

International
Institute of
Social Studies

Ezopus

Working Paper
No. 674

**Does research on economic sanctions suffer from
publication bias? A meta-analysis**

*Binyam A. Demena, Gabriela Benalcazar Jativa,
Alemayehu S. Reta, Patrick B. Kimararungu and
Peter A.G. van Bergeijk*

March 2021

ISSN 0921-0210

The International Institute of Social Studies is Europe's longest-established centre of higher education and research in development studies. On 1 July 2009, it became a University Institute of the Erasmus University Rotterdam (EUR). Post-graduate teaching programmes range from six-week diploma courses to the PhD programme. Research at ISS is fundamental in the sense of laying a scientific basis for the formulation of appropriate development policies. The academic work of ISS is disseminated in the form of books, journal articles, teaching texts, monographs and working papers. The Working Paper series provides a forum for work in progress which seeks to elicit comments and generate discussion. The series includes academic research by staff, PhD participants and visiting fellows, and award-winning research papers by graduate students.

Working Papers are available in electronic format at www.iss.nl/en/library

Please address comments and/or queries for information to:

Institute of Social Studies

P.O. Box 29776

2502 LT The Hague

The Netherlands

or

E-mail: wpapers@iss.nl

Table of Contents

ABSTRACT	4
1 INTRODUCTION	5
2 METHOD AND METHODOLOGY	8
3 DATA AND EMPIRICAL STRATEGY	10
3.1 Methods, protocols and data construction	10
3.2 Meta-data	11
4 META-ANALYSIS AND BIAS	13
4.1 Graphical inspections	13
4.2 Meta regression analysis	14
4.3 Sources of bias	15
5 CONCLUSIONS AND ISSUES FOR FURTHER RESEARCH	21
REFERENCES	23

Abstract

We meta-analyse 36 primary studies on determinants of the effectiveness of economic sanctions published over the years 1985-2018, using the Protocol of the Meta-Analysis in Economics Research-network. We investigate the impact of trade linkage, sanction duration and prior relations on sanction success. While the descriptive analysis and weighted averages suggest that the impact of the three variables of interest is significant and conforms to a priori theoretical expectations, our econometric analysis uncovers significant publication bias in the results. Bias is significant and large for the three variables of interest and the genuine impact of these variables on success and failure of sanctions after correction for publication bias is insignificant. Moreover, we find that bias in this literature increases over time.

Keywords

Economic sanctions, success, failure, meta-analysis, trade linkage, sanction duration, prior relations, heterogeneity.

JEL Classification: F14

Acknowledgements

This is the pre-peer review version of Chapter 6 of the *Research Handbook on Economic Sanctions* (Edward Elgar, forthcoming 2021). This research is based on a MA project that was awarded the ISS prize for the best Student Research Project in 2018. A preliminary version of this paper was presented at the 19th Jan Tinbergen Peace Science Conference and summaries appear as van Bergeijk et al. (2019) and van Bergeijk (2019). We would like to thank discussants and readers for useful comments.

Does research on economic sanctions suffer from publication bias? A meta-analysis

1 Introduction

Empirical research on economic sanctions started in 1985 with the seminal publication of *Economic Sanctions Reconsidered* by Hufbauer and Schott (1985, see also Hufbauer and Jung 2021).¹ This was the first time that a so-called large- N (sample size) data set became available – until then sanction research dealt with (a handful of) individual case studies (e.g., Wallensteen, 1968). The availability of the HSE data set stimulated numerous articles that used it to test theories that determine success and failure of economic sanctions. The set of sanction cases was significantly extended with the third edition (Hufbauer et al., 2007). More importantly, new data sets went beyond the implemented sanctions at the macro level of the Hufbauer et al. study, notably the Threats and Imposition of Sanctions (TIES, see Morgan et al. 2009; Morgan et al., 2014; Morgan et al. 2021) and the UN targeted sanctions dataset (Biersteker et al., 2018; Biersteker and Hudáková 2021). This working paper addresses the research puzzle that despite 35 years of empirical research no consensus has yet transpired on the sign and size of the impact of some key factors that theoretically shape the success of boycotts and embargoes (van Bergeijk 2019).

This puzzle is even more striking, because disagreement and inconclusiveness, as will become clear later on, *increased* despite significant improvements in data collection and analysis. In order to solve our research puzzle, we will perform a meta-analysis of 37 primary studies published over the years 1985 – 2018, inclusive. These primary studies aim to explain empirically under which conditions economic sanctions succeed. Our goal is not to cover all approaches to and explanations of economic sanctions; we will use a sub-sample of empirical studies and will only consider those primary studies that have included trade linkage, sanction duration and/or prior relations amongst the explanatory variables of the success and failure of economic sanctions.²

In the meta-analysis each parameter estimate that is reported in a primary study is an observation of the dependent variable. In our study this is the reported coefficient of a determinant of sanction success. The meta-analysis enables us to estimate the ‘true’ underlying meta-effects and to test for publication bias as a potential explanation for the heterogeneity of findings in the literature. Publication bias occurs if certain types of statistical results have a higher probability to be reported than other results. It includes – often unconsciously – selection of research findings that satisfy prior beliefs, theoretical expectations or statistical significance (Demena, 2017). This distorts research and our knowledge of sanctions, as large and/or statistically significant, conventional

¹ This data set is often referred to by the acronym HSE or HSEO.

² These studies define success as sanctions that “result in either full target compliance or at least partial policy change in line with the stated policy objectives of senders” (Peksen 2019, p. 637).

results would be over-represented whereas controversial, small and/or insignificant findings are likely to be under-reported.

Rather than focussing on only one variable of interest we meta-analyse three variables simultaneously because by dealing with three meta-analyses on related variables of interest we are able to investigate whether our findings indicate a common trend of this subset of the literature or a finding that is specific for, e.g., duration only. The literature on economic sanctions provides a great many potential determinants of sanction success including characteristics of the sanction goal, political characteristics and economic aspects. It is for practical reasons impossible to cover all these potential determinants in one study so we have to be selective and we will focus on three key economic determinants. We analyse three important determinants of the outcome of sanction episodes for which we provide a meta-analysis each:

- pre-sanction trade linkage between sanction sender and target,
- duration of the sanction episode and
- prior relations between sanction sender and target.

The selection of these variables is based on three reasons. Firstly, the findings for these variables are robust with respect to the different vintages of the Hufbauer et al. (1990, 1995 and 2007) data set that has habitually been used in empirical sanction research and is often included in the more recent datasets (van Bergeijk and Siddiquee, 2017). Secondly, we can unambiguously ground these variables in the economic theory of trade. Thirdly, the pattern of estimated coefficients for each determinant reveals continued disagreement among the primary studies and this makes a meta-analysis more valuable. Figure 1 illustrates this disagreement, as it plots on the vertical axis the t -value of the reported coefficients for each of our three variables of interest and on the horizontal axis the year of publication of the primary study. We report t -values, because this helps us to directly focus on sign and statistical significance and also to see if disagreement is substantial. Note that as an additional benefit, t -values are dimensionless parameters and therefore the comparison is not compromised by differences in operationalization of the variables.

Moreover, Figure 1 provides an overview of the development of the estimated parameters over time. While the majority of these estimates are in line with our view that successful economic sanctions are associated with larger trade linkage, shorter duration and better prior relations, the figure also clearly reveals substantial and persistent controversy with respect to both the significance and the sign of each of the three parameters.³ Since meta-analysis is a relatively unknown methodology for the study of sanctions, we first discuss the methodology in Section 2. Section 3 discusses data collection and our empirical strategy in the context of the MAER-Net protocol. Section 4 provides the meta-analysis and establishes the extent of bias. Section 5 then delves into the

³ Bapat et al. (2013), for example, have observed similar inconclusiveness and used a theory-free extreme bounds analysis of 18 potential determinants of sanction success to investigate the sensitivity of the empirical findings regarding the choice of variables in regressions that are used to explain sanction success.

determinants of publication bias. The final section provides a discussion on the implications of our findings and offers suggestions for methodological changes in the field.

FIGURE 1
Reported t-values for three determinants of sanction success
and year of publication of the primary studies (1985-2018)



2 Method and methodology

The sanctions literature mainly relies on so-called narrative evaluations or reviews of literature. Such reviews may or may not be based on a systematic review, that is: an identification of all potentially relevant articles based on keywords and typically using search engines such as the *Web of Science* or *Google Scholar*. Part of the systematic review consists of the long-standing best practice to identify potentially relevant work using citation analysis and an inspection of authors identified in the references of the primary studies. This body of literature is summarized and evaluated by an expert (or group of experts) with the aim to identify the state of the art regarding a topic. The evaluators consider the quality and reliability of research findings (in principle including the risks for bias) and often identify promising avenues for research. These narrative reviews are extremely important and relevant for scientific progress, also because they can combine and evaluate different forms of knowledge including qualitative, quantitative and formal (mathematical) findings.

Meta-analysis was developed in medicine where it was used to establish the true effect out of many small- N samples of drug tests. The method quantitatively investigates empirical studies and statistically controls for biases. Therefore, a meta-analysis enables one to establish objectively the central tendency in the empirical literature and to make corrections for bias introduced by intrinsic motivation, bias of the authors of the primary studies themselves as well as bias introduced by editors and reviewers during the referee. The meta-analysis both comprises and extends beyond a traditional systematic review (van Bergeijk and Lazzaroni, 2015). A well-defined meta-analysis augments a traditional review of the literature: a meta-analysis is less susceptible to subjective predisposition and more transparent than a narrative review of the literature, as it methodically analyses causes of (quantitative) discrepancies of the primary studies.

Our meta-analysis is the first to deal with economic sanctions. This is a relevant topic for meta-analysis in view of the apparent disagreement on the success of sanctions that does not only differ by ‘school’, but also over time shows significant fluctuation in policy discussions that often are strongly driven by recent high-profile cases.⁴ In this context subjectivity can become problematic.

We will focus on three key economic determinants of sanction impact. Underlying our analysis is the assumption that costs and benefits of considered policies motivate decision makers and therefore we focus on economic welfare (aggregate utility) as a driver of behaviour of the sanction target. (We do of course not argue that economic costs and benefits are the only relevant variables and recognize that many other non-economic variables can determine the outcome of concrete sanction episodes.) An appendix with the mathematical modelling details is available upon request. Basically, the standard neoclassical model in the context of trade uncertainty can be used to motivate the choice of the variables of interest for the meta-analysis, and also to derive theoretical expectations from core economic theory regarding their signs. Our

⁴ See Wallenstein (2000) on the 30-year cycle of sanction research.

a priori expectations are that sanctions are more likely to succeed for larger trade linkage, shorter duration and better prior relations. Based on a standard economic model we have therefore strong priors regarding the sign and significance of the variables of interest – and this makes the findings of the meta-analysis even more challenging.

3 Data and empirical strategy

3.1 Methods, protocols and data construction

The construction of the dataset strictly follows the guidelines of the Meta-Analysis in Economics Research-network (MAER-Net); see Stanley et al. (2013). We conducted a comprehensive literature search using Google Scholar supplemented with the Web of Science. Moreover, we checked the lists of the references of recent primary empirical studies and reviews. The search included all potentially relevant published and unpublished empirical primary studies from 1985 up to and including 2018. We searched using different broad keyword combinations on the three variables of interest (trade-linkage, duration and prior relations) as illustrated in Table 1. Studies are included if they satisfy the following selection criteria: English language, empirical investigations that are conducted on the success (failure) of economic sanctions and include at least one of our variables of interest (either as variable of interest in the primary study or as controlling variable) and that report regression-based coefficients, sample size, t-statistics or standard errors.⁵ The application of these criteria resulted in 33 (trade-linkage), 15 (duration) and 24 (prior-relations) empirical studies for coding. The multiple search process ended February 2019. All authors of this working paper were involved in the construction of the dataset providing for double coding and consistency check.

TABLE 1
Search list of keywords, selected primary studies and collected observations

Database	Category	Keywords	Results returned (selected)	N
Google Scholar & Web of Science (ISI Web of Knowledge)	trade linkage	Economic sanctions, economic-coercion*, sanction* threat, success or failure*, work, sanction and outcomes*, episodes, determinant*, cost*, result*, culture*, effectiveness*	198 (33)	174
	duration	economic sanctions, sanctions*, success of economic sanctions, sanction*outcome*, sanction* duration, sanction time, sanctions episode*, sanctions imposition*, length sanction episode*	210 (15)	77
	prior relations	Economic sanctions, economic coercion, sanction*, episodes, determin*, success*, fail*, effect*, work, out-comes, result*, cost*, sender state, target state, foreign, *politic*, democratic*, autocrat*, *leader*, *stability, empirical analysis, sensitivity analysis, approach, econometric analysis, modelling	360 (24)	83

⁵ We excluded descriptive and qualitative studies and studies that consider economic sanctions as exogenous variable (e.g., see Afesorgbor, 2019). For missing data, we contacted the authors of the primary studies, enabling us to include 22 observations.

We use the so-called *all-set*, constructed through coding all relevant regressions in all studies, because the number of available studies is limited, we want to use within-study information and using ‘preferred specifications’ introduce selection bias based on the author’s preferences that can be avoided by simply taking all reported coefficients (Demena, 2015; Demena and van Bergeijk, 2017; Floridi et al., 2020).

3.2 Meta-data

Our dataset consists of 334 observations derived from 37 primary studies for which the required date (estimated coefficients) are provided. Only 2 of the 37 studies concern a country-specific study. All other studies are large- N studies. The earliest study in our sample was published in 1985 and the most recent in 2018. The median study in our sample appeared in 2007 so that half of the primary studies was published in the last ten years, illustrating again the topicality of this research field. The median number of parameter estimates taken from a primary study is 5 coefficients. The minimum, the mean and the maximum number of observations (regressions) per study are 1, 9 and 48 coefficients, respectively.

The data set includes 30 peer-reviewed journal articles, illustrating that the primary studies are predominantly published in peer-reviewed journals.⁶ In total this literature had received 4523 citations in Google Scholar⁷ as of March 2019. The study *Economic Sanctions Reconsidered: History and Current Policy* (all editions combined) received most citations (2,065 citations). The second most cited article is Drezner’s 2000 *Bargaining* article (310 citations). Slightly more than forty percent of the reported coefficients were published in A-journals.⁸ Table 2 provides a summary of the qualitative results of the empirical studies published in A-journals and again illustrates the lack of consensus in the literature – even for the leading journals – with contradictory significant signs and many insignificant coefficients.

A meta-analysis employs both visual inspection of graphs and statistical analysis, as prescribed by the MAER-net protocol. We use funnel plots (scatter diagrams with the effect size on the horizontal axis and its precision on the vertical axis) to get a first indication about publication bias. In a non-biased literature, the estimates will vary randomly and symmetrically around the median: imprecise estimates will show up scattered widely and symmetrically at the lower part of the funnel and more precise estimates will be compactly distributed at the upper part of the funnel. An asymmetric funnel plot indicates bias. Next, we will use more objective statistical procedures to assess

⁶The other studies were 5 books, 1 PhD dissertation and 1 ‘cautionary note’.

⁷ Google Scholar also covers books and grey literature and thus gives a better indication of impact.

⁸ An A journal is here defined as a journal with a listing in the top third of an ISI category in 2017.

publication selection bias. An appendix is available upon request with the technical econometric details.

TABLE 2
Qualitative findings of empirical studies published in high-quality journal

Study	Journal	Findings		
		Trade linkage	Prior relations	Duration
Dashti-Gibson et al. (1997)	<i>American Journal of Political Science</i>	*		*
Drury (1998)	<i>Journal of Peace Research</i>	*	*	
Hart (2000)	<i>Political Research Quarterly</i>	+	*	
Nooruddin (2002)	<i>International Interactions</i>	*		*
Jing et al (2003)	<i>Journal of Peace Research</i>		+	
Lektzian and Souva (2007)	<i>Journal of Conflict Resolution</i>	*		
Ang and Peksen (2007)	<i>Political Research Quarterly</i>	*	+	*
Bapat and Morgan (2009)	<i>International Studies Quarterly</i>	+	*	*
Major (2012)	<i>International Interactions</i>	+		
Whang et al (2013)	<i>American Journal of Political Science</i>	+		
Lektzian and Patterson (2015)	<i>International Studies Quarterly</i>	*	+	*
Bapat and Kwon (2015)	<i>International Interactions</i>	+		
van Bergeijk and Siddiquee (2017)	<i>International Interactions</i>	*	+	–
Kleinberg (2018)	<i>Journal of Peace Research</i>	*	–	
Chan (2009)	<i>International Political Science Review</i>	–		–
Drezner (2000)	<i>International Organization</i>	+	+	
Woo and Verdier (2014)	<i>Journal of Semantics</i>	*	–	
Peterson (2018)	<i>Conflict Management and Peace Science</i>	–		

Notes: * variable included but not significant, - negative and significant at the 10% level and better, + positive and significant at the 10% level and better. Blank cells indicate that this variable is not covered in the primary study.

4 Meta-analysis and bias

Figure 1 and Table 2 already indicated that there is no obvious convergence in the literature towards a consensus parameter value. In this section we investigate if this is due to publication bias.

4.1 Graphical inspections

We start our meta-analysis with Figure 2 that provides the funnel plots with the inverse of standard error (the most commonly used measure of precision) on the vertical axis and the effect size on the horizontal axis. The funnel plot for prior relations is skewed to the right, suggesting bias for reporting positive estimates. For duration the funnel plot is skewed to the left indicating a preference to report negative estimates. The trade linkage funnel plot is skewed to right (this is not directly clear from visual inspection).

FIGURE 2
Funnel plots of the reported estimates on economic sanction success

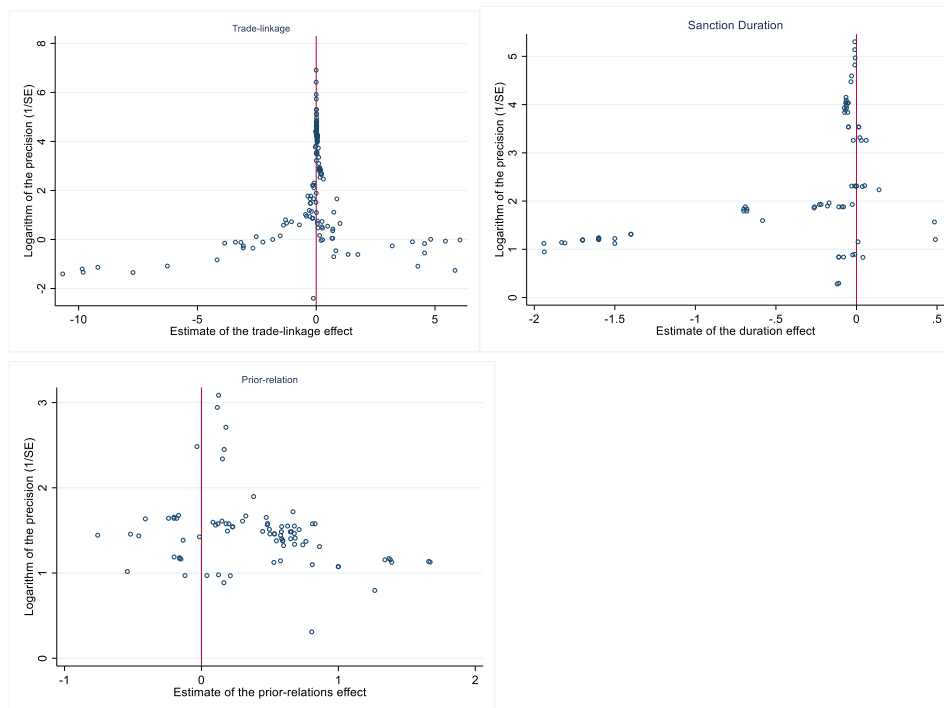


Table 3 offers an additional perspective by means of the summary statistics of the unweighted (simple) average and the weighted average (using the inverse variance; $1/se$) of the coefficients reported in the primary studies. The simple and weighted averages of duration are significant at 1% and suggest that longer duration is associated with sanction failure. The weighted average effect for prior relations is significant at 1% and suggests that better prior relations are

associated with sanction success. Both averages are insignificant for trade linkage. These findings motivate our next step to move to a more formal method to statistically test the issue of publication bias.

TABLE 3
Estimates of the overall reported coefficients on economic sanction

Method	Effect size	Standard Error	N	95% confidence interval	
Trade-linkage					
Simple average effect	-0.203	0.164	174	-0.527	0.120
Weighted average effect	0.014	0.016	174	-0.016	0.045
Duration					
Simple average effect	-0.440**	0.077	77	-0.5948	-0.2852
Weighted average effect	-0.075**	0.028	77	-0.1310	-0.0186
Prior-relations					
Simple average effect	0.400	0.055	83	0.2916	0.5089
Weighted average effect	0.312**	0.047	83	0.2192	0.4042

Notes: the simple average is the arithmetic mean of the reported estimates and the weighted average uses inverse variance as weight. **, * stands for 1 and 5% level of significance, respectively.

4.2 Meta regression analysis

Table 4 provides the results of the MRA for each of the three determinants of sanction success or failure. We report results of mixed-effects multilevel model (MEM) that accounts for both between-study dependence and within-study correlation, the OLS standard errors clustered data analysis (CDA) that accounts for within-study correlation only (see Demena and Afesorgbor, 2020 on the importance of controlling for between-study dependence via the multilevel model).

The findings of the MRA are consistent: for each variable we find substantial publication bias. The first column reports the so-called MEM model: the publication bias is significant at 1% in all cases exaggerating the effect of the variable of interest. In terms of magnitude, the bias ranges from 0.907 to 1.384 in absolute values, implying substantial bias. Doucouliagos and Stanley (2013) argue that selectivity is ‘little to modest’ if the bias is insignificant or less than 1; ‘substantial’ if between 1 and 2; and ‘severe’ if it exceeds 2. Hence the publication bias in this literature is substantial.

The Clustered Data Analysis (CDA) in Table 4 supports the finding of publication bias although at a lower level of significance (Trade linkage is significant at the 10% level). The CDA reports a higher magnitude of the publication bias for duration and prior-relations and a lower magnitude for trade-linkage. In all cases the verdict is that the literature suffers from substantial bias. The genuine effect, moreover, is insignificant in all regressions in Table 4. Therefore, the overall simple and uncorrected weighted effects

reported in Table 3 reflect a publication bias, again implying that the reported estimates of the primary studies are likely to be exaggerated and to substantially overstate the actual impact of the determinants. By way of example: van Bergeijk (1989) reports a coefficient of 1.83 for sanction duration that is significant at a 95% confidence level, but in view of the literature that developed later this finding needs careful reinterpretation in view of the estimated bias of 1.38 (MEM) and 1.76 (CDA). The result also remains significant if we take the bias for the whole empirical literature into account, but it is with hindsight only just large enough to support the original conclusion of the 1989 article. Our next task is to explore the potential source and determinates of the identified substantial publication bias.

TABLE 4
MRA for FAT-PET: publication bias and true effect

Panel A: Trade-linkage				
Variables	MEM		CDA	
	Coefficient	<i>t</i> -value	Coefficient	<i>t</i> -value
Bias (FAT)	0.907**	3.32	0.727	1.96
Genuine effect (PET)	-0.0003	-0.19	0.002	0.66
Observations	174		174	
Studies	33		33	
Panel B: Duration				
Variables	MEM		CDA	
	Coefficient	<i>t</i> -value	Coefficient	<i>t</i> -value
Bias (FAT)	-1.384**	-3.01	-1.7624*	-2.25
Genuine effect (PET)	0.004	0.59	-0.005	-0.51
Observations	77		77	
Studies	15		15	
Panel C: Prior-relations				
Variables	MEM		CDA	
	Coefficient	<i>t</i> -value	Coefficient	<i>t</i> -value
Bias (FAT)	0.893**	3.09	0.988*	2.43
Genuine effect (PET)	0.004	1.05	0.043	0.90
Observations	83		83	
Studies	24		24	

Notes: **, * stands for 1 and 5% level of significance, respectively. All estimates use the inverse variance as weights and standard errors are clustered at study level. Reported *t*-values are from cluster-robust standard errors. MEM is mixed-effects multilevel estimated via restricted maximum likelihood; CDA: clustered data analysis (robust standard errors clustered at study level)

4.3 Sources of bias

Next, we turn to the relevance of characteristics that at the individual study level could be associated with publication bias. A similar approach was employed by Doucouliagos and Stanley (2013) in their meta-analysis

investigating the determinants of publication bias in economics. We focus of course on a much narrower literature, namely the success (failure) of economic sanctions.

We define the dependent variable (publication bias) as the absolute value of the estimate-level of bias. This is the difference between on the one hand, the underlying genuine meta-effect estimated in Table 4 and on the other hand, the individual reported coefficients of each regression of the primary empirical studies. Due to reporting characteristics of the field, it is not possible to estimate study-level publication bias using the intercept reported in Table 4 because only in 8 of the 36 studies included in our sample reported more than 10 parameter estimates per study and the vast majority of studies provided between 1 and 5 estimates. Importantly, our method does allow us to deal with the heterogeneity of the individual estimates that may not be uncovered by the study-level approach. Table 5 reports potential sources of bias and their summary statistics.

TABLE 5
Summary statistics of explanatory variables

Variables	Mean	Std. Dev.	Min	Max
Study was published in a peer-reviewed journal	0.689	0.464	0	1
No. of citations of the lead or corresponding author ^a	1.815	3.120	0	11.548
No. of observations used by the study ^a	0.340	0.537	0.019	3.658
A study is co-authored	0.275	0.447	0	1
A cited author is affiliated with US based institution	0.626	0.485	0	1
A cited author is affiliated with an academic institution	0.793	0.405	0	1
A cited author has completed PhD ≥ 5 years ago	0.602	0.490	0	1
High journal rank, 2017 ISI impact factor	0.425	0.495	0	1
Homogeneous gender (not mixed)	0.907	0.290	0	1
A cited author is a political scientist (base, economist/others)	0.692	0.463	0	1
The publication year of the study (base, 1985)	23.620	8.393	0	33
Idem squared	628.14	335.14	0	1089

^a Mean, standard deviation and max are divided by a thousand to make the figures easier to read.

Peer review: Peer review is often expected to improve the quality and reliability of the reported results, but it could also be conservative and avoid results that challenge consensus. We test this factor by means of a binary dummy variable that assumes the value 1 if a study was published in a peer-reviewed journal (76% of the estimates was published in a peer-reviewed study). We also include the quality of the outlets using the ISI impact factor of the journals to test if lower and higher impact factors are systematically associated with the pattern of publication selectivity. High quality (A-journals) ranked form the top quartile cited outlets of the 2017 ISI impact factor.

Research team: Research teams are more likely to consider results in a balanced way than single authors because the team is able to provide within-team peer review. Therefore, we include a binary dummy variable that assumes the value 1 if the studies are co-authored (72% of estimates were obtained from single-authorship). We further investigate this issue by considering the gender composition of the team assuming that single-sex teams are less heterogeneous and may therefore reach less balanced results (9% of the teams is mixed). Jarrell and Stanley (2004), provide evidence that gender (composition) is an important factor of bias (see, however, Medoff, 2003 for a contrary opinion)

Author characteristics: An important issue could be that differences in academic field may result in different reporting standards (Moons and van Bergeijk, 2017). We test this assumption with a binary dummy that assumes the value 1 if the author is a political scientist (opposed to other professions such as, economist, sociologist, etc.). In 69% of the reported estimates researchers have a political science background). Doucouliagos and Paldam (2010) have shown the importance of the affiliation of researchers. Following Stanley (2005), we create binary dummies that assume the value 1 if an author is affiliated with an US based institution and if a study author works at an academic institution, respectively. The former accounted for two third of the reported estimates, and the latter obtained nearly in four of the five observations. We collected this from the affiliations provided by the authors at the moment of the publication of the studies.

Next, we consider the PhD status of the cited authors of the studies at the date of the publication of a particular study. Authors of the studies are categorized into two groups: those who have completed their PhD more than or equal to 5 years ago at the time when their study was published (60%), and those who have completed less than 5 years ago including those who have not completed PhD (40%). We expect the reported estimates published by authors who are at their early stage of their career to exhibit more selectivity/publication bias. Finally, we consider the reputation of the lead/corresponding author of the primary studies by including the log of total citations for each author that we collect from their Google Scholar profile.

Study characteristics⁹: We also consider the number of observations of the study to test for systematic variation between small and large-N. We expect studies with large samples to show less selectivity bias as these studies are more likely to produce significant estimates without much search for econometrics specification to support prior believe or expectation.

Finally, we control for the publication year of the study. Goldfarb (1995) formulated the research-cycle hypothesis related to the issue of novelty and fashion in academic research. The hypothesis proposes that while seminal studies often produce large and significant estimates, sceptical studies will

⁹ A technical appendix with robustness analyses is available upon request.

follow up implying a downward trend in the bias over time. However, if this is not the case, we will observe a positive trend, that is: a pattern by which reported estimates become more biased over time.

Table 6 reports the results of the reduced multivariate MRA using general-to-specific (GTS) modelling. GTS modelling starts with a general specification (4) in which all potential moderator variables are included. Next, one at a time, the statistically most insignificant variables are removed, until we arrive at a reduced specification that contains significant variables only. During GTS, we note that nine of the 13 determinants of publication bias included in the analysis are not statistically significant. The joint insignificance of these variables yields $F_{(9, 319)} = 1.07$, while the joint test of the four included determinants of publication bias rejects the null hypothesis of a zero joint effect with $F_{(4, 329)} = 8.24$.

TABLE 6
Determinants of publication bias (restricted maximum likelihood)
Dependent variable = $|\beta_{1j}|$

Moderator variable	(1)	(2)	(3)	(4)	(5)
	Specific	Specific	MEM	Robust se	CDA
Peer Reviewed	-0.467* (0.174)	-0.329* (0.172)	-0.188 (0.361)	-0.329* (0.142)	-0.329 (0.277)
Co-authored	-0.519* (0.180)	-0.489* (0.175)	-0.468 (0.294)	-0.489** (0.154)	-0.489 (0.302)
Journal Rank	0.695** (0.177)	0.487* (0.178)	0.447 (0.332)	0.487** (0.154)	0.487 (0.287)
Publication year (base 1985)	0.018* (0.009)	-0.181** (0.045)	-0.140* (0.071)	-0.181** (0.046)	-0.181* (0.076)
Idem squared		0.005** (0.001)	0.004* (0.001)	0.005** 0.001	0.005* (0.002)
Observations	334	334	334	334	334
Number of studies	37	37	37	37	37

Notes: **, * stand for 1% and 5% level of significance, respectively. Constant term included but not reported. (Standard errors in brackets).

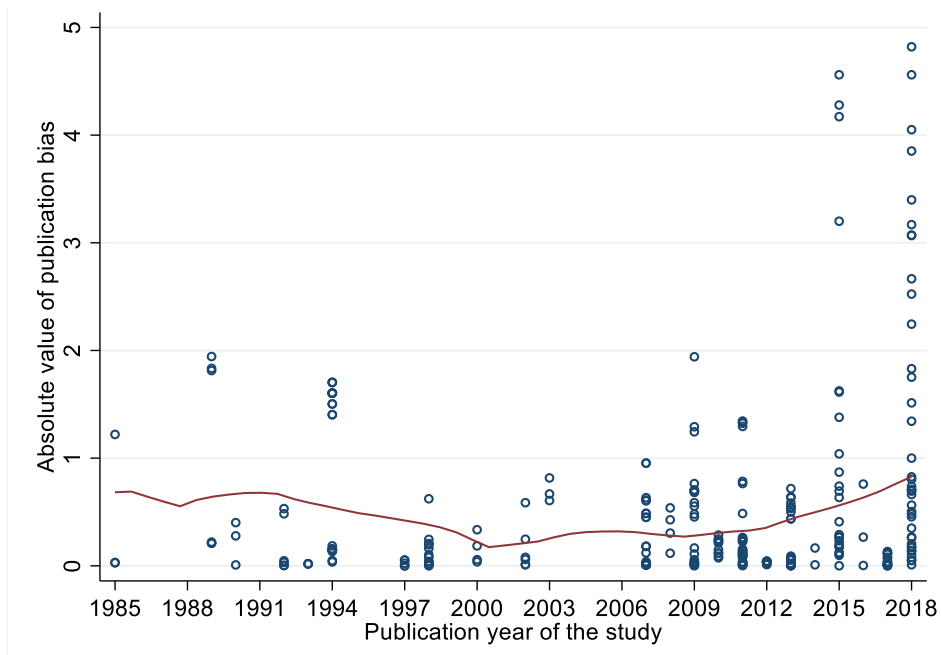
^a Coefficients and standard errors are divided by a thousand to make the figures easier to read.

The dependent variable is the absolute value of the estimate-level bias derived from the difference between the prediction of regression of Table 4 for PET and the individual reported coefficients of each regression of the primary empirical studies. Columns 1 and 2 are the reduced/specific model using the general-to-specific approach without adjusting standard error. Column 3 is mixed-effects multilevel model using the restricted maximum likelihood, Columns 4 and 5 gives robust standard error and clustered standard error data analysis at study-level. All columns include dummies (not-reported) of trade-linkage and duration with prior relation as a reference category.

The joint test of the 4 included variables in column 1 rejects the null hypothesis of a zero joint effect $F_{(4, 329)} = 8.24$ at any conventional level (p -value=0.000). The joint test of the other 9 excluded variables supports the joint insignificance of these variables - $F_{(9, 319)} = 1.07$ with p-value 0.3820. Table 6 Columns 1 and 2 give the results of the specific model. This model is then re-estimated using the preferred MEM model (Column 3) and, for comparison and robustness check, with robust standard errors (Column 4) and CDA (column 5). In all the regressions, we use prior relations as the reference category.¹⁰

Columns 1-2 for the specific model show that peer review is always negative suggesting at first sight that peer review reduces bias, but we find no significant impact of peer review when we control the with-in and between-study correlations (Columns 3 and 5). In a sense this is disappointing, given the efforts of the referees, but – on the positive side – it also indicates that it is not the peer-review process per se that creates the bias in this literature. The same is true for most of the other variables that are often considered to be of relevance in the literature on publication bias. Our most important finding is that bias over time *increases* in this literature.

FIGURE 3
Kernel plot of the development of bias 1985-2018



¹⁰ We included dummy variables for trade linkage studies and duration studies to correct for potential dissimilarity across variables of interest. Regardless of the reference variable, the coefficients are always insignificant, and they drop out in the General to Specific procedure in column 5.

In order to test the later conclusion, we also use a non-linear relationship by including the square of publication year in column 2. The non-linear specification shows a 'U-shaped' pattern of bias over time: initially bias reduces as suggested by the research cycle hypothesis. However, since 2000 the literature moves in a different direction and publication bias increasingly becomes a serious concern as illustrated in Figure 3.

5 Conclusions and issues for further research

The results of our meta-analysis are far-reaching. While the descriptive analysis and weighted averages suggest that the impact of each of the three variables of interest is significant and conforms to our *a priori* expectations, the econometric investigation uncovers significant bias in the reported findings of the original studies. This distortion is so sizeable that we cannot even determine the signs of the trade linkage, sanction duration and prior relations. Based on our meta-analysis the conclusion is therefore that the primary studies on the whole have not established a significant impact of the variables of interest on the success and failure of economic sanctions. This is a challenging finding that goes much further than Pape's (1997) argument on the classification of (in)effectiveness of sanctions in the HSE data set.

Does this mean that we do not know anything? First of all, we now know that "we know less" than we thought we did, and that means that we have actually learned that the studies published in the 1990s were less convincing than initially thought. The early studies, for example, suggest a very strong link between trade linkage and sanction success. Up till 1995 all reported coefficients are positive and most are quite significant, but the new methods and data sets that have been introduced over time lead to studies that on average tend to report more negative and/or less significant relationships. This is one reason why publication bias increases when we expand the time period from which we meta-analyse our studies. In this sense the research on sanctions has been at least to some extent cumulative.

The topic of success and failure of economic sanctions fits in the recent trend where research synthesis and/or replication fail to reproduce well-accepted results in the literature; the so-called 'credibility crisis' that is apparent in many scientific disciplines (Gerber et al. 2001; Ioannidis 2005; Christensen and Miguel 2018).

We reject Goldfarb's research cycle hypothesis, as Table 6 and Figure 3 find that bias increases over time, especially so in more recent times. Initially only a few studies exist that appear to provide guidance and establish consensus (up till the mid-1990s the evidence is clearly supporting our *a priori* expectations for the variables of interest) but starting in 2000 and especially since 2010 the heterogeneity of results has increased sharply. This makes publication bias an urgent issue for the literature on sanctions. Indeed, our findings suggest that the credibility crisis has reached sanction studies and that action is necessary and we would like to suggest some possible solutions.

First, Doucouliagos and Stanley (2013) argue that publication bias is widespread and imminent if competition among rival theories is not sufficiently strong. Alternative heterodox theories of sanctions should therefore get a better chance of publication. More competition and especially debate between rival/alternative theories could reduce selectivity and improve inference.

Second, the standard of evidence needs to be increased. Presently the standard is to accept results at the 90% confidence level and better, but this may not be a sufficiently stringent statistical test and therefore a higher confidence level should be required.

Third, the literature needs an alternative data source. The empirical literature, as rightly observed by Peksen (2019), is by and large based on three data collections. The available large- N studies mainly use the Hufbauer and Schott data base, the Threats and Imposition of Economic Sanctions data base and their derivatives. The large- N studies are the dominant research technology in the field, but the sample is still relatively small and empirical studies typically revisit the existing data sets. An important new data set, the Global Sanctions Dataset, has recently become available for research (Feldbermayr et al. 2019; Kirilakha et al. 2021) and this will be an important improvement, but an alternative methodology would still be very much needed. In order to make progress towards an alternative individual country studies should be stimulated. The body of individual country studies would then provide a methodological alternative for the existing large- N data sets. Meta-analysis and other forms of research synthesis could be used to determine the meta-effect based on these studies.

Finally, we suggest that future research needs to carefully understand the issue or pattern of publication bias and we underscore that policy makers and researchers that use empirical findings need to consider the extent of bias that we have found.

References

Studies included in the meta-analysis has been marked

- Afesorgbor, S.K., 2019, 'The impact of economic sanctions on international trade: How do threatened sanctions compare with imposed sanctions?', *European Journal of Political Economy* **56** (January), pp. 11-26.
- # Allen, S.H., 2008, 'Political institutions and constrained response to economic sanctions', *Foreign Policy Analysis* **4** (3), pp.255-274.
- # Ang, A. U-J. and D. Peksen, 2007, 'When do economic sanctions work? Asymmetric perceptions, issue salience, and outcomes', *Political Research Quarterly* **60** (1), pp. 135-145.
- # Bapat, N.A. and T. C. Morgan, 2009, 'Multilateral versus unilateral sanctions reconsidered: A test using new data', *International Studies Quarterly* **53** (4), pp.1075-94.
- Bapat, N.A., T. Heinrich, Y. Kobayashi, and T. C. Morgan, 2013, 'Determinants of sanctions effectiveness: Sensitivity analysis using new data', *International Interactions* **39**, (1) (2013), pp. 79-98.
- # Bapat, N.A., and B.R. Kwon, 2015, 'When are sanctions effective? A bargaining and enforcement framework', *International Organization* **69** (2), pp. 131-162.
- # Bergeijk, P.A.G. van, 1989, 'Success and failure of economic sanction', *Kyklos* **42** (3), pp. 385-404.
- # Bergeijk, P.A.G. van and M.S.H. Siddiquee, 2017, 'Biased sanctions? Methodological change in economic sanctions reconsidered and its implications', *International Interactions* **43** (5), pp. 879-893.
- # Bergeijk, P.A.G. van 1994, *Economic diplomacy, trade and commercial policy: Positive and negative sanctions in the new world order*. Edward Elgar: Cheltenham.
- # Bergeijk, P.A.G. van (2009) *Economic diplomacy and the geography of international trade*, Edward Elgar: Cheltenham.
- Bergeijk, P.A.G. van, 2012, 'Opârg fitfara eft ksvurf, *Ekonomiy ur Dônopros*', **72** (8), p. 6.
- Bergeijk, P.A.G., van, 2019, 'Can the sanction debate be resolved?', *CESifo Forum* **20** (4), pp. 3-8.
- Bergeijk, P.A.G. van and S. Lazzaroni, 2015, 'Macroeconomics of natural disasters: strengths and weaknesses of meta-analysis versus review of literature', *Risk Analysis*, **35** (6), pp. 1050-72.
- Bergeijk, P.A.G. van and C. van Marrewijk, 1995, 'Why do sanctions need time to work? Adjustment, learning and anticipation', *Economic Modelling* **12** (2), pp. 75-86.
- Bergeijk, P.A.G. van, Demena, B.A., Reta, A., Jativa, G.B. and Kimararungu, P., 2019, 'Could the literature on the economic determinants of sanctions be biased?' *Peace Economics, Peace Science and Public Policy*, **25**(4), doi: 10.1515/peps-2019-0048
- Biersteker, T.J., S.E. Eckert, M. Tourinho and Z. Hudáková, 2018, 'UN targeted sanctions datasets (1991–2013)', *Journal of Peace Research* **55**(3)), pp. 404-412.
- # Bonetti, S., 1998, 'Distinguishing characteristics of degrees of success and failure in economic sanctions episodes', *Applied Economics* **30** (6), pp. 805-813.
- # Chan, S., 2009, 'Strategic anticipation and adjustment: Ex ante and ex post information in explaining sanctions outcomes', *International Political Science Review* **30** (3), pp. 319-338.

- Christensen, G. and E., Miguel, 2018, 'Transparency, reproducibility, and the credibility of economics research', *Journal of Economic Literature* **56** (3), pp. 920-80.
- # Dashti-Gibson, J., P. Davis and B. Radcliff, 1997, 'On the determinants of the success of economic sanctions: An empirical analysis', *American Journal of Political Science* **41** (2), pp. 608-18.
- # Dehejia, R.H. and B. Wood, 1992, 'Economic sanctions and econometric policy evaluation: a cautionary note', *Journal of World Trade* **26** (1), p. 73. #
- Demena, B.A. and P.A.G. van Bergeijk, 2017, 'A meta-analysis of FDI and productivity spillovers in Developing countries', *Journal of Economic Surveys* **31** (2), pp. 546-71.
- Demena, B.A., 2015, 'Publication bias in FDI spillovers in developing countries: a meta-regression analysis', *Applied Economics Letters* **22** (14), pp. 1170-74.
- Demena, B.A., 2017, *Essays on intra-industry spillovers from FDI in developing countries: a firm-level analysis with a focus on Sub-Saharan Africa*, PhD diss., Erasmus University, The Hague.
- Demena, B.A. and S.K. Afesorgbor, 2020, 'The effect of FDI on environmental emissions: Evidence from a meta-analysis', *Energy Policy* **138**.
- Dizaji, S.F. and P.A.G. van Bergeijk, 2013, 'Potential early phase success and ultimate failure of economic sanctions: a VAR approach with an application to Iran', *Journal of Peace Research* **50** (6), pp. 721-36.
- Doucouliaqos, C. and T.D. Stanley, 2013, 'Are all economic facts greatly exaggerated? Theory competition and selectivity', *Journal of Economic Surveys* **27** (2), pp. 316-39.
- Doucouliaqos, C. and M. Paldam, 2010, 'Conditional aid effectiveness: a meta-study', *Journal of International Development* **22** (4), pp. 391-410.
- # Drezner, D.W., 2000, 'Bargaining, enforcement, and multilateral sanctions: when is cooperation counterproductive?', *International Organization* **54** (1), pp. 73-102.
- # Driscoll, D., D. Halcoussis and A.D. Lowenberg, 2010, 'Economic sanctions and culture', *Defence and Peace Economics* **22** (4), pp. 423-48.
- # Drury, A.C., 1998, 'Revisiting economic sanctions reconsidered', *Journal of Peace Research* **35** (4), pp. 497-509.
- # Early, B.R., 2011, 'Unmasking the black knights: Sanctions busters and their effects on the success of economic sanctions', *Foreign Policy Analysis* **7** (4), pp. 381-402.
- Felbermayr, G., C. Syropoulos, E. Yalcin and Y.V. Yotov, 2019, 'On the effects of sanctions on trade and welfare: New evidence based on structural gravity and a new database', *Kiel Working Paper, No. 2131*, Kiel Institute for the World Economy: Kiel.
- Floridi, A., Demena, B.A. and Wagner, N., 2020. 'Shedding light on the shadows of informality: A meta-analysis of formalization interventions targeted at informal firms'. *Labour Economics*, **67**, p.101925.
- Gerber, A.S., D.P. Green, and D. Nickerson, 2001, 'Testing for publication bias in political science', *Political Analysis* **9** (4), pp. 385-92.
- Goldfarb, R.S., 1995, 'The economist-as-audience needs a methodology of plausible inference', *Journal of Economic Methodology* **2** (2), pp. 201-22.
- # Hart Jr, R.A., 2000, 'Democracy and the successful use of economic sanctions', *Political Research Quarterly* **53** (2), pp. 267-84. #
- # Hufbauer, G.C. and J.J. Schott, 1985, *Economic sanctions reconsidered. History and Current Policy*. Peterson Institute: Washington DC.

- Hufbauer, G.C., J.J. Schott and K.A. Elliott, 1990, *Economic sanctions reconsidered: History and current policy* (Vol. 1). Peterson Institute: Washington DC.
- # Hufbauer, G.C., J.J. Schott, K.A. Elliott and B. Oegg, 2007, *Economic Sanctions Reconsidered: [in two volumes]* (2nd edition) Institute for International Economics, Washington DC.
- Hufbauer, G.C. and E. Jung, 2021, 'Economic sanctions in the 21st century' in: P.A.G. van Bergeijk (ed.) *Research Handbook on Economic Sanctions*, Edward Elgar: Cheltenham, Chapter 2
- Ioannidis, J.P.A., 2005, 'Why most published research findings are false', *PLoS medicine* **2** (8), p. e124.
- Jarrell, S.B. and T.D. Stanley, 2004, 'Declining bias and gender wage discrimination? A meta-regression analysis', *Journal of Human Resources* **39** (3), pp. 828-38.
- # Jing, C., W.H. Kaempfer and A.D. Lowenberg, 2003, 'Instrument choice and the effectiveness of international sanctions: a simultaneous equations approach', *Journal of Peace Research*, **40** (5), pp. 519-35.
- # Kim, H.M., 2009, 'Determinants of the success of economic sanctions: an empirical analysis', *Journal of International and Area Studies* **16** (1), pp. 29-51. #
- Kirilakha, A., G.J. Felbermayr, C. Syropoulos, E. Yalcin, and Y.V. Yotov, 2021, 'The Global Sanctions Data Base: An update to include the years of the Trump presidency' in: P.A.G. van Bergeijk (ed.) *Research Handbook on Economic Sanctions*, Edward Elgar: Cheltenham, Chapter 4
- # Kleinberg, K.B., 2018, Trade, credit, and the effectiveness of sanction threats, <http://www.katjakleinberg.com/uploads/3/7/8/9/37890105/jprsubmission.pdf>
- # Lam, S.L., 1990, 'Economic sanctions and the success of foreign policy goals: a critical evaluation', *Japan and the World Economy* **2** (3), pp. 239-48.
- # Lektzian, D. and D. Patterson, 2015, 'Political cleavages and economic sanctions: The economic and political winners and losers of sanctions', *International Studies Quarterly* **59** (1), pp. 46-58.
- # Lektzian, D. and M. Souva, 2007, 'An institutional theory of sanctions onset and success', *Journal of Conflict Resolution* **51** (6), pp. 848-71.
- # Major, S., 2012, 'Timing is everything: Economic sanctions, regime type, and domestic instability', *International Interactions* **38** (1), pp. 79-110.
- Medoff, M.H., 2003, 'Collaboration and the quality of economics research', *Labour Economics* **10** (5), pp. 597-608.
- Moons, S.J.V. and P.A.G. van Bergeijk, 2017, 'Does economic diplomacy work? A meta-analysis of its impact on trade and investment', *The World Economy* **40** (2), pp. 336-68.
- Morgan, T. C., N. Bapat and V. Krustev (2009) 'The threat and imposition of economic sanctions, 1971–2000', *Conflict Management and Peace Science* **26** (1), pp. 92–110.
- Morgan, T.C, N. Bapat and Y. Kobayashi, 2014, 'The threat and imposition of sanctions: Updating the TIES data set', *Conflict Management and Peace Science* **31** (5), pp. 541–58.
- Morgan, T.C., N. Bapat, and Y. Kobayashi, 2021, 'The threat and imposition of economic sanctions data project: A retrospective' in: P.A.G. van Bergeijk (ed.) *Research Handbook on Economic Sanctions*, Edward Elgar: Cheltenham, Chapter 3
- # Nooruddin, I., 2002, 'Modeling selection bias in studies of sanctions efficacy', *International Interactions* **28** (1), pp. 59-75.

- Pape, R.A., 1997, 'Why economic sanctions do not work', *International security* **22** (2), pp. 90-136.
- Peksen, D., 2019, 'When do imposed economic sanctions work? A critical review of the sanctions effectiveness literature', *Defence and Peace Economics* **30** (6), pp. 635-647.
- # Peterson, T.M., 2018, 'Reconsidering economic leverage and vulnerability: Trade ties, sanction threats, and the success of economic coercion', *Conflict Management and Peace Science* doi: 10.1177/0738894218797024.
- # Shagabudinova, E. and J. Berejikian, 2007, 'Deploying sanctions while protecting human rights: Are humanitarian "smart" sanctions effective?' *Journal of Human Rights* **6** (1), pp. 59-74.
- Stanley, T.D., 2005, 'Beyond publication bias', *Journal of Economic Surveys* **19** (3), pp. 309-45.
- Stanley, T.D., et al., 2013, 'Meta-analysis of economics research reporting guidelines', *Journal of Economic Surveys* **27** (2), pp. 390-394.
- Stanley, T.D. and C. Doucouliagos, 2012, *Meta-regression analysis in economics and business*, Routledge: Oxford.
- Wallensteen, P., 1968, 'Characteristics of economic sanctions', *Journal of Peace Research* **5** (3), pp. 248-267.
- Wallensteen, P., 2000, A century of economic sanctions: A field revisited Department of Peace and Conflict Research: Uppsala.
- # Whang, T., E.V. McLean and D.W. Kuberski, 2013, 'Coercion, information, and the success of sanction threats', *American Journal of Political Science* **57** (1), pp. 65-81.