HOW CREDIBLE IS TRADE UNION RESEARCH? FORTY YEARS OF EVIDENCE ON THE MONOPOLY–VOICE TRADE-OFF

HRISTOS DOUCOULIAGOS, RICHARD B. FREEMAN, PATRICE LAROCHE, AND T. D. STANLEY*

This article is the second in a series to celebrate the 70th anniversary of the *ILR Review*. The series features articles that analyze the state of research and future directions for important themes this journal has featured over many years of publication.

In this article, the authors assess the credibility of research that has tested the theoretical contests between the monopoly and the collective voice model of unions developed by Freeman and Medoff in *What Do Unions Do?* The authors go beyond prior analyses by examining more than 2,000 estimates that consider the effects of unions on a broad range of organizational and individual outcomes, including productivity, productivity growth, capital investment, profits, and job satisfaction. They advance our understanding of the current empirical findings and credibility of this research by using meta-statistical analysis to evaluate research quality, publication selection bias, statistical power, and heterogeneity. The authors conclude that compared to other areas of economics, research on union effects has lower bias but larger problems of statistical power. They argue that Freeman and Medoff's monopoly–collective voice model helped produce more credible results, and they suggest ways to reduce the power and heterogeneity problems in existing research.

In *What Do Unions Do?*, Freeman and Medoff (1984) contested the traditional economics view of labor unions as monopolies that adversely affect workplace performance. In addition, Freeman and Medoff argued in support of a collective voice and institutional response dimension to unions. The

*Hristos Doucouliagos is Alfred Deakin Professor and Chair in Economics at Deakin University. Richard B. Freeman is the Herbert Ascherman Professor of Economics at Harvard University. Patrice Laroche is a Professor of Human Resource Studies and Labor Relations at ESCP Europe, Paris. T. D. Stanley is Professor of Meta-Analysis at Deakin University.

In Doucouliagos, Freeman, and Laroche (2017), we offer a comprehensive meta-regression analysis of the impact of unions on various outcomes: productivity, productivity growth, physical and intangible capital investment, turnover, job satisfaction, and profitability. This article further expands the analysis in the 2017 book to consider the credibility of the exit-voice/union monopoly trade-off research agenda. For information, please address correspondence to the authors at douc@deakin.edu.au.

Keywords: monopoly–collective voice model, unions, meta-statistical analysis

ILR Review, 71(2), March 2018, pp. 287–305
DOI: 10.1177/0019793917751144. © The Author(s) 2017
Journal website: journals.sagepub.com/home/ilr
Reprints and permissions: sagepub.com/journalsPermissions.nav

collective voice face can reduce pay inequality, improve communication channels, retain more productive employees, and otherwise increase productivity.[1] Consequently, unions can have a net positive effect on workplace performance and economic efficiency.

More than three decades have passed since the publication of *What Do Unions Do?* and a considerable body of research has emerged that challenges Freeman and Medoff's ideas and re-examines the monopoly–voice trade-off. The scope of trade union research has expanded in many ways from an initial focus on wage effects to productivity and profitability, to non-wage effects and various indirect channels through which unions affect employee and investment behavior. The research base has also broadened from a historic basis of data from the United States to Europe and more recently to emerging countries, and from manufacturing to services industries. Furthermore, the sophistication of the data analytic techniques has evolved from cross-sectional OLS studies to panel data studies that allow researchers to remove person or workplace fixed effects and to quasi-experimental design.

Studies of trade unions, labor–management relations, and related topics constitute millions of Google Scholar entries.[2] The impressive growth in trade union research has created a challenge: How can scholars draw any generalizable conclusions from such a heterogeneous empirical evidence base? Fortunately, the mode of summarizing the findings of diverse studies into a statistically valid assessment of the evidence has also advanced considerably. Research synthesis methods and the meta-analysis of research now enable researchers to combine and analyze the results from diverse empirical studies (Doucouliagos and Laroche 2003; Stanley and Doucouliagos 2012).

Although a part of the union literature is descriptive or prescriptive and thus cannot be easily quantified, a substantial part consists of econometric estimates of the effects of trade unions on worker well-being, most often measured by wages or on enterprise productivity. We integrate the findings of the quantitative research using meta-statistical analysis to understand its main lessons and to assess the quality of trade union research.

Are the estimated union effects sufficiently similar across studies to determine a credible central tendency of union effects? To what extent are the differences among studies explicable in terms of statistical design or the economic situation facing the union and the employer? What does the research say about Freeman and Medoff's *What Do Unions Do?* claim that unions operate as an institution of collective voice representing workers' interests and knowledge in firm decision making as well as a monopoly agent raising workers' earnings? More broadly, are empirical studies of

[1]The collective voice and institutional response research agenda commenced with Freeman (1976). *What Do Unions Do?* summarizes trade union research by Richard Freeman and James Medoff.

[2]In October 2017, Google Scholar listed 1.7 million results on trade unions, 2.4 million on labor–management relations, 3.0 million on labor relations, 4.0 million on industrial relations, 1.1 million on collective bargaining, and 0.9 million on labor disputes.

trade union effects sufficiently credible to give policymakers confidence in using reported findings to formulate policies and to give researchers confidence in building new theories and analyses based on the shoulders of existing findings?

We analyze these questions in steps, moving from what data comprise our analyses to which tools determine and define credibility of empirical union research, and concluding with suggestions on future meta-analyses—all of which build on our knowledge base of what unions do.

The Studies under Study

Statistical analysis of union effects has a long history in empirical economics because unions have traditionally been the key labor institution in capitalist economies and because their effects on the economy have generated widespread controversy. On one side is an institutional tradition that stresses the ways in which unions affect outcomes in labor markets that deviate substantially from the competitive ideal. On the other side is a pure competition tradition that stresses the inefficiencies that unions can create in labor markets that fit the competitive ideal.

The growth of unions in the midst of the Great Depression through World War II generated studies on wage determination in unionized settings, most notably by John Dunlop (1944). In the 1950s and 1960s, H. Gregg Lewis (1963) and his students at Chicago used area and industry data to contrast earnings between industries or areas differing in union density. As data sets with individuals became widely available, researchers made more refined analyses of differences in hourly or weekly earnings between union and nonunion workers with comparable demographic characteristics. Lewis (1986) reviewed a large part of this literature. Freeman and Medoff's (1984) research program widened the analysis from earnings to diverse other outcomes on the grounds that unions affected workers and firms in ways beyond simply changing the price of labor. Bennett and Kaufman (2007) reviewed the ensuing two decades of work using traditional expert reviews as opposed to the meta-statistical approach that we use in this article.

Our analysis covers microeconomic studies of the relationship between unions and seven outcome variables—productivity in manufacturing, productivity in non-manufacturing, productivity growth, investment in physical capital, investment in intangible capital, job satisfaction, and profits—holding other factors that affect the outcome variable constant. We investigate 2,242 comparable estimates of trade union effects from 301 studies compiled by Doucouliagos, Freeman, and Laroche for their 2017 publication.[3] They transformed estimates of the relationship between unionism

[3]Estimates of more than one outcome from a single study (e.g., productivity and profitability) are treated as separate "studies."

Table 1. Trade Union Studies, Estimates, and Average Effect Size

Dimension	*Number of studies (1)*	*Number of estimates (2)*	*Meta-average effect, US (3)*
Direct effects			
Productivity, manufacturing	52	324	−0.02
Productivity, construction	63	386	0.20*
Productivity growth	42	268	0.01
Profits	44	478	−0.13*
Channels			
Physical capital	20	343	−0.23*
Intangible capital	25	208	−0.34*
Job satisfaction	59	235	−0.03

Source: Doucouliagos et al. (2017).
*Denotes statistical significance at the 1% level.

and the outcome variables reported in various studies—which often followed different disciplinary norms about reporting statistics—into partial correlations comparable across studies. Although almost certainly incomplete,[4] this compilation is the largest population of comparable estimates of the microeconomic effects of trade unions as of 2016. Most estimates come from samples of hundreds or thousands of workers or firms but some are from more aggregated data.

The majority of estimates come from the United States, in part because empirical economic analysis developed rapidly there and in part because the lack of centralized bargaining in the United States makes union–nonunion comparisons a more natural way to analyze what unions do (as compared to countries in which unions and management sign national or sectoral bargaining agreements that cover all or nearly all workers, as is common in Europe). With collective bargaining regulating both union and nonunion worker pay, union–nonunion comparisons are difficult to interpret and likely miss the effect of unions on outcomes.

Translating union effects reported in different ways across studies into partial correlations allows us to directly compare the magnitudes and standard deviations of the largest possible number of estimates. In some cases, we would have preferred to report estimated elasticities of effects, but many statistical studies do not report elasticities nor do they provide their statistics in ways that allow us to confidently calculate elasticities.[5] The wide diversity of measures of trade union effects dictates our focus on partial correlations to assess broadly the credibility of trade union research.

Table 1 presents the number of studies (column (1)) and the number of estimates (column (2)) for six key outcomes: productivity, productivity

[4]This database does not include foreign language papers, privately distributed working papers, student papers or theses, and relatively obscure journals.

[5]Elasticity is preferable for identifying the economic significance of union effects. However, by providing a larger pool of data, partial correlations better enable assessment of credibility.

growth, physical capital investment, intangible capital investment, job satisfaction, and profits. We divide the productivity estimates into manufacturing, the focus of a disproportionate number of studies, and other industries. Column (3) reports the meta-average effect for the United States, expressed as a partial correlation.[6] Doucouliagos et al. (2017) concluded that unions have no effect on productivity in manufacturing but have a positive effect in some other industries, most notably construction and education. Unions depress investment in both physical and intangible capital and reduce profits because union wage increases exceed the smaller productivity effect. At the same time, unions reduce worker turnover, which partially offsets adverse investment effects.

Nearly all of trade union effect studies come from observational data on workplaces. Such data are gathered by government or other survey organizations, and thus they suffer from the well-known problems of interpreting causality from non-experimental evidence. They leave open a potential omitted-variable problem in the form of unmeasured attributes of workplaces correlated with unionism. Depending on the data set, they cover some outcome variables but not others. Given that neither firms nor workers are likely to accept random assignment of union conditions, as an ideal experiment might demand, researchers are left with exogenous changes in laws or other variation to identify causality. Responses to legal or other changes may differ among workers and firms, however, producing results that may not generalize from one setting to another, which can produce heterogeneity among estimates and raise further questions about why the same exogenous change produces variation in outcomes among persons or firms.

Evaluating Credibility

Credibility is critical for assessing the validity and reliability of inferences drawn from evidence (Ioannidis and Doucouliagos 2013; Ioannidis, Stanley, and Doucouliagos 2017). In social psychology, for example, credibility has been called a "crisis" because of the inability to exactly duplicate highly regarded research results (Pashler and Wagenmakers 2012; Stroebe and Strack 2014; Stroebe 2016). Such variation may be attributable to poor protocol descriptions, questionable research practices, or failure to report all the results. Interest in the credibility of economic research (Camerer et al. 2016; Ioannidis et al. 2017; Christensen and Miguel 2017) is also increasing. For example, failure to replicate key economic findings raises the possibility that economic results may be fragile to precise specification, deletion of outliers, and potential confounders (Leamer 2010; Christensen and Miguel 2017). But given that much economic analysis is based on observational

[6]Column (3) meta-averages are weighted averages of all comparable effect estimates using inverse variance weights, adjusted for publication selection and model misspecification biases (Doucouliagos et al. 2017).

data, for which differences over time presumably reflect genuine changes in the economy and differences among economic units can produce heterogeneity in union effects, the credibility problem has not become a major concern. Still, analysis of observational data has its own credibility problems that researchers and policymakers must address along with their findings.

One way in which researchers have traditionally assessed credibility is based on the quality of the journal that published the work: Higher quality journals have stricter peer review, which makes findings in accepted papers more likely to be valid. Ratings of journal quality can be subjective, however, depending on disciplinary focus, refereeing policies, editorial board, or place of publication, as can be seen in the differences of "top quality journals" lists that some departments use in assessing researchers. The most widely used objective metric of a journal is its impact factor, which often depends on its having one or two highly cited papers. Citations, however, may be manipulated by journal publication policies and practices. The citations to an article provide an alternative to determining quality by the journal of publication: more citations suggest higher quality, ceteris paribus. But citations are imperfect because they depend on the network of researchers to whom the authors of a paper are connected as well as their "intrinsic" value.

Another indicator of quality is the reputation of the researchers or of the institutions where they work. If X is known as a very careful analyst whereas Y is known as someone prone to oversell results, it makes sense to trust X's findings over Y's findings. If X works for a world-class research institution whereas Y is employed at a community college, most researchers will trust X's claims more than Y's. Relying on reputation, however, risks downplaying the findings by younger or less-known researchers compared to those of senior researchers as well as confusing popularity with research quality.

Meta-analyses

Meta-analysis uses the actual statistics reported in each study, most often the precision (inverse standard error or variance) of an estimate to assign quality to each individual estimate, rather than using the impact factor, citations, or reputation of researchers in judging the credibility of research. An estimate that is more precise (has lower variance) is given greater weight in a synthesis of the evidence base, regardless of whether it appears in a journal of high or low impact, in a paper with many or few citations, or by researchers with dissimilar reputations.

To assess the credibility of an entire literature, as we do, requires additional consideration of the extent to which results in one study either replicate previously reported findings or are replicated in ensuing work.

Replication takes several forms, from using the same data and methods to *reproduce* previously reported findings to analyzing new data or using new methods to estimate the same parameters as in earlier findings, with due allowance for the uncertainty of estimates as indicated by their confidence intervals. In this article, we do not attempt to replicate primary studies using the same data and models nor do we provide new primary data estimates. Instead, we use the tools of meta-regression analysis to assess whether the results from numerous studies converge to a central tendency and whether credible inferences can be drawn from the accumulated evidence base.

We do not expect exact replication in our meta-analysis given that observational studies will reflect genuine heterogeneity and researchers will often make different choices about the inclusion or exclusion of particular control variables, about the exact model specification, about missing data, and about outliers. Rather, for credibility, we expect results to have the approximate magnitude as prior studies and the same sign. A finding that is credible should hold for moderately different specifications across data sets, with variation that is sufficiently balanced around the mean value. The key replicability issue is robustness of findings.

Another dimension of credibility relates to *publication selection bias,* which is often signaled by a distribution of reported outcomes around a mean value that is notably imbalanced. The suspicion is that authors, journal editors, or reviewers of journals selectively publish papers that report some particular empirical result, perhaps a statistically significant one, and omit others. For example, some researchers may have a preference to report only statistically significant negative (or positive) union effects. Researchers in psychology report that having a paper published in some journals requires significant results around a clear story, which leads them to hold back ambiguous or insignificant specifications of their analysis (Stroebe 2016). Omission from the public record of statistically insignificant findings or findings at odds with researchers' priors can negatively affect the synthesis of the evidence base, producing parameter estimates that can be biased and often inflate the magnitudes and significance of effects.

Statistical power is another criterion for credibility. It refers to the probability that a study can detect the underlying effect. By definition, low-powered studies will produce high rates of false negatives, but they can also cause high rates of false positives, producing evidence for nonexistent effects (Ioannidis 2005). Statistical power is a function of the desired level of statistical significance, sample size, and the size of the underlying "true" effect of unions on an outcome. Studies with small samples testing weak associations will be *under-powered.* This can lead to failures to replicate and may also lead to greater use of specification searching and questionable research practices, leading to publication selection bias. By examining a large body of studies, meta-analyses can reduce bias by averaging across estimates with potentially different biases while increasing power by pooling under-powered studies together.

Finally, empirical studies report a wide range of estimates of a single union effect. This variation stems from four sources:

1) sampling error due in part to the number of observations in a study;
2) research design methods, for example, specification of econometric model;

Table 2. Quality of Trade Union Effects Studies

Dimension	*Percentage published in leading journals* (1)	*Percentage published in ILR Review* (2)	*Mean (median) citations* (3)	*Mean (median) journal impact factor* (4)
Productivity, manufacturing	48	8	349 (79)	2.31 (1.49)
Productivity, other	42	17	291 (34)	1.73 (0.81)
Productivity growth	53	4	239 (52)	1.39 (0.81)
Physical capital	70	5	97 (57)	1.97 (1.77)
Intangible capital	81	17	258 (45)	1.73 (1.54)
Job satisfaction	43	14	176 (96)	1.36 (1.33)
Profits	60	11	343 (68)	1.96 (1.37)

Notes: See footnote 7 for list of leading journals. Unit of analysis is estimates in columns (1) and (2) and studies in columns (3) and (4). Column (3) presents Google Scholar citations as of November 12, 2017. Column (4) presents 2016 SSCI 5-year impact factors.

3) measurement error of outcome and unionization variables; and
4) inherent heterogeneity due to different union effects across sectors, countries, or time.

Sources (1) to (3) are features of the underlying data or research process, which can produce dissimilar estimated effects due to low power and high bias. By contrast, source (4) reflects actual differences in union effects in various economic or institutional settings. In this case, the effects are truly different and do not constitute a failure to replicate. For example, unions will likely have different effects on wages or productivity in manufacturing than they would in education, or in boom times rather than in recessions.

Credibility of Union Research

We begin our analysis of the credibility of studies of trade union effects by examining the quality of the research record. What proportion of studies appear in leading field journals compared to journals of less prestige? What are the impact factors of the journals that publish studies of trade union effects? Do the studies receive many or few citations?

Table 2 presents data on the journal of publication and citations to the relevant articles. Column (1) records the proportion of the estimates published in what are widely regarded as the leading field journals in economics, industrial relations, and management.[7] By considering only estimates

[7]We classify the following as leading field journals: *Academy of Management Journal, American Economic Review, Bell Journal of Economics, British Journal of Industrial Relations, Brookings Papers: Microeconomics, Canadian Journal of Economics, Economic Journal, Economica, Economics Letters, European Economic Review, Human Relations, Industrial and Labor Relations Review (ILR Review), Industrial Relations, Journal of Banking and Finance, Journal of Business, Journal of Human Resources, Journal of Industrial Economics, Journal of International Economics, Journal of Labor Economics, Journal of Law and Economics, Journal of Political Economy, Management Science, Oxford Bulletin of Economics and Statistics, Oxford Economic Papers, Quarterly Journal of Economics, Review of Economics and Statistics,* and *Strategic Management Journal.* For articles on job satisfaction, we also include *American Sociological Review* and *Human Resource Management.*

from these sources, we are taking a conservative view of studies published in books and other journals, in contrast to Doucouliagos et al.'s (2017) assessment that deems estimates published in any peer-reviewed journal or in a book to be of sufficient quality for analysts to take seriously.

Our narrow measure of research quality finds that a fairly large proportion of the union effect estimates are published in leading field journals. In two areas of this research, union effects on physical capital investment and intangible capital investment, 70% and 81% of the estimates, respectively, are published in leading field journals listed in footnote 7. A likely reason for that finding is that data on intangible and physical capital are harder to access, so those analyses are more novel and thus more attractive to leading journals. Direct estimates of productivity and productivity growth, for which data are easier to access, are published in a wider set of places, with 42 to 53% appearing in the leading journals. For analyses of profits, 60% are in leading journals. The percentage of publications in leading journals is 43% for studies of job satisfaction. Some of these data are based on particular occupations or professions, which are of interest to persons in those fields and are thereby published in an array of narrower publication outlets.

Column (2) reports the percentage published in the *Industrial and Labor Relations Review* (*ILR Review*), where the percentage of productivity studies outside manufacturing is particularly high.

Next, we consider citations. We use citations from Google Scholar, as Scopus does not cover many of the earlier studies in our sample. Column (3) presents the mean and median number of citations received. Union effects studies draw a relatively large number of citations, with the median productivity study receiving nearly 60 citations, which exceeds typical citations in economics.[8] Column (4) reports the mean and median journal impact factor of the journal in which estimates are published. These are alternative measures of research quality. In most cases, the median impact factor exceeds 1, suggesting a literature that finds a place in well-regarded journals.

In Table 3 we turn to the quality of estimates themselves. Here we treat all estimates and their estimated precision equally, regardless of where the article was published. In column (1) we compare the mean and median degrees of freedom. Studies with larger degrees of freedom potentially convey more information. The median degree of freedom is relatively small for most dimensions that deal with firm- or industry-level data. The literature on job satisfaction is an exception, as this deals with individual workers, for whom data sets can be quite large. In column (2) we compare statistical precision as an objective measure of quality. Statistical precision is defined here as the inverse of the estimated standard error of the estimated union effect. In meta-analysis, more precise estimates are deemed to be of higher quality, ceteris paribus, and are used to weight estimates. The job satisfaction data

[8]As citations take time to accumulate, older papers usually draw more citations.

Table 3. Quality of Trade Union Effects Estimates

Dimension	*Mean (median) degrees of freedom (1)*	*Mean (median) precision (2)*	*Percentage using panel data (3)*	*Percentage treating endogeneity (4)*
Productivity, manufacturing	2,296 (250)	31.4 (15.8)	73	12
Productivity, other	1,346 (137)	22.9 (11.8)	35	5
Productivity growth	732 (177)	20.5 (12.6)	100	9
Physical capital	2,266 (294)	26.5 (17.1)	76	11
Intangible capital	774 (296)	22.3 (17.0)	62	24
Job satisfaction	11,556 (2,776)	81.7 (53.2)	23	19
Profits	1,028 (344)	24.6 (17.6)	55	11

have higher precision, on average, reflecting the larger sample sizes associated with individual worker data sets.

Econometric studies of workers and firms from panel data are often preferable to those from cross-section data as the longitudinal information enables researchers to control for unobserved factors that might influence performance and thus produce biased results. A typical panel study of productivity adds establishment, firm, or industry fixed effects whereas a typical panel study of job satisfaction adds individual worker effects. In most cases, this approach reduces the sample size for estimating the union effect as the estimate is based exclusively on the firms or persons who change union status, with the fixed effects absorbing the outcomes of firms or persons who do not change status. Column (3) shows that the proportion of panel data studies differs across outcomes. The proportion is highest in studies of productivity in manufacturing, productivity growth, and investment in physical capital and lowest for job satisfaction, suggesting that the former studies have used stronger econometric methods and thus are more credible than the typical job satisfaction study that does not use panel methods.

Another critical aspect of the quality of an estimate is the information it provides about the extent to which estimates can be interpreted in a causal manner. Causality is almost always difficult to establish in observational research. Unions may choose to unionize more productive firms or firms with market power, and unionized firms with greater productivity or market power may have longer life spans than do unionized firms with lower productivity or market power. This tendency leads to biased estimates in favor of positive union productivity effects. Alternatively, workers in struggling and unprofitable firms may seek union protection, in which case failing to correct for reverse causality will give downward-biased estimates and inflate adverse union effects.

Some studies attempt to control for unobservable factors using firm fixed effects estimated with panel data. Others try to establish causality using event study and regression discontinuity methods (e.g., DiNardo and Lee 2004). Good instruments are often hard to come by, however, and studies with access to short panels restrict the use of lags as a way to probe causality. Column (4) reports the percentage of estimates from studies that treat for endogeneity in some way in an attempt to establish causality between unionization and an outcome variable. Most studies do not tackle this issue and consequently causality remains a challenge to this literature.

Publication Selection Bias

To deal with potential publication bias, we use the Funnel Asymmetry Test-Precision Effect Test (FAT-PET), meta-regression analysis (MRA) (Stanley 2005, 2008; Stanley and Doucouliagos 2012). The overall, unconditional version of these tests involves regressing union effects upon a constant and the standard error of the partial correlations: $r_i = \beta_0 + \beta_1 SE_i + \varepsilon_i$, where r denotes the union effect, SE is the standard error of the union effect, β_0 is the measure of the genuine empirical effect corrected for publication selection bias, and $\beta_1 SE$ approximates publication bias. Testing for publication bias takes the form of testing H_0: $\beta_1 = 0$.

What might appear as bias, however, might alternatively be heterogeneity. Hence, we also estimate the conditional multiple MRA version of the FAT-PET that considers heterogeneity in trade union effects along with potential reporting bias. This more complex approach expands FAT-PET to include a vector of moderator variables that reflect heterogeneity, misspecification, and research design choices. For these more complex MRAs, we do not provide new estimates here, but rather rely on the findings from Doucouliagos et al. (2017).

Table 4 presents the funnel asymmetry test (FAT) for publication selection bias. Columns (1), (2), and (3) report the estimated unconditional FAT coefficient from the FAT-PET for all estimates, and separately for the US and non-US estimates, respectively. Column (4) reports the conditional FAT coefficient, which controls for heterogeneity in data and methods.[9] The results presented in column (4) suggest preferential reporting bias in favor of positive productivity effects for manufacturing, in favor of adverse productivity growth effects, and for positive intangible capital effects.

Doucouliagos and Stanley (2013) offered suggestions for the interpretation of the size of publication bias, in which FAT coefficients less than 1 indicate a small degree of bias. From this perspective, the Table 4 estimates that the FAT coefficient is always a little over or less than 1 suggest modest selection bias, per the column (5) conclusions for each variable. Although there may be some pockets of residual publication selection bias in trade

[9]Doucouliagos et al. (2017) provided the details of the variables and methods used.

Table 4. Publication Selection Bias, Trade Union Effects

	Unconditional			Conditional	
Dimension	*FAT, all* *(1)*	*FAT, USA* *(2)*	*FAT, non-USA* *(3)*	*FAT, all* *(4)*	*Magnitude of bias* *(5)*
Productivity, manufacturing	0.453	0.685	−0.107	1.018**	Small to moderate
	(1.59)	(1.68)	(−0.23)	(2.52)	
Productivity, other	0.243	1.008**	−1.941**	0.432	None
	(0.61)	(2.45)	(−2.21)	(0.68)	
Productivity growth	−1.170**	−1.275**	−0.897	−0.834**	Small
	(−2.47)	(−2.18)	(−1.35)	(−2.20)	
Physical capital	−0.430	−0.779	−0.336	−0.106	None
	(−0.78)	(−0.87)	(−0.88)	(−0.66)	
Intangible capital	0.031	0.638**	−1.961**	0.795**	Small
	(0.04)	(2.06)	(−2.50)	(2.41)	
Job satisfaction	−0.431	−0.443	−0.006	0.439	None
	(−0.97)	(−0.73)	(−0.01)	(0.84)	
Profits	−1.329***	−0.919	−0.506	−0.422	None
	(−3.25)	(−1.53)	(−0.97)	(−1.15)	

Source: Doucouliagos et al. (2017).
Notes: Columns (1) to (3) present unconditional FAT coefficients from the FAT-PET regressions. Column (4) presents conditional FAT coefficients from the FAT-PET regressions controlling for other covariates. FAT-PET, funnel asymmetry test-precision effect test.
***and ** denote statistically significant at the 1% and 5% level, respectively.

union research, in general, we find little overall bias in this research relative to other areas of economics (Doucouliagos and Stanley 2013; Ioannidis et al. 2017).

The Effect of Contestability on Credibility

Doucouliagos, Laroche, and Stanley (2005) and Doucouliagos and Stanley (2013) argued that when a research field is contested, publication bias is less likely. In the union literature, the Freeman and Medoff (1984) two-faces view of trade unions stated that unions can have either a positive or a negative effect, or no effect at all, on the various dimensions of performance, depending on the specifics of union activities and the response of management. This finding is analogous to studies of the effect of wages on hours worked, in which income and substitution effects work in opposite directions to affect the leisure–work choice. The monopoly side of unions may dominate, or the collective voice may offset monopoly effects. This circumstance paves the way for researchers to report all of their estimates, which, in turn, creates a research literature that should be relatively free of bias. By challenging the traditional monopoly view of trade unions, the two-faces view of unions reduces publication bias in this field of research.

One dimension of contestability is the concentration of studies. A field dominated by a small group of scholars might be less competitive than one

Table 5. Research Contestability, Trade Union Effects Research

Dimension	*Number of authors* *(1)*	*Largest share (%)* *(2)*	*C4 (%)* *(3)*	*HHI* *(4)*
Productivity, manufacturing	84 (1.6)	12	36	540
Productivity, other	92 (1.5)	11	33	457
Productivity growth	62 (1.5)	13	41	615
Physical capital	33 (1.7)	40	82	2301
Intangible capital	35 (1.4)	36	75	1929
Job satisfaction	90 (1.5)	7	25	330
Profits	68 (1.5)	42	63	2023

Notes: Parentheses report average number of authors per paper. C4, top four authors or groups of coauthors; HHI, Herfindahl-Hirschman Index.

with a large number of "competitor" authors. Table 5 reports several measures of research concentration. Column (1) reports the number of authors and coauthors who published in each research dimension. Column (2) reports the largest share of estimates by a single author or group of coauthors. Column (3) reports the proportion of estimates reported by the top four authors or groups of coauthors (C4). Column (4) reports the Herfindahl-Hirschman Index (HHI) as a measure of the share of estimates or dominance of a field.

Concentration is greatest in studies of physical and intangible capital and profits, in which one group of scholars contributes between 36% and 42% of the estimates. C4 and HHI confirm that union impact on physical and intangible capital and profits are the most concentrated areas, with the top four authors or groups of coauthors producing between 63% and 82% of the estimates. At one level, having different groups competing may be important to the market for ideas, although we note that Ioannidis (2005) argued that having numerous groups competing can increase the chance of false positives by increasing the incentives to exaggerate findings.

Virtually all the studies included in our data cite Freeman and Medoff's *What Do Unions Do?* or similar texts relating to the two-faces view of unions. Hence, most authors are aware that theory allows a range of conflicting results. Consequently, we expect a relatively high degree of competition between authors and little bias. This premise is confirmed by comparing Tables 4 and 5, which shows no obvious correlation between bias and concentration.

Statistical Power

We assess statistical power by using meta-averages from our feasible population of estimates on a union effect to assess power. Following Ioannidis et al. (2017), Stanley and Doucouliagos (2015), and Stanley, Doucouliagos, and Ioannidis (2017), we use the simple unrestricted weighted least squares weighted average (WLS) as the estimate of "true" effect and the

Table 6. Statistical Power, US and Non-US Estimates

	Unconditional			Conditional	
Dimension	*% with adequate statistical power, all (1)*	*% with adequate statistical power, US (2)*	*% with adequate statistical power, non-US (3)*	*% with adequate statistical power, US (4)*	*% with adequate statistical power, UK (5)*
Productivity, manufacturing	0	0	0	7	54
Productivity, other[a]	5	1	16	33	0
Productivity growth	0	0	0	0	0
Physical capital[b]	4	6	0	85 (24)	—
Intangible capital[b]	17	36	0	97 (78)	—
Job satisfaction	4	0	23	10	12
Profits	15	9	8	27	—

Notes: Columns (1), (2), and (3) report the percentage of estimates that are adequately powered, using the unconditional WLS mean to estimate the "true" effect. Columns (4) and (5) use the conditional WLS mean as the estimate of the true effect. WLS, weighted least squares.
[a]Columns (4) and (5) assessed for the construction industry.
[b]Column (4) US estimates at industry level with firm-level estimates reported in parentheses.

conventional 80% as threshold for adequate power.[10] Ioannidis et al. (2017) demonstrated how an econometric estimate will have adequate power when its standard error is smaller than the absolute value of the unrestricted WLS weighted average divided by 2.8.[11] If selective reporting bias occurs in either direction, then WLS will also be biased in the same direction and hence overestimate the true effect. In this case, power will also be overestimated, identifying more studies as being *adequately* powered and hence giving a conservative estimate of power (Stanley et al. 2017).[12]

Table 6 reports our findings for statistical power. Column (1) reports results for all studies combined. Columns (2) and (3) give the results for the US and for all non-US studies, taken as a group. Columns (4) and (5) report the percentage adequately powered using the conditional mean for the United States and for the United Kingdom, the two countries for which we have enough studies to treat separately. In these two columns we use the estimated multiple meta-regressions to derive a best estimate for the underlying effect and treat this as the estimated true effect. The conditional

[10]The unrestricted WLS weighted average must give the exact same point estimate as the conventional meta-analysis fixed-effect weighted average. They differ only in that WLS has a different variance (and confidence interval) that accommodates heterogeneity should there be any evidence of it in the research record. Because only the point estimate is used to calculate power, our power calculations will be exactly the same if one uses the point-estimate equivalent, the fixed-effect weighted average.

[11]This 2.8 value is derived from conventional standards for statistical significance and power. The 2.8 is the sum of the number of standard deviations (1.96) from the null hypothesis required for statistical significance and the fact that it takes 0.84 standard deviations for the cumulative normal distribution to have a 20/80% split required for 80% power (Cohen 1988).

[12]Absent selective reporting bias, any weighted average (including WLS) is an unbiased estimate of the true effect.

Table 7. Time Trends in Power

Dimension	*Sample size* *(1)*	*Standard error* *(2)*
Productivity, manufacturing	255	−0.003
	(2.24)**	(−5.81)***
Productivity, other	158	−0.002
	(2.56)**	(−2.77)***
Productivity growth	7	−0.002
	(0.32)	(−1.24)
Physical capital	154	0.002
	(0.71)	(0.83)
Intangible capital	14	−0.003
	(0.34)	(−2.01)*
Job satisfaction	836	−0.001
	(2.76)***	(−2.09)**
Profits	196	−0.003
	(2.54)**	(−4.29)***

Notes: Figures in parentheses are *t*-statistics using standard errors adjusted for clustering of estimates within studies. Cells report the coefficient on the year the study was published.
***, **, and * denote statistically significant at the 1%, 5%, and 10% level, respectively.

means are larger and hence provide the best-case scenario for statistical power in this literature.

In the case of productivity growth, not a single estimate is adequately powered. The percentage adequately powered increases when conditional means are used, but the percentages are still small. For the branch of literature that has received most of the attention—estimates of productivity effect of unions in the United States—only 7% of the estimates for manufacturing are adequately powered. The percentage is highest in US construction, but even here only one-third of the estimates are adequately powered. The two exceptions are US studies of the impact of unions on physical and intangible capital at the industry level, for which most of the estimates are adequately powered, with lower power when analysis turns to capital investment at the firm level.

To what extent is power improving over time? Table 7, columns (1) and (2) report the results of regressing the sample size and standard error, respectively, on the year the study was published. The positive coefficient on sample size and the negative coefficient on standard errors suggest that statistical power is rising over time. Power has risen in all areas, except for studies of productivity growth and physical capital.

Heterogeneity

Heterogeneity occurs when there is no single union effect but rather a distribution of true union effects. That is, the union effect might vary over time, among industries, or among nations. A widely used estimate of heterogeneity is I^2, which measures the proportion of observed variation among

Table 8. Heterogeneity

Dimension	I^2 *(1)*	τ *(2)*
Productivity, manufacturing	87	0.12
Productivity, other	86	0.16
Productivity growth	84	0.16
Physical capital	80	0.09
Intangible capital	89	0.11
Satisfaction	82	0.03
Profits	84	0.11

Notes: I^2 estimates the proportion of observed variation among reported union effects not due to sampling error. τ estimates between-study standard deviation.

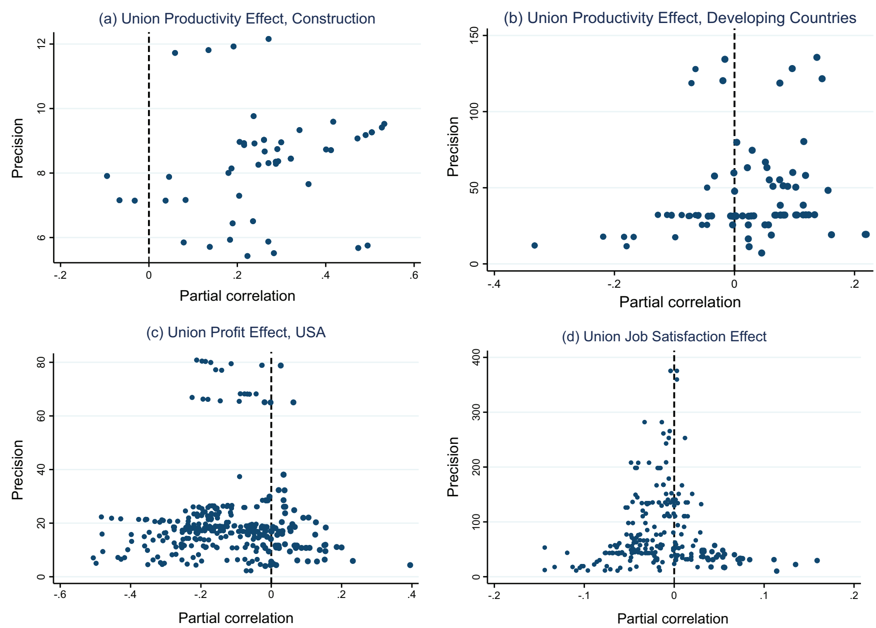
reported union effects *not* due to sampling error (Higgins and Thompson 2002). Another popular indicator of heterogeneity is τ^2, a measure of between-study variance (Rücker, Schwarzer, Carpenter, and Schumacher 2008).

Table 8 registers these estimates of heterogeneity with column (1) reporting I^2 and column (2) reporting the estimated τ (between-study standard deviation). I^2 indicates a very large degree of heterogeneity. The value of τ suggests large variation relative to the estimated effect sizes; recall Table 1, column (3). With such a large heterogeneity among the true effects, the true effect will frequently be in the opposite direction as the reported mean union effects reported in Table 1, column (3). For example, the job satisfaction average effect is the exact same size as τ (-0.03 compared to 0.03), implying, for example, that job satisfaction will actually be positively related to unionization 16% of the time even if this -0.03 were the true mean effect. Only the average negative effects on physical capital and intangible capital are reliable (probability of opposite true effect $< .05$), relative to these high levels of heterogeneity.

Figure 1 illustrates heterogeneity in parts of this literature in the form of funnel plots. Panels (a) and (b) illustrate the effect of unions on productivity in construction and in developing countries, respectively. Panels (c) and (d) illustrate the effect of unions on profit in the United States and the impact on job satisfaction, respectively. Reported estimates vary widely across these dimensions.

The funnel plots also inform the ease by which trade union research can be replicated. For example, in the case of construction, the majority of estimates are in the same direction. That is, while they differ in their estimates of the magnitude of the union effect, studies have essentially been able to replicate the *directional* effect of trade unions on productivity in construction. By contrast, less directional consistency occurs in terms of productivity in developing countries. Point estimates, however, can be a misleading indicator of replication. Confidence intervals provide a better indicator of the degree of replication, and this is what meta-analysis considers when pooling estimates from different studies.

Figure 1. Heterogeneity in Trade Union Effects



Source: Authors' construction using data from Doucouliagos et al. (2017).

Conclusion and Suggestions

Our review of the credibility of the trade union effects literature finds that it has relatively low publication selection bias, which makes the literature reasonably credible. However, trade union research suffers from low statistical power and high heterogeneity of estimated effects. Researchers can improve credibility by conducting studies with greater statistical power (with more observations and better statistical methods, when possible) and by explicitly analyzing the high heterogeneity of estimated union effects (for a better understanding of why effects vary in magnitude across units).

Are the trade union research studies more or less credible than other quantitative economics? Although assessment of the credibility of economics research is still a new area of analysis, available surveys suggest that trade union research is more credible than other areas of economics in terms of publication selection bias, but it has problems in terms of power. Doucouliagos and Stanley (2013) surveyed publication selection in 87 areas of economics. They found an average degree of selection bias of 1.64 (measured as the average estimate of β_1, with a median value of 1.54.[13] By contrast, trade union research has a significantly lower degree of selection bias,

[13]Doucouliagos and Stanley (2013) included five union meta-analyses in their survey. We remove these to determine estimates of selection bias in empirical economics excluding trade union research. Selection bias terms are converted into absolute values to enable comparison.

with a mean estimated β_1 = 0.58 and a median of 0.43. The difference between these means is statistically significant (p value = 0.0001).

To compare the power in union studies with that in other parts of economics, we rely on Ioannidis et al.'s (2017) survey of statistical power in economics across 159 research areas. They reported a median proportion that is adequately powered across economics is 10.5% or less. Nonetheless, Table 6, column (1) reports that trade union research typically has even fewer studies adequately powered, 4%.

Last, few studies control for endogeneity even though almost all trade union effects may be subject to reverse causation. Systemic low power and the frequent failure to adequately control for potential endogeneity question the credibility of trade union research.

On the basis of this analysis, we offer four suggestions for primary research on quantitative union effects. First, we recommend increasing the sample size and power of estimated union effects, particularly at the establishment level. One way to determine the likely union status of an establishment from existing data would be to exploit links between the Current Population Survey (CPS), the Longitudinal Employment and Household Data, and Censuses and Surveys of establishments to match workers at an establishment to the reports of those workers on CPS about union status. Given that union membership in the US private sector is virtually coterminous with collective bargaining, even a small number of CPS-establishment matched workers could identify the union status of an establishment.

Second, we recommend estimating union effects outside of manufacturing, a dimension for which the share of employment continually falls. The few studies of union impacts on the growing education and health sectors find positive union effects, but the number of such studies needs to expand to provide reliable research synthesis and to deepen our understanding of how a collective voice institution affects performance in service sectors more broadly.

Third, trade union research needs to embrace experimental and quasi-experimental designs (e.g., regression discontinuity, natural experiments, fixed-effect panel models) that control for potential reverse causation. Greater effort to find good instruments for unionization (admittedly easier said than done) might also go a long way toward ensuring that estimated effects are not the artifact of ignored endogeneity.

Fourth, we recommend greater uniformity in modes of presenting estimates of union effects. The Doucouliagos et al. (2017) database that we use required a massive effort to pull the partial correlation coefficients out of empirical results given in different ways in different disciplines. If researchers and journals in the labor relations field could come to some agreement for uniformity of reporting results (in technical appendices if not in the main body of papers), it would greatly facilitate meta-analysis of the next decade's research.

References

Bennett, James T., and Bruce E. Kaufman (Eds.). 2007. *What Do Unions Do? A Twenty-Year Perspective.* New Brunswick, NJ: Transaction Publishers.

Camerer, Colin F., Anna Dreber, Eskill Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan, Emma Heikensten, Felix Holzmeister, Taisuke Imai, Siri Isaksson, Gideon Nave, Thomas Pfeiffer, Michael Razen, and Hang Wu. 2016. Evaluating replicability of laboratory experiments in economics. *Science* 351(6280): 1433–36.

Christensen, Garret S., and Edward Miguel. 2017. Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature.* Forthcoming.

Cohen, Jacob. 1988. *Statistical Power Analysis in the Behavioral Sciences*, 2nd edition. Hillsdale, NJ: Erlbaum.

DiNardo, John, and David S. Lee. 2004. Economic impacts of new unionization on private sector employers: 1984–2001. *Quarterly Journal of Economics* 119(4): 1383–441.

Doucouliagos, Hristos (Chris), and Patrice Laroche. 2003. What do unions do to productivity? A meta-analysis. *Industrial Relations: A Journal of Economy and Society* 42: 650–91.

Doucouliagos, Hristos (Chris), and T. D. Stanley. 2013. Are all economic facts greatly exaggerated? Theory competition and selectivity. *Journal of Economic Surveys* 27(2): 316–39.

Doucouliagos, Hristos (Chris), Patrice Laroche, and Tom D. Stanley. 2005. Publication bias in union-productivity research? *Relations Industrielles/Industrial Relations* 60: 320–44.

Doucouliagos, Hristos (Chris), Richard B. Freeman, and Patrice Laroche. 2017. *The Economics of Trade Unions: A Study of a Research Field and Its Findings.* Oxon, UK: Routledge.

Dunlop, John T. 1944. *Wage Determination Under Trade Unions.* New York: Macmillan.

Freeman, Richard B. 1976. Individual mobility and union voice in the labor market. *American Economic Review: Papers and Proceedings* 66: 361–68.

Freeman, Richard B., and James L. Medoff. 1984. *What Do Unions Do?* New York: Basic Books.

Higgins, Julian, and Simon G. Thompson. 2002. Quantifying heterogeneity in meta-analysis. *Statistics in Medicine* 21(11): 1539–58.

Ioannidis, John P. A. 2005. Why most published research findings are false. *PLoS Medicine* 2(8): e124.

Ioannidis, John P. A., and Hristos (Chris) Doucouliagos. 2013. What's to know about the credibility of empirical economics? *Journal of Economic Surveys* 27(5): 997–1004.

Ioannidis, John P. A., T.D. Stanley, and Hristos (Chris) Doucouliagos. 2017. The power of bias in economics research. *Economic Journal* 127: F236–F265.

Leamer, Edward E. 2010. Tantalus on the road to Asymptopia. *Journal of Economic Perspectives* 24(2): 31–46.

Lewis, H. Gregg. 1963. *Unionism and Relative Wages in the United States: An Empirical Inquiry.* Chicago: University of Chicago Press.

———. 1986. *Union Relative Wage Effects: A Survey.* Chicago: University of Chicago Press.

Pashler, Harold, and Eric-Jan Wagenmakers. 2012. Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science* 7(6): 528–30.

Rücker, Gerta, Guido Schwarzer, James R. Carpenter, and Martin Schumacher. 2008. Undue reliance on I^2 in assessing heterogeneity may mislead. *BMC Medical Research Methodology* 8(1): 79.

Stanley, T. D. 2005. Beyond publication bias. *Journal of Economic Surveys* 19(3): 309–45.

———. 2008. Meta-regression methods for detecting and estimating empirical effect in the presence of publication bias. *Oxford Bulletin of Economics and Statistics* 70: 103–27.

Stanley, T. D., and Hristos (Chris) Doucouliagos. 2012. *Meta-Regression Analysis in Economics and Business.* Oxon, UK: Routledge.

———. 2015. Neither fixed nor random: Weighted least squares meta-analysis. *Statistics in Medicine* 34: 2116–27.

Stanley, T. D., Hristos (Chris) Doucouliagos, and John P. A. Ioannidis. 2017. Finding the power to reduce publication bias. *Statistics in Medicine* 36: 1580–98.

Stroebe, Wolfgang. 2016. Are most published social psychological findings false? *Journal of Experimental Social Psychology* 66: 134–44.

Stroebe, Wolfgang, and Fritz Strack. 2014. The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science* 9(1): 59–71.