Post-Instrument Bias

Adam N. Glynn and Miguel R. Rueda

October 5, 2017

Abstract

Post-instrument covariates are often included in IV analyses to address a violation of the exclusion restriction. We demonstrate that even in linear constant-effects models with large samples: 1) invariance between IV estimates (with and without post-instrument covariates) does not imply that the exclusion restriction holds with respect to the post-instrument covariate, 2) OLS with an omitted variable will often have less bias than IV with the post-instrument covariate, 3) measurement error in the post-instrument covariate does not necessarily lead to attenuation, and 4) the bias of OLS and IV are related. Therefore, if used, IV with a post-instrument covariate should be paired with OLS, and results should be discussed in concert. We illustrate these points with a re-analysis of Acemoglu, Johnson, and Robinson (2001), showing that for the paper's claims to be valid, at least 35% of the variance in the causal explanatory variable must be due to measurement error.

Introduction

Post-instrument covariates are often included in instrumental variables analyses (IV) to justify the exclusion restriction. For example, of the 155 papers using IV published since 2010 in the *American Political Science Review*, the *American Journal of Political Science*, and the *Journal of Politics*, we have identified 29 that use post-instrument covariates to justify the exclusion restriction.[1] Furthermore, an important econometrics textbook seems to advocate the use of this approach (Wooldridge 2010, 94), and the approach was used in one of the most cited social science papers of the last 20 years (Acemoglu, Johnson and Robinson 2001). Because detailed discussion of the exclusion restriction is a relatively recent phenomenon, it is difficult to determine how often post-instrument covariates were used in the past, but if more than 1/6 of all IV papers use the technique, this constitutes a very large number of papers in social science journals.

Despite the widespread use of the technique, to our knowledge, formal justification of its use has been limited to Appendix A of Acemoglu, Johnson, and Robinson (2001), which provides a large sample bias formula based on a linear simultaneous equations model with constant effects. In this paper, we use linear constant effects models where bias in the technique is induced either by unmeasured common causes of outcome and the post-instrument covariate or by measurement error in the post-instrument covariate. This approach allows us to compare the bias of this technique to the bias of OLS and IV without the post-instrument covariate.

The next section presents the model with and without measurement error and presents large sample bias formulas for the estimators: IV with the post-instrument covariate, IV without the post-instrument covariate, and OLS. It also presents some general results regarding the comparison of these biases. In particular, we show that even when the violation of the exclusion restriction is due entirely to a measured post-instrument covariate, invariance between IV estimates (with and without this post-instrument covariate) does not imply that the exclusion restriction holds. We also show 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. This finding implies that the use of a post-instrument covariate essentially turns an experimental study (perhaps a natural experiment) into an observational study. Finally, we provide bias formulas that include the

[1] We found that out of the 155 papers using IV, 116 explicitly discussed the exclusion restriction. In 25% of them a covariate was included to address potential violations of the exclusion restriction.

effects of measurement error on the causal 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.

The application section presents an illustration of these points by re-analyzing the Acemoglu, Johnson, and Robinson (2001) (AJR from here on) analysis of the effect of protection against expropriation on GDP. We show that our model without measurement error confirms AJRs prediction that an IV estimator with ethnic fractionalization as a post-instrument covariate would be likely to have negative bias due to unmeasured common causes of ethnic fractionalization and GDP. However, we also show that the inclusion of measurement error (for ethnic fractionalization) in the model is likely to result in positive bias. It is uncertain whether this positive bias due to measurement error would outweigh the negative bias due to unmeasured common causes. We also point out that even if we accept AJRs statement of negative bias, the OLS estimate is smaller than the IV estimate, and therefore OLS would also have to have negative bias. As AJR suggest, most unmeasured common causes of expropriation and GDP would lead to positive bias, so any negative bias in the OLS estimator would be due to measurement error in the expropriation variable. We use our measurement error model to show that these combined assumptions imply that at least 35% of the variance in the expropriation variable must be due to measurement error. If we instead take AJRs preferred IV estimate (without inclusion of ethnic fractionalization) as the true value of the effect, then nearly 50% of the variance in the expropriation variable must be due to measurement error. If we additionally allow for some of the potential unmeasured common causes of GDP and expropriation variable suggested by AJR, the percentage of variance in the expropriations variable due to measurement error would need to be greater that 50% in order to rationalize the results.

In the final section, we discuss the implications of these results for practice. We first discuss our result that conditioning on a post-instrument covariate turns an experiment or a natural experiment into an observational study, and we discuss some alternatives to this approach. We next discuss reporting standards that should be upheld when such an observational study is deemed worthwhile. Finally, we highlight the fact that all of the results in this paper are in the context of linear constant effects models. We point readers to literature that addresses exclusion restriction violations in heterogeneous effects models.

Models for bias formulas and comparison

Model without measurement error

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, classical measurement error in x. In the simplest linear model case, we have

$$y = \beta_0 + \beta_x x + \epsilon_0, \tag{1}$$

with $E[\epsilon_0] = 0$ and $cov(x, \epsilon_0) \neq 0$.

In order to address this problem, the researcher considers using a variable z as an instrument. For an instrumental variables (IV) regression to give consistent estimates of β_x, two conditions must hold: 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 second 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 several studies of political attitudes and behaviour (e.g. Bergan 2009; Erickson and Stoker 2011; Davenport 2015). The author is interested in estimating the effect of serving in the Vietnam war on the earnings of men. 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 in an earnings linear model.

Although, the number that determines draft eligibility is chosen randomly, there could be some concerns about the validity of the exclusion restriction. Those who received a low number in the lottery could have chosen to stay in school to obtain a deferment (Angrist 1990, 330). Alternatively, employers of low number holders could have invested less in training these employees knowing that they were more likely to be drafted (Wooldridge 2010, 94). Both of these situations create a direct link between salaries and lottery numbers, which invalidates the exclusion restriction. If information on job training or education were available, should we control for those variables in the earnings equation? Wooldridge (2010) points out that not doing so would violate the condition $cov(z, \epsilon_0) = 0$ that is needed for

consistent estimation of the effect of interest (Wooldridge 2010, 94, 95). Can this justify the inclusion of these variables as additional controls?

As we will see below, the fundamental problem with this approach is the possibility that w, the variable that captures the direct link between the instrument and the outcome, is itself affected by the error term. This possibility is represented by the dashed arrow from ϵ to w in Figure 1, which summarizes the situation of interest. In the previous example, measures of on the job training taken some time after the lottery or educational attainment could take the role of w. Arguably, both of these variables are also influenced by innate ability, parents' levels of education, and other unobserved determinants of income.

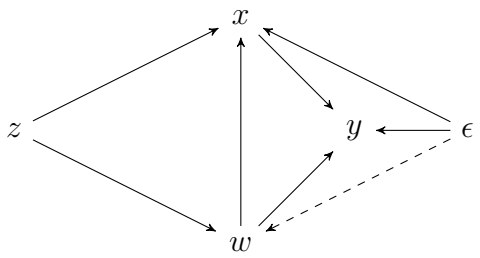

Figure 1: Model's graphical representation

In order to asses the relative value of "fixing" the IV estimation by controlling for w, we also consider as an alternative ordinary least squares (OLS) estimates. The new model for y used in both the fixed IV and the OLS approach is then

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

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, then β_x could not be interpreted as the total effect of x on y. Second, while we do have an arrow from w to x, this could be replaced with an unmeasured common cause of x and w without changing the results that follow.

Results without measurement error

Let σ_i denote the standard deviation of variable i and ρ_{ij} denote the correlation between variables i and j, where $i, j \in \{w, x, y, z\}$. The next result allows us to compare the probability limits of the estimates of β_x.

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:*

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

The result indicates that there are two conditions that make $\widehat{\beta}_x^{IV\ \mathrm{w}}$ consistent for β_x: if w and ϵ are uncorrelated, or if w and z are uncorrelated. The first of these is hypothetically possible. The second is ruled out by the assumption that the instrument is related to the dependent variable through w. The result shows that if w is affected by the error, 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. Notice also, that many of the parameters of these bias terms can be estimated with data. In fact, the only parameters that cannot be estimated are β_w, σ_ϵ, $\rho_{x\epsilon}$, and $\rho_{w\epsilon}$. Furthermore, σ_ϵ appears in the numerator of the bias terms that include w, so in order to assess relative bias of these two estimators, one only need to provide a statement about the relative values of $\rho_{x\epsilon}$ and $\rho_{w\epsilon}$. Knowledge of the sign of these correlations and the fact that they are in the interval $[-1, 1]$ can also provide the researcher with the ability to determine bounds for the relative biases.

Even in the absence of a particular data set, we can say some things about the relative size of these bias terms. First, when the scaled effect of w on y equals the negative of the scaled confounding of w and y, the *plim* of the IV without w will equal the *plim* of the IV with w.

Corollary 1. *If* $\frac{\sigma_w \beta_w}{\rho_{zx}} = -\frac{\sigma_\epsilon \rho_{w\epsilon}}{(\rho_{zx} - \rho_{xw}\rho_{zw})}$, *then* $plim\ \widehat{\beta}_x^{IV\,no\ w} = plim\ \widehat{\beta}_x^{IV\ w}$.

Importantly, it is possible for these *plim*s to be equal with $\rho_{zw} \neq 0$ and $\beta_w \neq 0$, and therefore it is possible for IV with and without w to produce similar estimates, even when there is an exclusion restriction violation through w.

Second, consider the simple case where $|\rho_{x\epsilon}| = |\rho_{w\epsilon}|$ and $|\rho_{zx}| = |\rho_{zw}|$, and where we remove any arrows between x and w, the first condition states that confounding is equally bad for x and w, and the second condition would hold if z is an equally strong instrument for x and w. We obtain the following result:

Corollary 2. *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 implies that OLS has less large-sample bias than IV with a post-instrument covariate when x and w are similar in the sense described above. Therefore, 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 $\rho_{x\epsilon}$. We have never seen such an analysis conducted. Additionally, what this result makes clear is that IV with a post-instrument covariate is in the same class of methods as OLS with a potentially omitted variable, and therefore should be considered a type of observational study.

Measurement error

In this section we explore situations in which measurement error affects the explanatory variables x and w while maintaining all other relationships between variables as described by the model in Figure 1. In particular,

$$\begin{aligned} \tilde{x} &= x + u_x \\ \tilde{w} &= w + u_w, \end{aligned} \tag{2}$$

where the tilde denotes observed variables, and u_x and u_w are zero-mean measurement errors of x and w 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 ($E[xu_x] = E[wu_w] = E[wu_x] = E[xu_w] = 0$). 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 ($E[u_x u_w] = E[u_x\epsilon] = E[u_w\epsilon] = E[zu_w] = E[zu_x] = 0$).

It is well known that instrumental variables regression with a valid instrument (one that satisfies $E[zx] \neq 0$ and $E[z\epsilon_0] = E[zu_x] = 0$) can give us consistent estimates of β_x with a population model (1) 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 approach.

Proposition 2. *The probability limits of IV and OLS estimates of β_x including $\tilde{w}$ when*

1. *Explanatory variables are measured with error according to* (2)*,*
2. *Measurement errors are not correlated with true values of explanatory variables,*

3. *Measurement errors are not correlated with the instrument, error term, ϵ, nor each other, and*

4. *All other variable relationships are as described by the model in Figure 1*

are:

$$\begin{aligned} plim\ \widehat{\beta}_x^{OLS\ w} &= \beta_x \left(1 - \frac{\sigma^2_{u_x}}{\sigma^2_{\tilde{x}}(1-\rho^2_{\tilde{x}\tilde{w}})}\right) + \frac{\sigma_\epsilon(\rho_{\tilde{x}\epsilon} - \rho_{\tilde{x}\tilde{w}}\rho_{\tilde{w}\epsilon})}{\sigma_{\tilde{x}}(1-\rho^2_{\tilde{x}\tilde{w}})} + \frac{\sigma^2_{u_w}\rho_{\tilde{x}\tilde{w}}\beta_w}{\sigma_{\tilde{x}}\sigma_{\tilde{w}}(1-\rho^2_{\tilde{x}\tilde{w}})}, \\ plim\ \widehat{\beta}_x^{IV\ w} &= \beta_x + \frac{\sigma^2_{u_w}\rho_{z\tilde{w}}\beta_w}{\sigma_{\tilde{x}}\sigma_{\tilde{w}}(\rho_{z\tilde{x}} - \rho_{\tilde{x}\tilde{w}}\rho_{z\tilde{w}})} - \frac{\sigma_\epsilon\rho_{z\tilde{w}}\rho_{\tilde{w}\epsilon}}{\sigma_{\tilde{x}}(\rho_{z\tilde{x}} - \rho_{\tilde{x}\tilde{w}}\rho_{z\tilde{w}})}. \end{aligned}$$

Note that, when using OLS, the estimate is not going to be necessarily biased towards zero as in the typical case of classical measurement error when $\tilde{w}$ is not included.[2] 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. Note that the sign of this term depends on β_w and not β_x, hence unlike with OLS, including $\tilde{w}$ in the analysis will not necessarily bias the IV estimate toward zero. This indicates that a researcher trying to address a potential 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. In the next section we discuss how these results can be used in an applied setting.

Illustrative Application

AJR 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. As discussed in their article, an analysis based on OLS regressions is unlikely to give accurate estimates of the effect of institutions on GDP per capita, since: 1) it is difficult to account for all common causes of institutions and economic performance, 2) there is the possibility that economic performance can determine the ability

[2] In that case the relevant expression is

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

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

	OLS	IV	
Explanatory variables	(1)	(2)	(3)
Protection against expropriation	0.46	0.94	0.74
Ethnolinguistic fragmentation	-1.3		-1.02
First stage results			
Log European settler mortality		-0.71	-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.

of governments to protect property rights, and 3) the index against expropriation is measured with error and cannot capture all institutional arrangements that lead to property right protection and subsequent economic prosperity. 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 were Europeans faced higher mortality rates, Europeans could not settle and were more likely to impose extractive economic institutions on the native population—the majority in those places. Institutional persistence explains why this instrument would still explain variation in current indexes of expropriation.

AJR are aware of potential violations of the exclusion restriction and their paper presents as robustness tests a number of results in which they include as controls variables that provide alternative links between the settler mortality rate and GDP per capita (other than through economic institutions). To illustrate how to apply our results, we consider AJR's estimates from Tables 4 and 6 and Appendix A and focus on the regressions which include as a control a measure of ethnolinguistic fragmentation, which takes the role of w. AJR point out that ethnolinguistic fragmentation can be considered endogenous in a GDP per capita regression, which makes it a good candidate for this illustration.[3] Table 1 reproduces their relevant results here.

Using their data, we compute ρ_{zw} finding it positive and $\rho_{z\tilde{x}} - \rho_{\tilde{x}\tilde{w}}\rho_{z\tilde{w}}$, which is negative.[4]

[3]Unlike many 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.

[4]The computed values of these correlations and standard deviations are: $\rho_{\tilde{x}\tilde{w}} = -0.22, \rho_{z\tilde{w}} = 0.49, \rho_{z\tilde{x}} = -0.52, \sigma_{\tilde{x}} = 1.47, \sigma_{\tilde{w}} = 0.32$, which imply $\rho_{z\tilde{x}} - \rho_{\tilde{x}\tilde{w}}\rho_{z\tilde{w}} = -0.41$.

Therefore, if we assume, 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 instrumental variables analysis that includes ethnolinguistic fragmentation should have negative bias.[5] However, if we allow for the possibility of measurement error in ethnolinguistic fragmentation, then this result is not so straightforward. We see from Proposition 2 that we will still get negative bias from the third term of the *plim* for the IV estimate with $\tilde{w}$. However, if we also believe that ethnolinguistic fragmentation has a negative effect on GDP ($\beta_w < 0$), then the second term of this *plim* will be positive. This means that the overall sign of the large sample bias will depend on the relative magnitudes of the effect (β_w), the confounding ($\rho_{w\epsilon}$), and the measurement error ($\sigma^2_{u_w}$).

Our results that account for measurement error also allows us to relate the previous observations to the bias in OLS and to explain possible sources of differences between OLS and IV estimates. First note in Table 1 that AJR's OLS estimate of the effect of institutions is smaller than the IV estimate that conditions on ethnic fractionalization and also that of the model that does not include this variable as an additional regressor. The only explanation AJR give for negative bias in OLS is the result of attenuation bias in the OLS estimates caused by measurement error in the institutions variable. Using Proposition 2, we examine under what conditions measurement error in the institution variable 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 terms in the expression of the *plim* for OLS (which are likely positive as described above) to zero. We then use the AJR IV estimate in the model that includes ethnolinguistic fragmentation in place of β_x and solve for $\sigma^2_{u_x}$ to obtain 0.77. This means that, given that the overall variance of the institutions variable is the sum of the true variance and the measurement error variance, 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 observed 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

[5]Their analysis of the effects of including an endogenous regressor, unlike ours, assumes the main explanatory variable of interest to be strictly exogenous.

instead AJR's baseline estimate of 0.91, the measurement error in the institutions variable must make up nearly 50% ($\approx \frac{1.05}{1.47^2}$) of the total variance. These observations indicate that for classical measurement error to account for an OLS estimate in our model as small as the one obtained by AJR, such measurement error has to be quite large.

It is important to note that the previous analysis relies on assumptions highlighted in Proposition 2 that are inline with classical measurement error. Other assumptions regarding correlations of measurement errors and the explanatory variables and error ϵ have the potential to affect these observations. We do believe, however, that this assumptions provide a first step to analyze this application in a way that is consistent with reasonable scenarios for the situation of interest as well as with AJRs own assumptions.

Discussion and conclusion

Instrumental variables regression methods allow researchers to address estimation challenges like unobserved heterogeneity, classical measurement error, and reverse causality. As many have pointed out (e.g. Bartels 1991; Sovey and Green 2011), 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 the exclusion restriction. We find that although it is possible for researchers to fix violations of the exclusion restriction or provide robustness by controlling for additional variables, doing so requires a number of 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 requires a number of strong theoretical claims, and therefore, researchers may want to use an alternative approach. One approach is to focus on the estimated effect of the instrument which will not be invalidated by an exclusion restriction violation. In many cases, the effect of the instrument (sometimes known as the intent-to-treat effect in experimental studies) will have some theoretical or policy relevance, and therefore when the IV analysis is questionable, this effect should 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 it is possible that results will be robust to violations of the exclusion restriction.

Second, if conditioning on post-instrument covariates is conducted, the results should be discussed in a manner similar to other observational studies. Measurement error and unmea-

sured common causes need to be discussed. Additionally, the corresponding OLS analysis (or analogous selection on observables analysis) should be reported and the measurement error and unmeasured common causes in both approaches should be discussed in concert. The formulas presented in this paper should help with this discussion.

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.

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.

Bartels, Larry. 1991. "Instrumental and 'Quasi Instrumental' variables." *American Journal of Political Science* 35(3):777–800.

Bergan, Daniel E. 2009. "The Draft Lottery and Attitudes Towards the Vietnam War." *Public Opinion Quarterly* 73(2):379–384.

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.

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.

Giles, David E.A. 1984. "Instrumental Variables Regressions Involving Seasonal Data." *Economic Letters* 14:339–343.

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.

Sovey, Allison J. and Donald P. Green. 2011. "Instrumental Variables Estimation in Political Science: A Readers' Guide." *American Journal of Political Science* 55(1):188–200.

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

Derivations

Proof of Proposition 1. Assume that all variables have mean zero.[6] We start by writing the expression of the probability limit of $\widehat{\beta}^{OLS}$ in matrix notation,

$$plim\ \widehat{\beta}^{OLS} = \begin{bmatrix} \beta_x \\ \beta_w \end{bmatrix} + E \begin{bmatrix} x^2 & xw \\ xw & w^2 \end{bmatrix}^{-1} E \begin{bmatrix} x\epsilon \\ w\epsilon \end{bmatrix}.$$

After computing the inverse of the variance-covariance matrix, we have

$$\begin{aligned} plim\ \widehat{\beta}_x^{OLS} &= \beta_x + \frac{E[w^2]E[x\epsilon] - E[xw]E[w\epsilon]}{E[x^2]E[w^2] - E[xw]^2} \\ &= \beta_x + \frac{E[w^2]E[x\epsilon] - E[xw]E[w\epsilon]}{\sigma_x^2\sigma_w^2(1-\rho_{xw}^2)} \\ &= \beta_x + \frac{\sigma_w^2\sigma_x\sigma_\epsilon(\rho_{x\epsilon} - \rho_{xw}\rho_{w\epsilon})}{\sigma_x^2\sigma_w^2(1-\rho_{xw}^2)} \\ &= \beta_x + \frac{\sigma_\epsilon(\rho_{x\epsilon} - \rho_{xw}\rho_{w\epsilon})}{\sigma_x(1-\rho_{xw}^2)}. \end{aligned}$$

Similarly for the $plim\ \widehat{\beta}^{IV}$,

[6] This is without loss of generality, as shown by Giles (1984).

$$
\begin{aligned}
plim\ \widehat{\beta}^{IV} &= \begin{bmatrix} \beta_x \\ \beta_w \end{bmatrix} + E\begin{bmatrix} zx & zw \\ xw & w^2 \end{bmatrix}^{-1} E\begin{bmatrix} z\epsilon \\ w\epsilon \end{bmatrix} \\
plim\ \widehat{\beta}_x^{IV} &= \beta_x + \frac{E[w^2]E[z\epsilon] - E[zw]E[w\epsilon]}{E[zx]E[w^2] - E[zw]E[xw]} \\
&= \beta_x + \frac{-E[zw]E[w\epsilon]}{\sigma_x\sigma_w^2\sigma_z(\rho_{zx} - \rho_{xw}\rho_{zw})} \\
&= \beta_x + \frac{-\sigma_w^2\sigma_z\sigma_\epsilon\rho_{zw}\rho_{w\epsilon}}{\sigma_x\sigma_w^2\sigma_z(\rho_{zx} - \rho_{xw}\rho_{zw})} \\
&= \beta_x + \frac{-\sigma_\epsilon\rho_{zw}\rho_{w\epsilon}}{\sigma_x(\rho_{zx} - \rho_{xw}\rho_{zw})},
\end{aligned}
$$

where we used the fact that $E[z\epsilon] = 0$. Finally for the $plim\ \widehat{\beta}^{IVnow}$,

$$
\begin{aligned}
plim\ \widehat{\beta}^{IVnow} &= \begin{bmatrix} \beta_x \end{bmatrix} + E\begin{bmatrix} zx \end{bmatrix}^{-1} E\begin{bmatrix} z\epsilon_0 \end{bmatrix} \\
plim\ \widehat{\beta}_x^{IVnow} &= \beta_x + \frac{E[z\epsilon_0]}{E[zx]} \\
&= \beta_x + \frac{E[z(\beta_w w + \epsilon]}{E[zx]} \\
&= \beta_x + \frac{E[z\beta_w w]}{E[zx]} \\
&= \beta_x + \frac{\beta_w\sigma_z\sigma_w\rho_{zw}}{\sigma_z\sigma_x\rho_{zx}} \\
&= \beta_x + \frac{\beta_w\sigma_w\rho_{zw}}{\sigma_x\rho_{zx}}
\end{aligned}
$$

where we used the fact that $E[z\epsilon] = 0$.

□

Proof of Corollary 2. First we examine the case $\rho_{x\epsilon} = \rho_{w\epsilon}$ and $\rho_{zx} = \rho_{zw}$. The OLS and IV biases are

$$
\begin{aligned}
\left|Bias(\widehat{\beta}_x^{OLS})\right| &= \left|\frac{\sigma_\epsilon\rho_{w\epsilon}}{\sigma_x(1+\rho_{xw})}\right|, \\
\left|Bias(\widehat{\beta}_x^{IV})\right| &= \left|\frac{\sigma_\epsilon\rho_{w\epsilon}}{\sigma_x(1-\rho_{xw})}\right|.
\end{aligned}
$$

Note that $\rho_{xw} = \rho_{zx} \cdot \rho_{zw} > 0$, and so $\left|Bias(\widehat{\beta}_x^{OLS})\right| < \left|Bias(\widehat{\beta}_x^{IV})\right|$.

Consider the case $\rho_{x\epsilon} = \rho_{w\epsilon}$ and $\rho_{zx} = -\rho_{zw}$. The OLS and IV biases are the same

$$\left|Bias(\widehat{\beta}_x^{OLS})\right| = \left|Bias(\widehat{\beta}_x^{IV})\right| = \left|\frac{\sigma_\epsilon \rho_{w\epsilon}}{\sigma_x(1+\rho_{xw})}\right|.$$

Similarly, if $\rho_{x\epsilon} = -\rho_{w\epsilon}$ and $\rho_{zx} = \rho_{zw}$, we have

$$\left|Bias(\widehat{\beta}_x^{OLS})\right| = \left|Bias(\widehat{\beta}_x^{IV})\right| = \left|\frac{\sigma_\epsilon \rho_{w\epsilon}}{\sigma_x(1-\rho_{xw})}\right|.$$

Finally, if $\rho_{x\epsilon} = -\rho_{w\epsilon}$ and $\rho_{zx} = -\rho_{zw}$, $\rho_{xw} = \rho_{zx} \cdot \rho_{zw} < 0$ and

$$\begin{aligned}\left|Bias(\widehat{\beta}_x^{OLS})\right| &= \left|\frac{\sigma_\epsilon \rho_{w\epsilon}}{\sigma_x(1-\rho_{xw})}\right|, \\ \left|Bias(\widehat{\beta}_x^{IV})\right| &= \left|\frac{\sigma_\epsilon \rho_{w\epsilon}}{\sigma_x(1+\rho_{xw})}\right|,\end{aligned}$$

giving us the same strict inequality of the first case. □

Proof of Proposition 2. We start by writing the expression of the probability limit of $\widehat{\beta}^{OLS}$ in matrix notation as before,

$$plim\ \widehat{\beta}^{OLS} = \begin{bmatrix} \beta_x \\ \beta_w \end{bmatrix} + E\begin{bmatrix} \tilde{x}^2 & \tilde{x}\tilde{w} \\ \tilde{x}\tilde{w} & \tilde{w}^2 \end{bmatrix}^{-1} E\begin{bmatrix} \tilde{x}\left(-u_x\beta_x - u_w\beta_w + \epsilon\right) \\ \tilde{w}\left(-u_x\beta_x - u_w\beta_w + \epsilon\right) \end{bmatrix},$$

where we used the fact that $y = (\tilde{x} - u_x)\beta_x + (\tilde{w} - u_w)\beta_w + \epsilon$. After computing the inverse of the variance-covariance matrix, we have

$$
\begin{aligned}
plim\ \widehat{\beta}_x^{OLS} &= \beta_x + \frac{E[\tilde{w}^2]E[\tilde{x}\left(-u_x\beta_x - u_w\beta_w + \epsilon\right)] - E[\tilde{x}\tilde{w}]E[\tilde{w}\left(-u_x\beta_x - u_w\beta_w + \epsilon\right)]}{E[\tilde{x}^2]E[\tilde{w}^2] - E[\tilde{x}\tilde{w}]^2} \\
&= \beta_x + \frac{E[\tilde{w}^2]\left(-E[\tilde{x}u_x]\beta_x - E[\tilde{x}u_w]\beta_w + E[\tilde{x}\epsilon]\right) - E[\tilde{x}\tilde{w}]\left(-E[\tilde{w}u_x]\beta_x - E[\tilde{w}u_w]\beta_w + E[\tilde{w}\epsilon]\right)}{E[\tilde{x}^2]E[\tilde{w}^2] - E[\tilde{x}\tilde{w}]^2} \\
&= \beta_x + \frac{\sigma_{\tilde{w}}^2(-\sigma_{u_x}^2\beta_x + \sigma_{\tilde{x}}\sigma_\epsilon\rho_{\tilde{x}\epsilon}) - \sigma_{\tilde{x}}\sigma_{\tilde{w}}\rho_{\tilde{x}\tilde{w}}(-\sigma_{u_w}^2\beta_w + \sigma_{\tilde{w}}\sigma_\epsilon\rho_{\tilde{w}\epsilon})}{\sigma_{\tilde{x}}^2\sigma_{\tilde{w}}^2(1-\rho_{\tilde{x}\tilde{w}}^2)} \\
&= \beta_x\left(1 - \frac{\sigma_{u_x}^2}{\sigma_{\tilde{x}}^2(1-\rho_{\tilde{x}\tilde{w}}^2)}\right) + \frac{\sigma_\epsilon(\rho_{\tilde{x}\epsilon} - \rho_{\tilde{x}\tilde{w}}\rho_{\tilde{w}\epsilon})}{\sigma_{\tilde{x}}(1-\rho_{\tilde{x}\tilde{w}}^2)} + \frac{\sigma_{u_w}^2\rho_{\tilde{x}\tilde{w}}\beta_w}{\sigma_{\tilde{x}}\sigma_{\tilde{w}}(1-\rho_{\tilde{x}\tilde{w}}^2)}.
\end{aligned}
$$

Similarly for the $plim\ \widehat{\beta}^{IV}$,

$$
\begin{aligned}
plim\ \widehat{\beta}^{IV} &= \begin{bmatrix} \beta_x \\ \beta_w \end{bmatrix} + E\begin{bmatrix} z\tilde{x} & z\tilde{w} \\ \tilde{x}\tilde{w} & \tilde{w}^2 \end{bmatrix}^{-1} E\begin{bmatrix} z\left(-u_x\beta_x - u_w\beta_w + \epsilon\right) \\ \tilde{w}\left(-u_x\beta_x - u_w\beta_w + \epsilon\right) \end{bmatrix} \\
plim\ \widehat{\beta}_x^{IV} &= \beta_x + \frac{E[\tilde{w}^2]E[z\left(-u_x\beta_x - u_w\beta_w + \epsilon\right)] - E[z\tilde{w}]E[\tilde{w}\left(-u_x\beta_x - u_w\beta_w + \epsilon\right)]}{E[z\tilde{x}]E[\tilde{w}^2] - E[z\tilde{w}]E[\tilde{x}\tilde{w}]} \\
&= \beta_x + \frac{\sigma_{\tilde{w}}\sigma_z\rho_{z\tilde{w}}\sigma_{u_w}^2\beta_w}{\sigma_{\tilde{x}}\sigma_{\tilde{w}}^2\sigma_z(\rho_{z\tilde{x}} - \rho_{x\tilde{w}}\rho_{z\tilde{w}})} + \frac{-\sigma_{\tilde{w}}^2\sigma_z\sigma_\epsilon\rho_{z\tilde{w}}\rho_{\tilde{w}\epsilon}}{\sigma_{\tilde{x}}\sigma_{\tilde{w}}^2\sigma_z(\rho_{z\tilde{x}} - \rho_{\tilde{x}\tilde{w}}\rho_{z\tilde{w}})} \\
&= \beta_x + \frac{\sigma_{u_w}^2\rho_{z\tilde{w}}\beta_w}{\sigma_{\tilde{x}}\sigma_{\tilde{w}}(\rho_{z\tilde{x}} - \rho_{\tilde{x}\tilde{w}}\rho_{z\tilde{w}})} - \frac{\sigma_\epsilon\rho_{z\tilde{w}}\rho_{\tilde{w}\epsilon}}{\sigma_{\tilde{x}}(\rho_{z\tilde{x}} - \rho_{\tilde{x}\tilde{w}}\rho_{z\tilde{w}})}.
\end{aligned}
$$

□