THE POWER OF BIAS IN ECONOMICS RESEARCH*

John P. A. Ioannidis, T. D. Stanley and Hristos Doucouliagos

We investigate two critical dimensions of the credibility of empirical economics research: statistical power and bias. We survey 159 empirical economics literatures that draw upon 64,076 estimates of economic parameters reported in more than 6,700 empirical studies. Half of the research areas have nearly 90% of their results under-powered. The median statistical power is 18%, or less. A simple weighted average of those reported results that are adequately powered (power ≥ 80%) reveals that nearly 80% of the reported effects in these empirical economics literatures are exaggerated; typically, by a factor of two and with one-third inflated by a factor of four or more.

> Statisticians routinely advise examining the power function, but economists do not follow the advice.
>
> McCloskey (1985, p. 204)

Good policy and practice is built on the foundations of reliable scientific knowledge. Unfortunately, there are long-held suspicions that much of what passes as evidence in economics, medicine or in psychology (and possibly other fields) lacks sufficient credibility (De Long and Lang, 1992; Ioannidis, 2005*b*; Leamer, 1983; Ioannidis and Doucouliagos, 2013; Maniadis *et al.*, 2017). For example, it has been difficult to reproduce and verify significant bodies of observational and experimental research independently (Ioannidis, 2005*a*; Begley and Ellis, 2012; Begley and Ioannidis, 2015; Duvendack *et al.*, 2015; Nosek *et al.*, 2015). Moreover, empirical research is plagued by a range of questionable practices and even the fabrication of results. Consequently, some argue that science is experiencing a credibility crisis. This crisis of confidence in research permeates multiple scientific disciplines. While there are discipline-specific nuances, there are also many shared experiences and distorted incentives. Just as declining credibility may spill over from one discipline to another, successful strategies and practices can benefit other disciplines. Hence, a multidisciplinary approach may advance all sciences.

Statistical power is a critical parameter in assessing the scientific value of an empirical study. Power's prominence increases with policy importance. The more pressing it is to have evidence-based policy, the more critical it is to have the evidence base adequately powered and thereby credible. By definition, adequate power means that the empirical methods and data should be able to detect an effect, should it be there. Low power means high rates of false negatives. However, as Ioannidis (2005*b*) has argued, low power also causes high rates of false positives, where non-existent effects are seemingly detected. Aside from the prior probability that a given economic proposition is true (a magnitude that would likely cause endless

debate among economists), the key parameters for assessing the validity of any given reported research result are: statistical power and the proportion of reported non-null results that are the artefact of some bias (e.g. misspecification bias and publication selection bias).

How credible is empirical economics? Is empirical economics adequately powered? Many suspect that statistical power is routinely low in empirical economics. However, to date, there has been no large-scale survey of statistical power widely across empirical economics. The main objectives of this article are to fill this gap, investigate the implications of low power on the magnitude of likely bias and recommend changes in practice that are likely to increase power, reduce bias and thereby increase the credibility of empirical economics.

For many researchers, a key consideration is whether a particular research project is publishable. In contrast, from a social welfare perspective, the more important consideration is the contribution that the research inquiry makes to science.[1] The validity and credibility of empirical economics has long been questioned. For example, Leamer (1983) famously pointed out that empirical economics is vulnerable to a number of biases and, as a result, produces rather fragile results that few economists take seriously. De Long and Lang (1992) found evidence of publication selection bias among the top economic journals. Ziliak and McCloskey (2004) searched papers in the *American Economic Review* (*AER*) and found that only 8% of the empirical studies published in the 1990s actually consider statistical power.[2] Doucouliagos and Stanley (2013) quantitatively surveyed 87 empirical economics areas and found evidence of widespread publication selection bias. Ioannidis and Doucouliagos (2013, p. 997) recently reviewed and summarised available evidence of prevalent research practices and biases in the field and called into question the credibility of empirical economics, arguing that overall 'the credibility of the economics literature is likely to be modest or even low'.

We are not suggesting that power analysis is never done in economics. Indeed, the importance of power is recognised in several areas of empirical economics research. For example, in assessing randomisation of microcredit programmes, Banerjee *et al.* (2015, p. 3) conclude that 'statistical power still poses a major challenge to microcredit impact studies'. Power analysis is frequently used in the revealed preference approach to consumption or production decisions (Andreoni *et al.*, 2013). Nevertheless, in spite of its widely recognised importance, there are currently no large-scale surveys of statistical power in empirical economics nor a careful quantification of the consequences of ignoring power.[3]

---

[1] We do not mean to suggest that publication is the sole or even dominant motive for the majority of researchers. Indeed, the quest for scientific evidence or support likely inspires the vast majority of researchers. Nevertheless, career concerns and misaligned incentives in the scientific reward system may cause many researchers to supply research findings that are more likely to be published.

[2] This, however, was a marked increase from only 4% of papers published in the *AER* in the 1980s (McCloskey and Ziliak, 1996). McCloskey and Ziliak merely asked: 'Does the paper mention the power of the tests? (Consider the power of the test?)'. They did not assess whether studies were adequately powered, but only if power was mentioned.

[3] Recently in an unpublished paper, Zhang and Ortmann (2013) use a single meta-analysis to calculate the statistical power of dictator game experiments. They find that the median power of these experiments is 25%, slightly higher than what our survey of 159 meta-analyses finds.

Prior studies discuss power or bias only for leading economics journals (De Long and Lang, 1992; McCloskey and Ziliak, 1996), or where a wider range of journals is surveyed, only bias is considered (Doucouliagos and Stanley, 2013). In order to validate the claims of the lack of credibility of economics and to quantify the likely magnitude of bias, it is necessary to investigate the broader evidence base more rigorously. Accordingly, we survey two dimensions of the credibility of empirical economics research: statistical power and bias. Our survey is based on a statistical examination of 159 meta-analyses that provide over 64,000 estimates of key parameters (the estimated effect size and its estimated standard error) drawn from approximately 6,700 empirical studies. Using these data, we calculate: the proportion of reported findings that are adequately powered for a given area of economics research, the median power of that area of research, the estimate of effect that emerges when only adequately powered estimates are considered, and the proportion of the typical reported effect that is likely to be the result of some type of bias or artefact. We find that economics research is generally underpowered and most economics research is afflicted with substantial residual bias. Half of the areas of economics research have 10.5% or fewer of their reported results with adequate power.[4] We also find that 20% or more of research literatures have no single study that is adequately powered. In spite of this low power, most studies still report statistically significant effects. While these results cast a shadow on the credibility of economics research, not all is lost. At least one adequately powered study is available in most economics literatures that we examined. Moreover, meta-analysis can synthesise the results from numerous underpowered studies, filter out various biases and thereby suggest better estimates of underlying empirical economic parameters, necessary for valid inferences. Hence, even if the credibility of economics research is much lower than desirable, a careful systematic review and meta-analysis may improve statistical inference and offer some policy guidance.

Our second contribution is to present a new approach to correcting bias in empirical economics research, a weighted average of the adequately powered (WAAP). This estimator employs an unrestricted weighted least squares weighted average calculated only on the adequately powered estimates - in contrast to conventional meta-analysis that uses all available estimates. We show that using only the adequately powered studies, WAAP may give a credible and defensible estimate of the empirical effect in question. Should some type of publication selection, reporting or small-sample bias be present in the research record, WAAP is quite likely to reduce it. At the very least, the weighted average of the adequately powered offers a validation of corrected empirical effect estimated by other meta-regression analysis methods. An advantage of WAAP is that it makes no assumption about the shape, cause, or model of publication or selective reporting bias.

The ensuing Section provides a review of statistical power and bias. Our survey data are discussed in Section 2. Section 3 discusses our findings. The article is concluded in Section 4.

---

[4] Power is also low in other social science disciplines but often higher than economics. In industrial psychology, it ranges between 40% and 60% (Schmidt and Hunter, 2015) but has an average of 35% more broadly in psychology (Bakker *et al.*, 2012). In neuroscience, the median power is quite low, 21%, but still somewhat higher than what we find for economics (Button *et al.*, 2013).

1. Power and Bias

> That power which erring men call Chance.
>   John Milton (1631: 1909–14, Comus, a Mask, lines 601-02, *L'Allegro*).

A study is adequately powered when there is a high likelihood that it will detect a genuine empirical effect. Low statistical power results in high Type II error. Statistical power is influenced by sample size, the size of the effect and the desired level of statistical significance. Thus, for example, studies investigating small effects with small sample sizes will, by necessity, suffer from low power. *Ceteris paribus*, the larger the effect, the easier it will be to detect. The larger the sample, the greater the statistical power to detect a given genuine effect.[5]

Since Cohen (1965), adequate power in most social sciences has been conventionally defined as 80%. That is, the probability of a Type II error should be no larger than four times the probability of the conventional Type I error (0.05).[6] Like the conventional significance level, $\alpha = 0.05$, this convention is arbitrary and yet routinely followed across different sciences. Investigators in the experimental sciences are routinely required by funding bodies and journal editors to design their experiments to achieve this conventional level of power. Conventional formulas for sample sizes are often based on achieving this level of power.[7]

If we adopt the conventional 5% level of statistical significance and 80% power level, as well, then the 'true effect' will need to be 2.8 standard errors from zero to discriminate it from zero. The value of 2.8 is the sum of the usual 1.96 for a significance level of 5% and 0.84 that is the standard normal value that makes a 20/80% split in its cumulative distribution. Hence, for a study to have adequate power, its standard error needs to be smaller than the absolute value of the underlying effect divided by 2.8. We make use of this relationship to survey adequate power in economics.

All that remains to calculate power are the values of the standard error and an estimate of 'true' effect. Because our survey of empirical economics produced 64,076 effect size estimates and their associated standard errors from 159 meta-analyses (see Section 2 and Appendix A), we have much information from which to work. We could calculate power using each of these estimates and their standard errors but to do so would be circular and tell us little beyond the reported p-values. By this circular calculation of power, if the estimate is reported to be statistical significant, then power

---

[5] The average effect size among the data sets we survey (see Section 3 below), measured largely as partial correlations and elasticities, is only 0.17. By Cohen's (1988) widely used benchmarks or by economic significance, this is a relatively small effect. Hence, *ceteris paribus*, it will be difficult to detect effects for the typical economic phenomenon.

[6] A 20% Type II error can still be considered to be rather high (Schmidt and Hunter, 2015). Others have argued that the optimal pair of type I and II errors vary according to the circumstances and aim of a study (Ioannidis *et al.*, 2013).

[7] In economics, one area that should match the requirements of statistical theory is experimental economics. Here, the researcher has some control over sample size. Yet, it is doubtful whether sample size is determined in order to achieve adequate statistical power in experimental economics. Indeed, anecdotal evidence suggests that sample size is driven primarily by available budget and/or the researchers' time rather than statistical power. As mentioned previously, low power has been found (25%) in experiments involving the dictator game, Zhang and Ortmann (2013).

will be least 50%; 80% if the reported p-value is 0.005 or less. But what if statistically significant estimates are preferentially reported and/or published?

All economists know that empirical research is subject to many sources of potential bias. From dozens of past meta-analyses, it appears that misspecification biases are a major source of the excess variation routinely observed among reported economic estimates. As a result, it is conventional practice to code for omitted-variable bias and most economic meta-analyses do (Stanley and Doucouliagos, 2012; Stanley *et al.*, 2013). All empirical estimates potentially contain sampling error, random misspecification biases and various other selection biases. In most fields of applied econometrics, there are nearly limitless allowed combinations of methods, models and data choices. Nonetheless, only a few possible model specifications can actually be correct. Thus, most available variations to methods and models produce, by definition, misspecification biases.

> Empirical results reported in economics journals are selected from a large set of estimated models. Journals, through their editorial policies, engage in some selection, which in turn stimulates extensive model searching and pre-screening by prospective authors. Since this process is well known to professional readers, the reported results are widely regarded to overstate the precision of the estimates, and probably to distort them as well. As a consequence, statistical analyses are either greatly discounted or completely ignored (Leamer and Leonard, 1983, p. 306).

Since Leamer's (1983) 'Let's take the con out of econometrics', economists have been acutely aware of the high likelihood that the inadvertent selection of misspecification bias may overcome low statistical power to produce apparent statistically significant outcomes. When power is low, reported statistically significant findings are quite likely to be artefacts from chance and bias (Ioannidis, 2005*b*; Ioannidis and Trikalinos, 2007).

If any single estimate of 'true' effect is questionable, how can we break the circularity of a power calculation that depends on some assessment of 'true' effect? One way to break the circle is to calculate power hypothetically; that is, for a given effect deemed to be 'practically significant'. For example, if we were to agree that an elasticity greater in magnitude than 0.1 is of sufficient size to have practical policy implications, it would be easy to calculate the power needed to detect this threshold of practical significance if we already know the associated standard error. But would any such definition of 'practical significance' hold up to scrutiny? If we are estimating the elasticity for a specific brand that has strong competition, is such a small price elasticity, -0.1, relevant? Perhaps, in this case, it might be sufficient to know whether the elasticity is greater than or less than -1.0? Or, what about income elasticities? Should we demand as much precision in estimating income elasticity as price elasticity? Without some widely accepted standard for practically significant effect size, this circle cannot be broken.

Alternatively, we can escape this circle, if we estimate effect from all reported estimates in the research record, rather than only one. Doing so could also greatly reduce the sampling error and potentially also some of the biases that are likely to be present in any individual empirical estimate. This is precisely the role of meta-analysis. Meta-regression analysis is the statistical analysis of an entire empirical economics

research literature (Stanley and Jarrell, 1989). It seeks to summarise and analyse the full empirical record on a given empirical economic question, phenomenon or effect. With a meta-analysis estimate of a given empirical effect, we can eliminate much of the potential bias in single estimates.

Simple weighted or unweighted averages of all reported estimates do much to eliminate sampling error and random misspecification bias, because the average number of estimates per meta-analysis in our survey is 403 (median = 191). As long as the estimation error and bias are random, the central limit theorem ensures that the average across all estimates in given area of research will provide an adequate estimate of 'true' effect. Furthermore, when we can calculate 159 such averages across as many areas of economics research,[8] *random* estimation error and bias would be virtually eliminated. The important exception to such desirable statistical properties of meta-averages occurs if there is *systematic* bias.[9] That is, if reported estimates contain selected systematic bias, then any average will also be biased. Such biases in single estimates may also carry over their impact onto meta-analysis summary estimates. This is one reason why it is important to find meta-analysis methods that are less susceptible to the impact of selective reporting.

1.1. *Publication and Reporting Bias*

For at least a half century, 'publication bias' has been widely recognised as a potential threat to the validity of empirical science (Sterling, 1959; Tullock, 1959; Rosenthal, 1979; Glass *et al.*, 1981; Lovell, 1983; Hedges and Olkin, 1985; Begg and Berlin, 1988; De Long and Lang, 1992; Card and Krueger, 1995; Sterling *et al.*, 1995; Copas, 1999; Ioannidis and Trikalinos, 2007; Stanley, 2008; Stanley and Doucouliagos, 2012, 2014; to cite but a few relevant references). Publication selection bias is the tendency by some researchers, referees or editors to report, selectively, statistically significant findings or those consistent with conventional theory (Card and Krueger, 1995). For the purposes of this study, we consider the tendency to selectively report statistical significant findings in the 'right' direction as 'publication bias', regardless of whether the specific cause is selection in publication-review process, the confusion of statistical significance with scientific importance, small-samples, or p-value hacking (Stanley and Doucouliagos, 2012). Such selection of biases and sampling errors tend to increase the reported magnitude of effects, making the average notably larger than the 'true' effect (Stanley, 2008; Stanley *et al.*, 2010). Among 87 areas of economic research, Doucouliagos and Stanley (2013) find that the majority exhibit patterns suggestive of 'substantial' or 'severe' publication selection bias. Thus, we cannot rule out the

---

[8] These 159 meta-analyses actually represent fewer areas of economic research, because a few areas of research have more than one meta-analysis. Nonetheless, more than one hundred distinct areas of economics research, distributed widely across micro, macro, labour and international economics, are represented in our survey.

[9] If there is genuine heterogeneity of the effect (e.g. across different countries, industries, income levels etc.), then no single summary estimate can represent an entire research literature. Conventional practice is to employ meta-regression with numerous moderators to map out this distribution and to accommodate any systematic heterogeneity. However, in this study, we are interested in documenting the proportion of underpowered studies/estimates that we find across many areas of economics research. Hence, we focus on the mean of the distribution of effect sizes.

possibility that a given empirical economics research record is not a selected and thereby skewed sample of estimated effects.

Several statistical methods have been developed to identify and accommodate potential publication and related reporting biases (Stanley, 2005, 2008; Ioannidis and Trikalinos, 2007; Stanley *et al.*, 2010; Stanley and Doucouliagos, 2014) and others have proposed methods to detect and evaluate the extent of p-value hacking (Simonsohn *et al.*, 2014). With information from 159 meta-analyses, these statistical methods can be used to approximate the genuine empirical effect, or at the least, to filter out some of the selection bias should it be present in a given area of research. We acknowledge that any correction for these reporting biases is based on assumptions or approximations and that the 'true' effect is impossible to know. To be sure that our approach does not contribute to bias, we use several conservative methods that can only reduce publication bias, on average, if it is present. Furthermore, we use multiple meta-methods to ensure that our overall assessment of empirical economics is robust. If several conservative approaches reveal similar general patterns of power and imputed bias across 159 meta-analyses, 6,700 studies with over 64,000 estimates, then these revealed patterns are likely to reflect some genuine features of economics research.

1.2. *Meta-estimates of True Effect*

To be conservative and robust, we use four alternate meta-analytical approaches to approximate 'true' effect. First, we use a simple weighted average of all estimates in a given area of research, the so-called 'fixed-effect'. The fixed-effect weighted average employs optimal weights, weights that are the same as those that economists use for weighted least squares (WLS) (Hedges and Olkin, 1985; Cooper and Hedges, 1994; Stanley and Doucouliagos, 2015). We then compute adequate power by comparing the standard error of each estimate to the absolute value of this fixed-effect weighted average (*WLS-FE*) divided by 2.8. If the standard error is less than this threshold, we can conclude that that estimate is adequately powered to detect an effect size suggested by the weighted average of all estimates in this area of research. Similarly, median power can be calculated from the median standard error and *WLS-FE*. Of course, these calculations are only an approximation, because the true effect remains unknown. We simply ask whether the power would have been adequate had the 'true' effect been this weighted average of all estimates observed across the entire area of research in question. We prefer the fixed-effect weighted average over the random-effects weighted average, because random-effects are widely known to be more biased than fixed-effects when there is publication selection bias, reporting bias or small-study effects, i.e. larger estimates in smaller studies (Poole and Greenland, 1999; Sutton *et al.*, 2000; Henmi and Copas, 2010; Stanley *et al.*, 2010; Stanley and Doucouliagos, 2014, 2015).

We consider this a very conservative approach to calculating power, because *WLS-FE* is known to be exaggerated when there is some type of reporting bias, but its bias is less than the unweighted average of the reported effects or the random-effects weighted average (Stanley, 2008; Stanley and Doucouliagos, 2014). If there is selective reporting bias, *WLS-FE* will tend to be too large and thereby give a power that is too high. If there is no selective reporting, this fixed-effect weighted average is known to be unbiased and

thus statistical power will be appropriately assessed. If anything, *WLS-FE* will tend, on average, to overestimate power; thus, this approach serves, on average, as an empirical upper bound for power.

For the sake of robustness, we consider three additional approaches to estimating the underlying effect from a meta-analysis that may get closer to the 'true' effect, on average. The first of these is the top 10% (*Top 10%*), which calculates the same fixed effect weighted average on the most precise (smallest standard errors) 10% of the reported estimates (again using inverse variance WLS weights). The advantage of *Top 10%* is that the most precise estimates may be less vulnerable to selection for statistical significance or small-sample biases (Stanley *et al.*, 2010). Our next robustness approach is to follow (Ioannidis, 2013) and use the single most precise estimate (*Top 1*). *Top 1* is the single most precise estimate, not the top 1% of most precise estimates. The most precise study is not necessarily the best in quality or the most well-protected from bias but it carries more information and, other things being equal, is likely to be less susceptible than smaller studies to diverse biases, for example, selective reporting, questionable research practices or (much rarer) fraud.

Finally, we use the conditional corrected estimate, precision-effect test and precision-effect estimate with standard error (*PET-PEESE*), to identify the meta-average (Stanley, 2008; Stanley and Doucouliagos, 2014). This involves regressing the reported effect sizes on a constant and either the standard errors or on the standard errors squared. These meta-regression models are Taylor polynomial approximations to the expected value of an incidentally truncated normal distribution (Stanley and Doucouliagos, 2014). Simulations reveal that the simple meta-regression model of reported estimated effects and their standard errors can identify and quantify publication selection bias (funnel-asymmetry test, *FAT*) and provide evidence of a genuine empirical effect beyond publication or reporting bias using the precision-effect test (*PET*) (Egger *et al.*, 1997; Stanley, 2008; Stanley *et al.*, 2010; Stanley and Doucouliagos, 2014). When there is evidence of such a genuine empirical effect from *PET*, the *PEESE* estimate (from the meta-regression model that uses the standard errors squared) is used as a correction for potential publication or reporting biases, because *PEESE* has been shown to have smaller bias and mean square error (Stanley and Doucouliagos, 2014).[10]

The first three methods of calculating the overall effect size make no assumption about how best to model or accommodate publication bias should it be present. They are 'conservative', in the sense that that they are likely to err on the side that overestimates the statistical power of reported empirical economics results. With the exception of *PET-PEESE*, these methods are known to be biased upward in the presence of the selective reporting of statistically significant results (Stanley, 2008; Stanley *et al.*, 2010; Stanley and Doucouliagos, 2014). As discussed above, the larger the magnitude of the 'true' effect, the higher the power of each reported estimate or test to detect it. Thus, at least three of these methods will tend to overestimate power, on average, with *WLS-FE* being the most prone to overestimation.

---

[10] See Stanley and Doucouliagos (2012, 2014) for a more detailed description and justification of this conditional *PET-PEESE* correction for publication selection or reporting bias.

2. Data

In order to assess power and bias in economics, we need detailed statistics over a wide range of economic research areas. Survey articles, systematic reviews and meta-analyses are all leading candidates for the requisite data. While survey articles and systematic reviews are informative, they rarely report sufficiently systematic quantitative information (e.g. standard errors for each estimate) from which power can be calculated. Hence, we rely on the existing meta-analyses of empirical economics.

To be included in our survey, a meta-analysis had to report some comparable effect size and its standard error for each study, or other information from which we can estimate the standard error. We collected our data from three sources. First, we used various search engines (e.g. Econlit, Scopus, JSTOR and Google Scholar) and we downloaded meta-data that was available on web pages and internet sites. Second, in several cases, we were able to extract the necessary data from tables or appendices in the article that reports the meta-analysis. Third, we contacted all researchers affiliated with MAER-Net or who were known to have conducted a meta-analysis in economics. In total, we contacted over 200 authors of economic meta-analyses. We include both published and unpublished meta-analyses, as well as several that were not published in economics journals but dealt with an economics-relevant topic. The inclusion criteria are that the meta-analysis covered some economics area of research, broadly conceived (e.g. labour, health, international, growth, development, or any other area that is covered by the *Journal of Economic Literature* classification codes) and had sufficient information to calculate statistical power.[11] The survey was conducted from March to June 2015.

The above search strategy resulted in 159 meta-analyses. In total, these 159 meta-analyses include approximately 6,730 primary studies collectively reporting 64,076 empirical estimates.[12] The 159 areas that make up our survey are reported in Appendix A. Of these, 18% relate to labour economics, 10% international economics, 23% micro-economics, 17% macro-economics and finance, 27% development and growth and 5% is other. We oversample from the more recent meta-analyses, because the authors of some of older studies did not respond to our request for data or the data were no longer available. Receiving the data often required the cooperation of the authors of these meta-studies.

Nearly all the estimates included in our survey involve tests to reject a null hypothesis. In a handful of cases, the empirics may involve theories that favour

---

[11] Thirty five (22%) of the 159 meta-analyses were unpublished at the time of the survey, while 32 (20%) were published in fields other than mainstream economics (e.g. political science, management and human resource management). Purely theoretical areas of research, by definition, will not contain the necessary empirical information upon which to calculate power.

[12] The most common effect size in these meta-analyses of economics research is elasticities, semi-elasticities and partial correlations. We do not have clear information on the exact number of studies, as opposed to estimates, for three of these 159 meta-analyses, because the available data did not always provide study IDs. Our best estimate is that the number of studies is approximately 6,730. Thirty eight of the 159 meta-analyses use a single estimate from each included empirical study. The remaining 121 meta-analyses include multiple estimates. While some of these multiple estimates are statistically independent, most are not. The standard treatment of non-independent samples in meta-regression is through clustered adjusted standard errors or modelling the dependence through linear hierarchical models or panel data estimators. Within-study dependence affects estimated variances. Hence, it is not an issue for our survey, as we are not calculating confidence intervals nor conducting hypothesis tests.

accepting the null hypothesis. Nevertheless, even if we expect little or no publication bias, the *FAT-PET* approach remains valid. In a survey of publication bias in economics, we found that nearly 30% of economics research shows little or no reporting selection (Doucouliagos and Stanley, 2013). The estimates in our survey come predominantly from reduced form models that explore a single parameter. However, our framework for calculating statistical power applies equally for structural and reduced form estimates, depending on which hypothesis is of central interest.

3. Findings

Table 1 reports the percentage of empirical economics findings that have 'adequate power', defined by the widely accepted convention that power is adequate if it is 80% or higher. It is clear that most of empirical economics is underpowered. The median proportion that is adequately powered is only 10.5%, using the most generous approach for estimating the 'true' effect – *WLS-FE*, column (1). That is, half of the areas of economics have approximately 10% or fewer of their estimates with adequate power. And, this median is smaller (1.9–6.5%) if any other approach to estimating the 'true' effect is employed.

Perhaps more disconcerting is the proportion of economics research areas that do not contain a single study or estimate that has adequate power to detect the effect in question. Between 19.5% (*WLS-FE*) and 35.8% (*Top 1*) of the areas of empirical economics are entirely comprised of underpowered studies. According to both *Top 10%* and *PET-PEESE*, 29.6% of these 159 research areas do not have an adequately powered study and over a quarter are 99% underpowered when using *WLS-FE*. Nonetheless, even in areas of research without a single adequately powered estimate, many estimates are reported to be statistically significant – 44%, on average, and ranging from 18% to 74%. Although this lack of adequate power should be a serious concern to economists, other fields are as bad or worse. Among 14,886 meta-analyses in the Cochrane Database of Systematic Reviews of medical research, Turner *et al.* (2013) find that 86% do not contain a single study with adequate power (≥80%) when the random-effects estimate serves as the summary estimate of effect.

Histograms of the proportion of estimates that are adequately powered in these 159 research areas reveal how poorly powered economics is, regardless of which approach

Table 1

Proportions of Empirical Economic Results with Adequate Statistical Power
(n = 159 Research Areas)

	WLS-FE	Top 10%	Top 1	PET-PEESE
	(1)	(2)	(3)	(4)
Median proportion (%)	10.5	6.5	1.9	5.8
Mean proportion (%)	21.9	20.1	22.1	20.1

Notes. WLS-FE denotes weighed least squares fixed effect. *Top 10%* is the top 10% of most precise estimates. *Top 1* is the most precise estimate. *PET-PEESE* denotes the precision-effect test and precision-effect estimate with standard error conditional correction for publication bias.
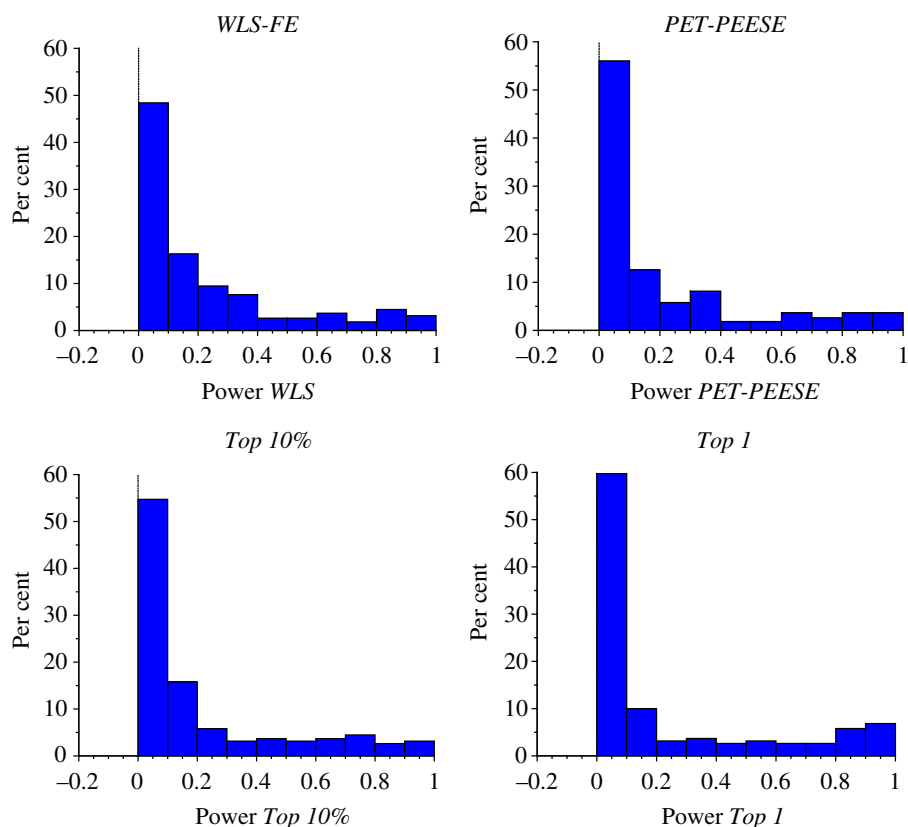
Fig. 1. *Histograms of Adequately Powered Estimates in Empirical Economics*
Note. Colour figure can be viewed at wileyonlinelibrary.com.

is taken (see Figure 1). Taking the more conservative approach first, *WLS-FE*, nearly half of economics research areas have 10% or fewer of their estimates adequately powered (or, in other words, nearly 90% are underpowered) and about two-thirds have 80% of their findings underpowered. Correcting for potential bias using the conditional *PET-PEESE* approach tells roughly the same story. The proportion underpowered worsens, using either *Top 10%*, where 55% have at least 90% of their estimates underpowered and 70% have at least 80% of their estimates underpowered or *Top 1* (60% and 70% respectively). Only 14% to 21% of economics research areas have adequate power for at least half of the reported estimates.

We also calculate the median power of each of these 159 areas of economics using the same four approaches to estimating the underlying 'true' effect: *WLS-FE, PET-PEESE, Top 10%, and Top 1* – see Figure 2. The median power of these 159 median powers is between 10% and 18%, while the mean of these medians are 30–33%. As before, these median powers reveal that a substantial proportion of empirical economics has very low statistical power. The typical power in empirical economics is under 10% in 35% to 52% of these areas of research; between 53% and 64% have median power less than 20%. In Figure 2, we report the typical power found across
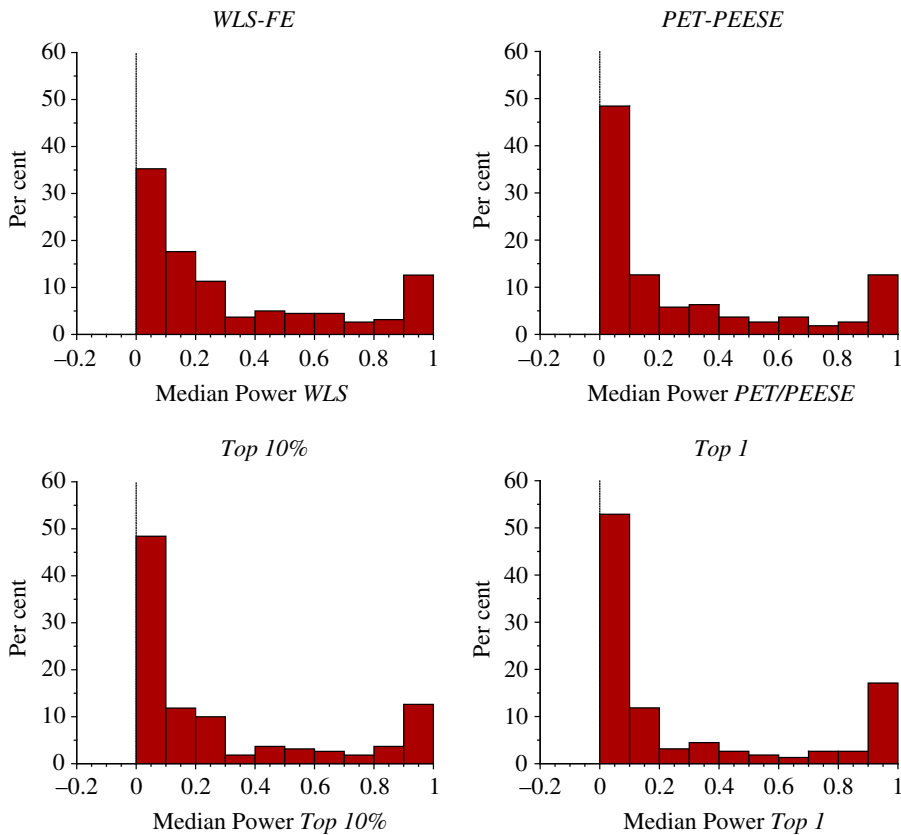
Fig. 2. *Histograms of Median Power in Empirical Economics*
Note. Colour figure can be viewed at wileyonlinelibrary.com.

these 64,076 estimates, 6,700 studies and 159 meta-analyses, rather than the proportion of them that have adequate power. The vast majority empirical economics estimates have less than a 50/50 chance of identifying the phenomenon that it investigates. In only about one-quarter of the areas of empirical economics surveyed does the typical estimate have a 50/50 or better chance of finding what it seeks (*WLS-FE*). This situation is reminiscent of the low power that also characterises many biomedical research fields, e.g. half of the studies in neuroscience have average statistical power less than 21% (Button *et al.*, 2013). When random effects is used as a proxy for the true effect, the median power among 14,886 meta-analyses in the Cochrane Database of Systematic Reviews to detect this effect is only 8% (Turner *et al.* (2013). Thus, medical research has lower power overall than economics because random effects will, on average, give a larger estimate than *WLS-FE* and hence higher power, *ceteris paribus* (Stanley and Doucouliagos, 2015).

While in many disciplines there has been mounting attention to the issue of statistical power (e.g. Maxwell, 2004), this has yet to develop in economics. Hence, there is little pressure on researchers to raise the statistical power of their research

findings. On the other hand, sample sizes have increased, especially with the availability of panel data and large surveys. This will increase power, *ceteris paribus*.[13]

We are able to explore whether power has changed over time by regressing the standard error on a linear time trend for a sub-sample of 55 meta-analyses that reported the average year of the data used, involving 2,992 empirical studies with 36,245 estimates of partial correlations or elasticities. For 55% of these 55 meta-analyses, we find a statistically significant increase in statistical power, and for a further 11% we find a statistically non-significant increase in statistical power. Interestingly, there is a statistically significant reduction in statistical power for 16% of these areas and a further 18% with a statistically non-significant reduction in statistical power. Needless to say, these are merely descriptive time trends and thereby say nothing about potential causes.

3.1. *Finding the power to reduce bias*

> However, it is typically important to have accurate estimates of the size of effects; it is rarely sufficient merely to know that an effect exists.
>
> Schmidt and Hunter (2015, p. 515).

If the absence of power is at the core of publication and reporting bias, can power be its solution? We propose a simple WLS weighted average that uses only adequately powered estimates to correct or reduce reporting or selection biases should such artefacts be present. That is, we calculate *WLS-FE* on that sub-sample of the research record that is adequately powered. This weighted average of the adequately powered estimator (*WAAP*) goes against the grain of conventional meta-analysis, which regards all comparable estimates as precious and combines all estimates so that aggregate power can be increased.[14] But not all estimates are equally biased or equally informative. *WAAP* provides a conservative approach to accommodating potential bias using only the most informative and potentially least biased portion of the research record. If there is selective reporting bias, then *WAAP* will also be biased in the same direction but less so than those summary statistics that include all the low-powered estimates. If there are no selective biases in a research literature, the weighted average of the adequately powered will remain unbiased and lose only a modest amount of efficiency (Stanley et al., 2017).[15] The motivation for *WAAP* is that the reduction of bias will outweigh the efficiency lost through discarding low-powered estimates when there

---

[13] We find no statistically significant differences in power between studies published in economics and non-economics journals (e.g., management or business related journals), no difference by the journal impact factor (using SSCI journal impact factors), or between alternate measures of effect (e.g. between partial correlations and elasticities). However, we do find that microeconomics research has more power, on average, than other research areas; fields that run multiple tests have slightly less power than fields that run a single test.

[14] Indeed, one advantage of meta-analysis is that it can provide a meaningful summary of the evidence base even if each individual study lacks adequate power (Schmidt and Hunter, 2015, p. 11). Against this grain, Stanley *et al.* (2010) document how it is likely to be better to 'discard 90% of the data' when there is the threat of publication bias.

[15] It is unlikely that larger sample studies will be correlated with larger bias. On the contrary, we find a clear pattern where larger studies report findings of a smaller magnitude; that is, one nearer to zero. This pattern is the opposite side of publication, reporting or small-sample bias. Authors of large-sample studies are under less pressure to report larger effects to obtain statistical significance.

is publication selection or reporting bias.[16] At a minimum, *WAAP* offers a robustness check for existing meta-regression publication bias reduction methods (e.g. *PET-PEESE*).

Recall that bias is the expected value of the difference between an estimate or estimator and the 'true' effect. We seek to approximate this theoretical magnitude, empirically. Empirically, expected values are approximated by simple averages and bias by the average difference of many estimates from some proxy of 'true' effect. Here, we use *WAAP* as a conservative benchmark for 'true' effect. One might question such an empirical assessment of bias as a mere coincidence if it came from a single study or even from a single meta-analysis across some hundreds of studies. However, a clear pattern of bias found across 6,700 studies, 64,000 estimates and 159 areas of economics research cannot be dismissed as a mere fluke.

As discussed above, using *WAAP* as the proxy for true effect provides a conservative assessment of bias. *WAAP* will be biased in the same direction as a single estimate if there is selective reporting or small-sample bias, only less so, because *WAAP* gives less weight or no weight to the lower powered estimates that tend to have larger reporting bias (Stanley *et al.*, 2017). Thus, the average difference between *WAAP* and all of the reported estimates will under-estimate the 'true' magnitude of these reporting or publication biases when some selected residual bias remains. For the lack of a better term, we call the systematic exaggeration of the average reported estimate relative to *WAAP*, 'research inflation', see Figure 3. It is well known that selection for statistical significance in the 'right' direction or for results with a sign consistent with dominant theory will cause the magnitude of reported effects to be exaggerated (Stanley, 2005, 2008; Ioannidis and Trikalinos, 2007; Stanley *et al.*, 2010; Doucouliagos and Stanley, 2013; Stanley and Doucouliagos, 2014). 'Research inflation' is the empirical lower bound of the magnitude of this exaggeration of reported results for a given area of economics research.
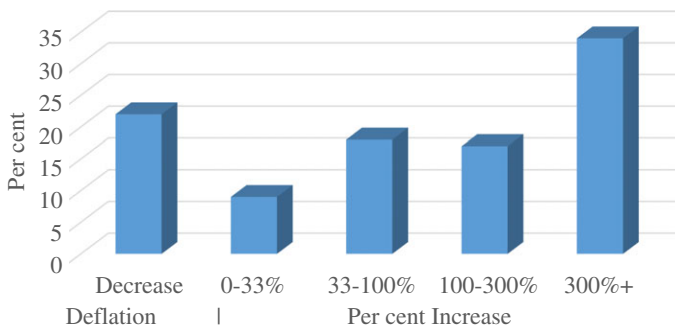

Fig. 3. *Research Inflation, Relative to WAAP*
Note. Colour figure can be viewed at wileyonlinelibrary.com.

---

[16] Simulations confirm that *WAAP* consistently reduces bias and mean square error relative to weighted or unweighted averages that use the entire research record when there is selective reporting bias. When there is no selective reporting, there are no practical differences among these weighted averages, including *WAAP* (Stanley et al., 2017).

We compare this conservative, corrected meta-average (*WAAP*) to the average of all reported results in each area of research. *WAAP* can be calculated in only 128 of 159 meta-analyses, because the remaining 31 do not have a single adequately powered study. We find that 34% of these average effects are exaggerated by a factor of 4 or more (research inflation ≥300%) compared with *WAAP*, 51% are exaggerated by factor of 2 or more (research inflation ≥100%), and another 18% are inflated somewhere between 33% and 100%, see Figure 3. Needless to say, revising reported effects downwards by a factor two, which is needed in most of economics, or by four, which is required for one-third of these areas of economics research, would often have large practical consequences for policy and practice.

Of the 128 areas of research where *WAAP* can be calculated, 17% of the effect sizes are revised upward. That is, *WAAP* finds a 'deflation' of research rather than an exaggeration in 17% of these areas of research. However, the magnitude of deflation is typically quite small and of little practical consequence. In only four areas of research is *WAAP* enough larger than the mean by enough to have practical implications on how we might regard the empirical effect in question. For the effect of ownership on corporate hedging (Arnold *et al.*, 2014), a negligible average correlation (−0.06) is revised to be a small one (−0.27). Also in Arnold *et al.* (2014), the average effect of R&D expenditures on corporate hedging increases the correlation and reverses its sign, from −0.25 to 0.32. However, in both of these cases, the revision is based on only one adequately powered estimate. As a result, *WAAP* is unreliable and its confidence interval is undefined.[17] For Nataraj *et al.* (2013), a negligible impact of government regulations on employment (partial correlation = 0.008) becomes small (0.14) but statistically insignificant. Thus, for these three above areas where *WAAP* revises an effect size upward by a potentially notable magnitude, the revision makes no statistically significant difference. The lone exception involves Doucouliagos and Paldam's (2006) meta-analysis of the effect of foreign aid on investment. Here, *WAAP* increases a small average correlation (0.13) to 0.55, which is a substantial practical and statistically significant, upward revision.[18]

To understand the practical consequences of accounting for power in economic research better, it might be instructive to consider some specific areas. For example, of the 1,474 reported estimates of the employment elasticity of a US minimum wage increase (Doucouliagos and Stanley, 2009), 96% are underpowered and the median power is 8.5%. The weighted average elasticity of the 60 adequately powered estimates is −0.0113, less than one-tenth (6.5%) of the reported average (−0.19) across all of these 1,474 estimates. As a second example, consider the 39 estimates of the value of a statistical life (Doucouliagos *et al.*, 2012), 74% are underpowered. The *WAAP* estimate of the 10 adequately powered studies is $1.47 million compared to the simple average of $9.5 million across all 39. Of the 110 reported price elasticities of residential water

---

[17] We believe that it is unwise to calculate WAAP on a single estimate. We do so here for nine meta-analyses in an effort to be as inclusive and as comprehensive as possible. In practice, we recommend that meta-analysts calculate a *WAAP* only if there are two or more adequately powered estimates. Otherwise, its standard error is undefined (df = 0) and its confidence interval cannot be computed.

[18] Not to dismiss this 'exception that proves the rule', *WAAP*'s large upward revision can be traced to just three values all from a single study that estimates the impact of aid on investment controlling only for savings. Hence, the aid-investment relation is most likely misspecified in this article.

demand (Dalhuisen *et al.*, 2003), 84% are underpowered. The weighted average elasticity of the 10 estimates that have adequate power is -0.1025, while the average across all 110 is -0.378. This means that for these three research areas, 94%, 84.5% and 72.9%, or more, of the average reported effects are likely to be biased.

For a second assessment of systematic residual bias in the economic research record, we compare the *PET-PEESE* correction for publication bias to the average reported effect in these 159 meta-analyses. One advantage of *PET-PEESE* is that it can be applied to all areas of research, regardless of statistical power or whether any single estimate is adequately powered. Generally, these two approaches give practically equivalent overall estimates of effect. The median absolute difference between them is 0.009. The exact meaning of all of the absolute differences between *PET-PEESE* and *WAAP* is somewhat mixed, because different areas of research use different metrics for empirical effect (e.g. partial correlations, elasticities, millions of dollars etc.). In spite of these distinctions in measurement, differences between *WAAP* and *PET-PEESE* that we observe here are all the more remarkably small. Eighty per cent of the empirical estimates from these 159 meta-analyses have been converted into either elasticities or to correlations by the authors to maximise coverage and comparability (Stanley and Doucouliagos, 2012). Among elasticities, the median absolute difference is 0.009, among correlations this difference is 0.011, and it is 0.010 when these 103 meta-analyses of correlations and elasticities are mixed together. In economics research, there is remarkable consistency between these two approaches to reducing or accommodating publication bias. This, perhaps, is the most surprising finding in our present study. To see this consistency more clearly, we use *PET-PEESE* to predict *WAAP* (Figure 4). *PET-PEESE* explains 97% of the variation in *WAAP* across all of these research areas, and the slope is 0.996.

Figure 5 employs *PET-PEESE* to gage research inflation and to serve as a robustness check for the similar assessment of residual bias when *WAAP* is employed. Now, 38%,
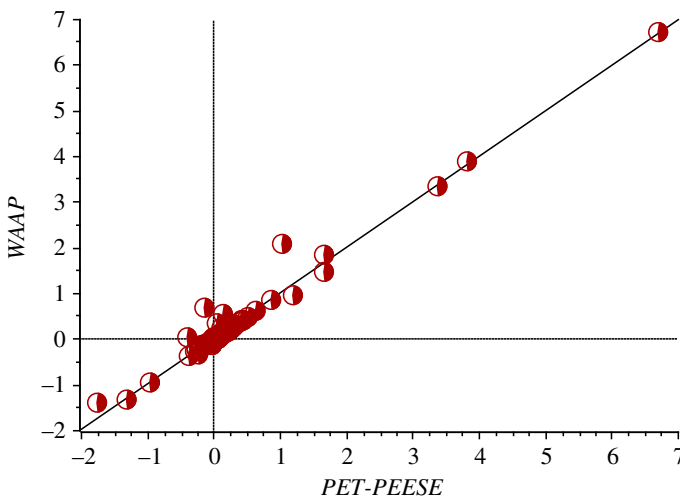

Fig. 4. *Scatter Diagram of the WAAP and PET-PEESE Corrected Estimates (n = 128)*
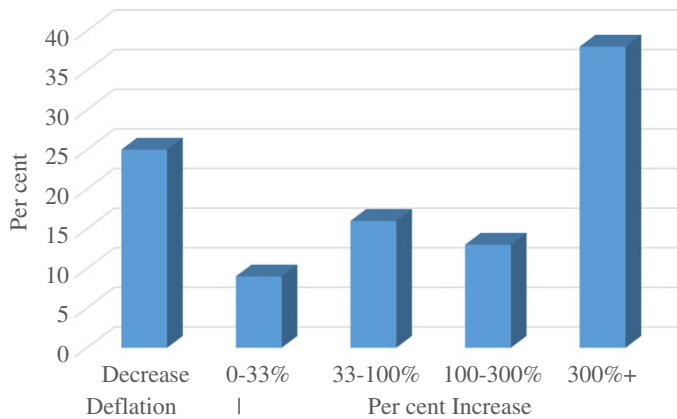*Note.* Colour figure can be viewed at wileyonlinelibrary.com.

Fig. 5. *Research Inflation, Relative to PET-PEESE*
Note. Colour figure can be viewed at wileyonlinelibrary.com.

rather than 34%, of these areas of research are exaggerated by factor of 4 or more, and both approaches to accommodating potential selective reporting find that exactly 51% are exaggerated by factor of 2 or more (see Figure 5). By either assessment, over one-third of the average results of economics research are exaggerated by a factor of more than four, and the majority of reported research is at least twice too large. Such high inflation of research results has practical consequences to what we believe to be true about economic phenomena and policy.

PET-PEESE revises 25% of empirical economics upward. Like *WAAP*, these upward revisions are usually quite small. However, for four areas of research, this upward revision can be considered of practical consequence and thereby deserves specific attention. In three of these cases, the number of estimates combined in the meta-analysis is very small, $n = \{4, 6 \text{ and } 10\}$. In two of these small meta-analyses, *PET-PEESE* revises a negligible average effect to be a small one. In the third of these small-sample meta-analysis (Chliova *et al.*, 2014), *PET-PEESE* revises a small positive effect of microcredit on venture survival into a small negative one. However, with only four observations, the *PET-PEESE* confidence interval includes anything from large negative to large positive effects. Lastly, *PET-PEESE* revises an average price elasticity of -0.80 for food consumption upward to -0.96 but it remains statistically inelastic (Green *et al.*, 2013). Thus, as before, upward corrections are relatively modest compared to the typical reduction that *PET-PEESE* makes.

What is the source of the systematic exaggeration of empirical economics? Although sampling error, misspecification biases, heterogeneity and small-sample bias can make these two sets of estimates (the reported simple average *versus* a meta-average, *WAAP* or *PET-PEESE*) somewhat different for any given area of research, only small-sample bias, selected bias and/or selection bias can cause the most precise estimates to be systematically smaller than the rest. Our survey of 159 quantitative surveys of economics justifies the widespread concern in the potentially distorting effect of publication, selection or, more generally, reporting biases that may be more prominent in the less precise studies.

4. Conclusion

Our survey of 159 meta-analyses of economics reveals that empirical economics research is often greatly underpowered. Regardless of how 'true' effect is estimated, typical statistical power is no more than 18% and nearly half of the areas surveyed have 90% or more of their reported results stemming from underpowered studies. This survey also identifies widespread bias. The majority of the average effects in the empirical economics literature are exaggerated by a factor of at least 2 and at least one-third are exaggerated by a factor of 4 or more.

Impotence begets bias. Is bias an unintended consequence of a preference for statistical significance? Without power, how else can statistical insignificance be overcome? Or as Ziliak and McCloskey (2004) find, researchers may not be cognisant of statistical power. Either way, the outcome is the same. Empirical economics has low power and much residual bias. However one chooses to view our findings, the typical economics result reported by any single study is not very credible, and its magnitude needs to be reduced, typically by half or more, rather than taken at face value.

People respond to incentives and economic researchers know that the incentives for publication and promotion are often perverse. Fostering a culture of replication is of critical importance to genuine scientific advancement. Creating incentives that reward replication is an important necessary step to increasing credibility. Many options are available to move forward, as summarised by Ioannidis (2014) and Maniadis *et al.* (2015). For example, sharing data and codes could be valued by promotion and tenure committees. While sharing in economics is already occurring (e.g. as a pre-requisite for publication in many major journals, such as the *American Economic Review* and *Econometrica*), it is unclear whether sharing policies are fully enforced and how often data and codes are, in fact, shared (Alsheikh-Ali *et al.*, 2011; Duvendack *et al.*, 2015). Publishing, by itself, could be allotted zero value in the absence of replication and validation. Quality and reproducibility of published work may need to be valued over simple quantity of publications before meaningful change would likely occur (Ioannidis and Khoury, 2014). Researchers could be rewarded for taking the effort to construct new data sets and to share them. Unwanted consequences of sharing may also need to be considered (Ioannidis, 2016; Maniadis *et al.*, 2015). Encouraging researchers to combine their data would increase sample size and thereby statistical power (Button *et al.*, 2013). Combined analyses using raw, individual-level data from all teams working in a field, perhaps as part of large collaborations, networks or consortia, could also help eliminate selective reporting biases. Selection and reporting biases could be reduced further still if all research teams were to agree upfront on how models are to be specified and how data are to be analysed and reported. Such collaborative analyses could still respect the fact that different studies on the same topic deal with different populations, settings or other factors that introduce heterogeneity. These sources of heterogeneity can be formally explored in a more efficient manner in large-scale collaborative analyses than in single studies.

Altering incentives towards increasing statistical power is important for economic science. Economics needs a power boost. Besides promotion committees at institutions which are likely to change their practices slowly, other stakeholders may affect research (e.g. funders and journals). Funders could routinely request power calculations for

proposed studies and should avoid funding underpowered studies, unless there is some specific justification for an underpowered study. Power calculations would need to be realistic and should consider systematically past studies that may inform the design of the proposed study. Systematic review and meta-analysis of past studies are necessary regardless. They can indicate whether a new study is even needed. Funders could also request investigators to specify whether the proposed work is exploratory or not. If not exploratory, pre-specification of the protocol, including model specifications and the analyses to be conducted might greatly reduce selective reporting biases downstream.

Journals could also adopt similar standards and report whether a study was exploratory or not, whether a systematic review of the field has been performed (if so, what it shows) and whether power calculations had been performed (if so, report the assumptions and the parameters upon which they are based). Several options have been proposed on how journals can respond to enhance reproducible research (Nosek *et al.*, 2015).

We are not suggesting that underpowered studies go unpublished; such a strategy would put pressure on investigators to report unrealistic and inflated power estimates based on spurious assumptions. We merely seek greater disclosure. Study quality is multidimensional and an underpowered study can provide important insights. Indeed, techniques such as meta-analysis can increase power by combining underpowered studies. It has been argued that it is sometimes preferable to have two or more studies with modest power (e.g. 50–60%) rather than just a single adequately powered study with greater than 80% power (IntHout *et al.*, 2016). However, grossly underpowered studies need some justification for their existence and most currently underpowered studies have no such justification. From a meta-analysis perspective, examination of all studies (regardless of their power) might allow the meta-analyst to identify an important source of misspecification bias or genuine heterogeneity widely contained in the research record. Nevertheless, we urge authors and journals to caution readers if a study is substantially underpowered.

Furthermore, careful meta-analyses are helpful in summarising the evidence, identifying sources of systematic heterogeneity and perhaps even reducing biases. Our study cannot address how often such meta-analyses are feasible in the economics literature. The topics that we analysed here are those where meta-analyses have already been conducted. For some economics topics, there is probably only one or a few studies and for most economic hypotheses there is as yet no meta-analysis available. Moreover, replication remains rare in economics (Hamermesh, 2007; Duvendack *et al.*, 2015). For example, Duvendack *et al.* (2015) surveyed 333 economics journals, finding that only 10 explicitly welcome replication studies. The lack of replication studies is common in many other social science fields, for example, psychological science (Ioannidis, 2012). The credibility of the results on topics with one study or with evidence that has not been systematically reviewed and meta-analysed may vary on a case-by-case basis. However, on average, credibility is quite likely to increase with additional studies and meta-analyses.

Replication studies are unlikely to increase unless there are incentives for their conduct, i.e. funds for their performance and the ability to publish them in respectable journals. Psychology offers a transformative example in this regard, since replication has now become highly popular, while it was previously given little or no credit. A replication effort of 100 high-profile psychology studies has been hailed as one of the

most influential papers in the history of psychological science (Open Science Collaboration, 2015). Similar replication efforts, albeit at smaller scale, have already started and prominently published in experimental economics (Kammerer *et al.*, 2016).

Although important, meta-analysis is not a panacea. Its results may have only modest credibility, especially when they come from largely small and/or biased studies (Pereira and Ioannidis, 2011). The corrections that we proposed are not perfect. The integrity of a given area of research must remain in some doubt until large studies, using unassailable methods and reliable data are conducted. High precision does not necessarily equate with high quality or reliability of the data. Moreover, as we documented, even among the topics where meta-analyses have been performed, a sizable proportion have no studies that are adequately powered. All these factors suggest that the credibility of the economics literature may be even worse than what we have estimated from 159 meta-analyses or than what was suggested recently by Ioannidis and Doucouliagos (2013). Proactive improvements in the design of single studies and in the setting of the wider research agenda to improve power, reduce opportunities for biases and to enhance a culture of replication and reproducibility would help.

Appendix A. Surveyed Areas

International economics	Labour economics	Growth and development	Microeconomics	Macroeconomics	Finance and public economics	Other
Common currency [65, 105] (95)	CEO pay [42, 107] (768)	Aid [10, 45–47] (3392)	Air travel demand [59] (532)	Bank competition & stability [124] (598)	Capital structure & tax [56] (1021)	Audit committee [73, 102] (925)
Distance & trade [39, 69] (2738)	Efficiency wages [83] (75)	Convergence [2] (610)	Alcohol demand [97–98, 100, 114] (1810)	Business cycle correlation [57] (460)	Corporate hedging [8] (113)	Corporate social responsibility [7] (82)
Exchange rates [64] (1255)	Employee participation [40] (67)	Education [1, 17, 58, 63, 110] (2838)	Cigarette demand [96] (32)	Capital-energy substitution [82] (631)	FDI & taxation [55] (704)	Determinants of voting [90] (641)
Foreign direct investment [19–20, 38, 61, 67–68, 76–78] (10383)	Immigration [89] (315)	Exports [116] (432)	Firm performance [14, 22, 104, 111, 113, 120–122] (942)	Central bank independence [80] (366)	Financial forecast errors [60] (26)	Incentives in surveys [11] (363)
Globalisation & tax [3] (233)	Labour demand [87] (890)	Finance [21, 88, 119] (1973)	Food demand [54, 62, 110] (447)	Habit formation [72] (558)	Property tax [94] (42)	Poverty and crime [75] (76)
Trade agreements [4, 28–30, 94] (2944)	Labour supply [6, 24] (53)	Gender [103] (16)	Gasoline demand [70–71] (903)	Inter-temporal substitution [66] (2734)	Stock market efficiency [79] (1719)	
	Marriage wage premium [36] (661)	ICT [115] (415)	Health [9, 13, 91–92, 112] (169)	Okun's law [101] (264)	Inter-jurisdictional competition [32–33] (1333)	
	Minimum wage [16, 25, 37, 48] (4418)	Inequality [35] (409)	Innovation [23, 51, 106, 118] (1491)	Price puzzle [108] (640)		
	Regulation [95] (93)	Institutions [5, 50, 53, 85] (857)	Ultimatum game [117] (97)	Public capital productivity [18] (578)		
	Unions [41, 43–44, 84] (570)	Microcredit [26, 52, 123] (232)	Value of statistical life [15, 49] (140)	Taylor rule [27] (1673)		
	Wage flexibility [12, 31, 99] (2004)	Natural disasters [86] (1872)	Water demand [34] (109)	Uncertainty and investment [81] (767)		
		Population [74] (480)				

Notes. Figures in square (round) brackets refer to the meta-analysis reference number (see Appendix B) and number of estimates respectively. Listed areas are broad categories. Listed areas sum to 124 not 159, as some areas report several meta-analyses.

Appendix B. List of Articles with Included Meta-analyses (*denotes article reporting more than one meta-analysis)

1. Abdullah, A.J., Doucouliagos, C.(H.) and Manning, L. (2015). 'Does education reduce inequality? A meta-regression analysis', *Journal of Economic Surveys,* vol. 29(2), pp. 301–16.

2. Abreu, M., de Groot, H.L.F. and Florax, R.J.G.M. (2005). 'A meta-analysis of β-convergence: the legendary 2%', *Journal of Economic Surveys,* vol. 19(3), pp. 389–420.

3. Adam, A., Kammas, P. and Lagou, A. (2013). 'The effect of globalization on capital taxation: what have we learned after 20 years of empirical studies?', *Journal of Macroeconomics,* vol. 35(4), pp. 199–209.

4. Afesorgbor, S.K. (2013). 'Revisiting the effectiveness of African economic integration: a meta-analytic review and comparative estimation methods', Aarhus University, Department of Economics and Business Working Paper 2013-13.

5. Ahmadov, A.K. (2014). 'Oil, democracy, and context: a meta-analysis', *Comparative Political Studies,* vol. 47(9), pp. 1238–67.

6. Akgunduz, Y.E. and Plantenga, J. (2011). 'Child care prices and female labour force participation: a meta-analysis', Tjalling C. Koopmans Research Institute, Discussion Paper Series 11-08.

7. Allouche, J. and Laroche, P. (2005). 'A meta-analytical investigation of the relationship between corporate social and financial performance', *Revue de Gestion des Ressources Humaines,* vol. 57, pp. 18.

8. *Arnold, M.M., Rathgeber, A.W. and Stöckl, S. (2014). 'Determinants of corporate hedging: a (statistical) meta-analysis', *Quarterly Review of Economics and Finance,* vol. 54(4), pp. 443–58.

9. Asenso-Boadi, F., Peters, T.J. and Coast, J. (2008). 'Exploring differences in empirical time preference rates for health: an application of meta-regression', *Health Economics,* vol. 17(2), pp. 235–48.

10. *Askarov, Z. and Doucouliagos, H. (2013). 'Does aid improve democracy and governance? A meta-regression analysis', *Public Choice,* vol. 157(3), pp. 601–28.

11. Auspurg, K. and Schneck, A. (2014). 'What difference makes a difference? A meta-regression approach on the effectiveness conditions of incentives in self-administered surveys', Paper presented at the MAER-Colloquium, Athens, September 2014.

12. Babecky, J., Ramos, R. and Sanromá, E. (2008). 'Meta-analysis on microeconomic wage flexibility (Wage Curve)', *Sozialer Fortschritt,* vol. 57(10), pp. 273–79.

13. Bassani, D.G., Arora, P., Wazny, K., Gaffey, M.F., Lenters, L. and Bhutta, Z.A. (2013). 'Financial incentives and coverage of child health interventions: a systematic review and meta-analysis', *BMC Public Health,* vol. 13(3), pp. S30.

14. Bel, G., Fageda, X. and Warner, M.E. (2010). 'Is private production of public services cheaper than public production? A meta-regression analysis of solid waste and water services', *Journal of Policy Analysis and Management,* vol. 29(3), pp. 553–77.

15. Bellavance, F., Dionne, G. and Lebeau, M. (2009). 'The value of a statistical life: a meta-analysis with a mixed effects regression model', *Journal of Health Economics,* vol. 28(2), pp. 444–64.

16. Belman, D. and Wolfson, P.J. (2014). *What does the minimum wage do?* Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

17. Benos, N. and Zotou, S. (2014). 'Education and economic growth: a meta-regression analysis', *World Development,* vol. 64(December), pp. 669–89.

18. Bom, P.R.D. and Ligthart, J.E. (2014). 'What have we learned from three decades of research on the productivity of public capital?', *Journal of Economic Surveys,* vol. 28(5), pp. 889–916.

19. *Bruno, R.L. and Campos, N.F. (2011). 'Foreign direct investment and economic performance: a systematic review of the evidence uncovers a new paradox', Final report for the Department for International Development systematic reviews programme. February 28 2011.

20. Bruno, R.L. and Cipollina, M. (2014). 'The impact of foreign direct investment on economic performance in the enlarged Europe: a meta-regression analysis', in (Y. Temouri and C. Jones, eds), *International Business and Institutions after the Financial Crisis*, Palgrave Macmillan, pp. 228–50.

21. Bumann, S., Hermes, N. and Lensink, R. (2013). 'Financial liberalization and economic growth: a meta-analysis', *Journal of International Money and Finance,* vol. 33(March), pp. 255–81.

22. Cano, C.R., Carrillat, F.A and Jaramillo, F. (2004). 'A meta-analysis of the relationship between market orientation and business performance: evidence from five continents', *International Journal of Research in Marketing,* vol. 21(2), pp. 179–200.

23. Castellacci, F. and Lie, C.M. (2015). 'Do the effects of R&D tax credits vary across industries? a meta-regression analysis', *Research Policy,* vol. 44(4), pp. 819–32.

24. Chetty, R., Guren, A., Manoli, D.S. and Weber, A. (2011). 'Does indivisible labor explain the difference between micro and macro elasticities? A meta-analysis of extensive margin elasticities', NBER Working Paper No. 16729.

25. Chletsos, M. and Giotis, G.P. (2015). 'The employment effect of minimum wage using 77 international studies since 1992: a meta-analysis', MPRA Paper 61321, University Library of Munich, Germany.

26. *Chliova, M., Brinckmann, J. and Rosenbusch, N. (2014). 'Is microcredit a blessing for the poor? A meta-analysis examining development outcomes and contextual considerations', *Journal of Business Venturing*, vol. 30(3), pp. 467–87.

27. *Chortareas, G. and Magonis, G. (2008). 'What do we learn from Taylor Rule estimations? A meta-analysis', *Ekonomia,* vol. 11(2), pp. 112–38.

28. Cipollina, M. and Pietrovito, F. (2011). 'Trade impact of EU preferential policies: a meta-analysis of the literature', in (L.D. Benedictis and L. Salvatici, eds), *The Trade Impact of European Union Preferential policies: An Analysis through Gravity Models*, Berlin: Springer, Chapter 5, pp. 91–109.

29. Cipollina, M. and Salvatici, L. (2010). 'Reciprocal trade agreements in gravity models: a meta-analysis', *Review of International Economics,* vol. 18(1), pp. 63–80.

30. Cirera, X., Willenbockel, D. and Lakshman, R. (2011). 'What is the evidence of the impact of tariff reductions on employment and fiscal revenue in developing countries? A systematic review', Technical report. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

31. Clar, M., Dreger, C. and Ramos, R. (2007). 'Wage flexibility and labour market institutions: a meta-analysis', *Kyklos,* vol. 60(2), pp. 145–63.

32. Costa-Font, J., De-Alburquerqe, F. and Doucouliagos, C. (2014). 'Do jurisdictions compete on taxes? A meta-regression analysis', *Public Choice,* vol. 161(3), pp. 451–70.

33. Costa-Font, J., De-Alburquerqe, F. and Doucouliagos, H. (2015). 'Does inter-jurisdictional competition engender a "race to the bottom"? A meta-regression analysis, *Economics and Politics,* vol. 27(3), pp. 488–508.

34. Dalhuisen, J.M., Florax, R.J.G.M., de Groot, H.L.F. and Nijkamp, P. (2003). 'Price and income elasticities of residential water demand: a meta-analysis', *Land Economics,* vol. 79(2), pp. 292–308.

35. de Dominicis, L., Florax, R.J.G.M. and de Groot, H.L.F. (2008). 'A meta-analysis of the relationship between income inequality and economic growth', *Scottish Journal of Political Economy,* vol. 55(5), pp. 654–82.

36. de Linde Leonard, M. and Stanley, T.D. (2015). 'Married with children: what remains when observable biases are removed from the reported male marriage wage premium', *Labour Economics,* vol. 33(April), pp. 72–80.

37. de Linde Leonard, M., Stanley, T.D. and Doucouliagos, H. (2014). 'Does the UK minimum wage reduce employment?', *British Journal of Industrial Relations,* vol. 52(3), pp. 499–520.

38. Demena, B.A. (2015). 'Publication bias in FDI spillovers in developing countries: a meta-regression analysis', *Applied Economics Letters*, vol. 22(14), pp. 1170–4.

39. Disdier, A-C. and Head, K. (2008). 'The puzzling persistence of the distance effect on bilateral trade', *Review of Economics and Statistics,* vol. 90(1), pp. 37–48.

40. *Doucouliagos, C.(H.) (1995). 'Worker participation and productivity in labor-managed and participatory capitalist firms: a meta-analysis', *Industrial and Labor Relations Review,* vol. 49(1), pp. 58–77.

41. *Doucouliagos, C., Freeman, R. and Laroche, P. (2016). *The Economics of Trade Unions: A Study of a Research Field and its Findings,* Oxford: Routledge.

42. Doucouliagos, C.(H.), Haman, J. and Stanley, T.D. (2012). 'Pay for performance and corporate governance reform', *Industrial Relations: A Journal of Economy and Society,* vol. 51(3), pp. 670–703.

43. Doucouliagos, C.(H.) and Laroche, P. (2013). 'Unions and innovation: new insights from the cross-country evidence', *Industrial Relations: A Journal of Economy and Society,* vol. 52(2), pp. 467–91.

44. Doucouliagos, C.(H.), Laroche, P. and Stanley, T.D. (2005). 'Publication bias in union-productivity research?' *Relations Industielles/Industrial Relations*, vol. 60(2), pp. 320–44.

45. *Doucouliagos, C.(H.) and Paldam, M. (2006). 'Aid effectiveness on accumulation: a meta study', *Kyklos,* vol. 59(2), pp. 227–54.

46. Doucouliagos, C.(H.) and Paldam, M. (2013*a*). 'The robust result in meta-analysis of aid effectiveness: a response to Mekasha and Tarp', *Journal of Development Studies,* vol. 49(4), pp. 584–7.

47. Doucouliagos, C.(H.) and Paldam, M. (2013*b*). 'Explaining development aid allocation by growth', *Journal of Entrepreneurship and Public Policy,* vol. 2(1), pp. 21–41.

48. Doucouliagos, C.(H.) and Stanley, T.D. (2009). 'Publication selection bias in minimum-wage research? a meta regression analysis', *British Journal of Industrial Relations,* vol. 47(2), pp. 406–28.

49. Doucouliagos, C.(H.), Stanley, T.D. and Viscusi, W.K. (2014). 'Publication selection and the income elasticity of the value of a statistical life', *Journal of Health Economics,* vol. 33(January), pp. 67–75.

50. Doucouliagos, C.(H.) and Ulubasoglu, M.A. (2008). 'Democracy and economic growth: a meta-analysis', *American Journal of Political Economy,* vol. 52(1), pp. 61–83.

51. Duran, P., Kammerlander, N., van Essen, M. and Zellweger, T. (2016). 'Doing more with less: innovation input and output in family firms', *Academy of Management Journal,* vol. 59(4), pp. 1224–1264.

52. Duvendack, M., Palmer-Jones, R. and Vaessen, J. (2014). 'Meta-analysis of the impact of microcredit on women's control over household decisions: methodological issues and substantive findings', *Journal of Development Effectiveness,* vol. 6(2), pp. 73–96.

53. Efendic A., Pugh, G. and Adnett, N. (2011). 'Institutions and economic performance: a meta-regression analysis', *European Journal of Political Economy,* vol. 27(3), pp. 586–99.

54. Escobar, M.A.C., Veerman, J.L., Tollman, S.M., Bertram, M.Y. and Hofman, K.J. (2013). 'Evidence that a tax on sugar sweetened beverages reduces the obesity rate: a meta-analysis', *BMC Public Health*, vol. 13, pp. 1072.

55. Feld, L.P. and Heckemeyer, J.H. (2011). 'FDI and taxation: a meta-study', *Journal of Economic Surveys,* vol. 25(2), pp. 233–72.

56. Feld, L.P., Heckemeyer, J.H. and Overesch, M. (2013). 'Capital structure choice and company taxation: a meta-study', *Journal of Banking and Finance,* vol. 37(8), pp. 2850–66.

57. Fidrmuc, J. and Korhonen, I. (2006). 'Meta-analysis of the business cycle correlation between the Euro Area and the CEECs', *Journal of Comparative Economics,* vol. 34(3), pp. 518–37.

58. Fleury, N. and Gilles, F. (2015). 'A meta-regression analysis on intergenerational transmission of education: publication bias and genuine empirical effect', TEPP Working Paper, No. 15-2.

59. Gallet, C.A. and Doucouliagos, C.(H.) (2014). 'The income elasticity of air travel: a meta-analysis', *Annals of Tourism Research,* vol. 49(November), pp. 141–55.

60. García-Meca, E. and Sánchez-Ballesta, J.P. (2006). 'Influences on financial analyst forecast errors: a meta-analysis', *International Business Review,* vol. 15(1), pp. 29–52.

61. Görg, H. and Strobl, E. (2001). 'Multinational companies and productivity spillovers: a meta-analysis', ECONOMIC JOURNAL, vol. 111(475), pp. F723-39.

62. Green, R., Cornelsen, L., Dangour, A.D., Turner, R., Shankar, B., Mazzocchi, M. and Smith, R.D. (2013). 'The effect of rising food prices on food consumption: systematic review with meta-regression', *British Medical Journal*, 346: f3703.

63. *Guerrero, G., Leon, J., Zapata, M. and Cueto, S. (2013). 'Getting teachers back to the classroom. A systematic review on what works to improve teacher attendance in developing countries', *Journal of Development Effectiveness,* vol. 5(4), pp. 466–88.

64. Haile, M.G. and Pugh, G. (2013). 'Does exchange rate volatility discourage international trade? A meta-regression analysis', *Journal of International Trade and Economic Development,* vol. 22 (3), pp. 321–50.

65. Havránek, T. (2010). 'Rose effect and the euro: is the magic gone?', *Review of World Economics,* vol. 146(2), pp. 241–61.

66. Havránek, T. (2015). 'Measuring intertemporal substitution: the importance of method choices and selective reporting', *Journal of the European Economic Association*, vol. 13(6), pp. 1180–204.

67. Havránek, T. and Irsova, Z. (2011). 'Estimating vertical spillovers from FDI: why results vary and what the true effect is', *Journal of International Economics,* vol. 85(2), pp. 234–44.

68. Havránek, T. and Irsova, Z. (2013). 'Determinants of horizontal spillovers from FDI: evidence from a large meta-analysis', *World Development,* vol. 42, pp. 1–15.

69. Havránek, T. and Irsova, Z. (2015). 'Do borders really slash trade? A meta-analysis', William Davidson Institute Working Paper No. 1088, University of Michigan.

70. Havránek, T., Irsova, Z. and Janda, K. (2012). 'Demand for gasoline is more price-inelastic than commonly thought', *Energy Economics*, vol. 34(1), pp. 201–7.

71. Havránek, T. and Kokes, O. (2015). 'Income elasticity of gasoline demand: a meta-analysis', *Energy Economics,* vol. 47(January), pp. 77–86.

72. Havránek, T., Rusnak, M. and Sokolova, A.V. (2015). Habit formation in consumption: a meta-analysis. Czech National Bank and Charles University, Prague.

73. Hay, D. (2014). 'Meta-regression in auditing research: evaluating the evidence on the big firm premium', University of Auckland – Business School manuscript.

74. Headey, D.D. and Hodge, A. (2009). 'The effect of population growth on economic growth: a meta-regression analysis of the macroeconomic literature', *Population and Development Review,* vol. 35(2), pp. 221–48.

75. Hsieh, C-C. and Pugh, M.D. (1993). 'Poverty, income inequality, and violent crime: a meta-analysis of recent aggregate data studies', *Criminal Justice Review,* vol. 18(2), pp. 182–202.

76. Iamsiraroj, S. (2008). 'A comprehensive analysis of foreign direct investment and economic growth', PhD Dissertation, School of Accounting, Economics and Finance, Deakin University.

77. Iamsiraroj, S. and Doucouliagos, C.(H.) (2015). 'Does growth attract FDI?', *Economics: The Open-Access, Open-Assessment E-Journal,* vol. 9(July), pp. 1–35.

78. Iwasaki, I. and Tokunaga, M. (2014). 'Macroeconomic impacts of FDI in transition economies: a meta-analysis', *World Development,* vol. 61(September), pp. 53–69.

79. Kim, J., Doucouliagos, H., Stanley, T.D. 2014. 'Market efficiency in Asian and Australasian stock markets: a fresh look at the evidence', Deakin University Economics Working Paper, 2014/9.

80. Klomp, J. and de Haan, J. (2010). 'Inflation and central bank independence: a meta-regression analysis', *Journal of Economic Surveys,* vol. 24(4), pp. 593–621.

81. Koetse, M.J., de Groot, H.L.F. and Florax, R.J.G.M. (2006). 'The impact of uncertainty on investment: a meta-analysis', Tinbergen Institute Discussion Paper TI 2006-060/3.

82. Koetse, M.J., de Groot, H.L.F. and Florax, R.J.G.M. (2008). 'Capital-energy substitution and shifts in factor demand: a meta-analysis', *Energy Economics*, vol. 30(5), pp. 2236–51.

83. Krassoi-Peach, E. and Stanley, T.D. (2009). 'Efficiency wages, productivity and simultaneity: a meta-regression analysis', *Journal of Labor Research,* vol. 30(3), pp. 262–68.

84. Laroche P. (2016). 'A meta-analysis of the union-job satisfaction relationship', *British Journal of Industrial Relations*, vol. 54(4), pp. 709–741.

85. *Lawry, S., Samii, C., Hall, R., Leopold, A., Hornby, D. and Mtero, F. (2014). The impact of land property rights interventions on investment and agricultural productivity in developing countries: a systematic review. *Campbell Systematic Reviews*, 2014:1, doi: 10.4073/csr.2014.1.

86. Lazzaroni, S. and van Bergeijk, P.A.G. (2014). 'Natural disasters' impact, factors of resilience and development: a meta-analysis of the macroeconomic literature', *Ecological Economics,* vol. 107(November), pp. 333–46.

87. Lichter, A., Peichl, A. and Siegloch, S. (2015). 'The own-wage elasticity of labor demand: a meta-regression analysis', *European Economic Review*, vol. 80, pp. 94–119.

88. Ljungwall, C. and Tingvall, P.G. (2012*b*). 'Is China different? A meta-analysis of China's financial sector', *Applied Economics Letters,* vol. 20(7), pp. 715–8.

89. *Longhi, S., Nijkamp, P. and Poot, J. (2010). 'Joint impacts of immigration on wages and employment: review and meta-analysis', *Journal of Geographical Systems,* vol. 12(4), pp. 355–87.

90. *Ludvigsen, S. (2009). 'Post-mortem of the VP function? Meta-regression analyses of economic voting in the United Kingdom', PhD Dissertation, Department of Political Science, Aarhus University.

91. Lye, J. and Hirschberg, J. (2010). 'Alcohol consumption and human capital: a retrospective study of the literature', *Journal of Economic Surveys,* vol. 2(4), pp. 309–38.

92. Maidment, C.D., Jones, C.R., Webb, T.L., Hathway, A.E. and Gilbertson, J.M. (2014). 'The impact of household energy efficiency measures on health: a meta-analysis', *Energy Policy,* vol. 65, pp. 583–93.

93. Martin, I.W. (2015). 'What property tax limitations do to local finances: a meta-analysis', UCLA Institute for Research on Labor and Employment. Working Paper No. 2015-12.

94. Moons, S. and van Bergeijk, P.A.G. (2013). A meta-analysis of economic diplomacy and its effect on international economic flows. ISS Working Papers, General Series, No. 566, The Hague: International Institute of Social Studies.

95. Nataraj, S., Perez-Arce, F., and Kumar, K.B. (2014). 'The impact of labor market regulation on employment in low-income countries: a meta-analysis', *Journal of Economic Surveys,* vol. 28(3), pp. 551–72.

96. Nelson, J.P. (2006). 'Cigarette advertising regulation: a meta-analysis', *International Review of Law and Economics,* vol. 26(2), pp. 195–226.

97. Nelson, J.P. (2011). 'Alcohol marketing, adolescent drinking and publication bias in longitudinal studies: a critical survey using meta-analysis', *Journal of Economic Surveys,* vol. 25(2), pp. 191–232.

98. Nelson, J.P. (2014). 'Estimating the price elasticity of beer: meta-analysis of data with heterogeneity, dependence, and publication bias', *Journal of Health Economics,* vol. 33(January), pp. 180–7.

99. Nijkamp, P. and Poot, J. (2005). 'The last word on the wage curve', *Journal of Economic Surveys,* vol. 19(3), pp. 421–50.

100. Oczkowski, E. and Doucouliagos, C. (2015). 'Wine prices and quality ratings: a meta-regression analysis', *American Journal of Agricultural Economics,* vol. 97(1), pp. 103–21.

101. *Perman, R., Stephen, G. and Tavéra, C. (2015). 'Okun's law: a meta-analysis', *Manchester School*, vol. 83(1), pp. 101–26.

102. Pomeroy, B. and Thornton, D.B. (2008). 'Meta-analysis and the accounting literature: the case of audit committee independence and financial reporting quality', *European Accounting Review,* vol. 17(2), pp. 305–30.

103. Quisumbing, A.R. (1995). 'Gender differences in agricultural productivity: a survey of empirical evidence', International Food Policy Research Institute, FCND Discussion Paper No. 5, July 1995.

104. Rhoades, D.L., Rechner, P.L. and Sundaramurthy, C. (2001). 'A meta-analysis of board leadership structure and financial performance: are "two heads better than one"?', *Corporate Governance,* vol. 9(4), pp. 311–9.

105. Rose, A.K. and Stanley, T.D. (2005). 'A meta-analysis of the effect of common currencies on international trade', *Journal of Economic Surveys,* vol. 19(3), pp. 347–65.

106. Rosenbusch, N., Brinckmann, J. and Bausch, A. (2011). 'Is innovation always beneficial? A meta-analysis of the relationship between innovation and performance in SMEs', *Journal of Business Venturing,* vol. 26(4), pp. 441–57.

107. Rost, K. and Osterloh, M. (2009). 'Management fashion pay-for-performance for CEOs', *Schmalenbach Business Review,* vol. 61(4), pp. 119–49.

108. Rusnak, M., Havránek, T. and Horvath, R. (2013). 'How to solve the price puzzle? A meta-analysis', *Journal of Money, Credit and Banking,* vol. 45(1), pp. 37–70.

109. *Saavedra, J.E. and Garcia, S. (2012). 'Impacts of conditional cash transfer programs on educational outcomes in developing countries: a meta-analysis', Rand Working Paper, No. WR-921-1.

110. Santeramo, F.G. and Shabnam, N. (2015). 'The income-elasticity of calories, macro- and micro-nutrients: what is the literature telling us?', *Food Research International*, 76(4), pp. 932–7.

111. Shen, Y-C., Eggleston, K., Lau, J. and Schmid, C.H. (2007). 'Hospital ownership and financial performance: what explains the different findings in the empirical literature?', *Inquiry,* vol. 44(1), pp. 41–68.

112. Sinnott, S-J., Buckley, C., O'Riordan, D., Bradley, C. and Whelton, H. (2013). 'The effect of copayments for prescriptions on adherence to prescription medicines in publicly insured populations: a systematic review and meta-analysis', *PloS One,* vol. 8(5): e64914, pp. 1–11.

113. Stam, W., Arzlanian, S. and Elfring, T. (2014). 'Social capital of entrepreneurs and small firm performance: a meta-analysis of contextual and methodological moderators', *Journal of Business Venturing,* vol. 29(1), pp. 152–73.

114. *Stanley, T.D. and Doucouliagos, C(H.). (2012). *Meta-Regression Analysis in Economics and Business*, Oxford: Routledge.

115. Stanley, T.D., Doucouliagos, C., Steel, P. (2015). 'Does ICT generate economic growth? A meta-regression analysis', Deakin University, Department of Economics, Economics Working Paper, No. 2015/9.

116. Tingvall, P.G. and Ljungwall, C. (2012a). 'Is China different? A meta-analysis of export-led growth', *Economics Letters,* vol. 115(2), pp. 177–9.

117. Tisserand, J.C. (2014). 'The ultimatum game: a meta-analysis of three decades of experimental research', MAER-net Colloquium 2014 Athens, Greece, September.

118. Ugur, M., Solomon, E., Guidi, F. and Trushin, E. (2014). 'R&D and productivity in OECD firms and industries: a hierarchical meta-regression analysis', Evaluation of Research and Development (R&D) Expenditures, Firm Survival, Firm Growth and Employment: UK Evidence in the OECD Context. Reference no ES/K004824/1.

119. Valickova, P., Havránek, T. and Horvath, R. (2015). 'Financial development and economic growth: a meta-analysis', *Journal of Economic Surveys*, vol. 29(3), pp. 506–26.

120. van Essen, M., Carney, M., Gedajlovic, E.R. and Heugens, P.P.M.A.R. (2015). 'How does family control influence firm strategy and performance? A meta-analysis of US publicly-listed firms', *Corporate Governance: An International Review,* vol. 23(1), pp. 3–24.

121. Wagner III, J.A., Stimpert, J.L. and Fubara, E.I. (1998). 'Board composition and organization performance: two studies of insider/outsider effects', *Journal of Management Studies,* vol. 35(5), pp. 655–77.

122. Wang, K. and Shailer, G. (2015). 'Ownership concentration and firm performance in emerging markets: a meta-analysis', *Journal of Economic Surveys,* vol. 29(2), pp. 199–229.

123. Yang, M. and Stanley, T.D. (2012). 'Micro-credit and income: A literature review and meta-analysis', *Bulletin of Economics and Meta-Analysis.* https://www.hendrix.edu/maer-network/default.aspx?id=15206. Accessed August 6th, 2015.

124. Zigraiova, D. and Havránek, T. (2016). 'Bank competition and financial stability: much ado about nothing?', *Journal of Economic Surveys*, vol. 30(5), pp. 944–81.

Stanford University
Hendrix College
Deakin University

Additional Supporting Information may be found in the online version of this article:

Data S1.

References

Alsheikh-Ali, A.A., Qureshi, W., Al-Mallah, M.H. and Ioannidis, J.P.A. (2011). 'Public availability of published research data in high-impact journals', *PLoS ONE*, vol. 6(9), p. e24357.

Andreoni, J., Gillen, B.J. and Harbaugh, W.T. (2013). The power of revealed preference tests: ex post evaluation of experimental design', manuscript, http://people.hss.caltech.edu/~bgillen/papers/GARPPower.pdf. (last accessed: 8 January 2017).

Bakker, M., van Dijk, A. and Wicherts, J.M. (2012). 'The rules of the game called psychological science', *Perspectives on Psychological Science*, vol. 7(6), pp. 543–54.

Banerjee, A., Karlan, D. and Zinman, J. (2015). 'Six randomized evaluations of microcredit: introduction and further steps', *American Economic Journal: Applied Economics*, vol. 7(1), pp. 1–21.

Begg, C.B. and Berlin, J.A. (1988). 'Publication bias: a problem in interpreting medical data', *Journal of the Royal Statistical Society, Series A*, vol. 151(3), pp. 419–63.

Begley, C.G. and Ellis, L.M. (2012). 'Drug development: raise standards for preclinical cancer research', *Nature*, vol. 483(March), pp. 531–3.

Begley, C.G. and Ioannidis, J.P.A. (2015). 'Reproducibility in science: improving the standard for basic and preclinical research', *Circulation Research* vol. 116(1), pp. 116–26.

Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J. and Munafò, M.R. (2013). 'Power failure: why small sample size undermines the reliability of neuroscience', *Nature Reviews: Neuroscience*, vol. 14(May), pp. 365–76.

Card, D. and Krueger, A.B. (1995). 'Time-series minimum-wage studies: a meta-analysis', *American Economic Review: Papers and Proceedings*, vol. 85(2), pp. 238–43.

Cohen, J. (1965). 'Some statistical issues in psychological research', in (B.B. Wolman, ed.), *Handbook of Clinical Psychology*, pp. 95–121, New York: McGraw-Hill.

Cohen, J. (1988). *Statistical Power Analysis in the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Erlbaum.

Cooper, H.M. and Hedges, L.V., (eds.). (1994). *Handbook of Research Synthesis*, New York: Russell Sage.

Copas, J. (1999). 'What works?: selectivity models and meta-analysis', *Journal of the Royal Statistical Society, Series A*, vol. 162(1), pp. 95–109.

Dalhuisen, J.M., Florax, R.J.G.M., de Groot, H.L.F. and Nijkamp, P. (2003). 'Price and income elasticities of residential water demand: a meta-analysis', *Land Economics*, vol. 79(2), pp. 292–308.

De Long, J.B. and Lang, K. (1992). 'Are all economic hypotheses false?', *Journal of Political Economy*, vol. 100(6), pp. 1257–72.

Doucouliagos, C.(H.) and Stanley, T.D. (2009). 'Publication selection bias in minimum-wage research? A meta-regression analysis', *British Journal of Industrial Relations*, vol. 47(2), pp. 406–28.

Doucouliagos, C.(H.) and Stanley, T.D. (2013). 'Are all economic facts greatly exaggerated? Theory competition and selectivity', *Journal of Economic Surveys*, vol. 27(2), pp. 316–39.

Doucouliagos, C.(H.)., Stanley, T.D. and Giles, M. (2012). 'Are estimates of the value of a statistical life exaggerated?' *Journal of Health Economics*, vol. 31(1), pp. 197–206.

Duvendack, M., Palmer-Jones, R.W. and Reed, W.R. (2015). 'Replication in economics: a progress report', *Econ Journal Watch*, vol. 12(2), pp. 164–91.

Egger, M., Smith, G.D., Schneider, M. and Minder, C. (1997). 'Bias in meta-analysis detected by a simple, graphical test', *British Medical Journal*, vol. 315(629), pp. 629–34.

Glass, G.V., McGaw, B. and Smith, M.L. (1981). *Meta-analysis in Social Research*, Beverly Hills, CA: Sage.

Hamermesh, D.S. (2007). 'Replication in economics', NBER Working Paper No. 13026.

Hedges, L.V. and Olkin, I. (1985). *Statistical Methods for Meta-analysis*, Orlando, FL: Academic Press.

Henmi, M. and Copas, J.B. (2010). 'Confidence intervals for random effects meta-analysis and robustness to publication bias', *Statistics in Medicine*, vol. 29(29), pp. 2969–83.

IntHout, J., Ioannidis, J.P.A. and Borm, G.F. (2016). 'Obtaining evidence by a single well-powered trial or several modestly powered trials', *Statistical Methods in Medical Research*, vol. 25(2), pp. 538–52.

Ioannidis, J.P.A. (2005*a*). 'Contradicted and initially stronger effects in highly cited clinical research', *Journal of the American Medical Association*, vol. 294(2), pp. 218–28.

Ioannidis, J.P.A. (2005*b*). 'Why most published research findings are false', *PLoS Medicine*, vol. 2(8). doi: 10. 1371/journal.pmed.0020124.

Ioannidis, J.P.A. (2012). 'Why science is not necessarily self-correcting', *Perspectives on Psychological Science*, vol. 7(6), pp. 645–54.

Ioannidis, J.P.A. (2013). 'Clarifications on the application and interpretation of the test for excess significance and its extensions', *Journal of Mathematical Psychology*, vol. 57(5), pp. 184–7.

Ioannidis, J.P.A. (2014). 'How to make more published research true', *PLoS Medicine*, vol. 11(10). doi: 10. 1371/journal.pmed.1001747

Ioannidis, J.P.A. (2016). 'Anticipating consequences of sharing raw data and code and of awarding badges for sharing', *Journal of Clinical Epidemiology*, vol. 70, pp. 258–60.

Ioannidis, J.P.A. and Doucouliagos, C.(H.) (2013). 'What's to know about the credibility of empirical economics?', *Journal of Economic Surveys*, vol. 27(5), pp. 997–1004.

Ioannidis, J.P.A. and Khoury, M.J. (2014). 'Assessing value in biomedical research: the PQRST of appraisal and reward', *Journal of the American Medical Association*, vol. 312(5), pp. 483–4.

Ioannidis, J.P.A., Hozo, I. and Djulbegovic, B. (2013). 'Optimal type I and type II error pairs when the available sample size is fixed', *Journal of Clinical Epidemiology*, vol. 66(8), pp. 903–10.

Ioannidis, J.P.A. and Trikalinos, T.A. (2007). 'An exploratory test for an excess of significant findings', *Clinical Trials*, vol. 4(3), pp. 245–53.

Leamer, E.E. (1983). 'Let's take the con out of econometrics', *The American Economic Review*, vol. 73(1), pp. 31–43.

Leamer, E.E. (2010). 'Leamer on the state of econometrics', EconTalk Episode with Ed Leamer. Available at: http://www.econtalk.org/archives/2010/05/leamer_on_the_s.html (last accessed: 3 August 2015).

Leamer, E.E. and Leonard, H.B. (1983). 'Reporting the fragility of regression estimates', *Review of Economics and Statistics*, vol. 65(2), pp. 306–17.

Lovell, M.C. (1983). 'Data mining', *Review of Economics and Statistics*, vol. 65(1), pp. 1–12.

Maniadis, Z., Tufano, F. and List, J.A. (2017). 'To replicate or not to replicate? Exploring reproducibility in economics through the lens of a model and a pilot study', ECONOMIC JOURNAL, vol. 127(605), pp. F209–35.

Maniadis, Z., Tufano, F. and List, J.A. (2015). 'How to make experimental economics research more reproducible: lessons from other disciplines and a new proposal', in (C. Deck, E. Fatas and T. Rosenblat, eds.), *Replication in Experimental Economics (Research in Experimental Economics, Volume 18)*, pp. 215–230. Bingley: The Emerald Group Publishing.

Maxwell, S.E. (2004). 'The persistence of underpowered studies in psychological research: causes, consequences, and remedies', *Psychological Methods*, vol. 9(2), pp. 147–63.

McCloskey, D.N. (1985). 'The loss function has been mislaid: the rhetoric of significance tests', *American Economic Review: Papers and Proceedings*, vol. 75(2), pp. 201–5.

McCloskey, D.N. and Ziliak, S.T. (1996). 'The standard error of regressions', *Journal of Economic Literature*, vol. 34(1), pp. 97–114.

Milton, J. (2001). *Complete Poems*. "Comus, a Mask." The Harvard Classics, pp. 1909–14. New York: P.F. Collier & Son. www.bartleby.com/4/. (last accessed 1 August 2017).

Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., Ishiyama, J., Karlan, D., Kraut, A., Lupia, A., Mabry, P., Madon, T., Malhotra, N., Mayo-Wilson, E., McNutt, M., Miguel, E., Levy Paluck, E., Simonsohn, U., Soderberg, C., Spellman, B. A., Turitto, J., VandenBos, G., Vazire, S., Wagenmakers, E. J., Wilson, R. and Yarkoni, T. (2015). 'Promoting an open research culture', *Science*, vol. 348(6242), pp. 1422–5.

Pereira, T.V. and Ioannidis, J.P.A. (2011). 'Statistically significant meta-analyses of clinical trials have modest credibility and inflated effects', *Journal of Clinical Epidemiology*, vol. 64(10), pp. 1060–9.

Poole, C. and Greenland, S. (1999). 'Random-effects meta-analyses are not always conservative', *American Journal of Epidemiology*, vol. 150(5), pp. 469–75.

Rosenthal, R. (1979). 'The "file drawer problem" and tolerance for null results', *Psychological Bulletin*, vol. 86 (3), pp. 638–41.

Schmidt, F.L. and Hunter, J.E. (2015). *Methods of Meta-analysis: Correcting Error and Bias in Research Findings*, 3rd ed. Thousand Oaks, CA: Sage.

Simonsohn, U., Nelson, L.D. and Simmons, J.P. (2014). 'P-curve: a key to the file drawer', *Journal of Experimental Psychology: General*, vol. 143(2), pp. 534–47.

Stanley, T.D. (2005). 'Beyond publication bias', *Journal of Economic Surveys*, vol. 19(3), pp. 309–45.

Stanley, T.D. (2008). 'Meta-regression methods for detecting and estimating empirical effect in the presence of publication selection', *Oxford Bulletin of Economics and Statistics*, vol. 70(1), pp. 103–27.

Stanley, T.D. and Doucouliagos, C.(H.). (2012). *Meta-Regression Analysis in Economics and Business*, Oxford: Routledge.

Stanley, T.D. and Doucouliagos, C.(H.). (2014). 'Meta-regression approximations to reduce publication selection bias', *Research Synthesis Methods*, vol. 5(1), pp. 60–78.

Stanley, T.D. and Doucouliagos, C.(H.). (2015). 'Neither fixed nor random: weighted least squares meta-analysis', *Statistics in Medicine*, vol. 34(13), pp. 2116–27.

Stanley, T.D., Doucouliagos, C. and Ioannidis, J.P.A, (2017). 'Finding the power to reduce publication bias', *Statistics in Medicine*, vol. 36(10), pp. 1580–98.

Stanley, T. D., Doucouliagos, C.(H.), Giles, M., Heckemeyer, J.H., Johnson, R.J., Laroche, P., Nelson, J.P., Paldam, M., Poot, J., Pugh, G., Rosenberger, R.S. and Rost, K. (2013). 'Meta-analysis of economics research reporting guidelines', *Journal of Economic Surveys*, vol. 27(2), pp. 390–4.

Stanley, T.D. and Jarrell, S.B. (1989). 'Meta-regression analysis: a quantitative method of literature surveys', *Journal of Economic Surveys*, vol. 3(2), pp. 161–70.

Stanley, T.D., Jarrell, S.B. and Doucouliagos, C.(H). (2010). 'Could it be better to discard 90% of the data? A statistical paradox', *The American Statistician*, vol. 64(1), pp. 70–7.

Sterling, T.D. (1959). 'Publication decisions and their possible effects on inferences drawn from tests of significance or vice versa', *Journal of the American Statistical Association*, vol. 54(285), pp. 30–4.

Sterling, T.D., Rosenbaum, W.L. and Weinkam, J.J. (1995). 'Publication decisions revisited: the effect of the outcome of statistical tests on the decision to publish and vice versa', *American Statistician*, vol. 49(1), pp. 108–12.

Sutton, A.J., Song, F., Gilbody, S.M. and Abrams, K.R. (2000). 'Modelling publication bias in meta-analysis: a review', *Statistical Methods in Medical Research*, vol. 9(5), pp. 421–45.

Tullock, G. (1959). 'Publication decisions and tests of significance – a comment', *Journal of the American Statistical Association*, vol. 54(287), p. 593.

Turner, R.M., Bird, S.M. and Higgins, J.P.T. (2013). 'The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews', *PLoS ONE*, vol. 8(3). doi: 10.1371/journal.pone.0059202

Zhang, Le and Ortmann, A. (2013). 'Exploring the meaning of significance in experimental economics'. UNSW Australian School of Business Research Paper No. 2013-32.

Ziliak, S.T. and McCloskey, D.N. (2004). 'Size matters: the standard error of regressions in the *American Economic Review*', *Journal of Socio-Economics*, vol. 33(5), pp. 527–46.