Measuring Measurement Error*

N. Aaron Pancost† and Garrett Schaller‡

This version: November 1, 2021

Abstract

Although proxy variables are pervasive in empirical work, the quality of proxy variables—in terms of how closely they track underlying economic forces—is not known. We derive novel regression specifications to infer the severity of measurement error using a sample of 2,552 instrumental variables regressions from 323 papers published in top economics and finance journals. We estimate that over 30% of the variation in the average regressor is white noise. For some proxies, our estimates exceed 95%. Our findings suggest that measurement error is a severe, pervasive, and understated source of bias in economics and finance.

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

*We thank Andrew Y. Chen (discussant), Jonathan Cohn, John Griffin, Michael Hasler (discussant), John Hatfield, Travis Johnson, Samuel Kruger, and Daniel Neuhann, and seminar participants at the University of Texas, the 2020 Midwest Finance Association Annual Meeting, Florida State University, the 2020 Lone Star Finance Conference, the University of Maryland (Smith), the Federal Reserve Bank of Boston, the Federal Reserve Bank of Chicago, and the Federal Reserve Board. Any remaining errors are our own. We are grateful to Tanner Bevill, Shobhan Bhatia, Quentin Brau, Caroline Carruthers, Jessica Castillo, Justin Chen, Samantha Chen, Nasouh Dadouch, Patricio Dieck, Mason Gross, Aarushi Khandelwal, Allison Hanzlik, Sisir Inagandla, Taylor Luce, Madison McBride, Divyank Mehta, Jorge Mendoza, Chi Pham, Ali Saffouri, and Changyong Song for outstanding research assistance. Jack Burchett, Nishal Desai, Andrew Kendrick, Lavanya Makdani, Saanya Pherwani, and Abigail Sawyer also provided helpful research assistance. We acknowledge financial support from the Salem Center for Policy at the University of Texas at Austin McCombs School of Business.

†University of Texas at Austin, McCombs School of Business. Email address: aaron.pancost@mccombs.utexas.edu

‡Colorado State University, College of Business. Email address: garrett.schaller@colostate.edu

1 Introduction

Proxy variables are ubiquitous in empirical research: Tobin's q proxies for investment opportunities, R&D spending and patent counts proxy for innovation, the Herfindahl index proxies for market power, Solow residuals proxy for productivity, years of education proxy for human capital investment, house prices proxy for household wealth, and realized returns proxy for expected returns, to give a few examples.

But just how accurate are these proxies? Erickson and Whited (2000, 2012) show that Tobin's q is wracked with measurement error: on average, around 50% of the variation in Tobin's q is pure noise. This much noise severely biases estimated coefficients and leads to spurious inference. We show that this problem goes beyond q: across all disciplines in economics and finance, we find that 39.0% of the variation in the average proxy variable is noise.

We measure measurement error across studies by exploiting the popularity of instrumental variables in published research. While most papers use instruments to address omitted variable bias (Angrist and Pischke, 2009), instrumental variable techniques can also purge the measurement error from an endogenous regressor. In fact, some of the earliest known work on instrumental variables was in the context of measurement error (Reiersøl, 1941; Geary, 1943; Aldrich, 1993). In a standard linear setting with omitted variables, simultaneity, and classical measurement error, a valid instrumental variables procedure solves *all three* problems. In this paper, we compare IV estimates to their OLS counterparts to gauge the extent of measurement error in the average regression.

Note that "measurement error" refers to the noisy measurement of the *economic* variable of interest. For example, many studies consider the effects of years spent in prison on recidivism. If years spent in prison is not the economic variable of interest *per se*, but rather proxies for exposure to criminality, then improving the measure of years spent in prison—e.g., by using administrative records as opposed to self-reported surveys—will not fix the measurement error problem; an instrument is needed. While many papers discuss the issues of omitted variable bias in recidivism, the same attention has not been paid to whether the measured variable is indeed a good proxy of the underlying economic variable.

To understand how we identify measurement error using pairs of OLS and IV estimates, consider the following regressions. Stroebel and Vavra (2019) regress changes in retail prices on changes in house prices, where changes in house prices proxy for local demand shocks, and estimate an elasticity of 6.80%; after instrumenting for house price changes, their estimate increases to 15.3%. Backus (2020) regresses the log of establishment-level productivity on the log number of establishments per square mile, where the number of establishments is

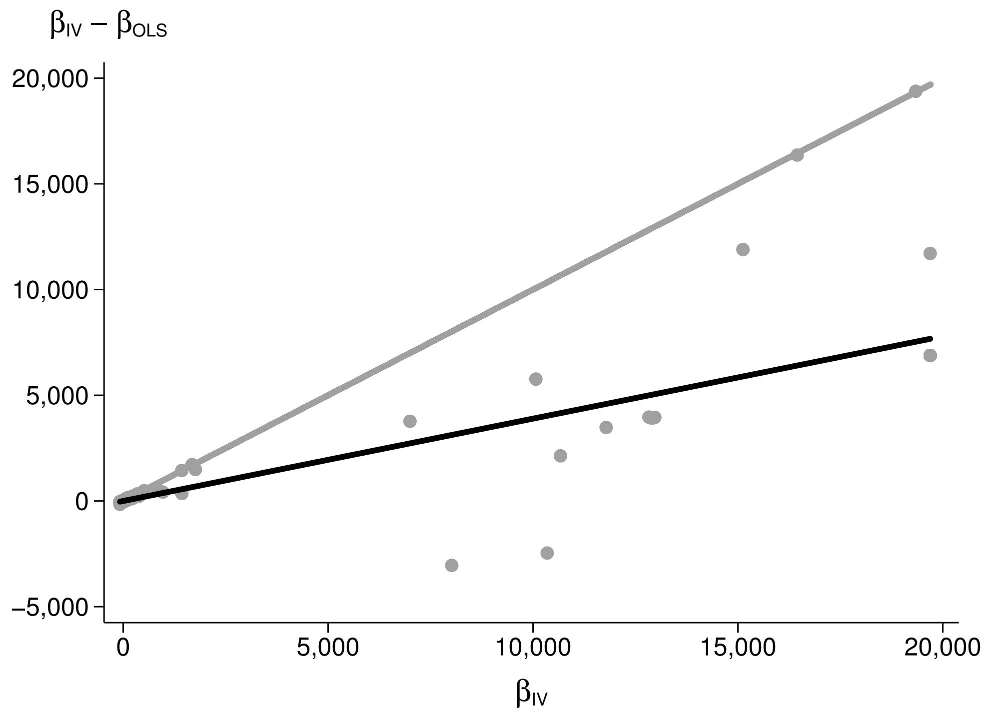

Figure 1. $\beta_{\text{IV}} - \beta_{\text{OLS}}$ vs. β_{IV}
The figure plots $\beta_{\text{IV}} - \beta_{\text{OLS}}$ against β_{IV} for our full dataset of 2,552 observations from 323 papers. The black line is the fitted regression line estimated from equation (3) and reported in column 1 of Table 2. The gray line is a 45-degree line.

a proxy for competitive pressure, and estimates an elasticity of 3.50%; after instrumenting for the number of establishments, his estimate increases to 4.68%. The IV exceeds the OLS estimate in both cases, but the difference between the two is larger for the larger IV coefficient. Because measurement error bias is proportional to the IV coefficient, we can use these two examples together to infer that, on average, $\frac{(15.3\%-6.80\%)-(4.68\%-3.50\%)}{15.3\%-4.68\%} \approx 68.9\%$ of the variance of these proxies is noise. This is effectively a regression on two data points; Figure 1 extends this logic to the 2,552 observations in our sample.

The measurement error we estimate is important for three reasons. First, our findings suggest that measurement error affects *all* OLS regressions. Even when authors are not concerned with omitted variable or simultaneity bias, the possibility of measurement error implies that their estimates are still biased. If valid instruments are not available to correct this bias, other techniques, such as higher-order moment estimators (Erickson and Whited, 2002; Erickson, Jiang and Whited, 2014), should be applied.

Second, our estimates suggest that researchers should treat measurement error with the same urgency as omitted variable bias. Thanks in part to the credibility revolution in

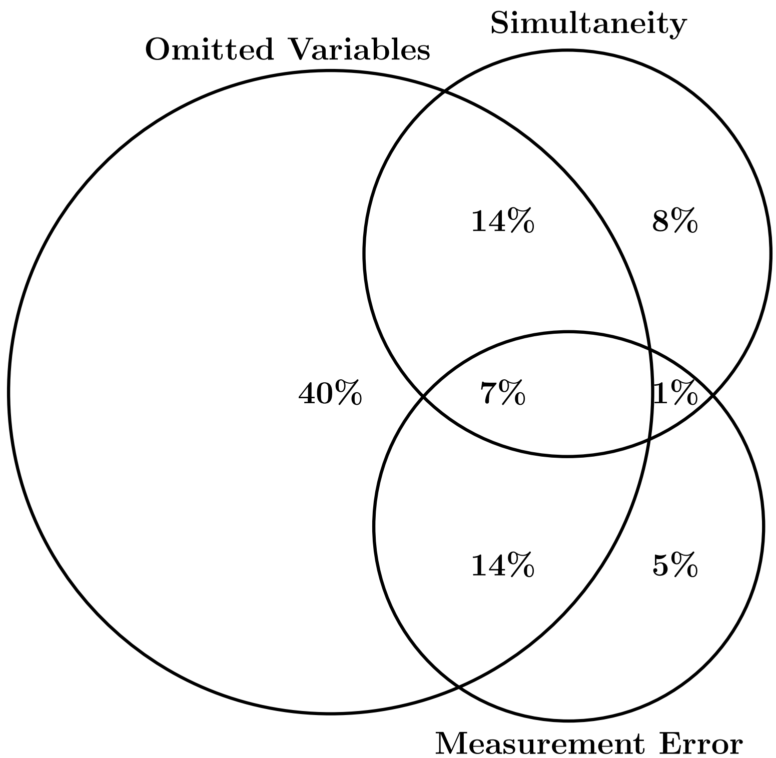

Figure 2. Motivation for Instrument
The figure shows the potential reasons for using instrumental variables. We report the percent of papers in our sample that appeal to each of these reasons. The area of each circle is proportional to the sum of the numbers inside it. The numbers do not add up to 100% because some papers do not report any motivation for using instrumental variables.

econometrics (Angrist and Pischke, 2010), modern empirical work is acutely aware of omitted variable bias; the same cannot be said for measurement error. Figure 2 shows the disparity in researchers' endogeneity concerns in our data. While 75% of papers in our data mention omitted variable bias as the reason they instrument their variable of interest, only 27% mention measurement error. In fact, 40% of papers mention *only* omitted variable bias, compared to 5% that mention *only* measurement error; indeed, Erickson, Jiang and Whited (2014) note a lack of attention to measurement error, despite the prevalence of proxy variables in the finance literature. Moreover, as we note in section 3, the presence of measurement error in addition to omitted variable bias makes it more likely that the observed OLS coefficient will have the opposite sign of the true effect. Quantitative empirical work must contend with the fact that proxies are poorly measured.

Third, classical measurement error explains why IV estimates are generally larger than OLS estimates, even when we expect the opposite to be true. Card (2001) has observed this pattern in the returns to education literature, and Jiang (2017) in the finance literature; as we show in Figure 3, this pattern emerges across disciplines in economics and finance. We separate regressions based on the expected direction of the OLS bias; following Jiang (2017), we say this bias is "corrective" when we expect the OLS coefficient to exceed the IV coefficient, and "affirmative" in the reverse case. Even when theory suggests the IV coefficient should shrink towards zero, 81% of the time it does the opposite; moreover, on average, it is more than six times larger than the OLS coefficient.

One explanation for the evidence in Figure 3 is publication bias, by which we mean selective publication of statistically significant results. Jiang (2017) shows in a simple example how such p-hacking can generate a large IV-OLS ratio. Figure 4 suggests that p-hacking is not an important driver of our results: restricting attention to the set of statistically insignificant results does not alter the conclusions from Figure 3.

In addition, we estimate a selection model that incorporates both measurement error and publication bias. We find that publication bias has virtually no effect on our estimate of measurement error, and measurement error can account for 99.1% of the median IV-OLS ratio; we target the median IV-OLS ratio since the mean can diverge if, for example, the IV and OLS coefficients are normally distributed.

Using our data, we quantify a simple explanation for the observed OLS bias: classical measurement error. Measurement error not only explains the direction of the OLS bias, it also predicts that the OLS bias will be positively correlated with the magnitude of the IV coefficient, a prediction with strong support in the data. Other explanations for the preponderance of IV estimates in excess of their OLS counterparts—such as weak instruments, publication bias, or differences between average and local-average treatment effects—do not speak to this additional evidence.

When researchers estimate an IV coefficient in excess of their OLS coefficient, they should use the meta-OLS estimator we propose in this paper to determine whether measurement error is driving their results. As we demonstrate with Mian and Sufi (2014), when the OLS coefficients are unexpectedly small, it may be that the measurement error bias has overwhelmed the expected omitted variable bias. The meta-OLS estimator can reveal the source of this disparity.

2 Literature Review

The econometric technique that we now call two-stage least squares has been used extensively to deal with omitted variables, simultaneity, and measurement error. The earliest use of instrumental variables was most likely Wright (1928), who sought to identify parameters in

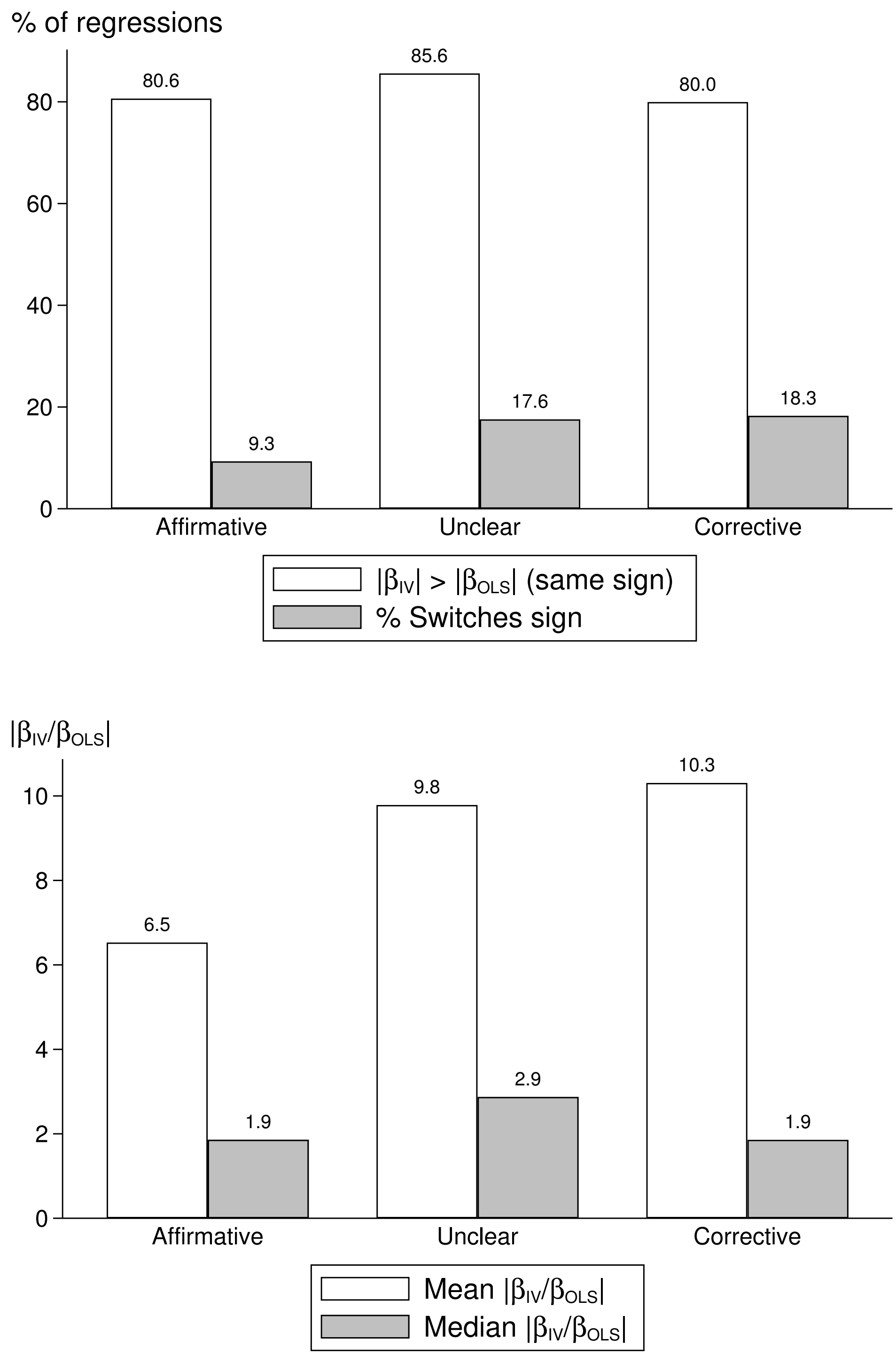

Figure 3. Comparing β_{IV} and β_{OLS}
The white bars in the top panel plot the percentage of regressions in which $|\beta_{\mathrm{IV}}| > |\beta_{\mathrm{OLS}}|$ when $\operatorname{sign}(\beta_{\mathrm{IV}}) = \operatorname{sign}(\beta_{\mathrm{OLS}})$, while the gray bars plot the percentage of regressions in which $\operatorname{sign}(\beta_{\mathrm{IV}}) \neq \operatorname{sign}(\beta_{\mathrm{OLS}})$. The white bars in the bottom panel plot the average value, across regressions, of the IV-OLS ratio $\left|\beta_{\mathrm{IV}}/\beta_{\mathrm{OLS}}\right|$, while the gray bars plot the median. "Affirmative" refers to coefficients for which theory predicts $|\beta_{\mathrm{IV}}| < |\beta_{\mathrm{OLS}}|$, "corrective" the reverse, and "unclear" refers to IV-OLS pairs in which the authors noted the direction of bias was ambiguous or declined to sign the bias.

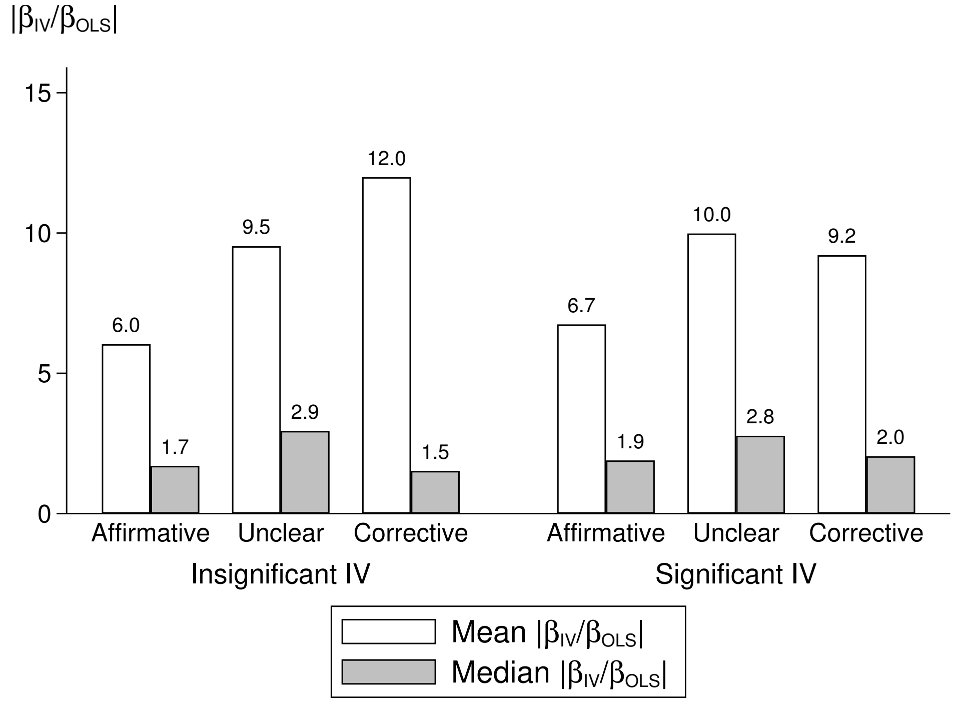

Figure 4. Comparing β_{IV} and β_{OLS}, by $\left|t\left(\beta_{\mathrm{IV}}\right)\right|$
The figure replicates the results reported in the bottom panel of Figure 3, but split by the statistical significance of β_{IV}. The white bars plot the average value, across regressions, of the IV-OLS ratio $\left|\beta_{\mathrm{IV}}/\beta_{\mathrm{OLS}}\right|$, while the gray bars plot the median. "Affirmative" refers to coefficients for which theory predicts $|\beta_{\mathrm{IV}}| < |\beta_{\mathrm{OLS}}|$, "corrective" the reverse, and "unclear" refers to IV-OLS pairs in which the authors noted the direction of bias was ambiguous or declined to sign the bias.

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 measurement error. Griliches and Hausman (1986) propose IV estimators for dealing with measurement error in panel data when the measurement error terms are independent and identically distributed; Biørn (2000) extends their estimator to deal with non-i.i.d. errors. DiTraglia and García-Jimeno (2020) derive cross-equation restrictions in the presence of omitted variables, measurement error, and instrument invalidity; they also demonstrate how to formally incorporate prior beliefs into

[1]See Aldrich (1993) for a history of econometric research in the 1940s as it relates to modern instrumental variables and measurement error. Stock and Trebbi (2003) provide an overview of the history and authorship of Appendix B in Wright (1928).

their analysis.

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 OLS bias and finds that IV coefficients tend to be larger than their OLS counterparts, particularly when theory suggests they should be smaller. Jiang (2017) considers papers published in the top three finance journals, while we consider papers published in the top three finance journals and the top five economics journals; her sample covers 2003 to 2014, while ours covers 2013 to 2019. We also focus on an alternative explanation for the OLS bias than the three she considers. Lal et al. (2021) replicate over 60 papers from top political science journals and also find that IV coefficients are often much larger than OLS coefficients. Another study similar in spirit to ours is Berg and Streitz (2019), who show that spillover effects in common 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 Bartik instruments reveal causal effects.

Other papers have estimated the extent of measurement error in particular cases. Card (2001) 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 instruments were employed because the authors expected the IV estimate to be smaller than the OLS estimate. Applying our method to the 21 unique IV-OLS pairs in his Table II yields an average measurement error ratio of 49%, which is 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 Van 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's variance is noise. Bound and Krueger (1991) compare survey and administrative data on earnings and find that the share of measurement error in earnings is between 8–18% in levels and 19–35% in first differences.

Several papers use higher-order moment estimators in lieu of instrumental variables to correct for measurement error; see Erickson and Whited (2000, 2002, 2012), Whited (2001), and Erickson, Jiang and Whited (2014). These techniques can be adapted to nonlinear settings (Schennach and Hu, 2013) or environments with conditional heteroskedasticity (Hahn, Hausman and Kim, 2020). See Schennach (2016) for a comprehensive review of techniques

[2] Removing two observations from Ichino and Winter-Ebmer (2004), whose values of β_{IV} are between five and nine times the average of the other estimates, reduces the share of measurement error to 39%.

for dealing with measurement error.

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 simultaneity. All proofs are in Appendix A.

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

$$\begin{aligned} x^* &\equiv \alpha_y y^* + f(w, z^*, \varepsilon_x), \\ y^* &\equiv \beta_x x^* + g(w, \varepsilon_y), \end{aligned} \tag{1}$$

such that $(w, z^*, \varepsilon_x, \varepsilon_y)$ are independently-distributed and $|\alpha_y \beta_x| < 1$.[3] Suppose further that we only have data on x, y, and z, where

$$\begin{aligned} x &\equiv x^* + u_x, \\ y &\equiv y^* + u_y, \\ z &\equiv z^* + u_z, \end{aligned}$$

such that u_x, u_y, and u_z denote mean-zero, independently-distributed measurement errors with variances $\sigma^2_{u_x}$, $\sigma^2_{u_y}$, and $\sigma^2_{u_z}$, respectively. Throughout this paper, we are interested in discerning what fraction of the variation in x is due to noise,

$$\theta \equiv \frac{\sigma^2_{u_x}}{\sigma^2_x} = 1 - \frac{\sigma^2_{x^*}}{\sigma^2_x},$$

which we refer to as the measurement error ratio.

An OLS regression of y on x would fail to recover β_x for three reasons. First, the unobservable w affects both x and y, which leads to omitted variable bias. Second, the coefficient α_y propogates feedback effects, which generates 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. More concretely, given the system of equations (1), the instrument z is relevant if z^* is correlated with $f(\cdot)$, and satisfies the exclusion restriction if z^* is uncorrelated with $g(\cdot)$.

[3] The latter assumption is required in order for the system of equations (1) to have a causal, structural interpretation; if $|\alpha_y \beta_x| \geq 1$, then shocks to either ε_x or ε_y would cause both x and y to diverge to $\pm\infty$.

Indeed, as the following proposition demonstrates, a standard two-stage least squares regression yields a consistent estimate of β_x, even if z contains measurement error.

Proposition 1. *The OLS regression of y on x does not yield a consistent estimate of β_x:*

$$\operatorname{plim} \beta_{OLS} = \beta_x \left(\frac{\sigma^2_{x^*}}{\sigma^2_x} \right) + \frac{1}{1 - \alpha_y \beta_x} \left(\frac{\sigma_{fg} + \alpha_y \sigma^2_g}{\sigma^2_x} \right) \tag{2}$$

where σ_{ab} denotes the covariance of a and b, and σ^2_a denotes the variance of a, where a and b may be functions of other variables.

The IV regression, by contrast, yields a consistent estimate of β_x:

$$\operatorname{plim} \beta_{IV} = \beta_x.$$

Equation (2) shows the consequences of running an OLS regression in the presence of an omitted variable ($\sigma_{fg} \neq 0$), simultaneity ($\alpha_y \neq 0$), and measurement error ($\sigma^2_{x^*} \neq \sigma^2_x$). In some settings, we can use economic intuition to infer the direction of the OLS bias ($\beta_{\text{IV}} - \beta_{\text{OLS}}$). 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 β_{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 β_{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 β_{OLS} towards zero ($\sigma^2_{x^*} < \sigma^2_x$), even in the absence of omitted variables or simultaneity.

It is also worth emphasizing that a *valid* instrument z is one that is uncorrelated with the measurement error u_x; indeed, if z and u_x were correlated, that could create a non-classical measurement error problem in which $\operatorname{cov}(x^*, u_x) \neq 0$, and the bias is not necessarily attenuative. When analyzing the validity of an instrument, it is therefore important to acknowledge when an instrument is solving an endogeneity problem but not a measurement-error problem (Roberts and Whited, 2013). For example, many papers use house price elasticities to instrument for changes in house prices. To the extent that house prices are meant to proxy for household wealth, this instrument is valid insofar as house price elasticities are uncorrelated with changes in non-housing wealth.

Note that the OLS coefficient (2) is a mixture of the true effect and omitted variable bias, where the true effect is diluted by measurement error bias. This reveals a pernicious interaction between these biases: as measurement error bias increases, the observed OLS coefficient will be less reflective of the true effect and more reflective of the omitted variable bias. This is particularly concerning when the true effect and omitted variable bias have opposite signs: if the measurement error is sufficiently severe, then the OLS coefficient will take on the sign of the omitted variable bias instead of the true effect.

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); moreover, if $\alpha_y = 0$, the right-hand term in equation (2) will not contain β_x. Exploiting this result leads to a simple and elegant method for estimating the average measurement error ratio, which we demonstrate in the following proposition.

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

Then, given a dataset containing pairs of OLS and IV regression coefficients $(\beta_{OLS}, \beta_{IV})$, we can consistently estimate the average measurement error ratio $\mathbb{E}[\theta] \equiv \mathbb{E}\left[\sigma^2_{u_x}/\sigma^2_x\right]$ from the slope of the regression

$$\beta_{IV} - \beta_{OLS} = a_{OLS} + b_{OLS}\beta_{IV} + v. \tag{3}$$

I.e.,

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

In Proposition 2, we could equivalently regress β_{OLS} on β_{IV} and test if the estimated slope $\mathbb{E}[1-\theta]$ 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 α_y is potentially nonzero, we can instead estimate $\mathbb{E}[\theta]$ using the following two-step approach.

Proposition 3. *Suppose that the direct effect of x^* on y^*, β_x, and the measurement error ratio, θ, are generated independently of all other parameters.*

Then, given a dataset containing triplets of regression coefficients $(\beta_{OLS}, \beta_{IV}, \beta_{FS})$, where β_{FS} is the first-stage coefficient found by regressing x on the instrument z, we can recover the *average measurement error ratio* $\mathbb{E}[\theta]$ *from the following two-stage least squares regression:*

$$\begin{aligned} \beta_{IV}\left|\beta_{FS}\right| &= a_{FS} + b_{FS}\beta_{IV} + v_1 && \textit{(first stage)}, \\ \left(\beta_{IV} - \beta_{OLS}\right)\left|\beta_{FS}\right| &= a_{IV} + b_{IV}\left(\widehat{\beta_{IV}\left|\beta_{FS}\right|}\right) + v_2 && \textit{(second stage)}, \end{aligned} \tag{4}$$

where $\widehat{\beta_{IV}\left|\beta_{FS}\right|}$ *is the fitted* $\beta_{IV}\left|\beta_{FS}\right|$ *estimated from the first stage. I.e.,*

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

Intuitively, this procedure is designed to eliminate the $\left(1-\alpha_y\beta_x\right)^{-1}$ term on the right-hand side of equation (2). Based on the system of equations (1), the reduced-from equation for x^* is scaled by $\left(1-\alpha_y\beta_x\right)^{-1}$, which implies the first-stage coefficient β_{FS} is proportional to $\left(1-\alpha_y\beta_x\right)^{-1}$. We can therefore multiply equation (2) through by $|\beta_{\text{FS}}|$, which will confine β_x to the term associated with measurement error. We use $|\beta_{\text{FS}}|$ rather than β_{FS}, since the latter can vary in sign across regressions—weakening the first stage of equation (4)—but the $\left(1-\alpha_y\beta_x\right)^{-1}$ term is always positive (see footnote 3).

In other words, Proposition 3 uses the first-stage coefficient β_{FS} to excise terms associated with simultaneity bias. Some of the regressions in our data are over-identified, so that there are multiple first-stage coefficients to choose from. The following corollary establishes that any linear combination of these coefficients will yield a consistent estimate of $\mathbb{E}[\theta]$.

Corollary. *Suppose that, in addition to* $z_1 \equiv z$*, some observations use* $n_z - 1$ *other instruments,* $z_2, z_3, ..., z_{n_z}$*, such that each* z_i *satisfies the same assumptions as* z*. In this case, we can implement the meta-IV procedure in equation (4) using any linear combination of first-stage regression coefficients, provided this combination is independent of* β_x *and* θ*.*

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

$$\widetilde{k} \equiv k - \mathbb{E}\left[c'c\right]^{-1}\mathbb{E}\left[c'k\right]c \text{ for } k \in \{x, z\}, \tag{5}$$

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

4 Data

We collect every instrumental-variables paper published from 2011 to 2019 from the following journals: the *American Economic Review*, *Econometrica*, the *Journal of Political Economy*, the *Quarterly Journal of Economics*, the *Review of Economic Studies*, the *Journal of Finance*, the *Journal of Financial Economics*, and the *Review of Financial Studies*. We identify instrumental-variables papers by searching for the word "instrument," and keep any paper which reports at least one pair of IV and OLS regressions with a single endogenous regressor.[4] The vast majority of papers that we analyze run more than one IV regression; we include in our dataset all IV-OLS regression pairs that appear in the main text.[5] This leaves us with a dataset of 2,552 coefficient pairs from 323 papers. Where available, we also recorded the first-stage coefficients, F-statistics, 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 five 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 it rarely reports the endogenous OLS regressions. The *Quarterly Journal of Economics* is a relatively small part of the sample because it publishes fewer articles. In 2019, the *Review of Economic Studies* released more issues than usual, which partly explains why it constitutes a larger portion of our sample for that year.

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 β_{IV}. The median $|t_{IV}|$ is 2.3; in fact, 37% of the β_{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; indeed, in a later section, we show that our results are robust to alternative treatments of observations with statistically insignificant β_{IV}.

Note that the 2,145 reported OLS coefficients in panel C refers only to distinct OLS regressions; it is not uncommon for the same OLS regression to apply to multiple IV regressions with, for example, different instruments. Likewise, the 1,309 reported first-stage coefficients and 935 reported F-stats in panel C refer only to distinct observations of these

[4]Our sample includes papers which have multiple non-instrumented regressors, i.e., control variables. In future work, we hope to include regressions with multiple instrumented regressors, though this poses additional challenges for data collection.

[5]This includes 11 papers published in the *American Economic Review* for which we can recover unreported OLS coefficients using replication files from the journal's website.

A. Paper Count by Journal/Year									
Journal	2011	2012	2013	2014	2015	2016	2017	2018	2019
American Economic Review	10	12	13	13	3	9	7	8	12
Econometrica	1	0	0	4	2	0	2	0	1
Journal of Political Economy	3	4	4	2	2	2	4	1	5
Quarterly Journal of Economics	4	3	6	2	5	2	1	4	4
Review of Economic Studies	2	5	1	5	4	2	1	3	12
Journal of Finance	2	0	5	2	3	5	5	2	2
Journal of Financial Economics	7	6	8	2	7	9	8	5	9
Review of Financial Studies	7	6	5	8	3	6	4	8	6

B. Observation Count by Journal/Year									
Journal	2011	2012	2013	2014	2015	2016	2017	2018	2019
American Economic Review	134	224	203	205	11	59	64	145	121
Econometrica	4	0	0	26	6	0	9	0	3
Journal of Political Economy	10	27	17	9	11	26	24	2	36
Quarterly Journal of Economics	20	21	9	22	57	6	2	25	16
Review of Economic Studies	10	29	44	117	8	20	2	4	74
Journal of Finance	4	0	23	12	28	23	59	13	7
Journal of Financial Economics	20	26	39	7	43	36	45	20	35
Review of Financial Studies	21	15	59	36	4	35	10	35	35

C. Variable Statistics						
Variable	Observations	Mean	Std Dev.	p10	p50	p90
β_{OLS}	2,145	195	1,508	-0.52	0.03	1.3
β_{IV}	2,552	268	2,171	-1.2	0.09	2.2
β_{FS}	1,309	0.53	4.5	-0.7	0.05	1.1
F-statistic	935	4,276	74,349	4.5	19	702
$\lvert t_{IV}\rvert$	2,511	28	1,116	0.59	2.3	6.1
Regressions per paper	323	7.9	14	1	4	17

Table 1. Summary Statistics
The table reports summary statistics for our sample of publish instrumental variables regressions. Panels A and B report the number of papers and regressions in our sample by journal and year, respectively. The first five rows of panel C report summary statistics at the regression level; some β_{OLS} and β_{FS} are paired with more than one β_{IV}, and some papers do not report first-stage results. The last row of panel C is at the paper level.

values. In our sample, 39% of IV regressions do not report an F-statistic, and 49% do not report first-stage coefficients. Of course, these percentages apply only to papers that report both OLS and IV results; we suspect that papers that fail to report OLS results are even less likely to report F-statistics or first-stage results.

As panel C of Table 1 shows, there is substantial variation in the magnitudes of both β_{IV} and β_{OLS}, and one concern with estimating equation (3) on the data directly is that observations with very high values of β_{IV} may be driving our results. To reduce the potential impact of outliers on our results, we follow Jiang (2017) and winsorize the coefficient values in our data at 1%.

Finally, the last row of panel C of Table 1 reports statistics on the number of IV-OLS pairs per paper. Each paper reports an average of 7.9 IV-OLS pairs. 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.

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 are estimates of the amount of measurement error in the average regressor. In section 5.2, we estimate equation (3) separately for each x variable that appears in a paper with multiple y variables, which yields regressor-level estimates of the measurement error for each of these x variables.

5.1 Average Results

Table 2 reports our main results. Panel A shows our estimates of the average measurement error ratio, $\mathbb{E}[\theta]$, using the regression prescribed by Proposition 2. Column 1 of panel A reports the results of our benchmark specification, where we estimate equation (3) through ordinary least squares using all 2,552 IV-OLS pairs in our data. The resulting estimate of $\mathbb{E}[\theta]$ is 0.390. In the second column of Table 2, we perform our estimation using only one IV-OLS pair per paper, similar to Jiang (2017). In particular, we use the main result for each paper, i.e., the regression that is emphasized in the abstract and introduction to the paper. The resulting estimate of $\mathbb{E}[\theta]$ is similar to the result in column 1.

A major source of variation in β_{IV}, even after winsorization, comes from the fact that the

	(1)	(2)	(3)	(4)
Dependent Variable: $\beta_{\mathrm{IV}} - \beta_{\mathrm{OLS}}$				
β_{IV}	0.390	0.336	0.666	0.535
	(0.042)	(0.014)	(0.044)	(0.067)
R^2	0.818	0.996	0.834	0.746
Observations	2,552	323	1,015	111
Number of papers	323	323	116	111
Normalized	No	No	Yes	Yes

Table 2. Meta-OLS Results
The reports estimates from the meta-OLS regression (3). Columns 1 and 3 include all regressions in each paper, while Columns 2 and 4 include only one regression per paper. Columns 3–4 report results where observations of β_{IV} and β_{OLS} are multiplied by σ_x/σ_y. Standard errors, clustered at the paper level, are reported in parentheses below the coefficients.

dependent and independent variables from the 2,552 regressions in our sample have different units. Figure 1 illustrates this in a scatterplot of the data used in the first column of Table 2. The dispersion of β_{IV} is massive, almost entirely due to a few papers which choose units for their x and y variables that happen to result in large coefficients; for example, Feyrer, Mansur and Sacerdote (2017) regress county-level changes in wages per capita on the total value of extracted oil and natural gas in millions of dollars. The estimated coefficients are on the order of tens of thousands and can be easily seen in Figure 1.

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

Perhaps surprisingly, authors do not routinely report the standard deviations of their dependent and independent variables. Even in the subset of papers that report summary statistics, the regressions are often run using a transformation of the reported variable, making the summary statistics unusable for our purposes. Nevertheless, we do observe σ_x and σ_y for 1,015 regression pairs in 116 papers. Figure 5 plots the normalized $\beta_{\mathrm{IV}} \times \sigma_x/\sigma_y$ against $(\beta_{\mathrm{IV}} - \beta_{\mathrm{OLS}}) \times \sigma_x/\sigma_y$, and column 3 of Table 2 reports the meta-OLS estimate of equation (3) for the normalized sample. Because normalization shrinks our sample, we do not view normalized estimates as being more accurate than non-normalized estimates; we present them to illustrate that our benchmark results are not driven by variation in the size of coefficients across papers. In fact, normalizing the data increases our benchmark

	(1)	(2)	(3)
Dependent Variable: $\beta_{\mathrm{IV}} - \beta_{\mathrm{OLS}}$			
β_{IV}	0.658	0.781	0.625
	(0.045)	(0.088)	(0.073)
R^2	0.837	0.832	0.751
Observations	829	185	27
Number of papers	86	29	4
Normalized	Yes	Yes	Yes
Measurement Error	No	Yes	Only

Table 3. Meta-OLS Results by Motivation

The table reports estimates from the meta-OLS regression (3). The first column includes only papers where the authors do not mention measurement error as a reason they are instrument for x. The second column includes papers where authors do mention measurement error, potentially in addition to other reasons. The third column includes papers in which authors mention only measurement error as the reason for instrumenting. All columns report results where observations of β_{IV} and β_{OLS} are multiplied by σ_x/σ_y. Standard errors, clustered at the paper level, are reported in parentheses below the coefficients.

estimate of $\mathbb{E}[\theta]$, from 0.390 to 0.666. The final column of Table 2 reports the results of the normalized regression with a single observation per paper; there are fewer papers than in Column 3 because in some cases the main regression of a paper cannot be normalized. The estimated $\mathbb{E}\left[\theta\right]$ is still substantially larger than in Columns 1 and 2.

Papers that mention measurement error as an explicit motivation for instrumenting have similar levels of measurement as those that do not, suggesting that measurement error is pervasive even though many authors seem to be unaware of it. Table 3 reports the results of estimating equation (3) on three sub-samples of the data: papers that do not mention measurement error as the reason they instrument for x (column 1), papers that do (column 2), and papers that *only* mention measurement error (column 3). The final category, as can be seen from Figure 2, is quite small. Nevertheless all three subsamples display similar levels of measurement error.

Table 4 reports the meta-IV estimates of $\mathbb{E}\left[\theta\right]$. As discussed in Section 3, if $\alpha_y \neq 0$, then we need to adjust our estimation procedure to account for simultaneity bias in the underlying regressors; Proposition 3 describes how the first-stage regression coefficients can be used to address this issue in a two-stage least squares procedure. The top part of panel B reports the second-stage estimates from regressing $(\beta_{\mathrm{IV}} - \beta_{\mathrm{OLS}})\left|\beta_{\mathrm{FS}}\right|$ on the instrumented $\widehat{\beta_{\mathrm{IV}}\left|\beta_{\mathrm{FS}}\right|}$, the middle part reports the first-stage coefficients from regressing $\beta_{\mathrm{IV}}\left|\beta_{\mathrm{FS}}\right|$ on β_{IV}, and, for

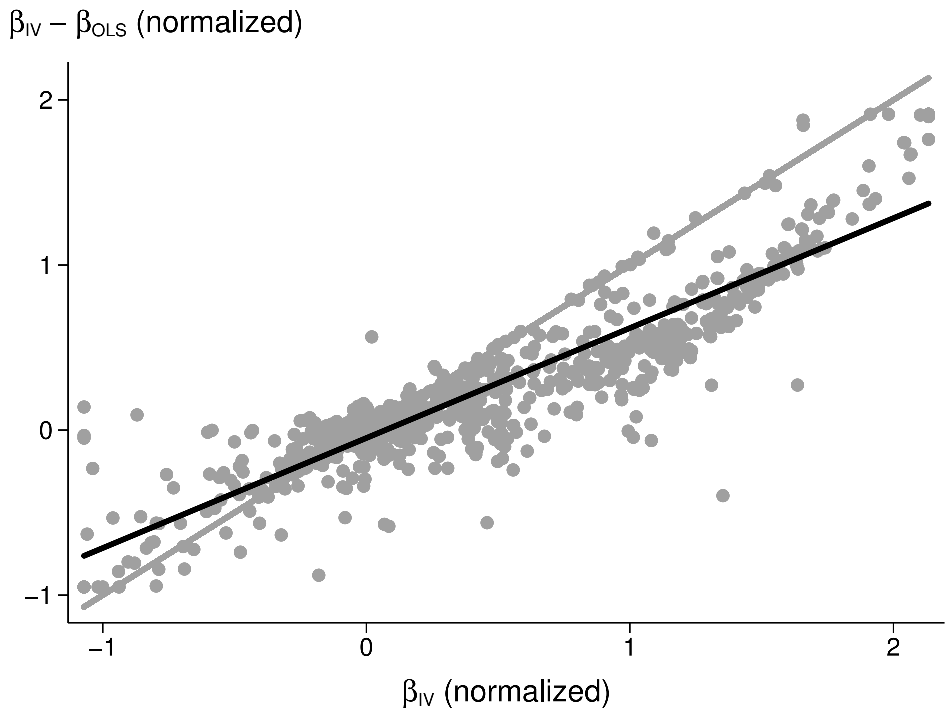

Figure 5. $(\beta_{\text{IV}} - \beta_{\text{OLS}})\frac{\sigma_x}{\sigma_y}$ vs. $\beta_{\text{IV}}\frac{\sigma_x}{\sigma_y}$
The figure plots $(\beta_{\text{IV}} - \beta_{\text{OLS}}) \times \sigma_x/\sigma_y$ against $\beta_{\text{IV}} \times \sigma_x/\sigma_y$ for 1,015 observations from 116 papers. The black line is the fitted regression line estimated from equation (3) and reported in column 2 of Table 2. The gray line is a 45-degree line.

the sake of completeness, the bottom part reports the OLS counterpart to the second stage.

The first three columns of Table 4 report the results of estimating equation (4). For most of the papers in our sample, there is only a single instrument, and thus a single first-stage coefficient to report; however, 129 observations from 40 papers report a vector of first-stage coefficients from an over-identified model. As the corollary to Proposition 3 shows, if there are multiple first-stage coefficients, then any linear combination of these coefficients can be used to obtain a consistent estimate of $\mathbb{E}[\theta]$. We therefore estimate equation (4) using three different combinations of the first-stage coefficients: a precision-weighted average of all reported first-stage coefficients (column 1); only the first reported first-stage coefficient (column 2); or just the just-identified observations (column 3). The estimates of $\mathbb{E}[\theta]$ are similar in all three cases.

The last three columns of Table 4 reports the results of estimating equation (4) on normalized coefficients; that is, where we multiply β_{IV} and β_{OLS} by σ_x/σ_y, and $|\beta_{\text{FS}}|$ by σ_z/σ_x. The sample is considerably smaller than in the first three columns of Table 4 since few papers report summary statistics of x, y, and z. Nevertheless, we still have over 50

	(1)	(2)	(3)	(4)	(5)	(6)
	Second-Stage Dependent Variable: $(\beta_{\text{IV}} - \beta_{\text{OLS}})\,|\beta_{\text{FS}}|$					
$\widehat{\beta_{\text{IV}}\beta_{\text{FS}}}$	0.352	0.352	0.352	0.548	0.547	0.548
	(0.000)	(0.000)	(0.000)	(0.055)	(0.055)	(0.058)
R^2	0.911	0.911	0.91	0.391	0.392	0.377
	First-Stage Dependent Variable: $\beta_{\text{IV}}\,|\beta_{\text{FS}}|$					
β_{IV}	2.363	2.363	2.363	0.410	0.410	0.410
	(0.001)	(0.001)	(0.000)	(0.068)	(0.068)	(0.071)
R^2	0.955	0.955	0.955	0.231	0.23	0.221
F-Stat	4814	4796	3919	36.5	36.1	33.7
	OLS Dependent Variable: $(\beta_{\text{IV}} - \beta_{\text{OLS}})\,|\beta_{\text{FS}}|$					
$\beta_{\text{IV}}\beta_{\text{FS}}$	0.346	0.346	0.346	0.341	0.341	0.338
	(0.000)	(0.000)	(0.000)	(0.097)	(0.097)	(0.099)
R^2	0.911	0.911	0.911	0.618	0.618	0.613
Observations	1,508	1,511	1,307	677	677	617
Number of papers	194	196	166	58	58	50
Normalized	No	No	No	Yes	Yes	Yes
Multiple z	Weighted	First	None	Weighted	First	None

Table 4. Meta-IV Results

The table reports estimates from the meta-IV regression (4). "Multiple z" refers to the way in which we use first stage coefficients from over-identified models: "Weighted" means that β_{FS} is a precision-weighted average of the reported coefficients, "First" means that only the first reported first-stage coefficient is used, and "None" means these observations are dropped. Columns 4–6 report results where observations of β_{IV} and β_{OLS} are multiplied by σ_x/σ_y, and observations of $|\beta_{\text{FS}}|$ are multiplied by σ_z/σ_x. Standard errors, clustered at the paper level, are reported in parentheses below the coefficients.

Dependent Variable: $\beta_{\text{IV}} - \beta_{\text{OLS}}$			
	(1)	(2)	(3)
A. Sample Restricted by $\lvert t_{IV}\rvert$			
	Insignificant $\beta_{\text{IV}} \to 0$	$\lvert t_{IV}\rvert \in (1.96, 3.1)$	$\lvert t_{IV}\rvert \geq 3.1$
β_{IV}	0.393	0.589	0.373
	(0.046)	(0.236)	(0.021)
R^2	0.622	0.686	0.905
Observations	2,552	788	786
Number of papers	323	210	189

B. Sample Restricted by $\lvert\beta_{\text{IV}}\rvert$			
	$\lvert\beta_{\text{IV}}\rvert < 50$	$\lvert\beta_{\text{IV}}\rvert < 5$	$\lvert\beta_{\text{IV}}\rvert < 1$
β_{IV}	0.547	0.615	0.505
	(0.075)	(0.135)	(0.036)
R^2	0.587	0.216	0.375
Observations	2,447	2,270	1,813
Number of papers	320	303	265

Table 5. Robustness
Both panels report robustness checks on the benchmark estimation equation (3). The first column of panel A includes all observations, but sets β_{IV} to 0 on both the left- and right-hand-side whenever its t-statistic is less than 1.96. Columns 2 and 3 restrict the sample to statistically significant β_{IV} observations with t-statistics less than or greater than the median t-statistic of 3.1. Panel B restricts the sample according to the magnitude of β_{IV}, including only observations with $|\beta_{\text{IV}}|$ less than 50, 5, and 1 in columns 1, 2, and 3, respectively.

papers in the sample, and the results are similar to Table 2 and the first three columns of Table 4: measurement error is substantial, and the estimate is even larger when we normalize coefficients.

Table 5 reports two sub-sample analyses as robustness checks. Panel A of Table 5 presents results split by the absolute value of the t-statistic on β_{IV}.[6] The first column of panel A shows that our results are not driven by imprecisely-estimated β_{IV}: the estimated $\mathbb{E}[\theta]$ is almost unchanged if we set $\beta_{\text{IV}} = 0$ for the 37% of the observations in which β_{IV} is statistically insignificant at the 5% level. Columns 2 and 3 instead split the sample based on the median

[6]Splitting the sample by the standard error of β_{IV} would be almost identical to splitting by the magnitude of β_{IV}, which we do in panel B.

$|t_{IV}|$ and re-estimate equation (3); while the estimate in column 2 exceeds the estimate in column 3, both are well within the range of estimates reported in of Table 2.

Another possibility is that our estimate of $\mathbb{E}[\theta]$ is driven by observations far in the tails of the distribution of β_{IV}; panel C of Table 1 shows that the standard deviation of β_{IV} in our sample is 2,171, while only 10% of β_{IV} observations are greater than 2.2. It may be that the estimated θ of 0.390 is entirely driven by a small number of papers where β_{IV} is very large.

Panel B of Table 5 shows that this is not an issue: restricting attention to sub-samples based on $|\beta_{\mathrm{IV}}|$ leads to estimates that are comparable to our benchmark estimates from Table 2. Exclusively using the 2,270 observations with $|\beta_{\mathrm{IV}}| < 5$ leads to an estimated measurement error of over 75%; further restricting the sample leads to estimates of $\mathbb{E}[\theta]$ that are still within the range of values in Table 2.

5.2 Regressor-Level Results

In section 5.1, we estimated the average measurement error ratio $\mathbb{E}[\theta]$ across all regressors in our sample. The same methodology can be applied to estimate the measurement error ratio θ of a *single* regressor x, provided we observe multiple IV-OLS pairs with the same regressor x but different dependent variables y.

Equation (2) shows how we can identify θ. Intuitively, omitted variable and simultaneity bias are features of x and y jointly, whereas measurement error depends only on x. Fixing the choice of x while varying the choice of y will therefore generate variation in β_{OLS} and β_{IV} while leaving θ unchanged, thereby permitting the use of meta-regressions to identify θ for a single x. Incidentally, the intercept for each of these meta-regressions will be the average omitted variable or simultaneity bias across the different y variables.

We identify 72 separate x variables from 49 papers, representing 677 IV-OLS regression pairs, for which 6 or more distinct y variables are regressed on x. We are careful to exclude cases in which the same y variable is regressed multiple times on the same x variable, for example, with different sets of controls or instruments. We estimate equation (3) separately for each of these x variables, yielding 48 positive estimates of θ that are statistically significant at the 1% level; Table 6 reports summary statistics for these regressions.

The first row in panel A of Table 6 shows that the mean and median θ, at 0.540 and 0.537, are within the range of the baseline estimates reported in Table 2. The median θ and t-statistic are both very high despite the fact that these 72 regressors include the negative and statistically insignificant estimates of θ.

Panels B and C split the 72 estimates into the 48 that are positive and statistically significant at the 1% level, and the 24 that are not. The θ estimates reported in panel B are

Statistic	Min	p10	Mean	Median	p90	Max
A. All 72 Regressors						
Estimated θ	-1.58	-0.005	0.540	0.537	0.970	3.56
t-statistic	-3.13	-0.010	55.5	7.55	44.9	1300
R^2	0.000	0.050	0.667	0.905	0.998	1.00
# IV-OLS pairs	6	6	9	7	17	28
B. All 48 Significantly-Positive Regressors						
Estimated θ	0.188	0.438	0.711	0.749	0.981	1.11
t-statistic	2.82	5.70	83.1	11.0	88.0	1300
R^2	0.285	0.821	0.918	0.944	0.999	1.00
# IV-OLS pairs	6	6	9	8	13	28
C. All 24 Remaining Regressors						
Estimated θ	-1.58	-0.247	0.198	0.183	0.462	3.56
t-statistic	-3.13	-1.24	0.378	0.759	1.95	2.11
R^2	0.000	0.002	0.164	0.126	0.395	0.526
# IV-OLS pairs	6	6	10	6	18	26

Table 6. Regressor-Level Results
The table reports summary statistics from 72 separate estimations of equation (3), where the sample is restricted to the same x variable but at least 6 separate y variables. The top panel includes statistics from all regressors; panel B restricts the sample to the 48 regressors that are statistically significantly positive at the 1% level, while panel C restricts the sample to the remaining 24 regressors. Each row reports a distribution across all the indicated regressors; for example, the minimum t-statistic does not necessarily correspond to the minimum estimated θ or R^2.

somewhat better-behaved than their imprecisely-estimated counterparts in panel C:[7] they range from 0.188 to 1.11 and explain quite a large variation in the OLS bias, with R^2 values ranging from 0.285 to 1.00.

Figure 6 shows how θ can be estimated precisely with so few observations; these panels plot the relationship between β_{IV} and the bias $\beta_{\text{IV}} - \beta_{\text{OLS}}$ for 2 of the 72 regressions reported Table 6. The top panel plots 8 IV-OLS regression pairs from Mian and Sufi (2014), where all 8 pairs use the same x variable: a function of county-level stock and bond holdings in 2006, county-level total debt owed in 2006, and county-level median house prices, population, and homeownership rates in 2000—projected to 2006 using a house-price index and changes in

[7] There are two regressions in which θ is significantly negative, with t-statistics of -3.13 and -2.39. When running 72 separate regressions, the probability of getting a few significantly-negative t-statistics when the true $\theta = 0$ is reasonably high.

population and homeownership rates from 2000 to 2006—to proxy for changes in county-level housing net worth from 2006 to 2009. The bottom panel of Figure 6 plots the 6 regression pairs reported in Becker and Pascali (2019), where they use the share of Protestants in Prussian counties to proxy for anti-Semitism.

In both cases, the potential for measurement error is undeniable; using equation (3), we can strongly reject the hypothesis that these proxies are free of noise. But we can do more than that: we can precisely quantify the extent of this measurement error. For Mian and Sufi (2014), our estimate suggests that 57.1% of the variation in their measure of housing net worth is noise, with a standard error of 6.2%. For Becker and Pascali (2019), we estimate that 52.6% of the variation in their measure of anti-Semitism is noise, with a standard error of 2.4%.

The bias in Mian and Sufi (2014) also serves to illustrate how omitted variables interact with measurement error and our estimator. Mian and Sufi are interested in the effects of housing net worth on employment growth, but note that OLS estimators will likely be biased upwards due to omitted variables:

> [Growth in housing net worth] may be spuriously correlated with supply-side industry-specific shocks that impact both employment and housing net worth. In particular, certain industries may be harder hit during the recession, and counties with greater exposure to these industries may naturally experience both a larger decline in housing net worth and larger fall in employment.

In terms of equation (1), the authors are arguing that $\sigma_{fg} > 0$, which implies $\beta_{\text{IV}} < \beta_{\text{OLS}}$ in the absence of measurement error. Yet the data defy this prediction: over 60% of the time, the IV estimates exceed their OLS counterparts. Measurement error can explain the discrepancy. As the intercept in the top panel of Figure 6 shows, the constant term in our meta-OLS regression is indeed negative, as expected when $\sigma_{fg} > 0$, and statistically significant with $t = -2.62$. In other words, omitted variables do indeed bias their OLS estimates upwards, but this effect is small relative to the downward bias induced by measurement error in their proxy variable.

Another way to understand our identification of θ is to treat equation (2) as an estimating equation. In this case, absent measurement error, a regression of β_{OLS} on β_{IV} would produce a slope coefficient of 1. Consider the six IV-OLS pairs from Becker and Pascali (2019):

Becker and Pascali (2019) IV-OLS Pairs						
β_{IV}	0.0282	0.0994	0.131	0.208	0.453	0.598
β_{OLS}	0.0383	0.0638	0.109	0.0802	0.233	0.311

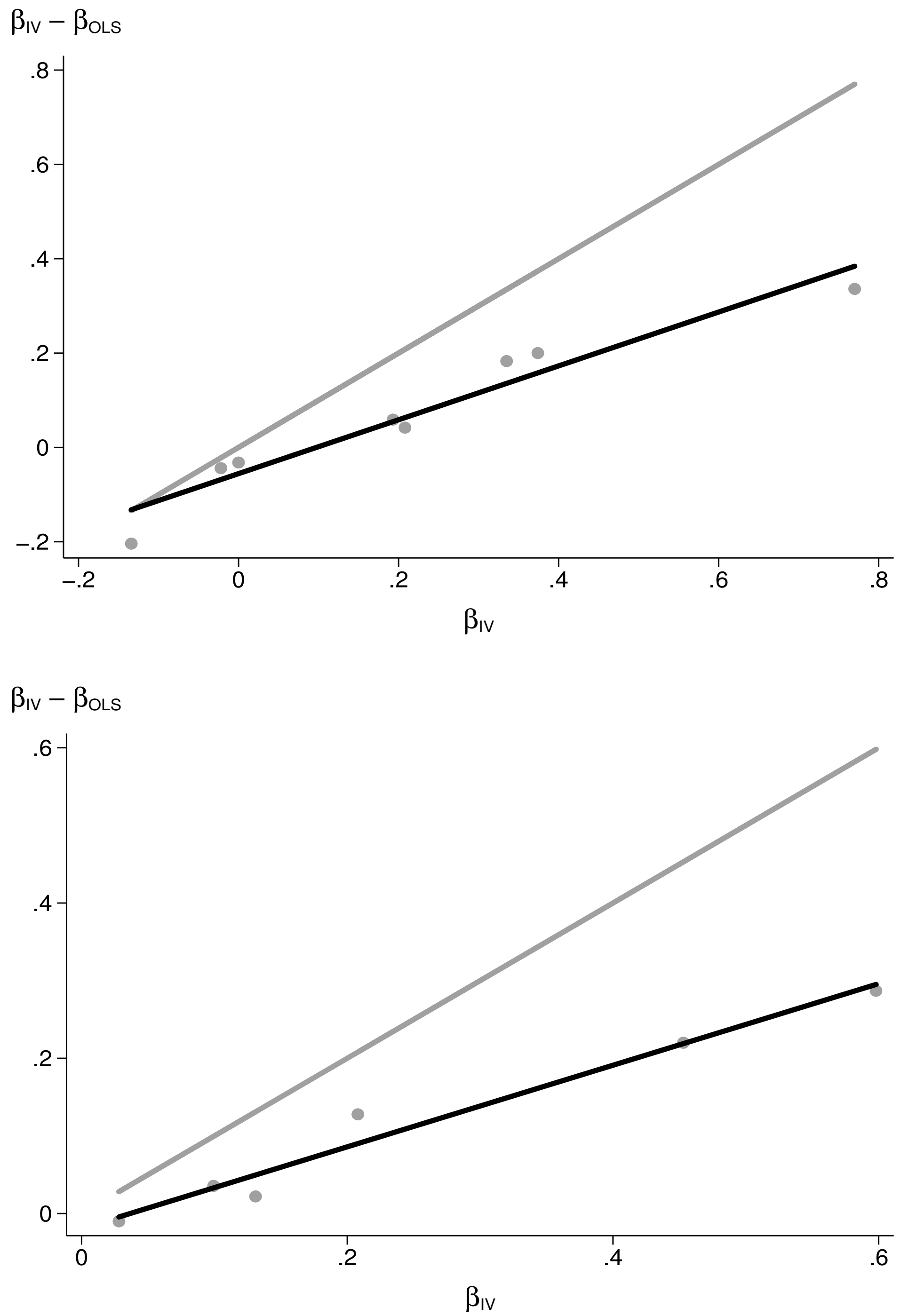

Figure 6. Within-Regressor Scatterplots
Both panels plot $\beta_{\text{IV}} - \beta_{\text{OLS}}$ against β_{IV}. The top panel plots 8 regressions reported in Mian and Sufi (2014), where the x variable is county-level growth in housing net worth (instrumented with the Saiz 2010 housing supply elasticity), 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 black lines are the estimated regression lines, while the gray lines are 45-degree lines.

Clearly, variation in β_{IV} does not translate one-for-one into variation in β_{OLS}: as β_{IV} ranges from near-zero to almost 0.6, β_{OLS} ranges from near-zero to approximately 0.3, resulting in an estimated θ of about 50%. Any level difference in the IV and OLS coefficients is absorbed by the constant term, and we estimate a positive θ because the OLS coefficients are considerably less variable than the IV coefficients. The same dynamic is apparent in Table IV of Mian and Sufi (2014).

Thus, we identify measurement error using the co-movement of OLS and IV coefficients; the average difference between them does not affect our estimate. This distinction is important for understanding how our results compare to Jiang (2017), who considers explanations for the average ratio of IV and OLS coefficients. We show below that her explanations require additional assumptions to account for the attenuated co-movement of OLS and IV coefficients.

6 Alternative Explanations

Using our sample of 2,552 IV-OLS regression pairs collected from 323 published papers, we have established two facts: first, that the IV coefficient is almost always larger in absolute value than its OLS counterpart, and second, that the OLS bias is positively related to the magnitude of the estimated IV coefficient. Measurement error in the endogenous regressor explains both facts simultaneously; in this section, we consider alternative hypotheses.

Jiang (2017) lists three reasons why we might observe $|\beta_{\text{IV}}| \gg |\beta_{\text{OLS}}|$, even when theory predicts the opposite. First, instrumental variable estimators return a local average treatment effect (LATE), while OLS returns an average treatment effect (ATE); it is possible that, on average, the difference between "local" and "global" is more important than the endogeneity problem itself. Second, weak instruments can exacerbate the bias associated with a slightly-invalid instrument and make β_{IV} explode relative to β_{OLS}. Finally, because papers generally require statistically significant results in order to be published, specification search and p-hacking can result in estimates of β_{IV} that are further from zero than β_{OLS}. We address each of these alternative explanations below, but before doing so, we note that only one of these three stories (publication bias) predicts $|\beta_{\text{IV}}| > |\beta_{\text{OLS}}|$ without additional assumptions, and none of them predict that the OLS bias is increasing in the level of β_{IV}.

6.1 Local-Average vs. Average Treatment Effects

It is well-known that an IV regression produces a local average treatment effect (LATE), while an OLS regression—in the absence of omitted variables, simultaneity, and measurement error—delivers an average treatment effect (ATE). Consider the following twist on the system of equations (1):[8]

$$y^* = \beta_x(\alpha_z)\,x^* + \beta_w w + \varepsilon_y,$$
$$x^* = \alpha_z z^* + \alpha_w w + \varepsilon_x,$$

where $\alpha_z \in \{0,1\}$ and $\beta_x(\cdot)$ are heterogeneous coefficients, and $\beta_x(\cdot)$ depends on α_z. In this case, the IV estimate β_{IV} will converge to $\beta_x(1)$. If $\alpha_w\beta_w = 0$ and $x = x^*$, the OLS estimate β_{OLS} will converge to a weighted average of $\beta_x(1)$ and $\beta_x(0)$, where the weights are determined by the distribution of α_z; if we further assume that $\left|\beta_x(1)\right| > \left|\beta_x(0)\right|$, then it immediately follows that $|\beta_{\mathrm{IV}}| > |\beta_{\mathrm{OLS}}|$. Indeed, even if $\alpha_w\beta_w \neq 0$ and $x \neq x^*$, it is possible that the difference between $\beta_x(1)$ and $\beta_x(0)$ is so overwhelming that it ensures $|\beta_{\mathrm{IV}}| > |\beta_{\mathrm{OLS}}|$.

The situation described above is plausible for many applications of instrumental variables. For example, Chetty, Hendren and Katz (2016) study the impact on adulthood wages of moving to a lower-poverty neighborhood when young; their instrument is an indicator for random assignment to receiving a voucher to move to a lower-poverty neighborhood. It seems natural that those families who would most benefit from moving away from their current neighborhood would "comply" with the lottery and move, so that $\left|\beta_x(1)\right| > \left|\beta_x(0)\right|$. Similar logic applies to many papers using experimental treatments, for example Dobbie and Fryer (2015).[9] In such cases, the existence of compliers with the treatment naturally generates a LATE that is much larger than the ATE, regardless of the endogeneity bias.

However, this explanation based on compliers is not general enough to explain the preponderance of $|\beta_{\mathrm{IV}}| \gg |\beta_{\mathrm{OLS}}|$, because it does not apply to a wide variety of cases. For example, a common instrumental variable in the finance literature is an indicator for inclusion in a stock market index, usually to instrument for institutional ownership (Fich, Harford and Tran, 2015; Schmidt and Fahlenbrach, 2017). There is no natural explanation for why, e.g., index inclusion would happen to pick up a larger local average treatment effect of institutional ownership on board composition. Likewise, many papers in our data use geographic distance to instrument for trade or immigration costs (Nunn and Wantchekon, 2011; Souza-Rodrigues, 2018; Karadja and Prawitz, 2019). Again, there is no obvious mechanism involving compliers in these settings. Thus, while in some cases we can expect $\left|\beta_x(1)\right| \gg \left|\beta_x(0)\right|$, perhaps to such an extent that IV coefficients are larger than OLS when the endogeneity problem suggests the reverse should be true, it is not as general an explanation as classical

[8] We thank Andrew Y. Chen for this concise formulation of the LATE vs. ATE issue.

[9] Unfortunately neither Chetty, Hendren and Katz (2016) nor Dobbie and Fryer (2015) are included in our sample, because they report the reduced-form regression of y on z, rather than the endogenous OLS regression of y on x.

	Dependent Variable: $\beta_{\text{IV}} - \beta_{\text{OLS}}$		
	(1)	(2)	(3)
	A. Sample Split by F-statistic		
	F-stat reported	F-stat < 24.48	F-stat ≥ 24.48
β_{IV}	0.397	0.386	0.411
	(0.049)	(0.001)	(0.106)
R^2	0.816	0.930	0.721
Observations	1,545	757	788
Number of papers	172	92	123

	B. Publication Bias		
	$\lvert t_{IV}\rvert < 1.96$	First 33%	Last 33%
β_{IV}	0.374	0.349	0.420
	(0.000)	(0.002)	(0.091)
R^2	0.724	0.996	0.628
Observations	937	604	729
Number of papers	172	125	125

Table 7. 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 presents results for the sub-sample of observations that report an 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 24.48. The first column of panel B restricts the sample to observations for which β_{IV} is statistically significant at the 5% level; 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.

measurement error.

Second, even if on average $\left|\beta_x(1)\right| \gg \left|\beta_x(0)\right|$, it would not explain the central and novel fact of our paper, which is that the average OLS bias $\beta_{\text{IV}} - \beta_{\text{OLS}}$ is increasing in β_{IV}. To generate a positive covariance between the OLS bias and β_{IV}, a story rooted in the difference between LATE and ATE would require an additional assumption: that β_{IV} is increasing in the extent of heterogeneity $\beta_x(1) - \beta_x(0)$. Thus, a LATE vs. ATE story requires two additional assumptions to explain two facts in the data, while classical measurement error neatly explains both.

6.2 Weak Instruments

Jiang (2017) notes that when an instrument is slightly invalid, the IV coefficient may differ from the true β_x, which can generate $|\beta_{\text{IV}}| \gg |\beta_{\text{OLS}}|$ if the instrument is also weak. To address this possibility, column 1 of panel A of Table 7 estimates equation (3) on the 1,545 observations which report a first-stage F-statistic. The estimate of $\mathbb{E}[\theta]$ 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 24.48.[10] The estimated $\mathbb{E}[\theta]$ for the weaker instruments is almost unchanged; for the stronger instruments, our estimate is even larger. Weak instruments are not driving our estimated measurement error ratio.

6.3 Publication Bias

Weak instruments by themselves are not enough to explain why $|\beta_{\text{IV}}| \gg |\beta_{\text{OLS}}|$, and they could just as easily predict $|\beta_{\text{IV}}| < |\beta_{\text{OLS}}|$. However, weak instruments become a stronger explanation when combined with both the inherent selection that goes into the publication process and the fact that IV standard errors tend to be much larger than their OLS counterparts. Jiang (2017) provides a simple example in which a bias towards publishing statistically significant results leads to β_{IV} that are substantially larger than β_{OLS}, even if the reverse is true in the population of attempted and unpublished studies.

In this section, we address the issue of publication bias in two ways. In Section 6.3.1 we present reduced-form evidence from our data that suggests that publication bias is not the entire story behind our results. Then, in Section 6.3.2 we estimate a structural model of publication bias, extending the model of Andrews and Kasy (2019) to incorporate both OLS and IV estimates as well as measurement error. We find that publication bias has little effect on our estimate of the extent of measurement error, and that measurement error by itself can explain over 90% of the median IV-OLS ratio.

6.3.1 Reduced-Form Evidence

The high percentage of statistically insignificant β_{IV} in our data (37%), alongside the fact that the OLS bias is so large and of the wrong sign even for statistically insignificant β_{IV} (Figure 4), suggest that publication bias cannot be the whole story. Nevertheless, in this

[10] 24.48 is the median F-statistic by IV regression, for which we have 1,545 observations, whereas the median F-statistic by first-stage regression (reported in Table 1) is 19, for which we have 935 observations. The difference is due to the fact that the same first stage is often used in multiple separate IV regressions; results in Panel A of Table 7 are similar using the alternative median F-statistic as a cutoff.

section we perform two additional tests to determine the extent to which publication bias drives our results. First, when estimating $\mathbb{E}[\theta]$, we restrict our sample to estimates of β_{IV} that are statistically insignificant at conventional levels. Second, we split the sample based on a regression's location in a paper, since results reported later on are less likely to be as important for the publication decision.

In column 1 of panel B of Table 7, we restrict the sample to observations of β_{IV} that are statistically insignificant at the 5% level. The estimated θ is again statistically significant, and similar to many of the estimates in Tables 2 and 5.

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 eight IV-OLS regression pairs, while the median paper has 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 7 we restrict the sample to those papers with at least 6 regressions, and include only IV-OLS pairs for which the IV regression occurs either in the first or last 33% of the paper's IV regressions, respectively. For example, 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 $\mathbb{E}[\theta]$ should be very different from our benchmark estimates; yet, we find that they are well within the range of estimates reported in Tables 2 and 4. Although $\mathbb{E}[\theta]$ is somewhat lower for results appearing later on in papers, it is still quite large at 0.420. Thus, publication bias does not unduly influence our estimate of the average measurement error.

6.3.2 Structural Evidence

We estimate a selection model to discern the effects of publication bias on both the average in-sample measurement error ratio $\mathbb{E}[\theta]$ and the median IV-OLS ratio Median $\left(|\beta_{IV}/\beta_{OLS}|\right)$.[11]

We assume that statistically significant results are q times as likely to be published as statistically insignificant results. Formally, we assume that the probability $p(\cdot)$ that a given

[11] Hannah R. Rothstein (2006) and Christensen and Miguel (2016) discuss methods for uncovering publication bias. Recent applications include Andrews and Kasy (2019) and Chen and Zimmermann (2019).

IV-OLS pair is published is given by

$$p\left(t_{IV}\right)=\begin{cases}p_1 & \text{if } |t_{IV}|<1.96\\ p_2 & \text{if } |t_{IV}|\geq 1.96,\end{cases} \tag{6}$$

where t_{IV} denotes the t-statistic of the instrumental variables estimate β_{IV}, p_1 is the probability that statistically-insignificant results are published, p_2 is the probability that statistically-significant results are published, and $q=p_2/p_1$. If $q>1$, for example, then statistically significant results are more likely to be published than statistically insignificant results. We estimate q alongside the joint distribution of the OLS and IV coefficients and standard errors using the simulated method of moments; see Appendix B for details on our distributional assumptions, moment conditions, and parameter estimates. To ensure that all coefficients and standard errors are comparable, we restrict attention to the $1,015$ observations that we can normalize by σ_x/σ_y.

After accounting for publication bias, we estimate that the average measurement ratio $\mathbb{E}\left[\theta\right]$ is 0.459 with a bootstrapped standard error of 0.030. For comparison, the in-sample meta-OLS estimate is 0.666 with a standard error of 0.044, which corresponds to the results reported in column 2 of Table 2. This suggests that publication bias has only a minor effect on our estimate of $\mathbb{E}\left[\theta\right]$, although we do find evidence of publication bias: our point estimate for q is 1.294, implying that statistically significant estimates are almost 30% more likely to be published than insignificant results.

Figure 7 decomposes the effects of publication bias and measurement error on the median IV-OLS ratio Median $\left(|\beta_{IV}/\beta_{OLS}|\right)$. For these 1,015 observations, the median ratio is 2.091. As Jiang (2017) suggests, publication bias does indeed play a role in magnifying this ratio: absent publication bias, IV coefficients would only be 1.637 times as large as OLS coefficients. However, measurement error plays a larger role: if there were no measurement error, IV coefficients would be 1.037 times the size of OLS coefficients. Absent both publication bias and measurement error, IV coefficients would be 1.027 the size of OLS coefficients. Put differently, publication bias can explain $(2.089-1.637)/(2.089-1.027)\approx 42.6\%$ of this magnification, while measurement error can explain $(2.089-1.037)/(2.089-1.027)\approx 99.1\%$.

If we restrict attention to a single regression per paper, as in column 4 of Table 2, we find that significant results are 53% more likely to be published; however, this bias has no statistically-significant effect on our estimate of $E[\theta]$. In this sample, publication bias can explain 35.6% of the IV-OLS ratio, while measurement error can explain over 99.0%.

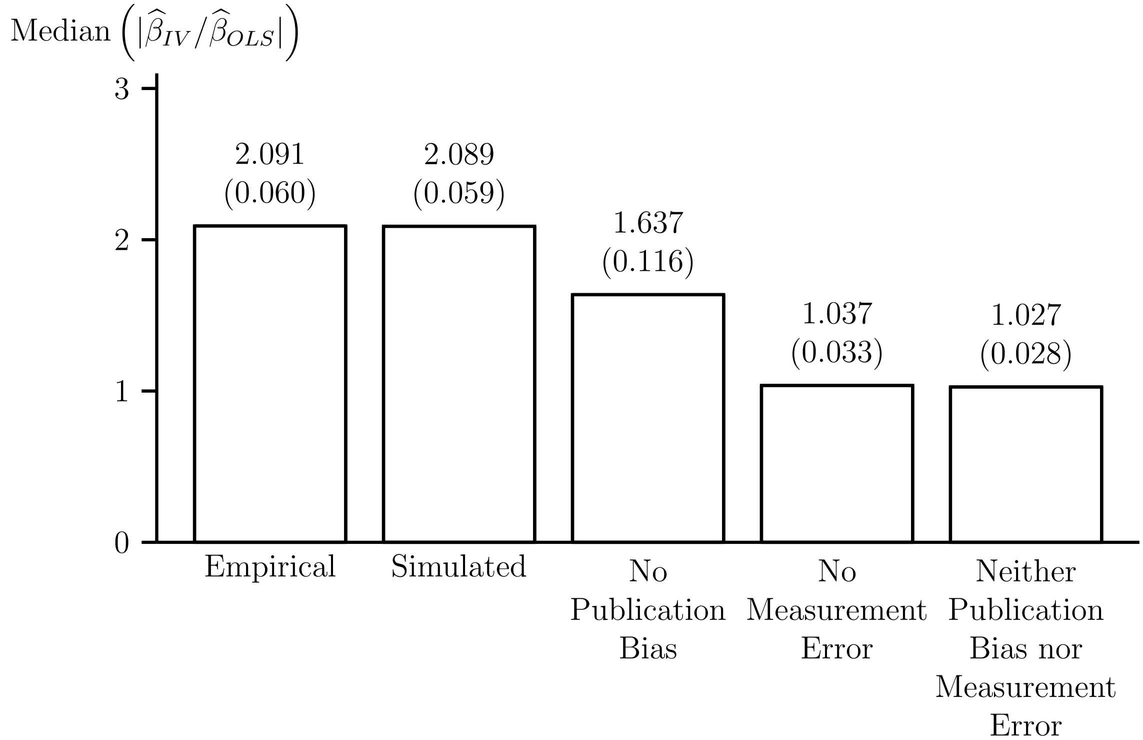

Figure 7. Median IV-OLS Ratio

This figure plots various estimates of the median IV-OLS ratio Median $\left(|\widehat{\beta}_{IV}/\widehat{\beta}_{OLS}|\right)$. "Empirical" is the in-sample median coefficient ratio. In the remaining four columns, we use the parameters estimated by the simulated method of moments: "Simulated" is the simulated in-sample median coefficient ratio; "No Publication Bias" shows what this ratio would be in the absence of publication bias; "No Measurement Error" shows what this ratio would be in the absence of measurement error; and "Neither Publication Bias nor Measurement Error" shows what this ratio would be in the absence of both publication bias and measurement error. Bootstrapped standard errors are in parentheses.

7 How to Use the Meta-OLS Estimator to Estimate the Average Omitted Variable Bias

In over 80% of regressions for which researchers expect to find an IV coefficient smaller than the OLS coefficient, they in fact find the opposite (Figure 3). Using the meta-OLS estimator, a researcher can discern whether this indicates a flaw in their story, or if their regressor is simply subject to measurement error.[12] A researcher who wishes to estimate the average omitted variable bias can run a meta-OLS regression as follows:

1. Collect N distinct triplets $\{x_i, y_i, z_i\}_{i=1}^{N}$ where each x_i is an endogenous regressor, y_i an outcome variable, z_i an instrument or vector of instruments, and i indexes variable triplets as opposed to individual observations of those variables.[13]

2. For each i, estimate an OLS regression of y_i on x_i and the controls. Collect these N coefficients $\{\beta_{\text{OLS}}^{i}\}_{i=1}^{N}$.

3. For each i, estimate a 2SLS regression of y_i on x_i and the controls, using instrument(s) z_i. Collect these N coefficients $\left\{\beta_{\text{IV}}^{i}\right\}_{i=1}^{N}$.

4. Estimate an OLS regression of β_{OLS}^{i} on β_{IV}^{i} and a constant; this is a single meta-OLS regression with N observations. The estimated slope coefficient is equal to 1 minus the average measurement error ratio (i.e., $1 - \mathbb{E}[\theta]$). The estimated constant term is the average omitted variable bias $\mathbb{E}[\text{OVB}_i]$ across all N regressions.[14]

In cases where β_{OLS}^{i} is systematically less than β_{IV}^{i}, but the researcher anticipated the opposite, the meta-OLS estimator will reveal if this discrepancy is due to measurement error as opposed to a flaw in either the instrument or the researcher's intuition for the omitted variable bias. The fact that so many of the estimates reported in Mian and Sufi (2014) have $\beta_{\text{IV}}^{i} > \beta_{\text{OLS}}^{i}$ might lead a reader to discount the paper, when in fact the results are consistent with their intuition regarding the direction of the omitted variable bias. As we show in Section 5.2, the discrepancy between their intuition and their finding of $\beta_{\text{IV}}^{i} > \beta_{\text{OLS}}^{i}$ is purely due to measurement error.

If a researcher uses the same $x_i = x$ variable while only varying y_i, then they will obtain an estimate of the measurement error for that single regressor x; i.e., they will obtain an estimate of θ as opposed to $\mathbb{E}[\theta]$. This will allow them to estimate the individual omitted variable bias OVB_i for each regression i, where $\text{OVB}_i = \beta_i^{\text{OLS}} - (1-\theta)\,\beta_i^{\text{IV}}$. It may therefore be worthwhile for authors to seek out distinct y variables for no other reason than to estimate the degree of measurement error and omitted variable bias.

[12] This section applies to cases in which a researcher is concerned with omitted variable bias. If the researcher is also concerned with simultaneity, they would also need multiple instruments z in order to conduct a meta-IV regression.

[13] These triplets need to be distinct from each other, but a given x_i can appear multiple times. Indeed, in our regressor-level estimates in Section 5.2, we use the same $x_i = x$ and $z_i = z$ while only varying y_i.

[14] This is different from the analysis in the rest of the paper, where we regress $\beta_{\text{IV}}^{i} - \beta_{\text{OLS}}^{i}$ on β_{IV}^{i}. That yields a slope estimate of $\mathbb{E}[\theta]$, as opposed to $1 - \mathbb{E}[\theta]$, but an intercept estimate of $-\mathbb{E}[\text{OVB}_i]$, as opposed to $\mathbb{E}[\text{OVB}_i]$.

8 Conclusion

Proxy variables are noisy and pervasive. Using a sample of 2,552 pairs of IV and OLS coefficients from 323 papers published in top economics and finance journals, we estimate that over 30% of the variation in the average regressor is pure noise. While over 82% of the papers in our sample are explicitly concerned with omitted variables or simultaneity, the reality is that most estimates are suffering from attenuation bias *regardless* of the supposed endogeneity problem (Jiang, 2017). Indeed, even when researchers expect omitted variables or simultaneity to inflate their estimates, the opposite occurs 81% of the time; this incessant attenuation bias, that respects neither sign nor story, is a hallmark of measurement error.

Instrumental variables can extirpate this error, provided the instrument addresses measurement error and not just omitted variables or simultaneity (Roberts and Whited, 2013). With such an instrument, our meta-OLS estimator can be used to separate attenuation bias from the average omitted variable bias; we recommend this estimator to authors who find that their IV coefficients are larger than their OLS coefficients, as it can uncover a sizeable amount of measurement error in the proxy variable. Measurement error matters, and this noise demands attention.

References

Aldrich, John. 1993. "Reiersøl, Geary And The Idea Of Instrumental Variables." *Economic and Social Review*, 24(3): 247–273. 2, 7

Andrews, Isaiah, and Maximilian Kasy. 2019. "Identification Of And Correction For Publication Bias." *American Economic Review*, 109(8): 2766–94. 28, 29, 42

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: an Empiricist's Companion.* Princeton University Press. 2

Angrist, Joshua D., and Jörn-Steffen Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics." *Journal of Economic Perspectives*, 24(2): 3–30. 4

Backus, Matthew. 2020. "Why Is Productivity Correlated With Competition?" *Econometrica*, 88(6): 2415–2444. 2

Becker, Sascha O., and Luigi Pascali. 2019. "Religion, Division Of Labor, And Conflict: Anti-Semitism In Germany Over 600 Years." *American Economic Review*, 109(5): 1764–1804. 23, 24

Berg, Tobias, and Daniel Streitz. 2019. "Handling Spillover Effects in Empirical Research: An Application using Credit Supply Shocks." 8

Biørn, Erik. 2000. "Panel Data With Measurement Errors: Instrumental Variables And Gmm Procedures Combining Levels And Differences." *Econometric Reviews*, 19(4): 391–424. 7

Bloom, Nicholas, and John Van Reenen. 2007. "Measuring And Explaining Management Practices Across Firms And Countries." *The Quarterly Journal of Economics*, 122(4): 1351–1408. 8

Bloom, Nicholas, Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Megha Patnaik, Itay Saporta-Eksten, and John Van Reenen. 2019. "What Drives Differences In Management Practices?" *American Economic Review*, 109(5): 1648–83. 8

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2018. "Quasi-Experimental Shift-Share Research Designs." National Bureau of Economic Research Working Paper 24997. 8

Bound, John, and Alan B. Krueger. 1991. "The Extent Of Measurement Error In Longitudinal Earnings Data: Do Two Wrongs Make A Right?" *Journal of Labor Economics*, 9(1): 1–24. 8

Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. "Measurement Error In Survey Data." In . Vol. 5 of *Handbook of Econometrics*, , ed. James J. Heckman and Edward Leamer, 3705–3843. Elsevier. 7

Card, David. 2001. "Estimating The Return To Schooling: Progress On Some Persistent Econometric Problems." *Econometrica*, 69(5): 1127–1160. 4, 8, 10

Chen, Andrew Y, and Tom Zimmermann. 2019. "Publication Bias and the Cross-Section of Stock Returns." *The Review of Asset Pricing Studies*, 10(2): 249–289. 29

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *American Economic Review*, 106(4): 855–902. 26

Christensen, Garret S, and Edward Miguel. 2016. "Transparency, Reproducibility, and the Credibility of Economics Research." National Bureau of Economic Research Working Paper 22989. 29

DiTraglia, Francis J., and Camilo García-Jimeno. 2020. "A Framework For Eliciting, Incorporating, And Disciplining Identification Beliefs In Linear Models." *Journal of Business Economic Statistics*, 0(0): 1–16. 7

Dobbie, Will, and Roland G. Fryer. 2015. "The Medium-Term Impacts of High-Achieving Charter Schools." *Journal of Political Economy*, 123(5): 985–1037. 26

Erickson, Timothy, and Toni M. Whited. 2000. "Measurement Error and the Relationship between Investment and q." *The Journal of Political Economy*, 108(5): 1027–1057. 2, 8

Erickson, Timothy, and Toni M. Whited. 2002. "Two-Step Gmm Estimation Of The Errors-In-Variables Model Using High-Order Moments." *Econometric Theory*, 18(3): 776–799. 3, 8

Erickson, Timothy, and Toni M. Whited. 2012. "Treating Measurement Error In Tobin's Q." *The Review of Financial Studies*, 25(4): 1286–1329. 2, 8

Erickson, Timothy, Colin Huan Jiang, and Toni M. Whited. 2014. "Minimum Distance Estimation Of The Errors-In-Variables Model Using Linear Cumulant Equations." *Journal of Econometrics*, 183(2): 211–221. Analysis of Financial Data. 3, 4, 8

Feyrer, James, Erin T. Mansur, and Bruce Sacerdote. 2017. "Geographic Dispersion Of Economic Shocks: Evidence From The Fracking Revolution." *American Economic Review*, 107(4): 1313–34. 16

Fich, Eliezer M., Jarrad Harford, and Anh L. Tran. 2015. "Motivated Monitors: The importance of institutional investors's portfolio weights." *Journal of Financial Economics*, 118(1): 21 – 48. 26

Geary, R. C. 1943. "Relations Between Statistics: The General And The Sampling Problem When The Samples Are Large." *Proceedings of the Royal Irish Academy. Section A: Mathematical and Physical Sciences*, 49: 177–196. 2, 7

Goolsbee, A. 2000. "The Importance Of Measurement Error In The Cost Of Capital." *National Tax Journal*, 53(2): 215–228. 8

Griliches, Zvi, and Jerry A. Hausman. 1986. "Errors In Variables In Panel Data." *Journal of Econometrics*, 31(1): 93–118. 7

Hahn, Jinyong, Jerry Hausman, and Jeonghwan Kim. 2020. "A Small Sigma Approach to Certain Problems in Errors-in-Variables and Panel Data Models." *Working paper.* 8

Hannah R. Rothstein, Alexander J. Sutton, Michael Borenstein. 2006. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments.* Hoboken, NJ: WIley. 29

Hausman, Jerry. 2001. "Mismeasured Variables In Econometric Analysis: Problems From The Right And Problems From The Left." *The Journal of Economic Perspectives*, 15(4): 57–67. 7

Honoré, Bo E, and James L Powell. 1994. "Pairwise difference estimators of censored and truncated regression models." *Journal of Econometrics*, 64(1): 241 – 278. 42

Ichino, Andrea, and Rudolf Winter-Ebmer. 2004. "The Long-Run Educational Cost Of World War II." *Journal of Labor Economics*, 22(1): 57–87. 8

Jiang, Wei. 2017. "Have Instrumental Variables Brought Us Closer To The Truth." *The Review of Corporate Finance Studies*, 6(2): 127–140. 5, 8, 10, 15, 25, 28, 30, 33

Karadja, Mounir, and Erik Prawitz. 2019. "Exit, Voice, And Political Change: Evidence From Swedish Mass Migration To The United States." *Journal of Political Economy*, 127(4): 1864–1925. 26

Lal, Apoorva, Mackenzie William Lockhart, Yiqing Xu, and Ziwen Zu. 2021. "How Much Should We Trust Instrumental Variable Estimates in Political Science? Practical Advice based on Over 60 Replicated Studies." *Working Paper, Stanford University.* 8

Mian, Atif, and Amir Sufi. 2014. "What Explains The 2007–2009 Drop In Employment?" *Econometrica*, 82(6): 2197–2223. 5, 22, 23, 24, 25, 32

Nunn, Nathan, and Leonard Wantchekon. 2011. "The Slave Trade And The Origins Of Mistrust In Africa." *American Economic Review*, 101(7): 3221–52. 26

Oster, Emily. 2019. "Unobservable Selection And Coefficient Stability: Theory And Evidence." *Journal of Business & Economic Statistics*, 37(2): 187–204. 8

Reiersøl, Olav. 1941. "Confluence Analysis By Means Of Lag Moments And Other Methods Of Confluence Analysis." *Econometrica*, 9(1): 1–24. 2, 7

Roberts, Michael R, and Toni M Whited. 2013. "Endogeneity in Empirical Corporate Finance." In *Handbook of the Economics of Finance*. Vol. 2, 493–572. Elsevier. 10, 33

Saiz, Albert. 2010. "The Geographic Determinants Of Housing Supply." *The Quarterly Journal of Economics*, 125(3): 1253–1296. 24

Schennach, S. M., and Yingyao Hu. 2013. "Nonparametric Identification And Semiparametric Estimation Of Classical Measurement Error Models Without Side Information." *Journal of the American Statistical Association*, 108(501): 177–186. 8

Schennach, Susanne M. 2016. "Recent Advances In The Measurement Error Literature." *Annual Review of Economics*, 8(1): 341–377. 8

Schmidt, Cornelius, and Rudiger Fahlenbrach. 2017. "Do Exogenous Changes In Passive Institutional Ownership Affect Corporate Governance And Firm Value?" *Journal of Financial Economics*, 124(2): 285–306. 26

Souza-Rodrigues, Eduardo. 2018. "Deforestation In The Amazon: A Unified Framework For Estimation And Policy Analysis." *The Review of Economic Studies*, 86(6): 2713–2744. 26

Stock, James H., and Francesco Trebbi. 2003. "Retrospectives: Who Invented Instrumental Variable Regression?" *Journal of Economic Perspectives*, 17(3): 177–194. 7

Stroebel, Johannes, and Joseph Vavra. 2019. "House Prices, Local Demand, And Retail Prices." *Journal of Political Economy*, 127(3): 1391–1436. 2

Whited, Toni M. 2001. "Is It Inefficient Investment That Causes The Diversification Discount?" *The Journal of Finance*, 56(5): 1667–1691. 8

Wright, Philip G. 1928. *The Tariff on Animal and Vegetable Oils.* New York: Macmillan. 5, 7

A Proofs

Proof of Proposition 1

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

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

The OLS regression yields

$$\begin{aligned}
\text{plim } \beta_{\text{OLS}} &= \frac{\text{cov}(y, x)}{\text{var}(x)} \\
&= \frac{\text{cov}(\beta_x x^* + g, x^* + u_x)}{\text{var}(x)} \\
&= \frac{\beta_x \text{var}(x^*) + \text{cov}(g, x^*)}{\text{var}(x)} \\
&= \frac{\beta_x \text{var}(x^*) + (1 - \alpha_y \beta_x)^{-1} \text{cov}(g, f + \alpha_y g)}{\text{var}(x)} \\
&= \frac{\beta_x \text{var}(x^*) + (1 - \alpha_y \beta_x)^{-1} (\text{cov}(g, f) + \alpha_y \text{var}(g))}{\text{var}(x)} \\
&\equiv \beta_x \left(\frac{\sigma_{x^*}^2}{\sigma_x^2} \right) + \frac{1}{1 - \alpha_y \beta_x} \left(\frac{\sigma_{fg} + \alpha_y \sigma_g^2}{\sigma_x^2} \right).
\end{aligned}$$

The IV regression yields

$$\begin{aligned}
\text{plim } \beta_{\text{IV}} &= \frac{\text{cov}(y, z)}{\text{cov}(x, z)} \\
&= \frac{\text{cov}(\beta_x x^* + g, z)}{\text{cov}(x, z)} \\
&= \frac{\beta_x \text{cov}(x^*, z)}{\text{cov}(x, z)} \\
&= \beta_x.
\end{aligned}$$

□

Proof of Proposition 2

The assumption that β_x is generated independently of all other parameters that generate the data is sufficient, but not necessary; in fact, we only require that β_x satisfies

$$\begin{aligned}
\mathbb{E}\left[\beta_x \frac{\sigma_{fg}}{\sigma_x^2}\right] &= \mathbb{E}\left[\beta_x\right] \mathbb{E}\left[\frac{\sigma_{fg}}{\sigma_x^2}\right], \\
\mathbb{E}\left[\beta_x \theta\right] &= \mathbb{E}\left[\beta_x\right] \mathbb{E}\left[\theta\right], \\
\mathbb{E}\left[\beta_x^2 \theta\right] &= \mathbb{E}\left[\beta_x^2\right] \mathbb{E}\left[\theta\right].
\end{aligned}$$

The meta-OLS regression yields

$$\begin{aligned}
\text{plim } b_{\text{OLS}} &= \frac{\text{cov}\left(\beta_{\text{IV}} - \beta_{\text{OLS}}, \beta_{\text{IV}}\right)}{\text{var}(\beta_{\text{IV}})} \\
&= \frac{\text{cov}\left(\sigma_x^{-2}\left[\beta_x \sigma_{u_x}^2 - \sigma_{fg}\right], \beta_x\right)}{\text{var}(\beta_x)} \\
&= \frac{\text{cov}\left(\beta_x \theta, \beta_x\right)}{\text{var}(\beta_x)} \\
&= \mathbb{E}\left[\theta\right].
\end{aligned}$$

□

Proof of Proposition 3

The assumption that β_x and θ are generated independently of all other parameters is sufficient, but not necessary; in fact, we only require that they satisfy

$$\begin{aligned}
\mathbb{E}\left[\beta_x \frac{\sigma_{fz}}{\sigma_z^2}\left(1-\theta\right)\zeta\right] &= \mathbb{E}\left[\beta_x\right] \mathbb{E}\left[\frac{\sigma_{fz}}{\sigma_z^2}\left(1-\theta\right)\zeta\right], \\
\mathbb{E}\left[\beta_x \left(1-\alpha_y\beta_x\right)^{-1} \frac{\sigma_{fz}}{\sigma_z^2}\theta\right] &= \mathbb{E}\left[\beta_x \left(1-\alpha_y\beta_x\right)^{-1} \frac{\sigma_{fz}}{\sigma_z^2}\right] \mathbb{E}\left[\theta\right], \\
\mathbb{E}\left[\beta_x^2 \left(1-\alpha_y\beta_x\right)^{-1} \frac{\sigma_{fz}}{\sigma_z^2}\theta\right] &= \mathbb{E}\left[\beta_x^2 \left(1-\alpha_y\beta_x\right)^{-1} \frac{\sigma_{fz}}{\sigma_z^2}\right] \mathbb{E}\left[\theta\right],
\end{aligned}$$

where

$$\zeta \equiv \frac{\sigma_{fg} + \alpha_y \sigma_g^2}{\sigma_f^2 + 2\alpha_y \sigma_{fg} + \alpha_y^2 \sigma_g^2}.$$

The meta-IV regression yields

$$
\begin{aligned}
\text{plim } b_{IV} &= \frac{\text{cov}\left(\left(\beta_{IV}-\beta_{OLS}\right)\left|\beta_{FS}\right|,\beta_{IV}\right)}{\text{cov}\left(\beta_{IV}\left|\beta_{FS}\right|,\beta_{IV}\right)}\\
&= \frac{\text{cov}\left(\left(\beta_{IV}-\sigma_x^{-2}\left(\beta_x\sigma_{x^*}^2+\left(1-\alpha_y\beta_x\right)^{-1}\left(\sigma_{fg}+\alpha_y\sigma_g^2\right)\right)\right)\left|\beta_{FS}\right|,\beta_{IV}\right)}{\text{cov}\left(\beta_{IV}\left|\beta_{FS}\right|,\beta_{IV}\right)}\\
&= \frac{\text{cov}\left(\left(\beta_{IV}-(1-\theta)\left(\beta_x+\sigma_{x^*}^{-2}\left(1-\alpha_y\beta_x\right)^{-1}\left(\sigma_{fg}+\alpha_y\sigma_g^2\right)\right)\right)\left|\beta_{FS}\right|,\beta_{IV}\right)}{\text{cov}\left(\beta_{IV}\left|\beta_{FS}\right|,\beta_{IV}\right)}\\
&= \frac{\text{cov}\left(\left(\theta\beta_{IV}-(1-\theta)\left(\sigma_f^2+2\alpha_y\sigma_{fg}+\alpha_y^2\sigma_g^2\right)^{-1}\left(1-\alpha_y\beta_x\right)\left(\sigma_{fg}+\alpha_y\sigma_g^2\right)\right)\left|\beta_{FS}\right|,\beta_{IV}\right)}{\text{cov}\left(\beta_{IV}\left|\beta_{FS}\right|,\beta_{IV}\right)}\\
&= \frac{\text{cov}\left(\left(\theta\beta_{IV}-(1-\theta)\left(1-\alpha_y\beta_x\right)\zeta\right)\left|\beta_{FS}\right|,\beta_{IV}\right)}{\text{cov}\left(\beta_{IV}\left|\beta_{FS}\right|,\beta_{IV}\right)}\\
&= \frac{\text{cov}\left(\left(\theta\beta_{IV}-(1-\theta)\left(1-\alpha_y\beta_x\right)\zeta\right)\left(1-\alpha_y\beta_x\right)^{-1}\sigma_z^{-2}\left|\sigma_{fz}\right|,\beta_{IV}\right)}{\text{cov}\left(\beta_{IV}\left(1-\alpha_y\beta_x\right)^{-1}\sigma_z^{-2}\left|\sigma_{fz}\right|,\beta_{IV}\right)}\\
&= \frac{\text{cov}\left(\theta\beta_{IV}\left(1-\alpha_y\beta_x\right)^{-1}\sigma_z^{-2}\sigma_{fz}-(1-\theta)\,\sigma_z^{-2}\left|\sigma_{fz}\right|\zeta,\beta_{IV}\right)}{\text{cov}\left(\beta_{IV}\left(1-\alpha_y\beta_x\right)^{-1}\sigma_z^{-2}\left|\sigma_{fz}\right|,\beta_{IV}\right)}\\
&= \frac{\text{cov}\left(\theta\beta_{IV}\left(1-\alpha_y\beta_x\right)^{-1}\sigma_z^{-2}\left|\sigma_{fz}\right|,\beta_{IV}\right)}{\text{cov}\left(\beta_{IV}\left(1-\alpha_y\beta_x\right)^{-1}\sigma_z^{-2}\left|\sigma_{fz}\right|,\beta_{IV}\right)}\\
&= \mathbb{E}\left[\theta\right].
\end{aligned}
$$

□

Proof of Corollary

Let $\beta_{FS,j}$ denote an arbitrary first-stage coefficient obtained by instrumenting x with any subset of the vector $[z_1, z_2, ..., z_{n_z}]'$. If we use $\sum_j \omega_j \left|\beta_{FS,j}\right|$ in lieu of $\left|\beta_{FS}\right|$, the meta-IV will yield

$$
\begin{aligned}
\operatorname{plim} b_{IV} &= \frac{\operatorname{cov}\left(\left(\beta_{IV}-\beta_{OLS}\right)\sum_j \omega_j \left|\beta_{FS,j}\right|, \beta_{IV}\right)}{\operatorname{cov}\left(\beta_{IV}\sum_j \omega_j \left|\beta_{FS,j}\right|, \beta_{IV}\right)} \\
&= \frac{\operatorname{cov}\left(\sum_j \omega_j \beta_{IV}\left(1-\alpha_y\beta_x\right)^{-1}\sigma_{z_j}^{-2}\left|\sigma_{fz_j}\right|\theta - \sum_j \omega_j\left(1-\theta\right)\sigma_{z_j}^{-2}\left|\sigma_{fz_j}\right|\zeta, \beta_{IV}\right)}{\operatorname{cov}\left(\sum \omega_j \beta_{IV}\left(1-\alpha_y\beta_x\right)^{-1}\sigma_{z_j}^{-2}\left|\sigma_{fz_j}\right|, \beta_{IV}\right)} \\
&= \frac{\operatorname{cov}\left(\beta_{IV}\left(1-\alpha_y\beta_x\right)^{-1}\sum_j \omega_j \sigma_{z_j}^{-2}\left|\sigma_{fz_j}\right|\theta, \beta_{IV}\right)}{\operatorname{cov}\left(\beta_{IV}\left(1-\alpha_y\beta_x\right)^{-1}\sum_j \omega_j \sigma_{z_j}^{-2}\left|\sigma_{fz_j}\right|, \beta_{IV}\right)} \\
&= \mathbb{E}\left[\theta\right].
\end{aligned}
$$

□

B Estimating Publication Bias

For each IV-OLS regression pair in our dataset, we observe the estimated OLS coefficient β_{OLS} with standard error ς_{OLS} and the estimated IV coefficient β_{IV} with standard error ς_{IV}. Since we only observe published results, our dataset consists draws from the conditional distribution $F_{\beta_{OLS},\varsigma_{OLS},\beta_{IV},\varsigma_{IV}|\delta}\left(b_{OLS}, s_{OLS}, b_{IV}, s_{IV}|1\right)$, where δ is an indicator variable equal to one if the results $(b_{OLS}, s_{OLS}, b_{IV}, s_{IV})$ are published. We want to estimate the publication bias parameter $q = p_2/p_1$ from equation (6) alongside the parameters governing the unconditional joint distribution $F_{\beta_{OLS},\varsigma_{OLS},\beta_{IV},\varsigma_{IV}}\left(b_{OLS}, s_{OLS}, b_{IV}, s_{IV}\right)$; this unconditional joint distribution is the publication-bias-free joint distribution. We assume that

$$
\begin{aligned}
\theta &\sim Beta\left(a_\theta, b_\theta\right), \\
\gamma &\sim \mathbb{N}\left(\mu_\gamma, \sigma_\gamma\right), \\
\beta_{IV}^* &\sim \mathbb{N}\left(\mu_{\beta_{IV}}, \sigma_{\beta_{IV}}\right), \\
\left(\log\left(\varsigma_{OLS}\right), \log\left(\varsigma_{IV}\right)\right) &\sim \mathbb{N}\left(\mu_{\tilde{\varsigma}}, \Sigma_{\tilde{\varsigma}}\right),
\end{aligned}
$$

where $\beta_{OLS}^* \equiv \beta_{IV}^*(1-\theta)+\gamma$ per equation (2), and $\left(\beta_j - \beta_j^*\right) \sim \mathbb{N}\left(0, \varsigma_j\right)$ for $j \in \{OLS, IV\}$. In other words, β_j^* denotes the asymptotic regression coefficient, and the estimate β_j obeys the law of large numbers. We must therefore estimate the following parameter vector:

$$
\psi \equiv [a_\theta, b_\theta, \mu_\gamma, \sigma_\gamma, \mu_{\beta_{IV}}, \sigma_{\beta_{IV}}, \mu_{\tilde{\varsigma}_{OLS}}, \sigma_{\tilde{\varsigma}_{OLS}}, \mu_{\tilde{\varsigma}_{IV}}, \sigma_{\tilde{\varsigma}_{IV}}, \rho_{\tilde{\varsigma}}, q],
$$

where $\rho_{\tilde{\varsigma}}$ denotes the correlation between $\log(\varsigma_{OLS})$ and $\log(\varsigma_{IV})$.

We employ the simulated method of moments and match the empirical moments in panel A of Table 8. While most of these moments are intuitive, there are two points worth noting. First, recall that we want to study the effects of publication bias on the average measurement error ratio $\mathbb{E}[\theta]$ and the median IV-OLS ratio $\text{Median}\left(|\beta_{IV}/\beta_{OLS}|\right)$; we therefore seek to match the meta-OLS estimate alongside the in-sample median IV-OLS ratio. Second, to pin down the degree of publication bias q, we use the U-statistic

$$\mathbb{E}\left[p\left(t_{IV}\right)^{-1} p\left(t'_{IV}\right)^{-1} \left(\begin{array}{c} \mathbb{1}\left(|\beta_{IV}| < 1.96\varsigma_{IV}\right) - \\ \mathbb{1}\left(\left|\beta'_{IV} + \eta\sqrt{\varsigma_{IV}^2 - \left(\varsigma'_{IV}\right)^2}\right| < 1.96\varsigma_{IV}\right) \end{array}\right) \middle| \varsigma_{IV} > \varsigma'_{IV}\right], \tag{7}$$

$$\eta \sim \mathbb{N}(0,1),$$

from Andrews and Kasy (2019).[15] This particular moment is a pairwise difference estimator (Honoré and Powell, 1994) that uses the $\binom{1{,}015}{2} = 514{,}605$ pairs of observations

$$\left\{\left(\beta_{OLS}, \varsigma_{OLS}, \beta_{IV}, \varsigma_{IV}\right), \left(\beta'_{OLS}, \varsigma'_{OLS}, \beta'_{IV}, \varsigma'_{IV}\right)\right\}$$

for which $\varsigma_{IV} > \varsigma'_{IV}$. Instead of equating the actual and simulated U-statistics, this moment identifies parameters by setting (7) equal to 0.

For the other eleven moments, we equate each empirical moment with the corresponding simulated moment based on an average of ten simulations. More concretely, we compute the empirical moment $g_i(X)$ for $i = 1, 2, ..., 11$, where X denotes the matrix of empirical data. For each i, we also compute the corresponding simulated moment $g_i\left(Y_s(\psi)\right)$ for $s = 1, ..., 10$, where $Y_s(\psi)$ denotes the sth matrix of simulated data with parameters ψ. Our vector of twelve moments is therefore

$$G\left(X, Y_1(\psi), ..., Y_{10}(\psi), q\right) \equiv \left[\begin{array}{c} g_1(X) - \frac{1}{10}\sum_{s=1}^{10} g_1\left(Y_s(\psi)\right) \\ g_2(X) - \frac{1}{10}\sum_{s=1}^{10} g_2\left(Y_s(\psi)\right) \\ \vdots \\ g_{11}(X) - \frac{1}{10}\sum_{s=1}^{10} g_{11}\left(Y_s(\psi)\right) \\ U(X,q) \end{array}\right],$$

where $U(X,q)$ is the U-statistic (7). Note that the U-statistic is a function of the empir-

[15] Andrews and Kasy (2019) utilize maximum likelihood estimation for their main results, but include a moment-based estimation as a robustness check in their online appendix.

A. Moments		
	Empirical Moment	Simulated Moment
Mean of β_{OLS}	0.169	0.161
Variance of β_{OLS}	0.101	0.124
Mean of ς_{OLS}	0.070	0.061
Variance of ς_{OLS}	0.017	0.004
Mean of β_{IV}	0.373	0.377
Variance of β_{IV}	0.511	0.511
Mean of ς_{IV}	0.248	0.249
Variance of ς_{IV}	0.151	0.134
Correlation between ς_{OLS} and ς_{IV}	0.312	0.312
Meta-OLS estimate	0.666	0.688
Median IV-OLS ratio	2.091	2.089
U-statistic	N.A.	-0.001

B. Parameter Estimates

a_θ	b_θ	μ_γ	σ_γ	$\mu_{\beta_{IV}}$	$\sigma_{\beta_{IV}}$	$\mu_{\tilde{\varsigma}_{OLS}}$	$\sigma_{\tilde{\varsigma}_{OLS}}$	$\mu_{\tilde{\varsigma}_{IV}}$	$\sigma_{\tilde{\varsigma}_{IV}}$	$\rho_{\tilde{\varsigma}}$	q
1.163	1.378	-0.040	0.023	0.351	0.528	-3.175	-1.881	0.891	1.056	0.425	1.294
(--)	(--)	(0.024)	(0.014)	(0.031)	(0.037)	(0.222)	(0.103)	(0.239)	(0.044)	(0.094)	(0.325)

Table 8. Simulated Method of Moments Results
Panel A presents the empirical moments that we seek to match using the simulated method of moments. The simulated moments are those implied by the parameter estimates from panel B. To ensure that the coefficients have a comparable scale, we restrict our estimation to the $1,015$ observations that we can normalize by σ_x/σ_y. In panel B, bootstrapped standard errors are in parentheses; we compute these standard errors using $1,000$ bootstrapped samples. For a small number of bootstrap sub-samples, the shape parameters of the beta distribution do not converge in a reasonable amount of time, leading to implausible standard errors for a_θ and b_θ; this has no effect on the other standard errors.

ical data X and the parameter q; while the U-statistic does not explicitly use simulated data, we regard it as a simulated moment because q is estimated jointly alongside the other parameters. Our moment condition is $G(\cdot) = 0$, and we estimate parameters by solving

$$\widehat{\psi}\left(X, Y_1(\cdot), ..., Y_{10}(\cdot)\right) = \arg\min_{\psi} G(\cdot)' G(\cdot).$$

We compute bootstrapped standard errors by re-sampling the empirical and simulated data one-thousand times each, and utilize the first-order approximation

$$\mathrm{var}\left(\widehat{\psi}\left(X, Y_1(\cdot), ..., Y_{10}(\cdot)\right)\right) \approx \mathrm{var}\left(\widehat{\psi}\left(X_n, Y_1(\cdot), ..., Y_{10}(\cdot)\right)\right) + \mathrm{var}\left(\widehat{\psi}\left(X, Y_{1,n}(\cdot), ..., Y_{10,n}(\cdot)\right)\right),$$

where $n = 1, 2, ..., 1,000$ denotes the nth bootstrap sample. Panel B of Table 8 reports our parameter estimates; we discuss the relevant implications in Section 6.3.2.