Do Better Journals Publish Better Estimates?

David Slichter
Binghamton University (SUNY) and IZA

Nhan Tran
*Binghamton University (SUNY)**

May 2023

Abstract

Are estimates typically closer to the true parameter value when those estimates are published in highly-ranked economics journals? Using 14,387 published estimates from 24 large literatures, we find that, within literatures, the mean and variance of parameter estimates have little or no correlation with journal rank. Therefore, regardless of what the true parameter value is that a literature is attempting to estimate, it cannot be that estimates in higher-ranked journals are on average noticeably closer to it. We discuss possible explanations and implications.

JEL classification: C13, C18, A11
Keywords: Meta-analysis, scientific methods, publication, science of science

*Slichter: slichter@binghamton.edu. Tran: ntran7@binghamton.edu. We are grateful for helpful comments from Greg Caetano, Florian Kuhn, Sulagna Mookerjee, Eric Nielsen, Ronni Pavan, Sol Polachek, Alex Apt Smith, and participants at the Stockman Conference. We are particularly grateful to Ichiro Iwasaki and Xinxin Ma for providing the data used in their meta-analysis of schooling returns.

1 Introduction

Suppose you were interested in learning the value of a particular parameter—say, the effect of a 10% minimum wage increase on teen employment in the United States in the 1980s—but you had no data and could only guess the parameter value by learning a randomly selected published estimate of the parameter. How much happier should you be if the estimate you get to observe was published in a highly-ranked journal?

Specifically, consider what we will call "journal-estimators" of a parameter of interest, which consists of (i) randomly selecting an article about your topic of interest which is published in a journal at a specified rank, then (ii), from that paper, randomly selecting any estimate which would be suitable for a meta-analysis on your topic of interest. To make things concrete, we will focus on two journal estimators: one which samples estimates from papers published at the rank of the *Quarterly Journal of Economics* ("QJE-estimator") and one which samples at the rank of *Industrial and Labor Relations Review* ("ILRR-estimator"). We select the QJE because it is the highest-ranked journal in our data, and ILRR because it is a reputable journal but not the first place where leading economists would send what they consider to be their best work. Note that our empirical analysis focuses solely on journal rank rather than exact journal, so what follows should not be interpreted as a commentary on those two specific journals, but rather as a way to translate our results into relatable units.

Our main goals are to estimate (i) how the mean-squared error (MSE) of the QJE-estimator compares with the MSE of the ILRR-estimator, and (ii) the probability that a randomly chosen QJE-estimate is closer to the true parameter than a randomly chosen ILRR-estimate. Of course, we can only empirically evaluate the estimators under some assumption about the true parameter value. Our baseline estimates assume that the QJE-estimator is unbiased, and therefore interprets any difference in average estimates as an indicator that the ILRR-estimator is biased.

However, it turns out that our findings about the relative accuracy of journal-estimators depend little on the assumed true parameter value, for the simple reason that the distribution of estimates is so similar across journal ranks. In particular, we do not find evidence that either the average estimate or the variance of estimates differs appreciably across journal ranks within literatures. Therefore, the MSE of the QJE-estimator and the ILRR-estimator must be approximately the same, regardless

of the true parameter value.

Our baseline estimates indicate that the MSE of the ILRR-estimator is 1.09 times larger than the MSE of the QJE-estimator—i.e., estimates in the QJE are more accurate, but only incrementally so. As an illustration, if a true parameter were equal to 3, an MSE ratio of 1.1 corresponds with the difference between estimating the parameter to be 3.20 and estimating it to be 3.21. Across nearly all specification and data cleaning choices, we obtain the same qualitative result: There is no meaningful difference in the MSE of the QJE-estimator and ILRR-estimator. Similarly, the probability that a randomly chosen QJE-estimate is closer to the true parameter than a randomly chosen ILRR-estimate is 51% in our baseline estimates—i.e., approximately a coin flip—and lies between 45 and 55% in nearly all alternative specifications.

Next, we consider the implications of our findings. Our analysis can be motivated by three different purposes, and the strength of the conclusions which can be drawn from our findings varies by purpose.

One reason our analysis is useful is that many people are in a position very much like the one described in the opening paragraph of simply wanting to learn a parameter value from published estimates: journalists, policymakers, researchers conducting meta-analyses, researchers looking for a parameter value to calibrate a model, etc. For these audiences, our findings straightforwardly suggest that it isn't worth paying much attention to journal rank.

A second question related to our analysis is whether higher-ranked journals publish better papers. This matters, for example, because researchers are evaluated based on publication records. While our findings are related to one aspect of what might be more desirable about higher-ranked publications, there are other reasons why publications in better-ranked journals might be more valuable contributions: For instance, they might organize and communicate ideas more clearly, they might make theoretical or methodological contributions, and they tend to be published earlier in literatures (when the same estimate makes a greater marginal contribution to knowledge). Therefore our analysis does not establish that journal rank is not an informative signal about *any* aspect of paper quality or that journal rank should not be considered in personnel evaluation.

A third purpose of our analysis is to evaluate the scientific method in economics. Many scientific fields straightforwardly establish their credibility with out-of-sample predictions or technical achievements; whoever can build a hydrogen bomb must

surely understand *something*. However, most economics research does not lend itself to clear out-of-sample predictions or technological feats, and therefore it is difficult to assess whether economics research uses reliable methods.

Our analysis is a test of the scientific method because, if referees and editors can recognize when estimates are likely to be close to the truth or not, then we would expect more accurate estimates to be published in more selective journals. The fact that higher-ranked journals do not publish more accurate estimates therefore suggests the possibility that economists might be focused on aspects of empirical methods which are scientifically unimportant—or, alternatively, that some attributes which are prized in the publication process are beneficial but others are actually harmful. An example of a potentially harmful attribute is that surprising findings could be more likely to publish well, but also surprising for good reason (i.e., wrong).

We cannot definitively resolve this third core question. Nonetheless, because it is such an important question, we offer some speculative assessment. In Section 6, we hypothesize that this is because the publication process screens so strongly for some accuracy-related attributes (like addressing endogeneity) that virtually all published papers perform well on those dimensions, while failing to screen papers on other accuracy-related dimensions (like arbitrary data-cleaning choices and coding errors). The result is that variation among published estimates is driven by dimensions which are not screened for in the publication process and are therefore uncorrelated with journal rank.

We then empirically assess the plausibility of four alternative explanations for our results: (i) higher-ranked papers might be published earlier in literatures, when empirical standards are lower; (ii) literatures might play "follow the leader," where papers in low-ranked journals imitate papers in high-ranked journals; (iii) MSE might underweight the importance of bias when an estimate contributes to a large literature; and (iv) estimates in higher-ranked journals might be studying different populations. We do not find that any of these four mechanisms is important, though the evidence against (iii) and (iv) is less conclusive.

These findings suggest that economists should generally not treat individual empirical papers as definitive. Instead, our results favor humility: Even expert readers have a limited ability to discern between more and less accurate estimates.

Our paper is most closely related to a literature which studies the accuracy of published estimates in economics. The existing literature focuses on issues such as selective publication of significant results (e.g., Doucouliagos, 2005; Doucouliagos

and Stanley, 2009; Havránek, 2013; Demena, 2015; Brodeur et al., 2020), lack of statistical power (e.g., Ioannidis, 2005; Ioannidis et al., 2017), and non-reproducible published results (e.g., Dewald et al., 1986; Chang and Li, 2015). See Ioannidis and Doucouliagos (2013) for a summary of critiques. To our knowledge, no prior paper has systematically studied how parameter estimates vary by economics journal rank. Brembs et al. (2013) review literature from other disciplines about the relationship between journal rank and various measures of scientific quality, and argue against the view that more prestigious journals publish more reliable findings. In medicine, Siontis et al. (2011) compare parameter estimates from experimental trials by journal rank. They find that prestigious journals publish larger effect sizes, with the results driven by anomalously large estimates in small trials published early in literatures, consistent with publication bias.

The rest of the paper proceeds as follows. Section 2 gives a conceptual model to help understand the subsequent analyses. Section 3 describes the data. In Section 4, we measure differences in bias, variance, and MSE by journal rank. In Section 5, we estimate the probability that a randomly selected estimate published in a higher-ranked journal is more accurate than a randomly selected estimate published in a lower-ranked journal. Section 6 considers explanations for our results and Section 7 concludes.

2 Conceptual framework

Let $\widehat{\theta}_i$ denote a published estimate i. We will assume that there exists some underlying true parameter of interest $\theta_{l(i)}$ for the literature l that i is published in. However, each individual paper may have a slightly different claimed estimand, e.g. because it measures a causal effect on a particular subpopulation. Let ν_i denote the difference between i's claimed estimand and $\theta_{l(i)}$—essentially, an external validity adjustment. Additionally, let ξ_i denote the difference between the claimed estimand $\theta_{l(i)} + \nu_i$ and the actual estimand, i.e., the failure of internal validity. Finally, let ζ_i denote sampling error, i.e., the difference between the actual estimate and the parameter that is consistently estimated by study i's research design. Then

$$\widehat{\theta}_i = \theta_{l(i)} + \nu_i + \xi_i + \zeta_i.$$

Estimates within literature l will differ due to ν, ξ, and ζ. It is not clear exactly

which of these is worst. The presence of ν can be innocuous if readers are able to assess issues of external validity, but not if they are not. Papers typically report estimates of the magnitude of ζ (in the form of standard errors), so the magnitude of this form of error is comparatively transparent, but this is counterbalanced by the fact that, unlike ν and ξ, readers cannot use auxiliary information about study design to guess the exact realization of ζ. Furthermore, publication bias might systematically select estimates with particular realizations of ζ. The internal validity bias term ξ_i might be easy or hard for readers to ballpark, depending on context. For the sake of this paper, we will simply treat all three sources of variation as equally undesirable.

The bias of a journal-estimator is defined to be the expected value of $\nu_i + \xi_i + \zeta_i$. Less intuitively, the variance of a journal-estimator stems not only from the fact that different estimates have different sampling error realizations ζ_i, but also from the fact that they have different external and internal validity realizations ν_i and ξ_i. For example, if half of ILRR-estimates suffer from an internal validity error of $\xi = 1$ and the other half suffer from an internal validity error of $\xi = -1$, in what follows, that would be considered a source of variance for the ILRR-estimator but not a source of bias.

3 Data

We collect estimates from 24 literatures. Table 1 lists the literatures.

We obtain our estimates from meta-analyses with systematic literature reviews. The sole exception is that we draw estimates for the employment effect of the minimum wage from two systematic reviews. We collect meta-analyses from three main sources. The first source of data comes from databases from Deakin University.[1] The second source of data is from Charles University and the Czech Science Foundation.[2]. The last source of data is from meta-analysis papers. Appendix A describes the parameters and meta-analyses from which the parameter estimates are drawn. In general, we restrict our sample to papers which have been published since 1990.

To measure the rank of journals, we use the IDEAS/RePEc 10-year recursive discount factor.[3] We take the log of the discount factor to avoid results being driven solely by variation among the very highest-ranked journals. This metric gives an

[1] https://www.deakin.edu.au/business/research/delmar/databases

[2] https://meta-analysis.cz/

[3] https://ideas.repec.org/top/top.journals.rdiscount10.html

Table 1: List of literatures

Labor economics		Health economics		International economics	
Literature 1	Minimum wage and teen employment	Literature 3	Health spending and children's mortality	Literature 7	Remittances and education spending
Literature 2	Return to schooling	Literature 4	Health spending and life expectancy	Literature 16	Trade elasticity
Literature 6	Immigration and natives' wage	Literature 5	Education and mortality		
Literature 21	Elasticity of labor demand				
Literature 23	Wage curve				
Macroeconomics		**Energy economics**		**Financial economics and others**	
Literature 11	Capital and labor substitution	Literature 12	Social cost of carbon	Literature 8	Intergenerational transmission of schooling
Literature 13	Intertemporal substitution in consumption	Literature 15	Elasticity of water demand	Literature 9	Tuition fee and demand for higher education
Literature 14	Skilled and unskilled labor substitution	Literature 20	Price elasticity of gas	Literature 10	Individual discount rate
Literature 17	Sensitivity of consumption to income	Literature 22	Income elasticity of gas	Literature 18	Student's employment and education
				Literature 19	Price elasticity of beer demand
				Literature 24	Elasticity of taxable income

intuitive distance between journal tiers. For instance, in our data, this gives a value of 2.63 to the *Quarterly Journal of Economics*, a value of 1.16 to the *Journal of Labor Economics*, a value of $-.09$ to *Labour Economics*, and a value of $-.83$ to *Industrial and Labor Relations Review*. That is, typical classifications of tiers of journal (top 5, top field, second field, and so on) generally correspond to intervals in our journal rank of something like one unit, perhaps slightly more.

In our preferred specification, we also bottom-code journal ranks so that all journals outside of the top 500 journals are assigned the same journal rank as the 500th ranked journal. This avoids identifying our parameters of interest from variation among very low-ranked journals, which is useful for two reasons. First, presumably economists do not actually perceive ranking differences among journals they have never heard of, and we judge that very few journals outside the top 500 are well-known.[4] Second, this limits the role of extrapolation in fitting values at the rank of the QJE and ILRR using a linear conditional expectation function. In alternative specifications, we either do not bottom-code journal ranking at all, or we bottom-code more aggressively by bottom-coding beyond the 300th journal instead of the 500th.

Next, we merge our data based on journal name. We eliminate a small number of observations that we cannot match to the journal ranking list, predominantly because they were published in non-economics journals. Our final sample has 14,387 estimates drawn from 871 papers. The minima in any single literature are 92 estimates and 6 papers.

[4]We chose the ranking cutoff based on our own views combined with feedback from three research active PhD economists who were not involved in this paper, to whom we provided the journal ranking and asked, "What would you say is a reasonable cutoff rank such that you would say, 'below this, I've never heard of basically any journals?'"

Finally, to allow for comparisons across literatures, we normalize parameter estimates to have weighted mean of 0 and weighted standard deviation of 1 within each literature, where the weights are the inverse of the number of estimates in each paper.

4 MSE differences by journal rank

In this section, we estimate differences in MSE of the QJE-estimator and ILRR-estimator. The MSE of an estimator is equal to the sum of its variance and the square of its bias. We use the following procedures to estimate these parameters.

4.1 Estimation of squared bias

To study the bias of journal-estimators, we first study how average parameter estimates vary by journal rank within a literature. We estimate the following regression:

$$\widehat{\theta_i} = \sum_{l=1}^{24} 1\{\text{lit}_i = l\}\alpha_l + \sum_{l=1}^{24} 1\{\text{lit}_i = l\}\beta_l \text{rank}_i + \sum_{l=1}^{24} 1\{\text{lit}_i = l\}\eta_l \text{year}_i + \epsilon_i, \qquad (1)$$

where lit_i denotes the literature that estimate i is published in, rank_i denotes the rank of the journal that i was published in (as defined in Section 3), and year_i denotes the median year of data used for estimate i demeaned by literature. In this regression, each estimate i is weighted by the inverse of the number of estimates contained in the same paper as i, and standard errors are clustered by paper.

The parameter of interest in this regression is β_l, which captures how the average estimate varies by journal rank. The purpose of controlling for year of data is that this helps ensure that the parameter being estimated in each literature, θ_l, is adjusted to be as comparable as possible between estimates i. We also implement specifications which do not control for year.

Estimates of β_l for each literature l (controlling for year) are reported in Table 2 in Appendix C. The distribution of estimates is shown in Figure 4 in Appendix C.

To obtain the bias of journal-estimators, we would need to know both the average estimate at a given journal rank *and* the true θ_l for each literature. This requires an assumption about the true θ_l. The assumption we will impose is that the QJE-estimator is unbiased in every literature, i.e., θ_l is equal to the average $\widehat{\theta_i}$ conditional

on $\text{rank}_i = 2.63$. Under this assumption, the squared bias of the QJE-estimator is 0 in every literature, and the squared bias of the ILRR-estimator is $(3.46\beta_l)^2$, where 3.46 is the ranking difference between the QJE (2.63) and ILRR ($-.83$).

Now, β_l is estimated rather than known directly, and sampling error will tend to inflate the variance of the estimates $\widehat{\beta}_l$. This poses a problem for us because, to estimate the MSE of journal-estimators, we need the average across literatures of β_l^2, and sampling error in our estimates of β_l will create an upwards bias in our estimate of $E(\beta_l^2)$.

We therefore model the distribution of true β_l. Noise limits our ability to discern the exact shape of this true distribution, but the estimated values of β_l roughly resemble a bell shape (see Figure 4 in Appendix C) and are centered on approximately 0, so we model the true values of β_l as having a normal distribution with mean zero. We then attempt to estimate the parameters of this normal distribution using two different approaches.

Maximum likelihood estimation Let $\widehat{\beta}_l$ denote the value of β_l obtained from estimating equation 1. Define the sampling error in this estimate to be

$$\iota_l := \beta_l - \widehat{\beta}_l.$$

Based on the Central Limit Theorem, we assume that ι_l is (i) normally distributed with mean zero and with standard deviation equal to the standard error of $\widehat{\beta}_l$, and (ii) is independent of β_l.

Let σ_β denote the standard deviation of true β_l across literatures and let s_l denote the standard error of the estimate of $\widehat{\beta}_l$. Then, since $\widehat{\beta}_l = \beta_l + \iota_l$, we have that each value of $\widehat{\beta}_l$ is drawn from a normal distribution with mean 0 and standard deviation of $\sigma_\beta + s_l$. Using a dataset containing $\widehat{\beta}_l$ and s_l for each literature, we can then calculate the likelihood for each $\widehat{\beta}_l$ given a value of σ_β by evaluating the pdf of a normal distribution with mean zero and standard deviation $\sigma_\beta + s_l$.

Subtracting variances An alternative approach is to observe that, because ι_l is uncorrelated with β_l, then

$$var(\widehat{\beta}_l) = var(\beta_l) + var(\iota_l),$$

which gives that

$$var(\beta_l) = var(\widehat{\beta}_l) - var(\iota_l).$$

We can approximate $var(\widehat{\beta}_l)$ with the sample variance of $\widehat{\beta}_l$, and $var(\iota_l)$ with the sample average of s_l.

We prefer the MLE approach. First, it has an efficiency advantage, since it in effect treats values of $\widehat{\beta}_l$ as more informative when those estimates are more precise.[5] Second, the method of subtracting variances has the unfortunate property that it can in principle result in a negative estimated variance of β_l if values of $\widehat{\beta}_l$ from different literatures are coincidentally similar. This is not actually the case for our preferred estimates, but it is for some alternate specifications, for which we instead take as our estimate that $var(\beta_l) = 0$.

Finally, given an estimated distribution of β_l, we can compute the squared bias of the ILRR-estimator, $E[(3.46\beta_l)^2]$.

4.2 Estimation of variance

Next, we study the variance of journal-estimators. First, we obtain estimated residuals $\widehat{\epsilon}_i$ from Equation 1 and square them. Then we estimate the following regression:

$$\widehat{\epsilon}_i^2 = \sum_{l=1}^{24} 1\{\text{lit}_i = l\}\gamma_l + \sum_{l=1}^{24} 1\{\text{lit}_i = l\}\omega_l \text{rank}_i + \sum_{l=1}^{24} 1\{\text{lit}_i = l\}\lambda_l \text{year}_i + \tau_i. \qquad (2)$$

We evaluate the variance at a journal with rank k using the average coefficients on literature dummies γ and the average coefficients on the interaction ω term. That is, the variance of journal-estimator with rank k is estimated to be $\frac{1}{24}\sum_{l=1}^{24}(\gamma_l + \omega_l k)$. When year_i is included as a control, because it is demeaned by literature, this value should instead be interpreted as the variance of a journal-estimator with rank k using data from the average data year in the literature.

Finally, we estimate the overall MSE at journal rank k by adding the estimated squared bias to the estimated variance.

[5] Note, however, that β_l might not be independent of s_l. In that case, the variance of β_l weighted by s_l might differ from the unweighted variance of β_l, which is our parameter of interest. Therefore, if there is a strong enough relationship between β_l and s_l, the subtracting variances approach might be favored.

4.3 MSE results across specifications

Combining our estimates of squared bias and variance in our preferred specification—in which we control for year of data, use MLE, and assume that the QJE-estimator is unbiased—we obtain an estimate that the MSE of the QJE-estimator is 0.781 and the MSE of the ILRR-estimator is 0.854. That is, our preferred estimate is that the QJE-estimator is slightly more accurate than the ILRR-estimator, with the ILRR-estimator having MSE which is 1.09 times larger.

In Figure 1, we report the robustness of this conclusion to alternative ways of handling the data. Specifically, we report the results which would be obtained under every possible combination of the following choices:

- **Method for obtaining $var(\beta_l)$:** Either use the MLE approach or the subtracting variances approach described above. Specifications using MLE have a dark circle shaded in the row labeled "MLE."

- **True parameter value:** We either assume that the QJE-estimator is unbiased and that ILRR has some bias $3.46\beta_l$, or that the true parameter lies halfway between the expected values of the QJE-estimator and ILRR-estimator, such that each has bias $3.46\beta_l/2$. Specifications assuming that the QJE-estimator is unbiased have a dark circle in the "Bias at ILRR" row.

- **Year controls:** Either control for year_i or do not. Specifications including this control have a dark circle for "With year."

- **Order of publication:** Either control for the order in which a paper is published in a literature (e.g., estimates from the second paper published get a value 2) or do not.

- **Shrinkage:** Either apply a shrinkage adjustment to $\widehat{\beta}_l$ before estimating $\widehat{\epsilon}_i$ or do not. See Appendix B for details on how the shrinkage adjustment is performed.

- **Outliers:** Our data contain a substantial number of outliers which may affect inference. We consider three approaches. Our baseline, which we refer to as "with outliers," does not attempt to limit the effects of outliers. A second approach, which we call "percentile," converts every coefficient estimate into a number between 0 to 1 equal to the fraction of estimates in that literature

which are smaller than the given paper's estimate. This approach limits the importance of outliers without deleting them, and also distorts the scale to magnify the importance of small differences in parameter estimates in proportion to how many other parameter estimates are nearby. A third approach, which we call "without outliers," first deletes any observations which are more than 3 standard deviations away from the literature average, then recalculates the standard deviation of estimates within a literature in the trimmed sample and deletes any observations more than 4 standard deviations away from the average of remaining observations.

- **Cutoff:** We consider specification without bottom-coding of journal rank ("no cutoff") and with bottom-coding at the rank of the 300th and 500th-ranked journals.

Figure 1 reports the results in the form of a specification curve (Simonsohn et al., 2020). The top half reports point estimates and 95% confidence intervals for the ratio of the MSE of the ILRR-estimator to the MSE of the QJE-estimator. That is, a value of 1 implies that the MSE of the two journal-estimators is the same. The estimates from each specification are reported in order from the smallest to largest point estimate of this ratio. The bottom half of Figure 1 reports the exact choices made in each specification.

Confidence intervals are estimated using a block bootstrap, where blocks are literatures. We truncate the top end of confidence intervals at 2 to enhance readability. The rightmost point estimate lies above 2.

A few things are worth noting. First, the mean, median, and mode of estimates are all very close to 1; so, it could be said that a "typical specification" finds that the MSE of the QJE-estimator is about the same as that of the ILRR-estimator.

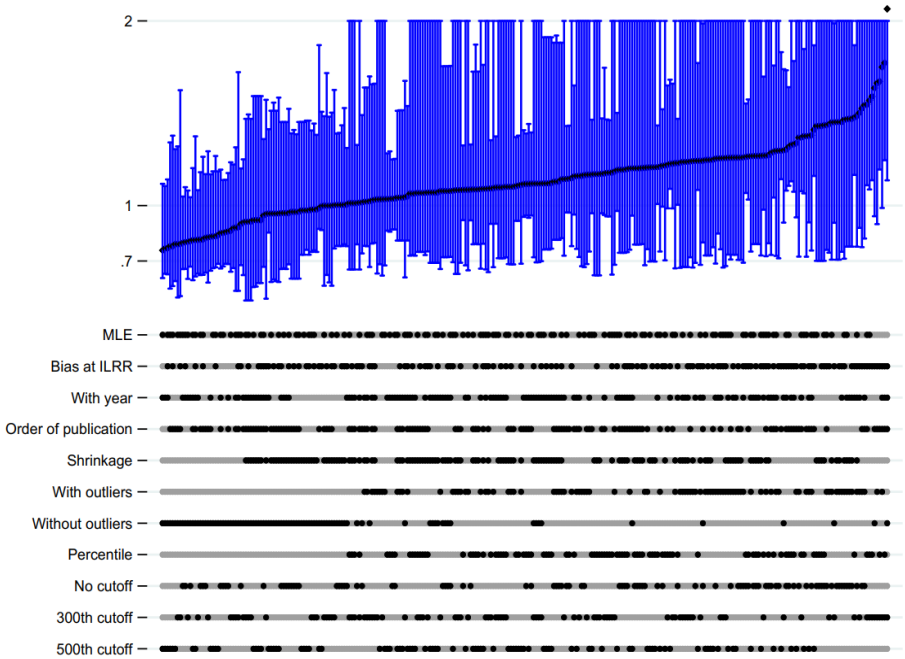
Second, the choice of target parameter makes little difference: While the estimates which assume that the QJE-estimator is unbiased are somewhat to the right of those which do not, the estimates are not strongly sorted on this aspect of the specification. In many cases, the two assumptions deliver identical estimates, since the variance of β_l is estimated to be 0.

Third, estimates that control for order of publication are not systematically larger (or smaller) than those that do not. We will return to this observation in Section 6.

Fourth, the most consequential aspect of the specification choice seems to be how outliers are handled. Fully dropping outliers tends to produce estimates favorable

to the ILRR-estimator. That is, when the most extreme estimates are deleted, the ILRR-estimator is generally estimated to have lower MSE. By contrast, both the baseline and percentile approaches produces estimates which favor the QJE-estimator.

Figure 1: The ratio of the MSE of ILRR-estimator to QJE-estimator



Notes: This figure reports the results for 288 different specifications. In the top half, we report the point estimate along with 95% confidence intervals for the ratio of the MSE of ILRR estimator to QJE estimator. We truncate the confidence interval at 2 for readability. In the bottom half, we present the choices in each specification.

5 Head-to-head comparisons of estimates by journal rank

Next, we estimate the probability that a randomly chosen QJE-estimate is closer to the true parameter value than a randomly chosen ILRR-estimate. This provides an alternate way of assessing the relative accuracy of the QJE-estimator and ILRR-estimator.

In our baseline analysis, we follow the same assumptions and modeling choices as those made in our baseline analysis of Section 4. That is, we assume that the

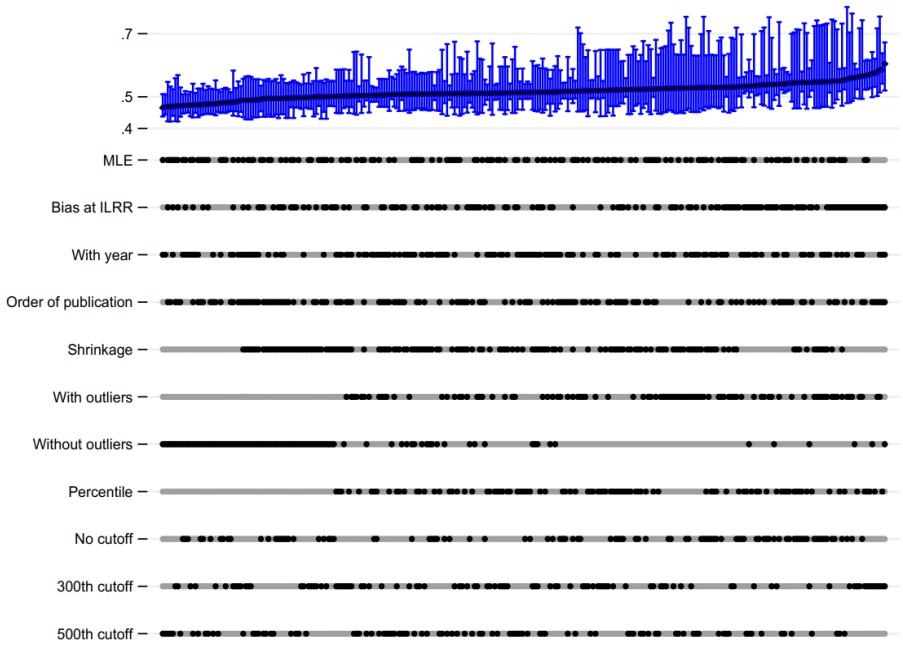
QJE-estimator is unbiased, we assume β_l is normally distributed and use MLE to estimate the parameters of the distribution, and we control for data year but not order of publication. In addition, we make a distributional assumption over ϵ: Based on plots of the estimated residuals $\widehat{\epsilon}$, we assume that ϵ follows a Laplace distribution.

In order to compare a randomly selected estimate in QJE versus in ILRR, we perform a simulation which mimics random realizations of the QJE-estimator and ILRR-estimator. We simulate the ILRR-estimate by drawing a random realization of β_l from the estimated distribution of β_l (again, assumed to be normal) and adding this to a random realization of ϵ, which is drawn from a Laplace distribution with mean zero and variance equal to the estimated variance of ϵ for the ILRR-estimator. For the QJE-estimator, we assume the bias term is 0 and add this to a draw from a Laplace distribution with mean zero and variance equal to the estimated variance of ϵ for the QJE-estimator. We then compare the absolute value of the simulated values; whichever is smaller is the winning estimate. We obtain win probabilities by running one million simulations and counting the fraction of wins for the QJE-estimator.

For additional specifications where we assume that the QJE-estimator and ILRR-estimator are equally biased, we simulate a bias term for each journal-estimator by drawing a value of β_l from a normal distribution and then assigning positive one-half times that value as the bias term for the ILRR-estimate, and negative one-half times that value as the bias term for the QJE-estimate.

Results across a range of specifications are shown in Figure 2. Our preferred specification estimates a win probability for the QJE-estimator of 51%. Regardless of the details of the specification, the win probability for the QJE-estimator never deviates far from 50%—i.e., the QJE-estimator is no better than the ILRR-estimator—and in many specifications the winning probability is almost exactly 50%. Unsurprisingly, specification choices which favored the QJE-estimator in Figure 1 also favor the QJE-estimator in Figure 2.

Figure 2: Winning probability of QJE-estimator



Notes: This figure reports the probability that randomly selected QJE-estimates are closer to the true parameter than randomly selected ILRR-estimates for 288 different specifications. In the top half, we report the point estimate along with 95% confidence intervals. In the bottom half, we present the choices in each specification.

In short, the results of Sections 4 and 5 suggest that, by two different metrics and across a wide variety of specifications, estimates published in higher-ranked economics journals are no closer to the true literature parameter θ_l than estimates published in lower-ranked journals.

6 Possible explanations

We now return to some of the broader questions which our analysis is related to.

The first question, which is relevant for audiences such as journalists, policymakers, and people calibrating models, was whether estimates published in high-ranked journals are likely to be closer to the truth than estimates published in lower-ranked journals. Based on our results, the answer is that they are not.

The second question, which is relevant for people involved in hiring or promotion, was whether papers published in higher-ranked journals are more scientifically valuable than paper published in lower-ranked journals. We cannot fully answer this

question because we only assess one dimension of scientific value. In addition to producing accurate estimates, for instance, papers can contribute to the scientific literature by making theoretical contributions, by making methodological contributions, or simply by explaining ideas clearly. Lastly, it is possible that estimates in higher-ranked journals make a greater contribution to knowledge even if they are not more accurate, as we discuss below. To assess whether papers in higher-ranked journals make a greater scientific contribution, we would need to measure every relevant dimension of scientific contribution, which is beyond the scope of this paper. However, in Appendix F, we find evidence supporting at least one sense in which papers in higher-ranked journals make a greater contribution, which is that they are published earlier in a literature, when a marginal estimate makes a greater contribution to knowledge.

The third question, which is relevant for anyone interested in the scientific credibility of economics, is whether expert referees and editors are able to detect which estimates are likely to be closer to the truth. While referees and editors value several attributes of papers, it is surely the case that, all else equal, referees and editors have a preference for papers which they believe do a better job of estimating the parameter of interest. Indeed, based on personal experience and conversations, we believe that referees' and editors' views about the appropriateness of methods and credibility of estimates is one of the most central criteria for evaluating papers.

If referees and editors were effective in this task, we would expect estimates published in higher-ranked journals to be more accurate. Therefore it is disturbing that higher-ranked journals do not publish more accurate estimates.

We devote the remainder of this section to discussing possible explanations for our results and what they imply about the scientific method in economics.

6.1 Possible explanation #1: Weak selection on accuracy

The most straightforward reading of our results is that, whatever dimensions explain differences in accuracy across published papers, the publication process simply does not select for them strongly enough for us to detect an appreciable relationship between journal rank and accuracy.

The theoretical case for this story is that, once we condition on the basic level of competence required to publish an estimate, (i) most attributes related to estimate accuracy are unobserved by referees and editors, and (ii) the publication process

anyhow does not solely select for accuracy.

Here are some reasons why the criteria used to determine publication might not be strongly related to the accuracy of estimates:

1. Papers are published not only for their parameter estimates, but also for theoretical and methodological contributions and for communicating ideas clearly. Publication outcomes might also depend in part on authors' reputations (Huber et al., 2022) or ability to anticipate the tastes of specific editors and referees.

2. Publication outcomes involve some element of chance conditional on a paper's attributes.

3. Typically the most important consideration related to accuracy is whether the paper has a credible approach for addressing endogeneity. Yet Young (2022) finds that, in a sample of papers published in leading journals using IVs, exogeneity of OLS is typically not statistically rejected, suggesting that endogeneity may be of limited importance in many applications—or at least, that it is small relative to sampling error. Furthermore, even papers published in low-ranked journals are expected to try to address endogeneity, so the difference in omitted variables bias between papers in high-ranked and low-ranked journals is likely to be smaller than the difference between IV and OLS estimands.

At the same time, there might be important sources of variation in estimates which are not typically important for a publication decision:

1. Sampling error is an important source of variation in estimates. Ioannidis et al. (2017) document that economics papers chronically lack statistical power, i.e., estimates are noisy. While referees and editors likely have a preference for papers with smaller standard errors, it could be that this preference is not very strong.

2. Most research involves making a series of minor choices where more than one option is defensible—the proverbial "garden of forking paths" (Gelman and Loken, 2013). Each choice may individually make little difference, but the cumulative effect of many choices can be substantial: Huntington-Klein et al. (2021) find that economics researchers given the same data and research design produce estimates which vary widely relative to the uncertainty implied by the

standard errors. The publication process likely selects little on the basis of these defensible choices.

3. Coding errors might be common. Authors who attempt to systematically replicate many published findings have had low success rates (e.g., Dewald et al., 1986; Chang and Li, 2015).

4. Estimates vary with data sources. It is common that descriptive statistics vary across surveys which purport to cover the same population, so other parameters probably also vary due to differences in sampling procedures, variable definitions, or measurement error.

5. External validity is difficult to evaluate and may be threatened for reasons which are difficult for referees and editors to recognize or assess.

6. Methodological issues may be unknown to referees and editors. For instance, for years, referees were unaware of the potential for negative weights in panel designs or IV models with controls.

This collection of arguments does not necessarily imply that referees and editors are making mistakes. Instead, there might simply be limits to what any reader can know about the accuracy of a given estimate.

This is also not an entirely nihilist explanation, in the sense that parameter estimates are not completely unrelated to the truth. The most extreme forms of nihilism do not fit the data; for instance, parameter estimates vary across literatures, meaning that economists must be producing estimates which in *some* way correspond to the question being asked. What this line of argument suggests instead is that there might be certain basic aspects of estimating a parameter which virtually all papers in a given literature get right, and variation in estimates beyond that is primarily due to factors which economists either cannot judge or are not even aware of.

If true, this line of argument suggests that the best way to learn parameters is to produce as many estimates as possible making independent choices (e.g., of methodologies, data sources, data cleaning procedures, and populations studied).

6.2 Possible explanation #2: Preference for surprising results

It is difficult to publish a paper in a leading journal finding that water is wet. But a paper which argues that water isn't wet might be interesting enough to have a shot, provided it made a persuasive case. If leading journals are more likely to publish findings which were *a priori* less likely to be true, then the findings there might still be less likely to be true *a posteriori*, even if the standards of evidence are higher at higher-ranked journals.

Unfortunately, it is difficult to collect empirical evidence related to this, because it is difficult to quantify what estimates would be considered "surprising". However, the empirical exercise described in Section 6.5 is related, and does not support an important role for this hypothesis.

6.3 Possible explanation #3: Different estimands

Returning to the conceptual framework of Section 2, estimates vary within a literature both because the estimand varies (ν_i) as well as because of the combined effects of internal validity and sampling error ($\xi_i + \zeta_i$).

One possible explanation of our results is that the QJE-estimator might provide substantially more accurate estimates of the paper's claimed estimand $\theta_{l(i)} + \nu_i$ but that higher-ranked journals are characterized by a counterbalancing substantially greater variation in ν_i. In this case, the publication process can be said in some sense to have sorted more accurate estimates to higher-ranked journals. However, this requires greater variation in claimed estimands to coincidentally approximately cancel out the decreased variation in $\xi_i + \zeta_i$. That is, if we accept that higher-ranked journals publish estimates with greater internal validity, we must also accept that they publish estimates where the estimand is harder to generalize to the sort of other contexts considered in the literature; and, the larger we think their internal validity advantage is, the larger must be the external validity disadvantage.

As discussed in Section 2, external validity issues may be more or less innocuous than internal validity or sampling error. On the one hand, populations may differ in ways which are clear to readers and have an obvious connection with parameters of interest. On the other hand, populations being studied can easily differ from populations of interest in ways which are difficult for readers to recognize or un-

derstand. Furthermore, if the parameter of interest is unknown, then how it varies across populations is likely to be unknown too.

Empirically, it is difficult to disentangle ν_i from $\xi_i + \zeta_i$, so we cannot fully evaluate this explanation. To some extent, we attempt to limit the role of ν by restricting to parameter estimates which someone thought were comparable enough to include in the same meta-analysis. Controlling for data year is also an attempt to limit the role of ν, but including this control does not generally favor the MSE of the QJE-estimator relative to the ILRR-estimator. This is not conclusive but it is most consistent with the view that the variance of ν is not different for the QJE-estimator and ILRR-estimator.

We can also attempt to disentangle the sampling error ζ_i from the sum of external and internal validity factors $\nu_i + \xi_i$ by looking at standard errors. This assumes that standard errors accurately reflect the role of sampling error—an assumption which would not hold, for instance, under the explanation in Section 6.2, where the realization of sampling error ζ_i would affect publication outcomes.

In Appendix D, we document differences in standard errors by journal rank. In short, the average estimate across specifications is that standard errors are slightly smaller for QJE-estimates, but only incrementally so in most specifications.

In short, we do not fully observe ν_i, ξ_i, and ζ_i, so we cannot draw strong conclusions. However, on the dimensions of these variables that we can observe, there are not appreciable differences by journal rank. An argument that better journals publish appreciably better estimates and that this is masked by differences in estimands must therefore explain (i) why better journals would have so much more variation in the external validity factor to fully offset advantages in internal validity and sampling error, (ii) why these advantages in internal validity and sampling error would not show up in the analyses above, and (iii) why variation in the external validity factor is more desirable or transparent than variation stemming from the other factors. We do not rule out that such an explanation exists, but we believe a simpler explanation is that estimates in higher- and lower-ranked journals have similar realizations of all three error components ν_i, ξ_i, and ζ_i, with no coincidental offsetting effects required.

6.4 Possible explanation #4: Order of publication

Another possibility is that the order of publication matters. In Appendix F, we estimate the relationship between publication order and journal rank and find bor-

derline significant evidence that higher-ranked publications are on average published earlier within a literature. Note that the literatures we study are large and in many cases date back decades prior to the start of our sample, which likely reduces the strength of this relationship.

It is important to distinguish between two stories. The first story is that earlier publications within a literature make a greater marginal contribution to knowledge at the time that they are published, which indicates that papers in higher-ranked journals are making a greater scientific contribution on average. However, this does not establish that referees and editors are able to evaluate the accuracy of estimates—only the novelty of the question.

A second story is that standards might be lower earlier in a literature. In this story, higher-ranked publications face higher standards because they are better-published but lower standards because they are published earlier. If this were the case, higher-ranked publications would still be more accurate than lower-ranked publications which are published at the same time. But this is not empirically supported: The specification curves in Sections 4 and 5 show that controlling for the order of publication within a literature does not produce appreciably different results.

6.5 Possible explanation #5: Follow the leader

Another possibility is that higher-ranked journals publish papers which introduce new methods or data sources to the literature, and lower-ranked journals publish papers with similar estimates because those papers subsequently adopt the same methods and data sources.

We can empirically test for the importance of "following the leader" as follows. To assess a paper's impact on a literature, we can compare the average parameter estimate published prior to that paper (call this $\bar{\theta}_i^{pre}$) to the average estimate published after that paper ($\bar{\theta}_i^{post}$). If the paper is influential, then, the higher estimate i is, then the higher the estimates published after i will be relative to those published before. The "follow the leader" hypothesis implies that estimates in high-ranked journals should be particularly influential.

We estimate the following regression:

$$\bar{\theta}_i^{post} - \bar{\theta}_i^{pre} = \psi + \pi\widehat{\theta}_i * \text{rank}_i + \delta\widehat{\theta}_i + \rho\text{rank}_i + \upsilon_i.$$

The coefficient of interest is π, which will take a positive value if results from high-ranked publications have a special influence on the subsequent literature.

Clustering by literature, we estimate π to be .016, with a standard error of .010. That is, our point estimate suggests that high-ranked publications might be more influential for subsequent estimates. However, the effect is not significant at the .05 level, and is not large enough to explain a significant fraction of variation in estimates.

Furthermore, for the "follow the leader" story to hold, it must be that papers in lower-ranked journals are more likely to follow. In Appendix E, we additionally consider whether influential papers have a greater influence on subsequent parameter estimates in high- vs. low-ranked journals. We do not find evidence that they do; our estimate points in the opposite direction but is not statistically significant.

The analysis in Appendix E also suggests that the explanation outlined in Section 6.2 may not be compelling: If surprising findings are more likely to be published in higher-ranked journals, then we would expect findings which differ from the previous literature to be published in higher-ranked journals. Therefore, for instance, the publication of an anomalously low estimate in a high-ranked (and therefore high-visibility) journal should increase the rank of journal that subsequent anomalously *high* estimates would publish in.

A related story which this empirical evidence does not address is as follows. Related to the line of argument in Section 6.6, when averaging estimates within a literature, it is better to have estimates which are independent. Therefore, papers which use unusual research designs and data may be particularly valuable. Estimates published in higher-ranked journals might be more likely to have these features. Because these estimates are independent, they may be particularly likely to differ from other estimates, therefore generating the effect that we observe.

The distinction between this argument and "follow the leader" is that this does not require that subsequent papers in lower-ranked journals adopt the same methods or data sources. Instead, it could simply be that highly-ranked journals publish papers which use methods and data that are both independent of past estimates and difficult enough to replicate (e.g., because they use experimental or confidential data) that they have little influence on the set of estimates produced later.

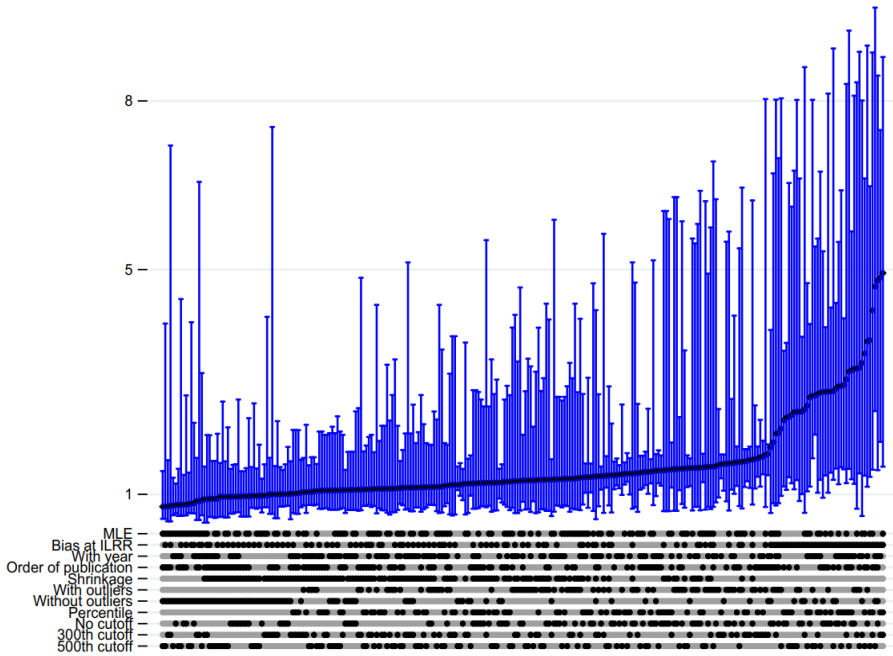
6.6 Possible explanation #6: Averaging out variance

A final possibility is that minimizing bias of journal-estimators is more important than minimizing variance because estimates are viewed in the wider context of a literature.

If we average, say, 10 independent estimates produced by a journal-estimator, the bias component of the MSE will not go away, but the variance component will be cut by a factor of 10. Therefore, for large literatures, the relative importance of minimizing bias vs. variance of journal-estimators might be different than in our main estimates.

In Figure 3 below, we replicate our main results of Figure 1, but dividing the variance of each journal estimator by 10 to estimate the MSE that would be obtained by averaging 10 independent estimates produced by a journal-estimator.

Figure 3: MSE ratio (ILRR/QJE), average of 10 independent estimates



Notes: This figure reports the ratio of the MSE of ILRR estimator to QJE estimator, average of 10 independent estimates for 288 different specifications. In the top half, we report the point estimate along with 95% confidence intervals. In the bottom half, we present the choices in each specification.

The figure shows that there exist some specifications where 10 independent QJE-estimates are substantially better than 10 independent ILRR-estimates. As one

would expect, these specifications all rely on the assumption that the QJE-estimator is unbiased, which is naturally favorable to the QJE-estimator. However, even among specifications which make this assumption, the substantial majority show little or no MSE advantage for the averaged QJE-estimator. This reflects that most specifications do not find a non-negligible bias difference between the QJE-estimator and ILRR-estimator.

7 Conclusion

We study the distributions of parameter estimates published within the same literatures in high-ranked vs. low-ranked economics journals. Our main conclusion is that estimates published in high-ranked journals are not appreciably closer to the true literature-wide parameter than estimates in low-ranked journals.

We offer an explanation that the publication process selects for some aspects of accuracy so strongly that there is little different on those dimensions between what is publishable in top journals as opposed to publishable anywhere, while barely selecting at all on other aspects of accuracy which then wind up quantitatively explaining most of the variation in estimates. We cannot rule out the possibility that there is some alternative explanation. However, we investigate a variety of alternative explanations and do not find clear evidence supporting them.

References

[1] Askarov, Z., & Doucouliagos, H. (2020). A meta-analysis of the effects of remittances on household education expenditure. *World Development*, 129, 104860.

[2] Bajzik, J., Havranek, T., Irsova, Z., & Schwarz, J. (2020). Estimating the Armington elasticity: The importance of study design and publication bias. *Journal of International Economics*, 127, 103383.

[3] Brembs, B., Button, K., & Munafo, M. (2013). Deep impact: Unintended consequences of journal rank. *Frontiers in Human Neuroscience*, 7: 291.

[4] Brodeur, A., Cook, N., & Heyes, A. (2020). Methods matter: P-hacking and publication bias in causal analysis in economics. *American Economic Review*, 110(11), 3634-3660.

[5] Chang, A. C., & Li, P. (2015). Is economics research replicable? Sixty published papers from thirteen journals say 'usually not'. Working paper.

[6] Dahl, C. A. (2012). Measuring global gasoline and diesel price and income elasticities. *Energy Policy*, 41, 2-13.

[7] Demena, B. A. (2015). Publication bias in FDI spillovers in developing countries: a meta-regression analysis. *Applied Economics Letters*, 22(14), 1170-1174.

[8] Dewald, W. G., Thursby, J. G., & Anderson, R. G. (1986). Replication in empirical economics: The journal of money, credit and banking project. *American Economic Review*, 587-603.

[9] Doucouliagos, C. (2005). Publication bias in the economic freedom and economic growth literature. *Journal of Economic Surveys*, 19(3), 367-387.

[10] Doucouliagos, H., & Stanley, T. D. (2009). Publication selection bias in minimum-wage research? A meta-regression analysis. *British Journal of Industrial Relations*, 47(2), 406-428.

[11] Fleury, N., & Gilles, F. (2018). The intergenerational transmission of education. A meta-regression analysis. *Education Economics*, 26(6), 557-573.

[12] Gallet, C. A., & Doucouliagos, H. (2017). The impact of healthcare spending on health outcomes: A meta-regression analysis. *Social Science & Medicine*, 179, 9-17.

[13] Gechert, S., Havranek, T., Irsova, Z., & Kolcunova, D. (2022). Measuring capital-labor substitution: The importance of method choices and publication bias. *Review of Economic Dynamics*, 45, 55-82.

[14] Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. Department of Statistics, Columbia University, 348, 1-17.

[15] Havránek, T. (2013). Publication bias in measuring intertemporal substitution (No. 15/2013). IES Working Paper.

[16] Havránek, T. (2015). Measuring intertemporal substitution: The importance of method choices and selective reporting. *Journal of the European Economic Association*, 13(6), 1180-1204.

[17] Havranek, T., Irsova, Z., & Janda, K. (2012). Demand for gasoline is more price-inelastic than commonly thought. *Energy Economics*, 34(1), 201-207.

[18] Havranek, T., Irsova, Z., Janda, K., & Zilberman, D. (2015). Selective reporting and the social cost of carbon. *Energy Economics*, 51, 394-406.

[19] Havranek, T., Irsova, Z., Laslopova, L., & Zeynalova, O. (2022). Publication and Attenuation Biases in Measuring Skill Substitution. *The Review of Economics and Statistics*, forthcoming.

[20] Havranek, T., Irsova, Z., Vlach, T. (2018). Measuring the income elasticity of water demand: the importance of publication and endogeneity biases. *Land Economics*, 94(2), 259-283.

[21] Havranek, T., Irsova, Z., & Zeynalova, O. (2018). Tuition fees and university enrolment: a meta-regression analysis. *Oxford Bulletin of Economics and Statistics*, 80(6), 1145-1184.

[22] Havranek, T., & Sokolova, A. (2020). Do consumers really follow a rule of thumb? Three thousand estimates from 144 studies say "probably not". *Review of Economic Dynamics*, 35, 97-122.

[23] Huber, J., Inoua, S., Kerschbamer, R., & Smith, V. L. (2022). Nobel and novice: Author prominence affects peer review. *Proceedings of the National Academy of Sciences*, 119(41), e2205779119.

[24] Huntington-Klein, N., Arenas, A., Beam, E., Bertoni, M., Bloem, J. R., Burli, P., Chen, N., Grieco, P., Ekpe, G., Pugatch, T., Saavedra, M., & Stopnitzky, Y. (2021). The influence of hidden researcher decisions in applied microeconomics. *Economic Inquiry*, 59(3), 944-960.

[25] Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, 2(8), e124.

[26] Ioannidis, J., & Doucouliagos, C. (2013). What's to know about the credibility of empirical economics?. *Journal of Economic Surveys*, 27(5), 997-1004.

[27] Ioannidis, J. P., Stanley, T. D., & Doucouliagos, H. (2017). The power of bias in economics research. *Economic Journal*, 127(605), F236-F265.

[28] Kroupova, K., Havranek, T., & Irsova, Z. (2021). Student Employment and Education: A Meta-Analysis (No. 28/2021). IES Working Paper.

[29] Lichter, A., Peichl, A., & Siegloch, S. (2015). The own-wage elasticity of labor demand: A meta-regression analysis. *European Economic Review*, 80, 94-119.

[30] Longhi, S., Nijkamp, P., Poot, J. (2008). Meta-analysis of empirical evidence on the labour market impacts of immigration. Working paper.

[31] Ma, X., Iwasaki, I. (2021). Return to schooling in China: A large meta-analysis. *Education Economics*, 29(4), 379-410.

[32] Matousek, J., Havranek, T., & Irsova, Z. (2022). Individual discount rates: a meta-analysis of experimental evidence. *Experimental Economics*, 25(1), 318-358.

[33] Neisser, C. (2021). The elasticity of taxable income: A meta-regression analysis. *The Economic Journal*, 131(640), 3365-3391.

[34] Nelson, J. P. (2013). Meta-analysis of alcohol price and income elasticities—with corrections for publication bias. *Health Economics Review*, 3(1), 1-10.

[35] Nijkamp, P., & Poot, J. (2005). The last word on the wage curve?. *Journal of Economic Surveys*, 19(3), 421-450.

[36] Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, 4(11), 1208-1214.

[37] Siontis, K. C. M., Evangelou, E., & Ioannidis, J. P. A. (2011). Magnitude of effects in clinical trials published in high-impact general medical journals. *International Journal of Epidemiology*, 40(5): 1280-1291.

[38] Xue, X., Cheng, M., & Zhang, W. (2021). Does education really improve health? A meta-analysis. *Journal of Economic Surveys*, 35(1), 71-105.

[39] Young, A. (2022). Consistency without inference: Instrumental variables in practical application. *European Economic Review*, 147, 104112.

Appendix

A Papers in meta-analysis

- **Literature 1: Minimum wage and employment**

  We draw estimates from Neumark and Wascher (2007), which reviews the literature at that time, and Neumark (2019), which updates the previous literature review. To promote comparability of estimates, we restrict to estimates of the elasticity of teen employment with respect to the minimum wage.

- **Literature 2: Return to schooling in China**

  The estimates in this literature are from the paper titled: "Return to schooling in China: A large meta-analysis" by Ma and Iwasaki (2021).

- **Literature 3: Health spending and children's mortality**

  We draw data for this literature from the paper titled: "The impact of health-care spending on health outcomes: A meta-regression" by Gallet and Doucouliagos (2017). These estimates reported the elasticity of children's mortality with respect to health care spending.

- **Literature 4: Health spending and life expectancy**

  The estimates in this literature were also taken from the same paper as Literature 3. However, in this literature, the estimates evaluate the elasticity of life expectancy with respect to health care spending.

- **Literature 5: Education and mortality**

  For this literature, we obtained data from the paper titled "Does education really improve health?" by Xue et al. (2021), which is publicly available at Open Science Framework (OSF). This literature includes estimates that measure the relationship between one's years of education and various health outcomes. In order to have comparable estimates, we only include estimates that quantify the effect of education on mortality.

- **Literature 6: Immigration and natives' wages**

In order to obtain estimates for this literature, we rely on a working paper titled "Meta-analysis of empirical evidence on the labour market impacts of immigration" by Longhi et al. (2008). We only include estimates that measure the effect of the stock of immigrants on natives' wages.

- **Literature 7: Remittances and education spending**

  For this literature, we draw data from the paper titled "The meta-analysis of effects of remittances on household education expenditure" by Askarov and Doucouliagos (2020), available on Deakin Lab for the Meta-Analysis of Research's database. The estimates evaluate the effect of households' remittances on their educational spending.

- **Literature 8: Intergenerational transmission of schooling**

  We obtain data on this literature by checking and collecting data from all studies included in the paper titled "The intergenerational transmission of education: A meta-regression analysis" by Fleury and Gilles (2018). The estimates in this literature measure the causal effect of parental education attainment on the educational attainment of their children.

- **Literature 9: Tuition and college enrollment**

  We obtain the data from the paper titled "Tuition fees and university enrolment: A meta-regression analysis" by Havránek et al. (2018). The estimates included in this literature evaluate the relationship between enrollment in a higher education institution and tuition, recalculated to partial correlation coefficients.

- **Literature 10: Individual discount rates**

  We draw estimates from the paper "Individual discount rates: A meta-analysis of experimental evidence" by Matousek et al. (2022). The estimates included in this meta-analysis are exclusively from experiments.

- **Literature 11: Capital and labor substitution**

  We draw the estimates from the paper "Measuring capital-labor substitution: The importance of method choices and publication bias" by Gechert et al. (2022). The estimates in this paper capture the elasticity of substitution between capital and labor.

- **Literature 12: Social cost of carbon**

  We draw estimates for this literature from the meta-analysis paper titled "Selective reporting and the social cost of carbon" by Havránek et al. (2015). The social cost of carbon is the approximate difference between present and future output as a result of carbon emissions, discounted back to the present time.

- **Literature 13: Elasticities of intertemporal substitution in consumption**

  We obtain estimates from "Measuring intertemporal substitution: The importance of method choices and selective reporting" by Havránek (2015). The elasticity of intertemporal substitution (EIS) in consumption is a measure of the willingness on the part of the consumer to substitute future consumption for present consumption.

- **Literature 14: Skilled & unskilled labor substitution**

  We draw data from the paper titled "Publication and attenuation biases in measuring skill substitution" by Havránek et al. (2022). Estimates in this literature measure the elasticity of substitution between skilled and unskilled labor.

- **Literature 15: Income elasticity of water demand**

  In this literature, we obtain estimates from the paper "Measuring the income elasticity of water demand: The importance of publication and endogeneity biases" by Havránek et al. (2018).

- **Literature 16: The elasticity of substitution of domestic and foreign goods**

  We draw the data from the paper "Estimating the Armington elasticity: The importance of study design and publication bias" by Bajzik (2020). In order to increase the comparability of estimates, we restrict the sample to papers that use the US as the domestic market.

- **Literature 17: Excess elasticity of consumption to income**

  We draw data from the paper titled "Do consumers really follow a rule of thumb? Three thousand estimates from 144 studies say 'probably not'" by

Havránek and Sokolova (2020). The parameter of interest is the estimate of consumption response to changes in income. Our data includes both micro and macro estimates.

- **Literature 18: Student employment and academic outcomes**

  We draw estimates from the paper titled "Student employment and education: A meta-analysis" (Kroupova et al., 2021). To make estimates comparable, we only include estimates that evaluate test scores. Moreover, the estimates by Kroupova et al. (2021) are converted to a comparable metric, the partial correlation coefficient (PCC).

- **Literature 19: Price elasticity of beer demand**

  We obtain a list of studies on price elasticity of beer demand from the paper titled "Meta-analysis of alcohol price and income elasticities—with corrections for publication bias" by Nelson (2013). The paper includes price elasticities for beer, wine and spirits. However, to increase comparability, we only consider the price elasticity of beer.

- **Literature 20: Price elasticity of gasoline demand**

  We draw the estimates from the paper titled "Demand for gasoline is more price-inelastic than commonly thought" by Havránek et al. (2012).

- **Literature 21: Elasticity of labor demand**

  We obtain data from the paper titled "The own-wage elasticity of labor demand: A meta-regression analysis" by Lichter et al. (2015). Data from this paper is derived from micro-level estimates of the elasticity of labor demand.

- **Literature 22: Income elasticity of gasoline demand**

  We draw data from the paper titled "Measuring global gasoline and diesel price and income elasticities" by Dahl (2012).

- **Literature 23: Wage curve**

  We draw data from the paper titled "The last word on the wage curve?" by Nijkamp and Poot (2005). The wage curve measures the elasticity of wage with respect to the unemployment rate in the local labor market.

- **Literature 24: Elasticity of taxable income**

  We draw data from the paper titled "The elasticity of taxable income: A meta-regression analysis" by Neisser (2021). The elasticity of taxable income measures the responsiveness of income to changes in the net-of-tax rate.

Appendix B

This appendix describes the shrinkage version of our main analysis.

The shrinkage version of our analysis limits the extent to which overfitting in Equation 1 will create noise in our estimates of regression residuals ϵ, which affects our main estimates by introducing noise into our estimates of Equation 2. To construct the shrinkage version of our main analysis, we follow these steps:

- **Step 1**: Construct a literature-demeaned journal rank by subtracting the mean journal rank by literature from the journal rank for each paper.
- **Step 2**: Estimate Equation 1 with the demeaned journal rank in lieu of rank_i.
- **Step 3**: Estimate the true variance of β_l using the subtracting variances approach as described in Section 4.
- **Step 4**: Construct a shrinkage factor S equal to the estimated $\frac{var(\beta_l)}{var(\widehat{\beta}_l)}$.
- **Step 5**: Multiply the slope for each literature j obtained in Step 2 by S and estimate fitted values and regression residuals from Equation 1 with these updated slopes.
- **Step 6**: Generate squared values of the residuals from **Step 5**.
- **Step 7**: Regress those squared residuals on journal rank.
- **Step 8**: Perform the main analysis with our shrinkage adjusted parameters.

The purpose of using demeaned journal ranks within literatures is to avoid the need for shrinkage adjustments to the literature-specific intercepts. (Use of demeaned journal ranks produces the same slopes of estimates with respect to ranks within each literature as in the baseline, but changes the literature intercepts to have the interpretation as the average $\widehat{\epsilon}^2$ for a paper of average rank.)

Appendix C

This appendix presents estimated slope coefficients from Equations 1 and 2 for each literature from our preferred specifications. These preferred specifications control for data year, do not address outliers, and winsorize rank at the 500th-ranked journal.

Table 2 reports the estimated bias coefficients β_l for each literature l from Equation 1—i.e., the literature-specific coefficient when regressing (normalized) parameter estimates on our measure of journal rank. See Table 1 for a list of literatures and Appendix A for more detailed description of the literatures.

Table 2: Literature-specific bias coefficients (Equation 1)

	Lit 1	**Lit 2**	**Lit 3**	**Lit 4**	**Lit 5**	**Lit 6**	**Lit 7**	**Lit 8**
β_l	-0.124	-0.158	-0.238	0.286	0.343	0.035	0.209	0.015
Robust SE	(0.132)	(0.127)	(0.134)	(0.193)	(0.127)	(0.103)	(0.179)	(0.061)
	Lit 9	**Lit 10**	**Lit 11**	**Lit 12**	**Lit 13**	**Lit 14**	**Lit 15**	**Lit 16**
β_l	0.015	0.003	-0.034	-0.069	0.011	-0.04	-0.296	0.302
Robust SE	(0.106)	(0.064)	(0.035)	(0.083)	(0.012)	(0.059)	(0.118)	(0.161)
	Lit 17	**Lit 18**	**Lit 19**	**Lit 20**	**Lit 21**	**Lit 22**	**Lit 23**	**Lit 24**
β_l	-0.107	0.299	-0.024	-0.253	-0.037	-0.045	-0.228	-0.063
Robust SE	(0.063)	(0.177)	(0.096)	(0.175)	(0.069)	(0.108)	(0.186)	(0.059)

Standard errors are clustered by paper.

Notes: The bias coefficients are reported from the regression of estimates on literature dummies, the interaction term between journal rank and literature dummies and average year of data.

Table 3 reports the variance coefficients ω_l for each literature l from Equation 2—i.e., the literature-specific coefficient when regressing squared estimated residuals from Equation 1 on our measure of journal rank.

Table 3: Variance coefficients with year control

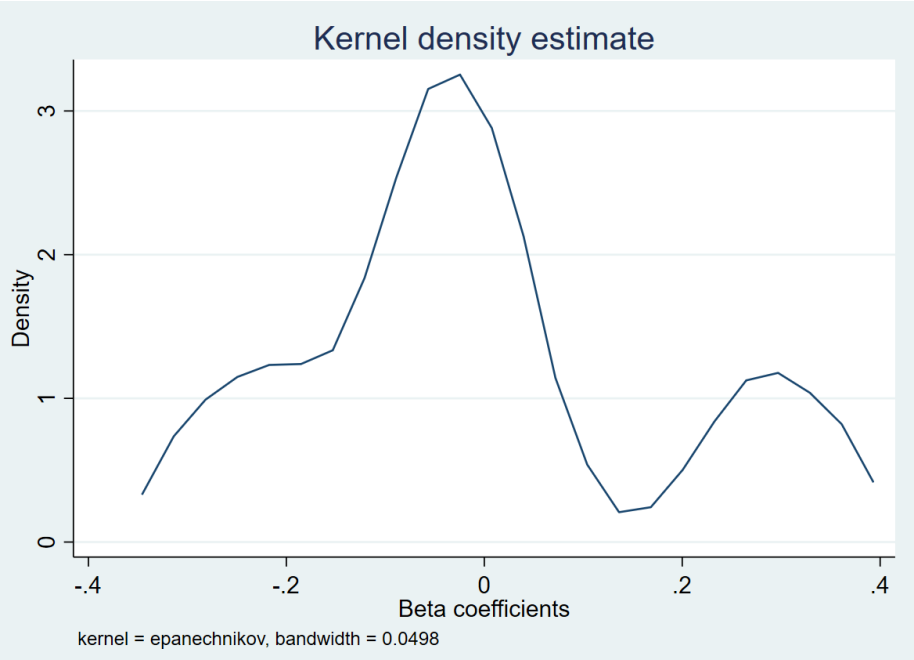
	Lit 1	**Lit 2**	**Lit 3**	**Lit 4**	**Lit 5**	**Lit 6**	**Lit 7**	**Lit 8**
ω_l	-1.186	-1.049	0.072	0.042	0.406	0.364	0.24	0.153
Robust SE	(1.064)	(0.928)	(0.143)	(0.123)	(0.229)	(0.208)	(0.220)	(0.163)
	Lit 9	**Lit 10**	**Lit 11**	**Lit 12**	**Lit 13**	**Lit 14**	**Lit 15**	**Lit 16**
ω_l	0.001	-0.05	-0.502	0.016	0.575	-0.423	-0.338	0.277
Robust SE	(0.168)	(0.143)	(0.454)	(0.332)	(0.574)	(0.413)	(0.275)	(0.048)
	Lit 17	**Lit 18**	**Lit 19**	**Lit 20**	**Lit 21**	**Lit 22**	**Lit 23**	**Lit 24**
ω_l	0.113	0.432	0.06	-0.144	0.138	-0.071	0.233	0.156
Robust SE	(0.094)	(0.332)	(0.189)	(0.328)	(0.364)	(0.098)	(0.454)	(0.194)

Standard errors are clustered by paper.

Notes: The variance coefficients are reported from the regression of estimates on literature dummies, the interaction term between journal rank and literature dummies and average year of data.

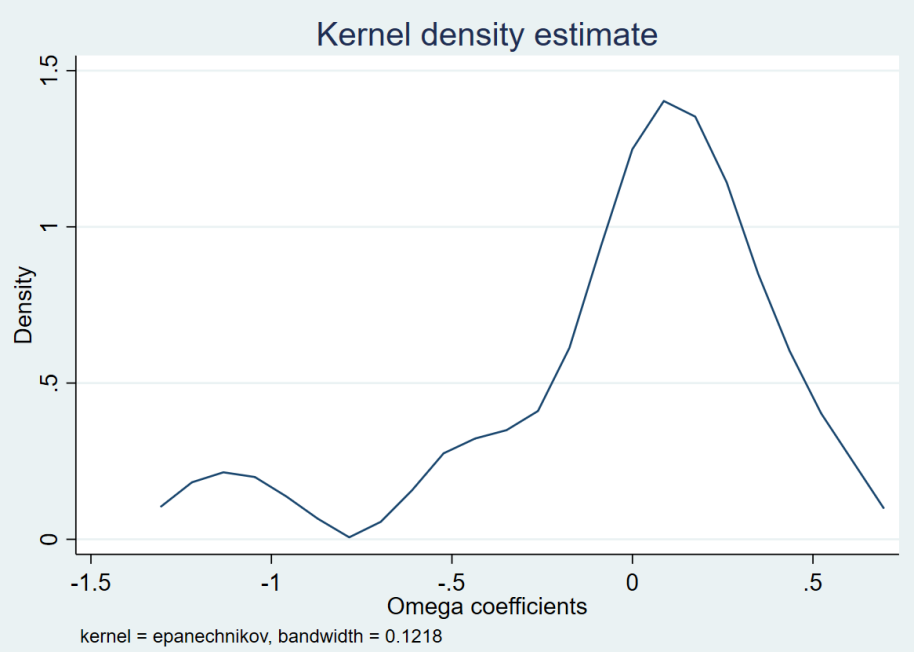
We also present the kernel density plots of $\widehat{\beta}_l$ and $\widehat{\omega}_l$ in Figures 4 and 5, respectively. That is, Figure 4 is a kernel density plot of the estimates contained in Table 2, and Figure 5 of the estimates in Table 3.

Figure 4: Kernel density of $\widehat{\beta}_l$



Notes: This figure presents the density of $\widehat{\beta}_l$ across 24 literatures obtained from estimating Equation 1. β_l is the slope coefficient relating journal rank to average parameter estimate.

Figure 5: Kernel density of $\widehat{\omega}_l$



Notes: This figure presents the density of $\widehat{\omega}_l$ across 24 literatures obtained from estimating Equation 2. ω_l is the slope coefficient relating journal rank to the variance of estimates.

Appendix D

Standard errors specification curve

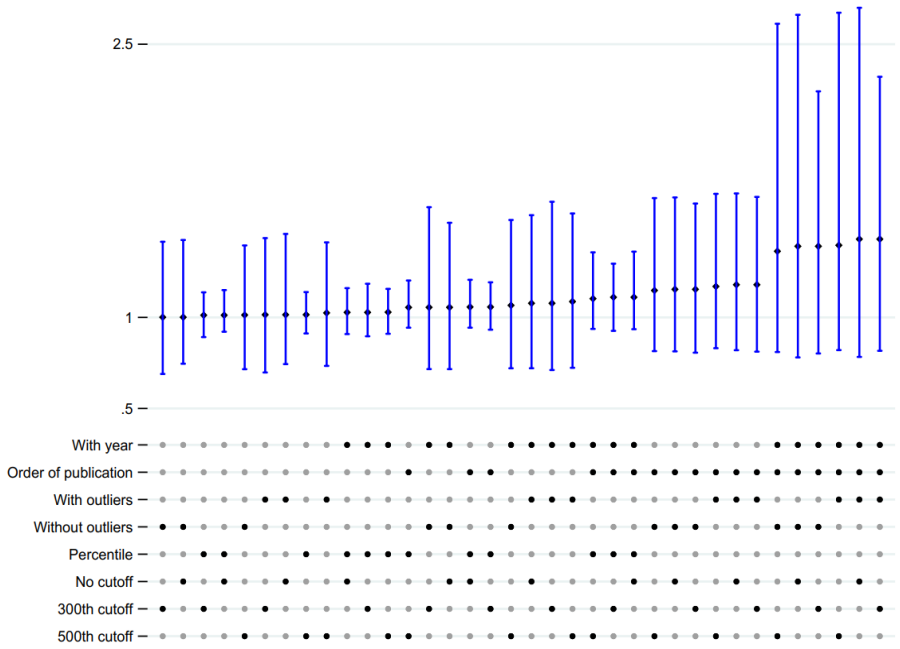
In this appendix, we estimate the relationship between journal rank and the magnitude of standard errors.

We work with the log of standard errors to avoid the possibility of estimating negative-valued standard errors due to extrapolation. Then, we run a regression analogous to Equation 1, but with log-transformed standard errors instead of estimates on the left-hand side. Next, we evaluate the estimated average log of standard errors at the rank of ILRR and the rank of the QJE. Finally, we transform those log forms of standard errors back to linear forms of standard errors and evaluate the ratio.

Figure 6 shows the ratio of average estimated standard errors at the rank of ILRR to the average estimated standard errors at the rank of the QJE across a variety of specifications. The most typical result is that the standard errors are comparable, i.e., the ratio is around 1, but with the QJE-estimator typically producing smaller

standard errors. However, as the specification curve shows, some specifications estimate ratios deviating substantially from 1.

Figure 6: The ratio of average standard error at ILRR to QJE



Notes: This figure reports the ratio of the standard errors of ILRR estimates to QJE estimates across 36 different specifications. The top half presents the point estimates along with 95% confidence interval. The bottom half presents the choices in each specification.

Appendix E

In this appendix, we report the results of regressions related to the "follow the leader" explanation.

Table 7 reports estimates of the parameters of the "follow the leader" regression described in Section 6. A positive coefficient on $\widehat{\theta}_i * \text{rank}_i$ indicates that estimates published in higher-ranked journals are more predictive of trends in a literature, consistent with the view that they are more influential.

Table 7: Influence of higher-ranked estimates on subsequent literature

	$\bar{\theta}_i^{post} - \bar{\theta}_i^{pre}$
$\widehat{\theta}_i * \text{rank}_i$	0.016
	(0.010)
$\widehat{\theta}_i$	0.003
	(0.009)
rank_i	-0.030
	(0.015)
constant	0.007
	(0.085)
R^2	0.021
N	13,048

Standard errors are clustered by literature

Table 8 reports the results of a similar analysis designed to determine whether high- or low-ranked journals are more influenced by prior estimates in a literature. Within each literature, we can estimate β_l using only observations which were published prior to estimate i; call this β_i^{pre}. Similarly, we can construct an estimated β_l using only papers published after i; call this β_i^{post}.

The relationship between $\beta_i^{post} - \beta_i^{pre}$ and $\widehat{\theta}_i$ is then informative about whether high- or low-ranked journal-estimators are more prone to following the leader. Suppose that an anomalously large estimate i would increase estimates published in low-ranked journals by more than it would increase estimates published in high-ranked journals. Then this means a high $\widehat{\theta}_i$ would lead to a decrease in β_i^{post}. We subtract β_i^{pre} to capture the difference in β_l caused by estimate i.

Table 8 specifically reports the results from regressing $\beta_i^{post} - \beta_i^{pre}$ on the same regressors as in Table 7. If publications in lower-ranked journals are more prone to following the leader, we would expect the coefficients on $\widehat{theta}_i$ and/or $\widehat{theta}_i * \text{rank}_i$ to be negative. Instead, the estimates are positive and insignificant.

Similarly, suppose that surprising results are easier to publish in high-ranked journals. Then the causal effect of a high result $\widehat{\theta}_i$ today on the publication outcome of future papers should be to increase the expected publication rank of papers with low estimates, and decrease the expected publication rank of papers with high esti-

mates. This would result in a negative relationship between $\widehat{\theta}_i$ and β_i^{post}, which is not supported by the results in Table 8.

Table 8: Differential influence on high- vs. low-ranked followers

	$\beta_i^{post} - \beta_i^{pre}$
$\widehat{\theta}_i * \text{rank}_i$	0.007
	(0.018)
$\widehat{\theta}_i$	0.013
	(0.040)
rank_i	-0.143
	(0.110)
constant	0.290
	(0.391)
R^2	0.00
N	11,130

Standard errors are clustered by literature.

Appendix F

This appendix reports the relationship between journal rank and order of publication, which we denote as order_i.

We estimate the following equation and report our coefficient of interest, χ, in Table 9. The first column of Table 9 reports the coefficient in our baseline analysis, which controls for data year. Because data year is likely to be correlated with publication year, and the unconditional relationship between journal rank and order of publication might be of interest, we also report the relationship without controlling for data year in the second column.

$$\text{order}_i = \sum_{l=1}^{24} 1\{\text{lit}_i = l\}\alpha_l + \chi \text{rank}_i + \sum_{l=1}^{24} 1\{\text{lit}_i = l\}\eta_l \text{year}_i + \epsilon_i, \tag{3}$$

Table 9: Regression of order of publication on journal rank

	With year control	**Without year control**
rank	-0.827	-1.127
	(0.510)	(0.625)
R^2	0.84	0.74
N	14,387	14,387

Standard errors are clustered by paper.