Bias and Methodological Change in Economic Sanction Reconsidered

Peter A.G. van Bergeijk and
Muhammad Shahadat Hossain Siddiquee

Abstract

The authors investigate the influence of case selection and (re)coding for two vintages of a key resource for research on economic sanctions: the Peterson Institute data base reported in Hufbauer et al. (2nd edition in 1990 and 3rd edition in 2007). The Peterson Institute has not reported transparently on these changes. At the level of individual case studies the authors uncover a tendency to inflate success scores, reclassifying failures into successes even when the evidence for doing so was not convincing. At the level of the aggregated case studies and general methodology they uncover positive bias (that is: methodological changes that make it more likely to find sanction success as indicated by a higher success score, either on average or in individual cases): splitting of episodes into cases and the changed definition of sanction contribution increase the success ratio in general and ultimately the share of sanctions that are judged to be a success. The authors also show the importance of the reclassification of 'destabilization cases' into 'regime changes'. Their probit analysis shows that the 3rd edition's methodology underestimates the contribution of certain sanction characteristics, including the positive impact of the costs of sanctions to the sender, duration of the sanctions and the sender's companion policies.

Authors

Peter A.G. van Bergeijk, ✉ International Institute of Social Studies of Erasmus University Rotterdam, The Netherlands, Bergeijk@iss.nl
Muhammad Shahadat Hossain Siddiquee, Department of Economics, University of Dhaka, Bangladesh

1. Introduction

This paper analyses the 2nd and 3rd (most recent) editions database on international economic sanctions of the Peterson Institute for International Economics, formerly the Institute for International Economics and the publisher of the publications of the studies by Hufbauer et al 1985, 1990 and 2007. Ever since the seminal publication *Economic Sanctions Reconsidered* by Hufbauer and Schott in 1985, the Peterson Institute database has been one of the most important databases used to analyse the effectiveness of economic sanctions and it has the status of accepted source amongst many researchers of economic coercion.[1] Two major new and extended revised editions appeared of the original study (Hufbauer and Schott 1985) so that we now have three vintages of this dataset reported in three editions of the studies by Hufbauer et al. that appeared in 1985, 1990 and 2007. Over the years the number of co-authors increased with each new edition: Hufbauer and Schott co-authored all three editions; Elliot co-authored the 2nd and 3rd edition and Oegg joined the team for the 3rd edition.

The 1st edition in 1985 contained 103 sanction episodes, providing the first comprehensive and systematic treatment of economic sanctions and offering a unique opportunity to put the sanction theory on an empirical footing. The 2nd edition was presented in 1990 as an update and contained 12 additional episodes and a critical re-evaluation of the other cases. It took 17 years before the 3rd edition was published in 2007. In contrast with the 2nd edition this was much more than an update as it included 65 post-1990 episodes enabling students of sanctions to investigate the more recent history which is relevant as the efficacy and efficiency of the sanction instrument may have been influenced by the sanctions against Saddam Hussein (a major sanction episode), by Détente and by the increasing speed of globalization (including the integration of China in the world economic system). All in all the data set has been extended substantially and moreover 38 old cases were revised, so that almost two thirds of the observations are new or revised. It is highly relevant that the 3rd edition enables an analysis that helps us to understand whether old findings and policy recipes are still valid.

Bergsten (Hufbauer et al., 2007, p. *x*) in the preface to the 3rd edition claims that "despite the dramatic changes in world politics and the world economy the conclusions reached by the authors a quarter century ago remain robust". In this article we show that underlying this apparent stability of both the analysis and the analytical results are shifts in methodology that may seem small at first sight but nevertheless are relevant, significant and create a bias in the 3rd

edition of Hufbauer et al. towards finding sanction success. Relatedly we find that the new methodology overestimates the impact of the costs of sanctions to the target and underestimates the contribution of certain sanction characteristics, including the positive impact of the costs of sanctions to the sender, duration of the sanctions and the sender's companion policies.

Hufbauer et al. (2007) have neither documented nor motivated the changes in methodology, case selection and judgements. For this reason the methodological changes so far have not generally been noticed. Our paper sets a first step at a better understanding of how the rules of measurement changed and how this could impact (future) analyses of economic sanctions and policy prescriptions by taking a detailed look at the apparent changes in methodology, episodes, case selection and evaluation that occured between the 2nd and 3rd (most recent) editions.

One implication of the changes in methodology is that it may complicate research synthesis when researchers combine or compare results of studies that are based on the different vintages of the database (examples of research synthesis include van Bergeijk 1994b; Kaempfer and Lowenberg 2007 and Bapat et al. 2013). By way of illustration Table 1 summarizes 15 empirical studies that have used the Peterson Institute database to investigate whether the extent of trade linkage is related to the success rate of sanctions. The majority of the studies use the 2nd edition, but this will of course change in the near future. The heterogeneity of the results in Table 1 illustrates the need to synthesize knowledge.[2]

<insert table 1 about here>

Many students of economic sanctions recognized the value of the Peterson Institute's sanction database from the start (also because it is the starting point of empirical research on the subject of economic sanctions as a tool of foreign policy), but this of course does not mean that the dataset has been accepted uncritically. Actually it has been criticised by many authors (see Kaempfer and Lowenberg 2007 for a detailed discussion). Prominent amongst these critics is Pape (1997), who has argued against many individual case assessments and is often misquoted for having established that the failure rate of sanctions is more than 95%[3]. Actually, the data set had a critical reception from the start, as for example, Bull (1984, p. 221) in his book review already questioned whether the data are sufficiently comparable to lend themselves to useful generalizations. Other authors have noted several sources of bias, including that failure cases and countries that have a comparative advantage in the application of sanctions could be expected to be over represented in the Hufbauer et al. databases (van Bergeijk 1989; Tsebellis 1990; Smith

3

1995 and Blake and Klemm 2006). Hufbauer et al are very much aware of these problems where they report (2007, 4-5) that sanctions between powers of second and third rank, sanctions for which the documentation was not available in the English language as well as sanctions imposed in comparative secrecy are not covered in their study and also in their frank statement that

> "We recognize that such assessments entail a good deal of subjective evaluation. Indeed, since foreign policy objectives often come in multiple parts, since objectives evolve over time, and since the contribution of sanctions to the policy outcome is often murky, judgment plays an important role in assigning a single number to each element of the success equation" – Hufbauer et al. 2007, p. 49

Given these caveats both in the literature and the Hufbauer et al. publications in particular, the user of the Peterson Institute's database has already been alerted to these weaknesses and therefore our primary interest is neither with individual case assessments nor with bias in the selection of cases. Our first goal is to make fully transparent that the Peterson Institute researchers have – without clear reporting – changed their methodology between the two recent editions of the Peterson Institute Database (Hufbauer et al. 1990 and 2007). Our second goal is to investigate how changes in methodology influence the likelihood of finding specific results, both regarding the success and failure of sanctions and the relevance and impact of sanction attributes. Thirdly, we want to know how the new cases that have been collected since the 1990s change our perspective on the efficacy and effectiveness of economic sanctions as a tool for foreign policy goals. Our research question, in a nutshell, concerns the following issues

- (How) did the methodology change between the 2nd and 3rd edition of the Peterson Institute's report on its sanction data base?
- (How) do these changes influence key findings?
- (How) do recently collected data (i.e. post-1990) differ from pre-1990 observations?

The remainder of this article is structured as follows. Section 2 by way of illustration provides a detailed discussion of four well-established sanction cases in Hufbauer et al. 2007, accompanying CD ROM[4]):

- case 48-5 the US and CoCom (Coordination Committee for Multilateral Export Controls) technology controls regarding the USSR and Comecon;
- case 75-2 the US sanctions against the USSR regarding emigration;
- case 77-2 the US sanctions against Guatemala; and

4

- case 81-2 the US sanctions against Poland 1981-1987.

The four cases are implemented by the US (a country where the researchers of the Peterson Institute have an informational advantage). The cases cover different policy goals: impairment of military potential (48-5), modest changes in target economy policies (75-2) modest changes in target economy policies according to the 2nd edition, but reclassified to regime change in the 3rd edition (77-2), and other major changes in target country policies according to the 2nd edition, but reclassified to regime change in the 3rd edition (81-2). All cases were no longer *de facto* relevant by the end of 1990 when the 2nd edition of Hufbauer et al. was published. At the time of the 2nd edition the four cases were generally considered to have failed. Interestingly, the 'success score' in the 3rd edition improved for all four cases (two cases 77-2 and 81-2 are reclassified as successes). We show that the new evidence presented in the cases would logically preclude such an improvement and in our opinion actually suggests a reduction in the success score. Moving beyond the level of case studies, Section 3 provides a detailed and comprehensive overview of the (general impact of) methodological changes between the 2nd and 3rd editions of the Hufbauer et al. (1990 and 2007, respectively) study. We show the impact of changes in the measurement both of success and of sanction attributes, including the involvement of other countries (regarding international assistance and also in relation to case definition) and the goals of the sanctions. While the important changes that occurred in the methodology of the 2nd and 3rd edition are thus on the table we need to go further and develop an analysis in a multivariate framework both to control for sanction attributes (*cf.* Drury 1998) and in order to illustrate the simultaneous impact of all changes. Section 4 sets out our research strategy that aims at uncovering the impact of the methodological changes using a binary (success/failure) choice model. This statistical model is a standard econometric tool that determines the probability that a sanction succeeds given the case characteristics. This model has been used by e.g. Bonetti (1991) and Hufbauer et al. (2007, pp. 181-192) have used a similar type of approach. In Section 5 we compare results for a common set of 105 sanction cases using data from the 2nd and the 3rd editions and also use this analysis as a bench mark to enrich our understanding of how newly included cases in the 3rd edition provide new information about the success and failure of economic sanctions. Using these findings as a benchmark we can disentangle differences between the statistical properties of 2nd and 3rd editions into differences due to methodological change and differences due to the availability of new observations. In the final section we move beyond methodology and deal with relevance for future research, hypothesis testing using the Hufbauer et al dataset and policy.

2. Case studies

Although the methodology of the Peterson Institute is well known in the field, it is useful to summarize its main characteristics before considering our four case studies.

Hufbauer et al. provide for each sanction case a detailed 'case history' that reports on (i) the chronology of key events, (ii) goals of the sender countries (that is the country or group of countries that imposed or threatened to impose the sanctions), (iii) the response of the target countries, (iv) attitudes of other countries, (v) legal notes, (vi) observed economic statistics, (vii) assessments by other persons than the Hufbauer et al. teams and (viii) the so called Author's' summary. The summary codes relevant attributes based on the case history and provides the 'value' for such items as companion policies, international cooperation with sender and target, type, duration and costs of the sanction. The most important variable throughout the different vintages of the Peterson Institute dataset, however, is the success score of a sanction which is defined (Hufbauer et al. 2007, pp. 49–50) as the product of the assessments of the policy result (the extent to which the policy outcome sought by the sender country was in fact achieved) and of the sanction contribution to this result (the contribution made by sanctions to a positive outcome). Both these assessments are each an integer from the interval 1 to 4 so the success score adopts the values 1 (complete failure), 2, 3, 4, 6, 8, 9, 12 and 16 (complete success).

In the remainder of this section we will compare the case histories as included in the 2nd and 3rd edition for four sanction cases. We selected these cases because the success score in these cases reportedly increased substantially between the 2nd and 3rd edition of the Hufbauer et al. study. These increases occur despite the fact that the original and additional material that the authors provide in their assessment in the 2nd and 3rd edition, respectively, does not support this improvement, suggesting that something else, i.e. methodological change, caused the change in the reported success scores.

The first case history (case 48-5 in Hufbauer et al.) describes the US and CoCom technology controls against the Soviet Union and Cocom that ran from 1948 to 1994. In the second edition the success score is 4 as reflected in a 1987 assessment by the National academy:

> "the recent performance of the U.S. government (…) has not been satisfactorily – and will be increasingly less so because of prevailing trends in international trade and technology diffusion – because it has tended to focus on tightening controls while giving little attention to their effectiveness and costs …. In the absence of appropriate corrective measures, these continuing problems will exact ever-higher tolls – on both Western economic vitality and innovative capacity and on the military security of the United States and its allies." (quoted in Hufbauer et al. 2007, accompanying CD).

Interestingly, in the 3rd edition the success score for case 48-5 becomes 6; a 50% increase. The chronology of key events is extended in the 3rd edition in order to cover the period December 1985 – December 1995. The thirty odd new key events mainly deal with relaxation of the sanctions and the dissolution of both the Soviet Union/Comecon and the CoCom. In the same vein do the five new items regarding the goals of the sending countries and the new response by the target countries deal with the issue of liberalizing export controls. The 2nd edition indicated that other countries helped the target but the 3rd edition no longer considers international assistance to the target to be present. Only a few new economic statistics are added increasing the per capita costs of the sanctions by 41 cents to a still negligible $3.85 per year. None of these new elements or changes suggests that the success score should increase so the only justification could come from the four new assessments added in the 3rd edition. One of the added assessments is the one by the National Academy that we quote above; the other three assessments deal with the less strict application of CoCom rules by allies of the US as well as the anachronistic character of the export controls after the end of the Cold War and their costs to the US economy. So also the assessments that were added in the 3rd edition seem to suggest that the sanctions and their application have been less effective, making the higher success score in the 3rd edition even more puzzling.[5]

The second case history (case 75-2) in Hufbauer et al.) describes the US sanctions against the Soviet Union aimed at increasing Jewish emigration that ran from 1975 to 1994. The success score in the 2nd edition is 4 reflecting the 1989 assessment by Stevenson and Fryre in *Foreign Affairs* that the sanctions were

"… conceptually flawed, and proved counterproductive in operation (and…) gutted the quiet diplomacy [of the Nixon administration] that had begun to yield significant increases in Jewish emigration." (quoted in Hufbauer et al. 2007, accompanying CD ROM)

As in the previous case the success score increased, doubling to 8 in the 3rd edition. Comparing the 2nd and 3rd edition case histories we note 22 new events over the period 1985-1997, three new entrees on sanction goals covering the Reagan, Bush and Clinton administrations and one new item regarding a Soviet counter threat. New statistics are provided that show a significant increase in Soviet emigration starting 1988, but also a 10 per cent lowering of per capita costs of the sanctions to a completely negligible $0.40. Again these new elements or changes do not suggest that the success score should increase. Hence the only justification could come from the

two new assessments added in the 3rd edition. The first assessment is by congressman Donker who in 1986 commented that the sanctions "proved disappointing for those who hoped that the use of MFN would in effect increase Jewish emigration from the Soviet Union" (quoted in Hufbauer et al. 2007, accompanying CD). The second new assessment in the 3rd edition is the one by Stevenson and Fryre quoted above; providing a similar puzzle as to why Hufbauer et al in the 3rd edition increased their assessment of the success score by 100 per cent.

The third case study (77-2 US versus Guatemala) is a protracted sanction (1977-2005) and this is reflected by the 54 new key events added in the 3rd edition. The new key events illustrate that some progress in human rights was indeed made after 1990, the cut-off point of the 2nd edition when the success score was set at a value of 4. However, results are not unambiguously positive as the 2000 *Twelfth Report on Human Rights of the United Nations Verification Mission in Guatemala* reportedly finds that

> "the State's response to past and present violations remains insufficient. Moreover, reports are included that claim that secret funds continued to be provided to the Guatemalan army despite official suspension of military aid. Of particular concern are impunity, intimidation of witnesses, judges, and prosecutors and continued existence of clandestine security units" (quoted by Hufbauer et al., 2007 accompanying CD ROM).

The most important change in the sender policy goals reportedly is Trudeau's 1993 observation that US aid to Guatemala was linked directly to Guatemala's support for the Contras in neighbouring Nicaragua as well as his assessment that "US officials continued to justify the military aid in terms of gaining leverage with the Guatemalan army, even though increasing amounts of support had prevented neither two major coup attempts in 1988 and 1989 nor increasing human rights violations during the Cerezo administration" (Hufbauer et al., 2007 accompanying CD ROM). Indeed also the 2005 US State Department's assessment is that "while government generally respected human rights, serious problems remained" (Hufbauer et al., 2007 accompanying CD ROM). The new economic data include an estimate of the annual per capita cost of the sanctions for the target that reduces from $3.18 over 1977-1986 to $1.15 over 1990-2004. All in all, the case history shows some concrete policy changes (especially towards the end of the period, but hardly provides evidence that the sanctions actually contribute to that result. Still both the scores for policy result and sanction contribution have been increased so that the success score becomes 9 (that is a 125% increase and a success score that is commonly seen to reflect a successful sanction).

8

The fourth case history (case 81-2 in Hufbauer et al. 2007) describes the US sanctions against Poland from 1981 to 1987. In the 2nd edition the success score was 6 reflecting the view that the sanctions had a marginal impact, possibly even stimulating the Polish economy to reintegrate into the Comecon. However, in the 3rd edition the success score increased by 50% to 9 (and therefore is relabelled from failure to success) due to a perceived larger contribution of sanctions (representing a shift from a 'limited contribution' in the 2nd edition to a 'modest contribution' in the 3rd edition). The case history does not provide any evidence in support of the increase in success score. Comparing the 2nd and 3rd edition we find five new key events that – importantly – do not deal with sanction contribution but only with policy results, a reduction of the number of reported sender goals (leaving basically out the promises by the Reagan administration to conditionally <u>end</u> the sanctions) and some new observations on the economic impact (that, however, did not change the estimated economic effect of sanctions *viz.* the 2nd edition). The assessments also do not provide a clue for the reasons to change the success score. The 3rd edition leaves out part of a 1984 *Wall Street Journal* assessment quoted in full in the 2nd edition: "The Reagan administration's decision [to lift some sanctions], although pegged to Polish liberalization, reflects a US acknowledgement that American sanctions haven't significantly altered Polish policies" (Hufbauer et al. 1990, p. 693) and the Authors' summary contains no longer the comment "Polish claims of the adverse impact of sanctions on their economy are undoubtedly exaggerated and, at least in part, an excuse for inept political and economic management in Poland" (Hufbauer et al. 1990, p. 695) possibly reflecting the deletion of earlier assessments by Solidarnosj observers from the case history. All in all, Hufbauer et al. 2007 in this case appear to have made a different selection of assessments which is problematic in view of their general claim that "by drawing on the views of other analysts we believe to have minimized the bias resulting from our personal views" (Hufbauer et al. 2007, p. 49).

The four updated case histories that we discussed in this section do not seem to support but rather contradict the increase in the success score. Our four case studies therefore suggest that the rules that govern the translation from the qualitative case history into a judgemental quantitative assessment by means of the success score have changed between the 2nd and 3rd edition of the Hufbauer et al. study. As is well known from other fields of research, 'intention' in international intervention is not only complex, multi-actor and overlapping, but also changing and eventually often illusive while there are many operationalization challenges in linking an intervention with a specific conflict and its outcome (see, e.g. Regan 1996). In addition, a possible explanation is that due to the changes of the team the requirement of inter-coder consistency was

violated. In the next section we move beyond the individual case studies and take a comparative look at the general characteristics of the Peterson Institute approach in 1990 and 2007.

3. Methodological changes

Table 2 shows how much value the 3rd edition adds: the number of sanction episodes covered by the Peterson Institute increased from 103 in the 2nd edition to 174 in the 3rd edition (a 69% increase). In addition to increase in the number of cases the content of the existing cases covers longer periods as could already be seen from the discussion in Section 2.

At yet another level the increase is even more impressive. Hufbauer et al. split some sanction episodes into separate cases, for example when the sanction goal(s) changed substantially during implementation or to show different impacts if the target comprises of more than one country. The number of cases in the database increased faster (from 108 in the 2nd to 205 in the 3rd editions: a 90% increase). A priori, empirical researchers would prefer to do statistical analysis using case level data rather than episode level data, since a larger number of observations increases the degrees of freedom in statistical analysis. However, the split may not be appropriate because purely bilateral cases (so with one sender and one target country) cannot be split by definition, so that the statistical weight of multi country cases increases due to splitting. Typically bilateral cases have a lower probability to succeed because trade diversion is easier than in the case of multilateral sanctions. In this sense case splitting can *a priori* be expected to increase the share of successful sanctions and if so it will add a positive bias to the analysis. Indeed, as illustrated in Table 3 (bottom rows) the share of successful sanctions in the episodes that have been split in the 3rd edition is well above the average success rate of 33% recorded for the 2nd edition.

<Insert tables 2 and 3 about here>

The 205 cases in the 3rd edition do not cover all 108 cases originally reported in the 2nd edition. Two episodes 65-4 (US versus Arab League) and 73-3 (US versus Chile) have been dropped (both are failures). These cases appear only in the 2nd edition and not in the 3rd edition. Two new pre-1990 cases have been added in the 3rd edition: 71-2 (UK versus Malta in 1971) and 77-2 (US versus Ethiopia in 1977–92); both cases are failures. It is unclear why the two episodes have been dropped, but clearly dropping these cases may induce additional bias, which is relevant also in view of the fact that the number of sanctions is still rather limited.

10

In conclusion, 101 common episodes amounting to 105 common cases in the 1914-1990 period are reported in both 2nd and 3rd editions. The other 100 cases of economic sanctions are completely new and of course can only be analyzed in the 3rd edition. Comparing the 2nd and 3rd edition, we identified that the changes in the methodology occur with respect to the measurement of, firstly, the success of economic sanctions (the dependent variable in most research) and, secondly, the sanction attributes (the explanatory variables in most research), including sanction goals, principal sender(s), target(s) and economic and political variables.

3.1 Changes in the measurement of success

Hufbauer et al. have always used the same overall definition of success. The methodology, however, for assessing the sanction contribution has been modified on two accounts. Firstly, the lowest score (1) was assigned to 'zero or negative sanction contribution' in the 2nd edition, but in the 3rd edition the lowest score is attributed to 'negative' contributions. The consequence is that a zero contribution would score '1' in the 2nd but '2' in the 3rd edition putting upward pressure on the success scores in general. Secondly, the highest score in the 3rd edition is assigned to 'substantial and decisive' sanction contributions whereas the top contribution category was 'modest and significant' contributions in the 2nd edition. Mainly due to these changes in coding we observe for the common 105 cases that the share of the bottom category reduces from 31% to 8% and the share of 'minor contributions' almost doubles to 52% so that the share of extreme cases with a sanction contribution score of 1 and 4 reduces from 45% to 16% respectively.

At the level of the success score, that combines sanction contribution and policy result, we observe that the average for the 105 common cases increases from 6.6 to 7.0 while the standard deviation decreases from 5.1 to 4.5 (the mean difference of the success score for the common cases in the 2nd and 3rd edition is, however, not statistically significant). Figure 1 presents the cumulative frequency distribution of the success score: in comparison to the 2nd edition, the 3rd edition clearly has less outright failure cases (the lines cross at a success score of 8, the commonly used cut off value for failure). From there on the cumulative frequency distribution for the 3rd edition exceeds the one for the 2nd edition illustrating that we will have more observations over this range (that is: more successes).

<insert Figure 1 about here>

3.2 Changes in sanction attributes

Let us start by noting that many changes that can be observed between the 2nd and 3rd editions are not problematic. A clear example of non-problematic change is the increase of the reported duration of sanctions from on average 5.3 years in the 2nd edition to 7.1 years in the 3rd edition. This increase simply reflects that some of the sanctions were still ongoing in 1990 (the cut-off point of the 2nd edition). Since the 3rd edition has the year 2000 as the cut-off point, the true duration of these ongoing sanctions can be reported in the later edition. This change between the 2nd and 3rd edition reflects that the actual duration of a sanction simply cannot be known while a sanction has not yet been formally ended. Other non-problematic changes include the introduction of new political variables that cover cooperating international organizations, the regime of target (democracy/autocracy) and political relations (prior to and during sanction episodes), and the addition of GDP growth and inflation in the 3rd edition. The reclassification of 'health and stability' as an economic instead of a political variable also is not problematic. These changes in sanction metrics are also easy to recognize for all users of the Peterson Institute's data set. For another group of variables the methodology has apparently not or only marginally changed between editions: international cooperation with the sender, costs in per cent of Gross National Product (GNP) and trade linkage (slightly but not significantly lower for the common cases in the 3rd edition) and the ratio of sender GNP to target GNP. The changes that do matter refer to the involved countries (sender, target and international assistance to the target) and to the goals of the sanctions to which we will now turn our attention

Involved countries

In five episodes the principal sender or target are either changed or modified between the 2nd and 3rd editions. An example is episode 82-2 (sanctions against Surinam): the Netherlands and United States are the principal senders in the 2nd edition, but only the Netherlands is a sender country in the 3rd edition. Other cases are: 39-1(Alliance Powers versus 'Alliance Powers and US' as senders), 56-3 (UK and France versus UK as targets), 61-3 (Western Allies versus US and Western Allies as senders), and 76-3 (Ethiopia versus Arab League as target).

Also regarding international assistance to the target shifts occur: the 2nd edition identifies international assistance to the target in 26% of the 105 common cases, but in the 3rd edition, this reduces to 22%. Changes with respect to prior relations between target and sender occur in 60-3 (US versus Cuba), 75-3 (US versus Eastern Europe), 75-4 (US versus South Africa), 75-5 (US

versus Kampuchea) and 76-3 (US versus Ethiopia).

Goals of the sanctions

The Peterson Institute identifies five major types of foreign policy objectives for the sanctions episodes. The foreign policy goal categories are

- modest changes in target-country policies,

- 'destabilization of target government' (2nd edition) or 'regime change' (3rd edition)

- disruption of military interventions other than major wars,

- impairment of military potential including major wars and

- other major changes in the target country.

In the 3rd edition the broader term 'regime change' is used. Hufbauer et al (2007, p. 13, footnote 14) explain that they use this term "to encompass not only the explicit targeting of a particular foreign leader but also structural changes that imply new leadership, such as broad human rights demands, and, in the 1990s, the embrace of democratic reform." Major changes have occurred with respect to the attribution of foreign policy goals to sanction episodes. Figure 2 provides an overview. In the 2nd edition, 41% of the common cases are considered to aim at modest policy changes in the target country whereas it is only 21% in the 3rd edition and where the most frequent sanction goal in the 2nd edition was a 'modest' goal, the mode in the 3rd edition is 'regime change'.[6] These are very large changes that are difficult to interpret the more so since cases with multiple goals were in general identified as being aimed at destabilization in the 2nd edition:

> "When another goal underlies or accompanies a destabilization case, we have generally listed these cases only in the destabilization group. (After all, a destabilization attempt presupposes policy disputes as well as personality differences.) We make an exception, however, when the sender country seeks to destabilize a government and to disrupt a military adventure; such cases are cross listed under both headings." (Hufbauer et al 1990, p. 43)

So on the one hand the 2nd edition favored classification as destabilization for sanctions with multiple targets; ending that procedure could be expected to reduce the share of destabilization cases.[7] On the other hand, the broadening of this goal to 'regime change' would increase that share. Given the observed shifts in Figure 2 the latter effect by far dominates the former with one important but unnoticed consequence: the content or meaning of the rest category modest policy goals by implication also changed between the 2nd and 3rd edition.

13

In addition to this shift, Hufbauer et al. (2007) added new goals in the 3rd edition and/or modified and changed some of the stated goals in the 2nd edition. For example, two goals are reported in case of 76-3 (US versus Ethiopia) in the 2nd edition: namely 'to settle expropriation claims and to improve human rights'. In the 3rd edition the goal becomes 'anti-boycott restrictions on US firms'. So for this particular case two goals are dropped and one new goal is added in the 3rd edition. In the 2nd edition, a total of 146 goals were set for 101 episodes. The number of goals is raised to 155 in the 3rd edition for the same 101 episodes (a 6% increase). The most important change is, however, not in the number of goals, but, as said, in the classification of goals.

<insert Figure 2 about here>

The conclusion is that many changes in methodology bias the results either in general (the definition of sanction contribution and the splitting of cases) or in specific cases (the new label regime change and the end of the rule to classify multiple goal sanctions as destabilization cases).

4. Research strategy

In the previous section we have shown the presence of various biases using bivariate analyses. In this section we will show that multivariate; simultaneous interactions effects extend these finding. Thereby we can see how, for example, the success core was related to a modest policy goal in the 2nd edition controlling for economic and political sanction attributes, and if and how that relation changes in the 3rd edition. Indeed, the reader will want to know how all these changes that occured simultaneously in the latest edition impact on research findings.

In order to investigate the potential impact of the methodological changes and the newly collected data we constructed 3 data sets. The first data set comprises of the data of the 105 common cases published in Hufbauer et al. (1990); as discussed this is the data set of the 2nd edition from which two cases have not been included because they are dropped in the 3rd edition. The second data set consists of the same 105 cases but uses the data reported in the 3rd edition (here we exclude the 2 pre-1990 case histories that were added in the 3rd edition because they are not reported in the 2nd edition). The third data set consists of all data from the 3rd edition. The comparison between the first and the second data set reveals the impact of changes in the methodology; the comparison between the second and the third data set reveals the impact of the newly added observations. We will use these three versions of the Peterson Institute dataset to

14

estimate the marginal effects in a probit model that estimates the probability of success. We define success in line with the empirical literature as a success score of 9 and higher. In our econometric investigation the binary variable y_i serves as the dependent variable.

$y_i = 1$, if the i-th sanction is a success (the Hufbauer et al. success score is 9, 12 or 16)

$y_i = 0$, if the i-th sanction is a failure (the Hufbauer et al. success score is 1, 2, 3, 4, 6 or 8)

The relation between the success score and the sanction attributes will be estimated with probit. Probit-analysis makes it possible to calculate the probability π that a specified sanction case ends successfully. If this probability exceeds 0.5, a success is 'predicted'; if not, a failure. In Section 5 we report the marginal impact for each sanction attribute.[8] If this impact is positive an increase (decrease) in the value of the sanction attribute implies that the probability that the sanction succeeds becomes higher (and for a negative number the implication is that it is less likely that a sanction succeeds). In practical terms we proceed as follows. Noting that Hufbauer et al. distinguish five foreign policy goals, we include binary (yes/no or 1/0) dummy variables for four of these goals (*modest policy change*, *destabilization/regime change*, *disruption of military adventures* and *military impairment*) in the probit regression analysis. By implication we report deviations from the reference case (in this formulation the reference case is a sanction that aims at 'other major changes in the target country'). A positive estimated coefficient for a policy goal dummy variable implies that the sanction is more likely to succeed for the respective policy goal and a negative coefficient implies that this is less likely. In our analysis the variable *companion policy* takes the value 1 if regular military action is considered and 0 otherwise. Likewise *international cooperation* is a dummy variable that takes values from 1 (no cooperation with the sender) to 4 (significant cooperation with the sender). If the target country receives any *international assistance*, it takes the value 1 and 0 otherwise. The variable *prior relations* takes values from 1 (antagonistic) to 3 (cordial). *Health and stability* is a dummy variable taking 1 for distressed and significant problem in the target economy and 0 otherwise. The reference case is a hypothetical sanction where all dummies are zero so that it describes a sanction that aims at 'other major changes in the target country' with no cooperation with the sender by other countries or an international organization or international support for the target and that is not distressed before the sanctions are implemented.

In this paper we do not want to test a particular theory of sanctions, as the aim of our paper is to investigate how changes in methodology and inclusion of new data may impact on empirical

15

findings (but note that we will discuss some research and policy implications of the findings in the final section). This does not mean that we think that theory does not matter. It does and any empirical research should start from a clear concept of what variables should be measured (and how). Our contribution, however, is not limited to one specific theoretical approach. We investigate for the variables in the Hufbauer et al. study whether methodological change matters. Researchers of economic sanctions can use our findings to decide whether the variables that are relevant from their theoretical perspective can be taken from Hufbauer et al. or need to be collected from other sources or, alternatively, whether a test of an hypothesis by a colleague that is based on the Hufbauer et al. 2007 data set is reliable and informative. Rather than limiting ourselves to a subset of the variables that we as economists would prefer to include in tests of our economic hypotheses we thus opt for what could be labelled a theory-free approach. This allows us to evaluate as many of the common cases as possible (so we optimize with respect to data availability in the 2nd and 3rd edition and do not include variables that appear in either the 2nd or the 3rd edition only). We also do not make corrections when variables from an economic-theoretical perspective are incorrectly specified in the Hufbauer et al. studies (an example is trade linkage that for obvious reasons should be measured in the year before the economic sanctions are implemented, but is measured in the year of the sanctions). Again our motivation is that we want to show the influence of the changes in methodology between the two editions for as a broad class of theories as possible.

In line with the reporting done by Hufbauer et al., we provide three specifications for the 'success equation' that differ with respect to the included (sets) of explanatory variables.

- Specification (1) includes only the foreign policy goals as explanatory variables. This model allows us to focus on the interaction between the changes in the definition of the success score (the dependent variable) and the newly defined policy goals (sanction attributes or explanatory variables). One important merit of this small model is that the data requirements are always met so that we use all 105 common cases in our analysis. However, we do not control for changes that occur in other sanction characteristics.

- Specification (2) therefore extends the first specification by including the so-called political variables. So we can now simultaneously analyse three issues: the redefinition of the success score, the new definition and assignment of policy goals and the political sanction attributes (companion policies, international cooperation/assistance, prior relations and duration).[9] However, while we include an important set of explanatory variables we also leave out an important set of attributes, namely the economic variables. The reason for doing so is that not all cases report (all) economic statistics. We miss 27

cases in the 2nd edition and 13 cases in the 3rd edition (coverage improved between the 2nd and 3rd edition) if we include the economic variables and therefore we follow a two-step procedure including first the political variables and later the economic variables.

- Specification (3), accordingly, as before includes the foreign policy goals and the political variables but adds economic variables (cost to target/sender, ratio of sender GNP to target GNP, absolute trade linkage and health and stability). We need to be cautious when we compare this model between the 2nd and 3rd edition because 14 common cases cannot be analysed due to the data deficiencies.

5. Empirical Results

Table 4 presents the results of applying the binary probit model. In this multivariate framework we can now show the impact of methodological change from a more comprehensive point of view. We start by comparing the marginal effects reported for the common cases in 2nd edition (columns 1-3) and the 3rd edition (columns 4-6). Column 1 and 4 provide the results for specification (1) for the 2nd and 3rd edition, respectively. Likewise specification (2) is reported in columns 2 and 4 and specification (3), that is the model with all sanctions characteristics, is reported in columns 3 and 6.

<center><insert Table 4 about here></center>

An intuitive interpretation of the results in columns 4-6 is as the reporting of results of the 2nd edition corrected for changes in methodology and case selection in the 3rd edition. For example, if we look at the findings for *disruption of military interventions* then we see consistently that the coefficient is positive and insignificant in all specifications (columns) for both the 2nd and 3rd edition. Here the methodological changes do not matter. Taking a look at *companion policies*, we see that the coefficients in the common cases are positive and significantly different from zero for the 2nd edition, but not for the 3rd edition. Since we use exactly the same cases, this difference can be attributed directly to changes in methodology. Indeed, whereas the 2nd edition finds that the companion policies increase the likelihood that a sanction succeeds, the 3rd edition for the common cases does refute that idea and therefore we can conclude that the change in methodology creates a bias against finding that companion policies are effective means to enhance sanction success. Now let us in the same manner take a look at the findings for each set of sanction characteristics: policy goals, political variables and economic variables

For *modest policy change* and *destabilization/regime change* we find consistently (so independent of specification) that the marginal effects are positive and significant implying that sanctions that

aim at these goals are more likely to be identified as successful. This result is the same for the 2ⁿᵈ and 3ʳᵈ edition. One difference, however is that the effect for destabilization/regime change is both larger and more significant in the 2ⁿᵈ edition (for example in column 3 the coefficient is 0.80 at the 99% confidence level while it is only 0.37 at the 90% confidence level in column 6). The implication is that the 3ʳᵈ edition's methodology in comparison to the 2ⁿᵈ edition's methodology is biased in favour of finding positive results for regime changes. As said before, the methodology has no bias for disruption of military interventions. For military impairment we see a difference in specification 3 as the coefficient is significant at the 95% for the 2ⁿᵈ edition data and insignificant in the 3ʳᵈ edition. The interpretation of this difference is complicated, however, due to the differing numbers of available observations with complete economic data.

Turning to the political variables we observe no difference for international cooperation and assistance so that methodological changes do not play havoc here. In contrast size and significance are consistently larger in the 2ⁿᵈ edition for companion policy and duration in years, reflecting the impact of methodological change and reclassification of protracted sanctions as a success (case 77-2 discussed in Section 2 is an example). For prior relations the evidence for any impact of methodological chance is less convincing in view of the fact that the confidence level for the coefficient in column 6 is only 90%.

The findings for *ratio of sender GNP to target GNP*, *trade linkage* and *health and stability* suggest no statistical impact of methodological change. The significant coefficient (95% confidence level) for the costs to the sender for 2ⁿᵈ edition data and methodology turns insignificant for the 3ʳᵈ edition suggesting that the new methodology is biased against finding this impact. Our probit analysis for the common cases in the 3rd edition assigns only marginal (90% confidence) significance to the cost to the target (column 6) so here our statistical evidence is not completely convincing.

<insert table 5 about here>

Now that we have identified bias induced by methodological change between the 2ⁿᵈ and 3ʳᵈ edition, the next step is to interpret the findings for the full sample of the 3ʳᵈ edition. Table 5 for ease of comparison repeats column 5 and 6 of Table 4 and summarizes the bias that we inferred from that table. Next the columns with label 7 and 8 report on specification 2 and 3 but then for all cases in the 3ʳᵈ edition, not just the common cases. This allows us not only to uncover how adding new cases changes the overall results, but also to see if these outcomes matter in view of the established bias.

We analyse the impact of the new observations by comparing the findings for the common cases reported in the 3ʳᵈ edition (columns with 5 and6 in Table 5) with the findings for

the full sample of the 3rd edition (columns labelled 7-8 in Table 5) for the two extended specifications. Focussing on the policy goals we start with *modest policy goals* and *disruption of military interventions,* because as reported in the column headed 'bias', we established in the probit analysis that the differences in methodology did not generate bias for these policy goals. Regarding *modest policy goals* we can see that whereas these were significant in the smaller sample of the 'old' common cases, these are no longer so in the full sample. Hence the new cases differ importantly (the lack of insignificance is even more telling form a statistical viewpoint because the number of observations is almost twice as large). For disruption of military interventions we see that the new cases suggest that this policy goal is more difficult to achieve given the negative coefficient (column 8) that is significant at a 90% confidence level. Then we turn to the policy goals where bias has been found. As to *military impairment* we continue to find insignificant coefficients, but we note that this finding may be biased due to methodological change. Finally for *regime change* we find that the new cases lead to a loss of significance. In conclusion then the findings demonstrate that differentiation by sanction goal adds less value for the whole sample then for the cases that were presented in the 2nd edition. Sanctions seem to work equally good or bad independently of the sanction goal with the exception of disruption of military interventions were it is more difficult to achieve good results.

Turning to the political sanction attributes we note that for the unbiased variables we find no different results. International cooperation and hindrance continue to be insignificant and the success enhancing impact of good *prior relations* is established at the 95% level and better. For *companion policies* we find no change, but note that this finding may be biased due to methodological change. For *duration* the new cases change the picture although the coefficients are significant only with a confidence level of 90%. However, our earlier finding of upward bias provides additional credibility of duration's impact: due to the methodological changes it has become more difficult to find a negative impact and the new methodological setting may therefore form at least part of the explanation why we find a highly significant negative impact for duration in the 2nd edition at the 99% confidence level and better and only at 90% in the full sample of the 3rd edition. As noted earlier, however, this also reflects more accurate information on duration and is not only a consequence of the change in methodology so that this should also be at least partially attributed to new information (both regarding new cases and regarding cases that were on-going in 1990).

Finally, regarding the economic sanction attributes we find no difference for trade linkage and health and stability (that were not corrupted by methodological change and continue to be insignificant) and for cost to the sender but note that the latter finding may be biased due to

19

methodological change. The other economic attributes experience changes in the level of significance but the impact on the probability that a sanction succeeds is very limited in view of their smallness. However the insignificance of the cost to the target is worth mentioning because it is established despite an upward bias due to methodological change and because the number of observations increases (on both counts one expects *a priori* a rise in significance so that the decrease is even more convincing).

6. (Why) does bias matter (and when)?

It is bad scientific practice to change methodology and not to report (the rationale for) the changes. Taking out cases and reclassifications of goals without further motivation entails the risk of data massaging. Changing your mind and arriving at new judgements is of course good scientific practice, but then the reasoning and weighting of evidence needs to be completely transparent. Our article pointed out the lack of transparency, reconstructed the changes and assessed the implications. But at the end of the day the question of course is "does it matter?" This is a relevant question because the cases in the 3rd edition have all been evaluated with the same methodology and therefore would seem to provide a framework than can be used to access success and fail factors as well as general trends of the efficacy of sanctions. Our analysis, however, has clarified that methodological change matters at all levels of analysis and thus can compromise research and policy advise both regarding case selection and the policy mix, including sanction attributes such as companion policies and trade coverage.

At the level of individual case studies (Section 2) we uncovered a tendency to inflate success scores, reclassifying failures into successes even when the evidence for doing so was not convincing. In a number of cases the case histories leave out information that was contradicting either policy result or sanction contribution. These changes in case history and general assessment will have an impact, because more success undoubtedly makes the sanction instrument more attractive to policy-makers (whether are interested in specific target countries or in the use of sanctions in general). Importantly, we have established these changes for cases where the US is the sender country. This is important for two reasons: the US is the most important sender of sanctions and the authors have an information advantage for the US. It is not unreasonable to expect that problems that we uncover in US case studies will at least be as relevant in cases involving second- and third-rank powers as a sender.

At the level of the aggregated case studies and general methodology (Section 3) we uncovered positive bias (so methodological changes that make it more likely to find sanction success as indicated by a higher success score, either on average or in individual cases). Splitting

of episodes into cases and the changed definition of sanction contribution increases the success ratio in general and ultimately the share of sanctions that are judged to be a success. We also show the importance of the reclassification of destabilization cases into regime change and of the end of the coding practice of lumping together multiple goal sanctions under that heading. An important, yet unnoticed, implication is that the definition and content of 'modest policy changes' (since it is a rest category) must have changed between the 2nd and 3rd edition.

Moving from a bivariate analysis to a multivariate analysis (Section 5) enabled us to take a closer look at sanction attributes: we were able to establish upward bias related to duration (an underestimate of the negative impact of duration) and cost to target (an overestimate of the positive impact of higher costs to the target) and downward bias for regime change, military impairment, companion policies and cost to the sender. These biases are important from a policy perspective because they influence the *a priori* assessment of cases where sanctions are contemplated and from a theoretical point of view whenever a sanction attribute is used to test a theory. An example is the negative bias that we find for cost to sender that implies that it is more easy to reject the hypothesis that domestic opposition induced by cost considerations may undercut the strictness of the sanctions.

We have scrutinized the latest study of Hufbauer, Schott, Elliott and Oegg especially in view of the fact that basically unreported methodological changes tend to bias one of the major data sources for applied sanction research. It should be noted, however, that a number of their finding and prescriptions survived our torment and by doing so testify of the robustness of these findings. Our results support the lower chance for success for disruption of military interventions and the importance of prior relations: friends are indeed more likely to comply. The policy recipe to 'slam the hammer instead of turning the screw' comes out even stronger since now we know that the methodological changes underestimate the negative impact of slow and piecemeal implementation. Bias *per se* is not bad, but we need to know its size and direction.

References

Bapat , N.A., T. Heinrich , Y. Kobayashi & T. C. Morgan, 2013, Determinants of Sanctions Effectiveness: Sensitivity Analysis Using New Data, *International Interactions: Empirical and Theoretical Research in International Relations*, 39:1, 79-98, DOI:10.1080/03050629.2013.751298

Bergeijk, P.A.G. van, 1989, Success and Failure of Economic Sanctions, *Kyklos* **42** (3), pp. 385-404. DOI: 10.1111/j.1467-6435.1989.tb00200.x

Bergeijk, P.A.G. van 1994a, *Economic Diplomacy, Trade and Commercial Policy; Positive and Negative Sanctions in a New World Order*, Edward Elgar: Cheltenham.

Bergeijk, P.A.G. van, 1994b, The Effectivity of Economic Sanctions: Illusion or Reality, *Peace Science, Peace Economics and Public Policy* (1994) 2 (1), pp. 24-35

Bergeijk, P. A. G. van, 2009, *Economic Diplomacy and the Geography of International Trade*, Edward Elgar, Cheltenham, UK.

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. DOI:10.1016/0264-9993(94)00003-N

Blake, C.H. and N. Klemm, 2006, Reconsidering the Effectiveness of International Economic Sanctions: An Examination of Selection Bias, *International Politics*, **43**, pp. 133–149. DOI:10.1057/palgrave.ip.8800137

Bonetti, S., 1998, Distinguishing characteristics of degrees of success and failure in economic sanctions episodes, *Applied Economics* 30.6 (1998): 805-813. DOI:10.1080/000368498325507

Bull, H.:, 1984, Economic Sanctions and Foreign Policy, *The World Economy*, **7** pp. 218-222. DOI: 10.1111/j.1467-9701.1984.tb00271.x

Dizaji, S. F, Bergeijk, P.A.G van., 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. 721-736. DOI: 10.1177/0022343313485487

Drury, A. C., 1998, Revisiting Economic Sanctions Reconsidered, *Journal of Peace Research,* Vol. **35**(4): pp. 497-509. DOI: 10.1177/0022343398035004006

Elliott, K.A., and P. P. Uimonen, 1993, The effectiveness of economic sanctions with application to the case of Iraq," *Japan and the World Economy* **5** (4): pp. 403-409. DOI:10.1016/0922-1425(93)90007-Q

Hart, R. A., 2000, Democracy and the successful use of economic sanctions, *Political Research Quarterly* **53** (2), pp. 267-284. DOI: 10.1177/106591290005300203

Hufbauer, G. C. and J. J. Schott (1985) 'Economic Sanctions Reconsidered: History and Current Policy', Institute for International Economics, Washington, D.C.

Hufbauer, G.C., J.J. Schott and K.A. Elliott (1990), *Economic Sanctions Reconsidered*, 2nd edition, Peterson Institute for International Economics: Washington, DC.

Hufbauer, G. C., J. J. Schott, K. A. Elliott and B. Oegg (2007) 'Economic Sanctions Reconsidered', Peterson Institute for International Economics (3rd eds.), Washington, D.C.

Kaempfer, W. H. and A. D. Lowenberg, 2007, The political economy of economic sanctions, *Handbook of Defense Economics* 2, pp. 867-911. DOI:10.1016/S1574-0013(06)02027-8

Kim, H.M., 2013, Determining the Success of Economic Sanctions, *Australian Journal of Political Science*, **48**(1), pp. 85-100 DOI:10.1080/10361146.2012.731488

McLean E. V. and T. Whang, 2010 Friends or Foes? Major Trading Partners and the Success of Economic Sanctions *International Studies Quarterly* **54**, (2), pp. 427–447. DOI: 10.1111/j.1468-2478.2010.00594.x

Miyagawa, M. (1992) *Do Economic Sanctions Work?* New York: St. Martin's Press.

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):92–110. DOI: 10.1177/0738894208097668

Nooruddin, I., 2002, Modeling selection bias in studies of sanctions efficacy, *International Interactions* 28.1 (2002), pp. 59-75. DOI:10.1080/03050620210394

Pape, R. 1997, 'Why Economic Sanctions Do Not Work', *International Security*, **22**, (2), 1997, pp. 90–136. http://www.mitpressjournals.org/doi/pdf/10.1162/isec.22.2.90

Regan, P. M., 1996, Conditions of successful third-party intervention in intrastate conflicts. *Journal of Conflict Resolution*, 40(2), pp. 336-359 DOI: 10.1177/0022002796040002006

Smith, A., 1995, The Success and Use of Economic Sanctions' *International Interactions*, **21** (3), pp 249–45. DOI:10.1080/03050629508434867

Tsebellis, G., 1990, 'Are Sanctions Effective: A Game Theoretic Analysis', *Journal of Conflict Resolution*, **34** (1), pp. 3–28. doi: 10.1177/0022002790034001001

Whang, T., 2010, Structural estimation of economic sanctions: From initiation to outcomes, *Journal of Peace Research* **47** (5), pp. 561-573 DOI: 10.1177/0022343310376868

Table 1

Summary statistics for impact of trade parameter in studies on economic sanctions that use Hufbauer et al.

Study	N	Period	Trade parameter	Estimate
Investigations based on the 1985 data set				
Hufbauer et al. (1985)	108	1914–1983	ATL	0.03
Van Bergeijk (1989)	80	1946–1983	PTL	0.21[$]
Lam (1990)	98	1914–1983	ATL	0.01
Investigations based on the 1990 data set				
Dehejia and Wood (1992)	115	1914–1989	ATL	0.00
Elliott and Uimonen (1993)	92	1914-1989	ATL	0.02[$]
Van Bergeijk (1994a)	103	1946–1989	ATL	0.04[#]
	92	1946–1989	PTL	0.15[$]
Bonetti (1998)	60	1946–1989	ATL	0.10-0.16[$]
			ATL^2	–0.00
	44		ATL	0.04
Drury (1998)	115	1914–1989	ATL	0.00-0.01
Hart (2000)	81	1914–1989	ATL	0.04[#]
Nooruddin (2002)	71	1946–1989	ATL	–0.01
	205	1946–1989	ATL	0.25
McLeany and Whang (2010)	106	1948-1992*	ATL	0.59[$]-0.85[$]
Kim (2013)	116	1950-1990	ATL	0.02[$]-0.69
Investigations based on the 2007 data set				
Hufbauer et al. (2007)	174	1914–1999	ATL	0.01
Van Bergeijk (2009)	166	1946–1999	ATL	0.01-0.02[#]
	161	1946–1999	PTL	0.04[#]
Whang (2010)	183	1903-2002	PTL	–0.65-0.94

Notes: $ significant at 99%,
\# significant at 95%,
* McLean and Whang (2010, p.434, footnote 6) use additional data sources to identify data and restrict their sample to cases in which the sender is a member of UN Security Council and/or G8
ATL is absolute trade linkage (bilateral trade in per cent of target's total trade)
PTL is proportional trade linkage (bilateral trade to target's GDP)

Table 2

Economic sanctions considered in the second and third editions

Edition	No of episodes	Cases (including split)	Coverage
Second	103	108	1914-1984
Third	174	205	1914-2000
Increase (%)	**69%**	**90%**	

Table 3

Details on changes in identified targets

Case ID in the 2nd edition	Success score in second edition	Case ID in the 3rd edition	Sender	Target in 2nd edition	Target in 3rd edition	Success score in third edition
32-1	6	32-1(1)	League of Nations	Paraguay and Bolivia	Paraguay	6
		32-1(2)			Bolivia	6
39-1	12	39-1(1)	Alliance powers	Germany, later Japan	Germany	8
		39-1(2)			Japan	8
		39-1(3)			Germany	12
		39-1(4)			Japan	12
48-3	2	48-3(1)	USSR	USA, UK. France	USA	2
		48-3(2)			UK	2
		48-3(3)			France	2
71-1	4	71-1 (1)	USA	India and Pakistan	Pakistan	4
		71-1 (2)			India	4
73-1	9	73-1(1)	Arab League	USA, Netherlands	USA	9
		73-1(2)			Netherlands	9
77-4	9	77-4 (1)	Canada	Japan, EC	EC	9
		77-4 (2)			Japan	9
P.M. Average	7.0					
P.M. Success rate	50%					

Table 4 Analysis of common cases: Probit regression results for the probability of 'success' of economic sanctions –

Variables	Common cases using 2nd edition data and definitions			Common cases using 3rd edition data and definitions		
	(1)	(2)	(3)	(4)	(5)	(6)
Policy goals						
Modest policy change	0.25*	0.32**	0.65***	0.36**	0.40**	0.55***
Destablilization/Regime change	0.37**	0.45**	0.80***	0.27*	0.35**	0.37*
Disruption of military interventions	0.13	0.12	0.37	0.16	0.23	0.16
Military impairment	0.02	0.34	0.77***	0.08	0.17	0.25
Political variables						
Companion policies (=1 if regular military action)		0.34*	0.50**		0.11	0.22
International cooperation (1 to 4)		–0.01	–0.14		–0.01	–0.04
International assistance (=1 if assistance to target)		–0.04	–0.12		–0.02	–0.01
Prior relation (1 to 3)		0.16**	0.15		0.18**	0.17*
Duration in years		– 0.03***	–0.03***		–0.01	–0.01
Economic variables						
Cost to target (in million $)			0.00			0.00*
Ratio of sender GNP to target GNP			0.00			0.00
Trade linkage (% of total trade)			0.00			0.00
Health and stability (=1 if distress or significant problems)			–0.06			0.08
Cost to sender (per cent of GDP)			0.05**			–0.02
Observations	**105**	**105**	**78**	**105**	**105**	**92**
Prob>chi2	**0.18**	**0.02**	**0.00**	**0.17**	**0.06**	**0.02**

***p<0.01, **p<0.05 and *p<0.1

Table 5 Analysis of old and new cases Probit regression results for the probability of 'success' of economic sanctions

| Variables	Common cases using 3rd edition data and definitions			Full sample (3rd edition data and definitions)	
	(5)	(6)	Bias	(7)	(8)
Policy goals					
Modest policy change	0.40**	0.55***	0	0.18	0.17
Regime change and democratization)	0.35**	0.37*	–	–0.01	–0.07
Disruption of military interventions	0.23	0.16	0	–0.112	–0.19*
Military impairment	0.17	0.25	(–)	0.08	–0.02
Political variables					
Companion policies (=1 if regular military action)	0.11	0.22	–	0.04	0.13
International cooperation (1 to 4)	–0.01	–0.04	0	0.03	0.03
International assistance (=1 if assistance to target)	–0.02	–0.01	0	-0.11	–0.07
Prior relation (1 to 3)	0.18**	0.17*	0	0.13***	0.12**
Duration in years	–0.01	–0.01	+	–0.01*	–0.01*
Economic variables					
Cost to target (in million $)		0.00*	(+)		0.00
Ratio of sender GNP to target GNP		0.00	0		–0.00*
Trade linkage (% of total trade)		0.00	0		0.00
Health and stability (=1 if distress or significant problems)		0.08	0		0.11
Cost to sender (per cent of GDP)		–0.02	–		–0.01
Observations	105	92	92	204	176
Prob>chi2	0.06	0.02	0.02	0.03	0.04

- downward bias

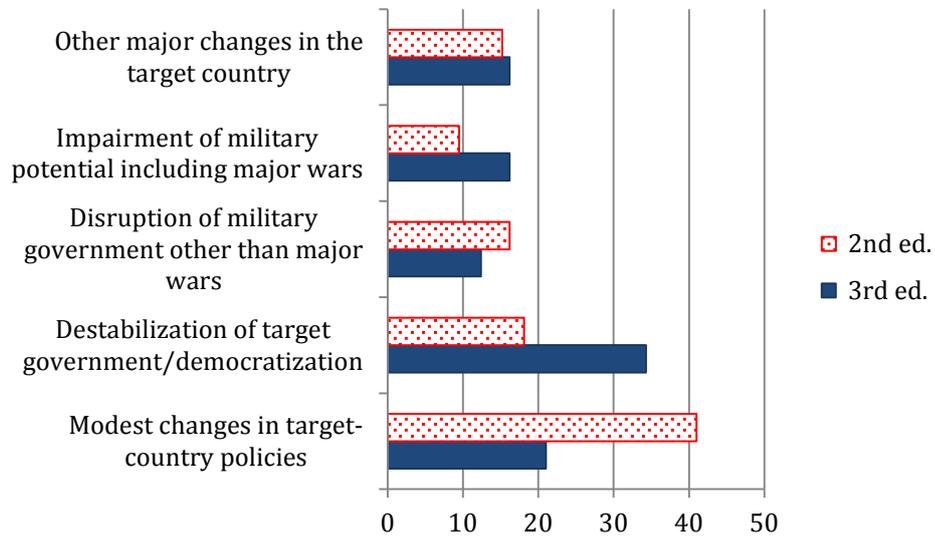0 no bias or insignificant

+ upward bias

(in brackets: bias established at 90% confidence only)

Figure 1
Cumulative frequency distributions of success score
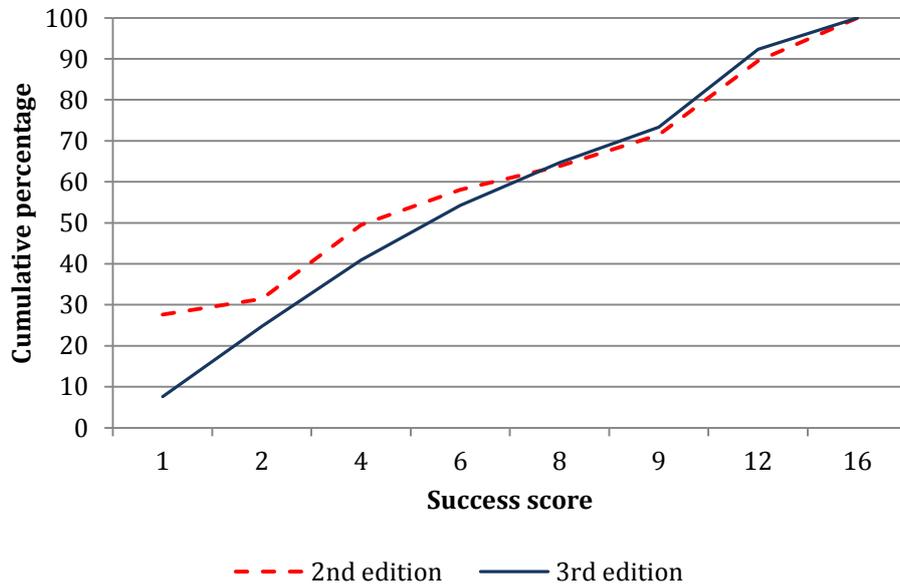
Note: No success score of 3 exists for the common cases.

Figure 2 Foreign Policy Goals (common cases, in per cent)

Endnotes

[1]   Recently researchers have constructed useful alternative data sets such as the Threat and Imposition of Economic Sanctions data set (Morgan et al 2009). The Peterson Institute database, however, is still one of the major empirical resources dealing with the occurrence and effectiveness of international economic sanctions as a tool of foreign policy.

[2]   Authors have adjusted and upgraded the Hufbauer et al. database (applying judgement both regarding the reliability of data in specific sanction cases, the need to consider different sets of explanatory variables (or, predictors) and the unit of observation, i.e., episodes, cases, years etc.

[3]   Pape (1997) proposes that only 5 out of 115 cases meet his criteria of successful sanctions. However, this does not imply that the success rate is less than five per cent, since the success rate is a ratio and Pape's arguments relate to both the numerator and the denominator. Using Pape's definitions on what sanctions should not be included reduces the population of relevant sanctions to 44 so that the success rate is at least 11% if one completely accepts Pape's verdicts on successes and failures of the outcomes of the sanctions.

[4]   We use the case identifiers in the Hufbauer et al. data set. The first number refers to the year in which the sanctions started.

[5]   The more so since Hufabuer et al (2007, p. 11) remark "It is doubtful whether these cases yielded positive results, not least because it is difficult to hamper the military capabilities of a major power by inflicting marginal degrees of economic deprivation".

[6]   Interestingly, also the judgement changes with regard to 'disruption of military government other than major wars' versus 'impairment of military potential including major wars'. All in all a very substantial change occurs in the distribution of the foreign policy goals of the common cases.

Hufbauer et al. (2007, p. 2, footnote 3) clarifies that such cases in the 3rd
edition have separate entries increasing the total number of cases so that the
share ceteris paribus decreases.

8

Here we have described the probability that $y_i = 1,$ or response probability as

$$\Pr(y_i = 1/x_i) = G(\beta' x_i)$$

Where $G(\beta' x_i)$ is a function with values on interval [0, 1] for all real numbers $\beta' x_i.$ The standard
normal distribution yields the probit model is

$$G(\beta' x_i) = \Phi(\beta' x_i) = \int_{-\infty}^{\beta' x_i} \frac{1}{\sqrt{2\Pi}} \exp\{-\frac{1}{2}t^2\}dt -----------(1)$$

From the probit model, marginal effects of Δx_i is calculated as

$$\frac{\partial \Phi(\beta' x_i)}{\partial x_{i,k}} = \phi(\beta' x_i)\beta_k.$$ Therefore, the effect of $\Delta x_{i,k}$ on y_i depends on x_i and $\beta_k.$

9 Duration is actually not a political variable. Duration is related to economics
as it determines the extent of substitution between traded and non traded
goods and services and thereby influences the annual costs for the target
(Dizaji and van Bergeijk, 2013).

Please note:

You are most sincerely encouraged to participate in the open assessment of this discussion paper. You can do so by either recommending the paper or by posting your comments.

Please go to:

http://www.economics-ejournal.org/economics/discussionpapers/2015-33

The Editor