantchekon (2011) estimate the relationship between African slave trades and the extent of mistrust in modern African countries. Spolaore and Wacziarg (2009) estimate the impact of genetic distance on income distance across countries.

## Appendix A: Theoretical Results

### A.1. Details of Proofs

I show the proof in this section for arbitrary  $R_{max}$  and  $\tilde{\delta}$ , the formulation described in Section 3. Note, however, that if  $R_{max} = 1$  then  $\tilde{\delta} = \delta$ , and the proof is identical. Therefore I show the proof for Corollary 1 below, noting that Proposition 1 is a special case.

**Proof of Lemma 1: Claim:**  $(\hat{\beta} - \tilde{\beta}) \xrightarrow{p} \sigma_{1X} \frac{\sigma_{11}^2 - \sigma_{1X}^2(\delta\sigma_{22} + \sigma_{11})}{\sigma_{11}(\sigma_{11} - \sigma_{1X}^2)}$

**Proof:** Observe that  $\hat{\lambda}_{w_1|X}$  converges in probability to  $\frac{Cov(W_1, X)}{V(X)} = \sigma_{1X}$ . By a similar logic,  $\hat{\lambda}_{W_2|X}$  converges to  $\sigma_{2X}$  and, under proportional selection, to  $\frac{\tilde{\delta}\sigma_{1X}\sigma_{22}}{\sigma_{11}}$ .  $\hat{\lambda}_{W_2|X, W_1}$  converges in probability to  $\frac{Cov(W_2, \tilde{X})}{Var(\tilde{X})}$  where  $\tilde{X}$  is the residual from a regression of  $X$  on  $W_1$ . Note that  $Var(\tilde{X})$  converges in probability to  $1 - \frac{\sigma_{1X}^2}{\sigma_{11}}$ . Therefore, again invoking proportional selection,  $\hat{\lambda}_{W_2|X, W_1}$  converges in probability to  $\frac{\tilde{\delta}\sigma_{22}\sigma_{1X}}{\sigma_{11} - \sigma_{1X}^2}$ . Subtracting and simplifying yields the result.

**Proof of Lemma 2: Claim:**  $(\tilde{R} - \hat{R})\hat{\sigma}_{yy} \xrightarrow{p} \frac{[\sigma_{11}^2 - \sigma_{1X}^2(\sigma_{11} + \tilde{\delta}\sigma_{22})]^2}{\sigma_{11}^2(\sigma_{11} - \sigma_{1X}^2)}$  and  $(R_{max} - \tilde{R})\hat{\sigma}_{yy} \xrightarrow{p} \frac{\sigma_{22}[\sigma_{11}^2 - \sigma_{1X}^2(\sigma_{11} + \tilde{\delta}^2\sigma_{22})]}{\sigma_{11}(\sigma_{11} - \sigma_{1X}^2)}$ .

**Proof:** Observe the following definitions. From the short regression coefficient,  $\hat{R} = \frac{(\beta + \hat{\lambda}_{w_1|X} + \hat{\lambda}_{W_2|X})^2}{\hat{\sigma}_{yy}}$ . By

Lemma 1, this converges in probability to  $\frac{(\beta + \frac{\sigma_{1X}(\delta\sigma_{22} + \sigma_{11})}{\sigma_{11}})^2}{\sigma_{yy}}$ . In the intermediate regression the calculation relies on the coefficient on  $X$  ( $\beta + \hat{\lambda}_{W_2|X, W_1}$ ) and the coefficient on  $W_1$ , which is also biased by the exclusion of  $W_2$  through the joint correlation with  $X$  and is equal to  $1 - \frac{\sigma_{1X}}{\sigma_{11}}\hat{\lambda}_{W_2|X, W_1}$ . Thus,

$$\tilde{R} = \frac{(\beta + \hat{\lambda}_{W_2|X, W_1})^2 + \sigma_{11}(1 - \frac{\sigma_{1X}}{\sigma_{11}}\hat{\lambda}_{W_2|X, W_1})^2 + 2\sigma_{1X}(\beta + \hat{\lambda}_{W_2|X, W_1})(1 - \frac{\sigma_{1X}}{\sigma_{11}}\hat{\lambda}_{W_2|X, W_1})}{\hat{\sigma}_{yy}}. \text{ By Lemma 1,}$$

$$\tilde{R} \xrightarrow{p} \frac{(\beta + \frac{\tilde{\delta}\sigma_{22}\sigma_{1X}}{\sigma_{11} - \sigma_{1X}^2})^2 + (1 - \frac{\sigma_{1X}}{\sigma_{11}}\frac{\tilde{\delta}\sigma_{22}\sigma_{1X}}{\sigma_{11} - \sigma_{1X}^2})^2\sigma_{11} + 2(\beta + \frac{\tilde{\delta}\sigma_{22}\sigma_{1X}}{\sigma_{11} - \sigma_{1X}^2})(1 - \frac{\sigma_{1X}}{\sigma_{11}}\frac{\tilde{\delta}\sigma_{22}\sigma_{1X}}{\sigma_{11} - \sigma_{1X}^2})\sigma_{1X}}{\sigma_{yy}}. \text{ Finally, observe that}$$

$$R_{max} = \frac{\beta^2 + \sigma_{11} + \sigma_{22} + 2\beta\sigma_{1X} + 2\beta\frac{\tilde{\delta}\sigma_{1X}\sigma_{22}}{\sigma_{11}}}{\sigma_{yy}}. \text{ Differencing these expressions appropriately yields the result.}$$

**Proof of Corollary 1. Claim :** Define:

$$\beta^{*'} = \begin{cases} \tilde{\beta} - \left[ \hat{\beta} - \tilde{\beta} \right] \frac{R_{max} - \tilde{R}}{\tilde{R} - \hat{R}} & \text{if } \tilde{\delta}=1 \\ \tilde{\beta} - \left[ \frac{\sqrt{[\hat{\beta} - \tilde{\beta}]^2 [\Theta^2 + \Theta(4\delta(1-\delta)[\hat{\beta} - \tilde{\beta}]^2 [R_{max} - \tilde{R}])]} - \Theta[\hat{\beta} - \tilde{\beta}]}{2(1-\delta)[\hat{\beta} - \tilde{\beta}]^2 [\tilde{R} - \hat{R}]} \right] & \text{if } \tilde{\delta} \neq 1, \sigma_{1X} \geq 0 \\ \tilde{\beta} - \left[ \frac{-\sqrt{[\hat{\beta} - \tilde{\beta}]^2 [\Theta^2 + \Theta(4\tilde{\delta}(1-\delta)[\hat{\beta} - \tilde{\beta}]^2 [R_{max} - \tilde{R}])]} - \Theta[\hat{\beta} - \tilde{\beta}]}{2(1-\tilde{\delta})[\hat{\beta} - \tilde{\beta}]^2 [\tilde{R} - \hat{R}]} \right] & \text{if } \tilde{\delta} \neq 1, \sigma_{1X} < 0 \end{cases}$$

where  $\Theta = \left( [\tilde{R} - \hat{R}]^2 \hat{\sigma}_{yy} + [\hat{\beta} - \tilde{\beta}]^2 [\tilde{R} - \hat{R}] \right) \cdot \beta^* \xrightarrow{p} \beta$ .

**Proof:** Recall that the object of interest – the bias – is  $\frac{\tilde{\delta}\sigma_{22}\sigma_{1X}}{\sigma_{11} - \sigma_{1X}^2}$ . There are three unknowns here:  $\sigma_{11}$ ,  $\sigma_{22}$  and  $\sigma_{1X}$ . Note that none of these can be calculated directly from the data. Lemmas 1 and 2 provide a system of three equations in these variables. Lemmas are stated in probability limits; for the proof I will write these as equalities to simplify notation, and return to the probability limit notation at the end. In addition, again to

simplify notation in the algebra, I will adopt single letter notation for each of the differences.

$$\begin{aligned}
A = \hat{\beta} - \tilde{\beta} &= \sigma_{1X} \frac{\sigma_{11}^2 - \sigma_{1X}^2(\tilde{\delta}\sigma_{22} + \sigma_{11})}{\sigma_{11}(\sigma_{11} - \sigma_{1X}^2)} \\
B = [\tilde{R} - \hat{R}] \sigma_{yy} &= \frac{[\sigma_{11}^2 - \sigma_{1X}^2(\sigma_{11} + \tilde{\delta}\sigma_{22})]^2}{\sigma_{11}^2(\sigma_{11} - \sigma_{1X}^2)} \\
C = [R_{max} - \tilde{R}] \sigma_{yy} &= \frac{\sigma_{22} [\sigma_{11}^2 - \sigma_{1X}^2(\sigma_{11} + \tilde{\delta}^2\sigma_{22})]}{\sigma_{11}(\sigma_{11} - \sigma_{1X}^2)}
\end{aligned}$$

The algebra differs slightly for the case of  $\tilde{\delta} = 1$  and the case of  $\tilde{\delta} \neq 1$  but only in a later step of the proof. I will note when the cases diverge below. The method of proof is simply to solve the system of simultaneous equations. Some algebraic steps are suppressed.

*Solve Equation (1) for  $\sigma_{22}$ :*

$$\begin{aligned}
A &= \sigma_{1X} \frac{\sigma_{11}^2 - \sigma_{1X}^2(\tilde{\delta}\sigma_{22} + \sigma_{11})}{\sigma_{11}(\sigma_{11} - \sigma_{1X}^2)} \\
\sigma_{22} &= \frac{1}{\tilde{\delta}} \left[ \frac{\sigma_{11}^2\sigma_{1X} - A\sigma_{11}(\sigma_{11} - \sigma_{1X}^2) - \sigma_{11}\sigma_{1X}^3}{\sigma_{1X}^3} \right]
\end{aligned}$$

*Solve Equation (2) for  $\sigma_{11}$  and  $\sigma_{22}$  in terms of  $\sigma_{1X}$ :*

$$\begin{aligned}
B &= \frac{[\sigma_{11}^2 - \sigma_{1X}^2(\sigma_{11} + \tilde{\delta}\sigma_{22})]^2}{\sigma_{11}^2(\sigma_{11} - \sigma_{1X}^2)\sigma_y} = \frac{[\sigma_{11}^2 - \sigma_{1X}^2(\sigma_{11} + \frac{\sigma_{11}^2\sigma_{1X} - A\sigma_{11}(\sigma_{11} - \sigma_{1X}^2) - \sigma_{11}\sigma_{1X}^3}{\sigma_{1X}^3})]^2}{\sigma_{11}^2(\sigma_{11} - \sigma_{1X}^2)} \\
\sigma_{11} &= \left[ \frac{\sigma_{1X}^2(A^2 + B)}{A^2} \right] \\
\sigma_{22} &= \frac{1}{\tilde{\delta}} \left[ \frac{\sigma_{1X}(B^2 + A^2B)[\sigma_{1X} - A]}{A^4} \right]
\end{aligned}$$

Note that for the bias calculation we do not require  $\sigma_{11}$  alone but only  $\sigma_{11} - \sigma_{1X}^2$  which, given values above, equals  $\frac{\sigma_{1X}^2B}{A^2}$  and allows us to collapse the bias calculation to  $\frac{\tilde{\delta}\sigma_{22}A^2}{\sigma_{1X}B}$ .

**Case 1:**  $\tilde{\delta} = 1$ . *Solve Equation (3) for  $\sigma_{1X}$ :*

$$\begin{aligned}
C &= \frac{\sigma_{22} [\sigma_{11}^2 - \sigma_{1X}^2(\sigma_{11} + \sigma_{22})]}{\sigma_{11}(\sigma_{11} - \sigma_{1X}^2)} = \frac{\left[ \frac{\sigma_{1X}(B^2 + A^2B)[\sigma_{1X} - A]}{A^4} \right] \left[ \frac{\sigma_{1X}^2(A^2 + B)}{A^2} - \sigma_{1X}^2 \left( \frac{\sigma_{1X}^2(A^2 + B)}{A^2} + \left[ \frac{\sigma_{1X}(B^2 + A^2B)[\sigma_{1X} - A]}{A^4} \right] \right) \right]}{\frac{\sigma_{1X}^2(A^2 + B)}{A^2} \left[ \frac{\sigma_{1X}^2(A^2 + B)}{A^2} - \sigma_{1X}^2 \right]} \\
\sigma_{1X} &= \frac{CA^3 + A(B^2 + A^2B)}{(B^2 + A^2B)} \\
\sigma_{22} &= [CA^3 + A(B^2 + A^2B)] \left[ \frac{C}{A(B^2 + A^2B)} \right]
\end{aligned}$$

Applying these values to the formula above, we have:

$$\frac{\tilde{\delta}\sigma_{22}\sigma_{1X}}{\sigma_{11} - \sigma_{1X}^2} = \frac{AC}{B} = \tilde{\delta} [\hat{\beta} - \tilde{\beta}] \frac{R_{max} - \tilde{R}}{\tilde{R} - \hat{R}}$$

which leads us to the  $\tilde{\delta} = 1$  result.

**Case 2:**  $\tilde{\delta} \neq 1$ . Solve Equation (3) for  $\sigma_{1X}$  :

$$C = \frac{\sigma_{22} [\sigma_{11}^2 - \sigma_{1X}^2 (\sigma_{11} + \tilde{\delta}^2 \sigma_{22})]}{\sigma_{11} (\sigma_{11} - \sigma_{1X}^2)}$$

$$C = \frac{1}{\tilde{\delta}} \frac{\frac{\sigma_{1X} (B^2 + A^2 B) [\sigma_{1X} - A]}{A^4} \left[ \left[ \frac{\sigma_{1X}^2 (A^2 + B)}{A^2} \right]^2 - \sigma_{1X}^2 \left( \frac{\sigma_{1X}^2 (A^2 + B)}{A^2} + \tilde{\delta} \frac{\sigma_{1X} (B^2 + A^2 B) [\sigma_{1X} - A]}{A^4} \right) \right]}{\frac{\sigma_{1X}^2 (A^2 + B)}{A^2} \left( \frac{\sigma_{1X}^2 (A^2 + B)}{A^2} - \sigma_{1X}^2 \right)}$$

This does not simplify to the extent that the  $\delta = 1$  case does, and solving for  $\sigma_{1X}$  requires the quadratic formula. Applying this, we find:

$$\sigma_{1X} = \frac{(A(B^2 + A^2 B)(1 - 2\tilde{\delta})) \pm \sqrt{(A^2(B^2 + A^2 B))^2 + 4(B^2 + A^2 B)(1 - \tilde{\delta})\tilde{\delta}CA^4}}{2(B^2 + A^2 B)(1 - \tilde{\delta})}$$

Note this has two roots. The positive root corresponds to the case where  $\sigma_{1X} \geq 0$ ; the negative root to the case where  $\sigma_{1X} < 0$ .

Given this and the resulting formula for  $\sigma_{22}$  we can complete the solution. If  $\sigma_{1X} \geq 0$ , we have:

$$\frac{\tilde{\delta}\sigma_{22}A^2}{\sigma_{1X}B} = \left[ \frac{(-A(B^2 + A^2 B) + \sqrt{(A^2(B^2 + A^2 B) [B^2 + A^2 B + 4\tilde{\delta}(1 - \tilde{\delta})CA^2]})}{2(1 - \tilde{\delta})BA^2} \right]$$

If  $\sigma_{1X} < 0$  we have:

$$\frac{\tilde{\delta}\sigma_{22}A^2}{\sigma_{1X}B} = \left[ \frac{(-A(B^2 + A^2 B) - \sqrt{(A^2(B^2 + A^2 B) [B^2 + A^2 B + 4\tilde{\delta}(1 - \tilde{\delta})CA^2]})}{2(1 - \tilde{\delta})BA^2} \right]$$

Substituting in the difference values for  $A$ ,  $B$  and  $C$  yields the result.

Note that calculating the values of  $R_{max}$  or  $\tilde{\delta}$  which produce  $\beta = 0$  simply involves rearranging these equations.

## A.2 Further Theoretical Results

This appendix discusses two additional issues related to the theory. Subsection A.2.1 below briefly contrasts the calculation of bias based on the coefficients to the calculation directly from the data suggested by Altonji, Elder and Taber (2005). Subsection A.2.2 discusses details of the case with **m**.

### A.2.1 Altonji, Elder and Taber (2005) Calculation

Recall the model:

$$Y = \alpha + \beta X + W_1 + W_2$$

Lemma 1 in the text demonstrates that, under the proportional selection relationship the bias on the intermediate regression coefficient  $\tilde{\beta}$  is  $\frac{\tilde{\delta}\sigma_{22}\sigma_{1X}}{\sigma_{11} - \sigma_{1X}^2}$ .

Altonji, Elder and Taber (2005) suggest that this bias might be calculated directly from the data. In particular, they propose:

1. Run the intermediate regression, which we will denote  $Y = \tilde{\beta}X + \Psi W_1 + \tilde{\epsilon}$ .

2. Calculate  $\Psi W_1$  and denote the variance of the residual  $V_{\bar{\epsilon}}$ .
3. Regress  $X$  on  $\hat{\Psi} W_1$ . Denote the coefficient on  $\hat{\Psi} W_1$  as  $\Gamma$ , and the variance of the residual  $V_{\hat{X}}$ .
4. Calculate the bias as  $\frac{\delta \Gamma V_{\bar{\epsilon}}}{V_{\hat{X}}}$

Recall that  $Var(X) = 1$ . Consider each of the elements of this in turn:

1.  $V_{\hat{X}}$

$$V_{\hat{X}} \xrightarrow{p} 1 - \frac{\sigma_{1X}^2}{\sigma_{11}}$$

2.  $\Gamma$ . Note first that  $\hat{\Psi} \rightarrow^p 1 - \frac{\sigma_{1X}}{\sigma_{11}} \frac{\delta \sigma_{22} \sigma_{1X}}{\sigma_{11}(1 - \sigma_{1X}^2)}$ .

$$\begin{aligned} \Gamma &= \frac{Cov(\hat{\Psi} W_1, X)}{Var(\hat{\Psi} W_1)} = \frac{Cov(W_1, X)}{\hat{\Psi} Var(W_1)} \\ \Gamma &\xrightarrow{p} \frac{\sigma_{1X}}{\left[1 - \frac{\sigma_{1X}}{\sigma_{11}} \frac{\delta \sigma_{22} \sigma_{1X}}{\sigma_{11}(1 - \sigma_{1X}^2)}\right] \sigma_{11}} = \frac{\sigma_{1X}(\sigma_{11} - \sigma_{1X}^2)}{\sigma_{11}(\sigma_{11} - \sigma_{1X}^2) - \delta \sigma_{22} \sigma_{1X}^2} \end{aligned}$$

3.  $V_{\bar{\epsilon}}$ .

$$V_{\bar{\epsilon}} \xrightarrow{p} \sigma_{22} - \frac{(\delta \sigma_{22} \sigma_{1X})^2}{\sigma_{11}(\sigma_{11} - \sigma_{1X}^2)}$$

Combining these, we find:

$$\frac{\delta \Gamma V_{\bar{\epsilon}}}{V_{\hat{X}}} \xrightarrow{p} \frac{\delta \sigma_{22} \sigma_{1X}}{\sigma_{11} - \sigma_{1X}^2} \left[ \frac{\sigma_{11}(\sigma_{11} - \sigma_{1X}^2) - \delta^2 \sigma_{22} \sigma_{1X}^2}{\sigma_{11}(\sigma_{11} - \sigma_{1X}^2) - \delta \sigma_{22} \sigma_{1X}^2} \right]$$

If  $\delta = 1$  the second term cancels, but in cases where  $\delta \neq 1$  it does not and this calculation is a close approximation to the bias.

## A.2.2 Additional Category Controls

Section 3 discusses extending the model to a case where there is an additional, orthogonal, category of controls, so the true model is

$$Y = \alpha + \beta X + W_1 + \tilde{W}_2 + m + \epsilon$$

I suggest in Section 3 that the appropriate procedure for recovering  $\beta$  if  $m$  is observed is to include  $m$  in both the short and intermediate regressions and preform the same procedure. It is trivial to see why this works. I have assumed that  $m$  is orthogonal to  $W_1$  and  $\tilde{W}_2$ . The only correlations are between  $m$  and  $X$  and  $Y$ . Consider regressing  $Y$  on  $m$  and taking residuals and doing the same for  $X$ . We can then run our original procedure on the residuals of  $X$  and  $Y$  to recover  $\beta$ . Including the  $m$  control in both regressions is equivalent to this exercise.

In the case where  $\mathbf{m}$  is not observed, I suggest that it is still possible to use this procedure to recover  $\bar{\beta}$  from this regression:

$$Y = \bar{\alpha} + \bar{\beta} X + \Psi W_1 + \Psi \tilde{W}_2 + \bar{\epsilon}$$

Although this will not be the causal effect, since  $m$  is omitted, it will be closer to the causal effect since it adjusts for the influence of  $\tilde{W}_2$ . The procedure for recovering  $\bar{\beta}$  differs from the main text only in that  $\Psi \neq 1$ .<sup>14</sup>

To prove this, we therefore work through a modified version of the proof in Section 2.

Short and intermediate regression coefficients are given below.

$$\begin{aligned} \hat{\beta} &= \bar{\beta} + \hat{\Psi} \hat{\lambda}_{w_1|X} + \hat{\Psi} \hat{\lambda}_{w_2|X} \\ \tilde{\beta} &= \bar{\beta} + \hat{\Psi} \hat{\lambda}_{w_2|X, W_1} \end{aligned}$$

<sup>14</sup>As mentioned in the text this is because they are biased by the exclusion of  $\mathbf{m}$  through the joint correlation with  $X$ .



By the same logic as Lemma 1 in the text, and the observation that  $\hat{\Psi} \rightarrow^p \Psi$  we observe that

$$\dot{\beta} - \tilde{\beta} \xrightarrow{p} \Psi \sigma_{1X} \frac{\sigma_{11}^2 - \sigma_{1X}^2 (\tilde{\delta} \sigma_{22} + \sigma_{11})}{\sigma_{11} (\sigma_{11} - \sigma_{1X}^2)}$$

Turning to the r-squared values, we observe  $\dot{R} \xrightarrow{p} \frac{(\bar{\beta} + \frac{\Psi \sigma_{1X} (\tilde{\delta} \sigma_{22} + \sigma_{11})}{\sigma_{11}})^2}{\sigma_{yy}}$ ,  
 $\tilde{R} \xrightarrow{p} \frac{(\bar{\beta} + \frac{\tilde{\delta} \sigma_{22} \sigma_{1X}}{\sigma_{11} - \sigma_{1X}^2})^2 + (\Psi - \frac{\sigma_{1X}}{\sigma_{11}} \frac{\tilde{\delta} \sigma_{22} \sigma_{1X}}{\sigma_{11} - \sigma_{1X}^2})^2 \sigma_{11} + 2(\bar{\beta} + \frac{\tilde{\delta} \sigma_{22} \sigma_{1X}}{\sigma_{11} - \sigma_{1X}^2})(\Psi - \frac{\sigma_{1X}}{\sigma_{11}} \frac{\tilde{\delta} \sigma_{22} \sigma_{1X}}{\sigma_{11} - \sigma_{1X}^2}) \sigma_{1X}}{\sigma_{yy}}$  and  
 $R_{max} = \frac{\bar{\beta}^2 + \Psi^2 \sigma_{11} + \Psi^2 \sigma_{22} + 2\bar{\beta} \Psi \sigma_{1X} + 2\bar{\beta} \Psi \frac{\tilde{\delta} \sigma_{1X} \sigma_{22}}{\sigma_{11}}}{\sigma_{yy}}$ . Algebraic simplification then yields:

$$\begin{aligned} [\tilde{R} - \dot{R}] \sigma_{yy} &\xrightarrow{p} \Psi^2 \frac{[\sigma_{11}^2 - \sigma_{1X}^2 (\sigma_{11} + \tilde{\delta} \sigma_{22})]^2}{\sigma_{11}^2 (\sigma_{11} - \sigma_{1X}^2)} \\ [R_{max} - \tilde{R}] \sigma_{yy} &\xrightarrow{p} \Psi \frac{\sigma_{22} [\sigma_{11}^2 - \sigma_{1X}^2 (\sigma_{11} + \tilde{\delta}^2 \sigma_{22})]}{\sigma_{11} (\sigma_{11} - \sigma_{1X}^2)} \end{aligned}$$

Combining, we replicate the results from Section 2.

## Appendix B: Appendix Tables

Table 1: **Summary Statistics: NLSY Wage Data**

	<i>Mean</i>	<i>Standard Deviation</i>	<i>Range</i>	<i>Sample Size</i>
Log Wages (1996-1998)	2.67	0.63	0-6.21	7496
Years of educ.	12.5	2.24	0-20	7496
Years of exper.	16.3	3.02	8-31	7496
Female	0.49	0.50	0-1	7496
Region of Residence	N/A	N/A	1-4	7496
White	0.64	0.47	0-1	7496
Married Codes	N/A	N/A	0-6	7496
Mother Educ (yrs)	11.0	3.00	1-20	7496
Father Educ (yrs)	11.2	5.27	1-20	7496
Mother Occup (codes)	N/A	N/A	0-984	7496
Father Occup (codes)	N/A	N/A	0-984	7496
Siblings (#)	3.8	2.6	0-22	7496

*Notes:* This table shows summary statistics for the data used in the NSLY wage analysis in Section 4. Data comes from the NLSY-79 cohort. Means are not reported for region, marital codes or occupation because they are not meaningful. All variables are controlled in the regressions as dummies. Wages are the max of 1996 and 1998 wages.

Table 2: **Summary Statistics: Early Life and Child IQ**

<b>Panel A: IQ Analysis</b>			
	<i>Mean</i>	<i>Standard Deviation</i>	<i>Sample Size</i>
IQ (PIAT Score, Standardized)	0.025	0.991	6962
Breastfeeding Months	2.40	4.63	6514
LBW + Preterm	0.049	0.217	6174
Mom Drink at all in Pregnancy	0.322	0.467	6537
Age	5.57	1.37	6962
Child Female	0.495	0.500	6962
Mother Black	0.282	0.450	6962
Mother Age	25.3	5.61	6962
Mother Education (years)	12.2	2.7	6962
Mother Income	\$41,294	\$80,735	6962
Mother Married	0.654	0.476	6962
<b>Panel B: Birth Weight Analysis</b>			
Birth Weight (grams)	3290.4	647.69	7686
Mom Smoke in Pregnancy	0.291	0.454	7686
Drinking Intensity (0-7)	0.638	1.16	7442
Child Female	0.486	0.499	7686
Mother Black	0.273	0.445	7686
Mother Age	24.4	5.49	7686
Mother Education (years)	12.0	2.7	7686
Mother Income	\$30,813	\$65,374	7686
Mother Married	0.667	0.471	7686

*Notes:* This table shows summary statistics for the data used in the analysis in Section 4. Drinking intensity is coded from 0 (never) to 7 (every day). Natality detail files are from 2001 and 2002. Data is from the NLSY Children and Young Adults panel.

## Appendix C: Alternative Parametrization of $R_{max}$

This appendix considers how the results in Section 5 would change if I used an alternative parametrization of  $R_{max}$ . The primary analysis in the paper uses  $R_{max} = \Pi \tilde{R}$  with varying  $\Pi$ . Here, I use  $R_{max} = \tilde{R} + \Pi(\tilde{R} - \hat{R})$ . I consider the same questions: the level of robustness of non-randomized results, the  $\Pi$  cutoff implied by the randomized data and what share of non-randomized results would survive the cutoff. For simplicity, I consider here only the primary robustness criteria of whether the identified set excludes zero, and therefore limit to results where inclusion of controls moves the effect size toward zero.

Figure 1a shows the robustness of non-randomized data under this parametrization, and Figure 1b shows the randomized robustness. The multiplier values in both cases are larger here, reflecting the fact that the increase in R-squared from  $\hat{R}$  to  $\tilde{R}$  is smaller in value than the level,  $\tilde{R}$ . The observation that the randomized data is more robust than the non-randomized is even more true. Table 1 replicates the form of Table 3 in the paper. I consider larger values of  $\Pi$  as cutoffs but, again, the conclusion of varying stability and higher stability of randomized results.

The robustness cutoff value implied by the randomized data is 4.95. With this value 67% of non-randomized results would survive, almost exactly the share in the main analysis. Perhaps more important, in 83% of cases the robustness conclusions overlap, suggesting that similar information is being provide by both parametrizations.

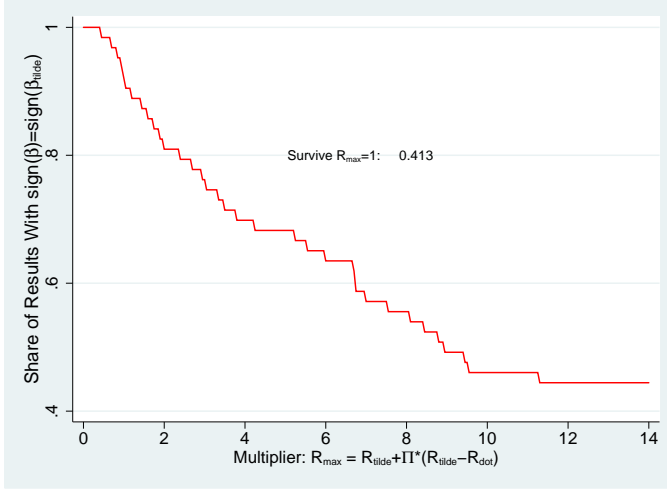
Table 1: **Robustness of Stability Results, Alternative  $R_{max}$**

Share of Results which Survive $\tilde{\delta} = 1$ , varying $R_{max}$ : Robustness is Identified Set Excludes Zero				
	$R_{max} = 1$	$R_{max} = \min\{\tilde{R} + 5(\tilde{R} - \hat{R}), 1\}$	$R_{max} = \min\{\tilde{R} + 3(\tilde{R} - \hat{R}), 1\}$	$R_{max} = \min\{\tilde{R} + 2(\tilde{R} - \hat{R}), 1\}$
Non-Randomized	41%	67%	75%	81%
Randomized	69%	94%	97%	100%

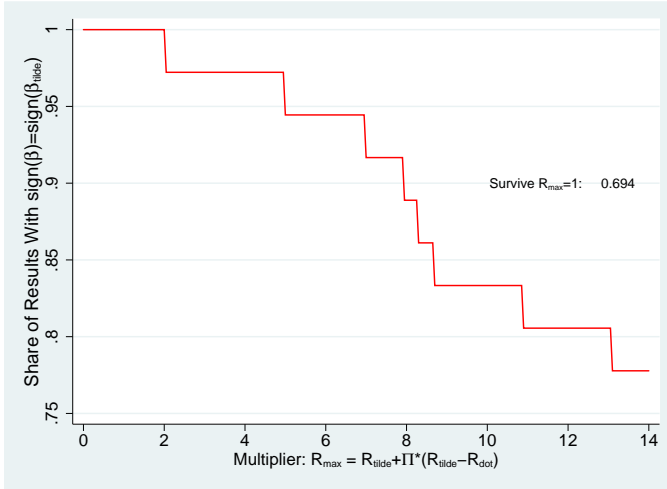
*Notes:* This table describes the survival of non-randomized and randomized results under the proportional selection adjustment with varying  $R_{max}$  using the alternative  $R_{max}$  parametrization. Both rows show the share of results which would survive  $\tilde{\delta}$  with varying  $R_{max}$  values with survival defined as the identified set does not include zero. The analysis includes only results which move toward zero when controls are added.

Figure 1: Stability Results Using Additive  $R_{max}$  Parametrization

(a) Rejection of Zero, Non-Randomized,  $R_{max} = \tilde{R} + \Pi(\tilde{R} - \hat{R})$ .



(b) Rejection of Zero, Randomized,  $R_{max} = \tilde{R} + \Pi(\tilde{R} - \hat{R})$ .



*Notes:* These graphs show the performance of non-randomized results (Sub-Figure a) and randomized results (Sub-Figure b) with under the proportional selection adjustment. Each figure graphs the share of results which would survive varying parametrizations of  $R_{max}$ , in all cases assuming  $\tilde{\delta} = 1$ . Each Sub-Figure indicates the share of results which would survive  $R_{max} = \tilde{R} + \Pi(\tilde{R} - \hat{R})$  for varying values of  $\Pi$ , with the survival in this case meaning the identified set does not include zero. This figure contains only relationships where the effect is significant with controls and adding the controls moves the coefficient toward zero.

## Appendix D: References for Section 5

- Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman**, “Reference Points and Effort Provision,” *AMERICAN ECONOMIC REVIEW*, APR 2011, *101* (2), 470–492.
- Acemoglu, Daron, Simon Johnson, James A. Robinson, and Pierre Yared**, “Income and democracy,” *AMERICAN ECONOMIC REVIEW*, JUN 2008, *98* (3), 808–842.
- Aghion, Philippe, Robin Burgess, Stephen J. Redding, and Fabrizio Zilibotti**, “The Unequal Effects of Liberalization: Evidence from Dismantling the License Raj in India,” *AMERICAN ECONOMIC REVIEW*, SEP 2008, *98* (4), 1397–1412.
- Aker, Jenny C., Christopher Ksoll, and Travis J. Lybbert**, “Can Mobile Phones Improve Learning? Evidence from a Field Experiment in Niger,” *AMERICAN ECONOMIC JOURNAL-APPLIED ECONOMICS*, OCT 2012, *4* (4), 94–120.
- Algan, Yann and Pierre Cahuc**, “Inherited Trust and Growth,” *AMERICAN ECONOMIC REVIEW*, DEC 2010, *100* (5), 2060–2092.
- Angrist, Joshua and Victor Lavy**, “The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial,” *AMERICAN ECONOMIC REVIEW*, SEP 2009, *99* (4), 1384–1414.
- Ashraf, Nava**, “Spousal Control and Intra-Household Decision Making: An Experimental Study in the Philippines,” *AMERICAN ECONOMIC REVIEW*, SEP 2009, *99* (4), 1245–1277.
- , **James Berry, and Jesse M. Shapiro**, “Can Higher Prices Stimulate Product Use? Evidence from a Field Experiment in Zambia,” *AMERICAN ECONOMIC REVIEW*, DEC 2010, *100* (5), 2383–2413.
- Ashraf, Quamrul and Oded Galor**, “Dynamics and Stagnation in the Malthusian Epoch,” *AMERICAN ECONOMIC REVIEW*, AUG 2011, *101* (5), 2003–2041.
- Bandiera, Oriana, Andrea Prat, and Tommaso Valletti**, “Active and Passive Waste in Government Spending: Evidence from a Policy Experiment,” *AMERICAN ECONOMIC REVIEW*, SEP 2009, *99* (4), 1278–1308.
- Beaman, Lori, Raghavendra Chattopadhyay, Esther Duflo, Rohini Pande, and Petia Topalova**, “POWERFUL WOMEN: DOES EXPOSURE REDUCE BIAS?,” *QUARTERLY JOURNAL OF ECONOMICS*, NOV 2009, *124* (4), 1497–1540.
- Beaudry, Paul, David A. Green, and Benjamin Sand**, “Does Industrial Composition Matter for Wages? A Test of Search and Bargaining Theory,” *ECONOMETRICA*, MAY 2012, *80* (3), 1063–1104.
- Becker, Sascha O. and Ludger Woessmann**, “WAS WEBER WRONG? A HUMAN CAPITAL THEORY OF PROTESTANT ECONOMIC HISTORY,” *QUARTERLY JOURNAL OF ECONOMICS*, MAY 2009, *124* (2), 531–596.
- Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott**, “Multiple-Product Firms and Product Switching,” *AMERICAN ECONOMIC REVIEW*, MAR 2010, *100* (1), 70–97.

- Bloom, Nicholas, Raffaella Sadun, and John Van Reenen**, “Americans Do IT Better: US Multinationals and the Productivity Miracle,” *AMERICAN ECONOMIC REVIEW*, FEB 2012, *102* (1), 167–201.
- Bohnet, Iris, Fiona Greig, Benedikt Herrmann, and Richard Zeckhauser**, “Betrayal aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States,” *AMERICAN ECONOMIC REVIEW*, MAR 2008, *98* (1), 294–310.
- Boivin, Jean, Marc P. Giannoni, and Ilian Mihov**, “Sticky Prices and Monetary Policy: Evidence from Disaggregated US Data,” *AMERICAN ECONOMIC REVIEW*, MAR 2009, *99* (1), 350–384.
- Broda, Christian, Nuno Limao, and David E. Weinstein**, “Optimal Tariffs and Market Power: The Evidence,” *AMERICAN ECONOMIC REVIEW*, DEC 2008, *98* (5), 2032–2065.
- Brunnermeier, Markus K. and Stefan Nagel**, “Do wealth fluctuations generate time-varying risk aversion? Micro-evidence on individuals’ asset allocation,” *AMERICAN ECONOMIC REVIEW*, JUN 2008, *98* (3), 713–736.
- Buera, Francisco J., Alexander Monge-Naranjo, and Giorgio E. Primiceri**, “Learning the Wealth of Nations,” *ECONOMETRICA*, JAN 2011, *79* (1), 1–45.
- Burde, Dana and Leigh L. Linden**, “Bringing Education to Afghan Girls: A Randomized Controlled Trial of Village-Based Schools,” *AMERICAN ECONOMIC JOURNAL-APPLIED ECONOMICS*, JUL 2013, *5* (3), 27–40.
- Bursztyn, Leonardo and Lucas C. Coffman**, “The Schooling Decision: Family Preferences, Intergenerational Conflict, and Moral Hazard in the Brazilian Favelas,” *JOURNAL OF POLITICAL ECONOMY*, JUN 2012, *120* (3), 359–397.
- Bustos, Paula**, “Trade Liberalization, Exports, and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinian Firms,” *AMERICAN ECONOMIC REVIEW*, FEB 2011, *101* (1), 304–340.
- Caballero, Ricardo J., Takeo Hoshi, and Anil K. Kashyap**, “Zombie Lending and Depressed Restructuring in Japan,” *AMERICAN ECONOMIC REVIEW*, DEC 2008, *98* (5), 1943–1977.
- Cai, Hongbin, Yuyu Chen, and Hanming Fang**, “Observational Learning: Evidence from a Randomized Natural Field Experiment,” *AMERICAN ECONOMIC REVIEW*, JUN 2009, *99* (3), 864–882.
- Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez**, “Inequality at Work: The Effect of Peer Salaries on Job Satisfaction,” *AMERICAN ECONOMIC REVIEW*, OCT 2012, *102* (6), 2981–3003.
- Carpenter, Jeffrey, Peter Hans Matthews, and John Schirm**, “Tournaments and Office Politics: Evidence from a Real Effort Experiment,” *AMERICAN ECONOMIC REVIEW*, MAR 2010, *100* (1), 504–517.
- Chaney, Eric**, “REVOLT ON THE NILE: ECONOMIC SHOCKS, RELIGION, AND POLITICAL POWER,” *ECONOMETRICA*, SEP 2013, *81* (5), 2033–2053.
- Charles, Kerwin Kofi and Jonathan Guryan**, “Prejudice and Wages: An Empirical Assessment of Becker’s The Economics of Discrimination,” *JOURNAL OF POLITICAL ECONOMY*, OCT 2008, *116* (5), 773–809.
- , **Erik Hurst, and Nikolai Roussanov**, “CONSPICUOUS CONSUMPTION AND RACE,” *QUARTERLY JOURNAL OF ECONOMICS*, MAY 2009, *124* (2), 425–467.
- Chen, Roy and Yan Chen**, “The Potential of Social Identity for Equilibrium Selection,” *AMERICAN ECONOMIC REVIEW*, OCT 2011, *101* (6), 2562–2589.

- Chetty, Raj, Adam Looney, and Kory Kroft**, “Salience and Taxation: Theory and Evidence,” *AMERICAN ECONOMIC REVIEW*, SEP 2009, *99* (4), 1145–1177.
- , **John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan**, “How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star,” *QUARTERLY JOURNAL OF ECONOMICS*, NOV 2011, *126* (4), 1593–1660.
- Cohen, Jessica and Pascaline Dupas**, “FREE DISTRIBUTION OR COST-SHARING? EVIDENCE FROM A RANDOMIZED MALARIA PREVENTION EXPERIMENT,” *QUARTERLY JOURNAL OF ECONOMICS*, FEB 2010, *125* (1), 1–45.
- Cohen, Lauren, Andrea Frazzini, and Christopher Malloy**, “The Small World of Investing: Board Connections and Mutual Fund Returns,” *JOURNAL OF POLITICAL ECONOMY*, OCT 2008, *116* (5), 951–979.
- Cole, Shawn, Xavier Gine, Jeremy Tobacman, Petia Topalova, Robert Townsend, and James Vickery**, “Barriers to Household Risk Management: Evidence from India,” *AMERICAN ECONOMIC JOURNAL-APPLIED ECONOMICS*, JAN 2013, *5* (1), 104–135.
- Cortes, Patricia**, “The effect of low-skilled immigration on U. S. prices: Evidence from CPI data,” *JOURNAL OF POLITICAL ECONOMY*, JUN 2008, *116* (3), 381–422.
- Dafny, Leemore, Mark Duggan, and Subramaniam Ramanarayanan**, “Paying a Premium on Your Premium? Consolidation in the US Health Insurance Industry,” *AMERICAN ECONOMIC REVIEW*, APR 2012, *102* (2), 1161–1185.
- Das, Jishnu, Stefan Dercon, James Habyarimana, Pramila Krishnan, Karthik Muralidharan, and Venkatesh Sundararaman**, “School Inputs, Household Substitution, and Test Scores,” *AMERICAN ECONOMIC JOURNAL-APPLIED ECONOMICS*, APR 2013, *5* (2), 29–57.
- Dell, Melissa**, “The Persistent Effects of Peru’s Mining Mita,” *ECONOMETRICA*, NOV 2010, *78* (6), 1863–1903.
- Djankov, Simeon, Oliver Hart, Caralee McLiesh, and Andrei Shleifer**, “Debt Enforcement around the World,” *JOURNAL OF POLITICAL ECONOMY*, DEC 2008, *116* (6), 1105–1150.
- Duflo, Esther, Michael Kremer, and Jonathan Robinson**, “Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya,” *AMERICAN ECONOMIC REVIEW*, OCT 2011, *101* (6), 2350–2390.
- , **Pascaline Dupas, and Michael Kremer**, “Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya,” *AMERICAN ECONOMIC REVIEW*, AUG 2011, *101* (5), 1739–1774.
- , **Rema Hanna, and Stephen P. Ryan**, “Incentives Work: Getting Teachers to Come to School,” *AMERICAN ECONOMIC REVIEW*, JUN 2012, *102* (4), 1241–1278.
- Dupas, Pascaline and Jonathan Robinson**, “Why Don’t the Poor Save More? Evidence from Health Savings Experiments,” *AMERICAN ECONOMIC REVIEW*, JUN 2013, *103* (4), 1138–1171.
- Duranton, Gilles and Matthew A. Turner**, “The Fundamental Law of Road Congestion: Evidence from US Cities,” *AMERICAN ECONOMIC REVIEW*, OCT 2011, *101* (6), 2616–2652.
- Estevadeordal, Antoni, Caroline Freund, and Emanuel Ornelas**, “DOES REGIONALISM AFFECT TRADE LIBERALIZATION TOWARD NONMEMBERS?,” *QUARTERLY JOURNAL OF ECONOMICS*, NOV 2008, *123* (4), 1531–1575.
- Fairlie, Robert W. and Jonathan Robinson**, “Experimental Evidence on the Effects of Home Computers on Academic Achievement among Schoolchildren,” *AMERICAN ECONOMIC JOURNAL-APPLIED ECONOMICS*, JUL 2013, *5* (3), 211–240.



- Ferraz, Claudio and Frederico Finan**, “Electoral Accountability and Corruption: Evidence from the Audits of Local Governments,” *AMERICAN ECONOMIC REVIEW*, JUN 2011, 101 (4), 1274–1311.
- Finan, Frederico and Laura Schechter**, “Vote-Buying and Reciprocity,” *ECONOMETRICA*, MAR 2012, 80 (2), 863–881.
- Finkelstein, Amy**, “E-ZTAX: TAX SALIENCE AND TAX RATES,” *QUARTERLY JOURNAL OF ECONOMICS*, AUG 2009, 124 (3), 969–1010.
- Gabaix, Xavier and Augustin Landier**, “Why has CEO pay increased so much?,” *QUARTERLY JOURNAL OF ECONOMICS*, FEB 2008, 123 (1), 49–100.
- Gentzkow, Matthew and Jesse M. Shapiro**, “Preschool television viewing and adolescent test scores: Historical evidence from the Coleman study,” *QUARTERLY JOURNAL OF ECONOMICS*, FEB 2008, 123 (1), 279–323.
- and —, “What Drives Media Slant? Evidence From US Daily Newspapers,” *ECONOMETRICA*, JAN 2010, 78 (1), 35–71.
- Gruber, Jonathan and Daniel M. Hungerman**, “The church versus the mall: What happens when religion faces increased secular competition?,” *QUARTERLY JOURNAL OF ECONOMICS*, MAY 2008, 123 (2), 831–862.
- Guiso, Luigi, Paola Sapienza, and Luigi Zingales**, “CULTURAL BIASES IN ECONOMIC EXCHANGE?,” *QUARTERLY JOURNAL OF ECONOMICS*, AUG 2009, 124 (3), 1095–1131.
- Guryan, Jonathan, Kory Kroft, and Matthew J. Notowidigdo**, “Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments,” *American Economic Journal: Applied Economics*, October 2009, 1 (4), 34–68.
- Heffetz, Ori and Moses Shayo**, “How Large Are Non-Budget-Constraint Effects of Prices on Demand?,” *AMERICAN ECONOMIC JOURNAL-APPLIED ECONOMICS*, OCT 2009, 1 (4), 170–199.
- Ifcher, John and Homa Zarghamee**, “Happiness and Time Preference: The Effect of Positive Affect in a Random-Assignment Experiment,” *AMERICAN ECONOMIC REVIEW*, DEC 2011, 101 (7), 3109–3129.
- Jayachandran, Seema and Adriana Lleras-Muney**, “LIFE EXPECTANCY AND HUMAN CAPITAL INVESTMENTS: EVIDENCE FROM MATERNAL MORTALITY DECLINES,” *QUARTERLY JOURNAL OF ECONOMICS*, FEB 2009, 124 (1), 349–397.
- Jensen, Robert**, “Do Labor Market Opportunities Affect Young Women’s Work and Family Decisions? Experimental Evidence from India,” *QUARTERLY JOURNAL OF ECONOMICS*, MAY 2012, 127 (2), 753–792.
- Jensen, Robert T. and Nolan H. Miller**, “Giffen Behavior and Subsistence Consumption,” *AMERICAN ECONOMIC REVIEW*, SEP 2008, 98 (4), 1553–1577.
- Karlan, Dean S. and Jonathan Zinman**, “Credit elasticities in less-developed economies: Implications for microfinance,” *AMERICAN ECONOMIC REVIEW*, JUN 2008, 98 (3), 1040–1068.
- Khwaja, Asim Ijaz and Atif Mian**, “Tracing the Impact of Bank Liquidity Shocks: Evidence from an Emerging Market,” *AMERICAN ECONOMIC REVIEW*, SEP 2008, 98 (4), 1413–1442.
- Kirwan, Barrett E.**, “The Incidence of US Agricultural Subsidies on Farmland Rental Rates,” *JOURNAL OF POLITICAL ECONOMY*, FEB 2009, 117 (1), 138–164.
- Kling, Jeffrey R., Sendhil Mullainathan, Eldar Shafir, Lee C. Vermeulen, and Marian V. Wrobel**, “Comparison Friction: Experimental Evidence from Medicare Drug Plans,” *QUARTERLY JOURNAL OF ECONOMICS*, FEB 2012, 127 (1), 199–235.

- Kremer, Michael, Jessica Leino, Edward Miguel, and Alix Peterson Zwane**, “SPRING CLEANING: RURAL WATER IMPACTS, VALUATION, AND PROPERTY RIGHTS INSTITUTIONS,” *QUARTERLY JOURNAL OF ECONOMICS*, FEB 2011, *126* (1), 145–205.
- Kroft, Kory, Fabian Lange, and Matthew J. Notowidigdo**, “Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment,” *QUARTERLY JOURNAL OF ECONOMICS*, AUG 2013, *128* (3), 1123–1167.
- Lalive, Rafael and Josef Zweimueller**, “HOW DOES PARENTAL LEAVE AFFECT FERTILITY AND RETURN TO WORK? EVIDENCE FROM TWO NATURAL EXPERIMENTS,” *QUARTERLY JOURNAL OF ECONOMICS*, AUG 2009, *124* (3), 1363–1402.
- Lavy, Victor**, “Performance Pay and Teachers’ Effort, Productivity, and Grading Ethics,” *AMERICAN ECONOMIC REVIEW*, DEC 2009, *99* (5), 1979–2011.
- Libecap, Gary D. and Dean Lueck**, “The Demarcation of Land and the Role of Coordinating Property Institutions,” *JOURNAL OF POLITICAL ECONOMY*, JUN 2011, *119* (3), 426–467.
- Mas, Alexandre and Enrico Moretti**, “Peers at Work,” *AMERICAN ECONOMIC REVIEW*, MAR 2009, *99* (1), 112–145.
- Mian, Atif and Amir Sufi**, “THE CONSEQUENCES OF MORTGAGE CREDIT EXPANSION: EVIDENCE FROM THE US MORTGAGE DEFAULT CRISIS,” *QUARTERLY JOURNAL OF ECONOMICS*, NOV 2009, *124* (4), 1449–1496.
- Michalopoulos, Stelios and Elias Papaioannou**, “PRE-COLONIAL ETHNIC INSTITUTIONS AND CONTEMPORARY AFRICAN DEVELOPMENT,” *ECONOMETRICA*, JAN 2013, *81* (1), 113–152.
- Muralidharan, Karthik and Venkatesh Sundararaman**, “Teacher Performance Pay: Experimental Evidence from India,” *JOURNAL OF POLITICAL ECONOMY*, FEB 2011, *119* (1), 39–77.
- Nunn, Nathan**, “The long-term effects of Africa’s slave trades,” *QUARTERLY JOURNAL OF ECONOMICS*, FEB 2008, *123* (1), 139–176.
- and **Leonard Wantchekon**, “The Slave Trade and the Origins of Mistrust in Africa,” *AMERICAN ECONOMIC REVIEW*, DEC 2011, *101* (7), 3221–3252.
- Olken, Benjamin A. and Patrick Barron**, “The Simple Economics of Extortion: Evidence from Trucking in Aceh,” *JOURNAL OF POLITICAL ECONOMY*, JUN 2009, *117* (3), 417–452.
- Pope, Devin G. and Maurice E. Schweitzer**, “Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes,” *AMERICAN ECONOMIC REVIEW*, FEB 2011, *101* (1), 129–157.
- Price, Joseph and Justin Wolfers**, “RACIAL DISCRIMINATION AMONG NBA REFEREES,” *QUARTERLY JOURNAL OF ECONOMICS*, NOV 2010, *125* (4), 1859–1887.
- Rockoff, Jonah E., Douglas O. Staiger, Thomas J. Kane, and Eric S. Taylor**, “Information and Employee Evaluation: Evidence from a Randomized Intervention in Public Schools,” *AMERICAN ECONOMIC REVIEW*, DEC 2012, *102* (7), 3184–3213.
- Schularick, Moritz and Alan M. Taylor**, “Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870–2008,” *AMERICAN ECONOMIC REVIEW*, APR 2012, *102* (2), 1029–1061.
- Shayo, Moses and Asaf Zussman**, “Judicial Ingroup Bias in the Shadow of Terrorism,” *QUARTERLY JOURNAL OF ECONOMICS*, AUG 2011, *126* (3), 1447–1484.
- Snyder, Jr. James M. and David Stromberg**, “Press Coverage and Political Accountability,” *JOURNAL OF POLITICAL ECONOMY*, APR 2010, *118* (2), 355–408.

- Song, Zheng, Kjetil Storesletten, and Fabrizio Zilibotti**, “Growing Like China,” *AMERICAN ECONOMIC REVIEW*, FEB 2011, *101* (1), 196–233.
- Spolaore, Enrico and Romain Wacziarg**, “THE DIFFUSION OF DEVELOPMENT,” *QUARTERLY JOURNAL OF ECONOMICS*, MAY 2009, *124* (2), 469–529.
- Sullivan, Daniel and Till von Wachter**, “JOB DISPLACEMENT AND MORTALITY: AN ANALYSIS USING ADMINISTRATIVE DATA,” *QUARTERLY JOURNAL OF ECONOMICS*, AUG 2009, *124* (3), 1265–1306.
- Voors, Maarten J., Eleonora E. M. Nillesen, Philip Verwimp, Erwin H. Bulte, Robert Lensink, and Daan P. Van Soest**, “Violent Conflict and Behavior: A Field Experiment in Burundi,” *AMERICAN ECONOMIC REVIEW*, APR 2012, *102* (2), 941–964.
- Wei, Shang-Jin and Xiaobo Zhang**, “The Competitive Saving Motive: Evidence from Rising Sex Ratios and Savings Rates in China,” *JOURNAL OF POLITICAL ECONOMY*, JUN 2011, *119* (3), 511–564.