

# Post-Instrument Bias in Linear Models

Adam N. Glynn\*      Miguel R. Rueda†      Julian Schuessler‡

July 1, 2021

## Abstract

Post-instrument covariates are often included as controls in IV analyses to address a violation of the exclusion restriction. However, we show that such analyses are subject to biases unless strong assumptions hold. Using linear constant-effects models, we present asymptotic bias formulas for three estimators (with and without measurement error): IV with post-instrument covariates, IV without post-instrument covariates, and OLS. In large samples and when the model provides a reasonable approximation, these formulas sometimes allow the analyst to bracket the parameter of interest with two estimators and allow the analyst to choose the estimator with the least asymptotic bias. We illustrate these points with a discussion of Angrist (1990) and a re-analysis of Acemoglu, Johnson, and Robinson (2001).

---

\*Associate Professor and Winship Distinguished Research Professor Department of Political Science and Quantitative Theory and Methods, Emory University. Email: adam.glynn@emory.edu

†Assistant Professor. Department of Political Science, Emory University. Email: miguel.rueda@emory.edu

‡Post-Doc, Department of Political Science, University of Aarhus. Email: julians@ps.au.dk.

Modern methodological research on instrumental variables (IV) has paid much attention to the fundamental role of ignorability and exclusion restriction (“no direct effect”) assumptions, leading to a more careful justification of the use of variables as instruments (Bollen 2012). While an additional focus has been on understanding IV estimators in models with heterogeneous causal effects, leading some analysts to cast doubt on the use of covariate-adjusted two-stage least-squares (2SLS) estimators (Morgan and Winship 2015, p.316), covariate adjustment in linear IV regressions remains popular. Indeed, this seems to be done not to examine heterogeneity, but in order to justify the more fundamental ignorability and exclusion restriction assumptions. However, it is often unclear whether such adjustment strategies lead to valid IV estimators.

For example, a major point of disagreement in the recent discussion between Polavieja (2015, 2017) and Chou (2017) about a method based on IV centered on exactly this issue. One position was that adjusting for “channeling variables”—those influenced by the instrument—can be detrimental (Polavieja 2017, p.448), while Chou notes that in addition to this leading to “post-treatment bias,” failure to control for these covariates may in some cases lead to violations of the exclusion restrictions, so that there would be “no fix” (Chou 2017, p.439). A leading econometrics textbook, on the other hand, seems to advocate statistical control in such situations (Wooldridge 2010, p.94), and one can find further examples of this in empirical practice (e.g., Sharkey, Torrats-Espinosa and Takyar 2017; de Vaan and Stuart 2019; VanHeuvelen 2020), including one of the most cited social science papers of the last 20 years (Acemoglu, Johnson and Robinson 2001).

Can covariate adjustments address violations of the exclusion restriction? If so, what are the conditions under which this is achieved? If such conditions are not met, how do the bias of OLS and IV estimators compare with and without the covariate adjustment that attempts to address such violation? This paper provides insights into these questions focusing on the linear model with constant-effects.

We first present large sample bias formulas for the IV and OLS estimators with the post-instrument covariate and the IV without this variable. Controlling for post-instrument covariates potentially induces a variant of “endogenous selection bias” (Elwert and Winship 2014). In causal graph terms, this is due to a “collider” structure. Accordingly, there is indeed a trade-off between this bias and bias from violations to the exclusion restriction when no such control is undertaken.

However, the formulas make clear that, in some applications, the effect of interest can be bounded by its estimates from models with and without the post-instrument covariate included as control. We illustrate this point with an example based on Angrist (1990), the study of the effects of serving in the military on earnings whose research design has inspired many others in economics, political science, and sociology (e.g. Erickson and Stoker 2011; Conley and Heerwig 2011; Heerwig and Conley 2013; Davenport 2015).

We also provide general results regarding the comparison of the biases of these estimators. We find that, without additional assumptions, invariance between IV estimates (with and without this post-instrument covariate) does not imply that the exclusion restriction holds or that the estimates are not biased. Moreover, we show that the IV will have bias at least as large as OLS when 1) the instrument has the same magnitude of effect on the causal variable and the post-instrument covariate and 2) the magnitude of the unmeasured confounding is the same for the causal variable and the post-instrument covariate.

In addition, we provide bias formulas that include the effects of measurement error in the central independent (treatment) variable and the post-instrument covariate. These bias formulas show that classical measurement error in the post-instrument covariate does not necessarily lead to attenuation, and relatedly, measurement error makes it difficult to predict the sign of the bias of the coefficient of interest.

We follow by illustrating how these points can inform researchers in an applied setting by re-analyzing the Acemoglu, Johnson, and Robinson (2001) (AJR) study of the effect

of protection against expropriation on GDP. The study's main conclusion is that better protection of property rights has a positive causal effect on economic development. Our model without measurement error confirms AJRs prediction that the IV estimated effect of the protection against expropriations index on GDP is understated due to unmeasured common causes of ethnic fractionalization (a post-instrument variable) and GDP. However, we also show that if ethnic fractionalization is measured with error, the IV estimate will be biased upwards.

We also show how our results can be used to understand the sources of bias. In AJR, the authors offer as one potential explanation for the difference between their OLS and IV results attenuation bias in the OLS estimates caused by classical measurement error. Using AJRs IV estimate in their baseline specification as the true effect and the fact that the OLS estimate is smaller than that of the IV, our analysis implies that at least 35% of the variance in the expropriation variable must be due to measurement error to rationalize their findings.

Finally, we comment on the implications of the paper's results, emphasizing alternatives to controlling for post-instrument variables. Our goal is to provide advice to practitioners facing different choices regarding which variables should be added as controls in their analysis and, in particular, we discuss reporting standards that should be upheld when addressing a violation of the exclusion restriction by adding a covariate is deemed worthwhile. To conclude, we highlight the fact that all of our results are in the context of linear constant-effects models and point readers to relevant literature that deals with heterogeneous effects. We use constant-effect models not because we believe that such parametric models are necessarily realistic, but because they allow for transparent and constructive derivations of biases (Pearl 2013; Elwert and Pfeffer 2020; Elwert and Segarra 2020). Furthermore, they naturally lead to estimation routines based on OLS or 2SLS which still are extremely popular in practice, and give us some identification power in the empirically relevant case when there is measurement error in the post-instrument variable.

Our paper belongs to the growing literature that studies the appropriate set of controls to achieve identification of causal effects. The consequences of adjusting for post-treatment variables—those affected by the treatment—has received the most attention (Rosenbaum 1984; Angrist and Pischke 2009; Montgomery, Nyhan and Torres 2018). Although this work highlights the risks linked to conditioning on post-treatment variables, recent findings identify conditions under which such adjustments could be bias-reducing (Elwert and Pfeffer 2020). Less attention has been given to the role of covariate adjustments in instrumental variables analysis. Swanson et al. (2015), Canan, Lesko and Lau (2017), and Elwert and Segarra (2020) study bias induced by controlling for a covariate affected by the treatment in instrumental variable settings, but do not study situations in which covariates are included as a way to address a violation of the exclusion restriction. Deuchert and Huber (2017) explore situations in which an instrument affects more than one variable that has effects on an outcome of interest and where researchers control for one of them. We extend this analysis by providing large sample bias formulas for the IV and OLS estimators with the post-instrument covariate which can help researchers bound the effects of interest. We also show that invariance between IV estimates with and without this post-instrument covariate—an informal test found in the literature to show robustness to violations of exclusion restrictions—is not indicative of consistency of estimates of causal effects, and analyse post-instrument variable adjustment when such variable is measured with error. Finally, our paper also contributes to work that explores violations of exclusion restrictions and sensitivity analyses to such violations (e.g. Conley, Hansen and Rossi 2012; Betz, Cook and Hollenbach 2018).

## Models for bias formulas and comparison

We are interested in situations in which a researcher wants to estimate the effect of an explanatory variable  $x$  on a dependent variable  $y$  but worries about an unmeasured common cause of  $x$  and  $y$  or classical measurement error in  $x$ . Suppose we have a linear model

$$y = \beta_0 + \beta_x x + \epsilon_0,$$

with  $E[\epsilon_0] = 0$  and  $cov(x, \epsilon_0) \neq 0$ , where the latter expression might be due to an omitted variable or measurement error. To address the endogeneity problem, the researcher considers using a variable  $z$  as an instrument. For an IV regression to give consistent estimates of  $\beta_x$ , three conditions must hold: the model must be correct - specifically,  $z$  cannot enter the equation for  $y$  (exclusion restriction), the instrument must be related to  $x$  ( $cov(x, z) \neq 0$ ), and it must not be related to other determinants of  $y$  ( $cov(z, \epsilon_0) = 0$ ). Unfortunately, the researcher is concerned that  $z$  violates the first condition, by having an effect on  $y$  through  $w$ , an observed variable available to the researcher. Our goal is to determine the consequences of including  $w$  as a control in the IV regression.

To fix ideas, consider an example from Angrist (1990), whose identification strategy has inspired multiple studies on the effects of military service on wages, marital stability, mortality, political attitudes, and turnout, among others (e.g. Erickson and Stoker 2011; Conley and Heerwig 2011, 2012; Heerwig and Conley 2013; Davenport 2015). In Angrist (1990), the author is interested in estimating the effect of serving in the Vietnam war on earnings. Angrist notes that men who have a low draft lottery number were more likely to serve in the war and uses functions of this number as instruments of military service.

Although the number that determines draft eligibility is chosen randomly, there could be some concerns about the validity of the exclusion restriction. For example, those who received a low lottery number could have chosen to stay in school to obtain a deferment (Angrist 1990, p. 330). This creates a direct link between salaries and lottery numbers, which invalidates the exclusion restriction. If information on post-lottery education was available, should we control for it? Wooldridge (2010) points out that not doing so would violate the condition  $cov(z, \epsilon_0) = 0$  that is needed for consistent estimation (Wooldridge

2010, p. 94, 95). Can this justify the inclusion of the education variable as a control?

The fundamental problem with this approach is the possibility that  $w$  is itself affected by the error term. This is represented by the dashed arrow from  $\epsilon$  to  $w$  in Figure 1, which summarizes the situation. In the example, measures of post-lottery educational attainment could take the role of  $w$ . Arguably, this variable is also influenced by innate ability, parents' levels of education, and other unobserved determinants of income.

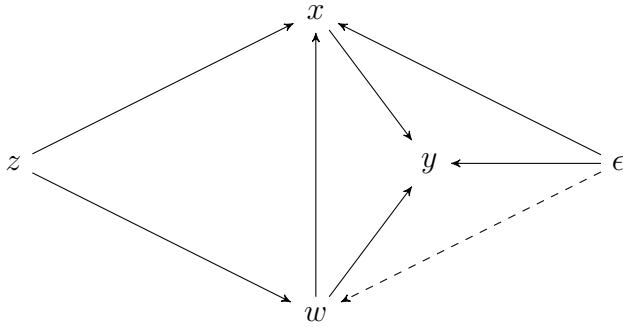


Figure 1: Model's graphical representation

There are a couple important observations about the model in Figure 1. First, we do not have an arrow from  $x$  to  $w$ . If we did, the graph would be cyclic, further complicating identification. Second, while we do have an arrow from  $w$  to  $x$ , this could be replaced (or complemented) with an unmeasured common cause of  $x$  and  $w$  without changing the results that follow.

In order to asses the relative value of “fixing” the IV estimation by controlling for  $w$ , we also consider as an alternative OLS estimates. The new causal model for  $y$  that serves as the basis for both the fixed IV and the OLS approach is then

$$y = \beta_0 + \beta_x x + \beta_w w + \epsilon.$$

Note that while this structural model makes a strong constant effects assumption, we

do not make any assumptions about the shape of the distributions of the involved variables (including  $\epsilon$ ). Furthermore, our analysis does not make any assumptions on the causal models for the other observed variables ( $z, x, w$ ). Indeed, the consistency of 2SLS estimation does not rely on a correctly specified first-stage model (Vansteelandt and Didelez 2018, Proposition 3).

Let  $\sigma_i$  denote the standard deviation of variable  $i$  and  $\rho_{ij}$  denote the correlation between variables  $i$  and  $j$ , where  $i, j \in \{w, x, y, z\}$ . Note that in Figure 1, correlations are induced by open paths between variables (Elwert 2013). The next result allows us to compare the probability limits of the estimates of  $\beta_x$ . All derivations are in the appendix.

**Proposition 1.** *The probability limits of IV estimates of  $\beta_x$  with and without  $w$  and the OLS estimates with  $w$  for the model in Figure 1 are:*

$$\begin{aligned} p\lim \widehat{\beta}_x^{IV \text{ no } w} &= \beta_x + \frac{\sigma_w \rho_{zw} \beta_w}{\sigma_x \rho_{zx}}, \\ p\lim \widehat{\beta}_x^{IV \text{ w}} &= \beta_x - \frac{\sigma_\epsilon \rho_{zw} \rho_{w\epsilon}}{\sigma_x (\rho_{zx} - \rho_{xw} \rho_{zw})}, \\ p\lim \widehat{\beta}_x^{OLS \text{ w}} &= \beta_x + \frac{\sigma_\epsilon (\rho_{x\epsilon} - \rho_{xw} \rho_{w\epsilon})}{\sigma_x (1 - \rho_{xw}^2)}. \end{aligned}$$

According to this result, there are two conditions that make  $\widehat{\beta}_x^{IV \text{ w}}$  consistent for  $\beta_x$ : if  $w$  and  $\epsilon$  are uncorrelated ( $\rho_{w\epsilon} = 0$ ), or if  $w$  and  $z$  are uncorrelated ( $\rho_{zw} = 0$ ). While the first of them is hypothetically possible, the second is ruled out by assumption, as the researcher thinks the instrument is impacting the outcome through  $w$ . If  $w$  is affected by the error ( $\rho_{w\epsilon} \neq 0$ ), the result tells us it is possible to obtain worse estimates by running an IV regression even when the instrument is not related to  $y$  through different channels than  $x$  and  $w$ .

This is a variant of endogenous selection bias (Elwert and Winship 2014):  $w$  is a collider on the path  $z \rightarrow w \leftarrow \epsilon \rightarrow y$ . Controlling empirically for  $w$  does block direct effects of the instrument on the outcome, but at the same time opens up this path, leading to an

alternative non-causal correlation.

However, note that with regards to the estimators that adjust for  $w$ , the only parameters that cannot be estimated are  $\sigma_\epsilon$ ,  $\rho_{x\epsilon}$ , and  $\rho_{w\epsilon}$ . Among them,  $\sigma_\epsilon$  appears in the numerator of the bias terms of the estimates that include  $w$ . So, if a researcher is interested in assessing relative bias of the OLS and IV estimators when including  $w$ , she only needs to provide a statement about the relative values of  $\rho_{x\epsilon}$  and  $\rho_{w\epsilon}$  (as  $\sigma_\epsilon$  cancels). Equipped with sample estimates of the  $\rho_{zx}$  and  $\rho_{xw}$ , a researcher could determine combinations of  $\rho_{x\epsilon}$  and  $\rho_{w\epsilon}$  that would make one estimator preferred to the other one. Clearly, substantive knowledge regarding the sources of endogeneity and, in particular, the sign of the unknown correlations could further limit the range of the values of  $\rho_{x\epsilon}$  and  $\rho_{w\epsilon}$  that are relevant for this comparison, as well as bounds for the ratio of the biases.

Similarly, note that knowledge of the sign of these correlations and the fact that they are in the interval  $[-1, 1]$  also allow the researcher to determine bounds of the effect of interest. To illustrate, we can return to the Vietnam-draft example. Recall that in this case,  $y$  captures earnings,  $z$  is a dummy variable that equals one if the person was draft eligible,  $x$  is a dummy variable that equals one if the person is a veteran, and  $w$  is a dummy variable that equals one if the person attended college. Since both, draft eligibility and unobserved ability, could encourage college enrollment,  $\rho_{zw} > 0$  and  $\rho_{w\epsilon} > 0$ . In addition, one would expect a positive effect of college attendance on earnings ( $\beta_w > 0$ ) as well as a positive correlation between draft eligibility and veteran status ( $\rho_{zx} > 0$ ). Given that the correlations between veteran status and college attendance is likely smaller than the one between draft eligibility and veteran status, the estimated IV coefficient on veteran status of the model that does not control for college attendance would overstate the true effect. On the other hand, the IV coefficient on veteran status in the model that controls for college attendance would understate it. That is, the IV estimate with  $w$  would bound the true effect from below, while the IV estimate without  $w$  would bound it from above.

More generally, the following corollary indicates when the adjusted IV estimates with  $w$  and without it can bound the effect of interest.

**Corollary 1.** *If  $\text{sgn}\left(\frac{\beta_w}{\rho_{zx}}\right) = \text{sgn}\left(\frac{\rho_{w\epsilon}}{\rho_{zx} - \rho_{xw}\rho_{zw}}\right)$ , then  $\beta$  is bounded by  $\widehat{\beta}_x^{\text{IV no } w}$  and  $\widehat{\beta}_x^{\text{IV } w}$ .*

However, finding an equality of adjusted and unadjusted estimates is not necessarily informative. It turns out that when the scaled effect of  $w$  on  $y$  equals the negative of the scaled confounding of  $w$  and  $y$ , IV with and without  $w$  produces similar estimates. Therefore, seeing that the IV estimate does not change after controlling for  $w$  does not mean estimates are consistent nor that the concerns about violating the exclusion restriction are unimportant:

**Corollary 2.** *If  $\frac{\sigma_w\beta_w}{\rho_{zx}} = -\frac{\sigma_\epsilon\rho_{w\epsilon}}{(\rho_{zx} - \rho_{xw}\rho_{zw})}$ , then  $\text{plim } \widehat{\beta}_x^{\text{IV no } w} = \text{plim } \widehat{\beta}_x^{\text{IV } w}$ .*

We can also consider the simple case where we remove any arrows between  $x$  and  $w$ , confounding is equally bad for  $x$  and  $w$ , and  $z$  is an equally strong instrument for  $x$  and  $w$ . Under those conditions, we can prove that OLS has less large-sample bias than IV.

**Corollary 3.** *If  $|\rho_{x\epsilon}| = |\rho_{w\epsilon}|$  and  $|\rho_{zx}| = |\rho_{zw}|$  then  $|Bias(\widehat{\beta}_x^{\text{OLS}})| \leq |Bias(\widehat{\beta}_x^{\text{IV}})|$ .*

This result gives a quick way to compare relative biases when assuming the absence of any causal relationships between the treatment and the post-instrument covariate is plausible. The result implies that even in linear constant-effects models, in order for IV with a post-instrument covariate to be preferred to OLS, one would need to estimate  $\rho_{zx}$  and  $\rho_{zw}$ , and then argue that  $\rho_{w\epsilon}$  was sufficiently small vis-a-vis  $\rho_{x\epsilon}$ .

## Measurement error

One reason why researchers employ instrumental variables is to address concerns regarding measurement error in their treatment variable and the biases that this induces in their estimates. We now explore situations in which measurement error affects the explanatory

variables  $x$  and  $w$  while maintaining all other relationships between the variables as described by the model in Figure 1.

Let  $\tilde{x}$  and  $\tilde{w}$  denote the observed covariates with  $\tilde{x} = x + u_x$  and  $\tilde{w} = w + u_w$ , where  $u_x$  and  $u_w$  are zero-mean measurement errors with variances  $\sigma_{u_x}^2$  and  $\sigma_{u_w}^2$ . We focus on the case of classical measurement error, in which the measurement errors are not related to the true value of the explanatory variables. We further assume that the measurement error terms are not correlated with the error term in the population model,  $\epsilon$ , the instrument,  $z$ , nor each other.

It is well known that instrumental variables regression with a valid instrument can give us consistent estimates of  $\beta_x$  even when we only observe  $\tilde{x}$ . We are now interested in studying the question of how the IV estimates perform when there is a violation of the exclusion restriction and we control for an additional variable  $\tilde{w}$  that captures the alternative link (other than through  $x$ ) between the instrument and the outcome. The following proposition gives expressions characterizing large sample bias of such an approach.

**Proposition 2.** *If  $E[xu_x] = E[xu_w] = E[wu_x] = E[wu_w] = E[u_xu_w] = E[u_x\epsilon] = E[u_w\epsilon] = E[zu_w] = E[zu_x] = 0$ , the probability limits of IV estimates of  $\beta_x$  with and without  $w$  and the OLS estimates with  $w$  are:*

$$\begin{aligned} \text{plim } \hat{\beta}_x^{IV \text{ no } w} &= \beta_x + \frac{\sigma_{\tilde{w}}\rho_{z\tilde{w}}\beta_w}{\sigma_{\tilde{x}}\rho_{z\tilde{x}}}, \\ \text{plim } \hat{\beta}_x^{IV \text{ w}} &= \beta_x + \frac{\sigma_{u_w}^2\rho_{z\tilde{w}}\beta_w}{\sigma_{\tilde{x}}\sigma_{\tilde{w}}(\rho_{z\tilde{x}} - \rho_{\tilde{x}\tilde{w}}\rho_{z\tilde{w}})} - \frac{\sigma_\epsilon\rho_{z\tilde{w}}\rho_{\tilde{w}\epsilon}}{\sigma_{\tilde{x}}(\rho_{z\tilde{x}} - \rho_{\tilde{x}\tilde{w}}\rho_{z\tilde{w}})}. \\ \text{plim } \hat{\beta}_x^{OLS \text{ w}} &= \beta_x \left(1 - \frac{\sigma_{u_x}^2}{\sigma_{\tilde{x}}^2(1 - \rho_{\tilde{x}\tilde{w}}^2)}\right) + \frac{\sigma_\epsilon(\rho_{\tilde{x}\epsilon} - \rho_{\tilde{x}\tilde{w}}\rho_{\tilde{w}\epsilon})}{\sigma_{\tilde{x}}(1 - \rho_{\tilde{x}\tilde{w}}^2)} + \frac{\sigma_{u_w}^2\rho_{\tilde{x}\tilde{w}}\beta_w}{\sigma_{\tilde{x}}\sigma_{\tilde{w}}(1 - \rho_{\tilde{x}\tilde{w}}^2)}, \end{aligned}$$

Note that, when using OLS, the estimate is not going to be necessarily biased towards zero. We also see that including  $\tilde{w}$  in the instrumental variable regression adds a term in

the bias expression that is proportional to the variance of the measurement error of  $w$  and to its effect on the outcome. This makes clear that including  $\tilde{w}$  will not necessarily bias the IV estimate toward zero either. It also indicates that a researcher trying to address a violation of the exclusion restriction by adding a regressor should not only be concerned about potential unmeasured common causes of  $y$  and  $w$  (or reverse causality), but also about measurement error in the added regressor.

The formula for the OLS asymptotic bias also makes clear that unlike in the case without measurement error, a comparison of the relative biases with  $w$  included between estimators is not straightforward. This is because the ratio of such biases depends on  $\beta_x$ ,  $\beta_w$ , and  $\sigma_{uw}$ , in addition to  $\sigma_\epsilon$ ,  $\rho_{\tilde{x}\epsilon}$  and  $\rho_{\tilde{w}\epsilon}$ . This further highlights the difficulties of attempting to address a violation of the exclusion restriction by adding a post-instrument covariate when measurement error is an additional concern.

## Application

To illustrate how these results can be applied, consider AJR, where the authors are interested in estimating the effects of institutions on economic performance. Their dependent variable,  $y$ , is GDP per capita in 1995. Their main explanatory variable,  $x$ , is an index of protection against expropriation. An analysis based on OLS regressions is unlikely to give accurate estimates, since: 1) it is difficult to account for all common causes of institutions and economic performance, 2) economic performance can determine the protection of property rights, and 3) the index against expropriation is measured with error and cannot capture all institutional arrangements that lead to property right protection. To address these issues, AJR propose as an instrument of the index of expropriation the mortality rates of settlers in the colonization period,  $z$ . AJR argue that in places where Europeans faced higher mortality rates, they could not settle and were more likely to impose extractive economic institutions—those that impose property right protections only to the elites. The fact that powerful economic and

Table 1: Point Estimates of the Effects of Institutions on Economic Performance

Explanatory variables	OLS	IV	
	(1)	(2)	(3)
Protection against expropriation ( $x$ )	0.46	0.94	0.74
Ethnolinguistic fragmentation ( $w$ )	-1.3		-1.02
<i>First stage results</i>			
Log European settler mortality ( $z$ )		-0.61	-0.64

Column (1) corresponds to Model (7) Table 6 Panel C in AJR, column (2) corresponds to Model (1) Table 4 Panels A and B in AJR, and column (3) corresponds to Model (7) from Table 6 Panels A and B in AJR.

political elites exert effective efforts to maintain the institutions that are favorable to them explain why the settler mortality instrument would impact current indexes of expropriation.

AJR are aware of potential violations of the exclusion restriction and present as robustness tests results in which they control for variables that capture alternative links between the settler mortality rate and GDP per capita. One way in which the exclusion restriction can be violated, for example, is through the potential impact of settler mortality on ethnolinguistic fragmentation. Because ethnolinguistic fragmentation has been linked to the provision of public goods, incidence of conflict, social capital, and directly to economic growth (Alesina et al. 2003; Montalvo and Reynal-Querol 2005; Bjørnskov 2008); a potential impact of settler mortality on fragmentation could bias the effect of interest. We consider AJR's estimates from Tables 4 and 6 and Appendix A, which we reproduce here in Table 1 and focus on the regressions which include as a control a measure of ethnolinguistic fragmentation.<sup>1</sup>

Using their data, we compute  $\rho_{zw}$ , finding it positive, and  $\rho_{z\tilde{w}} - \rho_{\tilde{x}\tilde{w}}\rho_{z\tilde{w}}$ , which is negative.<sup>2</sup>

---

<sup>1</sup>Unlike many other papers that add post-instrument controls, AJR's report OLS estimates as well as IV results for all specifications, discuss whether the added regressors can be correlated to the error term in the main model, and make available their data for replication.

<sup>2</sup>The computed values of these correlations and standard deviations are:  $\rho_{\tilde{x}\tilde{w}} =$

Therefore, if we assume based on the literature findings (and as in AJR) that  $\rho_{\tilde{w}\epsilon} < 0$  and that there is no measurement error in the added regressor ( $\sigma_{u_w} = 0$ ), Proposition 2 concurs with the AJR analysis that the IV analysis including ethnolinguistic fragmentation has negative bias.<sup>3</sup>

On the other hand, the difficulties of measuring fragmentation suggest that assuming no measurement error might be too strong (Okediji 2005). If  $\sigma_{u_w} \neq 0$ , Proposition 2 indicates that when ethnolinguistic fragmentation has a negative effect on GDP (as suggested by the literature), the second term of the *plim* of the IV with  $\tilde{w}$  will be positive. Hence it is not easy to determine the sign of the bias as it depends on the fragmentation effect ( $\beta_w$ ), the confounding ( $\rho_{we}$ ), and the measurement error ( $\sigma_{u_w}^2$ ).

Our results also allows us to explain possible sources of differences between OLS and IV estimates. AJR's OLS estimate is smaller than the IV estimate that conditions on ethnic fractionalization and also that of the model that does not. AJR explain this difference as a result of attenuation bias caused by measurement error in the institutions variable with OLS. Using Proposition 2, we examine under what conditions measurement error is consistent with those differences. An inspection of the *plim* for OLS in Proposition 2 together with our computed correlations indicates that the third term in that expression is positive. Moreover, the second term in the expression of the *plim* for OLS is likely positive as well, since  $\rho_{\tilde{x}\epsilon}$  will be positive and possibly larger than  $\rho_{\tilde{x}\tilde{w}} \cdot \rho_{\tilde{w}\epsilon}$  (we estimate  $\rho_{\tilde{x}\tilde{w}} = -0.22$ ). What would be the minimum variance in the measurement error of the institutional variable that is consistent with the observed differences between the OLS and IV estimates?

To answer that question, we set the two last (positive) terms in the expression of the *plim*

---

$-0.22, \rho_{z\tilde{w}} = 0.49, \rho_{z\tilde{x}} = -0.52, \sigma_{\tilde{x}} = 1.47, \sigma_{\tilde{w}} = 0.32$ , which imply  $\rho_{z\tilde{x}} - \rho_{\tilde{x}\tilde{w}}\rho_{z\tilde{w}} = -0.41$ .

<sup>3</sup>Their analysis, included in Appendix A page 1396, differs from ours as it treats the institutional variable as exogenous while the added covariate (ethnolinguistic fragmentation) is correlated with the error term from the second stage.

for OLS to zero. We then use AJR's IV estimate in the model that includes ethnolinguistic fragmentation in place of  $\beta_x$  and solve for  $\sigma_{u_x}^2$  to obtain 0.77. This means that measurement error must account for at least 35% ( $\approx \frac{0.77}{1.47^2}$ ) of the variation in the observed institutions variable to rationalize the difference between IV and the OLS estimates. Clearly, we could have chosen a different estimate of  $\beta_x$  to compute that minimum measurement error variance. If we use instead AJR's baseline estimate of 0.94, the measurement error in the institutions variable must make up nearly 50% ( $\approx \frac{1.05}{1.47^2}$ ) of the total variance.

It is important to note that the previous analysis relies on assumptions highlighted in Proposition 2 that are in line with classical measurement error. It also relies on the assumptions used in the linear constant-effect models of AJR and this paper. We do believe, however, that these assumptions provide a first step to analyze this application in a way that is consistent with reasonable scenarios for the situation of interest.

## Discussion and conclusion

Instrumental variables regression methods allow researchers to address estimation challenges like unobserved heterogeneity and classical measurement error. Whether the method delivers accurate results depends on the tenability of its assumptions. Here, we have studied one way in which researchers have dealt with potential violations of one of them, the exclusion restriction. We find that although it is possible for researchers to fix violations of the exclusion restriction by adding controls, doing so requires a number of strong assumptions. When these do not hold, the IV estimates can be worse than what the researcher would obtain running an OLS regression.

Our findings have a number of implications for practice. First, even when the linear constant-effects model is used, the inclusion of a post-instrument covariate in 2SLS estimation requires strong theoretical assumptions, and therefore, researchers may want to use

alternative approaches. One of them is to focus on the estimated effect of the instrument (the “reduced form effect”) which will not be invalidated by an exclusion restriction violation. In many cases, when the IV analysis is questionable, this effect will have some theoretical or policy relevance that could be emphasized. Another approach is to conduct a sensitivity analysis with respect to the exclusion restriction (e.g. see Conley, Hansen and Rossi (2012)), as the results might be robust to these violations.

Second, if conditioning on post-instrument covariates is conducted, the corresponding OLS analysis should be reported and the measurement error and unmeasured common causes in both approaches should be discussed in concert. The formulas presented here should help with this discussion and depending on the application of interest, might give bounds for the effect of interest.

Finally, we note that the entirety of this paper relies on the linear constant-effects model. If the constant-effects model is not a reasonable approximation, then there are many potential parameters of interest from an instrumental variables analysis (Imbens et al. 2014), and assessment of exclusion restriction violations becomes more complicated. Flores and Flores-Lagunes (2013) and Mealli and Pacini (2013) provide some strategies in this context and also surveys of the literature.

## References

- Acemoglu, Daron, Simon Johnson and James A. Robinson. 2001. “The Colonial Origins of Comparative Development: An Empirical Investigation.” *The American Economic Review* 91(5):1369 – 1401.
- Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat and Romain Wacziarg. 2003. “Fractionalization.” *Journal of Economic Growth* 8:155–194.
- Angrist, Joshua D. 1990. “Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records.” *The American Economic Review* 80(3):313–336.
- Angrist, Joshua David and Jörn-Steffen Pischke. 2009. *Mostly harmless econometrics: an empiricist’s companion*. Princeton University Press.
- Betz, Timm, Scott J Cook and Florian M Hollenbach. 2018. “On the use and abuse of spatial instruments.” *Political Analysis* pp. 1–6.
- Bjørnskov, Christian. 2008. “Social Trust and Fractionalization: A Possible Reinterpretation.” *European Sociological Review* 24(3):271–283.
- Bollen, Kenneth A. 2012. “Instrumental variables in sociology and the social sciences.” *Annual Review of Sociology* 38:37–72.
- Canan, Chelsea, Catherine Lesko and Bryan Lau. 2017. “Instrumental Variable Analyses and Selection Bias.” *Epidemiology* 28(3):396–398.
- Chou, Winston. 2017. “Culture remains elusive: on the identification of cultural effects with instrumental variables.” *American Sociological Review* 82(2):435–443.

- Conley, Dalton and Jennifer Heerwig. 2011. “The War at Home: Effects of Vietnam-Era Military Service on Postwar Household Stability.” *American Economic Review* 101(3):350–54.
- Conley, Dalton and Jennifer Heerwig. 2012. “The Long-Term Effects of Military Conscription on Mortality: Estimates From the Vietnam-era Draft Lottery.” *Demography* 49(3):841–55.
- Conley, Timothy G, Christian B Hansen and Peter E Rossi. 2012. “Plausibly exogenous.” *Review of Economics and Statistics* 94(1):260–272.
- Davenport, Tiffany C. 2015. “Policy-induced Risk and Responsive Participation: The Effect of a Son’s Conscription Risk on the Voting Behavior of his Parents.” *American Journal of Political Science* 59(1):225–241.
- de Vaan, Mathijs and Toby Stuart. 2019. “Does Intra-household Contagion Cause an Increase in Prescription Opioid Use?” *American Sociological Review* 84(4):577–608.
- Deuchert, Eva and Martin Huber. 2017. “A cautionary tale about control variables in IV estimation.” *Oxford Bulletin of Economics and Statistics* 79(3):411–425.
- Elwert, Felix. 2013. Graphical causal models. In *Handbook of causal analysis for social research*, ed. Stephen L Morgan. Springer pp. 245–273.
- Elwert, Felix and Christopher Winship. 2014. “Endogenous selection bias: The problem of conditioning on a collider variable.” *Annual Review of Sociology* 40:31–53.
- Elwert, Felix and Elan Segarra. 2020. Treatment Induced Selection Bias in Instrumental Variables Analysis: Exact Result. In *Probabilistic and Causal Inference: The Works of Judea Pearl*, ed. Hector Geffner, Rina Dechter and Joseph Y. Halpern. ACM Books.

Elwert, Felix and Fabian T. Pfeffer. 2020. “The Future Strikes Back: Using Future Treatments to Detect and Reduce Hidden Bias.” *Sociological Methods & Research* 0(0):0049124119875958. forthcoming.

Erickson, Robert and Laura Stoker. 2011. “Caught in the Draft: The Effects of Vietnam Draft Lottery Status on Political Attitudes.” *American Political Science Review* 105(2):221–237.

Flores, Carlos A and Alfonso Flores-Lagunes. 2013. “Partial identification of local average treatment effects with an invalid instrument.” *Journal of Business & Economic Statistics* 31(4):534–545.

Heerwig, Jennifer A. and Dalton Conley. 2013. “The Causal Effects of Vietnam-era Military Service on Post-war Family Dynamics.” *Social Science Research* 42(2):299–310.

Imbens, Guido W et al. 2014. “Instrumental Variables: An Econometrician’s Perspective.” *Statistical Science* 29(3):323–358.

Mealli, Fabrizia and Barbara Pacini. 2013. “Using secondary outcomes to sharpen inference in randomized experiments with noncompliance.” *Journal of the American Statistical Association* 108(503):1120–1131.

Montalvo, José G. and Marta Reynal-Querol. 2005. “Ethnic Polarization, Potential Conflict, and Civil Wars.” *American Economic Review* 95(3):796–816.

Montgomery, Jacob M, Brendan Nyhan and Michelle Torres. 2018. “How conditioning on posttreatment variables can ruin your experiment and what to do about it.” *American Journal of Political Science* 62(3):760–775.

Morgan, Stephen L and Christopher Winship. 2015. *Counterfactuals and causal inference*. Cambridge University Press.

Okediji, Tade O. 2005. “The Dynamics of Ethnic Fragmentation.” *American Journal of Economics and Sociology* 64(2):637–662.

Pearl, Judea. 2013. “Linear models: A useful “microscope” for causal analysis.” *Journal of Causal Inference* 1(1):155–170.

Polavieja, Javier G. 2015. “Capturing Culture: A New Method to Estimate Exogenous Cultural Effects Using Migrant Populations.” *American Sociological Review* 80(1):166–191.

Polavieja, Javier G. 2017. “Culture as a random treatment: a reply to Chou.” *American Sociological Review* 82(2):444–450.

Rosenbaum, Paul R. 1984. “The consequences of adjustment for a concomitant variable that has been affected by the treatment.” *Journal of the Royal Statistical Society. Series A (General)* pp. 656–666.

Sharkey, Patrick, Gerard Torrats-Espinosa and Delaram Takyar. 2017. “Community and the Crime Decline: The Causal Effect of Local Nonprofits on Violent Crime.” *American Sociological Review* 82(6):1214–1240.

Swanson, Sonja A, James M. Robins, Matthew Miller and Miguel A. Hernán. 2015. “Instrumental Variable Analyses and Selection Bias.” *American Journal of Epidemiology* 181(3):191–197.

VanHeuvelen, Tom. 2020. “The Right to Work, Power Resources, and Economic Inequality.” *American Journal of Sociology* 125(5):1255–1302.

Vansteelandt, Stijn and Vanessa Didelez. 2018. “Improving the robustness and efficiency of covariate-adjusted linear instrumental variable estimators.” *Scandinavian Journal of Statistics* 45(4):941–961.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.