Measuring Measurement Error

This version: October 1, 2019

Abstract

Although 2SLS is most often applied to deal with omitted variables or simultaneity, we show that in practice it also resolves a substantial amount of attenuation bias resulting from classical errors in variables. We analyze 398 pairs of OLS and 2SLS regressions in 57 papers published in top economics and finance journals; we find that even when theory suggests the OLS coefficient should be inflated relative to the 2SLS coefficient, the 2SLS coefficient is in fact larger 85% of the time. Classical measurement error, i.e. differences between the underlying economic variables and what is measured by the econometrician, can explain this phenomenon. We derive simple regression specifications to quantify the extent of measurement error and estimate that approximately 23% of the variation in the average instrumented regressor is white noise.

Keywords: Measurement error, instrumental variables, endogeneity.
JEL Classifications: C26, C36.
1 Introduction

Instrumental variables techniques were initially developed to estimate systems of simultaneous equations (Wright, 1928; Stock and Trebbi, 2003) or to address errors in variables (Reiersøl, 1941; Geary, 1943; Aldrich, 1993). In a standard linear setting with omitted variables, reverse causality, and classical errors in variables, a valid instrumental variables procedure solves all three problems. Although the vast majority of recently published papers that use instrumental variables do so to address the first two problems, particularly omitted variable bias (Angrist and Pischke, 2009), these instruments may also be incidentally solving a severe errors in variables problem.

One source of measurement error is misreporting; in survey data, for example, respondents may not perfectly remember their average number of hours worked over the past year. Measurement error can also arise when perfectly-measured regressors are imperfect proxies for the theoretical constructs of an economic model (Bound, Brown and Mathiowetz, 2001). Viewed in this light, it should not be surprising that measurement error is pervasive in applied empirical work. An important question is: how badly does measurement error bias empirical estimates?

In this paper, we analyze pairs of IV and OLS coefficients to infer the average amount of bias induced by classical measurement error. We collect 398 IV-OLS coefficient pairs from 57 papers published in top economics and finance journals and show that IV estimates are almost always further from zero than their OLS counterparts, even when theory suggests the opposite should be true. While this fact has been documented by others, including Card (2001) and Jiang (2017), we show that the IV-OLS bias is systematically related to the size of the IV coefficient, consistent with classical measurement error.

We illustrate our point by using two papers drawn from our sample. Stroebel and Vavra (2019) regress changes in retail prices on changes in house prices at the MSA level, and estimate an elasticity of 6.8%; they note, however, that local unobservable supply shocks could move both prices together, so that the OLS estimate does not represent the causal impact of house price shocks on retail prices through a local demand channel. After instrumenting for house price changes with the standard Saiz (2010) measure of local housing supply elasticity, their estimate increases to 15.3%; as is often the case, the standard errors on both the IV and OLS estimate are large enough that one cannot reject that the two are equal to each other.

Guo, Pérez-Castrillo and Toldrà-Simats (2019) regress changes in firm R&D activity on the number of equity analysts covering the firm. One might expect that firms with better future prospects have both more analysts and more R&D activity, leading to a positive
OLS coefficient even in the absence of a causal effect. On the other hand, analysts may exert pressure on managers to increase short-term earnings at the expense of long-term investments such as R&D. The authors use a standard instrument for analyst coverage and find that the second effect dominates: one additional analyst raises the probability of an R&D cut by 6.3 percentage points, while the endogenous OLS coefficient is only 3.5 percentage points.

In both cases, the IV estimate is further from zero than OLS, consistent with the economic theory in one case and contrary to the economic theory in another case. An alternative hypothesis, that there is classical measurement error in the instrumented regressor, predicts the direction of the IV-OLS bias in both cases. But measurement error predicts more than that: it also predicts that the difference between the coefficients is proportional to the IV coefficient itself. Relating the change in the coefficient to the size of the IV suggests that the average measurement error ratio is \( E[\theta] \equiv E\left[ \frac{\sigma^2_{\tilde{x}}}{\sigma^2_{\tilde{x}}} \right] = \frac{(15.3-6.8)-(6.3-3.5)}{15.3-6.3} \approx .63 \) in this case. This is a “regression” on two data points; extending this to the 398 data points in our sample suggests \( E[\theta] \approx .228 \).

Classical measurement error provides a parsimonious explanation for the troubling fact that IV estimates are almost always larger than OLS estimates, irrespective of the economic theory motivating the need for an instrument. The top panel of Figure 1 quantifies this fact in our data, where the vertical bars indicate the percentage of regressions in our sample for which \( |\beta_{IV}| > |\beta_{OLS}| \). The x-axis separates regressions based on the expected direction of endogeneity bias arising from omitted variables or simultaneity; following Jiang (2017), we term endogeneity bias “corrective” when it is expected to attenuate \( \beta_{OLS} \), and “affirmative” in the reverse case. The blue bars indicate, for the subsample for which \( \beta_{IV} \) and \( \beta_{OLS} \) have the same sign, the percent of the sample with \( |\beta_{IV}| > |\beta_{OLS}| \), while the red bars indicate the percent of the entire sample for which the two coefficients have opposite signs.

We find that IV coefficients are larger in magnitude than their OLS counterparts the vast majority of the time, consistent with the results in Jiang (2017) and the impressions of most economists with whom we’ve spoken. Even in the cases where theory suggests the IV coefficient should shrink towards zero, in 85% of cases it gets larger in absolute value.

The bottom panel of Figure 1 reports the mean and median values of \( \beta_{IV}/\beta_{OLS} \) in each of these subsamples. As the first bar indicates, the average \( \beta_{IV} \) coefficient is almost six times its OLS counterpart in cases where it should be smaller. Although there is substantial skewness in this estimate, the median value of \( \beta_{IV}/\beta_{OLS} \) is still massive at 2.6 times. While we do not find in our sample, as Jiang (2017) does in hers, that the bias is larger for the affirmative endogeneity cases relative to the corrective endogeneity cases, it is still disturbingly large. Notice that for “corrective” endogeneity cases the average \( \beta_{IV}/\beta_{OLS} \) could indeed be very large, for example because \( \beta_{OLS} \to 0 \).
Figure 1. Comparing $\beta_{IV}$ and $\beta_{OLS}$

The blue bars in the top panel plot the percentage of OLS-IV regressions in the sample in which $|\beta_{IV}| > |\beta_{OLS}|$ when $\text{sign}(\beta_{IV}) = \text{sign}(\beta_{OLS})$, while the red bars plot the percentage of regressions in which $\text{sign}(\beta_{IV}) \neq \text{sign}(\beta_{OLS})$. The blue bars in the bottom panel plot the average value, across regressions, of $\beta_{IV}/\beta_{OLS}$, while the red bars plot the median. The bottom panel restricts the sample to cases where $\text{sign}(\beta_{IV}) = \text{sign}(\beta_{OLS})$. “Affirmative” refers to coefficients in which theory suggests $|\beta_{IV}| < |\beta_{OLS}|$, “corrective” is the reverse, and “unclear” refers to IV-OLS pairs in which the authors noted the direction of bias was ambiguous.
When discussing the relative magnitudes of OLS and IV coefficients, a common refrain from the papers in our sample is “although the IV estimates are larger than the OLS, we cannot reject that the two are equal.” We overcome the lack of power when making inferences about any single IV-OLS pair by collecting a large sample of regression coefficients, and other statistics, across many papers. Using these data, we are able to quantify a simple explanation for the observed IV-OLS bias: classical measurement error. Measurement error is by far the most parsimonious explanation for the direction of the IV-OLS bias; other explanations such as weak instruments, publication bias, or the difference between average and local-average treatment effects, require additional assumptions in order to predict that the IV estimate is larger.

2 Literature Review

The econometric technique that we now call instrumental variables (or two-stage least squares) has been used extensively to deal with omitted variable bias, simultaneity, and measurement error. The earliest use of instrumental variables techniques was most likely Wright (1928), who sought to identify parameters in a system of simultaneous equations. The earliest studies proposing to employ instrumental variables to deal with measurement error were Reiersøl (1941) and Geary (1943), though neither is wholly recognizable as such.\(^1\)

The surveys of Bound, Brown and Mathiowetz (2001) and Hausman (2001) note the usefulness of instrumental variables in combating errors in variables. Griliches and Hausman (1986) propose IV estimators for dealing with measurement error in panel data when the measurement error terms are iid; Biørn (2000) extends their estimator to deal with non-iid errors.

A small number of recent papers analyze popular econometric methods and practices with fundamental insights from econometric theory. Of these, the closest to our study is Jiang (2017), who also analyzes the incidence and magnitude of the IV-OLS bias and finds that IV coefficients tend to be larger than analogous OLS coefficients, even (in fact, especially) when theory suggests they should be smaller. Apart from the difference in sample—Jiang (2017) collects 255 papers from the top three finance journals from 2003 to 2014, while we collect 398 regressions from 57 papers published in the top 6 finance and economics journals from 2017 to the present—we focus on an alternative explanation to the three that Jiang (2017) considers: classical measurement error. Another study similar in spirit to ours is

\(^1\)See Aldrich (1993) for a history of econometric research in the 1940s as it relates to modern instrumental variables techniques and measurement. Stock and Trebbi (2003) provide an overview of the history and authorship of Appendix B of Wright (1928).
Berg and Streitz (2019), who convincingly show that spillover effects in standard difference-in-difference exercises can bias estimated effects even with random treatment. Oster (2019) shows that the standard practice of using coefficient stability to make inference about omitted variable bias also requires examining the change in $R^2$. Borusyak, Hull and Jaravel (2018) derive explicit conditions under which the popular “Bartik” instrument reveals causal effects.

Several papers have estimated the extent of measurement error in particular cases. Card (2001) carefully analyzes eleven studies that report both IV and OLS estimates of the effect of years of schooling on wages, and finds the IV estimate usually exceeds the OLS estimate, despite the fact that the IV was employed because the authors expected the opposite (e.g. “affirmative endogeneity” in the terminology of Jiang 2017). Applying our method to the 18 unique IV-OLS pairs in his Table II yields an average measurement error ratio of 0.46, right in line with our benchmark estimates.² Goolsbee (2000) estimates measurement error in cost of capital models and finds that about 20% of the tax term’s variance is noise. Bloom and Reenen (2007) and Bloom et al. (2019) use repeated surveys to gauge measurement error in their management practices measure and find that about 50% of their survey measure variance is noise. Bound and Krueger (1991) compare survey and administrative data on earnings and find that the measurement error in earnings is between 8–18% in levels and 19–35% in first differences.

There is a literature in finance devoted to correcting for measurement error in Tobin’s $Q$, an important explanatory variable for corporate investment. Rather than instrumental variables, much of this literature uses higher-order moments and cumulants to control for measurement error following a basic insight by Lewbel (1987); see Erickson and Whited (2000, 2002, 2012) and Erickson, Jiang and Whited (2014). Almeida, Campello and Galvao (2010) compare higher-order cumulant estimators to standard panel-IV estimators (Arellano and Bond, 1991; Børn, 2000) on real and simulated data and find that although the IV estimators require stronger assumptions, they tend to be more robust and efficient. Peters and Taylor (2017) use panel-IV estimates to check the robustness of their higher-order cumulant estimators and find similar results.

3 Model

In this section, we start by describing the impact of endogeneity on OLS estimates. We then propose a methodology to disentangle measurement error from omitted variable bias and

²Removing Ichino and Winter-Ebmer (2004), whose value of $\beta_{IV}$ is almost nine times the average and four times the maximum of the other points, reduces the estimated measurement error variance share to 0.32.
simultaneity. All proofs are in Appendix A.

Suppose we are interested in estimating the effect of \( x^* \) on \( y^* \) where

\[
x^* \equiv \alpha_y y^* + f(w, z^*, \varepsilon_x),
\]
\[
y^* \equiv \beta_x x^* + g(w, \varepsilon_y),
\]

such that \( \varepsilon_x \) and \( \varepsilon_y \) are independently-distributed structural disturbances and \( |\alpha_y \beta_x| < 1 \).

Suppose further that we only have data on \( x, y, \) and \( z \), where

\[ i \equiv i^* + u_i, \text{ for } i \in \{x, y, z\}, \]

such that \( u_i \) denotes mean-zero, independently-distributed measurement error.

An OLS regression of \( y \) on \( x \) would fail to recover \( \beta_x \) for three reasons. First, the unobservable \( w \) affects both \( x \) and \( y \), which leads to omitted variable bias. Second, the coefficient \( \alpha_y \) generates feedback effects, which implies simultaneity bias. Third, we observe \( x \) instead of \( x^* \), and this measurement error induces attenuation bias.

We can rectify all three of these issues in an instrumental-variables framework, provided \( z \) is a relevant instrument which satisfies the exclusion restriction; given the system (1), \( z \) is relevant if \( z^* \) is correlated with \( f(\cdot) \), and satisfies the exclusion restriction if \( z^* \) is uncorrelated with \( g(\cdot) \). Indeed, as the following proposition demonstrates, standard two-stage least squares yields a consistent estimate of \( \beta_x \), even if \( z \) contains measurement error.

**Proposition 1.** The OLS regression of \( y \) on \( x \) does not yield a consistent estimate of \( \beta_x \):

\[
\text{plim } \beta_{\text{OLS}} = \beta_x \left( \frac{\sigma^2_{z^*}}{\sigma^2_x} \right) + \frac{1}{1 - \alpha_y \beta_x} \left( \frac{\sigma_{f,g} + \alpha_y \sigma^2_g}{\sigma^2_x} \right)
\]

where \( \sigma_{ab} \) denotes the covariance of \( a \) and \( b \), and \( \sigma^2_a \) denotes the variance of \( a \), where \( a \) and \( b \) may be functions of other variables.

On the other hand, the 2SLS regression yields a consistent estimate of \( \beta_x \):

\[
\text{plim } \beta_{\text{IV}} = \beta_x.
\]

Equation (2) shows the consequences of running an OLS regression in the presence of an omitted variable (\( \sigma_{f,g} \neq 0 \)), simultaneity (\( \alpha_y \neq 0 \)), and measurement error (\( \sigma^2_{x^*} \neq \sigma^2_x \)).

Omitted variable bias and simultaneity bias have similar effects on the OLS estimate.

\[ \text{3The latter assumption is required in order for the system (1) to have a causal, structural interpretation; } |\alpha_y \beta_x| \geq 1 \text{ would imply that a shock to either } \varepsilon_x \text{ or } \varepsilon_y \text{ would cause both } x \text{ and } y \text{ to diverge to } \pm \infty. \]
Moreover, in some settings, we can use economic intuition to infer the direction of these biases. To borrow an example from Jiang (2017), suppose that $y$ is adulthood wages and $x$ is years of education. The usual omitted variable in this regression is ability, which is unobservable and has a positive effect on both wages ($\sigma_{gw} > 0$) and years of education ($\sigma_{fw} > 0$), hence $\beta_{\text{OLS}}$ would be biased upwards ($\sigma_{fg} > 0$). At the same time, we might expect that individuals who anticipate higher future earnings will also have higher lifetime wealth, and can therefore afford luxuries like education irrespective of its effect on wages; once again, this will bias $\beta_{\text{OLS}}$ upwards ($\alpha_y > 0$).

Despite this intuition, it is well-known that OLS estimates of returns-to-schooling coefficients tend to be biased downwards (Card, 2001). One possibility is that there is some hitherto unknown effect which overwhelms the aforementioned mechanisms and forces $\sigma_{fg} < 0$ or $\alpha_y < 0$. Another explanation is measurement error; as equation (2) shows, measurement error in education will bias $\beta_{\text{OLS}}$ towards zero ($\sigma_{z^2}^2 < \sigma_x^2$), even in the absence of omitted variables or simultaneity.

An interesting feature of Proposition 1 is that the traditional endogeneity terms are collected in the second term on the right-hand side of equation (2); if $\alpha_y = 0$, this term will not contain $\beta_x$. Exploiting this result leads to a simple and elegant method for estimating the average attenuation bias, which we demonstrate in the following proposition.

**Proposition 2.** Suppose $\alpha_y = 0$, and assume that the direct effect of $x^*$ on $y^*$, $\beta_x$, is generated independently of all other parameters.

Then given a dataset containing regression coefficients $\beta_{\text{IV}}$ and $\beta_{\text{OLS}}$ generated from (1), we can consistently estimate $E[\theta] \equiv E[\sigma_u^2 / \sigma_x^2]$ from the slope of the regression

$$\beta_{\text{IV}} - \beta_{\text{OLS}} = a_{\text{OLS}} + b_{\text{OLS}} \beta_{\text{IV}} + v.$$  
(3)

I.e.,

$$\text{plim } b_{\text{OLS}} = E[\theta].$$

In Proposition 2, we could equivalently regress $\beta_{\text{OLS}}$ on $\beta_{\text{IV}}$ and test if the estimated slope $E[1 - \theta]$ differs from unity; however, because most readers are used to testing if coefficients differ from zero, we instead use ($\beta_{\text{IV}} - \beta_{\text{OLS}}$) as the regressand.

In cases where $\alpha_y$ is potentially nonzero, we can instead estimate $E[\theta]$ using the following two-step approach.

**Proposition 3.** Suppose that the direct effect of $x^*$ on $y^*$, $\beta_x$, and the measurement error ratio $\theta$ are generated independently of all other parameters.
Then, given a dataset containing regression coefficients $\beta_{IV}$, $\beta_{OLS}$, and $\beta_{FS}$ generated from (1), where $\beta_{FS}$ is the first-stage regression coefficient of $x$ on the instrument $z$, we can recover $\mathbb{E}[\theta]$ from the following two-stage least-squares regression:

$$
\beta_{IV}\beta_{FS} = a_{FS} + b_{FS}\beta_{IV} + v_1 \quad \text{(first stage)},
$$

$$
(\beta_{IV} - \beta_{OLS}) \beta_{FS} = a_{IV} + b_{IV}\left(\hat{\beta}_{IV}\beta_{FS}\right) + v_2 \quad \text{(second stage)},
$$

where $\hat{\beta}_{IV}\beta_{FS}$ is the fitted $\beta_{IV}\beta_{FS}$ estimated from the first stage. I.e.,

$$
\text{plim } b_{IV} = \mathbb{E}[\theta].
$$

Intuitively, this procedure is designed to eliminate the $(1 - \alpha_y\beta_x)^{-1}$ term on the right-hand side of equation (2). Based on system (1), the reduced-from equation for $x^*$ is scaled by $(1 - \alpha_y\beta_x)^{-1}$, which implies the first-stage coefficient $\beta_{FS}$ is proportional to $(1 - \alpha_y\beta_x)^{-1}$. We can therefore multiply equation (2) through by $\beta_{FS}$, which will confine $\beta_x$ to the term associated with measurement error.

Propositions 1-3 immediately extend to multivariate settings, in which case we replace $x$ and $z$ with

$$
\tilde{k} \equiv k - \mathbb{E}[Q'Q]^{-1}\mathbb{E}[Q'k]Q \quad \text{for } k \in \{x, z\},
$$

where $Q$ is an $n_Q \times 1$ vector of controls. In this case our estimated $\theta$ represents the variance of the measurement error term divided by the variance of the residualized $\tilde{x}$, rather than $x$ itself.

4 Data

We collected every instrumental-variables paper published since 2017 from the following journals: *Econometrica*, the *American Economic Review*, the *Journal of Political Economy*, the *Journal of Finance*, the *Review of Financial Studies*, and the *Journal of Financial Economics*. We identified instrumental-variables papers by searching for the word “instrument” and kept any paper which reported 2SLS regressions with a single endogenous regressor and instrument.4 The vast majority of papers that we analyze run more than one 2SLS regression; we include in our dataset all IV-OLS regression pairs that appear in the main text.

Our sample includes papers which have multiple non-instrumented regressors, i.e., controls. In future work, we hope to include regressions with multiple instruments and instrumented regressors; however, these pose an additional data-collection challenge.
This resulted in a dataset with 398 coefficient pairs from 57 papers. Where available, we also recorded first-stage coefficients, $F$-statistics, $R^2$ values, and standard deviations of relevant variables.

Table 1 reports summary statistics for our sample. Panels A and B report the paper and observation counts by year and journal, respectively. Among the three economics journals, the *American Economic Review* dominates our sample both in terms of the number of papers and the number of regressions. *Econometrica* is less represented primarily because most of its instrumental-variables papers have multiple instruments for each endogenous regressor; by and large, these are over-identified quantitative models.

Panel C of Table 1 reports summary statistics associated with the IV-OLS regression pairs. The penultimate row covers the absolute value of the t-statistics associated with $\beta_{IV}$.\(^5\) The median $|t|$ is 2.5; 32% of the $\beta_{IV}$ estimates in our sample are statistically insignificant at the 5% level. The large fraction of insignificant results suggests that specification search or $p$-hacking may not be a major driver of our results; we also show that our results are robust to alternative treatments of observations with statistically insignificant $\beta_{IV}$.

Finally, the last row of Panel C of Table 1 reports statistics on the number of IV-OLS pairs per paper. The average paper reports 7 IV-OLS pairs; the median is 4. For papers which report multiple IV-OLS pairs, we further address potential $p$-hacking by looking at the subsample of regressions that are reported later in each paper. If later results are comparatively less important in determining a paper’s publication, then this sub-sample is less likely to be affected by $p$-hacking.

For each of the 57 papers in our sample, we document the explicit motivation given for running 2SLS. An overwhelming 88% of papers use instruments to overcome omitted variable bias, while only 23% discuss simultaneity or reverse causality; 30% of papers mention measurement error, though this issue almost always coincides with omitted variable or simultaneity bias. Indeed, only 3 papers exclusively list measurement error as the sole reason for using an instrument.

## 5 Results

In this section we report our main regression results. In section 5.1, we estimate equations (3) and (4) on our full dataset and various sub-samples for robustness. These results estimate the measurement error in the average regression. In section 5.2 we estimate equation (3) separately for individual $x$ variables that appear in papers regressed on distinct $y$-variables, yielding estimates of the measurement error at the individual regressor level.

\(^5\)One paper in our sample does not report standard errors or $t$-statistics on $\beta_{IV}$.\)
Panel A: Paper Count by Journal/Year

<table>
<thead>
<tr>
<th>Journal</th>
<th>2017</th>
<th>2018</th>
<th>2019</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Econometrica</em></td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>American Economic Review</em></td>
<td>6</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td><em>Journal of Political Economy</em></td>
<td>6</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td><em>Journal of Finance</em></td>
<td>5</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><em>Review of Financial Studies</em></td>
<td>3</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td><em>Journal of Financial Economics</em></td>
<td>7</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Panel B: Observation Count by Journal/Year

<table>
<thead>
<tr>
<th>Journal</th>
<th>2017</th>
<th>2018</th>
<th>2019</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Econometrica</em></td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>American Economic Review</em></td>
<td>63</td>
<td>51</td>
<td>51</td>
</tr>
<tr>
<td><em>Journal of Political Economy</em></td>
<td>19</td>
<td>2</td>
<td>31</td>
</tr>
<tr>
<td><em>Journal of Finance</em></td>
<td>59</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td><em>Review of Financial Studies</em></td>
<td>6</td>
<td>31</td>
<td>4</td>
</tr>
<tr>
<td><em>Journal of Financial Economics</em></td>
<td>38</td>
<td>15</td>
<td>14</td>
</tr>
</tbody>
</table>

Panel C: Variable Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Observations</th>
<th>Mean</th>
<th>Std Dev.</th>
<th>p10</th>
<th>p50</th>
<th>p90</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{OLS}$</td>
<td>398</td>
<td>7,175</td>
<td>42,931</td>
<td>-.12</td>
<td>.07</td>
<td>13</td>
</tr>
<tr>
<td>$\beta_{IV}$</td>
<td>398</td>
<td>10,603</td>
<td>53,643</td>
<td>-.8</td>
<td>.15</td>
<td>75</td>
</tr>
<tr>
<td>$\beta_{FS}$</td>
<td>218</td>
<td>.58</td>
<td>2.1</td>
<td>-.44</td>
<td>.58</td>
<td>1.2</td>
</tr>
<tr>
<td><em>F</em>-statistic</td>
<td>223</td>
<td>131</td>
<td>289</td>
<td>10</td>
<td>31</td>
<td>303</td>
</tr>
<tr>
<td>OLS $R^2$</td>
<td>174</td>
<td>.41</td>
<td>.3</td>
<td>.05</td>
<td>.32</td>
<td>.86</td>
</tr>
<tr>
<td>First-stage $R^2$</td>
<td>102</td>
<td>.45</td>
<td>.3</td>
<td>.12</td>
<td>.35</td>
<td>.86</td>
</tr>
<tr>
<td>$</td>
<td>\beta_{IV} t$-statistic$</td>
<td>395</td>
<td>4.3</td>
<td>7.9</td>
<td>.99</td>
<td>2.5</td>
</tr>
<tr>
<td># of regressions</td>
<td>57</td>
<td>7</td>
<td>7.3</td>
<td>1</td>
<td>4</td>
<td>17</td>
</tr>
</tbody>
</table>

*Table 1. Summary Statistics*
5.1 Average Results

Table 2 reports our main results. Panel A shows our estimates of the average measurement error ratio, $E[\theta]$, using the regressions prescribed by Propositions 2 and 3. Column 1 of Panel A reports the results of our benchmark specification, where we estimate equation (3) through ordinary least squares. Using all 398 IV-OLS pairs in our data, our estimate for $E[\theta]$ is .228.

As discussed in Section 3, if $\alpha_y \neq 0$, then we will need to adjust our estimation procedure to account for the simultaneity bias in the underlying regressors. Accordingly, in Column 2 of Table 2, we report the results of estimating equation (4) using 2SLS; our first stage is a regression of $\beta_{IV}\beta_{FS}$ on $\beta_{IV}$, while our second stage is a regression of $(\beta_{IV} - \beta_{OLS}) \beta_{FS}$ on the fitted values from the first stage.

While the IV estimate in Column 2 is higher than the OLS estimate in Column 1, this is not another manifestation of the fact that IV estimates tend to be larger than OLS estimates; rather, this is a manifestation of the fact that only a subset of papers report the $\beta_{FS}$ which corresponds to their $\beta_{IV}$. To address this mismatch, we estimate equation (3) using the same 218 IV-OLS pairs from Column 2; the result is reported in Column 3 of Panel A. The OLS estimate of this regression, .565, is larger than the 2SLS estimate, which suggests that equation (3) suffers from affirmative endogeneity in this sub-sample.

As Panel C of Table 1 shows, there is substantial variation in the magnitudes of both $\beta_{IV}$ and $\beta_{OLS}$; hence one concern with estimating equations (3) and (4) is that observations with very high values of $\beta_{IV}$ may be driving our results; a major source of variation in $\beta_{IV}$ comes from the fact that the dependent and independent variables from the 398 regressions in our sample have different units. The top panel of Figure 2 illustrates this issue in a scatterplot of the points used to estimate the coefficient in Column 1 of Table 2.

To address the vastly different units across the many regressions in our sample, we normalize each regression pair to be effectively unit-less by multiplying $\beta_{IV}$ and $\beta_{OLS}$ by $\sigma_x/\sigma_y$, essentially inferring the coefficients that would have been reported had the authors of the original papers regressed $y/\sigma_y$ on $x/\sigma_x$ instead of $y$ on $x$. Notice that, in the univariate case, $\beta_{OLS} \times \sigma_x/\sigma_y$ is just the correlation coefficient $x$ and $y$.

Unfortunately, authors do not routinely report standard deviations of their dependent and independent variables. Even in the subset of papers that report summary statistics, the regressions are often run using some transformation of the reported variable, making the summary statistics unusable for our purposes. Nevertheless, we do observe $\sigma_x$ and $\sigma_y$ for 51 regression pairs in 14 papers. The bottom panel of Figure 2 plots the normalized $\beta_{IV} \times \sigma_x/\sigma_y$ against $\beta_{OLS} \times \sigma_x/\sigma_y$.

Columns 4–6 of Table 2 report the same regressions as columns 1–3, but for the normalized
### Panel A: OLS and Second Stage

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_{IV} )</td>
<td>0.228</td>
<td>0.565</td>
<td>0.627</td>
<td>0.716</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3.5e-03)</td>
<td></td>
<td>(1.84)</td>
<td>(0.41)</td>
<td></td>
<td>(0.74)</td>
<td></td>
</tr>
<tr>
<td>( \beta_{IV}\beta_{FS} )</td>
<td>0.342</td>
<td></td>
<td></td>
<td></td>
<td>0.627</td>
<td></td>
</tr>
<tr>
<td>(2.0e-03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.016)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>398</td>
<td>218</td>
<td>218</td>
<td>51</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>Number of Papers</td>
<td>57</td>
<td>37</td>
<td>37</td>
<td>14</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.541</td>
<td>0.657</td>
<td>0.71</td>
<td>0.909</td>
<td>0.771</td>
<td>0.945</td>
</tr>
</tbody>
</table>

### Panel B: First Stage

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_{IV} )</td>
<td>4.29</td>
<td></td>
<td></td>
<td></td>
<td>0.675</td>
<td></td>
</tr>
<tr>
<td>(2.2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.614)</td>
<td></td>
</tr>
<tr>
<td>Kleibergen-Paap F-Statistic</td>
<td>3.8</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.593</td>
<td>0.106</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Panel C: OLS

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_{IV}\beta_{FS} )</td>
<td>0.344</td>
<td></td>
<td></td>
<td></td>
<td>0.584</td>
<td></td>
</tr>
<tr>
<td>(3.0e-04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(4.0e-03)</td>
<td></td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.657</td>
<td>0.775</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2. Meta-OLS and meta-IV Results**

The first column reports the estimated value of \( E[\theta] \) from estimating equation (3) in the full sample. The second column reports the 2SLS estimate of \( E[\theta] \) from estimating equation (4): in the first stage, regress \( \beta_{IV}\beta_{FS} \) on \( \beta_{IV} \); in the second stage, regress \( (\beta_{IV} - \beta_{OLS})\beta_{FS} \) on the fitted values from the first stage. The third column reports OLS results for the same subsample as the IV results in column 2. Panel A reports the second-stage estimate of \( E[\theta] \), Panel B reports the first stage, and Panel C reports the corresponding OLS estimate from regressing \( (\beta_{IV} - \beta_{OLS})\beta_{FS} \) on \( \beta_{IV}\beta_{FS} \). Columns 4–6 repeat all three exercises, but multiplying observations of \( \beta_{IV} \) and \( \beta_{OLS} \) by \( \sigma_x/\sigma_y \). Standard errors clustered at the paper level are reported in parentheses below the coefficients.
Figure 2. $\beta_{IV}$ and $\beta_{OLS}$ Scatterplots
Both panels plot $\beta_{IV}$ on the $x$-axis against $\beta_{IV} - \beta_{OLS}$ on the $y$-axis. The top panel plots all 398 observations of $\beta_{IV}$ and $\beta_{OLS}$, while the bottom panel plots the 51 normalized observations $\beta_{IV} \times \sigma_x/\sigma_y$ against $\beta_{OLS} \times \sigma_x/\sigma_y$. The thick solid lines are the fitted regression lines estimated from equation (3) and reported in Columns 1 and 4 of Panel A of Table 2, while the thin dotted lines are 45-degree lines.
Dependent Variable: $\beta_{IV} - \beta_{OLS}$

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Sample Restricted by $\beta_{IV}$ t-statistics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insignificant $\beta_{IV} \to 0$</td>
<td>$</td>
<td>t</td>
<td>&lt; 3.2$</td>
</tr>
<tr>
<td>$\beta_{IV}$</td>
<td>.221</td>
<td>.289</td>
<td>.213</td>
</tr>
<tr>
<td></td>
<td>(2.1e-03)</td>
<td>(3.0e-04)</td>
<td>(9.0e-03)</td>
</tr>
<tr>
<td>Observations</td>
<td>398</td>
<td>135</td>
<td>134</td>
</tr>
<tr>
<td>Number of Papers</td>
<td>57</td>
<td>43</td>
<td>35</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.491</td>
<td>.752</td>
<td>.502</td>
</tr>
</tbody>
</table>

| **Panel B: Sample Restricted by $|\beta_{IV}|$** |           |           |           |
| $|\beta_{IV}| < 50$ | $|\beta_{IV}| < 5$ | $|\beta_{IV}| < 1$ |
| $\beta_{IV}$   | .796      | .452      | .53       |
|                 | (.13)     | (.14)     | (.099)    |
| Observations    | 346       | 320       | 253       |
| Number of Papers| 55        | 52        | 45        |
| $R^2$           | .825      | .525      | .494      |

Table 3. Robustness

Both panels report robustness checks on the benchmark estimation equation (3). The first column of Panel A includes all observations, but sets $\beta_{IV}$ to 0 (on both the right- and left-hand-side) whenever its $t$-statistic is less than 1.96. Columns 2 and 3 restrict the sample to observations with $t$-statistics less than and greater than 3.2 in absolute value, respectively. Panel B restricts the sample according to the magnitude of $\beta_{IV}$, including only observations with $|\beta_{IV}|$ less than 50, 5, and 1, respectively, in columns 1, 2, and 3.

In the case where $\beta_{IV}$ and $\beta_{OLS}$ have been multiplied by $\sigma_x/\sigma_y$. Because normalization requires that we observe additional statistics, it substantially shrinks the sample, down to only 29 observations for the normalized meta-IV regression. Thus, we do not view these normalized estimates as more accurate than our un-weighted estimates; we present them to illustrate that our benchmark results are not driven by variation in units across papers. Indeed, this variation suggests our benchmark results may be too conservative.

Table 3 reports two sub-sample analyses as robustness checks. In Panel A, we restrict the sample by the absolute value of the $t$-statistic on $\beta_{IV}$ to ensure the $\beta_{IV}$ are precisely-estimated, while in Panel B we restrict the sample based on the magnitude of $\beta_{IV}$ to determine

---

6We do not multiply $\beta_{FS}$ by $\sigma_z/\sigma_x$, because $\beta_{FS}$ does not exhibit nearly as much variation as $\beta_{IV}$ and $\beta_{OLS}$, and this would reduce the sample for the normalized meta-IV regressions to only 10 observations.
the extent to which our results are driven by influential outliers.

For our first robustness check, Panel A of Table 3 presents results split by the absolute value of the t-statistic on $\beta_{IV}$.

Splitting by the standard error would be almost identical to splitting by the magnitude of $\beta_{IV}$, which we do in Panel B.

Ignoring the fact that $\frac{1}{1-\theta}$ is convex in $\theta$, an estimated $E[\theta]$ of about 0.80 can almost fully explain why $\beta_{IV}$ is six times larger than $\beta_{OLS}$ in the top panel of Figure 1. Accounting for this convexity, Jensen’s inequality implies that the average $\beta_{IV}/\beta_{OLS}$ generated by this estimate is even larger.

We are careful to exclude cases in which the same $y$ variable is regressed multiple times on the same $x$

5.2 Regressor-Level Results

In section 5.1, we estimated the average measurement error share $E[\theta]$ across all papers in our sample. The same methodology can be applied to estimate the measurement error share $\theta$ of a given regressor $x$, provided the corresponding OLS-IV pairs all use the same regressor. Indeed, for a fixed regressor $x$, we can estimate the measurement error share $\theta$ of $x$ even if the OLS-IV pairs have distinct regressands $\{y\}$ and instruments $\{z\}$.

We identify 12 papers, representing 155 IV-OLS regression pairs, for which the same $x$ variable is used in at least 5 distinct regressions, and estimate equation (3) separately for each $x$ variable. This results in 14 regressions, in which 12 produce statistically-significant
Table 4. Regressor-Level Results

The table reports summary statistics from 12 separate estimations of equation (3), where the sample is restricted to the same \( x \) variable but at least 5 separate \( y \) variables. The table omits regressions in which the estimates of \( \theta \) were statistically insignificant; the estimates in these cases were -.02 and 0.10, with \( t \)-statistics of -0.63 and 1.93, respectively.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Min</th>
<th>Mean</th>
<th>Median</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated ( \theta )</td>
<td>.338</td>
<td>.724</td>
<td>.763</td>
<td>.982</td>
</tr>
<tr>
<td>( t )-statistic</td>
<td>2.82</td>
<td>19.4</td>
<td>13.1</td>
<td>54.1</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>.621</td>
<td>.892</td>
<td>.952</td>
<td>.995</td>
</tr>
<tr>
<td># OLS-IV pairs</td>
<td>5</td>
<td>11.5</td>
<td>8</td>
<td>28</td>
</tr>
</tbody>
</table>

The first row of Table 4 shows that, among the 12 statistically-significant measurement error shares \( \theta \), the smallest \( \theta \) is greater than 0.3, while the mean and median are above 0.70, both considerably higher than the baseline estimates reported in Table 2. Measurement error also explains much of the variation in the OLS-IV bias: the \( R^2 \) varies from more than 0.60, to over 0.99. The average \( t \)-statistic is also quite high, at almost 20, though it bears repeating that we have excluded the two statistically insignificant estimates of \( \theta \); the insignificant \( \theta \) estimates are -.02 and 0.10, with \( t \)-statistics of -0.63 and 1.93, respectively.

The high precision and explanatory power of \( \beta_{IV} \) for the OLS-IV bias reported in Table 4 is surprising, given that the median regression has only 8 observations. To help visualize this relationship, Figure 3 plots the data for two of the 12 regressions reported in Table 4. The top panel plots \( \beta_{IV} \) on the \( x \)-axis against \( \beta_{IV} - \beta_{OLS} \) on the \( y \)-axis for 28 different regression pairs in Adelino, Ma and Robinson (2017); all 28 regression pairs (IV and OLS) use the same endogenous independent regressor, commuting-zone level income growth, and different dependent variables. Adelino, Ma and Robinson (2017) has the most regressions for a single \( x \)-variable in our sample; the bottom panel of Figure 3 plots \( \beta_{IV} \) against \( \beta_{IV} - \beta_{OLS} \) for a more typical example, 6 regression pairs reported in Becker and Pascali (2019).

In both cases we can reject the hypothesis that \( \theta = 0 \). For Adelino, Ma and Robinson (2017), our estimate suggests that over 60% of their measure of commuting-zone-level income growth is noise, although this estimate is fairly imprecise: the standard error is 0.15 and

---

\(^{10}\) The 28 left-hand-side variables include CZ-level net employment creation at the aggregate level, as well as split by 4 different age categories (Table III), net change in non-tradable employment in the aggregate and split by 2 age categories (Table IV), gross job creation, destruction, and net job creation split by aggregate and 4 age categories (Table V), and net change in construction employment split by aggregate and 4 age categories (Table VI).
Figure 3. Within-Regressor Scatterplots
Both panels plot $\beta_{IV}$ on the $x$-axis against $\beta_{IV} - \beta_{OLS}$ on the $y$-axis. The top panel plots 28 regressions reported in Adelino, Ma and Robinson (2017), where the $x$ variable is commuting-zone-level income growth (instrumented with a Bartik manufacturing shock), while the bottom panel plots six regressions from Becker and Pascali (2019), where the $x$ variable is the Protestant share of Prussian counties in 1882 (instrumented with the distance to Wittenburg). The thick solid lines are the estimated regression lines, while the thin dotted lines are 45-degree lines.
the $R^2$ is only 0.62. For Becker and Pascali (2019), who analyze the effects of Protestant population shares on anti-Semitic politics, the estimated $\theta$ is 0.53 with a standard error of 0.024. Moreover, as seen in the bottom panel of Figure 3, the $R^2$ is very close to 1. Becker and Pascali (2019) are not directly interested in the effects of Protestant population shares, but rather use them as a proxy for the extent to which the Protestant Reformation led Jews and Christians into more direct economic competition, for example, in money-lending. Our estimates suggest that about half of this proxy is noise.

6 Alternative Explanations

Jiang (2017) lists three potential reasons why $|\beta_{IV}| \gg |\beta_{OLS}|$, even when theory suggests the opposite. First, 2SLS returns the local average treatment effect (LATE), while OLS returns the average treatment effect (ATE); it is possible that, on average, the difference between “local” and “global” is much more important than the endogeneity problem itself. Second, weak instruments can exacerbate a slightly-invalid instrument and make $\beta_{IV}$ explode relative to $\beta_{OLS}$. Finally, because papers generally require statistically-significant results in order to be published, specification search and $p$-hacking can result in estimates of $\beta_{IV}$ further from zero than $\beta_{OLS}$.

We note that of her three explanations, only the third predicts that $|\beta_{IV}| > |\beta_{OLS}|$ without additional assumptions, and even that explanation requires (as is true in the data) the standard errors of $\beta_{IV}$ to be much larger than the standard errors of $\beta_{OLS}$. On the other hand, classical measurement error requires no additional assumptions to predict the direction of the bias, and is therefore a much simpler explanation of the facts. However, it seems unlikely that measurement error on its own is large enough to explain all of the average bias illustrated in Figure 1; in addition, even if classical measurement error is present, it is possible that these other mechanisms contaminate our estimated measurement error.

Although our goal is not to explain the magnitude of the bias, but to gauge the importance of measurement error, all three of Jiang (2017)’s explanations have the potential to contaminate our estimated measurement error. In this section we present additional results that engage with two of her three explanations. Because equation (1) assumes linearity in the structural model, within our framework we have nothing to say about the difference between LATE and ATE. We do note, however, that if LATE is so different from ATE in the average regression, then 2SLS is not in general of much use. The fact that so many papers continue to be published using 2SLS methods suggests that authors and editors are more concerned with endogeneity bias (including measurement error) than they are with the difference between LATE and ATE.
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \beta_{IV} )</td>
<td>( \beta_{IV} )</td>
<td>( \beta_{IV} )</td>
</tr>
<tr>
<td></td>
<td>.225</td>
<td>.217</td>
<td>.345</td>
</tr>
<tr>
<td></td>
<td>(6.1e-03)</td>
<td>(.012)</td>
<td>(9.0e-03)</td>
</tr>
<tr>
<td>Observations</td>
<td>223</td>
<td>111</td>
<td>112</td>
</tr>
<tr>
<td>Number of Papers</td>
<td>34</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>.528</td>
<td>.494</td>
<td>.632</td>
</tr>
</tbody>
</table>

Panel A: Sample Split by \( F \)-statistic

<table>
<thead>
<tr>
<th></th>
<th>All Obs with ( F )-stat</th>
<th>( F )-stat &lt; 30.54</th>
<th>( F )-stat ( \geq ) 30.54</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \beta_{IV} )</td>
<td>( \beta_{IV} )</td>
<td>( \beta_{IV} )</td>
</tr>
<tr>
<td></td>
<td>.506</td>
<td>.224</td>
<td>.259</td>
</tr>
<tr>
<td></td>
<td>(7.0e-04)</td>
<td>(8.7e-03)</td>
<td>(2.4e-03)</td>
</tr>
<tr>
<td>Observations</td>
<td>126</td>
<td>83</td>
<td>106</td>
</tr>
<tr>
<td>Number of Papers</td>
<td>32</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>.914</td>
<td>.661</td>
<td>.708</td>
</tr>
</tbody>
</table>

Table 5. Alternative Explanations

This table reports estimates from equation (3) on sub-samples to explore alternative explanations of our results. Panel A splits the sample by \( F \)-statistic: the first column reports results for the sub-sample that reports the \( F \)-statistic, while the second and third columns split the sample according to whether the reported \( F \)-statistic is above or below the median value of 30.54. The first column of Panel B restricts the sample to only statistically-insignificant (at the 5% level) results; the second and third panels restrict the sample to regressions that appear in the first 33% or last 33% of the paper, respectively, in papers with at least 6 regressions.
To explore whether our results are driven by weak instruments, in Panel A of Table 5 we split the sample by the value of the reported $F$-statistic. Jiang (2017) notes that when an instrument is slightly invalid, the IV coefficient may differ from the true $\beta_x$, which can generate $|\beta_{IV}| \gg |\beta_{OLS}|$ if the instrument is also weak. To address this possibility, column 1 of Panel A of Table 3 estimates equation (3) via OLS on the 223 observations which report a first-stage $F$-statistic. The estimate of $E[\theta]$ is almost identical to our benchmark estimate in column 1 of Table 2. Columns 2 and 3 then estimate equation (3) on observations with $F$-statistics below and above the median value of 30.54, respectively. The estimated $E[\theta]$ for the weaker instruments is almost unchanged; for the stronger instruments, our estimate is even larger, although the range of estimates in Panel A is similar to the range in Panel A of Table 2.

Weak instruments by themselves are not enough to explain why $|\beta_{IV}| \gg |\beta_{OLS}|$; they could just as easily predict the reverse. However, weak instruments become a stronger explanation when combined with the inherent selection that goes into the publication process, and the fact that IV standard errors tend to be much larger than OLS ones. Jiang (2017) illustrates a simple example in which, because only papers with $t$-statistics on their 2SLS regressions over 2.5 are published, $\beta_{IV}$ is often substantially higher than $\beta_{OLS}$ even though in the population of attempted and unpublished studies, the reverse would be true.

To determine the extent to which publication bias drives our results, we perform two additional tests. First, we restrict attention to estimates of $\beta_{IV}$ that are statistically insignificant at conventional levels. To our surprise, and counter to the publication-bias story, 32% of our sample observations of $\beta_{IV}$ are statistically insignificant at the 5% level. This may be because some published results are meant to be negative, or because some estimates are robustness checks that are not as relevant for publication. For example, 19 of the 26 $\beta_{IV}$ estimates reported by Dobbie, Goldin and Yang (2018) are statistically insignificant, possibly because they have so many different dependent regressors that they use with their instrument.

In column 1 of Panel B of Table 5 we restrict the sample to observations of $\beta_{IV}$ that are statistically insignificant at the 5% level. The estimated $\theta$ is again statistically significant, and comparable to some of the estimates in Tables 2 and 3. Of course, we present this estimate only to show that a high value of $\theta$ is not driven solely by publication bias or $p$-hacking.

A second way we control for the effect of publication bias on our results is by restricting the sample to regressions that appear later on in papers. As shown in Panel C of Table 1, the average paper in our sample has almost seven IV-OLS regression pairs, while the median paper runs 4. If authors tend to report their most important results earlier in the paper, and these results are more important for the publication decision than later ones, then we
can reduce the effect of publication bias on our results by focusing on later regressions.

In columns 2 and 3 of Table 5 we restrict the sample to those papers with at least 6 regressions, and include only IV-OLS pairs that occur either in the first or last 33% of the paper, respectively. We define the percentage position of a regression in a paper as the rank of that regression, divided by the total number of regressions in that paper, times 100%. Thus a paper with six regressions would have its first two results reported in Column 2 and its last two results reported in Column 3, while any regression in a paper with five or fewer reported regressions would be dropped.

If publication bias severely affected our results, then these two estimates of $E[\theta]$ should be very different; however, we find that the two estimates are almost identical, and very similar to our baseline results. If anything, $E[\theta]$ is somewhat higher for results reported later. Thus while publication bias may be an important factor in explaining the magnitude of $\beta_{IV}/\beta_{OLS}$, it does not unduly influence our estimate of the average measurement error.

7 Conclusion

We analyze 398 pairs of IV-OLS coefficient estimates across 57 papers published in top economics and finance journals to quantify the extent of classical measurement error in the average independent variable. While the vast majority of papers use instruments in order to address omitted variable bias or simultaneity bias, we show that these instruments also resolve a substantial amount of attenuation bias arising from errors in variables. In our benchmark estimation, we find that approximately 23% of the variation in the average regressor is noise.

Our results suggest that instrumental variable methods are useful even in cases where authors suppose that there is neither omitted variable bias nor reverse causality. Of necessity, we can only analyze the average measurement error of regressors for which an instrument has been found; if we assume that non-instrumentable regressors are neither systematically more nor less well-measured than instrumentable ones, our analysis suggests that the average OLS coefficient is at least 23% larger than what is reported. Individual cases could be much higher or lower, as the range of our within-regressor estimates makes clear.

Thus, our analysis points to the need to instrument all regressions, at least when the magnitude of the estimated coefficient, and not just its sign, is of interest. For cases where omitted variable or simultaneity bias is not first-order, it may be comparatively easier to find instruments that satisfy the exclusion restriction. Measurement error matters, and this noise demands attention.
References


A Proofs

Proof of Proposition 1

We require only that equation (1) hold in expectation:

\[
\begin{align*}
\mathbb{E}[x^*] & \equiv \alpha_y \mathbb{E}[y^*] + \mathbb{E}[f(w, z^*, \varepsilon_x)], \\
\mathbb{E}[y^*] & \equiv \beta_x \mathbb{E}[x^*] + \mathbb{E}[g(w, \varepsilon_y)].
\end{align*}
\]

The OLS regression yields

\[
\text{plim } \beta_{\text{OLS}} = \frac{\text{cov}(y, x)}{\text{var}(x)} = \frac{\text{cov}(\beta_x x^* + g, x^* + u_x)}{\text{var}(x)} = \frac{\beta_x \text{var}(x^*) + \text{cov}(g, x^*)}{\text{var}(x)} = \frac{\beta_x \text{var}(x^*) + (1 - \alpha_y \beta_x)^{-1}\text{cov}(g, f + \alpha_y g)}{\text{var}(x)} = \beta_x \frac{\text{var}(x^*)}{\text{var}(x)}
\]

\[
\equiv \sigma_x^{-2} \left[ \beta_x \sigma_x^2 + (1 - \alpha_y \beta_x)^{-1}(\sigma_{fg} + \alpha_y \sigma_g^2) \right].
\]

The 2SLS regression yields

\[
\text{plim } \beta_{\text{IV}} = \frac{\text{cov}(y, z)}{\text{cov}(x, z)} = \frac{\text{cov}(\beta_x x^* + g, z)}{\text{cov}(x, z)} = \frac{\beta_x \text{cov}(x^*, z)}{\text{cov}(x, z)} = \beta_x.
\]

Proof of Proposition 2

The assumption that \(\beta_x\) is generated independently of all other parameters that generate the data is sufficient but not necessary; in fact we only require that \(\beta_x\) satisfies the following
assumptions:

\[ \mathbb{E} \left[ \frac{\beta_x \sigma_{fg}}{\sigma_x^2} \right] = \mathbb{E} [\beta_x] \mathbb{E} \left[ \frac{\sigma_{fg}}{\sigma_x^2} \right], \]

\[ \mathbb{E} [\beta_x \theta] = \mathbb{E} [\beta_x] \mathbb{E} [\theta], \]

and \( \mathbb{E} [\beta_x^2 \theta] = \mathbb{E} [\beta_x^2] \mathbb{E} [\theta]. \)

The meta-OLS regression yields

\[
\text{plim } b_{\text{OLS}} = \frac{\text{cov} (\beta_{\text{IV}} - \beta_{\text{OLS}}, \beta_{\text{IV}})}{\text{var}(\beta_{\text{IV}})} = \frac{\text{cov} \left( \sigma_{x}^{-2} [\beta_x \sigma_{fg}^{-2} - \sigma_{fg}], \beta_x \right)}{\text{var}(\beta_x)} = \frac{\text{cov} (\beta_x \theta, \beta_x)}{\text{var}(\beta_x)} = \mathbb{E} [\theta].
\]

**Proof of Proposition 3**

The assumption that \( \beta_x \) and \( \theta \) are generated independently of all other parameters is sufficient but not necessary; in fact we require only that they satisfy

\[ \mathbb{E} \left[ \beta_x \frac{\sigma_{fg}}{\sigma_x^2} \left( \frac{\sigma_{fg} + \alpha_y \sigma_{fg}^2}{\sigma_x^2 + \alpha_y \sigma_{fg}^2 + 2 \alpha_y \sigma_{fg}} \right) (1 - \theta) \right] = \mathbb{E} [\beta_x] \mathbb{E} \left[ \frac{\sigma_{fg}}{\sigma_x^2} \left( \frac{\sigma_{fg} + \alpha_y \sigma_{fg}^2}{\sigma_x^2 + \alpha_y \sigma_{fg}^2 + 2 \alpha_y \sigma_{fg}} \right) (1 - \theta) \right], \]

\[ \mathbb{E} \left[ \beta_x (1 - \alpha_y \beta_x)^{-1} \frac{\sigma_{fg}}{\sigma_x^2} \theta \right] = \mathbb{E} [\beta_x (1 - \alpha_y \beta_x)^{-1} \frac{\sigma_{fg}}{\sigma_x^2} \theta] \mathbb{E} [\theta], \]

and \( \mathbb{E} \left[ \beta_x^2 (1 - \alpha_y \beta_x)^{-1} \frac{\sigma_{fg}}{\sigma_x^2} \theta \right] = \mathbb{E} [\beta_x^2 (1 - \alpha_y \beta_x)^{-1} \frac{\sigma_{fg}}{\sigma_x^2} \theta] \mathbb{E} [\theta]. \)

The meta-2SLS regression yields

\[
\text{plim } b_{\text{IV}} = \frac{\text{cov} (\beta_{\text{IV}} - \beta_{\text{OLS}}, \beta_{\text{FS}}, \beta_{\text{IV}})}{\text{cov}(\beta_{\text{IV}} \beta_{\text{FS}}, \beta_{\text{IV}})} = \frac{\text{cov} (\beta_x (1 - \alpha_y \beta_x)^{-1} \frac{\sigma_{fg}}{\sigma_x^2} \theta, \beta_x)}{\text{cov} (\beta_x (1 - \alpha_y \beta_x)^{-1} \frac{\sigma_{fg}}{\sigma_x^2}, \beta_x)} = \mathbb{E}[\theta].
\]