Post-Instrument Bias in Linear Models

January 11, 2023

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 Acemoglu, Johnson, and Robinson (2001).

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 in order to justify the 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 an application of IV centered on this issue. One position was that adjusting for "channeling variables"—those influenced by the instrument—can be detrimental (Polavieja 2017, p.448), especially when the controls could be affected by the treatment. But Chou notes that while adjusting for them may indeed lead to bias, failure to do so 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 situations where the control is affected by the instrument (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*a*).

What are the conditions under which covariate adjustments can address violations of the exclusion restriction? If such conditions are not met, how do the bias of OLS and IV estimators compare with and without the covariate adjustment? 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 Acemoglu, Johnson and Robinson (2001*a*) (AJR) study of the effect of protection against expropriation on GDP.

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. This highlights the known fact that the exclusion restriction is not testable in just-identified models. 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 asymptotic bias formulas for the IV estimator 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 continue to re-analyze AJR to illustrate how these points can inform researchers in an applied setting. AJR'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 show that if ethnic fractionalization is measured with error, the IV estimate will be biased upwards.

We also illustrate 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. 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 (e.g., Flores and Flores-Lagunes 2013; Mealli and Pacini 2013; Schuessler, Glynn and Rueda 2021). 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) and they naturally lead to estimation routines based on OLS or 2SLS.

Our paper belongs to the literature that studies the appropriate set of controls to achieve identification of causal effects. The consequences of adjusting for variables 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 post-treatment covariate and sample selection 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 and where researchers control for one of them. Hughes et al. (2019) study several models where the instrument influences sample selection and there are variables that affect sample inclusion that could be adjusted for. We complement these analyses by providing large sample bias formulas for the IV and OLS estimators with the post-instrument covariate and analyze the 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 β_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.

As our running example, we consider Acemoglu, Johnson and Robinson (2001*a*). In this paper, the authors want to estimate the effect of protection of property rights (measured with an index of protection against expropriation), x, on GDP per capita, y. An analysis based on OLS regressions will not give accurate estimates, since: it is difficult to account for all common causes of economic institutions and economic performance, and any 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. AJR argue that in places were Europeans faced higher mortality rates, they could not settle in mass and were more likely to impose extractive economic institutions—those that give property right protections only to the elites. The fact that economic and political elites have an advantaged position in society to maintain the rules that favor them explains why the settler mortality in the colonial period 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 of them is ethnolinguistic fragmentation, z. 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. The question that arises is whether the authors should control for measures of ethnolinguistic fragmentation to alleviate a violation of the exclusion restriction.

The fundamental problem with this approach is the possibility that w is itself affected by the error term (Brito and Pearl 2002). Our interpretation of the situation studied by AJR is presented by the causal graph (Pearl 2009) in Figure 1.[1] The relationship between ϵ and w is represented by the dashed arrow. This is because, arguably, ethnolinguistic fragmentation is also influenced by cultural, geographical, and historical factors that are unobserved determinants of GDP per capita.

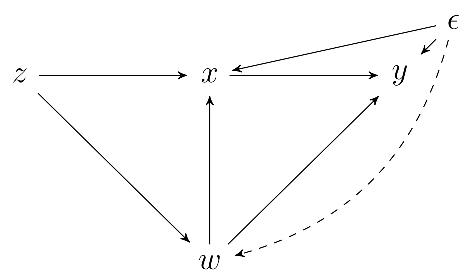

Figure 1: Model's graphical representation

Note that we do not have an arrow from x to w. If we did, the graph would be cyclic, further complicating identification. The existing arrow from w to x can capture how high ethnolinguistic fragmentation could lead to violations of property rights of minorities in cases where there is ethnic conflict.

In order to assess the relative value of "fixing" the IV estimation by controlling for w, we also consider as an alternative OLS estimates. In accordance with Figure 1, 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 ϵ). Furthermore, our analysis does not make any assumptions on the functional form of 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). If one introduces stronger linearity assumptions for the other observed variables, one can also derive our results directly from the causal graph using the rules of path-tracing (Pearl 2013).

Let σ_i denote the standard deviation of variable i and ρ_{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 β_x. All derivations are in the appendix.

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

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

According to this result, there are two conditions that make $\widehat{\beta}_x^{IV\ \mathrm{w}}$ consistent for β_x: if w and ϵ 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 (compared to OLS) 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 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 (see also Hughes et al. (2019) and Elwert and Segarra (2020) for similar cases where sample selection impacted by the instrument violates the exclusion restriction). Along these lines, w in Figure 1 does not d-separate y from z in the graph that removes the arrow from x to y, which is one of the sufficient conditions for identification of the effect in linear models (Brito and Pearl 2002).

However, note that with regards to the estimators that adjust for w, the only parameters that cannot be estimated are σ_ϵ, $\rho_{x\epsilon}$, and $\rho_{w\epsilon}$. Among them, σ_ϵ 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 σ_ϵ cancels). Equipped with sample estimates of the ρ_{zx} and ρ_{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, the expressions in Proposition 1 allow for a comparison of the biases between the IV adjusted and unadjusted models. This comparison only depends on the relative sizes of β_w and $\sigma_\epsilon\rho_{w\epsilon}$ after replacing the rest of parameters with their sample analogs.

Also note that knowledge of the sign of the correlations in the expressions of Proposition 1 and the fact that they are in the interval $[-1, 1]$ allow the researcher to determine asymptotic bounds of the effect of interest. To illustrate, we can return to the settler mortality example. Recall that in this case, y captures GDP per capita, z is colonial settler mortality, x is an index of protection against expropriation, and w is a measure of ethnolinguistic fragmentation. If higher settler mortality induces a more hierarchical social structure separating natives, slaves, and Europeans, $\rho_{zw} > 0$. Because ethnolinguistic fragmentation has been linked with more conflict and less social capital $\rho_{w\epsilon} < 0$ and, possibly, $\beta_w < 0$. In addition, if the correlations between fragmentation and protection against expropriation (ρ_{xw}) and between fragmentation and settler mortality (ρ_{zw}) are weaker than the correlation of settler mortality and protection against expropriation (ρ_{zx})—as they are in AJR data—, the IV coefficient on protection against expropriation in the model that controls for fragmentation would understate the effect. That is, the IV estimate with w would (asymptotically) 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* $sgn\left(\frac{\beta_w}{\rho_{zx}}\right) = sgn\left(\frac{\rho_{w\epsilon}}{\rho_{zx}-\rho_{xw}\rho_{zw}}\right)$, *then* β *is asymptotically bounded by* $\widehat{\beta}_x^{IV\, no\ w}$ *and* $\widehat{\beta}_x^{IV\ w}$.

In our discussion of the AJR example, we have assumed that all variables are measured *without error* to illustrate how a researcher could use Proposition 1. In the next section, we examine what happens if the x or w are measured with error, and return to the settler mortality example.

Proposition 1 also highlights that 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. The result is inline with the fact that it is not possible to test the exclusion restriction in just-identified models without further assumptions.

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

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* $\left|Bias(\widehat{\beta}_x^{OLS})\right| \leq \left|Bias(\widehat{\beta}_x^{IV})\right|$.

This result gives a way to compare relative biases in large samples 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 ρ_{zx} and ρ_{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 $\sigma^2_{u_x}$ and $\sigma^2_{u_w}$. 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, ϵ, the instrument, z, nor each other.

It is well known that instrumental variables regression with a valid instrument can give us consistent estimates of β_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* $E[xu_x] = E[xu_w] = E[wu_x] = E[wu_w] = E[u_x u_w] = E[u_x \epsilon] = E[u_w \epsilon] = E[zu_w] = E[zu_x] = 0$*, the probability limits of IV estimates of* β_x *with and without* w *and the OLS estimates with* w *are:*

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

It is known that OLS estimates will not be necessarily biased towards zero when other mismeasured variables are adjusted for (e.g, Garber and Klepper 1980). More interestingly, 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. Note also that even if $\tilde{w}$ and ϵ are uncorrelated, adjusting for this noisy covariate would not fully intercept the exclusion violation. This leads to an inconsistent estimator where the asymptotic bias will depend on the relationships between the instrument and the post-instrument variable, instrument and treatment, and treatment and post-instrument variable. In sum, the proposition highlights 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, 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 β_x, β_w, and

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

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

Outcome, y, is GDP per capita. 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.

σ_{u_w}, in addition to σ_ϵ, $\rho_{\tilde{x}\epsilon}$ 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, we return to the AJR's example. We now allow for the protection against expropriation and ethnolinguistic fragmentation variables to have an additive error as described in the previous section and label them $\tilde{x}$ and $\tilde{w}$, respectively. This is consistent with AJR's own motivation for using an IV strategy and the difficulties of measuring ethnic diversity and fragmentation (Okediji 2005). For ease of exposition, we reproduce AJR's estimates from their Tables 4 and 6 and Appendix A in Table 1, focusing on the regressions which include as a control the measure of ethnolinguistic fragmentation.[2]

Using their data (Acemoglu, Johnson and Robinson 2001*b*), the sample estimates of the correlations are $\rho_{z\tilde{w}} = 0.49$, $\rho_{z\tilde{x}} = -0.52$, and $\rho_{\tilde{x}\tilde{w}} = -0.22$. This implies that $\rho_{z\tilde{x}} - \rho_{\tilde{x}\tilde{w}}\rho_{z\tilde{w}} = -0.41$.[3] 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.[4]

However, with measurement error in the post-instrument covariate ($\sigma_{u_w} \neq 0$), 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 (β_w), the confounding ($\rho_{w\epsilon}$), and the variance of the error ($\sigma^2_{u_w}$).

Our results allow 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 than 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 $\rho_{\tilde{x}\tilde{w}} \cdot \rho_{\tilde{w}\epsilon}$. 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* for OLS to zero. We then use AJR's IV estimate in the model that includes ethnolinguistic fragmentation in place of β_x, 0.74, and solve for $\sigma^2_{u_x}$ to obtain 0.77.[5] 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 β_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. Finally, in Schuessler, Glynn and Rueda (2021) we provide a nonparametric analysis of the problem and develop a semi-parametric sensitivity analysis.

Author's note

Data and code to replicate all empirical results found in this paper can be found under http://dx.doi.org/10.17605/OSF.IO/Z38RX.

Notes

1. AJR do not provide a model's graphical representation.

2. 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 and code for replication.

3. The computed values of the standard deviations are: $\sigma_{\tilde{x}} = 1.47$, $\sigma_{\tilde{w}} = 0.32$.

4. Their analysis, included in Appendix A page 1396, differs from ours as it treats the protection against expropriation as exogenous while the added covariate (ethnolinguistic fragmentation) is correlated with the error term from the second stage.

5. That is, we solve for $\sigma^2_{u_x}$ in expression

$$plim\ \widehat{\beta}_x^{OLS\ \mathrm{w}} = \beta_x \left(1 - \frac{\sigma^2_{u_x}}{\sigma^2_{\tilde{x}}(1 - \rho^2_{\tilde{x}\tilde{w}})}\right),$$

where we assume $plim\ \widehat{\beta}_x^{OLS\ \mathrm{w}}$ is 0.46 (their OLS estimate that adjust for fractionalization), replace β_x by 0.74, and use the estimated correlation $\rho_{\tilde{x}\tilde{w}} = -0.22$ and variance $\sigma_{\tilde{x}} = 1.47$.

References

Acemoglu, Daron, Simon Johnson and James A. Robinson. 2001*a*. "The Colonial Origins of Comparative Development: An Empirical Investigation." *The American Economic Review* 91(5):1369 – 1401.

Acemoglu, Daron, Simon Johnson and James A. Robinson. 2001*b*. "Replication Data: The Colonial Origins of Comparative Development: An Empirical Investigation.". data retrieved from Daron Acemoglu's website, https://economics.mit.edu/faculty/acemoglu/data/ajr2001.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat and Romain Wacziarg. 2003. "Fractionalization." *Journal of Economic Growth* 8:155–194.

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.

Brito, Carlos and Judea Pearl. 2002. "Generalized Instrumental Variables." *Proceedings Conference on Uncertainty in Artificial Intelligence (UAI)* pp. 85–93.

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, Timothy G, Christian B Hansen and Peter E Rossi. 2012. "Plausibly exogenous." *Review of Economics and Statistics* 94(1):260–272.

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.

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.

Garber, Steven and Steven Klepper. 1980. "Extending the Classical Normal Errors-in-Variables Model." *Econometrica* 48(6):1541–1546.

Hughes, Rachael A., Neil M. Davies, George Davey Smith and Kate Tillinga. 2019. "Selection Bias When Estimating Average Treatment Effects Using One-sample Instrumental Variable Analysis." *Epidemiology* 30:350—-357.

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. 2009. *Causality.* Cambridge university press.

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 consquences 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.

Schuessler, Julian, Adam N Glynn and Miguel R Rueda. 2021. "Post-Instrument Bias." *Working Paper* .

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.